THE FLOW MAP OF THE FOKKER-PLANCK EQUATION DOES NOT PROVIDE OPTIMAL TRANSPORT

HUGO LAVENANT AND FILIPPO SANTAMBROGIO

ABSTRACT. In [Khrulkov and Oseledets. Understanding DDPM Latent Codes Through Optimal Transport. *arXiv preprint arXiv:2202.07477*] the authors conjecture that, by integrating the flow of the ODE given by the Wasserstein velocity in a Fokker-Planck equation, one obtains an optimal transport map. On the other hand this result was thought to be false in [Kim and Milman. A generalization of Caffarelli's contraction theorem via (reverse) heat flow. *Mathematische Annalen 354.3 (2012): 827-862*] but no proof was provided. In this note we show that the result claimed by Khrulkov and Oseledets cannot hold. This strengthens a counterexample which was built in [Tanana. Comparison of transport map generated by heat flow interpolation and the optimal transport Brenier map. *Communications in Contemporary Mathematics, 23(6)*, 2021].

On $\mathbb{R}^d$, let us consider μ_0 a probability measure having a smooth density with respect to the Lebesgue measure (still denoted by μ_0), and consider the following Fokker Planck equation

$$\frac{\partial \mu}{\partial t} - \nabla \cdot (x\mu) - \Delta \mu = 0 \tag{1}$$

with initial condition μ_0. It generates a curve $(\mu_t)_{t\geq 0}$ of probability measures. When $t \to +\infty$ the probability measure μ_t converges (for instance weakly) to the unit standard Gaussian γ, whose density is given by $(2\pi)^{-d/2}\exp(-|x|^2/2)$. This can be proved with many different techniques from several points of view, see for instance [1] or alternatively in [3].

Equation (1) can be written in the form

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (\mu v) = 0 \tag{2}$$

where $v = v(t,x)$ is given by

$$v(t,x) = -x - \nabla \log \mu_t(x) \tag{3}$$

and can be interpreted as the Wasserstein velocity of the curve $(\mu_t)_{t\geq 0}$ [1, Chapter 8]. Here and in the sequel, we tacitly assume that all measures have a density with respect to the Lebesgue measure and we identify a measure with said density.

We consider the flow of the (non autonomous) ODE generated by v, whose well-posedness is justified below. That is, we define the function $S : \mathbb{R}_{\geq 0} \times \mathbb{R}^d \to \mathbb{R}^d$ by

$$\begin{cases} \dfrac{\partial S}{\partial t} = v(t,S) = -S - \nabla \log \mu_t(S), \\ S(0,\cdot) = \mathrm{Id}. \end{cases} \tag{4}$$

DEPARTMENT OF DECISION SCIENCES AND BIDSA, BOCCONI UNIVERSITY, 20136 MILANO, ITALY
INSTITUT CAMILLE JORDAN, UNIVERSITÉ CLAUDE BERNARD LYON 1, 69622 VILLEURBANNECEDEX, FRANCE
E-mail addresses: hugo.lavenant@unibocconi.it, santambrogio@math.univ-lyon1.fr.
Date: May 30, 2022.
2020 *Mathematics Subject Classification.* 49Q22, 35B40, 34E99.

We use the shortcut S_t for $S(t,\cdot)$. Thanks to (2), we know that $S_t\#\mu_0 = \mu_t$ [10, Theorem 4.4]. Here $T\#\nu$, for a measurable map T and a measure ν, is the push-forward measure defined by $(T\#\nu)(A) = \nu(T^{-1}(A))$ for any Borel set A.

Now let us consider the limit $t \to +\infty$. The exponential decay of v in terms of t, see (6) below, makes it clear that all trajectories $S_t(x)$ do admit a limit $S_\infty(x)$ when $t \to \infty$. By continuity, the map S_∞ satisfies $S_\infty\#\mu_0 = \gamma$.

If we were to restrict to a bounded domain without drift in the Fokker Planck equation (1) the existence of S_∞ has also been established and such a technique has been used to find smooth maps pushing forward an arbitrary distribution μ_0 on the uniform distribution [2, 5]. We consider the following conjecture.

Conjecture 1. *The map S_∞ is the optimal transport map between μ_0 and γ, that is, solves the problem*

$$\min_T \left\{ \int_{\mathbb{R}^d} |T(x) - x|^2 \, \mathrm{d}\nu_1(x), \quad T \ \ \textit{Borel map such that} \ \ T\#\nu_1 = \nu_2 \right\} \tag{5}$$

for $(\nu_1, \nu_2) = (\mu_0, \gamma)$.

The goal of this note is to prove that generically this conjecture cannot hold. It cannot hold for all measures μ_0 having a smooth and fast decaying density with respect to the Lebesgue measure.

Comparison with a previous counterexample. In [11], that was pointed out to us after a first version of these notes was released, the author proves a similar result. In the setting of [11], the drift $-x$ in the Fokker-Planck equation (1) is replaced by $-Ax$ for a symmetric positive definite matrix A, so that the equilibrium measure γ is now a Gaussian with covariance matrix A^{-1}. Then, the author shows that Conjecture 1 cannot hold, even if one restricts the initial measure μ to Gaussian distributions. The argument of the author is quite similar to ours: she also exploits that if Conjecture (1) holds for all reasonable μ, then by the semi group property the map $S_\infty \circ S_t^{-1}$ is an optimal transport map for any $t \geq 0$.

On the other hand, the author restricts to Gaussian measures, heavily relying on explicit formulas for the optimal transport available in this case [7], and the counterexample relies on the non-commutativity between the covariance matrix of the initial measure μ, and the one of the equilibrium measure γ. Thus the argument fails if γ is the standard Gaussian as in [8] or in the present note. The author of [11] then suggests that a counterexample should also exist when the equilibrium measure is the standard Gaussian and shows it via a numerical computation, but no error estimate is presented in order to turn this intuition into a rigorous proof.

We argue that our article brings a rigorous proof for a stronger counterexample (as γ is the isotropic standard Gaussian), and also simplifies the structure of the counterexample by not relying on explicit formulas for Gaussian measures.

Bounds on v and well-posedness of the flow. Let us justify briefly that S is well defined and S_∞ indeed exists. By parabolic regularity the vector field v is smooth, but it also satisfies a bound depending on the initial datum: more precisely, we have

$$|v(t,x)| \leq e^{-t} ||v(0,\cdot)||_{L^\infty}. \tag{6}$$

This is a well-known fact about the Fokker-Planck equation, and the exponential decay depends on the uniform convexity of the potential $V(x) = |x|^2/2$ which characterizes the limit measure γ as $\gamma = ce^{-V}$. In order to see the validity of this estimate, the reader has the choice among many possible strategies. For instance in [6] the same estimate is proven for the Fokker-Planck flow on bounded convex domains deducing it from the discrete JKO approximation of Wasserstein gradient flows, but since no constant depends on the diameter of the domain it is easily seen that the same estimate is true on the whole space. It is also possible to directly deduce this exponential bound from the Bakry-Émery theory [3] for the

Ornsteil-Uhlenbeck operator, or to write a differential inequality satisfied by $w = |v|^2$. Indeed, taking in account that v is always a gradient field, we can find

$$\frac{\partial w}{\partial t} \leq \Delta w + \nabla \log \mu \cdot \nabla w - 2v.$$

By the maximum principle, assuming $v(0,\cdot) \in L^\infty(\mathbb{R}^d)$ then a simple comparison with a constant initial datum shows $w(t,x) \leq e^{-2t}||w(0,\cdot)||_{L^\infty}$.

Motivation. Though it may be surprising that it could be the true, Conjecture 1 was recently made in [8]. In such work, the authors make the link between S_∞ and "Denoising diffusion probabilistic models" (DDPMs) used in machine learning and state Conjecture 1. They prove it in the particular case where μ_0 is the law of a Gaussian vector, and they provide numerical evidence supporting the conjecture in two dimensions.

Moreover, in an older article [9], an other group of authors studied a generalized version of the map S_∞ (where the target measure γ is not only the standard Gaussian) and, though they wrote that Conjecture 1 is likely to be false, they did not find a concluding argument. In such work they were interested in the theoretical properties of the map S_∞ when both μ_0 and γ are log concave measures.

Our counterexample will be given as follows.

Proposition 2. *In $\mathbb{R}^2$, take a smooth and compactly supported function φ and define*

$$u(x) = \frac{|x|^2}{2} + \varepsilon\varphi(x).$$

Then for ε small enough u is convex, ∇u is a C^∞ diffeomorphism which coincides with the identity outside of a compact set. Moreover, we can choose φ such that, if we define $\mu_0 = (\nabla u)^{-1}\#\gamma$ and consider the curve $(\mu_t)_{t\geq 0}$ given by (1)*, then Conjecture 1 cannot hold when replacing μ_0 with any μ_t, $t \geq 0$.*

Note that we cannot exactly disprove Conjecture 1 for $\mu_0 = (\nabla u)^{-1}\#\gamma$, but for one of the μ_t for $t \geq 0$. However, all the μ_t have a smooth density, which is log concave [9, Theorem 1.2], and decays exponentially fast at infinity.

Proposition 2 is the main result of this note, and the rest of the present work is dedicated to its proof.

Notations and some results on optimal transport. For a function $u : \mathbb{R}^d \to \mathbb{R}$, we denote by ∇u its gradient and D^2u its Hessian matrix. Moreover, for a map $T : \mathbb{R}^d \to \mathbb{R}^d$, its Jacobian matrix is denoted by DT. We will denote by $S_d(\mathbb{R})$ the set of symmetric $d \times d$ matrices.

Definition 3. *Let ν_1, ν_2 two measures with quadratic second moments on $\mathbb{R}^d$. The map $T : \mathbb{R}^d \to \mathbb{R}^d$ is said to be an optimal transport map if it solves the problem* (5).

Optimal transport is a wide and rich theory, we refer to [1, 10] and references therein. For the purposes of this note, we rely only on the following characterization of optimal transport maps, at least in the smooth case.

Theorem 4 (Brenier's theorem, see [4] and, for instance Theorems 1.22, 1.48 and Section 1.7.6 in [10])**.** *Let ν_1, ν_2 be two probability measures both having density with respect to the Lebesgue measure, finite second moments, and supported on the whole $\mathbb{R}^d$. We also take $T : \mathbb{R}^d \to \mathbb{R}^d$ a smooth map such that $T\#\nu_1 = \nu_2$.*

The map T is optimal for Problem (5) *if and only if there exists a convex function $u : \mathbb{R}^d \to \mathbb{R}$ such that $T = \nabla u$. It is equivalent to DT being a symmetric semi positive definite matrix everywhere on $\mathbb{R}^d$.*

Moreover, if it is the case then u must satisfy the Monge-Ampère equation

$$\det D^2u(x) = \frac{\nu_1(x)}{\nu_2(\nabla u(x))} \tag{7}$$

everywhere on $\mathbb{R}^d$.

Note that in Conjecture 1 the optimality of S_t between μ_0 and μ_t is not addressed, which makes it difficult to use differential methods differentiating in time the condition for being optimal at each time t (i.e. being the gradient of a convex function). Indeed, should DS_t be a symmetric matrix for any t and any x, then it would have real eigenvalues, always different from 0 because of the determinant condition (7) which characterizs $S_t\#\mu_0 = \mu_t$, and these eigenvalues would be striclty positive by continuity since they are equal to 1 at $t = 0$. This would imply the optimality of S_t and is what actually happens in the Gaussian case studied in [8]. On the other hand [9] shows that generically DS_t is non-symmetric. However, it does not rule out the possibility that DS_∞ could be symmetric even if DS_t is not for any $t > 0$, and that makes the understanding of this conjecture much more complicated.

Yet, we will see that a different point of view on Conjecture 1 is possible and allows for a differential approach.

This different point of view will partially rely on the inversion of optimal transport maps. Indeed, we also need the following result: optimal transport map are stable by "inversion".

Proposition 5. *Let ν_1, ν_2 be two probability measures both having density with respect to the Lebesgue measure, finite second moments, and supported on the whole $\mathbb{R}^d$.*

Then a smooth diffeormorphism $T : \mathbb{R}^d \to \mathbb{R}^d$ is the optimal transport from ν_1 onto ν_2 if and only if T^{-1} is the optimal transport from ν_2 onto ν_1.

Proof. See Remark 1.20 in [10]. Alternatively, with such regularity and the help of Theorem 4, we can say that the matrix field DT is made of symmetric semi positive definite matrices if and only if the matrix field $(DT)^{-1}$ is made of symmetric semi positive definite matrices. □

A necessary condition for the answer to be positive. Let us derive a necessary condition for the answer to Conjecture 1 to be positive. We take μ_0 sufficiently smooth and quickly decaying at infinity, and we assume that Conjecture 1 holds, not only for μ_0, but for all $(\mu_t)_{t\geq 0}$. If we start from such μ_0, all the computations in the sequel are justified as, by parabolic regularity, all the functions are smooth enough and decay sufficiently fast at infinity.

In particular, it implies that for all $t \geq 0$, following the flow of v between instant t and $+\infty$ also yields an optimal transport map between μ_t and γ. That is, the map $S_\infty \circ S_t^{-1}$ is an optimal transport map between μ_t and γ. Using Proposition 5, its inverse $S_t \circ S_\infty^{-1}$ is also an optimal transport map between γ and μ_t. Let us denote by T the map S_∞^{-1}. Using Theorem 4, we see that

$$\forall t \geq 0, \qquad \text{the Jacobian matrix of } S_t \circ T \text{ belongs to } S_d(\mathbb{R}) \text{ at every point } x. \tag{8}$$

The Jacobian matrix reads $DS_t(T)DT$. Moreover, differentiating (4) with respect to x, we see that

$$\frac{\partial DS_t}{\partial t} = -DS_t - D^2 \log \mu_t(S_t) DS_t$$

together with $DS_0 = \mathrm{Id}$. Thus, differentiating (8) with respect to time,

$$\forall t \geq 0, \forall x \in \mathbb{R}^d, \qquad \left[-DS_t(T(x)) - D^2 \log \mu_t(S_t(T(x)))DS_t(T(x))\right] DT(x) \in S_d(\mathbb{R}).$$

We then evaluate this conclusion at $t = 0$: using that S_0 is the identity mapping we find that

$$\forall x \in \mathbb{R}^d, \quad \left[-\mathrm{Id} - D^2 \log \mu_0(T(x))\right] DT(x) \in S_d(\mathbb{R}).$$

Now, our goal is that $T = S_\infty^{-1}$ should be an optimal transport map, so that DT is a symmetric matrix. We conclude that the matrix fields $D^2 \log \mu_0(T)$ and DT should commute everywhere on $\mathbb{R}^d$. Composing with T^{-1} on the right, using $DT(T^{-1}) = [DS_\infty]^{-1}$, and that a symmetric matrix A commutes with an invertible symmetric matrix B if and only if it commutes with B^{-1}, we conclude that:

$$\forall x \in \mathbb{R}^d, \quad D^2 \log \mu_0(x) \text{ commutes with } DS_\infty(x),$$
$$\text{being } S_\infty \text{ the optimal transport map between } \mu_0 \text{ and } \gamma.$$

We know, thanks to Theorem 4, that $S_\infty = \nabla u$ for some convex function u which satisfies (taking the logarithm of the Monge-Ampère equation (7)):

$$\log\det D^2 u(x) = \log \mu_0(x) - \log \gamma(\nabla u(x)) = \log \mu_0(x) + \frac{1}{2}|\nabla u(x)|^2 + \text{const}.$$

As $DS_\infty = D^2 u$ we deduce that necessarily

$$\forall x \in \mathbb{R}^d, \quad D^2\left[\log\det D^2 u(x) - \frac{1}{2}|\nabla u(x)|^2\right] \text{ commutes with } D^2 u(x), \tag{9}$$
$$\text{being } \nabla u \text{ the optimal transport map between } \mu_0 \text{ and } \gamma.$$

Inspecting the expression, the matrix on the left depends of derivatives up to order 4 of u, while $D^2 u$ only contains second derivatives. Thus it could be surprising that (9) holds generically.

Building the counterexample. Now, assume that we can find a convex function u such that

$$D^2\left[\log\det D^2 u(x) - \frac{1}{2}|\nabla u(x)|^2\right] \text{ does not commute with } D^2 u(x) \text{ for some } x \in \mathbb{R}^d. \tag{10}$$

Then, defining $\mu_0 = (\nabla u)^{-1}\#\gamma$, we know, thanks to Theorem 4 and Proposition 5, that ∇u is the optimal transport between μ_0 and γ, and so (9) fails.

So to conclude the proof of Proposition 2 we will provide a function u such that (10) holds. It is not completely obvious: the reader can check that actually (9) holds in the following cases.

- If u is a convex quadratic function (which makes ∇u linear): it is consistent with [9] and [8] which prove Conjecture 1 if μ_0 is a Gaussian measure (in such case the optimal transport map is indeed linear [7]).
- If μ_0 has a radial symmetry, and so does u. Again [9] mentions that in this case Conjecture 1 holds.
- If u is separable, that is $u(x_1, x_2, \dots, x_d) = u_1(x_1) + u_2(x_2) + \dots u_d(x_d)$. With such symmetries, property (9) holds.

As we announced, our proposal it to take u of the form

$$u(x) = \frac{|x|^2}{2} + \varepsilon\varphi(x)$$

for a smooth compactly supported function φ and ε small, that is, to linearize around the identity. If ε is small enough, then u is convex and ∇u is a C^∞ diffeomorphism which coincides with the identity outside of a compact set.

On the one hand $D^2 u = \mathrm{Id} + \varepsilon D^2\varphi$, while if we do a Taylor expansion at order 1, as the differential of the determinant around Id is the trace operator:

$$\begin{aligned} D^2\left[\log\det D^2 u(x) - \frac{1}{2}|\nabla u(x)|^2\right] &= D^2\left[\log(1+\varepsilon\Delta\varphi) - \frac{1}{2}|x|^2 - \varepsilon x\cdot\nabla\varphi\right] + o(\varepsilon) \\ &= -\mathrm{Id} + \varepsilon\left\{D^2\Delta\varphi - D^2[x\cdot\nabla\varphi]\right\} + o(\varepsilon). \end{aligned}$$

So, to get a counterexample if ε is small it is enough to take φ such that $D^2\Delta\varphi - D^2[x\cdot\nabla\varphi]$ does not commute with $D^2\varphi$, at least for one point x. We will actually concentrate at $x = 0$, when $D^2[x\cdot\nabla\varphi](0) = 2D^2\varphi(0)$ and so commutes with $D^2\varphi(0)$. Thus we are left to find a smooth function φ such that the matrices $[D^2\Delta\varphi](0)$ and $D^2\varphi(0)$ do not commute. For instance we take, in $\mathbb{R}^2$, $\varphi(x_1, x_2) = x_1x_2 + x_1^4$ in such a way that

$$D^2\varphi = \begin{pmatrix} 12x_1^2 & 1 \\ 1 & 0 \end{pmatrix}, \qquad D^2\Delta\varphi = \begin{pmatrix} 24 & 0 \\ 0 & 0 \end{pmatrix},$$

and these two matrices do not commute when evaluated at $(x_1, x_2) = (0, 0)$. The function φ is not compactly supported, but by multiplying it by a smooth cutoff function which is constant and equal to 1 on a neighborhood of $(0, 0)$ we get our claim, and thus Proposition 2.

Concluding remarks. Numerical evidences in [8] indicate that, even if Conjecture 1 fails, the map S_∞ is almost optimal. Quantifying the defect of optimality of the S_∞ could be a direction for future research.

Acknowledgments. The authors thank Young-Heon Kim, Valentin Khrulkov, Ivan Oseledets and Giuseppe Savaré for stimulating discussions and pointing them to some references. The research leading to this paper has been done during a visit of the second author to the Department of Decision Sciences of Bocconi University, which is warmly acknowledged for the hospitality. The second author also acknowledges the support of the ANR project MAGA (ANR-16-CE40-0014) and of the Lagrange Mathematics and Computation Research Center project on Optimal Transportation.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Birkhäuser Verlag, 2008.

[2] Albert Avinyo, Joan Sola-Morales, and Marta Valencia. On maps with given Jacobians involving the heat equation. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, 54(6):919–936, 2003.

[3] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

[4] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.

[5] José Antonio Carrillo and Stefano Lisini. On the asymptotic behavior of the gradient flow of a polyconvex functional. *Nonlinear partial differential equations and hyperbolic wave phenomena*, 526:37–51, 2010.

[6] Vincent Ferrari and Filippo Santambrogio. Lipschitz estimates on the JKO scheme for the Fokker-Planck equation on bounded convex domains. *Appl. Math. Lett.*, 112:Paper No. 106806, 7, 2021.

[7] Matthias Gelbrich. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990.

[8] Valentin Khrulkov and Ivan Oseledets. Understanding DDPM Latent Codes Through Optimal Transport. *arXiv preprint arXiv:2202.07477*, 2022.

[9] Young-Heon Kim and Emanuel Milman. A generalization of Caffarelli's contraction theorem via (reverse) heat flow. *Mathematische Annalen*, 354(3):827–862, 2012.

[10] Filippo Santambrogio. Optimal transport for applied mathematicians. *Progress in Nonlinear Differential Equations and their applications*, 87, 2015.

[11] Anastasiya Tanana. Comparison of transport map generated by heat flow interpolation and the optimal transport Brenier map. *Communications in Contemporary Mathematics*, 23(6), 2021.