HAL
open science

Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm

Jean-David Benamou, Guillaume Carlier, Luca Nenna

To cite this version:

Jean-David Benamou, Guillaume Carlier, Luca Nenna. Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm. Numerische Mathematik, 2018. hal-01621311v2

HAL Id: hal-01621311

https://hal.science/hal-01621311v2

Submitted on 5 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm

Jean-David Benamou [*] [†] Guillaume Carlier [†] [*]
Luca Nenna [‡]

March 5, 2018

Abstract

Starting from Brenier's relaxed formulation of the incompressible Euler equation in terms of geodesics in the group of measure-preserving diffeomorphisms, we propose a numerical method based on Sinkhorn's algorithm for the entropic regularization of optimal transport. We also make a detailed comparison of this entropic regularization with the so-called Bredinger entropic interpolation problem (see [1]). Numerical results in dimension one and two illustrate the feasibility of the method.

Keywords: Incompressible Euler equations, generalized incompressible flows, multi-marginal optimal transport, entropic regularization, Sinkhorn algorithm.

MS Classification: 76M30, 65K10.

1 Introduction

The motion of incompressible inviscid fluids inside a bounded domain $\mathcal{D} \subset \mathbb{R}^d$ (or, as we shall often consider the periodic in space case, $\mathcal{D} = \mathbb{T}^d := \mathbb{R}^d/2\pi\mathbb{Z}^d$

[*]INRIA-Paris, MOKAPLAN, rue Simone Iff, 75012, Paris, France `Jean-David.Benamou@inria.fr`.

[†]Université Paris-Dauphine, PSL Research University, CEREMADE (UMR CNRS 7534), 75016 Paris, France `carlier@ceremade.dauphine.fr`

[‡]CNRS and Université Paris-Dauphine, PSL Research University, CEREMADE (UMR CNRS 7534), 75016 Paris, France `nenna@ceremade.dauphine.fr`

is the flat torus) without the action of external forces is governed by the equations introduced by Euler in 1755 [14]:

$$\begin{cases} \partial_t u + (u \cdot \nabla)u + \nabla p = 0 & \text{in } (0,T) \times \mathcal{D} \\ \operatorname{div}(u) = 0 & \text{in } (0,T) \times \mathcal{D} \\ u \cdot n = 0 & \text{on } (0,T) \times \partial\mathcal{D}, \end{cases} \tag{1.1}$$

where n denotes the unit normal to $\partial\mathcal{D}$, u denotes the velocity field and p is the pressure. As was first emphasized by Arnold [2], also see Arnold and Khesin [3], (1.1) can be seen, at least formally, in Lagrangian coordinates as the Euler-Lagrange for the minimization of the action

$$\mathcal{A}(X) := \int_0^T \|\dot{X}\|^2_{L^2(\mathcal{D})} \mathrm{d}t \tag{1.2}$$

subject to the constraint that $t \mapsto X(t,.)$ is a path in Sdiff, the group of Lebesgue-measure preserving diffeomorphisms of $\mathcal{D}$. Indeed, the incompressibility constraint translates in Eulerian terms as the requirement that the velocity field u associated with X, through $\partial_t X(t,x) = u(t, X(t,x))$ is divergence-free. The pressure p acts as a Lagrange multiplier for this constraint and the optimality equation for the minimization of $\mathcal{A}$ on paths constrained to remain in Sdiff leads to (1.1).

From now on, we shall consider Brenier's relaxed formulation [6], [7], [8], [9] of the minimizing geodesic problem between an initial and terminal configuration of the fluid. This formulation which allows splitting and crossing of particles, is based on the notion of generalized incompressible flow (GIF). Denoting by $\mathcal{L}$ the Lebesgue measure on $\mathcal{D}$ (normalized so as to be a probability measure on $\mathcal{D}$), by Ω the path space

$$\Omega := C([0,T], \mathcal{D})$$

and for $\omega \in \Omega$ and $t \in [0,T]$, the evaluation map at time t is defined by $e_t(\omega) := \omega(t)$, the set of generalized incompressible flows is by definition the set of probability measures Q on Ω such that $e_{t\#}Q = \mathcal{L}$ for every $t \in [0,T]$,

$$\mathrm{GIF} := \{Q \in \mathcal{P}(\Omega) \ : \ e_{t\#}Q = \mathcal{L}, \ \forall t \in [0,T]\}. \tag{1.3}$$

We are also given $\pi_{0,T} \in \mathcal{P}(\mathcal{D}\times\mathcal{D})$ a probability measure on $\mathcal{D}\times\mathcal{D}$ having $\mathcal{L}$ as marginals and which captures the joint distribution of particles at times 0 and T (one may think for instance the deterministic coupling $\pi_{0,T} := (\mathrm{id}, X_T)_\#\mathcal{L}$ where $X_T \in$ Sdiff represents the terminal Lagrangian configuration of the

fluid). The set of generalized incompressible flows compatible with $\pi_{0,T}$ is then given by

$$\mathrm{GIF}(\pi_{0,T}) := \{Q \in \mathrm{GIF} \ : \ (e_0, e_T)_\# Q = \pi_{0,T}\}. \tag{1.4}$$

For $\omega \in \Omega$ we denote by $E(\omega)$ its kinetic action:

$$E(\omega) := \begin{cases} \frac{1}{2}\int_0^T |\dot{\omega}(t)|^2 \mathrm{d}t \text{ if } \omega \in H^1((0,T),\mathcal{D}) \\ +\infty \text{ otherwise.} \end{cases} \tag{1.5}$$

Brenier's relaxation of Arnold's geodesic problem then reads as the infinite-dimensional linear-programming problem

$$\inf_{Q \in \mathrm{GIF}(\pi_{0,T})} \mathcal{E}(Q) := \int_\Omega E(\omega) \mathrm{d}Q(\omega). \tag{1.6}$$

This formulation can be viewed as an optimal transport problem with infinitely many marginal constraints corresponding to the incompressibility of the flow and an additional constraint corresponding to the prescribed joint initial/terminal distribution $\pi_{0,T}$. It is probably the first instance of the nowadays active field of multi-marginal optimal transport [25].

Mérigot and Mirebeau [22] recently produced a tractable numerical method for a non-convex Lagrangian formulation of (1.6). The marginal constraints are penalized using semi-discrete optimal transport for which fast solvers are now available (see [21], [19]).

In the present paper, we follow a different approach, based on the so-called entropic regularization which leads to a strictly convex problem. The entropic regularization approach, which goes back to Schrödinger [27] has deep connections with large deviations [13]. It has been extensively analyzed and developed by Mikami [23] and Léonard [17], [18] who in particular proved convergence of Schrödinger bridges to optimal transport geodesics as the noise intensity vanishes. It has also proved to be an efficient computational strategy for optimal transport by Cuturi [12] who made the connection with the simple but powerful Sinkhorn scaling algorithm which is equivalent to the famous iterative proportional fitting procedure (see [11], [26]), also see [5] for various applications to optimal transport.

Our aim in this paper is twofold: first showing that the entropic regularization leads to tractable simulations. We present in section 6 computations of a non classical two-dimensional GIF. Similar computations were presented recently in [22] but rely on a different relaxation and different numerical methods. Secondly, we connect this numerical scheme to the Schrödinger bridge framework which involves the relative entropy with respect to the

Wiener measure with a small variance parameter. We shall indeed show, that in the periodic case, $\mathcal{D} = \mathbb{T}^d$, our numerical approach is exactly the discretized in time counterpart of the entropic interpolation approach developed recently in [1] and reminiscent of Yasue's variational approach for Navier-Stokes equations [28].

The paper is organized as follows. After rewriting the time discretization of Brenier's problem as a multi-marginal optimal transport problem in section 2, we introduce its entropic regularization in section 3. In the periodic case, a detailed comparison with the time-discretization of Bredinger's problem[1] as well as a Γ-convergence result are given in section 4. In section 5, Sinkhorn's algorithm is described in the present setting. Finally, section 6 is devoted to numerical results in dimensions one and two.

2 Time discretization

In what follows, $\mathcal{D}$ is either a bounded convex domain of $\mathbb{R}^d$ or the flat torus, $\mathcal{D} = \mathbb{R}^d/2\pi\mathbb{Z}^d$, we denote by dist the distance on $\mathcal{D}$, that is the euclidean distance in the convex domain case and in the case of the torus:

$$\mathrm{dist}(x, y) := \inf_{k\in\mathbb{Z}^d} |x - y + 2\pi k| \ \ \forall (x, y) \in \mathbb{T}^d.$$

The path space $\Omega = C([0, T], \mathcal{D})$ is equipped with the topology of uniform convergence and $\mathcal{P}(\Omega)$ with the corresponding weak $*$ topology. Given $N \in \mathbb{N}$, $N \geq 1$, let $\mathcal{T}_N := \{k\frac{T}{N},\ k = 0, \dots, N\}$ and consider the time-discretization of Brenier's problem (1.6)

$$\inf_{Q\in \mathrm{GIF}_N(\pi_{0,T})} \mathcal{E}(Q), \tag{2.1}$$

where

$$\mathrm{GIF}_N(\pi_{0,T}) := \{Q \in \mathcal{P}(\Omega)\ :\ e_{t\#}Q = \mathcal{L},\ \forall t \in \mathcal{T}_N,\ (e_0, e_T)_\# Q = \pi_{0,T}\}.$$

It is well-known (see [6]) that both linear problems (1.6) and (2.1) admit solutions (which are not unique in general).

The discretized in-time problem (2.1) can easily be rewritten as an optimal transport problem with multi-marginal constraints as follows. Given $\mathbf{x}_N := (x_0, \cdots, x_N) \in \mathcal{D}^{N+1}$, let us denote by $\mathrm{proj}_{0,N}$ and proj_k the canonical projections:

$$\mathrm{proj}_k(\mathbf{x}_N) = x_k,\ k = 0, \cdots, N,\ \mathrm{proj}_{0,N}(\mathbf{x}_N) = (x_0, x_N).$$

[1] we adopt here the terminology of [1] where the name Bredinger is introduced as a contraction of Brenier and Schrödinger.

Defining the cost

$$c_N(\mathbf{x}_N) := \frac{N}{2T}\sum_{k=0}^{N-1}\mathrm{dist}^2(x_{k+1},x_k),\ \forall \mathbf{x}_N := (x_0,\cdots,x_N)\in\mathcal{D}^{N+1} \tag{2.2}$$

and the set of plans

$$\Gamma_N(\pi_{0,T}) := \{\gamma\in\mathcal{P}(\mathcal{D}^{N+1}) : \mathrm{proj}_{k\#}\gamma = \mathcal{L},\ k=0,...,N,\ \mathrm{proj}_{0,N\#}\gamma = \pi_{0,T}\} \tag{2.3}$$

let us consider the multi-marginal optimal transport problem:

$$\inf_{\gamma\in\Gamma_N(\pi_{0,T})}\int_{\mathcal{D}^{N+1}} c_N(\mathbf{x}_N)\mathrm{d}\gamma(\mathbf{x}_N). \tag{2.4}$$

Setting

$$P_N(\omega) := (\omega(t))_{t\in\mathcal{T}_N},\ \forall\omega\in\Omega$$

it is clear that $Q\in\mathcal{P}(\Omega)$ belongs to $\mathrm{GIF}_N(\pi_{0,T})$ if and only if $P_{N\#}Q$ belongs to $\Gamma_N(\pi_{0,T})$. Moreover, since

$$c_N(x_0,\cdots,x_N) = \inf\Big\{E(\omega)\ :\ \omega\in\Omega,\ P_N(\omega)=(x_0,\cdots,x_N)\Big\},$$

we easily deduce the following:

Lemma 2.1. *Problems (2.1) and (2.4) are equivalent in the sense that* $Q_N\in\mathrm{GIF}_N(\pi_{0,T})$ *solves (2.1) if and only if* $P_{N\#}Q_N$ *solves (2.4) and*

$$c_N(P_N(\omega)) = E(\omega),\ \textit{for } Q_N\textit{-a.e. } \omega.$$

Finally, let us emphasize that (2.1) approximates (1.6) in the sense of Γ-convergence. Let us denote by $\chi_{\mathrm{GIF}_N(\pi_{0,T})}$ and $\chi_{\mathrm{GIF}(\pi_{0,T})}$ the characteristic function of $\mathrm{GIF}_N(\pi_{0,T})$ and $\mathrm{GIF}(\pi_{0,T})$ respectively i.e. for $Q\in\mathcal{P}(\Omega)$,

$$\chi_{\mathrm{GIF}_N(\pi_{0,T})}(Q) = \begin{cases} 0 \text{ if } Q\in\mathrm{GIF}_N(\pi_{0,T}),\\ +\infty \text{ otherwise,}\end{cases},$$

$$\chi_{\mathrm{GIF}(\pi_{0,T})}(Q) = \begin{cases} 0 \text{ if } Q\in\mathrm{GIF}(\pi_{0,T}),\\ +\infty \text{ otherwise,}\end{cases},$$

we indeed have (we refer to chapter 6 of [24] for a proof):

Proposition 2.2. *The sequence of functionals on* $\mathcal{P}(\Omega)$ *(endowed with the weak* * *topology),* $\mathcal{E}+\chi_{\mathrm{GIF}_N(\pi_{0,T})}$ Γ*-converges as* $N\to+\infty$ *to* $\mathcal{E}+\chi_{\mathrm{GIF}(\pi_{0,T})}$.

Since $\mathcal{E}$ has relatively compact sublevel sets in Ω, the previous result in particular implies convergence of minimizers of (2.1) to minimizers of (1.6).

3 Entropy minimization

The equivalent linear programming problems (2.4) and (2.1) are extremely costly to solve numerically and a natural strategy, which has received a lot of attention recently, is to approximate these problems by strictly convex ones by adding an entropic penalization. First, let us set a few notations, given a Polish space X, and two probability measures on X, q and r the relative entropy of q with respect to r (a.k.a. Kullback-Leibler divergence) is given by

$$\mathrm{H}(q|r) := \begin{cases} \int_X \log\left(\frac{dq}{dr}\right)\mathrm{d}q \text{ if } q \ll r \\ +\infty \text{ otherwise} \end{cases}$$

where $\frac{dq}{dr}$ stands for the Radon-Nikodym derivative of q with respect to r. If $X = \mathcal{D}^m$ and $\mu \in \mathcal{P}(\mathcal{D}^m)$ we shall simply denote by $\mathrm{Ent}(\mu)$ the relative entropy of μ with respect to $\mathcal{L}^{\otimes m}$; slightly abusing notation, when $\mu \ll \mathcal{L}^{\otimes m}$ we shall identify μ with its density and simply write

$$\mathrm{Ent}(\mu) = \frac{1}{|\mathcal{D}|^m}\int_{\mathcal{D}^m} \log(\mu(x_1, \cdots, x_m))\mu(x_1, \cdots, x_m)\mathrm{d}x_1 \cdots \mathrm{d}x_m$$

where $|\mathcal{D}|$ denotes the Lebesgue measure of $\mathcal{D}$.

A first way to perform an entropic regularization of (2.4) is, given a small parameter $\varepsilon > 0$, to replace (2.4) by

$$\inf_{\gamma \in \Gamma_N(\pi_{0,T})} \int_{\mathcal{D}^{N+1}} c_N \ \mathrm{d}\gamma + \varepsilon\, \mathrm{Ent}(\gamma) \tag{3.1}$$

Note that for the previous problem to make sense, i.e. for the existence of at least one $\gamma \in \Gamma_N(\pi_{0,T})$ such that $\mathrm{Ent}(\gamma) < +\infty$ it is necessary and sufficient that

$$\mathrm{Ent}(\pi_{0,T}) < +\infty. \tag{3.2}$$

This, in particular, rules out the case[2] where $\pi_{0,T} = (\mathrm{id}, X_T)_\#\mathcal{L}$ with $X_T \in$ Sdiff. Defining the Gibbs measure on $\mathcal{D}^{N+1}$ associated to the cost c_N:

$$\eta_{N,\varepsilon}(\mathbf{x}_N) := \Lambda_{N,\varepsilon} \exp\Big(-\frac{c_N(\mathbf{x}_N)}{\varepsilon}\Big)$$

where $\Lambda_{N,\varepsilon}$ is the normalizing constant which makes $\eta_{N,\varepsilon}$ be a probability density, (3.1) can be rewritten as the Kullback-Leibler projection problem of $\eta_{N,\varepsilon}$ onto $\Gamma_N(\pi_{0,T})$:

$$\inf_{\gamma \in \Gamma_N(\pi_{0,T})} \varepsilon\, \mathrm{H}(\gamma|\eta_{N,\varepsilon}). \tag{3.3}$$

[2]However, as we shall see, one way to overcome this problem is by penalizing in the cost the condition $x_N = X_T(x_0)$.

It will follow from the proof of Theorem 4.3 (also see remark 4.4) that in the periodic case $\mathcal{D} = \mathbb{T}^d$, the finite entropy condition (3.2) guarantees the Γ-convergence of (3.1) to (2.4) as $\varepsilon \to 0$.

4 Comparison with Bredinger in the periodic case

4.1 Bredinger's problem and its time discretization

Another way to approximate the problem is to introduce some noise at the level of the path space i.e. in (1.6) as was done recently in [1][3] as follows. Throughout this section, we assume that $\mathcal{D} = \mathbb{T}^d := \mathbb{R}^d/2\pi\mathbb{Z}^d$. Assuming (3.2) and given a small parameter $\varepsilon > 0$ as before, let $R_\varepsilon \in \mathcal{P}(\Omega)$ be defined by

$$R_\varepsilon = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} \mathrm{Law}(x + \sqrt{\varepsilon} B_\varepsilon) \mathrm{d}x \tag{4.1}$$

where B_ε is the standard Brownian motion starting at 0 (that is the Markov process whose generator is $\frac{1}{2}\Delta$) on $\frac{1}{\sqrt{\varepsilon}}\mathbb{T}^d$. Following [1], we shall call

$$\inf_{Q \in \mathrm{GIF}(\pi_{0,T})} \mathrm{H}(Q|R_\varepsilon) \tag{4.2}$$

the Bredinger problem (with variance parameter ε). Let us now discretize in time the Bredinger problem (4.2) in a similar way as we deduced (2.1) from (1.6) i.e. consider

$$\inf_{Q \in \mathrm{GIF}_N(\pi_{0,T})} \mathrm{H}(Q|R_\varepsilon). \tag{4.3}$$

It is worth noting here that $R_\varepsilon \in \mathrm{GIF}$. Following Léonard [17], one can reduce (4.3) to an entropy minimization problem over $\Gamma_N(\pi_{0,T})$ as follows. Let us set

$$\theta_{N,\varepsilon} := P_{N\#} R_\varepsilon \tag{4.4}$$

and disintegrate R_ε with respect to $\theta_{N,\varepsilon}$ as

$$R_\varepsilon = \int_{(\mathbb{T}^d)^{N+1}} R_\varepsilon^{\mathbf{x}_N} \mathrm{d}\theta_{N,\varepsilon}(\mathbf{x}_N),$$

so that $R_\varepsilon^{\mathbf{x}_N}$ is the law of a Brownian bridge i.e. the law of a Brownian path conditional to the fact that its values at times in $\mathcal{T}_N$ are given by $\mathbf{x}_N$. In

[3] in connection with Navier-Stokes.

a similar way, given $Q \in \mathrm{GIF}_N(\pi_{0,T})$, setting $\gamma := P_{N\#}Q \in \Gamma_N(\pi_{0,T})$ and disintegrating Q with respect to γ:

$$Q = \int_{(\mathbb{T}^d)^{N+1}} Q^{\mathbf{x}_N} \mathrm{d}\gamma(\mathbf{x}_N),$$

we have

$$\mathrm{H}(Q|R_\varepsilon) = \mathrm{H}(\gamma|\theta_{N,\varepsilon}) + \int_{(\mathbb{T}^d)^{N+1}} \mathrm{H}(Q^{\mathbf{x}_N}|R_\varepsilon^{\mathbf{x}_N})\mathrm{d}\gamma(\mathbf{x}_N) \geq \mathrm{H}(\gamma|\theta_{N,\varepsilon})$$

with equality if and only if $Q^{\mathbf{x}_N} = R_\varepsilon^{\mathbf{x}_N}$ for γ-almost every $\mathbf{x}_N$. Hence we get the following entropic analogue of Lemma 2.1:

Proposition 4.1. *Let $Q^\star \in \mathrm{GIF}_N(\pi_{0,T})$ then $Q^\star$ solves (4.3) if and only if $\gamma^\star := P_{N\#}Q$ solves*

$$\inf_{\gamma \in \Gamma_N(\pi_{0,T})} \mathrm{H}(\gamma|\theta_{N,\varepsilon}) \tag{4.5}$$

and $Q^\star$ disintegrates with respect to $\gamma^\star := P_{N\#}Q^\star$ as

$$Q^\star = \int_{\mathbb{T}^{dN+1}} R_\varepsilon^{\mathbf{x}_N} \, d\gamma^\star(\mathbf{x}_N). \tag{4.6}$$

Proof. By strict convexity, (4.3) and (4.5) admit at most one solution. Thanks to

$$\mathrm{H}(Q|R_\varepsilon) = \mathrm{H}(\gamma|\theta_{N,\varepsilon}) + \int_{(\mathbb{T}^d)^{N+1}} \mathrm{H}(Q^{\mathbf{x}_N}|R_\varepsilon^{\mathbf{x}_N})\mathrm{d}\gamma(\mathbf{x}_N) \tag{4.7}$$

where $\gamma := P_{N\#}Q$ and the fact that $Q \in \mathrm{GIF}_N(\pi_{0,T})$ if and only if $\gamma := P_{N\#}Q \in \Gamma_N(\pi_{0,T})$ the minimization problem (4.3) can be rewritten as

$$\inf_{\gamma \in \Gamma_N(\pi_{0,T})} \left\{ \inf \left\{ \mathrm{H}(Q|R_\varepsilon) \mid Q \in \mathcal{P}(\Omega),\ P_{N\#}Q = \gamma \right\} \right\}. \tag{4.8}$$

With (4.7), the inner minimization is uniquely solved when $Q^{\mathbf{x}_N} = R_\varepsilon^{\mathbf{x}_N}$ for $\gamma-$almost every $\mathbf{x}_N \in \mathbb{T}^{dN+1}$ since $\mathrm{H}(Q^{\mathbf{x}_N}|R_\varepsilon^{\mathbf{x}_N}) = 0$ is the minimal value of the relative entropy. Therefore, for each $\gamma \in \Gamma_N(\pi_{0,T})$ we have

$$\inf \left\{ \mathrm{H}(Q|R_\varepsilon) \mid Q \in \mathcal{P}(\Omega),\ P_{N\#}Q = \gamma \right\} = \mathrm{H}(Q^\gamma|R_\varepsilon^{\mathbf{x}_N}) = \mathrm{H}(\gamma|\theta_{N,\varepsilon}),$$

where

$$Q^\gamma = \int_{(\mathbb{T}^d)^{N+1}} R_\varepsilon^{\mathbf{x}_N} \mathrm{d}\gamma(\mathbf{x}_N),$$

and the solution of (4.3) is

$$Q = Q^{\gamma^\star}$$

where $\gamma^\star$ is the unique solution of (4.5). □

Now, we see that (4.5) leads to another entropy regularized optimal transport problem, which can equivalently be rewritten as

$$\inf_{\gamma\in\Gamma_N(\pi_{0,T})} \int_{(\mathbb{T}^d)^{N+1}} c_{N,\varepsilon}\ \mathrm{d}\gamma + \varepsilon\,\mathrm{Ent}(\gamma) \text{ with } c_{N,\varepsilon} := -\varepsilon\log(\theta_{N,\varepsilon}). \tag{4.9}$$

The difference between the *naive* regularization (3.1) and (4.9) is that the cost which comes from the discretization of Bredinger's problem $c_{N,\varepsilon}$ is related to the heat kernel and not directly to the initial quadratic distance cost. We shall compare both costs and give a convergence result in the next paragraph. We conclude this section by summarizing in the following table all the variational models we have seen so far.

Brenier's problem
$$\inf_{Q\in\mathrm{GIF}(\pi_{0,T})} \mathcal{E}(Q)$$ where $$\mathcal{E}(Q) := \int_\Omega E(\omega)\mathrm{d}Q(\omega).$$
Discrete Brenier problem
$$\inf_{Q\in\mathrm{GIF}_N(\pi_{0,T})} \mathcal{E}(Q) \Longleftrightarrow \inf_{\gamma\in\Gamma_N(\pi_{0,T})} \int_{(\mathbb{T}^d)^{N+1}} c_N(\mathbf{x}_N)\mathrm{d}\gamma(\mathbf{x}_N).$$
Bredinger's problem
$$\inf_{Q\in\mathrm{GIF}(\pi_{0,T})} \mathrm{H}(Q\vert R_\varepsilon)$$
Discrete Bredinger problem
$$\inf_{Q\in\mathrm{GIF}_N(\pi_{0,T})} \mathrm{H}(Q\vert R_\varepsilon) \Longleftrightarrow \inf_{\gamma\in\Gamma_N(\pi_{0,T})} \mathrm{H}(\gamma\vert\theta_{N,\varepsilon}).$$

4.2 Convergence as noise vanishes

Our goal now is to establish (for fixed N) a convergence result as $\varepsilon \to 0$. In the first place, one needs to connect $c_{N,\varepsilon}$ and c_N which can be done thanks to classical Gaussian estimates for the heat kernel.

First we observe that the kernel $\theta_{N,\varepsilon}$ from (4.1)-(4.4) can be computed as

follows. First let us denote by p_t the heat kernel on $\mathbb{R}^d$:

$$p_t(z) := \frac{1}{(2\pi t)^{\frac{d}{2}}} \exp\Big(- \frac{|z|^2}{2t}\Big),\ t > 0,\ z \in \mathbb{R}^d, \tag{4.10}$$

the heat-kernel on $\mathbb{T}^d$ is obtained by its $2\pi\mathbb{Z}^d$ periodization:

$$g_t(x) := (2\pi)^d \sum_{k\in\mathbb{Z}^d} p_t(x + 2k\pi),\ x \in \mathbb{T}^d \simeq [-\pi, \pi]^d, \tag{4.11}$$

and the heat kernel on $\frac{1}{\sqrt{\varepsilon}}\mathbb{T}^d$ is likewise given by

$$g_t^\varepsilon(x) := (2\pi)^d \sum_{k\in\mathbb{Z}^d} p_t(x + \frac{2k\pi}{\sqrt{\varepsilon}}),\ x \in \frac{1}{\sqrt{\varepsilon}}\mathbb{T}^d. \tag{4.12}$$

Since

$$\varepsilon^{-\frac{d}{2}} g_t^\varepsilon\Big(\frac{x}{\sqrt{\varepsilon}}\Big) = g_{\varepsilon t}(x),\ x \in \mathbb{T}^d,$$

we have

$$\theta_{N,\varepsilon}(x_0, \cdots, x_N) = \varepsilon^{-\frac{dN}{2}} \prod_{k=0}^{N-1} g^\varepsilon_{\frac{T}{N}}\Big(\frac{x_{k+1} - x_k}{\sqrt{\varepsilon}}\Big) = \prod_{k=0}^{N-1} g_{\frac{\varepsilon T}{N}}(x_{k+1} - x_k)$$

so that

$$c_{N,\varepsilon}(x_0, \cdots, x_N) = -\varepsilon \sum_{k=0}^{N-1} \log\Big(g_{\frac{\varepsilon T}{N}}(x_{k+1} - x_k)\Big).$$

To compare this expression with the minimal quadratic cost c_N, a natural tool is the following Gaussian estimate for the heat kernel (see the self-contained notes of [20] for the case of the torus):

Theorem 4.2. *The heat kernel g_t on the torus satisfies for every $t > 0$ and $x \in \mathbb{T}^d$:*

$$\frac{\lambda}{(2\pi t)^{\frac{d}{2}}} \exp\Big(- \frac{\mathrm{dist}^2(x, 0)}{2t}\Big) \le g_t(x) \le \frac{\Lambda}{(2\pi t)^{\frac{d}{2}}} \exp\Big(- \frac{\mathrm{dist}^2(x, 0)}{2t}\Big)$$

where λ and Λ are two positive constants (depending on d).

Taking logarithms and summing over k we thus obtain

$$-N\varepsilon \log(\Lambda) + \frac{dN\varepsilon}{2} \log\Big(\frac{2\pi\varepsilon T}{N}\Big) \le c_{N,\varepsilon} - c_N \le -N\varepsilon \log(\lambda) + \frac{dN\varepsilon}{2} \log\Big(\frac{2\pi\varepsilon T}{N}\Big)$$

i.e. there is a constant M such that, for small enough ε, one has

$$|c_N - c_{N,\varepsilon}| \leq M(\varepsilon N|\log(\frac{\varepsilon T}{N})|). \tag{4.13}$$

From (4.13), we can easily deduce a Γ-convergence result as $\varepsilon \to 0$ (and N fixed) between (2.4) (equivalent to (2.1)) and its entropic regularization à la Bredinger (4.9) (equivalent to (4.3)). For $\gamma \in \mathcal{P}((\mathbb{T}^d)^{N+1})$, let us denote by $\chi_{\Gamma_N(\pi_{0,T})}$ the characteristic function of $\Gamma_N(\pi_{0,T})$ i.e.

$$\chi_{\Gamma_N(\pi_{0,T})}(\gamma) = \begin{cases} 0 \text{ if } \gamma \in \Gamma_N(\pi_{0,T}) \\ +\infty \text{ otherwise} \end{cases}$$

and define the functionals to be minimized in (2.4) and (4.9) respectively

$$J_N(\gamma) := \int_{(\mathbb{T}^d)^{N+1}} c_N \mathrm{d}\gamma + \chi_{\Gamma_N(\pi_{0,T})}(\gamma),$$

and

$$J_{N,\varepsilon}(\gamma) := \int_{(\mathbb{T}^d)^{N+1}} c_{N,\varepsilon} \mathrm{d}\gamma + \varepsilon \operatorname{Ent}(\gamma) + \chi_{\Gamma_N(\pi_{0,T})}(\gamma) = \varepsilon \operatorname{H}(\gamma|\theta_{N,\varepsilon}) + \chi_{\Gamma_N(\pi_{0,T})}(\gamma).$$

We then have

Theorem 4.3. *Let us assume the finite entropy condition (3.2), then $J_{N,\varepsilon}$, Γ-converges to J_N as $\varepsilon \to 0^+$ for the weak $*$ topology of $\mathcal{P}((\mathbb{T}^d)^{N+1})$.*

Proof. Let γ_ε converge weakly $*$ to γ, we first have to prove the Γ-liminf inequality

$$\liminf_{\varepsilon \to 0} J_{N,\varepsilon}(\gamma_\varepsilon) \geq J_N(\gamma). \tag{4.14}$$

We may assume that $\gamma_\varepsilon \in \Gamma_N(\pi_{0,T})$ for a vanishing sequence of ε otherwise there is nothing to prove. Since $\Gamma_N(\pi_{0,T})$ is weakly $*$ closed we then also have $\gamma \in \Gamma_N(\pi_{0,T})$. Using the fact that for $\gamma_\varepsilon \in \Gamma_N(\pi_{0,T})$, $\operatorname{Ent}(\gamma_\varepsilon) \geq \operatorname{Ent}(\mathcal{L}^{N+1}) = (N+1)\operatorname{Ent}(\mathcal{L})$ and the estimate (4.13) we get

$$J_{N,\varepsilon}(\gamma_\varepsilon) \geq -M(\varepsilon N|\log(\frac{\varepsilon T}{N})|) + \int_{(\mathbb{T}^d)^{N+1}} c_N \mathrm{d}\gamma_\varepsilon + (N+1)\varepsilon \operatorname{Ent}(\mathcal{L})$$

and since c_N is continuous, (4.14) immediately follows.

As for the Γ-limsup inequality, we have, given $\gamma \in \Gamma_N(\pi_{0,T})$, to find $\gamma_\varepsilon \in \Gamma_N(\pi_{0,T})$ converging weakly $*$ to γ and such that

$$\limsup_{\varepsilon \to 0} \int_{(\mathbb{T}^d)^{N+1}} c_{N,\varepsilon} \mathrm{d}\gamma_\varepsilon + \varepsilon \operatorname{Ent}(\gamma_\varepsilon) \leq \int_{(\mathbb{T}^d)^{N+1}} c_N \mathrm{d}\gamma. \tag{4.15}$$

We approximate $\gamma \in \Gamma_N(\pi_{0,T})$ by γ_ε as follows. First disintegrate γ with respect to its projection on the first and last components as

$$\gamma = \gamma^{x_0,x_N} \otimes \pi_{0,T}$$

and set

$$\gamma_\varepsilon = \gamma_\varepsilon^{x_0,x_N} \otimes \pi_{0,T} \text{ with } \gamma_\varepsilon^{x_0,x_N} = g_\varepsilon^{\otimes(N-1)} \star \gamma^{x_0,x_N}.$$

By construction $\mathrm{proj}_{0,N\#}\gamma_\varepsilon = \pi_{0,T}$ and $\mathrm{proj}_{k\#}\gamma_\varepsilon = g_\varepsilon \star \mathcal{L} = \mathcal{L}$ since $\mathcal{L}$ is invariant by the heat flow, we thus have $\gamma_\varepsilon \in \Gamma_N(\pi_{0,T})$ and γ_ε converges to γ as $\varepsilon \to 0$. To estimate $\mathrm{Ent}(\gamma_\varepsilon)$, we first observe that

$$\mathrm{Ent}(\gamma_\varepsilon) = \mathrm{Ent}(\pi_{0,T}) + \int_{\mathbb{T}^d\times\mathbb{T}^d} \mathrm{Ent}(\gamma_\varepsilon^{x_0,x_N})\mathrm{d}\pi_{0,T}(x_0,x_N)$$

and then, by construction of $\gamma_\varepsilon^{x_0,x_N}$ and the convexity of Ent

$$\begin{aligned}
\mathrm{Ent}(\gamma_\varepsilon^{x_0,x_N}) &= \mathrm{Ent}\Big(\int_{(\mathbb{T}^d)^{N-1}} \prod_{k=0}^{N-1} g_\varepsilon(x_k - \cdot)\mathrm{d}\gamma^{x_0,x_N}(x_1,\cdots,x_{N-1})\Big) \\
&\leq \int_{(\mathbb{T}^d)^{N-1}} \mathrm{Ent}(\prod_{k=0}^{N-1} g_\varepsilon(x_k - \cdot))\mathrm{d}\gamma^{x_0,x_N}(x_1,\cdots,x_{N-1}) \\
&= \mathrm{Ent}(g_\varepsilon^{\otimes(N-1)}) = (N-1)\,\mathrm{Ent}(g_\varepsilon)
\end{aligned}$$

thanks to Theorem 4.2 we have

$$\mathrm{Ent}(g_\varepsilon) \leq \log(\Lambda) - \frac{d}{2}\log(2\pi\varepsilon)$$

so, thanks to (3.2), we get $\mathrm{Ent}(\gamma_\varepsilon) = O(|\log(\varepsilon)|)$ hence $\limsup_{\varepsilon\to 0} \varepsilon\,\mathrm{Ent}(\gamma_\varepsilon) \leq 0$, thanks to (4.13) we thus deduce (4.15).

□

Remark 4.4. Of course, in the periodic case, there is also Γ-convergence as $\varepsilon \to 0^+$ if one uses c_N instead of $c_{N,\varepsilon}$ i.e. when considering (3.1) instead of (4.9). The case of a domain with boundary is less clear to us, in particular the Γ-limsup argument by convolution with the heat flow above does not work in this case.

5 Sinkhorn algorithm

Once ε and N are fixed, both entropic regularizations (3.1) (written as in (3.3)) and (4.5) can be written in a common way. It reads as a Kullback-Leibler projection on the set $\Gamma_N(\pi_{0,T})$, of a certain reference measure α whose

density is a product kernel:

$$\inf_{\gamma\in\Gamma_N(\pi_{0,T})} \mathrm{H}(\gamma|\alpha) \text{ with } \alpha(x_0,\cdots x_N) := \prod_{k=0}^{N-1} K(x_{k+1}-x_k) \tag{5.1}$$

where, in the case of (3.1), K is the Gaussian kernel

$$K(x) = \exp\Big(-\frac{N\,\mathrm{dist}^2(x,0)}{2\varepsilon T}\Big)$$

(we have omitted the normalizing constant which does play any role in the minimization problem) and in the Bredinger case (4.5), K is the heat kernel

$$K(x) = g_{\frac{\varepsilon T}{N}}(x).$$

There is a unique solution to (5.1) which is of the form

$$\gamma(x_0,\cdots,x_N) = \gamma_{a,b}(x_0,\cdots x_N) := b(x_0,x_N)\prod_{k=1}^{N-1} a_k(x_k)\alpha(x_0,\cdots x_N) \tag{5.2}$$

where $b = b(x_0,x_N)$ and $a = (a_1(x_1),\cdots,a_{N-1}(x_{N-1}))$ are positive potentials (exponentials of the Lagrange multipliers[4] associated to the marginal constraints). These positive potentials are (uniquely up to multiplicative constants with unit product) determined by the requirement that $\gamma_{a,b}\in\Gamma_N(\pi_{0,T})$ i.e. the relations

$$\pi_{0,T}(x_0,x_N) = b(x_0,x_N)\int_{\mathcal{D}^{N-1}}\prod_{k=1}^{N-1} a_k(x_k)\alpha(x_0,\cdots x_N)\mathrm{d}x_1\cdots\mathrm{d}x_{N-1}$$

and for $k=1,\cdots,N-1$

$$\frac{1}{|\mathcal{D}|} = a_k(x_k)\int_{\mathcal{D}^N}\prod_{j=1,j\neq k}^{N-1} a_j(x_j)b(x_0,x_N)\alpha(x_0,\cdots x_N)\mathrm{d}x_0\cdots\mathrm{d}x_{k-1}\mathrm{d}x_{k+1}\cdots\mathrm{d}x_N.$$

[4]this is formal since existence of Lagrange multipliers for the continuous problem cannot be taken for granted in infinite dimensions unless a demanding qualification-like assumption is met requiring that $\pi_{0,T}$ somehow lies in the interior of the domain of the relative entropy. Nevertheless, once discretized in space, the problem becomes a finite-dimensional smooth convex minimization with linear constraints so that the existence of such multipliers is guaranteed. To reduce the amount of notation here, we use the same notations for the continuous problem (5.1) as for the discretized one where integrals are replaced by finite sums.

The idea of Sinkhorn algorithm (also known as IPFP, *Iterated Proportional Fitting procedure*) is to update one multiplier so as to fit one marginal constraint at a time. The algorithm constructs inductively a sequence of measures with densities as follows. Start with

$$\gamma^{(0)} = \alpha = \gamma_{a^{(0)},b^{(0)}},\ a^{(0)} = (1, \cdots, 1),\ b^{(0)} = 1$$

and once $\gamma^{(l)} = \gamma_{a^{(l)},b^{(l)}}$ (with $a^{(l)} = (a_1^{(l)}, \cdots, a_{N-1}^{(l)})$) has been determined, compute $b^{(l+1)}$ by

$$b^{(l+1)}(x_0, x_N) = \frac{\pi_{0,T}(x_0, x_N)}{\displaystyle\int_{\mathcal{D}^{N-1}} \prod_{k=0}^{N-1} a_k^{(l)}(x_k)\alpha(x_0, \cdots x_N)\mathrm{d}x_1 \cdots \mathrm{d}x_{N-1}} \tag{5.3}$$

and then

$$a_k^{(l+1)}(x_k) = \frac{1}{|\mathcal{D}| A_k^{(l+1)}(x_k)}, k = 1, \cdots, N-1, x_k \in \mathcal{D}$$

with

$$A_1^{(l+1)}(x_1) = \int_{\mathcal{D}^N} \prod_{j=2}^{N-1} a_j^{(l)}(x_j) b^{(l)}(x_0, x_N)\alpha(x_0, \cdots x_N)\mathrm{d}x_0 \mathrm{d}x_2 \cdots \mathrm{d}x_N \tag{5.4}$$

and for $k = 2, \cdots, N-1$, setting $\mathbf{x}_{-k} := (x_0, \cdots, x_{k-1}, x_{k+1}, \cdots, x_N)$:

$$A_k^{(l+1)}(x_k) = \int_{\mathcal{D}^N} \prod_{j=1}^{k-1} a_j^{(l+1)}(x_j) \prod_{j=k+1}^{N-1} a_j^{(l)}(x_j) b^{(l)}(x_0, x_N)\alpha(x_0, \cdots x_N)\mathrm{d}\mathbf{x}_{-k} \tag{5.5}$$

Finally, set $\gamma^{(l+1)} = \gamma_{a^{(l+1)},b^{(l+1)}}$ with $a^{(l+1)} = (a_1^{(l+1)}, \cdots, a_{N-1}^{(l+1)})$. This algorithm can be viewed as Kullback-Leibler projecting α in an alternating way onto each linear marginal constraint [5]. In finite dimensions (hence for the discretized problem) it is well-known (see [4]) that this algorithm converges to the projection onto the intersection i.e. $\Gamma_N(\pi_{0,T})$.

Remark 5.1 (Implementation). The iterations of Sinkhorn might seem tedious at a first glance because of the integration against α, but due to the special form of α, these are just series of convolution with the kernel K. In the periodic case, this roughly amounts to compute efficiently Fourier coefficients. In addition, a useful property in practice is that the kernel K can be written in tensorized form

$$K(z_1, \cdots, z_d) = \prod_{j=1}^{d} k(z_j)$$

where k is a kernel in dimension one.

Remark 5.2 (Penalization of the terminal configuration). When the Lagrangian coupling between initial and final configuration is deterministic, i.e. $\pi_{0,T} := (\mathrm{id}, X_T)_\# \mathcal{L}$, it is possible to take into account this constraint as a penalization : first the constraint $\mathrm{proj}_{0,N\#}\gamma = \pi_{0,T}$ is removed from (2.3) which therefore does not depend anymore on $\pi_{0,T}$, secondly a least square penalization with a positive parameter β is added to (2.2) which becomes

$$c_{N,\beta}(\mathbf{x}_N) := \frac{N}{2T} \sum_{k=0}^{N-1} \mathrm{dist}^2(x_{k+1}, x_k) + \beta \ \mathrm{dist}^2(x_N, X_T(x_0))$$

The cost now depends on X_T and β.

The impact on Sinkhorn algorithm is small and the complexity unchanged : The Lagrange multiplier b disappears, take $b \equiv 1$ in (5.3-5.4-5.5). In the same equations, the kernel α (5.1) becomes also in (5.3-5.4-5.5)

$$\alpha_\beta(x_0, \cdots x_N) = \alpha(x_0, \cdots x_N) \ \exp\Big(- \frac{\beta |x_N - X_T(x_0)|^2}{2\varepsilon} \Big)$$

Of course, β has to be dimensionalized according to $\frac{T}{N}$.

6 Numerical results

6.1 One-dimensional experiments

We first reproduce a 1D result obtained with the method described in section 5 and taken from [5]. It is a simulation of a test case proposed (and solved with a Lagrangian method) in [6]. It provides a good warm up for the presentation of 2D results which are new.

The first test case is *not periodic*, it is set on $\mathcal{D} = [0, 1]$ equipped with the standard Euclidean distance. The final configuration is given by $X_T(x_0) = 1 - x_0$ which is not an orientation preserving diffeomorphism. Arnold problem (1.2) does not make sense since *classical* particles cannot cross.

We apply the algorithm of section 5 to Brenier GIF relaxation. The multimarginal transport plans gives a Eulerian measure of the mass movements. In order to track the underlying Lagrangian motion we represent in Figure 1 and for different times t_k the quantity

$$P_{t_k}(x_0, x_k) = \int_{\mathcal{D}^{N-2}} \gamma(x_0 \cdots x_N) \mathrm{d}x_1 \cdots \mathrm{d}x_{k-1} \mathrm{d}x_{k+1} \cdots \mathrm{d}x_N, \tag{6.1}$$

which represents the amount of mass which has traveled from x_0 to x_k between initial time and time t_k. If the solution was classical and deterministic (but it is not), it would correspond to $(\mathrm{id}, X_{t_k})_{\#}\mathcal{L}$.

The initial and final correlations $(\mathrm{id}, \mathrm{id})_{\#}\mathcal{L}$ and $(\mathrm{id}, X_T)_{\#}\mathcal{L}$ are prescribed. All marginals are the Lebesgue measure. The solution forces the mass to split and the "generalized particles" to mix and cross to minimize the kinetic energy of the GIF.

The results, displayed in figure 1 are consistent with the 1989 Lagrangian simulation of Brenier [6]. For this case we also plotted in figure 2 the the Hilbert projective metric between two consecutive iterations of Sinkhorn and for three different number of marginals. The Hilbert metric between two strictly positive vectors $p, q \in \mathbb{R}^M_{++}$ is defined as

$$\mathrm{Hil} := \log\left(\frac{\max_i \frac{p_i}{q_i}}{\min_i \frac{p_i}{q_i}}\right). \tag{6.2}$$

Since in this case $N+1$ marginals are considered and so $N+1$ Lagrange multipliers a_i, the Hilbert metric on the product space can be written as the sum of $N+1$ terms: if we consider the variables a_i at step l and $l+1$ the Hilbert metric is written as

$$\mathrm{Hil}^N_k(\otimes_{i=0}^N a_i^{(l+1)}, \otimes_{i=0}^N a_i^{(l)}) := \sum_{i=0}^N \mathrm{Hil}(a_i^{(l+1)}, a_i^{(l)}). \tag{6.3}$$

It has been showed in [15, 16] (and for a generalization of Sinkhorn in [10]) that the map defined by Sinkhorn iterations is indeed contractive in the Hilbert projective metric for the two-marginals case. As one can see in figure 2 this is still true in the multi-marginal generalization of Sinkhorn, even if it seems that the number of iterations required to convergence (for the three computations we have used Hil^N_k as stopping criteria with a tolerance $\eta = 10^{-4}$) depends on the number of marginals (or time steps) chosen.

We also present in figure 3 the numerical solution to the same problem on the flat torus $\mathbb{R}/\mathbb{Z}$. This can be simply achieved by using the periodization of the Euclidean distance $\mathrm{dist}(x, y) := \inf_{k\in\mathbb{Z}} |x - y + k| \ \forall (x, y) \in [0, 1]$. The kernel K is changed accordingly. The periodization induces a topological change in the solution which is classical in periodic optimal transport problems. In figure 4 we present a numerical solution (on the flat torus) for a discontinuous final configuration. The computations in figures 1, 3 and 4

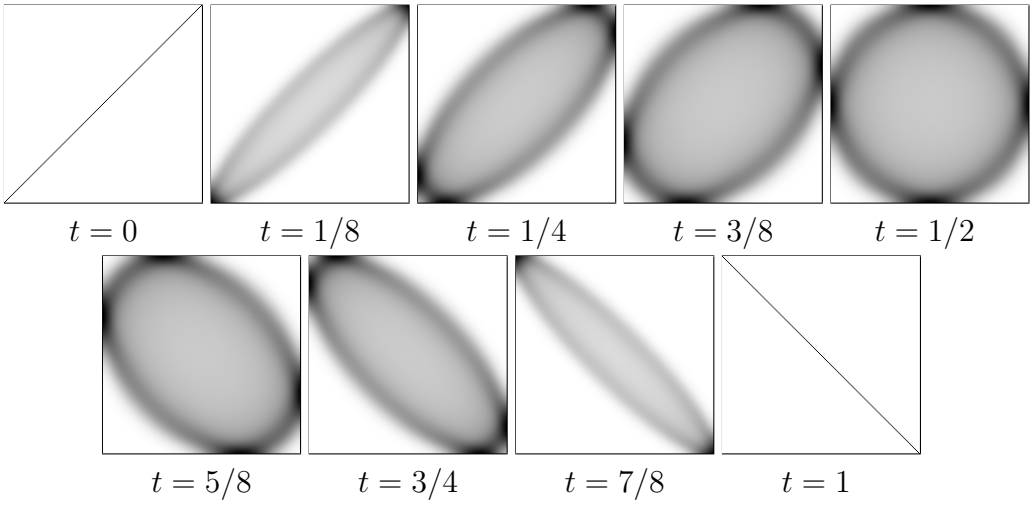

Figure 1: Non Periodic Case : Gray-map value of P_{t_k} (see (6.1)) for different times t_k. Horizontal axis is x_0 and vertical axis x_k.

are performed with a uniform discretization of $[0, 1]$ with $M = 200$ points, $\epsilon = 10^{-3}$ and $N = 16$.

6.2 Two-dimensional experiments: the Beltrami flow

Consider the unit square $\mathcal{D} = [0, 1]^2$ and the Beltrami flow obtained from the following time-independent velocity and pressure fields:

$$u(\tilde{x}_1, \tilde{x}_2) = (-\cos(\pi\tilde{x}_1)\sin(\pi\tilde{x}_2), \sin(\pi\tilde{x}_1)\cos(\pi\tilde{x}_2)),$$
$$p(\tilde{x}_1, \tilde{x}_2) = \frac{1}{2}(\sin(\pi\tilde{x}_1)^2 + \sin(\pi\tilde{x}_2)^2).$$

where $(\tilde{x}_1, \tilde{x}_2)$ denotes the 2D Cartesian coordinates One can verify that they solve the steady Euler equations.

It is possible to integrate the ODE $\partial_t X(t, x) = u(X(t, x)),\ \forall x \in \mathcal{D}$ and construct for any final time T classical solution in Sdiff to a (1.2). For the same final configuration $X_T = X(T, .)$ Brenier established in [6, Theorem 5.1]) the consistency with the generalized solution of (1.6) provided

$$\sup_{(t,x)\in[0,T]\times\mathcal{D}} \nabla_x^2 p(t, x) < \frac{\pi^2}{T^2}\, Id$$

in the sense of positive definite matrices. For the Beltrami flow, the maximum eigenvalue of the Hessian of the pressure is π^2 which suggests $T = 1$

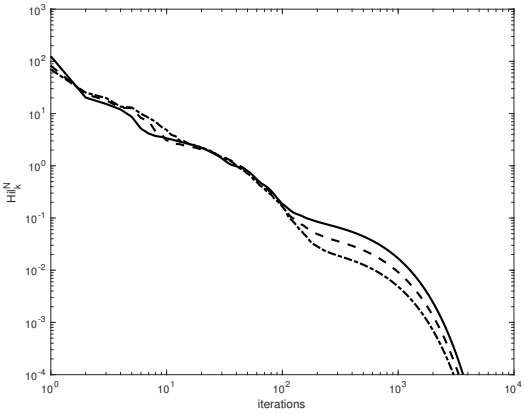

Figure 2: Distance (log scale) between two consecutive iterations of Sinkhorn by using the Hilbert projective metric for $N = 8$ (solid line), $N = 16$ (dashed line) and $N = 32$ (dotted-dashed line)

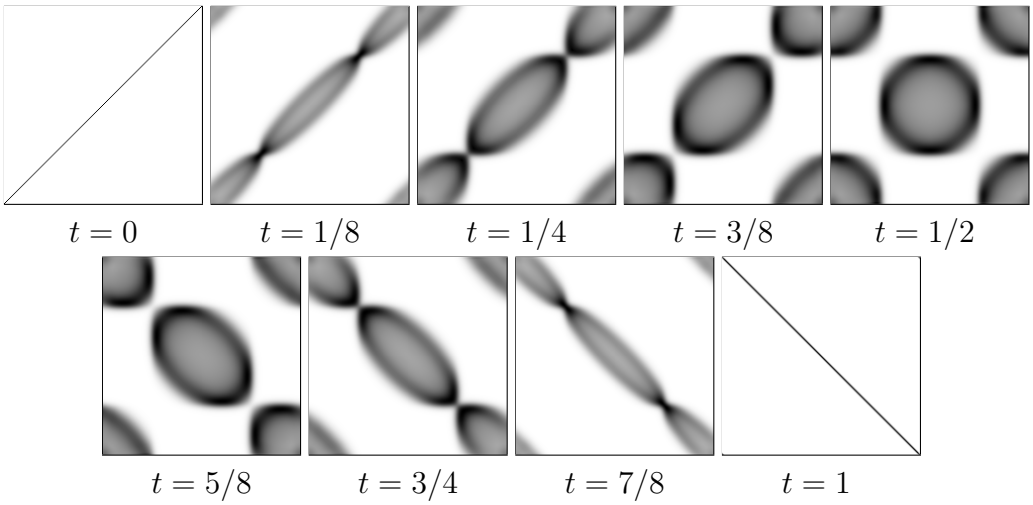

Figure 3: Periodic Case : Gray-map value of P_{t_k} (see (6.1)) for different times t_k. Horizontal axis is x_0 and vertical axis x_k.

is a critical time and that we may expect the optimal GIF to depart from the Sdiff solution for $T > 1$ and exhibit generalized behavior such as splitting/mixing/crossing of "generalized particles".

In order to track the Lagrangian behavior of the generalized solution, we extend the 1D representation technique as follows. We split the domain $\mathcal{D}$

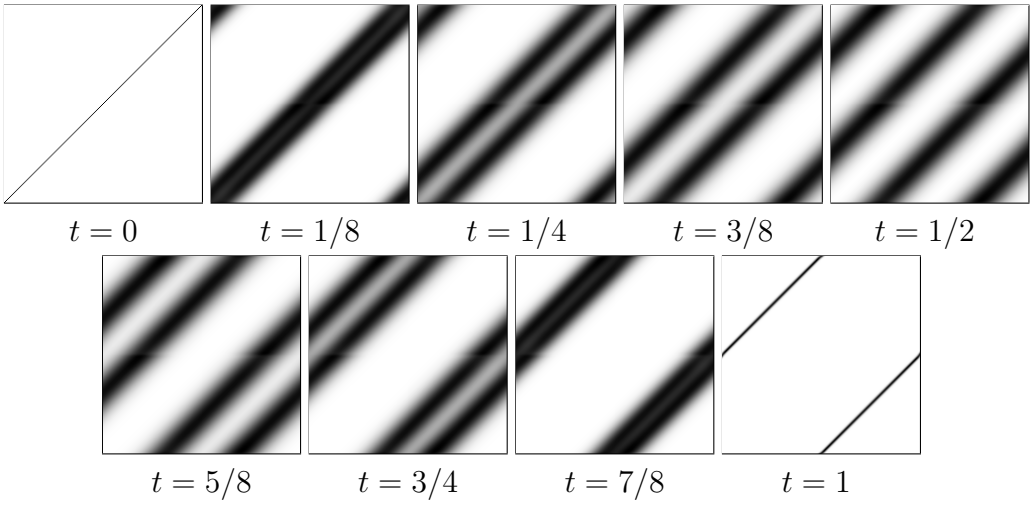

Figure 4: Periodic Case (X_T discontinuous) : Gray-map value of P_{t_k} (see (6.1)) for different times t_k. Horizontal axis is x_0 and vertical axis x_k.

at initial time into three colored sub-domains :

$$RED := [0, \frac{1}{3}] \times [0,1],\ GREEN := (\frac{1}{3}, \frac{2}{3}] \times [0,1],\ BLUE := (\frac{2}{3}, 1] \times [0,1]$$

Then we use again the definition (6.1) and plot for each time t_k and for all $x_k \in \mathcal{D}$ the 2D fields

$$P_{t_k,R}(x_k) = \int_{RED} P_{t_k}(x_0, x_k)\, \mathrm{d}x_0 \tag{6.4}$$

$$P_{t_k,G}(x_k) = \int_{GREEN} P_{t_k}(x_0, x_k)\, \mathrm{d}x_0 \tag{6.5}$$

$$P_{t_k,B}(x_k) = \int_{BLUE} P_{t_k}(x_0, x_k)\, \mathrm{d}x_0 \tag{6.6}$$

in the corresponding color with an *opacity* depending the actual value of the field. Each of these fields represent the amount of mass which has traveled from the initial RED/BLUE/GREEN region at time t_k. We also plot the sum of the three fields. The value at any time is the Lebesgue measure but because we use different colors with different opacities, it gives an idea of the mixing.

In figures 5, 6 and 7, we plot different final times T : the classical Lagrangian solution with no mixing in the first column, $P_R + P_G + P_B$ in the second column and $P_R/P_G/P_B$ in the remaining three columns.

For $T = 0.9$ (figure 5) the classical and GIF solution agree. One should keep in mind that we solve an Entropic regularization of the problem which should be accounted for some of the mass spreading.

For $T = 1.3$ (figure 6) we are past the critical time and the GIF exhibit a different behavior with a different pattern like a clockwise rotation in the middle of the domain. It is cheaper in terms of kinetic energy to send the mass across rather that doing the full counterclockwise rotation.

For $T = \pi$ (figure 7) , the GIF is again different but seems to produce less mixing.

Notice that our generalized Beltrami solutions are consistent with the solution in [22] which are computed using a non convex Lagrangian formulation.
All the computations are performed with a uniform discretization of $[0, 1]^2$ with $M = 64 \times 64$ points, $\epsilon = 10^{-4}$ and $N = 16$. All these simulation take approximately a CPU time of 3 hours, this means that the code must be parallelized in order to become competitive.

Acknowledgements: It is our pleasure to thank Christian Léonard and Yann Brenier for many fruitful discussions, we are also grateful to Christian Léonard for sharing a preliminary version of [1] with us. The authors are grateful to the Agence Nationale de La Recherche through the projects ISOTACE and MAGA.

References

[1] Marc Arnaudon, Ana Bela Cruzeiro, Christian Léonard, and Jean-Claude Zambrini. An entropic interpolation problem for incompressible viscid fluids. *arXiv preprint arXiv:1704.02126*, 2017.

[2] Vladimir Arnold. Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. In *Annales de l'institut Fourier*, volume 16, pages 319–361, 1966.

[3] Vladimir I. Arnold and Boris A. Khesin. *Topological methods in hydrodynamics*, volume 125 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1998.

[4] H. H. Bauschke and A. S. Lewis. Dykstra's algorithm with Bregman projections: a convergence proof. *Optimization*, 48(4):409–427, 2000.

[5] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[6] Yann Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989.

[7] Yann Brenier. The dual least action problem for an ideal, incompressible fluid. *Archive for rational mechanics and analysis*, 122(4):323–351, 1993.

[8] Yann Brenier. Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Communications on pure and applied mathematics*, 52(4):411–452, 1999.

[9] Yann Brenier. Generalized solutions and hydrostatic approximation of the Euler equations. *Physica D: Nonlinear Phenomena*, 237(14):1982–1988, 2008.

[10] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced transport problems. *arXiv preprint arXiv:1607.05816*, 2016.

[11] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Ann. Probability*, 3:146–158, 1975.

[12] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.

[13] Donald A. Dawson and Jürgen Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[14] L Euler. Principes généraux du mouvement des fluides. *Histoire de l'Académie de Berlin*, 1755.

[15] Joel Franklin and Jens Lorenz. Special issue dedicated to Alan J. Hoffman on the scaling of multidimensional matrices. *Linear Algebra and its Applications*, 114:717 – 735, 1989.

[16] Tryphon T Georgiou and Michele Pavon. Positive contraction mappings for classical and quantum Schrödinger systems. *Journal of Mathematical Physics*, 56(3):033301, 2015.

[17] Christian Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.

[18] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Systems, A*, 34(4):1533–1574, 2014.

[19] Bruno Lévy. A numerical algorithm for L_2 semi-discrete optimal transport in 3D. *ESAIM Math. Model. Numer. Anal.*, 49(6):1693–1715, 2015.

[20] Patrick Maheux. Notes on heat kernels on infinite dimensional torus. Lecture Notes, 2008. URL: =http://www.univ-orleans.fr/mapmo/membres/maheux/InfiniteTorusV2.pdf.

[21] Quentin Mérigot. A multiscale approach to optimal transport. *Computer Graphics Forum*, 30(5):1584–1592, August 2011. 18 pages.

[22] Quentin Mérigot and Jean-Marie Mirebeau. Minimal geodesics along volume-preserving maps, through semidiscrete optimal transport. *SIAM J. Numer. Anal.*, 54(6):3465–3492, 2016.

[23] Toshio Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields*, 129(2):245–260, 2004.

[24] Luca Nenna. *Numerical Methods for Multi-Marginal Optimal Transportation.* Theses, PSL Research University, December 2016.

[25] Brendan Pass. Multi-marginal optimal transport: theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1771–1790, 2015.

[26] Ludger Rüschendorf. Convergence of the iterative proportional fitting procedure. *Ann. Statist.*, 23(4):1160–1174, 1995.

[27] Erwin Schrödinger. *Über die umkehrung der naturgesetze.* Verlag Akademie der wissenschaften in kommission bei Walter de Gruyter u. Company, 1931.

[28] Kunio Yasue. A variational principle for the Navier-Stokes equation. *J. Funct. Anal.*, 51(2):133–141, 1983.

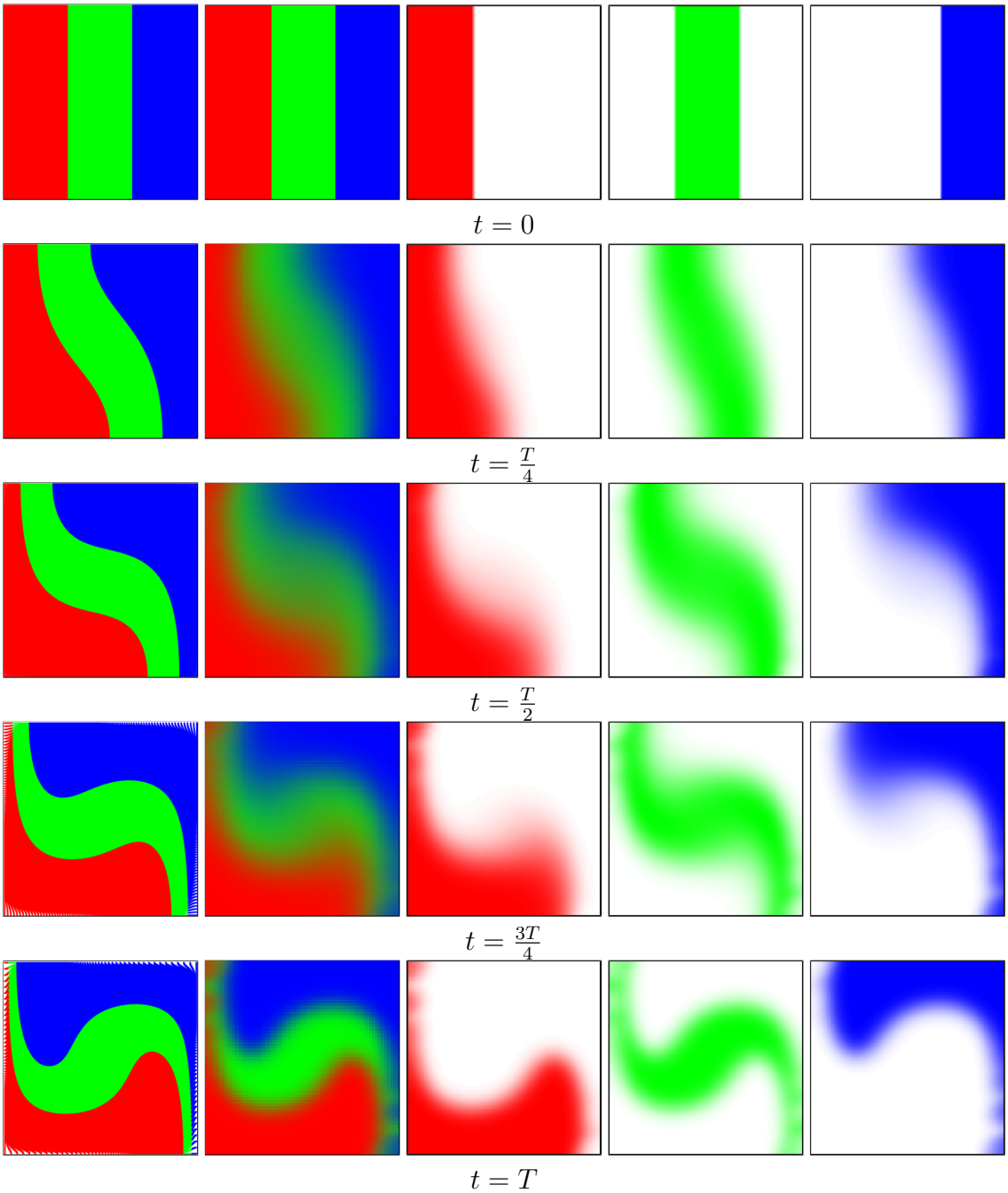

Figure 5: Final time : $T = 0.9$. Columns : Classical Color tracking of the Lagrangian solution with no mixing in the first column, $P_R + P_G + P_B$ in the second column and $P_R/P_G/P_B$ in the remaining three columns (see (6.4) for definitions). Rows : Time evolution. The final Lagrangian configuration at the bottom left is the final datum X_T in $\pi_{0,T} = (\mathrm{id}, X_T)_\# \mathcal{L}$.

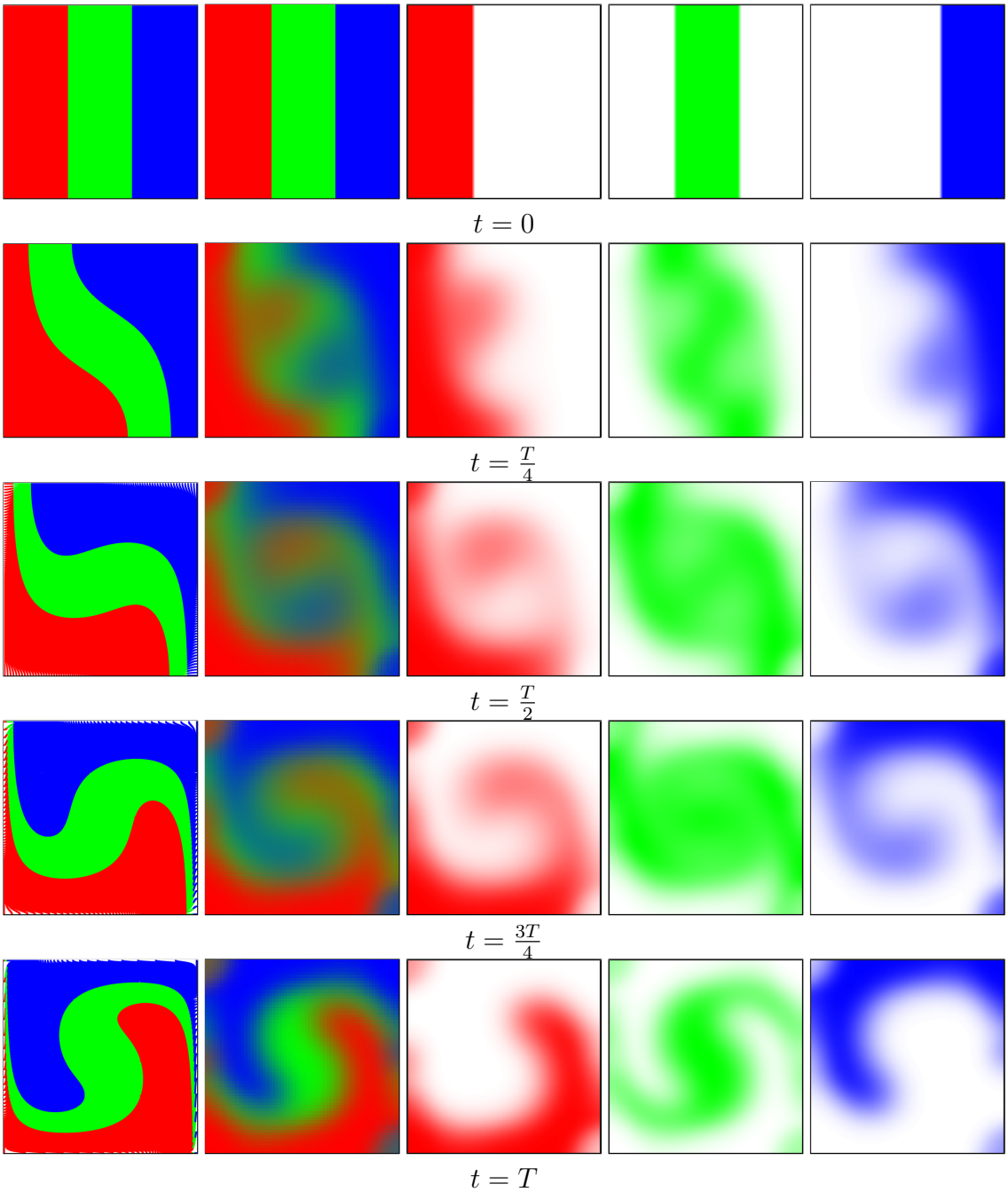

Figure 6: Final time : $T = 1.3$. Columns : Classical Color tracking of the Lagrangian solution with no mixing in the first column, $P_R + P_G + P_B$ in the second column and $P_R/P_G/P_B$ in the remaining three columns (see (6.4) for definitions). Rows : Time evolution. The final Lagrangian configuration at the bottom left is the final datum X_T in $\pi_{0,T} = (\mathrm{id}, X_T)_\# \mathcal{L}$.

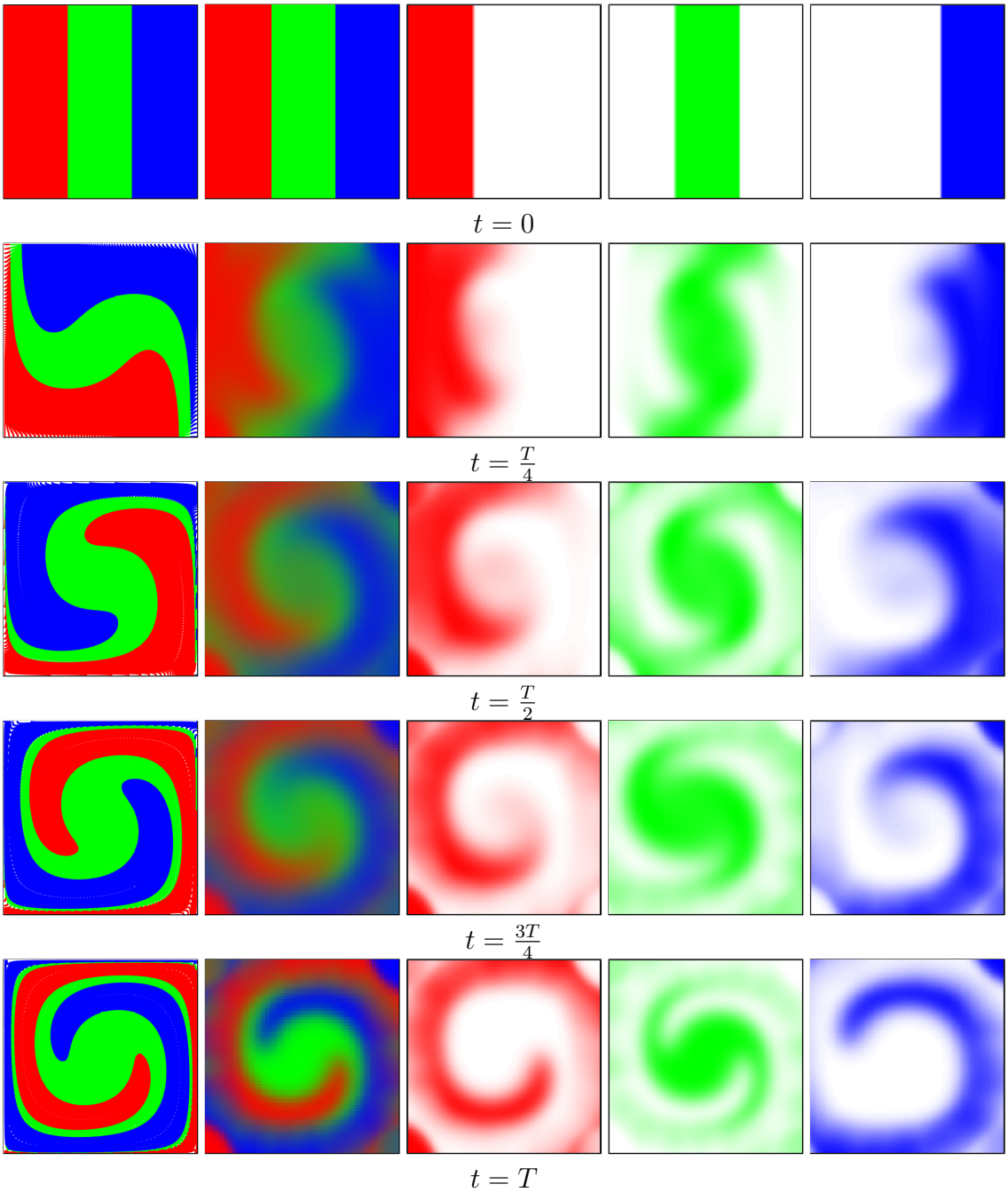

Figure 7: Final time : $T = \pi$. Columns : Classical Color tracking of the Lagrangian solution with no mixing in the first column, $P_R + P_G + P_B$ in the second column and $P_R/P_G/P_B$ in the remaining three columns (see (6.4) for definitions). Rows : Time evolution. The final Lagrangian configuration at the bottom left is the final datum X_T in $\pi_{0,T} = (\mathrm{id}, X_T)_\# \mathcal{L}$.