Long-Time Asymptotics of the Sliced-Wasserstein Flow

Giacomo Cozzi*, Filippo Santambrogio†

April 22, 2024

Abstract

The sliced-Wasserstein flow is an evolution equation where a probability density evolves in time, advected by a velocity field computed as the average among directions in the unit sphere of the optimal transport displacements from its 1D projections to the projections of a fixed target measure. This flow happens to be the gradient flow in the usual Wasserstein space of the squared sliced-Wasserstein distance to the target. We consider the question whether in long-time the flow converges to the target (providing a positive result when the target is Gaussian) and the question of the long-time limit of the flow map obtained by following the trajectories of each particle. We prove that this limit is in general not the optimal transport map from the starting measure to the target. Both questions come from the folklore about sliced-Wasserstein and had never been properly treated.

1 Introduction

Solving problems related to optimal transport, such as computing optimal transport maps and Wasserstein distances, is in general very expensive from the numerical point of view, when we consider measures on $\mathbb{R}^d$ with $d \geq 2$. The one-dimensional case, on the contrary, can be treated easily: explicit formulas for optimal transport maps, for example, are given by the cumulative distribution functions of the starting and target densities and their pseudo-inverses; also, the usually highly non linear Monge-Ampére equation, becomes an ODE in this case. Any kind of transport construction where the multi-dimensional computation of optimal transport maps is replaced by (many) one-dimensional computations are therefore of extreme interest. Among the first idea of this type we mention, for instance, the Knothe-Rosenblatt transport map (see [8] or [18, Section 2.3]).

A different construction was inspired by the so called Iterative Distribution Transfer (IDT) algorithm (first proposed in [15]), which was based on a sequence of one-dimensional optimal matchings between projections of the measures along different axes, chosen randomly at each time step. In order to avoid the dependence of this approximation scheme on the choice of the axes, M. Bernot later proposed an homogenization of the IDT procedure. We will see later that this approach is quite related to the notion of sliced-Wasserstein distance, which has been the object of many works in the last decade, starting from [16]. Yet, not a lot can be found in the

*Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova, Via Trieste, 63, 35131 Padova PD, Italy. cozzi@math.unipd.it

†Universite Claude Bernard Lyon 1, ICJ UMR5208, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne, France. santambrogio@math.univ-lyon1.fr

literature about this evolutionary procedure, which is now called sliced-Wasserstein flow and which we try to describe here, mainly based on [3] and [18, Section 2.5.2].

Suppose that ρ_0 and ν are absolutely continuous probability measures in $\mathbb{R}^d$ (we will use the same notation to indicate both an absolutely continuous probability and its density with respect to the Lebesgue measure $\mathscr{L}^d$), both with finite second moment. For any unitary direction $\vartheta \in \mathbb{S}^{d-1}$ we can define $\rho_0^\vartheta = (\pi^\vartheta)_\# \rho_0$ and $\nu^\vartheta = (\pi^\vartheta)_\# \nu$, where $\pi^\vartheta \colon \mathbb{R}^d \to \mathbb{R}$ is the canonical projection on the line $\vartheta\mathbb{R}$ and the symbol # indicates the push-forward of the measure. Since ρ_0^ϑ and ν^ϑ are one-dimensional probability measures (absolutely continuous, hence atomless), computing the optimal transport map T^ϑ between them is not difficult. Then, for every point $x \in \mathrm{spt}(\rho_0)$ and any direction ϑ, we want to move the point x in the direction ϑ of a displacement given by the displacement of the map T^ϑ at the point $x \cdot \vartheta$, and combine these displacements for all values of ϑ. We then define a vector field v via

$$v(x) := \fint_{\mathbb{S}^{d-1}} \left(T^\vartheta(x \cdot \vartheta) - x \cdot \vartheta \right) \vartheta \, d\mathscr{H}^{d-1}(\vartheta). \tag{1}$$

When needed, this vector field will be denoted by $v[\rho, \nu]$, as it depends on both the current measure ρ and on the target one ν, but when there is no ambiguity we will just use the notation v.

Fixing a small time step $\tau > 0$, we can define $\rho_1^\tau := (id + \tau v)_\# \rho_0$ and, iterating the construction, we can obtain a sequence $\{\rho_n^\tau\}_{n \in \mathbb{N}}$. Letting $\tau \to 0^+$ we recover an (absolutely continuous with respect to the 2–Wasserstein metric) curve of probabilities $(\rho_t)_{t \geq 0}$ solving the continuity equation

$$\partial_t \rho + \nabla \cdot (\rho v) = 0, \tag{2}$$

where $v = v[\rho, \nu]$ is obtained using the construction in (1).

The difference compared to the Iterative Distribution Transfer procedure is that in IDT we fix $\tau = 1$ and we only use finitely many ϑ, and more precisely we choose an orthonormal basis of $\mathbb{R}^d$. This continuous-in-time procedure is for sure smoother and more isotropic.

It can be seen that the above equation is indeed a gradient flow for the Wasserstein distance W_2 (see [1, 17]) of the following functional

$$\rho \mapsto \frac{SW_2^2(\rho, \nu)}{2},$$

where $SW_2(\rho, \nu)$ is the *2–sliced-Wasserstein distance* from ρ to ν, for which we refer to section 3.

From now on, we will call *Sliced-Wasserstein Flow (SWF)* the curve $(\rho_t)_{t \geq 0}$ obtained above. If the distance SW_2 has received much attention in the recent applied literature (because of its better computability than the standard W_2 or because of its better statistical properties, see for instance [16, 5, 2, 13] as well as [18, Section 5.5.4]), the flow itself has not. It was informally introduced by Bernot as a way to converge to the target measure, and few results are available on it: in [5] Bonnotte proved the existence of solutions to the SWF via JKO methods, and the same flow is also mentioned in [18, Sections 2.5.2 and 8.4.2], but uniqueness and questions related to the asymptotic behaviour of the flow still remain open. We cite [11] for a related flow, where diffusion is also added, and applications to generative flows. Yet, the presence of diffusion strongly changes the mathematical nature of these questions.

In this work we partially answer to two natural questions regarding the asymptotic behavior of the SWF:

- **Question 1.** (Long-time asymptotics of the flow) Is it true that the limit $\rho_\infty = \lim_{t \to \infty} \rho_t$ (in any suitable weak sense) exists and that we have $\rho_\infty = \nu$?

- **Question 2.** (Optimality of the flow map) Fix an initial particle $x \in \mathbb{R}^d$ and consider the flow map describing the characteristics of the flow. This map is obtained by solving, for every inital point x, the following system

$$\begin{cases} \dot{y_x} = v(t, y_x), \\ y_x(0) = x, \end{cases} \tag{3}$$

  where v is the velocity field defined in (1). Assume that $y_x(t)$, solution of the system above, exists and is well defined for any $t \geq 0$ and for every $x \in \mathbb{R}^d$, we will denote by Y_t the map $x \mapsto y_x(t)$. Moreover, assume that the limit $Y_\infty := \lim_{t\to\infty} Y(t)$ is well defined and that $\lim_{t\to\infty} \rho_t = \nu$. Is it true that the map Y_∞ is optimal in the sense of optimal transport between the initial datum ρ_0 and the target measure ν?

A similar question concerning the optimality of the flow map was already studied in the case of another flow map, the one obtained for the Fokker Planck equation. Negative results were produced first in [19], and then in [9]. The latter result also included the case where the target distribution is the standard Gaussian. Similarly to the problem studied in [19, 9], the flow map of the sliced-Wasserstein flow has also been the object of numerical investigation: in the case of discrete measures [15] suggests that the map Y_∞ obtained as in Conjecture 2 is a good approximation of the optimal one, but does not coincide with it in general. A similar result was also conjectured by Bernot [3].

This work is structured as follows: in the next section we recall some basic facts about optimal transport and Wasserstein spaces, then we present the definition and first properties of the sliced-Wasserstein distance and of the sliced-Wasserstein flow. In section 4 we prove estimates for the long-time behavior of the flow, and we discuss Conjecture 1, furnishing examples that show the flow may or may not converge, depending on the starting and target measures chosen: in particular, we prove convergence of the flow provided the target measure is the Gaussian distribution. This positive result, together with the nature of the negative ones suggest that one could have convergence to the target as soon as the initial measure is of finite entropy, but this general result is unfortunately unproven. Finally, section 5 contains the counterexample which shows that Conjecture 2 is false.

2 Notations and some results in optimal transport

For a function $u : \mathbb{R}^d \mapsto \mathbb{R}$, we denote by ∇u its gradient and D^2u its Hessian matrix. Moreover, for a map $T : \mathbb{R}^d \mapsto \mathbb{R}^d$, its Jacobian matrix is denoted by DT. Given any open subset $\Omega \subset \mathbb{R}^d$, we denote with $\mathscr{P}(\Omega)$ the set of probability measures with support in Ω. We denote with $\mathscr{P}_p(\Omega)$ the set of probability measures on Ω with finite moment of order p (in case Ω is unbounded), so that we have $\mathscr{P}_p(\Omega) = \{\mu \in \mathscr{P}(\Omega) : \int |x|^p d\mu < +\infty\}$.

Given $\mu, \nu \in \mathscr{P}(\Omega)$, we say that T is a transport map from μ to ν, and that ν is the pushforward of μ via T (and we write $\nu = T_\#\mu$) if for every Borel set $A \subset \Omega$ we have

$$\nu(A) = \mu(T^{-1}(A))$$

or, equivalently, if for any Borel map $\phi\colon \Omega \to \mathbb{R}$ it holds

$$\int_\Omega \phi(x) d\nu(x) = \int_\Omega \phi(T(x)) d\mu(x).$$

Definition 2.1. *Let $\mu, \nu \in \mathscr{P}(\Omega)$ and a continuous cost function $c : \Omega \times \Omega \to \mathbb{R}$. The map $T : \mathbb{R}^d \to \mathbb{R}^d$ is said to be an optimal transport map for the cost c if it solves the problem*

$$\min_T \left\{ \int_\Omega c(x, T(x)) d\mu(x), \quad T \text{ Borel map such that } T_\# \mu = \nu \right\}. \tag{4}$$

The map $\varphi : \mathbb{R}^d \to \mathbb{R}$ is said to be a Kantorovich potential if it solves the problem

$$\max_\varphi \left\{ \int_\Omega \varphi(x) d\mu(x) + \int_\Omega \varphi^c(x) d\nu(x) \right\}, \tag{5}$$

where $\varphi^c(x) := \inf_{y \in \mathbb{R}^d} \{c(x, y) - \varphi(y)\}$.

Remark. Problems (4) and (5) do not always admit a solution. However, when they exist, the two quantities defined by (4) with "min" replaced by "inf", and (5) with "max" replaced by "sup" coincide.

We are in particular interested in the case $\Omega = \mathbb{R}^d$ and $c(x, y) = \frac{1}{2}|x - y|^2$. In this case we have a relation between the optimal T and the optimal φ (which is a Lipschitz function, and hence diffentiable a.e.) which is

$$T(x) = x - \nabla\varphi(x). \tag{6}$$

In particular, we have $T = \nabla u$ where $u(x) = \frac{1}{2}|x|^2 - \varphi(x)$, and one can check that u is convex. This is summarized in the following celebrated theorem.

Theorem 2.1 (Brenier's theorem, see [6])**.** *Let μ, ν be two absolutely continuous probability measures in $\mathscr{P}_2(\mathbb{R}^d)$. A map $T: \mathbb{R}^d \to \mathbb{R}^d$ such that $T_\# \mu = \nu$ is optimal for the optimal transport problem* (4) *with the cost $c(x, y) = \frac{1}{2}|x - y|^2$ if and only if there exists a convex function $u: \mathbb{R}^d \to \mathbb{R}$ such that $T = \nabla u$ a.e. This is equivalent to DT being a symmetric positive semi-definite matrix valued measure on $\mathbb{R}^d$. Moreover if it is the case, then u must satisfy the Monge-Ampére equation*

$$\det D^2 u(x) = \frac{\nu(x)}{\mu(T(x))}$$

everywhere on $\mathbb{R}^d$. Finally, optimal transport maps are stable by "inversion": $T: \mathbb{R}^d \to \mathbb{R}^d$ is the optimal transport map from μ to ν if and only if T^{-1} is the optimal transport map from ν to μ.

For the one-dimensional problem, we are able to characterize optimal transport maps:

Theorem 2.2 (Brenier's theorem in $d = 1$, see [18], Chapter 2)**.** *Let μ, ν be two probability densities with finite second moments. If μ is atomless, then there exist a unique non-decreasing transport map from μ to ν, which is optimal and given by*

$$T(x) = F_\nu^{[-1]}\left(F_\mu(x)\right),$$

where $F_\mu(x) = \mu((-\infty, x])$ and $F^{[-1]}(y) = \inf\{t \in \mathbb{R}: F(t) \geq y\}$.

3 Sliced-Wasserstein distance and sliced-Wasserstein flow

3.1 Wasserstein spaces

Let $\Omega \subset \mathbb{R}^d$. Thanks to the transport value associated with the costs of the form $c(x, y) = |x - y|^p$ for $1 \leq p < \infty$, we can define a distance called *p–Wasserstein distance* over $\mathscr{P}_p(\Omega)$, the space of probability measures with support contained in Ω and finite p-moment.

Wasserstein distances play a key role in many fields of applications, and seem to be a natural way to describe distances between equal amounts of mass distributed on the same space.

Definition 3.1 (Wasserstein distance). *Let $1 \le p < \infty$, $\Omega \subset \mathbb{R}^d$. Given two absolutely ontinuous measures $\mu, \nu \in \mathscr{P}_p(\Omega)$ we define*

$$W_p(\mu,\nu) = \inf\left\{\int_\Omega |x - T(x)|^p d\mu(x) : \ T \ \text{Borel map such that} \ T_\#\mu = \nu\right\}^{\frac{1}{p}}.$$

The above definition could actually be extended to the whole set of measures in $\mathscr{P}_p(\Omega)$ (without the restriction to be absolutely continuous) but this requires to re-formulate the transport problem in terms of transport plans instead of transport maps. However, in this article this aspect will not be crucial and the only cases where non-absolutely continuous measures will be used concern some counter-examples where it is easy to understand the adaptations that need to be performed.

As a refence on the distance W_p we refer the reader to [18, Chapter 5] where it is proven, in particulat, that W_p is indeed a distance over $\mathscr{P}_p(\Omega)$. Moreover, the following theorem holds:

Theorem 3.1. *Fix $p \ge 1$, $q > p$ and $c > 0$. Then the set*

$$K := \left\{\mu \in \mathscr{P}_p\left(\mathbb{R}^d\right) : \int_{\mathbb{R}^d} |x|^q d\mu(x) \le c\right\}$$

is sequentially compact in the metric space $\left(\mathscr{P}_p\left(\mathbb{R}^d\right), W_p\right)$.

3.2 Sliced-Wasserstein

The following definition was introduced in [16], and provides a notion of distance on $\mathscr{P}_2(\mathbb{R}^d)$, alternative to the usual W_2 distance, based on the behavior of the measures "direction by direction".

Definition 3.2 (Sliced-Wasserstein distance). *Given $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, we define*

$$SW_2 := \left(\fint_{\mathbb{S}^{d-1}} W_2^2((\pi^\vartheta)_\#\mu, (\pi^\vartheta)_\#\nu) d\mathscr{H}^{d-1}(\vartheta)\right)^{1/2},$$

where $\pi^\vartheta : \mathbb{R}^d \to \mathbb{R}$ is the projection on the axis directed according to the unit vector ϑ, namely $\pi^\vartheta(x) = x \cdot \vartheta$, and $\mathscr{H}^{d-1}$ is the surface measure on $\mathbb{S}^{d-1}$.

It is not difficult to prove that, for any $\Omega \subset \mathbb{R}^d$, this is a distance on $\mathscr{P}_2(\Omega)$, that is always less than the usual 2-Wasserstein distance. The two distances generate the same topology (see [2]), with Hölder relation on compact sets (see [5], Chapter 5). Many interesting properties, pointing out surprising difference in the behavior of SW_2 compared to W_2, have been studied in the recent paper [14]. We also mention the possibility to define a distance SW_p, via $SW_p := \left(\fint_{\mathbb{S}^{d-1}} W_p^p((\pi^\vartheta)_\#\mu, (\pi^\vartheta)_\#\nu) d\mathscr{H}^{d-1}(\vartheta)\right)^{1/p}$, which is sometimes more convenient for sharp comparisons with the standard Wasserstein distances (see [7]).

To show that equation (2) derived in the introduction, describing the one-dimensional approximation scheme for a transport map, can be seen as the gradient flow of the functional $\mathscr{F} : \rho \mapsto \frac{1}{2} SW_2^2(\rho, \nu)$ in the Wasserstein space $(\mathscr{P}(\mathbb{R}^d), W_2)$ we need to convince ourselves that $v_t = -\nabla \frac{\delta \mathscr{F}}{\delta \rho}(\rho_t)$, where $\frac{\delta \mathscr{F}}{\delta \rho}(\rho)$ is the first variation of $\mathscr{F}$, defined by

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathscr{F}(\rho + \varepsilon\chi) = \int \frac{\delta \mathscr{F}}{\delta \rho}(\rho) d\chi$$

for any perturbation $\chi = \mu - \rho$ with $\mu \in L^\infty(\mathbb{R}^d) \cap \mathscr{P}(\mathbb{R}^d)$ and compactly supported. Since the first variation of the functional $\rho \mapsto \frac{1}{2} W_2^2(\rho, \nu)$ is the Kantorovich potential for the transport problem from μ to ν (see [18], proposition 7.17), we get

$$\frac{\delta \mathscr{F}}{\delta \rho}(\rho)(x) = \fint_{\mathbb{S}^{d-1}} \frac{1}{2} \frac{\delta W_2^2(\rho^\vartheta, \nu^\vartheta)}{\delta \rho}(x) d\mathscr{H}^{d-1}(\vartheta) = \fint_{\mathbb{S}^{d-1}} \varphi^\vartheta(x \cdot \vartheta) d\mathscr{H}^{d-1}(\vartheta),$$

where we denoted by φ^ϑ the Kantorovich potential between $\rho^\vartheta := (\pi^\vartheta)_\# \rho$ and $\nu^\vartheta := (\pi^\vartheta)_\# \nu$. Using also (6), this in turns leads to

$$-\nabla \frac{\delta \mathscr{F}}{\delta \rho}(\rho) = \fint_{\mathbb{S}^{d-1}} \left(T_t^\vartheta(x \cdot \vartheta) - x \cdot \vartheta \right) \vartheta d\mathscr{H}^{d-1}(\vartheta),$$

which is precisely the velocity field derived in (1).

4 Long-time asymptotics of the SWF

The question of the limit as $t \to \infty$ of the distribution ρ_t is very natural but surprisingly difficult, even if the SWF was probably mainly introduced as a way to approach the target distribution ν. Section 5.7 in [5] explains the difficulty of this question as it is not clear, except in the case of strictly positive densities, that critical points should coincide with the target. Note that [10] proves convergence results for a discrete (in time) version of a similar procedure, but among its assumption there is a very restrictive one: the existence of a compact set in the space of measures where the evolution takes place and which contains no other critical point than ν.

4.1 Bounds on the moments along the flow

In this section we provide some uniform bounds to the p-moments of the sliced-Wasserstein flow. We start by computing the derivative with respect to time of the p–moment ($p \geq 2$) of the sliced-Wasserstein flow ρ_t associated with a given target measure ν. We distinguish the case $p = 2$ and $p > 2$

For $p = 2$, using the equality

$$x \cdot \vartheta \left(T^\vartheta(x \cdot \vartheta) - x \cdot \vartheta \right) = -\frac{|T^\vartheta(x \cdot \vartheta) - x \cdot \vartheta|^2}{2} - \frac{|x \cdot \vartheta|^2}{2} + \frac{|T^\vartheta(x \cdot \vartheta)|^2}{2},$$

we have

$$\begin{aligned}
\partial_t \int \frac{|x|^2}{2} \rho_t &= \partial_t \int \frac{|x|^2}{2} \rho_t = -\int \frac{|x|^2}{2} \nabla \cdot (v \rho_t) = \int x \cdot v \rho_t \\
&= \fint \int x \cdot \vartheta \left(T^\vartheta(x \cdot \vartheta) - x \cdot \vartheta \right) \rho_t = \fint \int_{\mathbb{R}} \frac{|T_t^\vartheta(y)|^2}{2} \rho_t^\vartheta(y) \\
&- \fint \int_{\mathbb{R}} \frac{r^2}{2} \rho_t^\vartheta(y) - \frac{1}{2} \fint \int_{\mathbb{R}} |T_t^\vartheta(y) - y|^2 \rho_t^\vartheta(y) \\
&= \fint \int_{\mathbb{R}} \frac{|T_t^\vartheta(y)|^2}{2} \rho_t^\vartheta(y) - \fint \int_{\mathbb{R}} \frac{r^2}{2} \rho_t^\vartheta(y) - \frac{1}{2} SW_2^2(\rho_t, \nu) \\
&= \fint \int_{\mathbb{R}} \frac{r^2}{2} \nu^\vartheta(r) - \fint \int_{\mathbb{R}} \frac{r^2}{2} \rho_t^\vartheta(y) - \frac{1}{2} SW_2^2(\rho_t, \nu)
\end{aligned} \tag{7}$$

For $p > 2$, on the other hand, we have (defining $\mathcal{M}_p(\rho) := \int |x|^p \rho$)

$$\begin{aligned}
\partial_t \mathcal{M}_p(\rho_t) &= \partial_t \int \frac{|x|^p}{p} \rho_t = -\int \frac{|x|^p}{p} \nabla \cdot (v\rho_t) = \int |x|^{p-2} x \cdot v\rho_t \\
&= \int |x|^{p-2} x \cdot \left(\fint (T_t^{\vartheta}(x \cdot \vartheta) - x \cdot \vartheta)\vartheta \right) \rho_t = \fint \int |x|^{p-2} x \cdot \vartheta (T_t^{\vartheta}(x \cdot \vartheta) - x \cdot \vartheta) \rho_t \\
&\leq -\fint \int |x|^{p-2} \frac{|x \cdot \vartheta|^2}{2} \rho_t + \fint \int |x|^{p-2} \frac{|T_t^{\vartheta}(x \cdot \vartheta)|^2}{2} \rho_t
\end{aligned}$$

where we used the inequality

$$a \cdot (b - a) \leq \frac{1}{2}|b|^2 - \frac{1}{2}|a|^2.$$

Thus

$$\begin{aligned}
\partial_t \mathcal{M}_p(\rho_t) &\leq -\frac{1}{2} \int |x|^{p-2} \left(\fint |x \cdot \vartheta|^2 \right) \rho_t + \frac{1}{2} \fint \int |x|^{p-2} |T_t^{\vartheta}(x \cdot \vartheta)|^2 \rho_t \\
&\leq -\int |x|^{p-2} \frac{|x|^2}{2d} \rho_t + \frac{1}{2} \fint \left[\left(\int |x|^p \rho_t \right)^{\frac{p-2}{p}} \left(\int |T_t^{\vartheta}(x \cdot \vartheta)|^p \rho_t \right)^{\frac{2}{p}} \right]
\end{aligned}$$

using the equality

$$\fint_{\mathbb{S}^{d-1}} |x \cdot \vartheta|^2 d\mathscr{H}^{d-1}(\vartheta) = \frac{|x|^2}{d}.$$

This is a particular case of a more general family of equalities,

$$\fint |x \cdot \vartheta|^p = c_{p,d} |x|^p,$$

which are true for certain constants $c_{p,d} < 1$ (and indeed we have $c_{2,d} = 1/d$). Hence, finally, using Jensen's inequality, we get

$$\begin{aligned}
\partial_t \mathcal{M}_p(\rho_t) &\leq -\frac{1}{2d} \mathcal{M}_p(\rho_t) + \frac{1}{2} \mathcal{M}_p(\rho_t)^{1-\frac{2}{p}} \left(\int \fint |y \cdot \vartheta|^p \nu \right)^{\frac{2}{p}} \\
&\leq -\frac{1}{2d} \mathcal{M}_p(\rho_t) + \frac{1}{2} \mathcal{M}_p(\rho_t)^{1-\frac{2}{p}} c_{p,d}^{\frac{2}{p}} \mathcal{M}_p(\nu)^{\frac{2}{p}}.
\end{aligned} \tag{8}$$

We deduce

$$\mathcal{M}_p(\rho_t) \leq \max \left\{ \mathcal{M}_p(\rho_0), d^{p/2} c_{p,d} \mathcal{M}_p(\nu) \right\}, \tag{9}$$

since (8) gives a differential inequality of the form

$$x' \leq -\frac{1}{2d} x + x^{1-2/p} \frac{C^{2/p}}{2},$$

and the left hand side is negative for

$$x \geq C d^{p/2}.$$

By Theorem 3.1 we deduce that, if the target and starting measure have finite p-moments, the flow describes a compact curve in W_q for any $q < p$. Moreover, denoting by

$$R(\mu) := \inf \{ r > 0 \colon \operatorname{supp}(\mu) \subset \{|x| \leq r\} \} = \lim_{p \to \infty} \mathcal{M}_p^{1/p}(\mu),$$

we can see that we have

$$R(\rho_t) = \lim_{p\to\infty} \mathcal{M}_p(\rho_t)^{1/p} \le \max\left\{\lim_{p\to\infty} \mathcal{M}_p(\rho_0)^{1/p}, d^{1/2} \lim_{p\to\infty} c_{p,d}^{1/p} \mathcal{M}_p(\nu)^{1/p}\right\}$$
$$= \max\left\{R(\rho_0),\ \sqrt{d}R(\nu)\right\},$$

since

$$\lim_{p\to\infty} c_{d,p} = \lim_{p\to\infty} \|\vartheta\|_{L^p(\mathbb{S}^{d-1})} = \|\vartheta\|_{L^\infty(\mathbb{S}^{d-1})} = 1.$$

Obviously this estimate is useless when the initial and target measures are not compactly supported. This estimate is not sharp, and one can show that the constant $\sqrt{d}$ can be improved by finer arguments on the direction of the velocity field v on large balls far from the support of the target measure.

4.2 Gaussian-target case

We prove here that when the target measure ν is the standard Gaussian distribution, then $\lim_{t\to\infty} \rho_t$ in the Wasserstein topology exists and coincides with the Gaussian distribution ν itself. The argument we present can be adapted to isotropic Gaussian measures of arbitrary variance, but unfortunately does not generalize to other Gaussian distribution (whose covariance matrix is not a scalar multiple of the identity).

Let the entropy function $\mathcal{E}\colon \mathscr{P}(\mathbb{R}) \to \mathbb{R}$ be defined by

$$\mathcal{E}(\rho) = \begin{cases} \int_{\mathbb{R}^d} \rho \log \rho & \text{if } \rho \ll \mathscr{L}^d, \\ +\infty & \text{otherwise.} \end{cases}$$

Theorem 4.1. *Let ν be the standard Gaussian distribution on $\mathbb{R}^d$. If ρ_0 has finite entropy and finite second moment, then the flow $(\rho_t)_{t\ge 0}$ is well posed and we have*

$$SW_2^2(\rho_t, \nu) \le \frac{C}{t}$$

where $C = C(\rho_0) = 2(\mathcal{E}(\rho_0) + \mathcal{M}_2(\rho_0))$.

To prove this theorem we need the following preliminary result.

Lemma 4.2. *Given two absolutely continuous probability measures $\mu, \nu \in \mathscr{P}_2(\mathbb{R})$, consider the optimal transport map T from μ to ν. Then we have*

$$\int_{\mathbb{R}} \partial_y \mu(y)\,(T(y) - y)\,dy \le \mathcal{E}(\nu) - \mathcal{E}(\mu).$$

Proof. Consider the curve of measures $\omega(t) = ((1-t)id + tT)_\# \mu$. Such a curve is a constant-speed geodesic (with respect to the W_2 distance) connecting $\mu = \omega(0)$ to $\nu = \omega(1)$.

Since ω is a geodesic, we know it must solve the the continuity equation

$$\partial_t \omega + \partial_y(\mathbf{v}_t \omega) = 0,$$

where the velocity field is $\mathbf{v}(t,y) = T(S_t^{-1}(y)) - S_t^{-1}(y)$, with $S_t(y) := (1-t)y + tT(y)$. We thus have

$$\begin{aligned}
\partial_t \int_{\mathbb{R}} \omega_t \log \omega_t dr &= \int_{\mathbb{R}} \partial_t \omega_t \log \omega_t dr \\
&= -\int_{\mathbb{R}} \partial_r(\mathbf{v}_t \omega_t) \log \omega_t dr = \int_{\mathbb{R}} \mathbf{v}_t \partial_y \omega_t \\
&= \int_{\mathbb{R}} \left[T\left(S_t^{-1}(y)\right) - S_t^{-1}(y)\right] \partial_y \omega_t dy,
\end{aligned}$$

and valuating the expression at $t = 0$, we obtain:

$$\partial_t\Big|_{t=0} \mathcal{E}(\omega(t)) = \int_{\mathbb{R}} \partial_y \mu(y)\,(T(y) - y)\,dy.$$

We now use the convexity of the function $t \mapsto \mathcal{E}(\omega(t))$, where ω is a Wasserstein geodesic in $\mathscr{P}(\mathbb{R}^d)$. Indeed, the entropy is convex along these geodesics (see [12] and [18, Section 7.3.2]), and for any convex function f one has $f'(0) \le f(1) - f(0)$. This proves the claim. □

We are now ready to prove theorem 4.1

Proof. Since $(\rho_t)_{t\ge 0}$ solves the continuity equation (2), we have

$$\begin{aligned}
&\partial_t \int \rho_t \log \rho_t dx = \int [\partial_t \rho_t \log \rho_t + \partial_t \rho_t] dx = -\int [\nabla \cdot (v_t \rho_t) \log \rho_t - \nabla(v_t \rho_t)] dx \\
&= -\int \nabla \cdot (\rho_t v_t) \log \rho_t dx = \int \rho_t v_t \cdot \frac{1}{\rho_t} \nabla \rho_t dx = \int v_t \cdot \nabla \rho_t dx \\
&= \fint \left[\int \nabla \rho_t \cdot \vartheta \left(T_t^{\vartheta}(x\cdot\vartheta) - x\cdot\vartheta\right) dx\right] d\mathscr{H}^{d-1}(\vartheta) \\
&= \fint \int_{\mathbb{R}} \partial_y \rho_t^{\vartheta}(y) \left(T_t^{\vartheta}(y) - y\right) dy d\mathscr{H}^{d-1}(\vartheta) \le \fint \left[\mathcal{E}\left(\nu^{\vartheta}\right) - \mathcal{E}\left(\rho^{\vartheta}\right)\right] d\mathscr{H}^{d-1}(\vartheta).
\end{aligned} \tag{10}$$

The last inequality comes from the inequality in Lemma 4.2. We then obtain

$$\partial_t \mathcal{E}(\rho_t) \le \fint \mathcal{E}(\nu^{\vartheta}) - \fint \mathcal{E}(\rho^{\vartheta}). \tag{11}$$

Now notice that the Gaussian measure minimizes the functional

$$\rho \mapsto \mathcal{E}(\rho) + \int \frac{|x|^2}{2} \rho.$$

Indeed $f(s) = s \log s$ is a convex function, and thus $f(t) \ge f(s) + f'(s)(t-s) = f(s) + (\log s + 1)(t-s)$. Thus, for every absolutely continuous ρ, and if ν is the Gaussian distribution,

$$\begin{aligned}
\int \rho \log \rho &\ge \int \nu \log \nu + \int \log \nu(\rho - \nu) + \int (\rho - \nu) \\
&= \int \nu \log \nu - \int \frac{|x|^2}{2} \rho + \int \frac{|x|^2}{2} \nu,
\end{aligned}$$

and therefore

$$\int \left[\rho \log \rho + \frac{|x|^2}{2} \rho\right] \ge \int \left[\nu \log \nu + \frac{|x|^2}{2} \nu\right] = 0$$

Summing Equations (7) and (11), and using the fact that, for any $\vartheta \in \mathbb{S}^{d-1}$, ν^ϑ is still a (one-dimensional) standard Gaussian, we obtain

$$\begin{aligned}
&\partial_t \left(\mathcal{E}(\rho_t) + \int \frac{|x|^2}{2} \rho_t \right) \\
&\leq \fint \left(\mathcal{E}\left(\nu^\vartheta\right) - \mathcal{E}\left(\rho_t^\vartheta\right) \right) + \fint \int \frac{|y|^2}{2} \nu^\vartheta - \fint \int \frac{|y|^2}{2} \rho_t^\vartheta - \frac{1}{2} SW_2^2(\rho_t, \nu) \\
&= \fint \left[\left(\mathcal{E}\left(\nu^\vartheta\right) + \int \frac{|y|^2}{2} \nu^\vartheta \right) - \left(\mathcal{E}\left(\rho_t^\vartheta\right) + \int \frac{|y|^2}{2} \rho_t^\vartheta \right) \right] - \frac{1}{2} SW_2^2(\rho_t, \nu) \\
&\leq -\frac{1}{2} SW_2^2(\rho_t, \nu).
\end{aligned}$$

Hence, integrating with respect to time, and recalling that $\frac{1}{2} SW_2^2$ is a decrasing quantity along its gradient flow,

$$\begin{aligned}
\frac{T}{2} SW_2^2(\rho_T, \nu) \leq \frac{1}{2} \int_0^T SW_2^2(\rho_t, \nu) dt &\leq -\left[\mathcal{E}(\rho_T) - \mathcal{E}(\rho_0) \right] - \int \left[\frac{|x|^2}{2} \rho_T - \frac{|x|^2}{2} \rho_0 \right] dt \\
&\leq \mathcal{E}(\rho_0) + \int \frac{|x|^2}{2} \rho_0 =: C
\end{aligned}$$

for every $T \geq 0$. Thus

$$SW_2^2(\rho_t, \nu) \leq 2 \frac{\mathcal{E}(\rho_0) + \int \frac{|x|^2}{2} \rho_0}{t}.$$

□

The above proof is inspired by the proof of convergence of the IDT procedure in [15], presented as well in [5, Theorem 5.2.2].

Corollary 4.3. *Take $q \geq 2$. Then, if $\rho_0 \in \mathscr{P}_q(\mathbb{R}^d)$ has finite entropy, we have*

$$\rho_t \xrightarrow[t \to \infty]{W_p} \nu,$$

for all values of $p < q$.

Proof. The proof is easy once we observe that (9) and Theorem 3.1 provide compactness of the flow $(\rho_t)_{t \geq 0}$ in the spaces $(\mathscr{P}_p(\mathbb{R}^d), W_2)$ provided ρ_0 has finite moments. This turns the convergence in the distance SW_2 into a more standard W_p convergence. for any $q < p$. □

4.3 The flow may not converge to the target

We can provide examples in which the flow does not converge to the target.

Example 1. *Consider the following construction in $\mathbb{R}^2$: let the starting measure be*

$$\rho_0 = \frac{\delta_{(-1,0)}}{2} + \frac{\delta_{(1,0)}}{2}$$

and the target measure

$$\nu = \frac{\delta_{(0,a)}}{2} + \frac{\delta_{(0,-a)}}{2},$$

for a certain $a > 0$ to be defined. We can show that there exists $a > 0$ such that the configuration ρ_0 is stationary, i.e. the velocity field associated with the SWF at time zero is $v_0 = 0$ and therefore the gradient flow is $\rho_t = \rho_0$ for any $t \geq 0$, so that the flow does not converge to the target. Calling T^ϑ the optimal transport map from ρ_ϑ and ν_ϑ, we have

$$\begin{aligned}
v_0 &= \fint \left(T^\vartheta(x \cdot \vartheta) - x \cdot \vartheta\right) \vartheta \\
&= 2 \fint_0^{\frac{\pi}{2}} (a \sin\vartheta - \cos\vartheta) \begin{pmatrix} \cos\vartheta \\ \sin\vartheta \end{pmatrix} - 2 \fint_{-\frac{\pi}{2}}^0 (\cos\vartheta + a \sin\vartheta) \begin{pmatrix} \cos\vartheta \\ \sin\vartheta \end{pmatrix} \\
&= 2 \fint_0^{\frac{\pi}{2}} \begin{pmatrix} a \sin\vartheta \cos\vartheta - \cos^2\vartheta \\ a \sin^2\vartheta - \cos\vartheta \sin\vartheta \end{pmatrix} - 2 \fint_{\frac{-\pi}{2}}^0 \begin{pmatrix} \cos^2\vartheta + a \cos\vartheta \sin\vartheta \\ \cos\vartheta \sin\vartheta + a \sin^2\vartheta \end{pmatrix} \\
&= \begin{pmatrix} a \fint_0^{2\pi} \frac{\sin(2\vartheta)}{2} sgn(\vartheta) - \fint_0^{2\pi} \cos^2\vartheta \\ 0 \end{pmatrix}.
\end{aligned}$$

All we have to do now is imposing the first row of the above matrix equal to zero, obtaining:

$$a = \frac{\pi}{2}.$$

Note that the above example is not new at all and similar computations on the case where both the target and the starting measure are composed of two symmetric atoms are also illustrated in [20, Section 1.6] or [4, Figure 2].

Example 2. *Consider the following construction in $\mathbb{R}^2$: let the starting measure be*

$$\rho_0 = \frac{\mathscr{H}^1\left((a,b) \times \{0\}\right)}{|b-a|}$$

and the target measure ν be any radial distribution. Then, by the symmetry in the construction, we can see that we always have $\mathrm{supp}(\rho_t) \subset \mathbb{R} \times \{0\}$. *Thus, in general, this flow will not converge to the target measure.*

Notice that in both these examples we cannot assume ρ_0 to be absolutely continuous with respect to the Lebesgue measure. This is indeed coherent with the strategy used to prove the convergence of the Gaussian-target case, which relied on entropy methods.

5 The flow map of the SWF does not provide optimal transport

This section contains a counterexample to Conjecture 2 presented in the introduction. We show the following:

It is not true that, for any initial distribution ρ_0 and any target distribution ν, the sliced-Wasserstein flow converges to the target measure ν itself and the limit of the flow map arising from the Lagrangian description of the model exists and is the optimal transport map from ρ_0 to ν.

Using the idea developed in [9], we can write a necessary condition which must hold if we want Conjecture 2 to be true: take ρ_0 sufficiently smooth and quickly decaying at infinity, and assume that Conjecture 2 holds not only for ρ_0 but for all $(\rho_t)_{t\geq 0}$. This means that the flow Y_t is well defined for any $t \geq 0$, that the limit $\lim_{t\to\infty} Y_t =: T$ exists, and that it is the optimal transport map between ρ_0 and ν. Under these assumptions, we have that for any $t \geq 0$ the SWF provides

optimal transport from ρ_t to ν. That is, the map $T \circ Y_t^{-1}$ is an optimal transport map between ρ_t and ν. Thus, $Y_t \circ T^{-1}$ is also an optimal transport map between ν and ρ_t. Let us denote by S the map T^{-1}. Making use of theorem (2.1), we have that the following condition must hold

$$\forall t \geq 0, \forall x \in \mathbb{R}^d, \text{ the Jacobian of } Y_t \circ S(x) \text{ is a symmetric matrix .} \tag{12}$$

This Jacobian matrix reads $DY_t(S)DS$. Moreover, differentiating (3) with respect to x, we see that

$$\frac{\partial DY_t}{\partial t} = -\left(\fint_{\mathbb{S}^{d-1}} \varphi''_{t,\vartheta}(x \cdot \vartheta)\vartheta \otimes \vartheta \quad d\vartheta \right),$$

together with $DY_0 = \mathrm{Id}$. Thus, differentiating the Jacobian of $Y_t \circ S$ with respect to time, we have that condition (12) implies

$$\forall t \geq 0, \forall x \in \mathbb{R}^d, \qquad \left[\fint_{\mathbb{S}^{d-1}} \varphi''_{t,\vartheta}\left(S(x) \cdot \vartheta\right) \vartheta \otimes \vartheta \right] DS(x) \quad \text{is a symmetric matrix.}$$

Evaluating this expression at $t = 0$ and remembering that $DS(x)$ is symmetric for any x (since our assumption is that S is an optimal transport map between ν and ρ_0) we conclude that the matrices

$$\fint_{\mathbb{S}^{d-1}} \varphi''_{\vartheta}\left(S(x) \cdot \vartheta\right) \vartheta \otimes \vartheta \quad \text{and} \quad DS(x)$$

should commute for any x. Composing both the matrices with $S^{-1} = T$ on the right hand side and using the identity $DS(S^{-1}) = [DT]^{-1}$, and the fact that a symmetric matrix A commutes with an invertible matrix B if and only if it commutes with B^{-1}, we conclude that:

$$\forall x \in \mathbb{R}^d, \qquad \fint_{\mathbb{S}^{d-1}} \varphi''_{\vartheta}\left(x \cdot \vartheta\right) \vartheta \otimes \vartheta \quad \text{and} \quad DT(x) \quad \text{commute.}$$

Using again Theorem 2.1, namely the fact that $T = Du$ for a convex map $u\colon \mathbb{R}^d \to \mathbb{R}$, we can write

$$\forall x \in \mathbb{R}^d, \qquad \fint_{\mathbb{S}^{d-1}} \varphi''_{\vartheta}\left(x \cdot \vartheta\right) \vartheta \otimes \vartheta \quad \text{and} \quad D^2u(x) \quad \text{commute.}$$

For the sake of simplicity we will call $\rho_0 = \rho$ in the following. Denote also by $u_\vartheta\colon \mathbb{R} \to \mathbb{R}$ the map whose derivative is the transport map T^ϑ between ρ^ϑ and ν^ϑ. Writing the Monge-Ampére equation for ρ^ϑ and ν^ϑ we have

$$u''_\vartheta(z) = \frac{\rho^\vartheta(z)}{\nu^\vartheta\left(T^\vartheta(z)\right)} \quad \forall z \in \mathbb{R}.$$

Since we have the equality $\varphi''_\vartheta(z) = 1 - u''_\vartheta(z)$, we get

$$M_{i,j}(x) := \fint \left(1 - \frac{\rho^\vartheta(x \cdot \vartheta)}{\nu^\vartheta\left(T^\vartheta(x \cdot \vartheta)\right)}\right) \vartheta_i \vartheta_j = \frac{\delta_{ij}}{d} - \fint \frac{\rho^\vartheta(x \cdot \vartheta)}{\nu^\vartheta(T^\vartheta(x \cdot \vartheta))} \vartheta_i \vartheta_j.$$

If we choose ρ and ν to be even measures ($\rho(x) = \rho(-x)$, $\nu(x) = \nu(-x)$ for all $x \in \mathbb{R}^d$), we have $T^\vartheta(0) = 0$ for all $\vartheta \in \mathbb{S}^{d-1}$. Moreover, we also choose ν to be radial (which is stronger than even), so that the value $\nu^\vartheta(0)$ does not depend on ϑ. Our goal can be rewritten as follows: we want to show that, in general, the following two matrices do not commute:

$$\frac{\mathrm{Id}}{d} - M(0) = \fint \frac{\rho^\vartheta(0)}{\nu^\vartheta(0)} (\vartheta \otimes \vartheta) d\vartheta = c \fint \rho^\vartheta(0)(\vartheta \otimes \vartheta) d\vartheta \quad \text{and} \quad D^2u(0). \tag{13}$$

Since we are free to choose any even initial measure ρ, let us impose $\rho = (\nabla u)_{\#}^{-1}\nu$, for $u(x) = |x|^2/2 + \varepsilon\varphi(x)$, where φ is a smooth, compactly supported and even function. It is well known that, for ε small enough, u is convex and ∇u is a C^∞ diffeomorphism that coincides with the identity outside of a compact set which moreover, by Brenier's theorem, is the optimal transport map from ρ to ν. Finally the Monge-Ampére equation guarantees that, if φ is even, so is ρ. We seek now for the expression of $\rho^\vartheta(0)$ in terms of φ. Again by the Monge-Ampére equation for the transport problem from ρ to ν, we deduce that, for any $x \in \mathbb{R}^d$, we have

$$\rho(x) = \det D^2u(x)\nu(\nabla u(x)).$$

In our case $\nabla u(x) = x + \varepsilon\nabla\varphi(x)$ and $\det D^2u(x) = 1 + \varepsilon\Delta\varphi(x) + O(\varepsilon^2)$. We deduce that for any $x \in \mathbb{R}^d$

$$\begin{aligned}\rho(x) =& \left(1 + \varepsilon\Delta\varphi + O(\varepsilon^2)\right)(\nu(x) + \varepsilon D\nu(x)\cdot\nabla\varphi(x)) + O(\varepsilon^2)\\ =&\nu(x) + \varepsilon\left(D\nu(x)\cdot\nabla\varphi(x) + \Delta\varphi(x)\nu(x)\right) + O(\varepsilon^2).\end{aligned}$$

By the definition of ρ^ϑ, we get

$$\rho^\vartheta(0) = \int_{\vartheta^\perp}\rho(y)dy = \int_{\vartheta^\perp}\left(\nu(y) + \varepsilon\left(D\nu(y)\cdot\nabla\varphi(y) + \Delta\varphi(y)\nu(y)\right)\right)dy + O(\varepsilon^2).$$

We now impose $\varphi(x) = \varepsilon\psi(x) + \eta(x)$, with ψ, η even functions (so that φ is even) and such that $\mathrm{supp}(\psi) \subseteq B(0,1)$ and $\mathrm{supp}(\eta) \subseteq B(Re_d,1) \cup B(-Re_d,1)$ for a fixed $R >> 1$, e_d being the last vector of the canonical basis in $\mathbb{R}^d$. In this way $D^2u(0) = \mathrm{Id} + \varepsilon^2 D^2\psi(0)$ and so, by (13), our aim becomes to prove that the matrix $\fint \rho^\vartheta(0)\vartheta\otimes\vartheta$ does not commute with $D^2\psi(0)$. We have

$$\begin{aligned}\rho^\vartheta(0) =& \int_{\vartheta^\perp}\nu(y)dy + \varepsilon\int_{\vartheta^\perp}\left(D\nu(y)\cdot\nabla\eta(y) + \Delta\eta(y)\nu(y)\right)dy + O(\varepsilon^2)\\ =& \int_{\vartheta^\perp}\nu(y)dy + \varepsilon\int_{\vartheta^\perp}\nu(y)\eta_{\vartheta\vartheta}(y)dy + O(\varepsilon^2),\end{aligned}$$

after integration by parts. We deduce that our goal becomes now to prove the following fact: "$D^2\psi(0)$ does not commute with A", where

$$\begin{aligned}A =& \fint\int_{\vartheta^\perp}\nu(y)\eta_{\vartheta\vartheta}(y)(\vartheta\otimes\vartheta)dyd\vartheta\\ =& \fint\int_{\vartheta^\perp}\nu(y)(\vartheta\otimes\vartheta)D^2\eta(y)(\vartheta\otimes\vartheta)dyd\vartheta\\ =&2\fint\int_{\vartheta^\perp}\nu(y)(\vartheta\otimes\vartheta)D^2\tilde{\eta}(y)(\vartheta\otimes\vartheta)dyd\vartheta,\end{aligned}$$

and $\tilde{\eta}$ is the restriction of η to $B(Re_d,1)$. For convenience, with a slight abuse of notation, in the following we will call η this very function. Notice in particular that $\eta(x) = 0$ if $x_d \le 0$. Since ψ is any even function, in order to get our claim we just have to ensure that the matrix A is not a multiple of the identity matrix. To do so, we can for example check that the diagonal elements of A are not equal.

In the remaining part of this counterexample, for simplicity, we restrict to the case $d = 2$. In this case, we can just check that we have $A_{11} - A_{22} \neq 0$: take $\eta(y_1,y_2) = \eta^r(y_1,y_2) = a^r(y_1)b(y_2)$ for a^r and b positive functions, even on their support and of unitary integrals, supported respectively

in $[-r, r]$, $[R, R+1/2]$ (so that, if r is small, $\mathrm{supp}(\eta^r) \subseteq B(Re_2, 1)$). Observe that $a^r \rightharpoonup \delta_0$ when $r \to 0$.

Using the notation $\eta^r_{ij} = (D^2\eta^r)_{ij}$, we have

$$A_{11} = 2 \fint \int_{\vartheta^\perp} \vartheta_1^2(\vartheta_1^2\eta^r_{11} + 2\vartheta_1\vartheta_2\eta^r_{12} + \vartheta_2^2\eta^r_{22})\nu(y)dyd\vartheta,$$

$$A_{22} = 2 \fint \int_{\vartheta^\perp} \vartheta_2^2(\vartheta_1^2\eta^r_{11} + 2\vartheta_1\vartheta_2\eta^r_{12} + \vartheta_2^2\eta^r_{22})\nu(y)dyd\vartheta.$$

Since in dimension $d = 2$, for any $y \in \vartheta^\perp$ one has

$$(\vartheta_1, \vartheta_2) = \left(-\frac{y_2}{|y|}, \frac{y_1}{|y|}\right),$$

we have

$$\begin{aligned} A_{11} =& 2 \fint \int_{\vartheta^\perp} \frac{y_2^2}{|y|^4}(y_2^2\eta^r_{11} - 2y_1y_2\eta^r_{12} + y_1^2\eta^r_{22})\nu(y)dyd\vartheta, \\ A_{22} =& 2 \fint \int_{\vartheta^\perp} \frac{y_1^2}{|y|^4}(y_2^2\eta^r_{11} - 2y_1y_2\eta^r_{12} + y_1^2\eta^r_{22})\nu(y)dyd\vartheta. \end{aligned} \tag{14}$$

To conclude we need the following

Lemma 5.1. *Given $f\colon \mathbb{R}^d \to \mathbb{R}$ such that the following integrals are well defined, we have*

$$\fint_{\mathbb{S}^{d-1}} \int_{\vartheta^\perp} f(y)dyd\vartheta = \frac{(d-1)\omega_{d-1}}{d\omega_d} \int_{\mathbb{R}^d} \frac{f(x)}{|x|}dx$$

where ω_k stands for the measure of the unit ball in dimension k.

Proof. First restricting to compactly supported functions f and then removing this restriction, it is clear, by the Riesz representation theorem for the dual of C^0, that there exists a locally finite measure μ on $\mathbb{R}^d$ such that $\fint_{\mathbb{S}^{d-1}} \int_{\vartheta^\perp} f(y)dyd\vartheta = \int f d\mu$. It is also clear by symmetry reason that μ is a radial measure. Hence, in order to identify μ it is enough to compute the mass that it gives to every ball $B(0, R)$. Taking for f the indicator function of such a ball we obtain $\mu(B(0, R)) = \omega_{d-1}R^{d-1}$. This is the same as taking for μ the measure with density $\frac{(d-1)\omega_{d-1}}{d\omega_d}\frac{1}{|x|}$, which concludes the proof. □

Applying Lemma 5.1 to Equations (14), we deduce

$$\fint_{\mathbb{S}^1} \int_{\vartheta^\perp} f(y)dyd\vartheta = \frac{1}{\pi} \int_{\mathbb{R}^d} \frac{f(x)}{|x|}dx.$$

Using this formula, the difference between the diagonal elements of the 2×2 - matrix A, up to multiplication by the constant π, reads

$$A_{11} - A_{22} = \int_{\mathbb{R}^2} \left(\frac{x_2^2 - x_1^2}{|x|^5}(x_2^2\eta^r_{11} - 2x_1x_2\eta^r_{12} + x_1^2\eta^r_{22})\right)\nu(x)dx_1dx_2.$$

Since

$$h_i(x) := \frac{x_i^2}{|x|^5}\nu(x)$$

can be assumed smooth enough in $\mathbb{R}^2 \setminus \{0\}$, when $r \to 0$ the integrals defining A_{11} and A_{22} stay bounded (remembering that η^r is compactly supported). Setting

$$f(x_1, x_2) := h_2(x) - h_1(x) = \frac{x_2^2 - x_1^2}{|x|^5} \nu(x_1, x_2),$$

we have

$$A_{11} - A_{22} = \int_{\mathbb{R}^2} f(x)(x_2^2 \partial_{11} a^r b - 2x_1 x_2 \partial_1 a^r \partial_2 b + x_1^2 a^r \partial_{22} b) dx_1 dx_2.$$

We need to be sure that this quantity is not zero as soon as $r \to 0$. Integrating by parts with respect to the variable x_1, we obtain

$$A_{11} - A_{22} = \int \left(x_2^2 b \int \partial_{11} f a^r + 2x_2 \partial_2 b \int (\partial_1 f x_1 + f) a^r + \partial_{22} b \int f x_1^2 a^r \right) dx_2.$$

Sending $r \to 0$, this gives

$$A_{11} - A_{22} = \int x_2^2 b \partial_{11} f(0, x_2) + 2x_2 \partial_2 b f(0, x_2) dx_2.$$

Integrating by parts with respect to x_2,

$$A_{11} - A_{22} = \int b \left(\partial_{11} f(0, x_2) x_2^2 - 2f(0, x_2) - 2x_2 \partial_2 f(0, x_2) \right) dx_2.$$

Remembering that x_2 can be assumed strictly positive (due to the fact that η^r is symmetric and supported away from zero), we have

$$\partial_1 f = \nu x_1 \frac{-2|x|^2 - 5x_2^2 + 5x_1^2}{|x|^7} + \frac{x_2^2 - x_1^2}{|x|^5} \nu_1,$$

then

$$f_{11}(0, x_2) = -7 \frac{\nu}{|x|^5} + \frac{\nu_{11}}{|x|^3}$$

and, by similar computations,

$$f_2(0, x_2) = -3 \frac{\nu}{|x|^4} + \frac{\nu_2}{|x|^3}.$$

Therefore

$$A_{11} - A_{22} = \int b \left(-3 \frac{\nu}{x_2^3} + \frac{\nu_{11}}{x_2} - 2 \frac{\nu_2}{x_2^2} \right) dx_2.$$

Considering that ν has been chosen to be radial, we make use of the relations

$$\nu_i = \nu' \frac{x_i}{|x|}$$

(where ν' – and later ν'' – stands for the derivative of ν in the radial direction), and

$$\nu_{ij} = \frac{\nu'}{|x|} \delta_{ij} + \left(\frac{\nu''}{|x|^2} - \frac{\nu'}{|x|^3} \right) x_i x_j,$$

to get that $A_{11} - A_{22} = 0$ for any choice of b if and only if

$$-3\frac{\nu}{|x|^3} - \frac{\nu'}{|x|^2} = 0$$

holds for any $x \in \{0\} \times \mathbb{R}^+$: indeed, since b is free to be chosen, so is its support (which we defined to be of the form $[R, R + 1/2]$). Still remembering that x_2 can be assumed strictly positive, we get that this is equivalent to

$$3\nu + |x|\nu' = 0. \tag{15}$$

To conclude the counterexample it is sufficient to choose a radial target measure ν not satisfying (15). By the way, no probability density can solve (15) since a solution ν should satisfy $\nu \propto |x|^{-3}$, which implies that ν is not a probability distribution, as it is not integrable.

Acknowledgments The authors acknowledge the support of the Lagrange Mathematics and Computation Research Center, where the project started during a Master internship of the first author. The support of the European union via the ERC AdG 101054420 EYAWKAJKOS is also acknowledged.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.

[2] Erhan Bayraktar and Gaoyue Guo. *Strong equivalence between metrics of Wasserstein type*. 2021. arXiv: `1912.08247 [math.PR]`.

[3] Marc Bernot. Personal communications dating back to 2008-2012.

[4] Nicolas Bonneel et al. "Sliced and Radon Wasserstein Barycenters of Measures". In: *Journal of Mathematical Imaging and Vision* 51.1 (2015), pp. 22–45.

[5] Nicolas Bonnotte. "Unidimensional and evolution methods for optimal transportation". PhD thesis. Université Paris Sud-Paris XI; Scuola normale superiore (Pise, Italie), 2013.

[6] Yann Brenier. "Décomposition polaire et réarrangement monotone des champs de vecteurs". In: *CR Acad. Sci. Paris Sér. I Math.* 305 (1987), pp. 805–808.

[7] Guillaume Carlier et al. Work in preparation.

[8] H. Knothe. "Contributions to the theory of convex bodies". In: *Michigan Mathematical Journal* 4 (1957), pp. 39–52.

[9] Hugo Lavenant and Filippo Santambrogio. "The flow map of the Fokker–Planck equation does not provide optimal transport". In: *Applied Mathematics Letters* 133 (2022), p. 108225.

[10] Shiying Li and Caroline Moosmueller. *Measure transfer via stochastic slicing and matching*. 2023. arXiv: `2307.05705 [math.NA]`.

[11] Antoine Liutkus et al. "Sliced-Wasserstein Flows: Nonparametric Generative Modeling via Optimal Transport and Diffusions". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 4104–4113. URL: `https://proceedings.mlr.press/v97/liutkus19a.html`.

[12] Robert J McCann. "A convexity principle for interacting gases". In: *Advances in mathematics* 128.1 (1997), pp. 153–179.

[13] Kimia Nadjahi et al. "Statistical and Topological Properties of Sliced Probability Divergences". In: *ArXiv* abs/2003.05783 (2020). URL: https://api.semanticscholar.org/CorpusID:212675820.

[14] Sangmin Park and Dejan Slepčev. "Geometry and analytic properties of the sliced Wasserstein space". In: *arXiv preprint arXiv:2311.05134* (2023).

[15] François Pitié, Anil C Kokaram, and Rozenn Dahyot. "Automated colour grading using colour distribution transfer". In: *Computer Vision and Image Understanding* 107.1-2 (2007), pp. 123–137.

[16] Julien Rabin et al. "Wasserstein barycenter and its application to texture mixing". In: *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*. Springer. 2012, pp. 435–446.

[17] Filippo Santambrogio. "{Euclidean, metric, and Wasserstein} gradient flows: an overview". In: *Bulletin of Mathematical Sciences* 7 (2017), pp. 87–154.

[18] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Vol. 55. 58-63. Springer, 2015.

[19] Anastasiya Tanana. "Comparison of transport map generated by heat flow interpolation and the optimal transport Brenier map". In: *Communications in Contemporary Mathematics* 23.06 (2021), p. 2050025.

[20] Eloi Tanguy, Rémi Flamary, and Julie Delon. "Properties of Discrete Sliced Wasserstein Losses". In: *arXiv preprint arXiv:2307.10352* (July 2023).