ESAIM: COCV 30 (2024) 25
https://doi.org/10.1051/cocv/2024013

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

DISPLACEMENT SMOOTHNESS OF ENTROPIC OPTIMAL TRANSPORT

Guillaume Carlier[1,*], Lénaïc Chizat[2] and Maxime Laborde[3]

Abstract. The function that maps a family of probability measures to the solution of the dual entropic optimal transport problem is known as the Schrödinger map. We prove that when the cost function is $\mathcal{C}^{k+1}$ with $k \in \mathbb{N}^*$ then this map is Lipschitz continuous from the L^2-Wasserstein space to the space of $\mathcal{C}^k$ functions. Our result holds on compact domains and covers the multi-marginal case. We also include regularity results under negative Sobolev metrics weaker than Wasserstein under stronger smoothness assumptions on the cost. As applications, we prove displacement smoothness of the entropic optimal transport cost and the well-posedness of certain Wasserstein gradient flows involving this functional, including the Sinkhorn divergence and a multi-species system.

Mathematics Subject Classification. 49Q22, 49K40, 35A15.

Received October 4, 2022. Accepted February 27, 2024.

1. Introduction

The main goal of this paper is to study the regularity of the multi-marginal Entropic Optimal Transport (EOT) problem under "displacement" of the marginals, and to apply these results to prove the well-posedness of certain evolution equations and optimization methods involving EOT. For clarity of presentation, let us first present the context and our results in the classical two marginals case. Let $\mathcal{X}_1, \mathcal{X}_2 \subset \mathbb{R}^d$ be two compact convex sets, $\boldsymbol{\mu} = (\mu_1, \mu_2) \in \mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2)$ two probability measures and $c \in \mathcal{C}^k(\mathcal{X}_1 \times \mathcal{X}_2)$ a k-times continuously differentiable *cost* function. We consider the *entropic optimal transport* problem defined as

$$E(\mu_1, \mu_2) \coloneqq \min_{\gamma \in \Pi(\mu_1,\mu_2)} \int c(x_1, x_2) \mathrm{d}\gamma(x_1, x_2) + H(\gamma|\mu_1 \otimes \mu_2) \tag{1.1}$$

where $\Pi(\mu_1, \mu_2)$ is the set of *transport plans* between μ_1 and μ_2, that is probability measures on $\mathcal{X}_1 \times \mathcal{X}_2$ with marginals μ_1 and μ_2, and H is the *relative entropy* defined as $H(\mu|\nu) = \int \log(\mathrm{d}\mu/\mathrm{d}\nu)\mathrm{d}\mu$ if μ is absolutely continuous w.r.t. ν and $+\infty$ otherwise.

This problem can be seen as a regularization of the optimal transport problem [1, 2] that benefits from improved computational [3, 4] and statistical properties [5–7], at the expense of an approximation error that can

Keywords and phrases: Entropic optimal transport, Schrödinger map, Wasserstein gradient flows.

[1] Ceremade, Université Paris Dauphine, PSL, 7577 Paris, and Inria-Paris, Mokaplan, France.
[2] Institute of Mathematics, École polytechnique fédérale de Lausanne (EPFL), Station Z, CH-1015 Lausanne, Switzerland.
[3] Université Paris Cité and Sorbonne Université, CNRS, Laboratoire Jacques-Louis Lions (LJLL), F-75006 Paris, France.

[*] Corresponding author: carlier@ceremade.dauphine.fr

be quantified [8–13]. It is also tightly related to the Schrödinger bridge problem [14, 15], which is a modification of equation (1.1) obtained by replacing $\mu_1 \otimes \mu_2$ by the product Lebesgue measure in the relative entropy term.

1.1. Schrödinger map

Equation (1.1) defines a convex optimization problem which admits a dual concave maximization formulation

$$E(\mu_1, \mu_2) = \max_{\substack{\phi_1 \in \mathcal{C}^0(\mathcal{X}_1) \\ \phi_2 \in \mathcal{C}^0(\mathcal{X}_2)}} \int_{\mathcal{X}_1} \phi_1 \mathrm{d}\mu_1 + \int_{\mathcal{X}_2} \phi_2 \mathrm{d}\mu_2 + 1 - \int_{\mathcal{X}_1 \times \mathcal{X}_2} \mathrm{e}^{\phi_1(x_1)+\phi_2(x_2)-c(x_1,x_2)} \mathrm{d}\mu_1(x_1)\mathrm{d}\mu_2(x_2).$$

This dual problem admits solutions which satisfy, for $\mu_1 \otimes \mu_2$ almost every (x_1, x_2), the following first order optimality conditions, known as the *Schrödinger system*:

$$\begin{cases} \phi_1(x_1) = -\log \int_{\mathcal{X}_2} \mathrm{e}^{\phi_2(x_2)-c(x_1,x_2)} \mathrm{d}\mu_2(x_2) \\ \phi_2(x_2) = -\log \int_{\mathcal{X}_1} \mathrm{e}^{\phi_1(x_1)-c(x_1,x_2)} \mathrm{d}\mu_1(x_1) \end{cases}.$$

In this paper, our main object of interest is the particular solution (ϕ_1, ϕ_2) which satisfies the Schrödinger system *for every* $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. It is not difficult to see that this solution inherits the $\mathcal{C}^k$ regularity of c and is unique in the quotient space $\tilde{\mathcal{C}}^k := \mathcal{C}^k(\mathcal{X}_1) \times \mathcal{C}^k(\mathcal{X}_2)/\sim$ where the equivalence relation

$$(\phi_1, \phi_2) \sim (\psi_1, \psi_2) \quad \Leftrightarrow \quad \exists \kappa \in \mathbb{R} \text{ such that } \phi_1 = \psi_1 + \kappa \text{ and } \phi_2 = \psi_2 - \kappa$$

captures the trivial invariance of the dual problem. This particular choice of solution (ϕ_1, ϕ_2) is arguably the most natural to consider thanks to its stability. It is also useful in many contexts because it represents the differential of the functional E [16]. We refer to this special solution (ϕ_1, ϕ_2) as the *Schrödinger potentials* and we define the *Schrödinger map* $S : \mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2) \to \tilde{\mathcal{C}}^k$ as

$$S : \boldsymbol{\mu} = (\mu_1, \mu_2) \mapsto \boldsymbol{\phi} = (\phi_1, \phi_2). \tag{1.2}$$

1.2. Main result in the two-marginals case

Our main contribution is a proof that the Schrödinger map S is Lipschitz continuous with respect to the following distances:

- We endow $\mathcal{P}(\mathcal{X}_i)$ with the Wasserstein metric defined for two probability measures $\mu, \nu \in \mathcal{P}(\mathcal{X}_i)$ by

$$W_2(\mu, \nu) := \left(\min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathcal{X}_i \times \mathcal{X}_i} \|y - x\|^2 \mathrm{d}\gamma(x, y) \right)^{\frac{1}{2}}.$$

  and then we endow $\mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2)$ with the product Wasserstein $\boldsymbol{W_2}$ metric given for $\boldsymbol{\mu} = (\mu_1, \mu_2)$, $\boldsymbol{\nu} = (\nu_1, \nu_2) \in \mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2)$

$$\boldsymbol{W_2}(\boldsymbol{\mu}, \boldsymbol{\nu}) := \left(W_2(\mu_1, \nu_1)^2 + W_2(\mu_2, \nu_2)^2 \right)^{\frac{1}{2}}.$$

- We endow $\tilde{\mathcal{C}}^k$ with the product, quotient supremum $\mathcal{C}^k$ norm defined from the usual $\mathcal{C}^k$ norm $\|\cdot\|_{\mathcal{C}^k}$ as

$$\|(\phi_1, \phi_2)\|_{\tilde{\mathcal{C}}^k} := \inf_{\kappa \in \mathbb{R}} \|\phi_1 - \kappa\|_{\mathcal{C}^k} + \|\phi_2 + \kappa\|_{\mathcal{C}^k}.$$

Our Lipschitz stability result for the Schrödinger map – which is a particular case of the more general results Theorem 2.3 and Corollary 2.4 that cover the multi-marginal case and finer regularity results – reads as follows (where by convention $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$).

Theorem 1.1. *If $c \in \mathcal{C}^{k+1}(\mathcal{X}_1 \times \mathcal{X}_2)$ for $k \in \mathbb{N}^*$, then there exists $C > 0$ that only depends on $\|c\|_{\mathcal{C}^{k+1}}$ such that for all $\boldsymbol{\mu}, \boldsymbol{\mu}' \in \mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2)$,*

$$\|S(\boldsymbol{\mu}) - S(\boldsymbol{\mu}')\|_{\tilde{\mathcal{C}}^k} \leq C\, \boldsymbol{W_2}(\boldsymbol{\mu}, \boldsymbol{\mu}').$$

This translates into a useful regularity result for the functional E (see Thm. 4.1).

Corollary 1.2. *If $c \in \mathcal{C}^2(\mathcal{X}_1 \times \mathcal{X}_2)$, then given $(\mu_1^t)_{t\in[0,1]}$ and $(\mu_2^t)_{t\in[0,1]}$ two Wasserstein geodesics, the map $h : t \mapsto E(\mu_1^t, \mu_2^t)$ is differentiable and its derivative satisfies*

$$|h'(t) - h'(s)| \leq C|t - s|\, \boldsymbol{W_2}(\boldsymbol{\mu}^0, \boldsymbol{\mu}^1)$$

where $\boldsymbol{\mu}^t = (\mu_1^t, \mu_2^t)$ and $C > 0$ only depends on $\|c\|_{\mathcal{C}^2}$. In particular, E and $-E$ are displacement semi-convex.

As an application of these results, we will prove the well-posedness of Wasserstein gradient flows for several energies involving the functional E in Section 4, and also establish exponential convergence to equilibrium in some cases. Since the Wasserstein gradient of E is $\nabla S_1, \nabla S_2$ where $S = (S_1, S_2)$ is defined in (1.2), we have for example the following result (see Prop. 4.5), where $H(\mu) \coloneqq H(\mu|\mathrm{Leb})$ is the (convex) differential entropy.

Corollary 1.3. *Let $c \in \mathcal{C}^2(\mathcal{X}_1 \times \mathcal{X}_2)$ and $\boldsymbol{\mu}^0 = (\mu_1^0, \mu_2^0) \in \mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2)$. Then the functional F defined by*

$$F(\boldsymbol{\mu}) := E(\boldsymbol{\mu}) + H(\mu_1) + H(\mu_2),$$

admits a unique Wasserstein gradient flow starting from $\boldsymbol{\mu}^0$, i.e. there exists a unique absolutely continuous curve $\boldsymbol{\mu}^t = (\mu_1^t, \mu_2^t) \in \mathcal{P}(\mathcal{X}_1) \times \mathcal{P}(\mathcal{X}_2)$ for $\boldsymbol{W_2}$, satisfying

$$\begin{cases} \partial_t \mu_1 = \nabla \cdot (\mu_1 \nabla S_1(\boldsymbol{\mu})) + \Delta \mu_1 \\ \partial_t \mu_2 = \nabla \cdot (\mu_2 \nabla S_2(\boldsymbol{\mu})) + \Delta \mu_2 \\ \boldsymbol{\mu}_{|_{t=0}} = \boldsymbol{\mu}^0 \end{cases}$$

with no-flux boundary conditions. If in addition $H(\mu_1^0), H(\mu_2^0) < +\infty$, then $\boldsymbol{\mu}^t$ converges at an exponential rate to the unique equilibrium $\boldsymbol{\mu}^*$ (see Eq. (4.7)), in the sense that there exists $\kappa > 0$ independent of $\boldsymbol{\mu}^0$ such that

$$F(\boldsymbol{\mu}^t) - F(\boldsymbol{\mu}^*) \leq \mathrm{e}^{-\kappa t}(F(\boldsymbol{\mu}^0) - F(\boldsymbol{\mu}^*)).$$

Let us mention that the system of PDEs in the previous corollary may naturally appear as an evolution model for cities: μ_1 represents the distribution of agents, μ_2 the distribution of firms, S_2 the wage paid by firms to agents and the fact that S_1 and S_2 are given by (1.2) captures an equilibrium condition on the labour market. For more details about such models, we refer to [17] (for a gradient flow approach without entropic regularization) and to [18] (in the different context of mean-field games). For extensions to more than two species and more general functionals (typically $F + G$ where G is displacement convex), see Section 4.

More applications of Theorem 1.1 are developed in companion papers that study optimization dynamics for trajectory inference [19] and regularized Wasserstein barycenters [20], also involving the functional E.

1.3. Discussion of prior work

Several works have studied the stability of the unregularized optimal transport problem [21–23]. In particular, it is known that with the square-distance cost, the Kantorovich potential from μ_1 to μ_2 *(i.e.* the counterpart of the Schrödinger potential ϕ_1 in unregularized optimal transport) is a $\frac{1}{2}$-Hölder function of μ_2 from W_2 to $\dot{H}^1(\mu_1)$ under suitable assumptions on the fixed reference measure μ_1, and that this is the strongest regularity that one can hope for in general [23].

In [24], it is proved by inverse function arguments that S is Lipschitz continuous and smooth as a map from $L^\infty_{++} \to L^\infty$, given some fixed reference measures on the ambient space. Our results will use a similar strategy but in contrast to [24] and other follow-up works such as [25], Theorem 3, we do not consider stability under additive perturbations of the marginals, but under displacement perturbations by changing the parametrization of the problem. This leads to the stronger conclusion that E is smooth in Wasserstein distance which, as we shall see, is particularly useful in the context of gradient flows of energies involving E. Let us also mention that much less is known about stability in the non-compact case, see *e.g.* [26] that shows Lipschitz continuity of E, [27] where the continuity of the Schrödinger map for the topology of convergence in probability in a certain non-compact setting, see also [28] for further stability results for the primal variable. Our results are finer, but rely in an essential way on the compact setting.

Note that a result equivalent to Theorem 1.1 for $k = 0$ was already proved in [29]. Their elegant approach consists in showing that the Sinkhorn's iteration is stable under W_1 perturbations (or, equivalently in the compact setting, W_2 perturbations) of the marginals which, combined with the fact that this iteration is a contraction for the so-called Hilbert metric, leads to the conclusion. The strength of their analysis is that it applies to $k = 0$ (i.e. merely Lipschitz continuous costs); and from the result with $k = 0$, it is not difficult to prove the $k \geq 1$ case under regularity assumptions on the cost. However, their proof technique would likely not extend to the multi-marginal case (a well-know limitation of the Hilbert metric approach). Here, we propose an independent proof technique for all $k \geq 1$ (for displacement smoothness, we need the case $k = 1$) and our analysis also gives additional information on higher degrees of smoothness of the Schrödinger map and of E.

The rest of the paper is organized as follows. In Section 2, we introduce the multi-marginal setting and state the full version of our regularity results for the Schrödinger map. The proofs of those statements can be found in Section 3. Finally, in Section 4 we study the regularity of the functional E and apply our results to the analysis of certain Wasserstein gradient flows involving E.

2. The multi-marginal case

Notation and assumptions on the domain. Let $N \geq 2$ be the number of marginals, let $\mathcal{X}_i \subset \mathbb{R}^d$ be convex and compact, for $i \in [N] \coloneqq \{1, \dots, N\}$, and let $\mathcal{X} \coloneqq \prod_{i=1}^N \mathcal{X}_i$. Given $i \in [N]$, we denote $\mathcal{X}_{-i} = \prod_{j\neq i} \mathcal{X}_j$ and identify $\mathcal{X}$ to $\mathcal{X}_i \times \mathcal{X}_{-i}$, i.e. we denote $x = (x_1, \dots, x_N)$ as $x = (x_i, x_{-i})$.

For $k \geq 0$, let $\mathcal{C}^k(\mathcal{X}_i)$ be the space of k-times continuously differentiable functions over $\mathcal{X}_i$ (that is, functions defined on $\mathcal{X}_i$ that admit a $\mathcal{C}^k$ extension on $\mathbb{R}^d$) endowed with the usual supremum norm. Using the multi-index notation, this norm is defined as $\|f\|_{\mathcal{C}^k} \coloneqq \inf_{\tilde{f}} \sup_{|\alpha|\leq k} \|\tilde{f}^{(\alpha)}\|_\infty$ where the infimum is over functions $\tilde{f}$ that are extensions of f defined on $\mathbb{R}^d$. Endowed with this norm, $\mathcal{C}^k(\mathcal{X}_i)$ is a Banach space, see [30], Chapter 8, II for details.

We denote by $\mathcal{P}(\mathcal{X}_i)$ the space of Borel probability measures on $\mathcal{X}_i$, which we endow with the weak topology, characterized by its convergent sequences as $\mu_n \rightharpoonup \mu \Leftrightarrow \int \phi \mathrm{d}\mu_n \to \int \phi \mathrm{d}\mu$ for all $\phi \in \mathcal{C}^0(\mathcal{X}_i)$. Given $(\mu_1, \dots, \mu_N) \in \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$, we denote by $\boldsymbol{\mu}$ the N-tuple $(\mu_1, \dots, \mu_N)$ and by μ the product measure $\otimes_{i=1}^N \mu_i \in \mathcal{P}(\mathcal{X})$. For $\mu = \otimes_{i=1}^N \mu_i \in \mathcal{P}(\mathcal{X})$ we let $\mu_{-i} \coloneqq \otimes_{j\neq i}\mu_j$ and $\prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$ is endowed with the product Wasserstein $\boldsymbol{W_2}$ metric, given, for $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_N) \in \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$, by

$$\boldsymbol{W_2}(\boldsymbol{\mu}, \boldsymbol{\nu}) \coloneqq \left(\sum_{i=1}^N W_2(\mu_i, \nu_i)^2 \right)^{\frac{1}{2}}.$$

In the following, $k \in \mathbb{N}^*$ is arbitrary and always denotes the regularity of the output space $\tilde{\mathcal{C}}^k$ of the Schrödinger map S, while the regularity we require for the cost function c varies across statements.

In the proofs, we use $C, C', \dots$ to denote positive constants that may change from line to line and only depend on general characteristics of the problem such as N and other quantities that are specified when needed.

2.1. Multi-marginal Schrödinger system

The multi-marginal Schrödinger system arises as the optimality conditions for the multi-marginal *entropic optimal transport* problem, defined for $\boldsymbol{\mu} \in \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$ by

$$E(\boldsymbol{\mu}) := \min_{\gamma \in \Pi(\boldsymbol{\mu})} \int_{\mathcal{X}} c(x)\mathrm{d}\gamma(x) + H(\gamma|\mu). \tag{2.1}$$

where $\Pi(\boldsymbol{\mu})$ is the set of probability measures on $\mathcal{X}$ having marginals $(\mu_1, \dots, \mu_N)$. This convex problem admits a concave dual formulation in terms of the Lagrange multipliers for the marginal constraints

$$E(\boldsymbol{\mu}) = \max_{\boldsymbol{\phi} \in \prod_{i=1}^N \mathcal{C}^0(\mathcal{X}^i)} \sum_{i=1}^N \int_{\mathcal{X}_i} \phi_i(x_i)\mathrm{d}\mu_i(x_i) + 1 - \int_{\mathcal{X}} \mathrm{e}^{\sum_{i=1}^N \phi_i(x_i) - c(x)} \mathrm{d}\mu(x). \tag{2.2}$$

Notice that these problems are multi-marginal generalizations of those presented in Section 1. The primal-dual optimality conditions read

$$\gamma(\mathrm{d}x) = \mathrm{e}^{\sum_{i=1}^N \phi_i(x_i) - c(x)} \mu(\mathrm{d}x). \tag{2.3}$$

We refer to [15, 31] for the basic theory of entropic optimal transport and [24, 32] for the multi-marginal theory. The optimality conditions for equation (2.2) coincide with the condition that $\gamma \in \Pi(\boldsymbol{\mu})$ in equation (2.3), and lead to the Schrödinger system.

Definition 2.1 (Schrödinger system/potentials/map)**.** Consider the map $T : \prod_{i=1}^N \mathcal{C}(\mathcal{X}_i) \times \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i) \to \prod_{i=1}^N \mathcal{C}(\mathcal{X}_i)$ defined for $i \in [N]$ and $x_i \in \mathcal{X}_i$ as

$$T_i(\boldsymbol{\phi}, \boldsymbol{\mu})(x_i) := \log \Big(\int_{\mathcal{X}_{-i}} \mathrm{e}^{\sum_{j=1}^N \phi_j(x_j) - c(x_i, x_{-i})} \mathrm{d}\mu_{-i}(x_{-i}) \Big). \tag{2.4}$$

A function $\boldsymbol{\phi} = (\phi_1, \dots, \phi_N)$ is called a *Schrödinger potential* associated to $\boldsymbol{\mu}$ if it solves the *Schrödinger system*

$$T(\boldsymbol{\phi}, \boldsymbol{\mu}) = 0. \tag{2.5}$$

The *Schrödinger map* is the function S that maps $\boldsymbol{\mu}$ to its Schrödinger potential $\boldsymbol{\phi}$, i.e. that satisfies $T(S(\boldsymbol{\mu}), \boldsymbol{\mu}) = 0$ (Prop. 2.2 states that this map is well-defined in a suitable sense).

Let us stress that we require equation (2.5) to hold in the space of continuous functions, that is for *every* $x \in \mathcal{X}$, rather than only μ-a.e. which is the optimality condition of equation (2.2).

Clearly, if $\boldsymbol{\phi} = (\phi_1, \dots, \phi_N)$ solves (2.5) for some fixed $\boldsymbol{\mu}$, then so does every family of potentials of the form $(\phi_1 + \kappa_1, \dots, \phi_N + \kappa_N)$ where the $\kappa \in \mathbb{R}^N$ satisfies $\sum_{i=1}^N \kappa_i = 0$. This defines an equivalence relation $\sim$ and we define the quotient space

$$\tilde{\mathcal{C}}^k := \Big(\prod_{i=1}^N \mathcal{C}^k(\mathcal{X}_i) \Big) / \sim .$$

Endowed with the quotient norm (the infimum of the norm over all representatives in the equivalence class), $\tilde{\mathcal{C}}^k$ is a Banach space.

2.2. Existence and weak continuity of the Schrödinger map

Let us state some preliminary results about the Schrödinger map.

Proposition 2.2. *If $c \in \mathcal{C}^k(\mathcal{X})$ for $k \in \mathbb{N}^*$ then for any $\boldsymbol{\mu} \in \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$, there exists a unique $\boldsymbol{\phi} = \boldsymbol{\phi}^{\boldsymbol{\mu}} \in \tilde{\mathcal{C}}^k$ such that $T(\boldsymbol{\phi}, \boldsymbol{\mu}) = 0$, i.e. the Schrödinger map $S : \boldsymbol{\mu} \mapsto \boldsymbol{\phi}^{\boldsymbol{\mu}}$ is well-defined. Moreover,*

(i) for $i \in [N]$, $\phi_i^{\boldsymbol{\mu}}$ is L_i-Lipschitz continuous, where $L_i = \sup_{x\in\mathcal{X}} \|\nabla_{x_i} c(x)\|_2$,
(ii) the Schrödinger map $S : \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i) \to \tilde{\mathcal{C}}^k$ is weakly continuous and is bounded.

The continuity claim (ii) is not needed in the sequel – and is weaker than Theorem 2.3 – but it is instructive to recall this known result that can be obtained by elementary means, before delving into more technical proofs.

Proof. The existence of a unique solution to equation (2.5) in $\prod_{i=1}^N L^\infty(\mu_i)/\sim$, i.e. in the μ-almost-everywhere sense, is proved in [24], see also [32]. In order to prove the same in $\tilde{\mathcal{C}}^0$, i.e. in the everywhere sense, let us observe that T can be expressed as $T = \mathrm{Id} + \bar{T}$ with

$$\bar{T}_i(\boldsymbol{\phi}, \boldsymbol{\mu})(x_i) = \log\Big(\int_{\mathcal{X}_{-i}} \mathrm{e}^{\sum_{j\neq i}\phi_j(x_j) - c(x_i,x_{-i})} \mathrm{d}\mu_{-i}(x_{-i})\Big) \tag{2.6}$$

by factorizing $\mathrm{e}^{\phi_i(x_i)}$ out in the definition of T_i. Thus, given a representer of the L^∞ solution $\boldsymbol{\phi}^{L^\infty} \in \prod_{i=1}^N L^\infty(\mathcal{X}_i)$, one can *define* a solution $\boldsymbol{\phi}^{\mathcal{C}^k} \in \prod_{i=1}^N \mathcal{C}^k(\mathcal{X}_i)$ in the "everywhere sense" by setting

$$\phi_i^{\mathcal{C}^k}(x_i) := -\log\Big(\int_{\mathcal{X}_{-i}} \mathrm{e}^{\sum_{j\neq i}\phi_j^{L^\infty}(x_j) - c(x_i,x_{-i})} \mathrm{d}\mu_{-i}(x_{-i})\Big) = -\bar{T}_i(\boldsymbol{\phi}^{L^\infty}, \boldsymbol{\mu})(x_i). \tag{2.7}$$

Observe that $\boldsymbol{\phi}^{\mathcal{C}^k}$ inherits the $\mathcal{C}^k$ regularity of c. Moreover, by uniqueness in L^∞, any "everywhere" solution must coincide μ-a.e. with an "almost everywhere" solution and is thus of the form given by equation (2.7). Noticing that $\bar{T}(\boldsymbol{\phi}+\kappa, \boldsymbol{\mu}) = \bar{T}(\boldsymbol{\phi}, \boldsymbol{\mu}) - \kappa$ for any family of constants $\kappa \in \mathbb{R}^N$ such that $\sum_{i=1}^N \kappa_i = 0$ and quotienting by $\sim$, it follows that there exists a unique solution of the Schrödinger system (2.5) in $\tilde{\mathcal{C}}^k$.

The Lipschitz continuity constant of $x_i \mapsto \phi_i^{\boldsymbol{\mu}}(x_i)$ can be bounded by observing that differentiating $\boldsymbol{\phi}^{\boldsymbol{\mu}} = -\bar{T}(\boldsymbol{\phi}^{\boldsymbol{\mu}}, \boldsymbol{\mu})$ in x_i gives

$$\nabla\phi_i^{\boldsymbol{\mu}}(x_i) = \int_{\mathcal{X}_{-i}} \nabla_{x_i} c(x_i, x_{-i}) \mathrm{d}Q_{-i}(x_{-i}|x_i)$$

where $Q_{-i}(\cdot|x_i) \in \mathcal{P}(\mathcal{X}_{-i})$ is a probability measure whose expression is given later in equation (3.3).

Finally, to prove weak continuity of S, it is enough to prove that if $\boldsymbol{\mu}^n$ is a sequence weakly converging to $\boldsymbol{\mu}$, then $\boldsymbol{\phi}^{\boldsymbol{\mu}^n}$ converges to $\boldsymbol{\phi}^{\boldsymbol{\mu}}$. By the previous point, $\{\boldsymbol{\phi}^{\boldsymbol{\mu}^n}\}_n$ is uniformly Lipschitz continuous and one can choose a uniformly bounded sequence of representatives so by Ascoli-Arzelà Theorem we can extract a subsequence $\boldsymbol{\phi}^{m_n}$ which converges in $\tilde{\mathcal{C}}^0$ to some $\boldsymbol{\phi}^\infty$. Since the map T is jointly continuous, it follows that $T(\boldsymbol{\phi}^\infty, \boldsymbol{\mu}) = \lim_m T(\boldsymbol{\phi}^{\boldsymbol{\mu}^{m_n}}, \boldsymbol{\mu}^{m_n}) = 0$, hence $\boldsymbol{\phi}^\infty = \boldsymbol{\phi}^{\boldsymbol{\mu}}$. But this limit is unique, so the full sequence $(\boldsymbol{\phi}^{\boldsymbol{\mu}^n})_n$ converges to $\boldsymbol{\phi}^{\boldsymbol{\mu}}$ which proves the weak continuity in $\tilde{\mathcal{C}}^0$. Since one also has $\boldsymbol{\phi}^{\boldsymbol{\mu}^n} = -\bar{T}(\boldsymbol{\phi}^{\boldsymbol{\mu}^n}, \boldsymbol{\mu}^n)$ for all n and $\bar{T}$ is continuous as a function $\tilde{\mathcal{C}}^0 \times \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i) \to \tilde{\mathcal{C}}^k$, we have in fact that $\boldsymbol{\phi}^{\boldsymbol{\mu}^n}$ converges to $\boldsymbol{\phi}^{\boldsymbol{\mu}}$ in $\tilde{\mathcal{C}}^k$. Boundedness of S finally follows from the fact that it is weakly continuous on a weakly compact set. □

2.3. Main result: regularity of the Schrödinger map

In order to study regularity beyond the zero-th order, we bypass the lack of differentiable structure of $\mathcal{P}(\mathcal{X})$ by considering parametrized paths generated by transport plans.

Consider $\boldsymbol{\mu}^0 = (\mu_i^0)_{i=1}^N$ and $\boldsymbol{\mu}^1 = (\mu_i^1)_{i=1}^N$ two families of probability measures in $\prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$, and a family of transport plans[1] $\gamma = (\gamma_i)_{i=1}^N$ such that $\gamma_i \in \mathcal{P}(\mathcal{X}_i \times \mathcal{X}_i)$ has marginals μ_i^0 and μ_i^1. These transport plans define interpolations between μ_i^0 and μ_i^1, defined for $t \in [0,1]$ as

$$\mu_i^t = ((1-t)\pi^1 + t\pi^2)_\# \gamma_i \tag{2.8}$$

where π^1 (resp. π^2) is the projection on the first (resp. second) factor of $\mathcal{X}_i \times \mathcal{X}_i$. In other terms, μ_i^t is characterized by

$$\int_{\mathcal{X}_i} \varphi_i(x_i) \mathrm{d}\mu_i^t(x_i) = \int_{\mathcal{X}_i \times \mathcal{X}_i} \varphi_i((1-t)x_i + ty_i) \mathrm{d}\gamma_i(x_i, y_i), \qquad \forall \varphi_i \in \mathcal{C}(\mathcal{X}_i).$$

Our main result is as follows.

Theorem 2.3. *For $p, k \in \mathbb{N}^*$, $p \leq k$, if $c \in \mathcal{C}^{k+p}(\mathcal{X})$ then the parametrized Schrödinger map $t \mapsto \boldsymbol{\phi}^t := S(\boldsymbol{\mu}^t)$ belongs to $\mathcal{C}^p([0,1]; \tilde{\mathcal{C}}^k)$. Moreover, there exists $C > 0$ that only depends on $\|c\|_{\mathcal{C}^{k+1}}$ and N such that*

$$\| \boldsymbol{\phi}^t - \boldsymbol{\phi}^s \|_{\tilde{\mathcal{C}}^k} \leq C |t-s| \sqrt{\mathrm{cost}(\gamma)}$$

where $\mathrm{cost}(\gamma) := \sum_{i=1}^N \int_{\mathcal{X}_i \times \mathcal{X}_i} \|y_i - x_i\|^2 \mathrm{d}\gamma_i(x_i, y_i)$ is the L^2-transport cost associated with γ.

The detailed proof is postponed to the next section, the basic ingredient being the application of the Implicit Function Theorem to the map $G(\boldsymbol{\phi}, t) := T(\boldsymbol{\phi}, \boldsymbol{\mu}^t)$. We can make the following comments:

- Tracking the constants in the proof, it can be seen that C depends exponentially on the oscillation of the cost $\sup_x c(x) - \inf_x c(x)$ and polynomially on $\|c\|_{\mathcal{C}^{k+1}}$.
- From the primal-dual relation equation (2.3), one could easily deduce stability results for the primal variable γ from this theorem.
- The fact that the map $t \mapsto S(\boldsymbol{\mu}^t)$ belongs to $\mathcal{C}^p([0,1]; \tilde{\mathcal{C}}^k)$, also holds if $(\boldsymbol{\mu}^t)_{t \in [0,1]}$, instead of being of the form (2.8), is of the form $\mu_i^t = \xi_i(\cdot, t)_\# \mu_i^0$ for $i \in [N]$, for some $\mu_i^0 \in \mathcal{P}(\mathcal{X}_i)$ and a measurable $\xi_i(x_i, \cdot) \in \mathcal{C}^p([0,1]; \mathcal{X}_i)$ with a $\mathcal{C}^p$ norm uniformly bounded in x_i. This can be seen by suitably adapting the proof of Lemma 3.4.

Applying Theorem 2.3 by choosing γ_i as the optimal transport plan between μ_i and ν_i immediately leads to the following Lipschitz continuity result for the Schrödinger map.

Corollary 2.4. *For $k \in \mathbb{N}^*$, assume that $c \in \mathcal{C}^{k+1}(\mathcal{X})$. The Schrödinger map $S : \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i) \to \tilde{\mathcal{C}}^k$ is Lipschitz continuous, i.e. there exists $C > 0$ such that, for all $\boldsymbol{\mu}, \boldsymbol{\nu} \in \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$, letting $(\boldsymbol{\phi}^{\boldsymbol{\mu}}, \boldsymbol{\phi}^{\boldsymbol{\nu}}) = (S(\boldsymbol{\mu}), S(\boldsymbol{\nu}))$,*

$$\| \boldsymbol{\phi}^{\boldsymbol{\mu}} - \boldsymbol{\phi}^{\boldsymbol{\nu}} \|_{\tilde{\mathcal{C}}^k} \leq C\, \boldsymbol{W_2}(\boldsymbol{\mu}, \boldsymbol{\nu}).$$

Our approach will also enable us to deduce a control of the Schrödinger potentials and their derivatives in terms of negative Sobolev distances between the marginals (see Sect. 3.4 for detailed definitions):

Proposition 2.5. *Assume that $c \in \mathcal{C}^{k+p}(\mathcal{X})$ with $p > d/2$ and $p \in \mathbb{N}^*$. The Schrödinger map $S : \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i) \to \tilde{\mathcal{C}}^k$ is Lipschitz continuous in the negative Sobolev norm H^{-p}, i.e. there exists $C > 0$ such that, for all $\boldsymbol{\mu}, \boldsymbol{\nu} \in$*

[1] To be clear, we *do not* assume that γ_i is an *optimal* transport plan.

$\prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$, *letting* $(\boldsymbol{\phi}^{\boldsymbol{\mu}}, \boldsymbol{\phi}^{\boldsymbol{\nu}}) = (S(\boldsymbol{\mu}), S(\boldsymbol{\nu}))$,

$$\| \boldsymbol{\phi}^{\boldsymbol{\mu}} - \boldsymbol{\phi}^{\boldsymbol{\nu}} \|_{\tilde{\mathcal{C}}^k} \leq C \| \boldsymbol{\mu} - \boldsymbol{\nu} \|_{H^{-p}}.$$

Note that when $p > d/2 + 1$, by Morrey's Theorem (see Sect. 3.4) and our compactness assumption, there exists $C > 0$ that only depends on $\mathcal{X}$ such that $\| \boldsymbol{\mu} - \boldsymbol{\nu} \|_{H^{-p}} \leq C \cdot \boldsymbol{W_1}(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq C \cdot \boldsymbol{W_2}(\boldsymbol{\mu}, \boldsymbol{\nu})$. Thus, the conclusion of Proposition 2.5 is generally stronger than that of Corollary 2.4, but this is at the expense of requiring more regularity on the cost function.

For illustration purposes, let us explain how this inequality leads to nonasymptotic estimation guarantees for the Schrödinger potentials given random samples. In the two marginal case, this is essentially a known result, obtained *via* different means in [7, 33]. Specifically, suppose that $\hat{\boldsymbol{\mu}}$ is an empirical measure built by drawing n independent samples from each of the measures $\boldsymbol{\mu}_i$, $i \in N$. Then, since H^p is a Reproducible Kernel Hilbert Space with a bounded kernel for $p > d/2$ (by Morrey's Theorem again), Hoeffding's inequality shows that $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_{H^{-p}}$ is bounded by $C \cdot n^{-1/2}\sqrt{\log(1/\delta)}$ with probability $1 - \delta$ where here C depends only on $\mathcal{X}$. By Proposition 2.5, this directly translates into a high-probability bound on $\| \boldsymbol{\phi}^{\boldsymbol{\mu}} - \boldsymbol{\phi}^{\hat{\boldsymbol{\mu}}} \|_{\tilde{\mathcal{C}}^k}$. It might also be worth mentioning that Proposition 2.5 is obtained by linearly interpolating the marginals, as such, it is a slight departure from the rest of the paper which essentially focuses on displacement interpolation.

3. Proofs

The main tool to prove Theorem 2.3 is the Implicit Function Theorem. We will apply it to the function $G : \tilde{\mathcal{C}}^k \times [0, 1] \to \mathcal{C}^k$ defined as

$$G(\boldsymbol{\phi}, t) \coloneqq T(\boldsymbol{\phi}, \boldsymbol{\mu}^t) \tag{3.1}$$

whose expression[2] is, using the convention $y_i \coloneqq x_i$,

$$G_i(\boldsymbol{\phi}, t)(x_i) = \phi_i(x_i) + \log\Big(\int_{\mathcal{X}_{-i} \times \mathcal{X}_{-i}} \mathrm{e}^{\sum_{j\neq i} \phi_j((1-t)x_j + ty_j) - c((1-t)x + ty)} \mathrm{d}\gamma_{-i}(x_{-i}, y_{-i}) \Big).$$

For this purpose, in the next sections, we study the properties of the maps T and G.

3.1. Invertibility of the differential of T

Let us fix $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N) \in \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$ and $\mu = \otimes_{i=1}^N \mu_i \in \mathcal{P}(\mathcal{X})$ and study the map $\boldsymbol{\phi} \mapsto T(\boldsymbol{\phi}, \boldsymbol{\mu})$, which is a self-map of $\tilde{\mathcal{C}}^k$. Note that T is of class C^∞ in the first variable and its differential is given, for $\boldsymbol{h} \in \tilde{\mathcal{C}}^k$, by

$$D_\phi T_i(\boldsymbol{\phi}, \boldsymbol{\mu})(\boldsymbol{h})(x_i) = h_i(x_i) + \int_{\mathcal{X}_{-i}} \Big(\sum_{j \neq i} h_j(x_j) \Big) q_{-i}(x_{-i}|x_i) \mathrm{d}\mu_{-i}(x_{-i}).$$

where we have introduced the function q_{-i} defined, with the convention $x'_i = x_i$, by

$$q_{-i}(x_{-i}|x_i) \coloneqq \frac{\mathrm{e}^{\sum_{j\neq i} \phi_j(x_j) - c(x)}}{\int_{\mathcal{X}_{-i}} \mathrm{e}^{\sum_{j\neq i} \phi_j(x'_j) - c(x')} \mathrm{d}\mu_{-i}(x'_{-i})}.$$

[2] the maps G and T take values in $\mathcal{C}^k$ but it will sometimes be convenient to compose them from the left with the canonical projection $\mathcal{C}^k \to \tilde{\mathcal{C}}^k$, slightly abusing notations, we will still denote by G and T these maps with values in $\tilde{\mathcal{C}}^k$.

Note that q_{-i} depends on $\boldsymbol{\phi}$ and $\boldsymbol{\mu}$ although this is not explicit in the notation. Similarly, let

$$q(x) \coloneqq \frac{\mathrm{e}^{\sum_j \phi_j(x_j) - c(x)}}{\int_{\mathcal{X}} \mathrm{e}^{\sum_j \phi_j(x'_j) - c(x')} \mathrm{d}\mu(x')}, \qquad q_i(x_i) \coloneqq \frac{\int_{\mathcal{X}_{-i}} \mathrm{e}^{\sum_j \phi_j(x_j) - c(x)} \mathrm{d}\mu_{-i}(x_{-i})}{\int_{\mathcal{X}} \mathrm{e}^{\sum_j \phi_j(x'_j) - c(x')} \mathrm{d}\mu(x')}. \tag{3.2}$$

Observe that if $\boldsymbol{\phi}$ and c are of class $\mathcal{C}^k$ then the functions q, q_i, q_{-i} are of class $\mathcal{C}^k$ as well. These functions are densities of probability densities in the sense that it holds

$$Q \coloneqq q\mu \in \mathcal{P}(\mathcal{X}), \qquad Q_i \coloneqq q_i \mu_i \in \mathcal{P}(\mathcal{X}_i), \qquad Q_{-i}(\cdot|x_i) \coloneqq q_{-i}(\cdot|x_i)\mu_{-i} \in \mathcal{P}(\mathcal{X}_{-i}),\ \forall x_i \in \mathcal{X}_i. \tag{3.3}$$

By construction, for each i, Q_i is the i-th marginal of Q on $\mathcal{X}_i$ and Q_{-i} is the disintegration of Q with respect to this marginal, *i.e.*:

$$\mathrm{d}Q(x_i,\ x_{-i}) = \mathrm{d}Q_{-i}(x_{-i}|x_i)\mathrm{d}Q_i(x_i). \tag{3.4}$$

In the next lemma, we remark that these densities are uniformly bounded from above and below by positive quantities, a fact which we will often use in the following.

Lemma 3.1. *Let q be defined by (3.2). Then, for all $x \in \mathcal{X}$,*

$$\mathrm{e}^{-2(N\|\boldsymbol{\phi}\|_{\tilde{\mathcal{C}}^0} + \|c\|_{\mathcal{C}^0})} \leq q(x) \leq \mathrm{e}^{2(N\|\boldsymbol{\phi}\|_{\tilde{\mathcal{C}}^0} + \|c\|_{\mathcal{C}^0})}.$$

Moreover q_i and q_{-i} satisfy the same bounds.

Proof. From the definition of q, for all $x \in \mathcal{X}$,

$$q(x) = \frac{\mathrm{e}^{\sum_j \phi_j(x_j) - c(x)}}{\int_{\mathcal{X}} \mathrm{e}^{\sum_j \phi_j(x'_j) - c(x')} \mathrm{d}\mu(x')} \leq \frac{\mathrm{e}^{\sum_j \|\phi_j\|_\infty + \|c\|_{\mathcal{C}^0}}}{\mathrm{e}^{-\sum_j \|\phi_j\|_\infty - \|c\|_{\mathcal{C}^0}}} \leq \mathrm{e}^{2(N\|\boldsymbol{\phi}\|_{\mathcal{C}^0} + \|c\|_{\mathcal{C}^0})},$$

since $\mu \in \mathcal{P}(\mathcal{X})$. In addition, from the definition of q, we remark that q does not depend of the representative of $\boldsymbol{\phi}$ in $\tilde{\mathcal{C}}^0$ which gives the upper bound on q. We obtain the lower bound, as well as the result for q_i and q_{-i} with the same arguments. □

Let us also remark that in the previous bounds, the norm $\|c\|_{\mathcal{C}^0}$ can be replaced by $\inf_{\kappa \in \mathbb{R}} \|c + \kappa\|_{\mathcal{C}^0} = (\sup c - \inf c)/2$, i.e. half the *oscillation of* c (in fact, the Schrödinger map is invariant if c changes by an additive constant). The following lemma is central in our development and is an adaptation of [24], Proposition 3.1 (with different functional spaces). The first claim of invertibility appeared in a similar form in [34], Lemma 5 where it is key to prove a central limit theorem for EOT, but our proof (Step 1) is different as (i) in our context there is no natural way to get rid of the non-uniqueness of Schrödinger potentials so we work directly in the quotient space $\tilde{\mathcal{C}}^k$ and (ii) our approach leads to control on the norm on the inverse (the second part of the claim).

Lemma 3.2. *Let $k \in \mathbb{N}^*$ and assume that $\boldsymbol{\phi} \in \tilde{\mathcal{C}}^0$ and $c \in \mathcal{C}^k(\mathcal{X})$.*

Then $D_{\boldsymbol{\phi}} T(\boldsymbol{\phi}, \boldsymbol{\mu})$ is an invertible linear self-map of $\tilde{\mathcal{C}}^k$. Moreover, there exists $C > 0$ that only depends on N, $\|c\|_{\mathcal{C}^k}$ and $\|\boldsymbol{\phi}\|_{\tilde{\mathcal{C}}^0}$ such that

$$\|[D_{\boldsymbol{\phi}} T(\boldsymbol{\phi}, \boldsymbol{\mu})]^{-1}\|_{\tilde{\mathcal{C}}^k \to \tilde{\mathcal{C}}^k} \leq C.$$

Proof. We have $D_{\boldsymbol{\phi}}T(\boldsymbol{\phi},\boldsymbol{\mu}) = \mathrm{Id} + L$ with

$$L_i(\boldsymbol{h})(x_i) = \int_{\mathcal{X}_{-i}} \Big(\sum_{j\neq i} h_j(x_j)\Big) q_{-i}(x_{-i}|x_i)\mathrm{d}\mu_{-i}(x_{-i}).$$

Observe that since $c \in \mathcal{C}^k(\mathcal{X})$, $L_i(\boldsymbol{h}) \in \mathcal{C}^k(\mathcal{X}_i)$ with its derivatives up to order k equi-continuous when h runs through a bounded set of $\tilde{\mathcal{C}}^0$. It follows, by Arzelà-Ascoli Theorem, that $L : \tilde{\mathcal{C}}^0 \to \tilde{\mathcal{C}}^k$ is compact, and a fortiori $L : \tilde{\mathcal{C}}^k \to \tilde{\mathcal{C}}^k$ is compact too.

Step 1. Let us show that $\mathrm{id} + L$ is invertible. Let $\boldsymbol{h} \in \prod \mathcal{C}(\mathcal{X}_i)$ be such that $\boldsymbol{h} + L(\boldsymbol{h}) = 0$ in $\tilde{\mathcal{C}}^0$, i.e.

$$h_i(.) + \int_{\mathcal{X}_{-i}} \Big(\sum_{j\neq i} h_j(x_j)\Big)\mathrm{d}Q_{-i}(x_{-i}|.) = \lambda_i,\ i = 1,\dots,N,\ \sum_{j=1}^{N} \lambda_j = 0. \tag{3.5}$$

Integrating (3.5) with respect to Q_i, we deduce from (3.4), that

$$\sum_{k=1}^{N} \int_{\mathcal{X}_k} h_k \mathrm{d}Q_k = \lambda_i,\quad i = 1,\dots,N$$

so that all the λ_i's are equal to 0; hence $\boldsymbol{h} + L(\boldsymbol{h}) = 0$ in $\mathcal{C}^0$ (and not only in the quotient $\tilde{\mathcal{C}}^0$). Then, taking the dot product of $\boldsymbol{h}$ with $\boldsymbol{h} + L(\boldsymbol{h})$ in $\prod L^2(Q_i)$ and using (3.4), it follows, reasoning as in [24],

$$\begin{aligned}
0 &= \sum_{i=1}^{N} \int_{\mathcal{X}_i} h_i(x_i)\Big(h_i(x_i) + \int_{\mathcal{X}_{-i}} \Big(\sum_{j\neq i} h_j(x_j)\Big)\mathrm{d}Q_{-i}(x_{-i}|x_i)\Big)\mathrm{d}Q_i(x_i) \\
&= \sum_{i} \int_{\mathcal{X}_i} h_i(x_i)^2 \mathrm{d}Q_i(x_i) + \sum_{i\neq j} \int_{\mathcal{X}} h_i(x_i)h_j(x_j)\mathrm{d}Q(x) \\
&= \int_{\mathcal{X}} \Big(\sum_i h_i(x_i)\Big)^2 \mathrm{d}Q(x).
\end{aligned}$$

We deduce that $x \mapsto \sum h_i(x_i)$ is equal to 0 as a function in $L^2(Q)$ and hence in $L^2(\mu)$.

Now consider the space $\tilde{L}^2_{\boldsymbol{\mu}} := \prod_{i=1}^{N} L^2(\mu_i)/\sim$ which, endowed with the quotient space structure, is also a Hilbert space. By Lemma 3.3 (proved hereafter), it follows that $\boldsymbol{h} \sim 0$ in $\tilde{L}^2_{\boldsymbol{\mu}}$, *i.e.* there exists $\kappa \in \mathbb{R}^n$ such that $\sum \kappa_i = 0$ and $h_i(x_i) = \kappa_i$ for μ_i-a.e. x_i. It only remains to show that this equality holds in fact everywhere. Using $\boldsymbol{h} = -L(\boldsymbol{h})$ and $q_{-i}\mu_{-i} \in \mathcal{P}(\mathcal{X}_{-i})$, it holds for $x_i \in \mathcal{X}_i$

$$h_i(x_i) = -\int_{\mathcal{X}_{-i}} \Big(\sum_{j\neq i} h_j(x_j)\Big) q_{-i}(x_{-i}|x_i)\mathrm{d}\mu_{-i}(x_{-i}) = -\sum_{j\neq i} \kappa_j = \kappa_i$$

Thus $\boldsymbol{h} = 0$ in $\tilde{\mathcal{C}}^k$. Conversely, any $\boldsymbol{h} = 0$ in $\tilde{\mathcal{C}}^k$ also clearly belongs to $\ker(D_{\boldsymbol{\phi}}T(\boldsymbol{\phi},\boldsymbol{\mu}))$. Hence $\ker(D_{\boldsymbol{\phi}}T(\boldsymbol{\phi},\boldsymbol{\mu}))$ is precisely the equivalence class of 0 i.e. $D_{\boldsymbol{\phi}}T(\boldsymbol{\phi},\boldsymbol{\mu})$ is injective on $\tilde{\mathcal{C}}^k$. Since L is a compact operator of $\tilde{\mathcal{C}}^k$, it follows from the Fredholm Alternative Theorem [35], Chapter 6 that the range of $\mathrm{Id} + L$ is $\tilde{\mathcal{C}}^k$. Hence $D_{\boldsymbol{\phi}}T(\boldsymbol{\phi},\boldsymbol{\mu})$ is onto and therefore an invertible linear self-map of $\tilde{\mathcal{C}}^k$.

Step 2. Now let us estimate the operator norm of $D_{\boldsymbol{\phi}}T(\boldsymbol{\phi},\boldsymbol{\mu})^{-1}$ as a self-map of $\tilde{\mathcal{C}}^k$. Let $\boldsymbol{h} \in \tilde{\mathcal{C}}^k, \boldsymbol{g} \in \tilde{\mathcal{C}}^k$ be such that $\boldsymbol{h} + L(\boldsymbol{h}) = \boldsymbol{g}$. Let us choose the representative of $\boldsymbol{g}$ that satisfies

$$\int_{\mathcal{X}_1} g_1(x_1)\mathrm{d}Q_1(x_1) = \cdots = \int_{\mathcal{X}_N} g_N(x_N)\mathrm{d}Q_N(x_N). \tag{3.6}$$

Reasoning as above, it holds

$$\begin{aligned}\sum_{i=1}^N \int_{\mathcal{X}_i} g_i(x_i)h_i(x_i)\mathrm{d}Q_i(x_i) &= \sum_i \int_{\mathcal{X}_i} h_i(x_i)^2 \mathrm{d}Q_i(x_i) + \sum_{i,j,\, i\neq j} \int_{\mathcal{X}} h_i(x_i)h_j(x_j)\mathrm{d}Q(x) \\ &= \int_{\mathcal{X}} \Big(\sum_i h_i(x_i)\Big)^2 \mathrm{d}Q(x)\end{aligned}$$

where the first integral is unambiguously defined because, thanks to our choice of representative for g, it does not depend on the representative chosen for h. Let us choose the optimal representative for $\boldsymbol{h}$ in $\tilde{L}^2_{\boldsymbol{\mu}}$ appearing in Lemma 3.3. We have, for some $C > 0$ that may change from a line to another but only depends on N, $\|c\|_{\mathcal{C}_0}$ and $\|\,\boldsymbol{\phi}\,\|_{\tilde{\mathcal{C}}_0}$:

$$\begin{aligned}\|\,\boldsymbol{h}\,\|^2_{\prod L^2(\mu_i)} &\overset{(i)}{\leq} N \left\|\oplus_{i=1}^N h_i\right\|^2_{L^2(\mu)} \\ &\overset{(ii)}{\leq} C \left\|\oplus_{i=1}^N h_i\right\|^2_{L^2(Q)} \\ &\overset{(iii)}{\leq} C\|\,\boldsymbol{g}\,\|_{\prod L^2(Q_i)}\|\,\boldsymbol{h}\,\|_{\prod L^2(Q_i)} \\ &\overset{(iv)}{\leq} C\|\,\boldsymbol{g}\,\|_{\prod L^2(Q_i)}\|\,\boldsymbol{h}\,\|_{\prod L^2(\mu_i)}\end{aligned}$$

where we have used (i) Lemma 3.3 (where the notation $\oplus$ is defined) , (ii) and (iv) Lemma 3.1, and (iii) the previous computation and Cauchy-Schwarz inequality in $\prod L^2(Q_i)$. It follows, invoking once again Lemmas 3.3 and 3.1, that

$$\|\,\boldsymbol{h}\,\|_{\tilde{L}^2_{\boldsymbol{\mu}}} = \|\,\boldsymbol{h}\,\|_{\prod L^2(\mu_i)} \leq C\|\,\boldsymbol{g}\,\|_{\prod L^2(Q_i)} = C\|\,\boldsymbol{g}\,\|_{\tilde{L}^2_Q} \leq C\|\,\boldsymbol{g}\,\|_{\tilde{\mathcal{C}}^0}.$$

For the last equality, we have used the fact that the $\tilde{L}^2_Q$ norm is precisely the $\prod L^2(Q_i)$ norm of the representative that satisfies equation (3.6), by Lemma 3.3 (here $\tilde{L}^2_Q$ is defined similarly as $\tilde{L}^2_{\boldsymbol{\mu}}$ from the marginals Q_i of Q).

Step 3. We now improve the $\tilde{L}^2_{\boldsymbol{\mu}}$ control into a $\tilde{\mathcal{C}}^k$ control. Restarting from $\boldsymbol{h}+L(\boldsymbol{h}) = \boldsymbol{g}$, it holds

$$h_i(x_i) = g_i(x_i) - \int_{\mathcal{X}_{-i}} \Big(\sum_{j\neq i} h_j(x_j)\Big) q_{-i}(x_{-i}|x_i)\mathrm{d}\mu_{-i}(x_{-i}). \tag{3.7}$$

Thanks to our control on $\|\,\boldsymbol{h}\,\|_{\prod L^2(\mu_i)}$ by $\|\,\boldsymbol{g}\,\|_{\tilde{\mathcal{C}}^0}$, given constants κ_i, it follows from (3.7) that

$$\|h_i + \kappa_i\|_{\tilde{\mathcal{C}}^0(\mathcal{X}_i)} \leq \|g_i + \kappa_i\|_{\tilde{\mathcal{C}}^0(\mathcal{X}_i)} + C\|\,\boldsymbol{g}\,\|_{\tilde{\mathcal{C}}^0}$$

summing over i and minimizing with respect to the κ_i's summing to 0, we get

$$\|\,\boldsymbol{h}\,\|_{\tilde{\mathcal{C}}^0} \leq C\|\,\boldsymbol{g}\,\|_{\tilde{\mathcal{C}}^0}.$$

In a similar way, using the fact that $c \in \mathcal{C}^k$, successive differentiations of (3.7) yield

$$\|\,\boldsymbol{h}\,\|_{\tilde{\mathcal{C}}^k} = \|[D_{\boldsymbol{\phi}}T(\boldsymbol{\phi},\boldsymbol{\mu})]^{-1}(\boldsymbol{g})\|_{\tilde{\mathcal{C}}^k} \leq C\|\,\boldsymbol{g}\,\|_{\tilde{\mathcal{C}}^k}$$

for a constant C that only depends on N, $\|c\|_{\mathcal{C}^k}$ and $\|\,\boldsymbol{\phi}\,\|_{\tilde{\mathcal{C}}^0}$. □

To end this section, we prove Lemma 3.3 used in the previous proof.

Lemma 3.3. *For $\boldsymbol{h} \in \prod_{i=1}^N L^2(\mu_i)$, denoting $\oplus_{i=1}^N h_i : x \mapsto \sum_{i=1}^N h_i(x_i)$ it holds*

$$\| \boldsymbol{h} \|_{\tilde{L}^2_{\boldsymbol{\mu}}}^2 \leq \left\| \oplus_{i=1}^N h_i \right\|_{L^2(\mu)}^2 \leq N \| \boldsymbol{h} \|_{\tilde{L}^2_{\boldsymbol{\mu}}}^2.$$

Moreover, the quotient norm is achieved by the unique representative $\hat{\boldsymbol{h}} \sim \boldsymbol{h}$ that satisfies $\int_{\mathcal{X}_1} \hat{h}_1 \mathrm{d}\mu_1 = \cdots = \int_{\mathcal{X}_N} \hat{h}_N \mathrm{d}\mu_N$, i.e. it holds $\|h\|_{\tilde{L}^2_{\boldsymbol{\mu}}}^2 = \sum_{i=1}^N \|\hat{h}_i\|_{L^2(\mu_i)}^2$.

Proof. By definition,

$$\| \boldsymbol{h} \|_{\tilde{L}^2_{\boldsymbol{\mu}}}^2 = \min_{\substack{\kappa \in \mathbb{R}^N \\ \sum_i \kappa_i = 0}} \sum_{i=1}^N \int_{\mathcal{X}_i} (h_i(x_i) - \kappa_i)^2 \mathrm{d}\mu_i(x_i). \tag{3.8}$$

A vector $\kappa \in \mathbb{R}^N$ solves this problem iff $\sum_i \kappa_i = 0$ and there exists a Lagrange multiplier $\nu \in \mathbb{R}$ such that for $i \in [N]$,

$$0 = \int_{\mathcal{X}_i} (h_i(x_i) - \kappa_i) \mathrm{d}\mu_i(x_i) - \nu = \mathbf{E}_{\mu_i}[h_i] - \kappa_i - \nu.$$

with the shorthand $\mathbf{E}_\mu[h] \coloneqq \int h \mathrm{d}\mu$ and $\mathrm{Var}_\mu(h) \coloneqq \mathbf{E}_\mu[(h - \mathbf{E}_\mu[h])^2]$. It follows that $\nu = \frac{1}{N} \sum_{i=1}^N \mathbf{E}_{\mu_i}[h_i]$ and as a consequence

$$\begin{aligned} \| \boldsymbol{h} \|_{\tilde{L}^2_{\boldsymbol{\mu}}}^2 &= \sum_{i=1}^N \int_{\mathcal{X}_i} (h_i(x_i) - \mathbf{E}_{\mu_i}[h_i] + \nu)^2 \mathrm{d}\mu_i(x_i) \\ &= \sum_{i=1}^N \Big[\int_{\mathcal{X}_i} (h_i(x_i) - \mathbf{E}_{\mu_i}[h_i])^2 \mathrm{d}\mu_i(x_i) + \nu^2 \Big] \\ &= \sum_{i=1}^N \mathrm{Var}_{\mu_i}(h_i) + \frac{1}{N} \Big(\sum_{i=1}^N \int h_i \mathrm{d}\mu_i \Big)^2 \end{aligned}$$

where the second equality follows by expanding the square and observing that the cross-terms vanish. On the other hand, using the fact that $\sum_i \kappa_i = 0$, it holds

$$\begin{aligned} \left\| \oplus_{i=1}^N h_i \right\|_{L^2(\mu)}^2 &= \int_{\mathcal{X}} \Big(\sum_{i=1}^N (h_i(x_i) - \kappa_i) \Big)^2 \mathrm{d}\mu(x) \\ &= \sum_{i,j,i \neq j} \int_{\mathcal{X}_i \times \mathcal{X}_j} (h_i(x_i) - \mathbf{E}_{\mu_i}[h_i] + \nu)(h_j(x_j) - \mathbf{E}_{\mu_j}[h_j] + \nu) \mathrm{d}\mu_i(x_i) \mathrm{d}\mu_j(x_j) \\ &\quad + \sum_i \int_{\mathcal{X}_i} (h_i(x_i) - \mathbf{E}_{\mu_i}[h_i]) + \nu)^2 \mathrm{d}\mu_i(x_i) \\ &= N(N-1)\nu^2 + \| \boldsymbol{h} \|_{\tilde{L}^2_{\boldsymbol{\mu}}}^2 \\ &= \sum_{i=1}^N \mathrm{Var}_{\mu_i}(h_i) + \Big(\sum_{i=1}^N \int_{\mathcal{X}_i} h_i \mathrm{d}\mu_i \Big)^2 \end{aligned}$$

The first claim follows. For the second claim, observe that this representative $\hat{h}$ satisfies the optimality condition of equation (3.8). ☐

3.2. Differentiability of G

Let us first establish the regularity of the map G.

Lemma 3.4. *For $p, k \in \mathbb{N}^*$, $p \leq k$, if $c \in \mathcal{C}^{k+p}(\mathcal{X})$, then the map $G : \mathcal{C}^k \times [0,1] \to \mathcal{C}^k$ is of class $\mathcal{C}^p$.*

Proof. The i-th component of G can be expressed as

$$G_i(\boldsymbol{\phi}, t)(x_i) = \phi_i(x_i) + \log\Big(\int_{\mathcal{X}_{-i}\times\mathcal{X}_{-i}} \mathrm{e}^{\sum_{j\neq i}\phi_j((1-t)x_j+ty_j)-c(x_i,(1-t)x_{-i}+ty_{-i})}\mathrm{d}\gamma_{-i}(x_{-i}, y_{-i})\Big)$$

Fixing i and $(x_{-i}, y_{-i}) \in \mathcal{X}_{-i} \times \mathcal{X}_{-i}$, let us observe that when $c \in \mathcal{C}^{k+p}(\mathcal{X})$, the curve $t \in [0,1] \mapsto c(., (1-t)x_{-i} + ty_{-i}) \in \mathcal{C}^k(\mathcal{X}_i)$ is of class $\mathcal{C}^p$ and that its derivatives up to order p can be bounded independently of (x_{-i}, y_{-i}). Now, for $j \neq i$ (and fixed x_j and y_j in $\mathcal{X}_j$), consider the real-valued map $L_j : (\phi_j, t) \in \mathcal{C}^k(\mathcal{X}_j) \times [0,1] \mapsto \phi_j(x_j + t(y_j - x_j))$. For $k = 1$, this map admits partial derivatives with respect to t and ϕ_j which are given respectively by $\nabla\phi_j(x_j + t(y_j - x_j))^\top(y_j - x_j)$ and $L_j(., t)$, both being continuous (for the $\mathcal{C}^1$ norm for ϕ_j) so that $L_j \in \mathcal{C}^1(\mathcal{C}^1(\mathcal{X}_j) \times [0,1], \mathbb{R})$ note also that the first-order partial derivatives of L_j can be bounded by a constant depending on the $\mathcal{C}^1$ norm of ϕ_j but not on x_j, y_j.

For $k \geq 2$, we can argue inductively. Indeed, by the previous argument, showing k times continuous differentiability of L_j amounts to showing $k-1$ times continuous differentiability of L_j applied to $\nabla\phi_j(.)^\top(y_j - x_j)$ and t. This shows that $L_j \in \mathcal{C}^k(\mathcal{C}^k(\mathcal{X}_j) \times [0,1], \mathbb{R})$, with bounds on derivatives up to order k controlled by the $\mathcal{C}^k$ norm of ϕ_j independently of $(x_j, y_j) \in \mathcal{X}_j \times \mathcal{X}_j$. By the chain rule and differentiating under the integral sign by dominated convergence, we can readily conclude that G is of class $\mathcal{C}^{\min\{k,p\}} = \mathcal{C}^p$ from $\mathcal{C}^k(\mathcal{X}) \times [0,1]$ to $\mathcal{C}^k(\mathcal{X})$. ☐

We now give a quantitative regularity estimate for the partial derivative of G in its real variable t.

Lemma 3.5. *Let $k \in \mathbb{N}^*$ and assume that $c \in \mathcal{C}^{k+1}(\mathcal{X})$. Given $\boldsymbol{\phi} \in \tilde{\mathcal{C}}^1$, the partial differential of G in t satisfies*

$$\|D_t G(\boldsymbol{\phi}, t)\|_{\tilde{\mathcal{C}}^k} \leq C\sqrt{\mathrm{cost}(\gamma)}$$

where $C > 0$ only depends on $\|\,\boldsymbol{\phi}\,\|_{\tilde{\mathcal{C}}^1}$ and $\|c\|_{\mathcal{C}^{k+1}}$ and $\mathrm{cost}(\gamma)$ is the transport cost associated with γ as in Theorem 2.3.

Proof. By Lemma 3.4, G is differentiable in t. Using the shorthand $x^t \coloneqq (1-t)x + ty$ and again the convention $y_i = x_i$, it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}G_i(\boldsymbol{\phi}, t)(x_i) = \int_{\mathcal{X}_{-i}\times\mathcal{X}_{-i}} \Big(\sum_{j\neq i}(y_j - x_j)^\top(\nabla\phi_j(x_j^t) - \nabla_j c(x^t))\Big)\mathrm{d}Q^t_{-i}(x_{-i}, y_{-i}|x_i) \tag{3.9}$$

with $Q^t_{-i} \coloneqq q^t_{-i}\gamma_{-i} \in \prod_{j\neq i}\mathcal{P}(\mathcal{X}_j \times \mathcal{X}_j)$ and, posing $(x')^t_i = x^t_i$,

$$q^t_{-i}(x_{-i}, y_{-i}|x_i) \coloneqq \frac{\mathrm{e}^{\sum_{j\neq i}\phi_j(x^t_j)-c(x^t)}}{\int_{\mathcal{X}_{-i}\times\mathcal{X}_{-i}} \mathrm{e}^{\sum_{j\neq i}\phi_j((x')^t_j)-c((x')^t)}\mathrm{d}\gamma_{-i}(x'_{-i}, y'_{-i})}.$$

Reasoning as in Lemma 3.1, this function q^t_{-i} admits positive upper and lower bounds only depending on $\|\,\boldsymbol{\phi}\,\|_{\tilde{\mathcal{C}}_0}$ and $\|c\|_{\mathcal{C}^0}$. Let us now control equation (3.9), starting with a control in uniform norm. First, by Cauchy-Schwarz in $L^2(Q^t_{-i}(\cdot,\cdot|x_i))$, for $i \in [N]$ and $x_i \in \mathcal{X}_i$,

$$\Big|\frac{\mathrm{d}}{\mathrm{dt}}G_i(\boldsymbol{\phi},t)(x_i)\Big|^2 \\ \leq \left(\int_{\mathcal{X}_{-i}\times\mathcal{X}_{-i}} \|y-x\|^2 \mathrm{d}Q^t_{-i}(x_{-i},y_{-i}|x_i)\right)\left(\int_{\mathcal{X}_{-i}\times\mathcal{X}_{-i}} \|(\nabla\boldsymbol{\phi}-\nabla c)(x^t)\|^2 \mathrm{d}Q^t_{-i}(x_{-i},y_{-i}|x_i)\right) \quad (3.10)$$

where $\nabla\boldsymbol{\phi} \coloneqq (\nabla\phi_1,\ldots,\nabla\phi_N)$. Observe that the second factor is uniformly bounded for $x_i \in \mathcal{X}_i$ because $Q^t_{-i}(\cdot,\cdot|x_i)$ is a probability measure and both $\boldsymbol{\phi}$ and c are continuously differentiable on a compact set. It follows,

$$\begin{aligned}\Big\|\frac{\mathrm{d}}{\mathrm{dt}}G_i(\boldsymbol{\phi},t)\Big\|_{\mathcal{C}^0} &\leq C\sup_{x_i\in\mathcal{X}_i}\left(\int_{\mathcal{X}_{-i}\times\mathcal{X}_{-i}} \|y-x\|^2 \mathrm{d}Q^t_{-i}(x_{-i},y_{-i}|x_i)\right)^{1/2} \\ &\leq C'\left(\int_{\mathcal{X}_{-i}\times\mathcal{X}_{-i}} \|y-x\|^2 \mathrm{d}\gamma_{-i}(x_{-i},y_{-i})\right)^{1/2} \\ &\leq C'\sqrt{\mathrm{cost}(\gamma)}\end{aligned}$$

where C, C' depend on $\|c\|_{\mathcal{C}^1}$ and $\|\boldsymbol{\phi}\|_{\tilde{\mathcal{C}}^1}$ only. Moreover, one can further differentiate equation (3.9) in x_i and obtain analogous bounds because this variable only appears in the term $\nabla_j c$ which is of regularity $\mathcal{C}^k$ and in the factor q^t_{-i} which is of regularity $\mathcal{C}^{k+1}$. With this reasoning, it follows

$$\Big\|\frac{\mathrm{d}}{\mathrm{dt}}G_i(\boldsymbol{\phi},t)\Big\|_{\tilde{\mathcal{C}}^k} \leq C_k\sqrt{\mathrm{cost}(\gamma)}.$$

where C_k depends on $\|c\|_{\tilde{\mathcal{C}}^{k+1}}$ and $\|\boldsymbol{\phi}\|_{\tilde{\mathcal{C}}^1}$ only.

□

3.3. Proof of Theorem 2.3

Proof. Let us first observe that c can be extended in a $\mathcal{C}^{k+p}$ way to a convex open set containing $\mathcal{X}$. One can therefore extend by extrapolation the definition of $\boldsymbol{\mu}^t$ to an open time interval $(-\varepsilon, 1+\varepsilon)$, for some $\varepsilon > 0$ containing $[0,1]$. We shall then apply the Implicit Function Theorem (IFT) to $G : \tilde{\mathcal{C}}^k \times [0,1] \to \tilde{\mathcal{C}}^k$ defined in (3.1).

The existence and continuity on $[0,1]$ of the Schrödinger map $t \mapsto \boldsymbol{\phi}^t$ is guaranteed by Proposition 2.2.

In Lemma 3.4, we have shown that G is of class $\mathcal{C}^p$ and in Lemma 3.2, we have shown that $D_{\boldsymbol{\phi}}G(\boldsymbol{\phi}^t,t) = D_{\boldsymbol{\phi}}T(\boldsymbol{\phi}^t,\boldsymbol{\mu}^t)$ is an invertible linear self-map of $\tilde{\mathcal{C}}^k$. Thus all the hypotheses are gathered to apply the Implicit Function Theorem, see *e.g.* [36], Theorem 10.2.1: the map $t \mapsto \boldsymbol{\phi}^t$ is of class $\mathcal{C}^p$ on $[0,1]$ and its derivative is given by

$$D_t\boldsymbol{\phi}^t = -[D_{\boldsymbol{\phi}}T(\boldsymbol{\phi}^t,\boldsymbol{\mu}^t)]^{-1}\circ D_tG(\boldsymbol{\phi}^t,t).$$

Moreover, we have by respectively Lemma 3.2 and Lemma 3.5 that there exists $C > 0$ only depending on N, $\|c\|_{\mathcal{C}^{k+1}}$ and $\|\boldsymbol{\phi}^t\|_{\tilde{\mathcal{C}}^1}$ such that

$$\|[D_{\boldsymbol{\phi}}T(\boldsymbol{\phi}^t,\boldsymbol{\mu}^t)]^{-1}\|_{\mathrm{op}} \leq C \qquad \text{and} \qquad \|D_tG(\boldsymbol{\phi}^t,t)\|_{\mathrm{op}} \leq C\sqrt{\mathrm{cost}(\gamma)}.$$

Since we know by Proposition 2.2 that $\| \boldsymbol{\phi}^t \|_{\tilde{\mathcal{C}}^1}$ is a priori bounded by $\|c\|_{\mathcal{C}^1}$, it follows that $\|D_t \boldsymbol{\phi}^t \| \leq C' \sqrt{\mathrm{cost}(\gamma)}$ for some $C'_k > 0$ that only depends on $\|c\|_{\mathcal{C}^{k+1}}$. The Lipschitz estimate in Theorem 2.3 follows by the mean-value inequality. □

3.4. Proof of Proposition 2.5

Recall that the Sobolev space $H^p(\mathcal{X}_i)$ consists of all functions $f_i \in L^2(\mathcal{X}_i)$ whose partial derivatives up to order p belong to $L^2(\mathcal{X}_i)$ which is a Hilbert space for the norm

$$\|f_i\|^2_{H^p(\mathcal{X}_i)} := \sum_{\alpha\,:\,|\alpha|\leq p} \int_{\mathcal{X}_i} |\partial^\alpha f_i|^2.$$

If $p > d/2$, by Morrey's Theorem (see [35]), $H^p(\mathcal{X}_i)$ embeds continuously into the space of continuous functions, hence, by duality, measures belong to the dual space $H^{-p}(\mathcal{X}_i)$. We can therefore define

$$\| \boldsymbol{\mu} - \boldsymbol{\nu} \|_{H^{-p}} := \sum_{i=1}^N \|\mu_i - \nu_i\|_{H^{-p}}$$

where

$$\|\mu_i - \nu_i\|_{H^{-p}} := \sup\Big\{ \int_{\mathcal{X}_i} f_i \mathrm{d}(\mu_i - \nu_i) \ : \ \|f_i\|_{H^p(\mathcal{X}_i)} \leq 1 \Big\}.$$

To obtain the bound announced in Proposition 2.5, we simply consider the linear interpolation between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, $\boldsymbol{\mu}^t := \boldsymbol{\mu} + t(\boldsymbol{\nu} - \boldsymbol{\mu})$ for $t \in [0,1]$ and $G(\boldsymbol{\phi}, t) := T(\boldsymbol{\phi}, \boldsymbol{\mu}^t)$ as well as $\boldsymbol{\phi}^t := S(\boldsymbol{\mu}^t) \in \tilde{\mathcal{C}}^k$ i.e. $G(\boldsymbol{\phi}^t, t) = 0$. Recall that $\boldsymbol{\phi}^t$ is bounded in $\mathcal{C}^k$ by a constant that only depends on c. The same holds for the operator norm of $[D_{\boldsymbol{\phi}} T(\boldsymbol{\phi}^t, \boldsymbol{\mu}^t)]^{-1}$ in $\mathcal{C}^k$ as well. To conclude as before by the implicit function theorem, we have to differentiate G with respect to t and bound the $\mathcal{C}^k$ norm of $D_t G(\phi^t, t)$ by a constant depending on c times $\| \boldsymbol{\mu} - \boldsymbol{\nu} \|_{H^{-p}}$. To simplify notations, let us set

$$\xi_i(\boldsymbol{\phi}, t)(x_i) := \frac{1}{\int_{\mathcal{X}_{-i}} \mathrm{e}^{-c(x_i, x_{-i}) + \sum_{j\neq i} \phi_j(x_j)} \mathrm{d}\mu^t_{-i}(x_{-i})}$$

and observe that $\xi_i(\boldsymbol{\phi}^t, t)(.)$ has uniformly bounded derivatives up to order k (with bounds that depend on $\|c\|_{\mathcal{C}^k}$ only). If $N = 2$, we simply have

$$D_t G_1(\boldsymbol{\phi}, t)(x_1) = \xi_1(\boldsymbol{\phi}, t)(x_1) \int_{\mathcal{X}_2} \mathrm{e}^{-c(x_1, x_2) + \phi_2(x_2)} \mathrm{d}(\nu_2 - \mu_2)(x_2).$$

By Leibniz formula to bound the k first derivatives of $D_t G_1(\boldsymbol{\phi}^t, t)(x_1)$, we then just have to bound the k first derivatives of $x_1 \in \mathcal{X}_1 \mapsto \int_{\mathcal{X}_2} \mathrm{e}^{-c(x_1,x_2)+\phi_2^t(x_2)} \mathrm{d}(\nu_2 - \mu_2)(x_2)$ which are obviously controlled by $\|\nu_2 - \mu_2\|_{H^{-p}}$ times the $\mathcal{C}^{k+p}$ norm of $(x_1, x_2) \mapsto \mathrm{e}^{-c(x_1,x_2)+\phi_2^t(x_2)}$ which can in turn be bounded by a constant only depending on $\|c\|_{\mathcal{C}^{k+p}}$. Proceeding in the same way for $x_2 \mapsto D_t G_2(\boldsymbol{\phi}^t, t)(x_2)$ gives the desired result. The case $N \geq 3$ is slightly more tedious to write, for a fixed pair of indices $i \neq j$, we denote by $\mathcal{X}_{-(i,j)}$ the cartesian product of all the $\mathcal{X}_l$ but i and j, and write $x \in \mathcal{X}$ as $x = (x_i, x_j, x_{-(i,j)}) \in \mathcal{X}_i \times \mathcal{X}_j \times \mathcal{X}_{-(i,j)}$, likewise we write $\mu^t_{-(i,j)}$ for the tensor product of μ_l^t for $l \neq i$, $l \neq j$. Doing so, we have

$$D_t G_i(\boldsymbol{\phi}, t)(x_i) = \xi_i(\boldsymbol{\phi}, t)(x_i) \sum_{j \neq i} \int_{\mathcal{X}_j} h_{ij}(\boldsymbol{\phi}, t)(x_i, x_j) \mathrm{d}(\nu_j - \mu_j)(x_j)$$

where

$$h_{ij}(\boldsymbol{\phi}, t)(x_i, x_j) := \int_{\mathcal{X}_{-(i,j)}} \mathrm{e}^{-c(x_i, x_j, x_{-(i,j)}) + \phi_j(x_j) + \sum_{l \neq i, l \neq j} \phi_l(x_l)} \mathrm{d}\mu_{-(i,j)}^t(x_{-(i,j)})$$

so that by the same arguments as before, if $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq k$, we have for a constant C only depending on N and the $\mathcal{C}^{k+p}$ norm of c, possibly varying from one line to another:

$$|\partial^\alpha D_t G_i(\boldsymbol{\phi}^t, t)(x_i)| \leq C \sum_{j \neq i} \|\mu_j - \nu_j\|_{H^{-p}(\mathcal{X}_j)} \|\partial_{x_i}^\alpha h_{ij}(\boldsymbol{\phi}^t, t)(x_i, .)\|_{H^p(\mathcal{X}_j)} \leq C \| \boldsymbol{\mu} - \boldsymbol{\nu} \|_{H^{-p}}.$$

This enables us to conclude exactly as in the end of Section 3.3.

4. Smoothness of entropic optimal transport and Wasserstein gradient flows

In this section, we apply our main stability results to the analysis of Wasserstein gradient flows of functionals involving the entropic optimal transport functional E.

4.1. Displacement smoothness and gradient flows

Let $\boldsymbol{\mu}^0, \boldsymbol{\mu}^1, \gamma, \boldsymbol{\mu}^t$ and $\boldsymbol{\phi}^t$ be as in the beginning of Section 2.3. A consequence of Theorem 2.3 is that the functional E is as nice as one could hope for in the Wasserstein space.

Theorem 4.1. *If $c \in \mathcal{C}^{2k-1}(\mathcal{X})$ for some $k \geq 1$, then the function $t \mapsto E(\boldsymbol{\mu}^t)$ is of class $\mathcal{C}^k$. Moreover, if $c \in \mathcal{C}^2(\mathcal{X})$ then its derivative is $C\,\mathrm{cost}(\gamma)$-Lipschitz, for some $C > 0$ that only depends on N and $\|c\|_{\mathcal{C}^2}$. In particular both E and $-E$ are $(-C)$-displacement convex.*

Proof. It follows from the dual formulation (2.2) that $\boldsymbol{\mu} \mapsto E(\boldsymbol{\mu})$ is a convex function of $\otimes_i \mu_i$ (eventhough E is not convex) by optimality of $\boldsymbol{\phi}^t$ in this dual formulation E, setting $V_t(x) := \mathrm{e}^{-c(x) + \sum_{i=1}^N \phi_i^t(x_i)}$ it holds that for every t and s in $[0, 1]$, one has

$$\sum_i \int_{\mathcal{X}_i} \phi_i^t(\mu_i^s - \mu_i^t) - \int_{\mathcal{X}} V_t(\otimes_i \mu_i^s - \otimes_i \mu_i^t) \leq E(\boldsymbol{\mu}^s) - E(\boldsymbol{\mu}^t) \leq \sum_i \int_{\mathcal{X}_i} \phi_i^s(\mu_i^s - \mu_i^t) - \int_{\mathcal{X}} V_s(\otimes_i \mu_i^s - \otimes_i \mu_i^t) \quad (4.1)$$

Using the notation $x_i^t = (1-t)x_i + t y_i$ as before, remark that

$$\begin{aligned} \int_{\mathcal{X}_i} \phi_i^t \mathrm{d}(\mu_i^s - \mu_i^t) &= \int_{\mathcal{X}_i^2} \left(\phi_i^t(x_i^s) - \phi_i^t(x_i^t)\right) \mathrm{d}\gamma_i(x_i, y_i) \\ &= (s-t) \int_{\mathcal{X}_i^2} (y_i - x_i)^\top \nabla \phi_i^t(x_i^t) \mathrm{d}\gamma_i(x_i, y_i) + o(|s-t|). \end{aligned}$$

We now claim that the second term in the left hand side of (4.1) is $o(|s-t|)$. To prove this, we shall for notational simplicity restrict ourselves to the case $N=2$ (the general case is similar but more tedious),

$$\int_{\mathcal{X}} V_t(\mu_1^s \otimes \mu_2^s - \mu_1^t \otimes \mu_2^t) = \int_{\mathcal{X}} V_t(\mu_1^t \otimes (\mu_2^s - \mu_2^t) + (\mu_1^s - \mu_1^t) \otimes \mu_2^t) + \int_{\mathcal{X}} V_t(\mu_1^s - \mu_1^t) \otimes (\mu_2^s - \mu_2^t)$$

the first term in the right-hand side is 0 because the integral of $V_t(.,x_2)$ (respectively $V_t(x_1,.)$ with respect to μ_1^t (respectively μ_2^t) is constant equal to 1 and μ_2^s and μ_2^t (respectively μ_1^s and μ_1^t) have the same total mass. We are therefore left to show that the second term is $o(|t-s|)$. Defining for $x_1 \in \mathcal{X}_1$, $\xi_{s,t}(x_1) := \int_{\mathcal{X}_2} V_t(x_1,.)(\mu_2^s - \mu_2^t)$ and observing that since the 1-Wasserstein distance between μ_2^s and μ_2^t is bounded by $M|t-s|$ where M is the diameter of $\mathcal{X}_2$, it follows from the Kantorovich-Rubinstein inequality that

$$|\int_{\mathcal{X}} V_t(\mu_1^s - \mu_1^t) \otimes (\mu_2^s - \mu_2^t)| \leq M|t-s| \ \|\nabla \xi_{s,t}\|_{\mathcal{C}^0(\mathcal{X}_1)}.$$

Writing $\nabla \xi_{s,t}(x_1)$ as

$$\nabla \xi_{s,t}(x_1) = \nabla \phi_1^t(x_1) \xi_{t,s}(x_1) - \int_{\mathcal{X}_2} V_t(x_1,.) \nabla_{x_1} c(x_1,.)(\mu_2^s - \mu_2^t)$$

we deduce from the uniform continuity of $\nabla_{x_1} c$ and V_t and the weak $*$ continuity of $s \mapsto \mu_2^s$ that $\nabla \xi_{s,t}$ converges uniformly to 0 as $s \to t$, hence that

$$\int_{\mathcal{X}} V_t(\mu_1^s - \mu_1^t) \otimes (\mu_2^s - \mu_2^t) = o(|s-t|).$$

Thus dividing equation (4.1) by $|s-t|$ and using that $\boldsymbol{\phi}^s \to \boldsymbol{\phi}^t$ in $\tilde{\mathcal{C}}^1$ as s tends to t, we get:

$$\frac{\mathrm{d}}{\mathrm{d}t} E(\boldsymbol{\mu}^t) = \sum_{i=1}^N \int_{\mathcal{X}_i \times \mathcal{X}_i} (y_i - x_i)^\top \nabla \phi_i^t(x_i^t) \mathrm{d}\gamma_i(x_i, y_i). \tag{4.2}$$

From Theorem 2.3, we know that $t \mapsto \boldsymbol{\phi}^t$ is in $\mathcal{C}^{k-1}([0,1], \tilde{\mathcal{C}}^k)$ (note that the case $k=1$ is instead a consequence of Prop. 2.2) and hence $t \mapsto ((x_i, y_i) \mapsto \nabla\, \boldsymbol{\phi}_i^t(x_i^t))$ is in $\mathcal{C}^{k-1}([0,1], \mathcal{C}^0(\mathcal{X}_i \times \mathcal{X}_i))$.

It follows that $h : t \mapsto E(\boldsymbol{\mu}^t) \in \mathcal{C}^k([0,1])$. Notice how this argument uses the two notions of regularity of the Schrödinger map (indexed by p and k in Them. 2.3).

For the Lipschitz regularity of h', fixing $s,t \in [0,1]$, one has

$$\begin{aligned} |h'(t) - h'(s)| &\leq \left| \sum_{i=1}^N \int_{\mathcal{X}_i \times \mathcal{X}_i} (y_i - x_i)^\top \big(\nabla \phi_i^t(x_i^t) - \nabla \phi_i^s(x_i^s)\big) \mathrm{d}\gamma_i(x_i, y_i) \right| \\ &\leq \sum_{i=1}^N \int_{\mathcal{X}_i \times \mathcal{X}_i} \|y_i - x_i\| \|\nabla \phi_i^t(x_i^t) - \nabla \phi_i^t(x_i^s)\| \mathrm{d}\gamma_i(x_i, y_i) \\ &\quad + \sum_{i=1}^N \int_{\mathcal{X}_i \times \mathcal{X}_i} \|y_i - x_i\| \|\nabla \phi_i^t(x_i^s) - \nabla \phi_i^s(x_i^s)\| \mathrm{d}\gamma_i(x_i, y_i). \end{aligned}$$

Now, if $c \in \mathcal{C}^2(\mathcal{X})$, using the Lipschitz regularity of $x_i \mapsto \nabla\phi_i^t(x_i)$ and of $t \mapsto \nabla\,\boldsymbol{\phi}^t$, from Theorem 2.3, it follows that

$$\|\nabla\phi_i^t(x_i^t) - \nabla\phi_i^t(x_i^s)\| \le C|t-s|\|y_i - x_i\| \text{ and } \|\nabla\phi_i^t(x_i^s) - \nabla\phi_i^s(x_i^s)\| \le C|t-s|\sqrt{\mathrm{cost}(\gamma)},$$

for some C that only depends on N and $\|c\|_{\mathcal{C}^2}$. Then, we obtain

$$\begin{aligned}|h'(t) - h'(s)| &\le C|t-s|\sum_{i=1}^N \int_{\mathcal{X}_i\times\mathcal{X}_i} \|y_i - x_i\|^2 \mathrm{d}\gamma_i(x_i,y_i) \\ &\qquad + C|t-s|\sqrt{\mathrm{cost}(\gamma)}\sum_{i=1}^N \int_{\mathcal{X}_i\times\mathcal{X}_i} \|y_i - x_i\| \mathrm{d}\gamma_i(x_i,y_i) \\ &\le C|t-s|\,\mathrm{cost}(\gamma).\end{aligned}$$

In particular, this implies that $t \in [0,1] \mapsto h(t) + \frac{C\,\mathrm{cost}(\gamma)}{2}t^2$ is convex hence

$$E(\boldsymbol{\mu}^t) \le (1-t)E(\boldsymbol{\mu}^0) + tE(\boldsymbol{\mu}^1) + \frac{C\,\mathrm{cost}(\gamma)t(1-t)}{2}$$

and displacement semi-convexity follows by choosing γ_i to be an optimal transport plan between μ_i^0 and μ_i^1 for each $i \in [N]$ (see [38] for a definition). Displacement semi-convexity of $-E$ is obtained in the same way, observing that $t \in [0,1] \mapsto -h(t) + \frac{C\,\mathrm{cost}(\gamma)}{2}t^2$ is convex. □

Proposition 4.2. *If $c \in \mathcal{C}^1(\mathcal{X})$, we have that $S(\boldsymbol{\mu})$ is the gradient of $\boldsymbol{\mu} \mapsto E(\boldsymbol{\mu})$ and $x \mapsto \nabla_x S(\boldsymbol{\mu})(x)$ is its Wasserstein gradient, in the sense of [38] i.e. for $\boldsymbol{\mu}^0$ and $\boldsymbol{\mu}^1$ in $\prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$ and for any γ_i optimal plan between μ_i^0, and μ_i^1, one has*

$$E(\boldsymbol{\mu}^1) - E(\boldsymbol{\mu}^0) = \sum_{i=1}^N \int_{\mathcal{X}_i\times\mathcal{X}_i} (y_i - x_i)^\top \nabla\phi_i(x_i)\mathrm{d}\gamma_i(x_i,y_i) + o(\boldsymbol{W_2}(\boldsymbol{\mu}^0,\boldsymbol{\mu}^1))$$

where $\boldsymbol{\phi} := S(\boldsymbol{\mu}^0)$. If $c \in \mathcal{C}^2(\mathcal{X})$ then the error $o(\boldsymbol{W_2}(\boldsymbol{\mu}^0,\boldsymbol{\mu}^1))$ is in fact $O(\boldsymbol{W_2}(\boldsymbol{\mu}^0,\boldsymbol{\mu}^1)^2)$.

Proof. For the case $c \in \mathcal{C}^1(\mathcal{X})$, this follows by integrating (4.2) in time and Proposition 2.2(ii), which guarantees that S is weakly continuous as a function in $\tilde{\mathcal{C}}^1$. For the case $c \in \mathcal{C}^2(\mathcal{X})$, this follows from (4.2) and Theorem 2.3. □

Theorem 4.1 and the above identification of the Wasserstein gradient of E enable us to deduce from [38], Theorem 11.2.1 that E admits a unique Wasserstein gradient flow, which shows well-posedness of the Cauchy problem for the system of PDEs

$$\begin{cases} \partial_t\mu_i = \nabla\cdot(\mu_i\nabla S_i(\boldsymbol{\mu})), \quad i \in \{1,\dots,N\} \\ \boldsymbol{\mu}\,|_{t=0} = \boldsymbol{\mu}^0 \end{cases}$$

This system, as all the PDEs below, is understood in the sense of distributions with no-flux boundary conditions, i.e. for $i \in \{1,\dots,N\}$, for every $\psi \in \mathcal{C}_c^\infty([0,+\infty)\times\mathbb{R}^d)$ it holds

$$\int_0^\infty \int_{\mathcal{X}_i} \big(\partial_t\psi(t,x_i) + \nabla S_i(\boldsymbol{\mu})(x_i)^\top\nabla_{x_i}\psi(t,x_i)\big)\mathrm{d}\mu_i^t(x_i)\mathrm{d}t = -\int_{\mathcal{X}_i}\psi(0,x_i)\mathrm{d}\mu_i^0(x_i).$$

We also have that the fact that the gradient flow map $\boldsymbol{\mu}^0 \mapsto \boldsymbol{\mu}^t$ satisfies

$$\boldsymbol{W_2}(\boldsymbol{\mu}^t, \boldsymbol{\nu}^t)^2 \leq \mathrm{e}^{Ct}\, \boldsymbol{W_2}(\boldsymbol{\mu}^0, \boldsymbol{\nu}^0)^2.$$

Of course, adding to E a separable term of the form $\sum_{i=1}^N E_i(\mu_i)$ where each E_i is displacement semi-convex, we can deduce well-posedness for more general systems like

$$\partial_t \mu_i - \alpha_i \Delta \mu_i - \nabla \cdot (\mu_i \nabla S_i(\boldsymbol{\mu})) = 0, i = 1, \ldots, N, \quad \boldsymbol{\mu}\,|_{t=0} = \boldsymbol{\mu}^0$$

or

$$\partial_t \mu_i - \alpha_i \Delta \mu_i^{m_i} - \nabla \cdot (\mu_i \nabla S_i(\boldsymbol{\mu})) = 0, i = 1, \ldots, N, \quad \boldsymbol{\mu}\,|_{t=0} = \boldsymbol{\mu}^0$$

with $m_i \geq 1$ and $\alpha_i \geq 0$. For the sake of concreteness, we are going to detail three such examples with interesting additional structure in the next paragraphs.

4.2. Wasserstein gradient flow of the Sinkhorn divergence

We consider the Sinkhorn divergence functional [16], the gradient flow of which has been previously considered as a numerical method for density fitting. As a consequence of our analysis and of [38], Theorem 11.2.1 we have the following result.

Proposition 4.3. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact convex set, $c \in \mathcal{C}^2(\mathcal{X} \times \mathcal{X})$ and let $\mu^0, \nu \in \mathcal{P}(\mathcal{X})$. There exists a unique Wasserstein gradient flow starting from μ^0 of the Sinkhorn divergence (from ν) functional*

$$\mu \mapsto E(\mu, \nu) - \frac{1}{2} E(\mu, \mu) - \frac{1}{2} E(\nu, \nu).$$

Here, a Wasserstein gradient flow is a curve $(\mu^t)_{t\geq 0} \in \mathcal{P}(\mathcal{X})$ that is absolutely continuous for the W_2 metric and that satisfies

$$\partial_t \mu^t = \nabla \cdot (v^t \mu^t), \qquad v^t = \nabla S_1(\mu^t, \nu) - \frac{1}{2}(\nabla S_1(\mu^t, \mu^t) + \nabla S_2(\mu^t, \mu^t)), \qquad \mu|_{t=0} = \mu^0 \tag{4.3}$$

where we recall that S is the Schrödinger map.

An interesting open question is whether this dynamics can be provably shown to converge to the unique minimizer μ^*, which is $\mu^* = \nu$ for suitable choices of costs, *e.g.* for $c(x, y) = \|y - x\|^2$, as proved in [16].

4.3. Convergence to equilibrium for the Schrödinger bridge energy

Let us continue with another simple example that shows that our theory is also natural to deal with the Lebesgue measure as a reference in the definition of E equation (1.1), which is the original definition of the Schrödinger bridge problem. This alternative definition is equivalent (see *e.g.* [32]) to considering $E + H$ where $H(\mu) \coloneqq \int \log(\mu) \mathrm{d}\mu$ if μ is absolutely continuous and $+\infty$ otherwise is minus the differential entropy.

Proposition 4.4. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact convex set, $c \in \mathcal{C}^2(\mathcal{X} \times \mathcal{X})$ and let $\mu^0, \nu \in \mathcal{P}(\mathcal{X})$. There exists a unique Wasserstein gradient flow of*

$$\mu \mapsto E(\mu, \nu) + H(\mu)$$

starting from μ^0. Moreover, if $H(\mu^0) < \infty$ then this gradient flow converges at an exponential rate to the unique global minimizer μ^*. Specifically, there exists $\kappa > 0$ independent of μ^0, ν such that

$$F(\mu^t) - F(\mu^*) \leq \mathrm{e}^{-\kappa t}(F(\mu^0) - F(\mu^*)).$$

In addition, there exists a constant $C > 0$, independent of μ^0, ν such that

$$W_2(\mu^t, \mu^*)^2 \leq C\mathrm{e}^{-\kappa t}(F(\mu^0) - F(\mu^*)).$$

For this functional, the Wasserstein gradient flow $(\mu^t)_{t\geq 0} \in \mathcal{P}(\mathcal{X})$ solves

$$\partial_t \mu^t = \nabla \cdot (v^t \mu^t) + \Delta \mu^t, \qquad v^t = \nabla S_1(\mu^t, \nu). \tag{4.4}$$

Proof. We have semi-convexity along Wasserstein geodesics, by Theorem 4.1 for the first component and by a standard result due to Mc Cann [37] (see [2], Thm. 7.28) for the H component. Thus the general well-posedness results from [38], Theorem 11.2.1 applies. For the exponential convergence – in function value and in distance – we apply the result from [39], Theorem 3.2, see also [40] where the same argument was discovered independently (although stated on $\mathbb{R}^d$, the argument goes through on a compact domain).

The main assumptions to check are that (i) $\mu \mapsto E(\mu, \nu)$ is convex, which is clear from the dual formulation equation (2.2) which expresses this functional as a supremum of affine forms, (ii) that a global minimizer μ^* exists, which is not difficult here since $\mathcal{P}(\mathcal{X})$ is weakly compact, H is weakly lower-semicontinuous and E is weakly continuous and finally we need to check that the probability measure $\hat{\mu}_t \propto \mathrm{e}^{-S_1(\mu^t,\nu)} \in \mathcal{P}(\mathcal{X})$ satisfies a log-Sobolev inequality, uniformly in t (Asm. 3 in [39]).

Since $\mathcal{X}$ is bounded, the normalized Lebesgue measure satisfies a log-Sobolev inequality [41], Theorem 7.3. By the Holley-Stroock perturbation criterion [42] (see [41], Lem. 1.2), $\hat{\mu}_t$ satisfies it as well; this criterion applies here because $\sup_x S_1(\mu, \nu)(x) - \inf_x S_1(\mu, \nu)(x)$ is bounded, uniformly in $\mu, \nu \in \mathcal{P}(\mathcal{X})$ by Prop. 2.2. The convergence in Wasserstein distance is stated in [39], Corollary 3.3 and follows from the fact that log-Sobolev inequalities imply Talagrand inequalities [43], Theorem 1.

□

4.4. Convergence to equilibrium in the multi-species case

We now consider Wasserstein gradient flow of[3]

$$F(\boldsymbol{\mu}) := E(\boldsymbol{\mu}) + \sum_{i=1}^{N} H(\mu_i), \quad \boldsymbol{\mu} = (\mu_1, \ldots, \mu_N) \in \prod_{i=1}^{N} \mathcal{P}(\mathcal{X}_i)$$

that is

$$\partial_t \mu_i - \Delta \mu_i - \nabla \cdot (\mu_i \nabla S_i(\boldsymbol{\mu})) = 0, \quad i = 1, \ldots, N, \ \boldsymbol{\mu}\,|_{t=0} = \boldsymbol{\mu}^0. \tag{4.5}$$

Up to adding a constant to c (which does not affect the dynamics (4.5)) we may assume that

$$\int_{\mathcal{X}} \mathrm{e}^{-c(x)} \mathrm{d}x = 1. \tag{4.6}$$

[3] Note that $F(\boldsymbol{\mu})$ can also be written as the value of an entropic optimal transport problem but with the Lebesgue measure as reference measure i.e. $F(\boldsymbol{\mu}) = \min_{\gamma \in \Pi(\boldsymbol{\mu})} \int_{\mathcal{X}} c(x) \mathrm{d}\gamma(x) + H(\gamma)$.

With this normalization, we have

$$\inf_{\boldsymbol{\mu}\in\prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)} F(\boldsymbol{\mu}) = \inf_{\gamma\in\mathcal{P}(\mathcal{X})} H(\gamma|\mathrm{e}^{-c}) = 0$$

so that F admits $\boldsymbol{\mu}^*$, the marginals of $\gamma^* := \mathrm{e}^{-c}$, as unique minimizer

$$\mu_i^*(x_i) = \int_{\mathcal{X}_{-i}} \mathrm{e}^{-c(x_i,x_{-i})}\mathrm{d}x_{-i} \tag{4.7}$$

for every i and $x_i \in \mathcal{X}_i$ and $F(\boldsymbol{\mu}^*) = 0$. In the next proposition, we extend Proposition 4.4 to the multi-species case. In this case, the functional $\boldsymbol{\mu} \mapsto E(\boldsymbol{\mu})$ is not convex anymore but we can overcome this difficulty taking advantage of the form of F.

Proposition 4.5. *Let $\boldsymbol{\mu}^0 \in \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$. Then there exists a unique Wasserstein gradient flow of F starting from $\boldsymbol{\mu}^0$, that we call $\boldsymbol{\mu}^t$. Assume that $H(\mu_i^0) < +\infty$ for every i, then $\boldsymbol{\mu}^t$ converges at an exponential rate to the equilibrium $\boldsymbol{\mu}^*$, defined in equation (4.7), i.e. there exists $\kappa > 0$ independent of $\boldsymbol{\mu}^0$ such that*

$$F(\boldsymbol{\mu}^t) - F(\boldsymbol{\mu}^*) \leq \mathrm{e}^{-\kappa t}(F(\boldsymbol{\mu}^0) - F(\boldsymbol{\mu}^*)).$$

In addition, there exists a constant $C > 0$, independent of $\boldsymbol{\mu}^0$, such that

$$\boldsymbol{W_2}(\boldsymbol{\mu}^t, \boldsymbol{\mu}^*)^2 \leq C\mathrm{e}^{-\kappa t}(F(\boldsymbol{\mu}^0) - F(\boldsymbol{\mu}^*)).$$

Proof. The well-posedness of the Wasserstein gradient flow is proved as previously using the geodesic semi-convexity of F that follows from Theorem 4.1. For the convergence, first note the identities

$$E(\boldsymbol{\mu}) = \sum_{i=1}^N \int_{\mathcal{X}_i} S_i(\boldsymbol{\mu})\mathrm{d}\mu_i, \quad F(\boldsymbol{\mu}) = \sum_{i=1}^N H(\mu_i|\mathrm{e}^{-S_i(\boldsymbol{\mu})})$$

which hold for any $\boldsymbol{\mu} \in \prod_{i=1}^N \mathcal{P}(\mathcal{X}_i)$ and easily follow from (2.2) and (2.5). Let us then remark that, denoting by $\gamma(\boldsymbol{\mu})$, the optimal entropic plan

$$\mathrm{d}\gamma(\boldsymbol{\mu})(x) := \mathrm{e}^{-c(x)+\sum_{i=1}^N S_i(\boldsymbol{\mu})(x_i)}\mathrm{d}\mu_1(x_1)\cdots\mathrm{d}\mu_N(x_N)$$

and recalling that $\gamma^* = \mathrm{e}^{-c}$ we can conveniently rewrite F as a relative entropy with respect to the fixed probability measure γ^* on $\mathcal{X}$:

$$F(\boldsymbol{\mu}) = H(\gamma(\boldsymbol{\mu})|\gamma^*).$$

Since $H(\mu_i^0) < +\infty$ for every i, and denoting by $\boldsymbol{\mu}^t$ the Wasserstein gradient flow of F starting from $\boldsymbol{\mu}^0$, we have using the chain rule, (4.5) and an integration by parts:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}F(\boldsymbol{\mu}^t) &= \sum_{i=1}^N \int_{\mathcal{X}_i} (S_i(\boldsymbol{\mu}^t) + \log(\mu_i^t))\partial_t\mu_i^t = -\sum_{i=1}^N \int_{\mathcal{X}_i} \|\nabla \log \mu_i^t + \nabla S_i(\boldsymbol{\mu}^t)\|^2\mathrm{d}\mu_i^t \\ &= -\sum_{i=1}^N I_i(\mu_i^t|\mathrm{e}^{-S_i(\boldsymbol{\mu}^t)})\end{aligned}$$

where $I_i(\rho|\mathrm{e}^{-V})$, for $\rho \in \mathcal{P}(\mathcal{X}_i)$, stands for the relative Fisher information

$$I_i(\rho|\mathrm{e}^{-V}) := \int_{\mathcal{X}_i} \left\| \nabla \log\left(\frac{\rho}{\mathrm{e}^{-V}}\right)\right\|^2 \mathrm{d}\rho.$$

Defining $\gamma^t := \gamma(\boldsymbol{\mu}^t)$, we have $F(\boldsymbol{\mu}^t) = H(\gamma^t|\gamma^*)$ and

$$\begin{aligned} I(\gamma^t|\gamma^*) &:= \int_{\mathcal{X}} \left\| \nabla_x \log\left(\frac{\gamma^t(x)}{\mathrm{e}^{-c(x)}}\right)\right\|^2 \mathrm{d}\gamma^t(x) = \int_{\mathcal{X}} \sum_{i=1}^N \left\| \nabla_{x_i}(\log(\mu_i^t(x_i)) + S_i(\boldsymbol{\mu}^t)(x_i))\right\|^2 \mathrm{d}\gamma^t(x) \\ &= \sum_{i=1}^N I_i(\mu_i^t|\mathrm{e}^{-S_i(\boldsymbol{\mu}^t)}) \end{aligned}$$

where we used the fact that $\gamma^t \in \Pi(\boldsymbol{\mu}^t)$ in the last line. But since $\mathcal{X}$ is convex, it follows from the Holley-Stroock perturbation criterion [42] (see [41], Lem. 1.2), that γ^* satisfies a log-Sobolev inequality, hence

$$I(\gamma^t|\gamma^*) \geq \kappa H(\gamma^t|\gamma^*)$$

with $\kappa > 0$ depending only on $\mathcal{X}$ and c. We thus have

$$\frac{\mathrm{d}}{\mathrm{d}t} F(\boldsymbol{\mu}^t) = \frac{\mathrm{d}}{\mathrm{d}t} H(\gamma^t|\gamma^*) = -I(\gamma^t|\gamma^*) \leq -\kappa H(\gamma^t|\gamma^*) = -\kappa F(\boldsymbol{\mu}^t)$$

hence

$$F(\boldsymbol{\mu}^t) \leq \mathrm{e}^{-\kappa t} F(\boldsymbol{\mu}^0).$$

Thanks to Talagrand's inequality, which follows from the log-Sobolev inequality [43], Theorem 1, we get an exponential decay of $W_2(\gamma_t, \gamma^*)$ hence also an exponential decay in Wasserstein distance between the marginals of γ^t and γ^* i.e. of $\boldsymbol{W_2}(\boldsymbol{\mu}^t, \boldsymbol{\mu}^*)$. □

References

[1] C. Villani, Optimal Transport: Old and New, Vol. 338. Springer (2009).

[2] F. Santambrogio, Optimal transport for applied mathematicians. *Birkäuser, NY* **55** (2015) 94.

[3] J.J. Kosowsky and A.L. Yuille, The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Netw.* **7** (1994) 477–490.

[4] M. Cuturi, Sinkhorn distances: lightspeed computation of optimal transport. *Adv. Neural Inform. Process. Syst.* **26** (2013).

[5] A. Genevay, L. Chizat, F. Bach, M. Cuturi and G. Peyré, Sample complexity of Sinkhorn divergences, in *International Conference on Artificial Intelligence and Statistics.* PMLR (2019) 1574–1583.

[6] G. Mena and J. Niles-Weed, Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. *Adv. Neural Inform. Process. Syst.* **32** (2019).

[7] E. del Barrio, A. Gonzalez-Sanz, J.-M. Loubes and J. Niles-Weed, An improved central limit theorem and fast convergence rates for entropic transportation costs. arXiv preprint arXiv:2204.09105 (2022).

[8] S. Pal, On the difference between entropic cost and the optimal transport cost. arXiv preprint arXiv:1905.12206 (2019).

[9] J. Weed, An explicit analysis of the entropic penalty in linear programming, in *Conference On Learning Theory.* PMLR (2018) 1841–1855.

[10] L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard and G. Peyré, Faster wasserstein distance estimation with the sinkhorn divergence. *Adv. Neural Inform. Process. Syst.* (2020).

[11] G. Conforti and L. Tamanini, A formula for the time derivative of the entropic cost and applications. *J. Funct. Anal.* **280** (2021) 108964.

[12] G. Carlier, P. Pegon and L. Tamanini, Convergence rate of general entropic optimal transport costs (2022).
[13] S. Eckstein and M. Nutz, Convergence rates for regularized optimal transport via quantization (2022).
[14] E. Schrödinger, Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. Henri Poincaré* **2** (1932) 269–310.
[15] C. Léonard, From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262** (2012) 1879–1920.
[16] J. Feydy, T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouvé and G. Peyré, Interpolating between optimal transport and MMD using Sinkhorn Divergences, in *International Conference on Artificial Intelligence and Statistics*. PMLR (2019) 2681–2690.
[17] M. Laborde, Nonlinear systems coupled through multi-marginal transport problems. *Eur. J. Appl. Math.* **31** (2020) 450–469.
[18] C. Barilla, G. Carlier and J.-M. Lasry, A mean field game model for the evolution of cities. *J. Dyn. Games* **8** (2021) 299–329.
[19] L. Chizat, S. Zhang, M. Heitz and G. Schiebinger, Trajectory inference via mean-field langevin in path space. *Adv. Neural Inform. Process. Syst.*, in press. 2022.
[20] L. Chizat, Doubly regularized entropic wasserstein barycenters. arXiv preprint arXiv:2303.11844 (2023).
[21] A. Delalande and Q. Merigot, Quantitative stability of optimal transport maps under variations of the target measure. arXiv preprint arXiv:2103.05934 (2021).
[22] R.J. Berman, Convergence rates for discretized Monge–Ampère equations and quantitative stability of optimal transport. *Found. Computat. Math.* **21** (2021) 1099–1140.
[23] N. Gigli, On hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proc. Edinb. Math. Soc.* **54** (2011) 401–409.
[24] G. Carlier and M. Laborde, A differential approach to the multi-marginal Schrödinger system. *SIAM J. Math. Anal.* **52** (2020) 709–717.
[25] Z. Goldfeld, K. Kato, G. Rioux and R. Sadhu, Limit theorems for entropic optimal transport maps and the sinkhorn divergence. arXiv preprint arXiv:2207.08683 (2022).
[26] S. Eckstein and M. Nutz, Quantitative stability of regularized optimal transport and convergence of Sinkhorn's algorithm. arXiv preprint arXiv:2110.06798 (2021).
[27] M. Nutz and J. Wiesel, Stability of Schrödinger potentials and convergence of Sinkhorn's algorithm. arXiv preprint arXiv:2201.10059 (2022).
[28] P. Ghosal, M. Nutz and E. Bernton, Stability of entropic optimal transport and Schrödinger bridges. arXiv preprint arXiv:2106.03670 (2021).
[29] G. Deligiannidis, V. De Bortoli and A. Doucet, Quantitative uniform stability of the iterative proportional fitting procedure. arXiv preprint arXiv:2108.08129 (2021).
[30] H. Queffélec and C. Zuily, Analyse pour l'agrégation-Agrégation/Master Mathématiques. Dunod (2020).
[31] M. Nutz, Introduction to entropic optimal transport (2021).
[32] S. Di Marino and A. Gerolin, An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *J. Sci. Comput.* **85** (2020) 1–28.
[33] P. Rigollet and A.J. Stromme, On the sample complexity of entropic optimal transport. arXiv preprint arXiv:2206.13472 (2022).
[34] A. Gonzalez-Sanz, J.-M. Loubes and J. Niles-Weed, Weak limits of entropy regularized optimal transport; potentials, plans and divergences. arXiv preprint arXiv:2207.07427 (2022).
[35] Ha. Brézis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, Vol. 2. Springer (2011).
[36] J. Dieudonné, Foundations of Modern Analysis. Read Books Ltd (2011).
[37] R.J. McCann, A convexity principle for interacting gases. *Adv. Mathem.* **128** (1997) 153–179.
[38] L. Ambrosio, N. Gigli and G. Savaré, Gradient Flows: In Metric Spaces and In the Space of Probability Measures. Springer Science & Business Media (2005).
[39] L. Chizat, Mean-field langevin dynamics : Exponential convergence and annealing. *Trans. Mach. Learn. Res.* (2022).
[40] A. Nitanda, D. Wu and T. Suzuki, Convex analysis of the mean field langevin dynamics, in *International Conference on Artificial Intelligence and Statistics*. PMLR (2022) 9741–9757.
[41] M. Ledoux, Logarithmic Sobolev inequalities for unbounded spin systems revisited. *Séminaire de Probabilités XXXV* (2001) 167–194.
[42] R. Holley and D. Stroock, Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.* **46** (1987) 1159–1194.
[43] F. Otto and C. Villani, Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *J. Funct. Anal.* **173** (2000) 361–400.