



# An Optimal Transport Approach for the Schrödinger Bridge Problem and Convergence of Sinkhorn Algorithm

Simone Di Marino<sup>1,2</sup> · Augusto Gerolin<sup>3</sup>

Received: 14 January 2020 / Revised: 8 August 2020 / Accepted: 24 September 2020 /  
Published online: 19 October 2020  
© The Author(s) 2020

## Abstract

This paper exploit the equivalence between the Schrödinger Bridge problem (Léonard in *J Funct Anal* 262:1879–1920, 2012; Nelson in *Phys Rev* 150:1079, 1966; Schrödinger in *Über die umkehrung der naturgesetze*. Verlag Akademie der wissenschaften in kommission bei Walter de Gruyter u, Company, 1931) and the entropy penalized optimal transport (Cuturi in: *Advances in neural information processing systems*, pp 2292–2300, 2013; Galichon and Salanié in: *Matching with trade-offs: revealed preferences over competing characteristics*. CEPR discussion paper no. DP7858, 2010) in order to find a different approach to the duality, in the spirit of optimal transport. This approach results in a priori estimates which are consistent in the limit when the regularization parameter goes to zero. In particular, we find a new proof of the existence of maximizing entropic-potentials and therefore, the existence of a solution of the Schrödinger system. Our method extends also when we have more than two marginals: the main new result is the proof that the Sinkhorn algorithm converges even in the continuous multi-marginal case. This provides also an alternative proof of the convergence of the Sinkhorn algorithm in two marginals.

**Keywords** Schrödinger problem · Entropic regularization of optimal transport · Kantorovich duality · Sinkhorn algorithm · Iterative proportional fitting procedure

## 1 Introduction

Let  $(X, d_X)$  and  $(Y, d_Y)$  be Polish spaces,  $c : X \times Y \rightarrow \mathbb{R}$  be a cost function,  $\rho_1 \in \mathcal{P}(X)$  and  $\rho_2 \in \mathcal{P}(Y)$  be probability measures. We consider the *Schrödinger Bridge problem*

$$\text{OT}_\varepsilon(\rho_1, \rho_2) = \inf_{\gamma \in \Pi(\rho_1, \rho_2)} \varepsilon \text{KL}(\gamma | \mathcal{K}), \quad (1.1)$$

---

✉ Augusto Gerolin  
augustogerolin@gmail.com

Simone Di Marino  
simone.dimarino@unige.it

<sup>1</sup> INdAM, Unità SNS, Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, PI, Italy

<sup>2</sup> DIMA, Università degli Studi di Genova, Genoa, Italy

<sup>3</sup> Department of Theoretical Chemistry, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

where  $\mathcal{K}$  is the so-called *Gibbs Kernel* associated with the cost  $c$ :

$$\mathcal{K} = k(x, y)\rho_1 \otimes \rho_2 = e^{-\frac{c(x,y)}{\varepsilon}} \rho_1 \otimes \rho_2. \tag{1.2}$$

The function  $\text{KL}(\gamma|\mathcal{K})$  in (1.1) is the Kullback–Leibler divergence between the probability measures  $\gamma$  and  $\mathcal{K} \in \mathcal{P}(X \times Y)$  which is defined as

$$\text{KL}(\gamma|\mathcal{K}) = \begin{cases} \int_{X \times Y} \gamma \log \left( \frac{\gamma}{\mathcal{K}} \right) d(\rho_1 \otimes \rho_2) & \text{if } \gamma \ll \rho_1 \otimes \rho_2 \\ +\infty & \text{otherwise} \end{cases}.$$

Here, by abuse of notation we are denoting by  $\gamma$  the Radon-Nikodym derivative of  $\gamma$  with respect to the product measure  $\rho_1 \otimes \rho_2$ . Geometrically speaking, when we interpret the Kullback–Leibler divergence as a distance, the problem (1.1) defines the so called *Kullback–Leibler projection* of  $\mathcal{K}$  on the set  $\Pi(\rho_1, \rho_2)$ .

In the past years, theoretical and numerical aspects of (1.1) has been object of study in mathematical physics (e.g. [13–15,24,33,59–61,66,71–73]), probability (e.g. [19,47,52,53]), fluid mechanics (e.g. [3,6]), metric geometry (e.g. [41,45]), optimal transport theory (e.g. [5,16,18,20,21,32,44,51,54,55,58]), data sciences (e.g. [34,39,40,56,57,62,63] see also the book [28] and references therein).

The existence of a minimizer in (1.1) was obtain in different generality by Czisar, Ruschendorf, Borwein, Lewis and Nussbaum, Léonard, Gigli and Tamanini et al. [8,25,45,65]. In the most general case the kernel  $\mathcal{K}$  is not even assumed to be absolutely continuous with respect to  $\rho_1 \otimes \rho_2$ , as opposed to our assumption (1.2). In particular, under the assumption (1.2) on  $\mathcal{K}$  (see for example [53]), a unique minimizer for (1.1) exists and  $\gamma_{opt}^\varepsilon$  is the minimizer if and only if

$$\gamma_{opt}^\varepsilon = a^\varepsilon(x)b^\varepsilon(y)\mathcal{K}, \text{ where } a^\varepsilon, b^\varepsilon \text{ solve } \begin{cases} a^\varepsilon(x) \int_Y b^\varepsilon(y)k(x, y)d\rho_2(y) = 1 \\ b^\varepsilon(y) \int_X a^\varepsilon(x)k(x, y)d\rho_1(x) = 1 \end{cases}. \tag{1.3}$$

The functions  $a^\varepsilon(x)$  and  $b^\varepsilon(y)$  are called *Entropic potentials*. They are unique up to the trivial transformation  $a \mapsto a/\lambda, b \mapsto \lambda b$  for some  $\lambda > 0$ . The system solved by the Entropic potentials is called the *Schrödinger system*. Assuming  $\rho_1$  and  $\rho_2$  are everywhere positive and have finite entropy with respect to  $\mathcal{K}$ , the minimizer in (1.1) has a special form as is stated in the Theorem below [8, Corollary 3.9].

**Theorem 1.1** (J. M. Borwein, A. S. Lewis, and R. D. Nussbaum) *Let  $(X, d_X)$  and  $(Y, d_Y)$  be a Polish spaces,  $c : X \times Y \rightarrow [0, \infty[$  be a bounded cost function,  $\rho_1 \in \mathcal{P}(X)$ ,  $\rho_2 \in \mathcal{P}(Y)$  be probability measures such that  $\rho_1(x), \rho_2(y) > 0, \forall x \in X, y \in Y$  and  $\mathcal{K} = e^{-\frac{c(x,y)}{\varepsilon}} \rho_1 \otimes \rho_2$ . Then, if and  $\text{KL}(\rho_1 \otimes \rho_2|\mathcal{K}) < +\infty$ , then for each  $\varepsilon > 0$ , there exists a unique minimizer  $\gamma^\varepsilon \in \Pi(\rho_1, \rho_2)$  for the Schrödinger problem  $\text{OT}_\varepsilon(\rho_1, \rho_2)$  that can be written as*

$$\gamma_{opt}^\varepsilon(x, y) = a^\varepsilon(x)b^\varepsilon(y)\mathcal{K}(x, y), \quad \ln a^\varepsilon(x) \in L^1_{\rho_1}(X), \ln b^\varepsilon(x) \in L^1_{\rho_2}(Y).$$

In this paper, we are interested in the following questions:

- (1) *What is the regularity of the Entropic potentials  $a^\varepsilon$  and  $b^\varepsilon$ ?*
- (2) *Can we understand the structure and regularity of the minimizer in (1.1) if we consider the Schrödinger Bridge problem with  $N$  given marginals  $\rho_1, \rho_2, \dots, \rho_N$  instead of  $2$ ?*

The answers to the questions (1) and (2) relies on the Kantorovich duality formulation of (1.1) and its extension to the multi-marginal setting: we will exploit the parallel with optimal transport to give also a new (variational) proof for the existence of a solution to

the Schrödinger system. We believe that also this contribution is important since the only available proofs of that pass through abstract results of closure of “sum type” functions.

The multi-marginal Schrödinger Bridge problem, to be introduced in Sect. 4, has been recently considered in the literature from different viewpoints (e.g. [5,6,16,18,20,27,42,43]) as, for instance, the Wasserstein Barycenters, Matching problem in Economics, time-discretisation of Euler Equations and Density Functional Theory in computational chemistry.

Finally, we want to mention that G. Carlier and M. Laborde in [18] show the well-posedness (existence, uniqueness and smooth dependence with respect to the data) for the multi-marginal Schrödinger system in  $L^\infty$ —see (4.8) in Sect. 4—via a local and global inverse function theorems. This is a different approach and orthogonal result compared to the study presented in this paper; moreover their result is restricted to measures  $\rho_i$  which are absolutely continuous with respect to some reference measure, with density bounded from above and below.

### Computational Aspects and Connection with Optimal Transport Theory

In many applications, the method of choice for numerically computing (1.1) is the so-called *Iterative Proportional Fitting Procedure* (IPFP) or *Sinkhorn algorithm* [67]. The aim of the Sinkhorn algorithm is to construct the measure  $\gamma^\varepsilon$  realizing minimum in (1.1) by fixing the shape of the guess as  $\gamma_n^\varepsilon = a^n(x)b^n(y)\mathcal{K}$  (since this is the actual shape of the minimizer) and then alternatively updating either  $a^n$  or  $b^n$ , by matching one of the marginal distribution respectively to the target marginals  $\rho_1$  or  $\rho_2$ .

The IPFP sequences  $(a^n)_{n \in \mathbb{N}}$  and  $(b^n)_{n \in \mathbb{N}}$  are defined thus iteratively by<sup>1</sup>

$$\begin{aligned} a^0(x) &= 1, \\ b^0(y) &= 1, \\ b^n(y) &= \frac{1}{\int k(x, y)a^{n-1}(x)d\rho_1(x)}, \\ a^n(x) &= \frac{1}{\int k(x, y)b^n(y)d\rho_2(y)}. \end{aligned} \tag{1.4}$$

While  $a^n$  and  $b^n$  will be approximations of the solution of the Schrödinger system, the sequence of probability measures  $\gamma_n^\varepsilon = a^n(x)b^n(y)\mathcal{K}$  will approximate the minimizer  $\gamma^\varepsilon$ .

We stress that the IPFP procedure can be easily generalized in the multi-marginal setting, whose discussion will be detailed in Sect. 4.

#### (3) Can one prove convergence of the Sinkhorn algorithm in two and several marginals case?

In the two marginals case, the IPFP schemes was introduced by Sinkhorn [67]. The convergence of the iterates (1.4) was proven by Franklin and Lorenz [36] in the discrete case and by Ruschendorf [65] in the continuous one. An alternative proof in the continuous setting, based on the Franklin-Lorenz approach via the Hilbert metric, was also provided by Chen et al. Pavon [21], which in particular leads to a linear convergence rate of the procedure (in the Hilbert metric).

Despite the different approaches and theoretical guarantees obtained in the 2-marginal problem, in the multi-marginal case for continuous measures the situation changes completely. Theoretical results guaranteeing convergence and stability were unknown (even if in [65] it is claimed that with his methods the result can be extended to the multi-marginal case, but to our knowledge this has not been done yet). In the discrete case, the convergence has

<sup>1</sup> The iterations above system also appeared in [4,29,48,50,70] with different names (eg. RAS, IPFP).

been established by viewing it as a special case of iterative Bergman projection algorithm, whose global convergence is guaranteed (e.g. [23,49]).

One of the contributions of this paper is to extend the alternate maximization procedure of the multi-marginal Sinkhorn algorithm in the continuous setting. Our approach is an extension of dual block coordinate ascent in the continuous setting and is based on intuitions coming from Optimal Transport Theory. In particular, we exploit the regularity of Entropic potentials to prove by compactness the convergence of IPFP scheme (1.4).

*Connection with Optimal Transport Theory:* the problem (1.1) allow us to create very efficient numerical scheme approximating solutions to the Monge–Kantorovich formulation of optimal transport and its many generalizations. Indeed, notice that we can rewrite (1.1) as a functional given by the Monge–Kantorovich formulation of Optimal Transport with a cost function  $c$  plus an Entropic regularization parameter

$$\begin{aligned} \text{OT}_\varepsilon(\rho_1, \rho_2) &= \min_{\gamma \in \Pi(\rho_1, \rho_2)} \varepsilon \int_{X \times Y} \gamma \log \left( \frac{\gamma}{k} \right) d(\rho_1 \otimes \rho_2) \\ &= \min_{\gamma \in \Pi(\rho_1, \rho_2)} \int_{X \times Y} c d\gamma + \varepsilon \int_{X \times Y} \gamma \log \gamma d(\rho_1 \otimes \rho_2). \end{aligned} \tag{1.5}$$

In particular, one can show that [16,52,58] if  $(\gamma^\varepsilon)_{\varepsilon \geq 0}$  is a sequence of minimizers of the above problem, then  $\gamma^\varepsilon$  converges when  $\varepsilon \rightarrow 0$  to a solution of the Optimal Transport ( $\varepsilon = 0$ ), as depicted in Fig. 1. More precisely, let us define the functionals  $C_k, C_0 : \mathcal{P}(X \times Y) \rightarrow \mathbb{R} \cup \{+\infty\}$

$$\begin{aligned} C_k(\gamma) &= \begin{cases} \int_{X \times Y} c d\gamma + \varepsilon_k \int_{X \times Y} \rho_\gamma \log \rho_\gamma d(\rho_1 \otimes \rho_2) & \text{if } \gamma \in \Pi(\rho_0, \rho_1) \\ +\infty & \text{otherwise} \end{cases}, \\ C_0(\gamma) &= \begin{cases} \int_{X \times Y} c d\gamma & \text{if } \gamma \in \Pi(\rho_0, \rho_1) \\ +\infty & \text{otherwise} \end{cases}. \end{aligned}$$

Then in [16,53,58] it is shows that the sequence of functionals  $(C_k)_{k \in \mathbb{N}}$   $\Gamma$ -converges to  $C_0$  with respect to the weak\* topology. In particular the minima and minimal values are converging and so, in particular if  $c(x, y) = d(x, y)^p$ , then

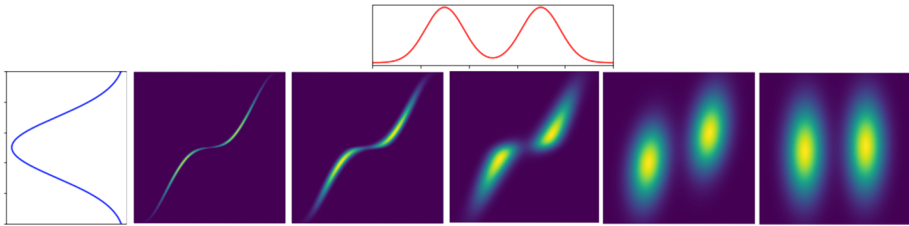
$$\lim_{k \rightarrow +\infty} \text{OT}_{\varepsilon_k}^p(\rho_1, \rho_2) = W_p^p(\rho_1, \rho_2),$$

where  $W_p(\rho_1, \rho_2)$  is the  $p$ -Wasserstein distance between  $\rho_1$  and  $\rho_2$ ,

$$W_p^p(\rho_1, \rho_2) = \min_{\gamma \in \Pi(\rho_1, \rho_2)} \int_{X \times Y} d^p(x, y) d\gamma(x, y).$$

In the context of Optimal Transport Theory, the entropic regularization was introduced by Galichon and Salanié [37] to solve matching problems in economics; and by Cuturi [26] in the context of machine learning and data sciences. Both seminal papers received renewed attention in understanding the theoretical aspects of (1.5) as well as had a strong impact in imagining, data sciences and machine learning communities due to the efficiency of the Sinkhorn algorithm.

Sinkhorn algorithm provides an efficient and scalable approximation to optimal transport. In particular, by an appropriate choice of parameters, the Sinkhorn algorithm is in fact a near-linear time approximation for computing OT distances between discrete measures [2]. However, as explained in [31,69], the Wasserstein distance suffer from the so-called *curse of dimensionality*. We refer to the recent book [28] written by Cuturi and Peyré for a complete presentation and references on computational optimal transport.



**Fig. 1** Support of the optimal coupling  $\gamma^\epsilon$  in (1.1) for the one-dimensional distance square cost with different values of  $\epsilon = 10^{-1}, 1, 10, 10^2, 10^3$ : the densities  $\rho_1 \sim N(0, 5)$  (blue) and  $\rho_2 = \frac{1}{2}\eta_1 + \frac{1}{2}\eta_2$  is a mixed Gaussian (red), where  $\eta_1 \sim N(-2, 0.8)$  and  $\eta_2 \sim N(2, 0.7)$ . The numerical computations are done using the POT library [35]

### 1.1 Main Contributions

In order to study the regularity of Entropic-potentials, we introduce the dual (Kantorovich) functional

$$D_\epsilon(u, v) = \int_X u(x)d\rho_1(x) + \int_Y v(y)d\rho_2(y) - \epsilon \int_{X \times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\epsilon}} d(\rho_1 \otimes \rho_2).$$

The Kantorovich duality of (1.1) is given by the following variational problem (see Proposition 2.12)

$$OT_\epsilon(\rho_1, \rho_2) = \sup_{u \in C_b(X), v \in C_b(Y)} D_\epsilon(u, v) + \epsilon. \tag{1.6}$$

The Entropy-Kantorovich duality (1.6) appeared, for instance, in [18,34,43–45,53]. The firsts contributions of this paper are (i) prove the existence of maximizers  $u^*$  and  $v^*$  (up to translation) in (1.6) in natural spaces; (ii) show that the Entropy-Kantorovich potentials inherit the same regularity of the cost function (see the precise statement in Proposition 2.4 and Theorem 2.8).

We then link  $u^*$  and  $v^*$  to the solution of the Schrödinger problem; as a byproduct of our results we are able to provide an alternative proof of the convergence of the Sinkhorn algorithm in the 2-marginal case via a purely optimal transportation approach (Theorem 3.1), seeing it as an alternate maximization procedure. The strength of this proof is that it can be easily generalized to the multi-marginal setting for continuous measures (Theorem 4.7).

### 1.2 Summary of Results and Main Ideas of the Proofs

Our approach follows ideas from Optimal Transport and relies on the study of the duality (Kantorovich) problem (1.6) of (1.1). Analogously to the optimal transport case, if one assume some regularity (boundedness, uniform continuity, concavity) of the cost function  $c$ , then we can obtain the same type of regularity of the Entropy potentials  $u$  and  $v$ .

The relation between solution of the dual problem (1.6) and the Entropic-Potentials solving the Schrödinger system was already pointed out by C. Léonard [53]. From our knowledge, the direct proof of existence of maximizers in (1.6) a new result.

Our approach to obtain the existence of Entropic-Kantorovich potentials, follow the direct method of Calculus of Variations. The key idea in the argument is to define a generalized notion of  $c$ -transform in the Schrödinger Bridge case, namely the  $(c, \epsilon)$ -transform. The main

duality result, in the most general case where we assume only that  $c$  is bounded, is given by the Theorem 2.8 and stated below.

**Theorem 1.2** *Let  $(X, d_X), (Y, d_Y)$  be Polish spaces,  $c : X \times Y \rightarrow \mathbb{R}$  be a Borel bounded cost,  $\rho_1 \in \mathcal{P}(X), \rho_2 \in \mathcal{P}(Y)$  be probability measures and  $\varepsilon > 0$  be a positive number. Then the supremum in (1.6) is attained for a unique couple  $(u_0, v_0)$  (up to the trivial transformation  $(u, v) \mapsto (u + a, v - a)$ ). Moreover we also have*

$$u_0 \in L^\infty(\rho_1) \quad \text{and} \quad v_0 \in L^\infty(\rho_2)$$

and we can choose the maximizers such that  $\|u_0\|_\infty, \|v_0\|_\infty \leq \frac{3}{2}\|c\|_\infty$ .

On the  $(c, \varepsilon)$ -transform: Given a measurable function  $u : X \rightarrow \mathbb{R}$  such that  $\int_X e^{u/\varepsilon} d\mu < +\infty$ , we defined the  $(c, \varepsilon)$ -transform of  $u$  by

$$u^{(c,\varepsilon)}(y) = -\varepsilon \log \left( \int_X e^{(u(x)-c(x,y))/\varepsilon} d\mu(x) \right).$$

One can show that this operation is well defined and, moreover,  $D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v), \forall u, v$  and  $D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v)$  if and only if  $v = u^{(c,\varepsilon)}$  (Lemma 2.6). If we assume additionally regularity for the cost function  $c$ , for instance that  $c$  is  $\omega$ -continuous or that it is merely bounded, the  $(c, \varepsilon)$ -transform is actually a compact operator, respectively from  $L^1(\rho_1)$  to  $C(Y)$  and from  $L^\infty(\rho_1)$  to  $L^p(\rho_2)$  (Proposition 2.4).

*IPFP/Sinkhorn algorithm:* As a byproduct of the above approach to the duality, we present an alternative proof of the convergence of the IPFP/Sinkhorn algorithm. The main idea in our proof is that we can rewrite the IPFP iteration substituting  $a^n = \exp(u^n/\varepsilon)$  and  $b^n = \exp(v^n/\varepsilon)$ ; in these new variables the iteration becomes

$$\begin{aligned} v^n(y)/\varepsilon &= -\log \left( \int_X k(x, y) \otimes \frac{u^{n-1}(x)}{\varepsilon} d\rho_1 \right) \\ u^n(x)/\varepsilon &= -\log \left( \int_Y k(x, y) \otimes \frac{v^n(x)}{\varepsilon} d\rho_2 \right). \end{aligned}$$

Or,  $v^n(y) = (u^{(n-1)})^{(c,\varepsilon)}$  and  $u^n(y) = (v^n)^{(c,\varepsilon)}$ . In particular we can interpret the IPFP in optimal transportation terms: at each step the Sinkhorn iterations (1.4) are equivalent to take the  $(c, \varepsilon)$ -transforms alternatively and therefore the IPFP can be seen as an alternating maximizing procedure for the dual problem (in view also of Lemma 2.6). Therefore, using the aforementioned compactness, it is easy to show that  $u^n$  and  $v^n$  converge to the optimal solution of the Kantorovich dual problem when  $n \rightarrow +\infty$ . A similar approach has been used also in the discrete case in [23,49], where however global convergence is guaranteed thanks to standard results in optimization.

**Theorem 1.3** *Let  $(X, d_X)$  and  $(Y, d_Y)$  be Polish metric spaces,  $\rho_1 \in \mathcal{P}(X)$  and  $\rho_2 \in \mathcal{P}(Y)$  be probability measures and  $c : X \times Y \rightarrow \mathbb{R}$  be a Borel bounded cost. If  $(a^n)_{n \in \mathbb{N}}$  and  $(b^n)_{n \in \mathbb{N}}$  are the IPFP sequences defined in (1.4), then there exists  $\lambda_n > 0$  such that*

$$a^n/\lambda_n \rightarrow a \text{ in } L^p(\rho_1) \quad \text{and} \quad \lambda_n b^n \rightarrow b \text{ in } L^p(\rho_2), \quad 1 \leq p < +\infty,$$

for  $a, b$  that solve the Schrödinger system. In particular, the sequence  $\gamma^n = a^n b^n \mathcal{K}$  converges in  $L^p(\rho_1 \otimes \rho_2)$  to  $\gamma_{opt}^\varepsilon$  in (1.1),  $1 \leq p < +\infty$ .

We recall that the argument in original proof of convergence of the Sinkhorn algorithm for the continuous case [36] (also in [21]) relies on defining the Hilbert metric on the projection cone of the Sinkhorn iterations. The authors show that the Sinkhorn iterates are a contraction

under this metric and therefore the procedure converges. This proof has the advantage of providing automatically the rate of convergence of the iterates; however it is not easily extendable in the several marginals case.

Our approach instead can be extended to obtain the existence and convergence results also in the multi-marginal setting for any probability measures  $\rho_1, \rho_2, \dots, \rho_N$  (however we don't get any quantitative convergence):

**Theorem 1.4** *Let  $(X_i, d_{X_i})$  be Polish metric spaces and  $\rho_i \in \mathcal{P}(X_i)$  be probability measures, for  $i \in \{1, \dots, N\}$  and  $c : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$  be a Borel bounded cost. If  $(a_i^n)_{n \in \mathbb{N}}$  are the multi-marginal IPFP sequences that will be defined (4.9), then there exist  $\lambda_i^n > 0$  with  $\prod_i \lambda_i^n = 1$  such that*

$$a_i^n / \lambda_i^n \rightarrow a_i \text{ in } L^p(\rho_i) \text{ for all } i \in \{1, \dots, N\}, 1 \leq p < +\infty,$$

where  $(a_1, \dots, a_N)$  solve the Schrödinger system. In particular, the sequence  $\gamma_N^n = \otimes_{i=1}^N a_i^n \mathcal{K}$  converges in  $L^p(\otimes_{i=1}^N \rho_i)$ ,  $1 \leq p < +\infty$ , to the optimal coupling  $\gamma_{N,opt}^\varepsilon$  solving the multi-marginal Schrödinger Bridge problem to be defined in (4.2).

### 1.3 Organization of the Paper

The remaining part of the paper is organized as follows: Sect. 2 contains the main structural results of the paper, namely Proposition 2.4 and Theorem 2.8. In particular, we define the main tools for showing the existence of maximizer of the Entropic-Kantorovich problem and prove regularity results of the Entropic-Kantorovich potentials via the  $(c, \varepsilon)$ -transform.

In the Sect. 3, we apply the above results to prove convergence of the Sinkhorn algorithm purely via the compactness argument and alternating maximizing procedure (Theorem 3.1) and, in Sect. 4, we extend the main results of the paper to the multi-marginal Schrödinger Bridge problem, including convergence of Sinkhorn algorithm in the multi-marginal case (Theorem 4.7).

### 1.4 The Role of the Reference Measures

In this subsection, we simply give a technical remark, discussing the role of the reference measures  $m_1$  and  $m_2$ . We stress that all the results of the paper can be extended while considering a kind of entropic optimal transport problem, where the penalization occurs with respect to some reference measures  $m_1, m_2$ .

For  $\varepsilon > 0$ , we in particular may look at the problem

$$\begin{aligned} \mathcal{S}_\varepsilon(\rho_1, \rho_2; m_1, m_2) &:= \min_{\gamma \in \Pi(\rho_1, \rho_2)} \left\{ \int_{X \times Y} c \, d\gamma + \varepsilon \text{KL}(\gamma | m_1 \otimes m_2) \right\} \\ &= \min_{\gamma \in \Pi(\rho_1, \rho_2)} \varepsilon \text{KL}(\gamma | \mathcal{K}). \end{aligned} \tag{1.7}$$

where  $\mathcal{K}$  is the Gibbs Kernel  $\mathcal{K} = e^{-\frac{c}{\varepsilon}} m_1 \otimes m_2$ .

While having a reference measure in some situations can be quite useful (for example the Schrödinger problem is set with  $m_1 = m_2 = \mathcal{L}^d$ ), in other it is the opposite, for example when we are considering  $\rho_1, \rho_2$  to be sums of diracs. In those cases it is a much better solution to consider  $m_1 = \rho_1$  and  $m_2 = \rho_2$ . Notice that in this case, we have that

$$\text{OT}_\varepsilon(\rho_1, \rho_2) = \mathcal{S}_\varepsilon(\rho_1, \rho_2; \rho_1, \rho_2).$$

Now we will see that in fact  $OT_\varepsilon$  is a *universal* reduction for  $S_\varepsilon$ , meaning that we can always assume  $m_1 = \rho_1$  and  $m_2 = \rho_2$ :

**Lemma 1.5** *Let  $(X, d, m_1)$  and  $(Y, d, m_1)$  be a Polish metric measure spaces and  $c : X \times Y \rightarrow [0, +\infty[$  be a cost function. Assume that  $\rho_1 \in \mathcal{P}(X)$ ,  $\rho_2 \in \mathcal{P}(Y)$ . Then we have*

$$S_\varepsilon(\rho_1, \rho_2; m_1, m_2) = OT_\varepsilon(\rho_1, \rho_2) + \varepsilon KL(\rho_1|m_1) + \varepsilon KL(\rho_2|m_2);$$

moreover, whenever one of the two sides is finite the minimizers of  $S_\varepsilon$  and  $OT_\varepsilon$  are the same.

**Proof** The key equality in proving this lemma is that, whenever  $\gamma \in \Gamma(\rho_1, \rho_2)$  one has

$$KL(\gamma|m_1 \otimes m_2) = KL(\gamma|\rho_1 \otimes \rho_2) + KL(\rho_1|m_1) + KL(\rho_2|m_2). \tag{1.8}$$

While this equality is clear whenever all the terms are finite, we refer to Lemma 1.6 below for a complete proof entailing every case. From this equality we can easily get the conclusions.  $\square$

**Lemma 1.6** *Let  $(X, \sigma_X)$  and  $(Y, \sigma_Y)$  be measurable spaces. Assume that  $\gamma \in \mathcal{P}(X \times Y)$ ,  $m_1 \in \mathcal{P}(X)$  and  $m_2 \in \mathcal{P}(Y)$ . Then we have*

$$KL(\gamma|m_1 \otimes m_2) = KL(\gamma|\rho_1 \otimes \rho_2) + KL(\rho_1|m_1) + KL(\rho_2|m_2), \tag{1.9}$$

where  $\rho_1 = (e_1)_\# \gamma$  and  $\rho_2 = (e_2)_\# \gamma$  are the projections of  $\gamma$  onto  $X$  and  $Y$  respectively.

**Proof** First we assume the right hand side of (1.9) is finite, and in particular this implies  $\gamma \ll \rho_1 \otimes \rho_2$ ,  $\rho_1 \ll m_1$  and  $\rho_2 \ll m_2$ . In particular we get  $\gamma \ll m_1 \otimes m_2$  and we can infer

$$\frac{d\gamma}{d(m_1 \otimes m_2)}(x, y) = \frac{d\gamma}{d(\rho_1 \otimes \rho_2)}(x, y) \cdot \frac{d\rho_1}{dm_1}(x) \cdot \frac{d\rho_2}{dm_2}(y).$$

We can now compute

$$\begin{aligned} KL(\gamma|m_1 \otimes m_2) &= \int_{X \times Y} \ln \left( \frac{d\gamma}{d(m_1 \otimes m_2)}(x, y) \right) d\gamma \\ &= \int_{X \times Y} \ln \left( \frac{d\gamma}{d(\rho_1 \otimes \rho_2)}(x, y) \right) d\gamma + \int_{X \times Y} \ln \left( \frac{d\rho_1}{dm_1}(x) \right) d\gamma \\ &\quad + \int_{X \times Y} \ln \left( \frac{d\rho_2}{dm_2}(y) \right) d\gamma \\ &= KL(\gamma|\rho_1 \otimes \rho_2) + \int_X \ln \left( \frac{d\rho_1}{dm_1}(x) \right) d\rho_1 + \int_Y \ln \left( \frac{d\rho_2}{dm_2}(y) \right) d\rho_2 \\ &= KL(\gamma|\rho_1 \otimes \rho_2) + KL(\rho_1|m_1) + KL(\rho_2|m_2). \end{aligned}$$

We assume now that the left hand side of (1.9) is finite. Thanks to the fact that  $KL(F_\# \mu | F_\# \nu) \leq KL(\mu | \nu)$  for every measurable function  $F$ , we immediately have that  $KL(\rho_1|m_1)$  and  $KL(\rho_2|m_2)$  are finite, and in particular  $\rho_1 \ll m_1$  and  $\rho_2 \ll m_2$ . Now let us introduce  $f = \frac{d\gamma}{d(m_1 \otimes m_2)}$ ,  $g_1 = \frac{d\rho_1}{dm_1}$  and  $g_2 = \frac{d\rho_2}{dm_2}$ ; let us then consider any measurable set  $A \subseteq X \times Y$  and assume that  $(\rho_1 \otimes \rho_2)(A) = 0$ . In particular we have

$$\int_{X \times Y} \chi_A(x, y) g_1(x) g_2(y) d(m_1 \otimes m_2) = (\rho_1 \otimes \rho_2)(A) = 0;$$

from this we deduce that  $A$  is  $m_1 \otimes m_2$ -essentially contained in the set  $B = \{g_1(x)g_2(y) = 0\} = B_x \cup B_y$ , where  $B_x = \{g_1(x) = 0\} \times Y$  and  $B_y = X \times \{g_2(y) = 0\}$ . However, by the

marginal conditions, we have  $\gamma(B_x) = \rho_1\{g_1(x) = 0\} = 0$  and similarly  $\gamma(B_y) = 0$ , which imply  $\gamma(B) = 0$ . In particular we have

$$\begin{aligned} \gamma(A) &= \int_{X \times Y} \chi_A(x, y) f(x, y) d(\mathbf{m}_1 \otimes \mathbf{m}_2) = \int_{X \times Y} \chi_{A \cap B}(x, y) f(x, y) d(\mathbf{m}_1 \otimes \mathbf{m}_2) \\ &\leq \int_{X \times Y} \chi_B(x, y) f(x, y) d(\mathbf{m}_1 \otimes \mathbf{m}_2) \leq \gamma(B) = 0. \end{aligned}$$

This proves that  $\gamma \ll \rho_1 \otimes \rho_2$  and so we can perform the same calculation we did before to conclude. □

## 2 Regularity of Entropic-Potentials and Dual Problem

In this section we will treat the case where  $c : X \times Y \rightarrow \mathbb{R}$  is a Borel *bounded* cost; of course everything extends also to the case when  $c \in L^\infty(\rho_1 \otimes \rho_2)$ . Some of the results extend naturally for unbounded costs (for example (i), (ii), (v) in Proposition 2.4), but we prefer to keep the setting uniform.

### 2.1 Entropy-Transform and a Priori Estimates

We start by defining the Entropy-Transform. First, let us define the space  $L_\varepsilon^{\text{exp}}$ , which will be the natural space for the dual problem.

**Definition 2.1** ( *$L_\varepsilon^{\text{exp}}$  spaces*) Let  $\varepsilon > 0$  be a positive number and  $(X, d_X)$  be a Polish space. We define the set  $L_\varepsilon^{\text{exp}}(X, \rho_1)$  by

$$\begin{aligned} L_\varepsilon^{\text{exp}}(X, \rho_1) &= \left\{ u : X \rightarrow [-\infty, \infty] : u \text{ is a measurable function in } (X, \rho_1) \text{ and } 0 < \int_X e^{u/\varepsilon} d\rho_1 < \infty \right\}. \end{aligned}$$

For  $u \in L_\varepsilon^{\text{exp}}(X, \rho_1)$  we define also  $\lambda_u := \varepsilon \log \left( \int_X e^{u/\varepsilon} d\rho_1 \right)$ .

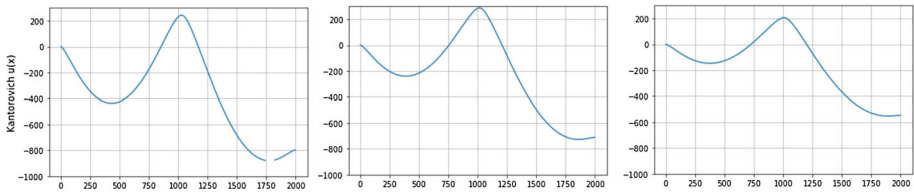
For simplicity, we will use the notation  $L_\varepsilon^{\text{exp}}(\rho_1)$  instead of  $L_\varepsilon^{\text{exp}}(X, \rho_1)$ . Notice that it is possible that  $u \in L_\varepsilon^{\text{exp}}(X, \rho_1)$  attains the value  $-\infty$  in a set of positive measure, but not everywhere, because of the positivity constraint  $\int_X e^{u/\varepsilon} d\rho_1 > 0$ . On the other hand, we have that  $u \in L_\varepsilon^{\text{exp}}(X, \rho_1)$  implies  $u^+ \in L^p(\rho_1)$  for every  $p \geq 1$ , where  $u^+(x) := \max\{u(x), 0\}$  denotes the positive part of  $u$ .

**Definition 2.2** (*Entropic  $c$ -transform or  $(c, \varepsilon)$ -transform*) Let  $(X, d_X), (Y, d_Y)$  be Polish spaces,  $\varepsilon > 0$  be a positive number,  $\rho_1 \in \mathcal{P}(X)$  and  $\rho_2 \in \mathcal{P}(Y)$  be probability measures and let  $c$  be a bounded measurable cost on  $X \times Y$ . The entropic  $(c, \varepsilon)$ -transform  $\mathcal{F}^{(c, \varepsilon)} : L_\varepsilon^{\text{exp}}(\rho_1) \rightarrow L^0(\rho_2)$  is defined by

$$\mathcal{F}^{(c, \varepsilon)}(u)(y) := -\varepsilon \log \left( \int_X e^{\frac{u(x) - c(x, y)}{\varepsilon}} d\rho_1(x) \right). \tag{2.1}$$

Analogously, we define the  $(c, \varepsilon)$ -transform  $\mathcal{F}^{(c, \varepsilon)} : L_\varepsilon^{\text{exp}}(\rho_2) \rightarrow L^0(\rho_1)$  by

$$\mathcal{F}^{(c, \varepsilon)}(v)(x) := -\varepsilon \log \left( \int_Y e^{\frac{v(y) - c(x, y)}{\varepsilon}} d\rho_2(y) \right). \tag{2.2}$$



**Fig. 2** Entropy-Kantorovich potentials  $u^\varepsilon(x) - \varepsilon \ln(\rho_1)$  associated to the densities  $\rho_1$  and  $\rho_2$  for different values of the regularization parameter:  $\varepsilon_1 < \varepsilon_2 < \varepsilon_3$  (from left to right). The densities  $\rho_1 \sim N(0, 5)$  and  $\rho_2 = \frac{1}{2}\eta_1 + \frac{1}{2}\eta_2$  is a mixed Gaussian, where  $\eta_1 \sim N(7, 0.9)$  and  $\eta_2 \sim N(14, 0.9)$

Whenever it will be clear we denote  $v^{(c,\varepsilon)} = \mathcal{F}^{(c,\varepsilon)}(v)$  and  $u^{(c,\varepsilon)} = \mathcal{F}^{(c,\varepsilon)}(u)$ , in an analogous way to the classical  $c$ -transform.

Notice that  $L_\varepsilon^{\text{exp}}(\rho_1)$  is the natural domain of definition for  $\mathcal{F}^{(c,\varepsilon)}$  because if  $u \notin L_\varepsilon^{\text{exp}}(\rho_1)$  we would have either  $\mathcal{F}^{(c,\varepsilon)}(u) \equiv -\infty$  or  $\mathcal{F}^{(c,\varepsilon)}(u) \equiv +\infty$ ; moreover, thanks to the positivity constraint  $\int_X e^{u/\varepsilon} d\rho_1 > 0$  we also have  $\mathcal{F}^{(c,\varepsilon)}(u)(y) \in \mathbb{R}$  almost everywhere. In fact we will show that  $\mathcal{F}^{(c,\varepsilon)}(u) \in L^\infty(\rho_2)$ .

We also remark that the  $(c, \varepsilon)$ -transform is consistent with the  $c$ -transform when  $\varepsilon \rightarrow 0$ :  $u^{(c,\varepsilon)} \rightarrow \max\{u(x) - c(x, y) : x \in X\}$ , when  $\varepsilon \rightarrow 0$ . In other words,  $u^{(c,\varepsilon)}(y) = u^c(y) + O(\varepsilon)$  (Fig. 2).

**Lemma 2.3** *Let  $(X, d_X), (Y, d_Y)$  be Polish spaces,  $u \in L_\varepsilon^{\text{exp}}(\rho_1), v \in L_\varepsilon^{\text{exp}}(\rho_2)$  and  $\varepsilon > 0$ . Then,*

(i)  $u^{(c,\varepsilon)}(y) \in L^\infty(\rho_2)$  and  $v^{(c,\varepsilon)}(x) \in L^\infty(\rho_1)$ . More precisely,

$$-\|c\|_\infty - \varepsilon \log \left( \int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1 \right) \leq u^{(c,\varepsilon)}(y) \leq \|c\|_\infty - \varepsilon \log \left( \int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1 \right)$$

(ii)  $u^{(c,\varepsilon)}(y) \in L_\varepsilon^{\text{exp}}(\rho_2)$  and  $v^{(c,\varepsilon)}(x) \in L_\varepsilon^{\text{exp}}(\rho_1)$ . Moreover  $|\lambda_{u^{(c,\varepsilon)}} + \lambda_u| \leq \|c\|_\infty$ .

**Proof** If  $u \in L_\varepsilon^{\text{exp}}(\rho_1)$  then

$$\begin{aligned} u^{(c,\varepsilon)}(y) &= -\varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1 \right) \\ &\leq -\varepsilon \log \left( e^{-\frac{\|c\|_\infty}{\varepsilon}} \int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1 \right) \\ &= \|c\|_\infty - \varepsilon \log \left( \int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1 \right). \end{aligned}$$

Moreover, we get a lower bound for the above quantity using  $c \geq -\|c\|_\infty$ :

$$u^{(c,\varepsilon)}(y) = -\varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1 \right) \geq -\|c\|_\infty - \varepsilon \log \left( \int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1 \right).$$

Then we proved that  $u^{(c,\varepsilon)} \in L^\infty(\rho_2)$ . The fact that  $v^{(c,\varepsilon)} \in L^\infty(\rho_1)$  is analogous. This shows the (i). Since  $u \in L_\varepsilon^{\text{exp}}(\rho_1)$ , by using the part (i) we have that

$$\int_Y e^{\frac{u^{(c,\varepsilon)}(y)}{\varepsilon}} d\rho_2(y) \leq \int_Y e^{\|c\|_\infty/\varepsilon} \left( \int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1(x) \right)^{-1} d\rho_2(y) < +\infty.$$

Therefore  $u^{(c,\varepsilon)} \in L_\varepsilon^{\text{exp}}(\rho_2)$  and in particular  $\lambda_{u^{(c,\varepsilon)}} \leq -\lambda_u + \|c\|_\infty$ ; the other inequality follows with a similar calculation and the same holds for  $v^{(c,\varepsilon)}$ , which proves (ii).  $\square$

Some of the following properties were already known for the softmax operator (for example in [39]) and they are used in order to get *a posteriori* regularity of the potentials but, up to our knowledge, were never used to get *a priori* results. Another very cleverly used properties of the  $(c, \varepsilon)$ -transform are in [32] in order to obtain a new proof of the Caffarelli’s contraction theorem [12].

**Proposition 2.4** *Let  $\varepsilon > 0$  be a positive number,  $(X, d_X)$  and  $(Y, d_Y)$  be Polish metric spaces,  $c : X \times Y \rightarrow [0, \infty]$  be a bounded cost function,  $\rho_1 \in \mathcal{P}(X)$ ,  $\rho_2 \in \mathcal{P}(Y)$  be probability measures and  $u \in L^\varepsilon_{\text{exp}}(\rho_1)$ . Then*

- (i) *if  $c$  is  $L$ -Lipschitz, then  $u^{(c,\varepsilon)}$  is  $L$ -Lipschitz;*
- (ii) *if  $c$  is  $\omega$ -continuous, then  $u^{(c,\varepsilon)}$  is  $\omega$ -continuous;*
- (iii) *if  $|c| \leq M$ , then we have  $|u^{(c,\varepsilon)} + \lambda_u| \leq M$ ;*
- (iv) *if  $|c| \leq M$ , then  $\mathcal{F}^{(c,\varepsilon)} : L^\infty(\rho_1) \rightarrow L^p(\rho_2)$  is a 1-Lipschitz compact operator.*
- (v) *if  $c$  is  $K$ -concave with respect to  $y$ , then  $u^{(c,\varepsilon)}$  is  $K$ -concave.*

**Proof** Of course we have that (ii) implies (i); let us prove directly (ii).

- (ii) Let  $u \in L^\varepsilon_{\text{exp}}(\rho_1)$ ,  $y_1, y_2 \in Y$ . We can assume without loss of generality that  $u^{(c,\varepsilon)}(y_1) \geq u^{(c,\varepsilon)}(y_2)$ ; in that case

$$\begin{aligned} & |u^{(c,\varepsilon)}(y_1) - u^{(c,\varepsilon)}(y_2)| \\ &= \varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y_2)}{\varepsilon}} d\rho_1 \right) - \varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}} d\rho_1 \right) \\ &= \varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y_1)+c(x,y_1)-c(x,y_2)}{\varepsilon}} d\rho_1 \right) - \varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}} d\rho_1 \right) \\ &\leq \varepsilon \log \left( e^{\frac{\omega(d(y_1,y_2))}{\varepsilon}} \int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}} d\rho_1 \right) - \varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}} d\rho_1 \right) \\ &= \omega(y_1, y_2). \end{aligned}$$

- (iii) This is a direct consequence of Lemma 2.3 (i);
- (iv) We first prove that  $\mathcal{F}^{(c,\varepsilon)}$  is 1-Lipschitz. In fact, letting  $u, \tilde{u} \in L^\infty(\rho_1)$ , we can perform a calculation very similar to what has been done in (ii): for every  $y \in Y$  we have

$$\begin{aligned} \mathcal{F}^{(c,\varepsilon)}(u)(y) &= -\varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1 \right) \geq -\varepsilon \log \left( \int_X e^{\frac{\tilde{u}(x)+\|u-\tilde{u}\|_\infty-c(x,y)}{\varepsilon}} d\rho_1 \right) \\ &= \mathcal{F}^{(c,\varepsilon)}(\tilde{u})(y) - \|u - \tilde{u}\|_\infty \end{aligned}$$

We can conclude that  $\|\mathcal{F}^{(c,\varepsilon)}(u) - \mathcal{F}^{(c,\varepsilon)}(\tilde{u})\|_p \leq \|\mathcal{F}^{(c,\varepsilon)}(u) - \mathcal{F}^{(c,\varepsilon)}(\tilde{u})\|_\infty \leq \|u - \tilde{u}\|_\infty$ . This proves in particular that  $\mathcal{F}^{(c,\varepsilon)} : L^\infty(\rho_1) \rightarrow L^p(\rho_2)$  is continuous. In order to prove that  $\mathcal{F}^{(c,\varepsilon)}$  is compact it suffices to prove that  $\mathcal{F}^{(c,\varepsilon)}(B)$  is precompact for every bounded set  $B \subset L^\infty(\rho_1)$ . We will use Proposition 5.1; since  $\mathcal{F}^{(c,\varepsilon)}$  is Lipschitz, for sure if  $B$  is bounded we have that  $\mathcal{F}^{(c,\varepsilon)}(B)$  is bounded in  $L^p(\rho_2)$ , so it remains to prove part (b) of the criterion of Proposition 5.1.

Let us consider  $\gamma = \rho_1 \otimes \rho_2$ . Since  $c \in L^\infty(\gamma)$ , by Lusin theorem we have that for every  $\sigma > 0$  there exists  $N_\sigma \subset X \times Y$ , with  $\gamma(N_\sigma) < \sigma$ , such that  $c|_{(N_\sigma)^c}$  is uniformly continuous, with modulus of continuity  $\omega_\sigma$ . We now try to mimic what we did in (ii), this time keeping also track of the remainder terms we will have.

For each  $y \in Y$ , we consider the slice of  $N_\sigma$  above  $y$ ,  $N_y^\sigma = \{x \in X : (x, y) \in N_\sigma\}$ ; then we consider the set of *bad*  $y \in Y$ , where the slice  $N_y^\sigma$  is too big:

$$N_b^\sigma = \left\{ y \in Y : \rho_1(N_y^\sigma) \geq \sqrt{\sigma} \right\}.$$

In particular, by definition if  $y \notin N_b^\sigma$  we have  $\rho_1(N_y^\sigma) \leq \sqrt{\sigma}$ , but thanks to Fubini and the condition  $\gamma(N_\sigma) < \sigma$  we have also that  $\rho_2(N_b^\sigma) \leq \sqrt{\sigma}$ .

Let us now consider  $y, y' \notin N_b^\sigma$ , and let us denote  $X^* = X \setminus (N_y^\sigma \cup N_{y'}^\sigma)$ . Then we have that for every  $x \in X^*$ ,  $|c(x, y) - c(x, y')| \leq \omega_\sigma(d(y, y'))$ . We can assume without loss of generality that  $u^{(c,\varepsilon)}(y) \geq u^{(c,\varepsilon)}(y')$  and we have

$$\begin{aligned} & |u^{(c,\varepsilon)}(y) - u^{(c,\varepsilon)}(y')| \\ &= -\varepsilon \log \left( \int_X e^{(u(x)-c(x,y))/\varepsilon} d\rho_1 \right) + \varepsilon \log \left( \int_X e^{(u(x)-c(x,y'))/\varepsilon} d\rho_1 \right) \\ &= \varepsilon \log \left( \frac{\int_X e^{(u(x)-c(x,y)+c(x,y)-c(x,y'))/\varepsilon} d\rho_1}{\int_X e^{(u(x)-c(x,y))/\varepsilon} d\rho_1} \right) \\ &\leq \varepsilon \log \left( e^{\frac{\omega_\sigma(d(y,y'))}{\varepsilon}} + \frac{\int_{N_y^\sigma \cup N_{y'}^\sigma} e^{(u(x)-c(x,y'))/\varepsilon} d\rho_1}{\int_X e^{(u(x)-c(x,y))/\varepsilon} d\rho_1} \right) \\ &\leq \varepsilon \log \left( e^{\frac{\omega_\sigma(d(y,y'))}{\varepsilon}} + \rho_1(N_y^\sigma \cup N_{y'}^\sigma) e^{2(\|c\|+\|u\|)/\varepsilon} \right) \\ &\leq \varepsilon \log \left( e^{\frac{\omega_\sigma(d(y,y'))}{\varepsilon}} + 2\sqrt{\sigma} e^{2(\|c\|+\|u\|)/\varepsilon} \right) \end{aligned}$$

Now we denote by  $A = 2e^{2(\|c\|+\|u\|)/\varepsilon}$  and thanks to the fact that if  $a, b \geq 0$  then  $e^a + b \leq e^{a+b}$ , we have

$$|u^{(c,\varepsilon)}(y) - u^{(c,\varepsilon)}(y')| \leq \omega_\sigma(d(y, y')) + \varepsilon\sqrt{\sigma}A \quad \forall y, y' \notin N_b^\sigma.$$

Then (having in mind also (iii) and that  $A$  depends only on  $\|u\|_\infty$ ), we have that also (b) of Proposition 5.1 is satisfied for  $\mathcal{F}^{(c,\varepsilon)}(B)$ , for every bounded set  $B \subset L^\infty(\rho_1)$ , granting then the compactness of  $\mathcal{F}^{(c,\varepsilon)}$ .

- (v) In this case we are assuming that  $Y$  is a geodesic space and that there exists  $K \in \mathbb{R}$  such that for each constant speed geodesic  $(y_t)_{t \in [0,1]}$  we have

$$c(x, y_t) \geq (1-t)c(x, y_0) + tc(x, y_1) + 2t(1-t)Kd^2(y_0, y_1) \quad \forall x \in X.$$

Then, setting  $f_t(x) = e^{(u(x)-c(x,y_t))/\varepsilon}$ , the  $K$ -concavity inequality for  $c$  implies

$$\begin{aligned} f_t(x) &= e^{(u(x)-c(x,y_t))/\varepsilon} \\ &\leq e^{(u(x)-(1-t)c(x,y_0)-tc(x,y_1)-2t(1-t)Kd^2(y_0,y_1))/\varepsilon} \\ &= e^{-2t(1-t)Kd^2(y_0,y_1)/\varepsilon} \cdot e^{((1-t)(u(x)-c(x,y_0))+t(u(x)-c(x,y_1)))/\varepsilon} \\ &= e^{-2t(1-t)Kd^2(y_0,y_1)/\varepsilon} \cdot f_0(x)^{1-t} \cdot f_1(x)^t \end{aligned}$$

Using this along with Hölder inequality we get

$$u^{(c,\varepsilon)}(y_t) = -\varepsilon \log \left( \int_X e^{(u(x)-c(x,y_t))/\varepsilon} d\rho_1 \right) = -\varepsilon \log \left( \int_X f_t(x) d\rho_1 \right)$$

$$\begin{aligned}
 &\geq -\varepsilon \log \left( \int_X e^{-2t(1-t)Kd^2(y_0, y_1)/\varepsilon} \cdot f_0(x)^{1-t} \cdot f_1(x)^t d\rho_1 \right) \\
 &= 2t(1-t)Kd^2(y_0, y_1) - \varepsilon \log \left( \int_X f_0(x)^{1-t} \cdot f_1(x)^t d\rho_1 \right) \\
 &\geq 2t(1-t)Kd^2(y_0, y_1) - \varepsilon \log \left( \left( \int_X f_0 d\rho_1 \right)^{1-t} \cdot \left( \int_X f_1 d\rho_1 \right)^t \right) \\
 &= 2t(1-t)Kd^2(y_0, y_1) + (1-t)u^{(c, \varepsilon)}(y_0) + tu^{(c, \varepsilon)}(y_1).
 \end{aligned}$$

□

**Remark 2.5** (Entropic  $c$ -transform for  $S_\varepsilon(\rho_1, \rho_2; m_1, m_2)$ ). It is possible to define the entropic  $c$ -transform also for the Schrödinger problem  $S_\varepsilon(\rho_1, \rho_2; m_1, m_2)$  with reference measures  $m_1 \in \mathcal{P}(X)$  and  $m_2 \in \mathcal{P}(Y)$ . In this case,

$$\mathcal{F}_*^{(c, \varepsilon)}(u)(y) = \varepsilon \log(\rho_2(y)) - \varepsilon \log \left( \int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} dm_1(x) \right), \quad \text{and} \tag{2.3}$$

$$\mathcal{F}_*^{(c, \varepsilon)}(v)(x) = \varepsilon \log(\rho_1(x)) - \varepsilon \log \left( \int_Y e^{\frac{u(x)-c(x,y)}{\varepsilon}} dm_2(y) \right). \tag{2.4}$$

It is easy to see that

$$\begin{cases} \mathcal{F}_*^{(c, \varepsilon)}(v) = \varepsilon \log \rho_1 + \mathcal{F}^{(c, \varepsilon)}(v - \varepsilon \log \rho_2) \\ \mathcal{F}_*^{(c, \varepsilon)}(u) = \varepsilon \log \rho_2 + \mathcal{F}^{(c, \varepsilon)}(u - \varepsilon \log \rho_1) \end{cases}$$

so that in fact the  $(c, \varepsilon)$ -transforms with reference measures are in fact the  $(c, \varepsilon)$ -trasforms conjugated by the addition of a function. In particular we can get exactly the same estimates we did in Lemma 2.3, up to translate in the appropriate manner. For example we would have if  $u \in L_\varepsilon^{\text{exp}}(m_1)$ , we would have then  $u_*^{(c, \varepsilon)}(y) - \varepsilon \log(\rho_2(y)) \in L^\infty(m_2)$ .

### 2.2 Dual Problem

Let  $u \in L_\varepsilon^{\text{exp}}(\rho_1)$ ,  $v \in L_\varepsilon^{\text{exp}}(\rho_2)$  and consider the Entropy-Kantorovich functional,

$$D_\varepsilon(u, v) = \int_X u(x)d\rho_1(x) + \int_Y v(y)d\rho_2(y) - \varepsilon \int_{X \times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2). \tag{2.5}$$

What are the minimal assumption on  $u, v$  in order to make sense for  $D_\varepsilon(u, v)$ ? First of all if  $u^+ \in L^1(\rho_1)$  and  $v^+ \in L^1(\rho_2)$  then  $D_\varepsilon(u, v) < \infty$  and in particular in order to have  $D_\varepsilon(u, v) > -\infty$  we need  $u \in L_\varepsilon^{\text{exp}}(\rho_1)$ ,  $v \in L_\varepsilon^{\text{exp}}(\rho_2)$  which is then a natural assumption (since we want to compute the supremum of  $D_\varepsilon$ ).

**Lemma 2.6** Let us consider  $D_\varepsilon : L_\varepsilon^{\text{exp}}(\rho_1) \times L_\varepsilon^{\text{exp}}(\rho_2) \rightarrow \mathbb{R}$  defined as in (2.5), then

$$D_\varepsilon(u, u^{(c, \varepsilon)}) \geq D_\varepsilon(u, v) \quad \forall v \in L_\varepsilon^{\text{exp}}(\rho_2), \tag{2.6}$$

$$D_\varepsilon(u, u^{(c, \varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c, \varepsilon)}. \tag{2.7}$$

In particular we can say that  $u^{(c, \varepsilon)} \in \text{argmax}\{D_\varepsilon(u, v) : v \in L_\varepsilon^{\text{exp}}(\rho_2)\}$ .

**Proof** By Fubini’s theorem and Eq. (2.1), we have

$$D_\varepsilon(u, v) = \int_X u(x)d\rho_1(x) + \int_Y v(y)d\rho_2(y) - \varepsilon \int_{X \times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2),$$

$$\begin{aligned} &= \int_X u(x)d\rho_1(x) + \int_Y v(y)d\rho_2(y) - \varepsilon \int_Y e^{\frac{v(y)}{\varepsilon}} \left( \int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1 \right) d\rho_2, \\ &= \int_X u(x)d\rho_1(x) + \int_Y v(y) - \varepsilon e^{\frac{v(y)-u^{(c,\varepsilon)}(y)}{\varepsilon}} d\rho_2(y). \end{aligned}$$

Therefore, for any  $v \in L_\varepsilon^{\text{exp}}(\rho_1)$ ,  $D_\varepsilon(u, v) \leq D_\varepsilon(u, u^{(c,\varepsilon)})$ , since the function  $g(t) = t - \varepsilon e^{(t-a)/\varepsilon}$  is strictly concave and attains its maximum in  $t = a$ . In particular,  $D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v)$  if and only if  $v = u^{(c,\varepsilon)}$ .  $\square$

**Lemma 2.7** *Let us consider  $u \in L_\varepsilon^{\text{exp}}(\rho_1)$  and  $v \in L_\varepsilon^{\text{exp}}(\rho_2)$ . Then there exist  $u^* \in L_\varepsilon^{\text{exp}}(\rho_1)$  and  $v^* \in L_\varepsilon^{\text{exp}}(\rho_2)$  such that*

- $D_\varepsilon(u, v) \leq D_\varepsilon(u^*, v^*)$ ;
- $\|v^*\|_\infty \leq 3\|c\|_\infty/2$ ;
- $\|u^*\|_\infty \leq 3\|c\|_\infty/2$ .

Moreover  $a \in \mathbb{R}$  such that  $u^* = (v + a)^{(c,\varepsilon)}$  and  $v^* = (u^*)^{(c,\varepsilon)}$ .

**Proof** Let us apply Proposition 2.4 (iii) to  $v$  and  $\tilde{u} = v^{(c,\varepsilon)}$ :

$$\begin{aligned} -\|c\|_\infty &\leq v^{(c,\varepsilon)} + \lambda_v \leq \|c\|_\infty \\ -\|c\|_\infty &\leq (v^{(c,\varepsilon)})^{(c,\varepsilon)} + \lambda_{v^{(c,\varepsilon)}} \leq \|c\|_\infty \end{aligned}$$

Let us define  $\tilde{u} = v^{(c,\varepsilon)}$  and  $\tilde{v} = (v^{(c,\varepsilon)})^{(c,\varepsilon)}$ . Then by Lemma 2.6 we have of course that  $D_\varepsilon(u, v) \leq D_\varepsilon(\tilde{u}, \tilde{v})$ ; now we know that  $D_\varepsilon(\tilde{u} - a, \tilde{v} + a) = D_\varepsilon(\tilde{u}, \tilde{v})$  for any  $a \in \mathbb{R}$  and moreover

$$\|\tilde{u} - a\|_\infty \leq \|c\|_\infty + |a + \lambda_v| \quad \|\tilde{v} + a\|_\infty \leq \|c\|_\infty + |\lambda_{v^{(c,\varepsilon)}} - a|.$$

We can now choose  $a^* = (\lambda_{v^{(c,\varepsilon)}} - \lambda_v)/2$  and, recalling Lemma 2.3 (ii) we can conclude that  $u^* = \tilde{u} - a^*$  and  $v^* = \tilde{v} + a^*$  satisfy the required bounds.  $\square$

**Theorem 2.8** *Let  $(X, d_X), (Y, d_Y)$  be Polish spaces,  $c : X \times Y \rightarrow \mathbb{R}$  be a Borel bounded cost,  $\rho_1 \in \mathcal{P}(X), \rho_2 \in \mathcal{P}(Y)$  be probability measures and  $\varepsilon > 0$  be a positive number. Consider the problem*

$$\sup \{ D_\varepsilon(u, v) : u \in L_\varepsilon^{\text{exp}}(\rho_1), v \in L_\varepsilon^{\text{exp}}(\rho_2) \}. \tag{2.8}$$

Then the supremum in (2.8) is attained for a unique couple  $(u_0, v_0)$  (up to the trivial transformation  $(u, v) \mapsto (u + a, v - a)$ ). In particular we have

$$u_0 \in L^\infty(\rho_1) \quad \text{and} \quad v_0 \in L^\infty(\rho_2);$$

moreover we can choose the maximizers such that  $\|u_0\|_\infty, \|v_0\|_\infty \leq \frac{3}{2}\|c\|_\infty$ .

**Proof** Now, we are going to show that the supremum is attained in the right-hand side of (2.8). Let  $(u_n)_{n \in \mathbb{N}} \subset L_\varepsilon^{\text{exp}}(\rho_1)$  and  $(v_n)_{n \in \mathbb{N}} \subset L_\varepsilon^{\text{exp}}(\rho_2)$  be maximizing sequences. Due to Lemma 2.7, we can suppose that  $u_n \in L^\infty(\rho_1), v_n \in L^\infty(\rho_2)$  and  $\|u_n\|_\infty, \|v_n\|_\infty \leq \frac{3}{2}\|c\|_\infty$ . Then by Banach-Alaoglu theorem there exists subsequences  $(u_{n_k})_{n_k \in \mathbb{N}}$  and  $(v_{n_k})_{n_k \in \mathbb{N}}$  such that  $u_{n_k} \rightharpoonup \bar{u}$  and  $v_{n_k} \rightharpoonup \bar{v}$ . In particular,  $\tilde{u}_{n_k} + \tilde{v}_{n_k} - c \rightarrow \bar{u} + \bar{v} - c$ .

First, notice that since  $t \mapsto e^t$  is a convex function, we have

$$\liminf_{n \rightarrow \infty} \int_{X \times Y} e^{\frac{u_n + v_n - c}{\varepsilon}} d(\rho_1 \otimes \rho_2) = \liminf_{n \rightarrow \infty} \int_{X \times Y} e^{\frac{u_n + v_n - c}{\varepsilon}} d(\rho_1 \otimes \rho_2)$$

$$\geq \int_{X \times Y} e^{\frac{\bar{u} + \bar{v} - c}{\varepsilon}} d(\rho_1 \otimes \rho_2).$$

Moreover,

$$\begin{aligned} \sup_{u, v} D_\varepsilon(u, v) &= \lim_{n \rightarrow \infty} \left\{ \int_X u_n d\rho_1 + \int_Y v_n d\rho_2 - \varepsilon \int_{X \times Y} e^{\frac{u_n + v_n - c}{\varepsilon}} d(\rho_1 \otimes \rho_2) \right\} \\ &\leq \lim_{n \rightarrow \infty} \left\{ \int_X u_n d\rho_1 + \int_Y v_n d\rho_2 \right\} - \varepsilon \liminf_{n \rightarrow \infty} \left\{ \int_{X \times Y} e^{\frac{u_n + v_n - c}{\varepsilon}} d(\rho_1 \otimes \rho_2) \right\} \\ &\leq \int_X \bar{u} d\rho_1 + \int_Y \bar{v} d\rho_2 - \varepsilon \int_{X \times Y} e^{\frac{\bar{u} + \bar{v} - c}{\varepsilon}} d(\rho_1 \otimes \rho_2) = D(\bar{u}, \bar{v}). \end{aligned}$$

So,  $(\bar{u}, \bar{v})$  is a maximizer for  $D_\varepsilon$ . By construction, we have also that  $\bar{u} \in L^\infty(\rho_1)$  and  $\bar{v} \in L^\infty(\rho_2)$ . Finally, the strictly concavity of  $D_\varepsilon$  and Lemma 2.6 implies that the maximizer is unique and, in particular  $\bar{v} = \bar{u}^{(c, \varepsilon)}$ .  $\square$

**Corollary 2.9** *Let  $(X, d_X, m_1), (Y, d_Y, m_2)$  be Polish metric measure spaces,  $c : X \times Y \rightarrow \mathbb{R}$  be a Borel bounded cost function,  $\rho_1 \in \mathcal{P}(X)$  and  $\rho_2 \in \mathcal{P}(Y)$  be probability measures such that  $\text{KL}(\rho_1 | m_1) + \text{KL}(\rho_1 | m_2) < \infty$ . Consider the dual functional  $\tilde{D}_\varepsilon : L_\varepsilon^{\text{exp}}(m_1) \times L_\varepsilon^{\text{exp}}(m_2) \rightarrow \mathbb{R}$ ,*

$$\begin{aligned} \tilde{D}_\varepsilon(u, v) &= \int_X u(x)\rho_1(x)d m_1(x) + \int_Y v(y)\rho_2(y)d m_2(y) \\ &\quad - \varepsilon \int_{X \times Y} e^{\frac{u(x) + v(y) - c(x, y)}{\varepsilon}} d(m_1(x) \otimes m_2(y)). \end{aligned}$$

Then the supremum

$$\sup \{ D_\varepsilon(u, v) : u \in L_\varepsilon^{\text{exp}}(m_1), v \in L_\varepsilon^{\text{exp}}(m_2) \},$$

is attained for a unique couple  $(u_0, v_0)$  and in particular we have

$$u_0 - \varepsilon \log \rho_1 \in L^\infty(m_1) \quad \text{and} \quad v_0 - \varepsilon \log \rho_2 \in L^\infty(m_2).$$

**Proof** The proof follows by the change of variable  $T : (u, v) \mapsto (u - \varepsilon \log \rho_1, v - \varepsilon \log \rho_2)$  which is such that  $\tilde{D}_\varepsilon(u, v) = D_\varepsilon(T(u, v)) + \varepsilon \text{KL}(\rho_1 | m_1) + \varepsilon \text{KL}(\rho_1 | m_2)$ , and Theorem 2.8. Another way is to apply same arguments of theorem (2.8) by using the Entropic  $c$ -transform  $u_{m_1}^{(c, \varepsilon)}$  described in Remark 2.5.  $\square$

In the following proposition an important concept will be that of bivariate transformation. Given  $\mathcal{K}$  a Gibbs measure,  $a(x)$  and  $b(y)$  two measurable function with respect to  $\kappa$ , such that  $a, b \geq 0$ , we define the bivariate transformation of  $\mathcal{K}$  through  $a$  and  $b$  as

$$\kappa(a, b) := a(x)b(y) \cdot \mathcal{K} \tag{2.9}$$

this is still a (possibly infinite) measure.

**Lemma 2.10** *Let  $\varepsilon > 0$  be a positive number,  $(X, d_X)$  and  $(Y, d_Y)$  be Polish metric spaces,  $c : X \times Y \rightarrow \mathbb{R}$  be a cost function (not necessarily bounded),  $\rho_1 \in \mathcal{P}(X)$ ,  $\rho_2 \in \mathcal{P}(Y)$  be probability measures and let  $\kappa$  as in (1.2). Then for every  $\gamma \in \Pi(\rho_1, \rho_2)$ ,  $u \in L_\varepsilon^{\text{exp}}(\rho_1)$  and  $v \in L_\varepsilon^{\text{exp}}(\rho_2)$  then we have*

$$\varepsilon \text{KL}(\gamma | \mathcal{K}) \geq D_\varepsilon(u, v) + \varepsilon, \quad \text{with equality iff } \gamma = \kappa(e^{u/\varepsilon}, e^{v/\varepsilon}), \tag{2.10}$$

where  $\kappa$  is defined as in (2.9).

**Proof** First of all we can assume  $\gamma \ll \mathcal{K}$ , otherwise the right hand side would be  $+\infty$  and so the inequality would be verified; then if we denote (with a slight abuse of notation)  $\gamma(x, y)$  the density of  $\gamma$  with respect to  $\mathcal{K}$ , we get

$$\begin{aligned} \varepsilon \text{KL}(\gamma|\mathcal{K}) &= \int_{X \times Y} cd\gamma + \varepsilon \int_{X \times Y} \gamma \log \gamma d(\rho_1 \otimes \rho_2) \\ &= \int_{X \times Y} (c + \varepsilon \log \gamma - u - v) \cdot \gamma d\rho_1 \otimes \rho_2 + \int_X ud\rho_1 + \int_Y vd\rho_2 \\ &= \int_X ud\rho_1 + \int_Y vd\rho_2 + \int_{X \times Y} (\varepsilon \log \gamma + c - u - v) \cdot \gamma d(\rho_1 \otimes \rho_2) \\ &\geq \int_X ud\rho_1 + \int_Y vd\rho_2 - \varepsilon \int_{X \times Y} e^{\frac{u+v-c}{\varepsilon}} d(\rho_1 \otimes \rho_2) + \varepsilon \\ &= D_\varepsilon(u, v) + \varepsilon, \end{aligned}$$

where we used  $ts + \varepsilon t \ln t - \varepsilon \geq -\varepsilon e^{-s/\varepsilon}$ , with equality if  $t = e^{-s/\varepsilon}$ . Notice in particular that, as we wanted, there is equality iff  $\gamma = e^{(u(x)+v(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2 = \kappa(e^{u/\varepsilon}, e^{v/\varepsilon})$ .  $\square$

**Proposition 2.11** (Equivalence and complementarity condition) *Let  $\varepsilon > 0$  be a positive number,  $(X, d_X)$  and  $(Y, d_Y)$  be Polish metric spaces,  $c : X \times Y \rightarrow \mathbb{R}$  be a bounded cost function,  $\rho_1 \in \mathcal{P}(X)$ ,  $\rho_2 \in \mathcal{P}(Y)$  be probability measures and let  $\kappa$  as in (1.2). Then given  $u^* \in L_\varepsilon^{\text{exp}}(\rho_1)$ ,  $v^* \in L_\varepsilon^{\text{exp}}(\rho_2)$ , the following are equivalent:*

1. (Maximizers)  $u^*$  and  $v^*$  are maximizing potentials for (2.8);
2. (Maximality condition)  $\mathcal{F}^{(c,\varepsilon)}(u^*) = v^*$  and  $\mathcal{F}^{(c,\varepsilon)}(v^*) = u^*$ ;
3. (Schrödinger system) let  $\gamma^* = \kappa(e^{u^*/\varepsilon}, e^{v^*/\varepsilon}) = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$ , then  $\gamma^* \in \Pi(\rho_1, \rho_2)$ ;
4. (Duality attainment)  $\text{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$ .

Moreover in those cases  $\gamma^*$ , as defined in 3, is also the (unique) minimizer for the problem (1.1).

**Proof** We will prove  $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 1$ .

1.  $\Rightarrow$  2. This is a straightforward application of Lemma 2.6. In fact thanks to (2.6) we have  $D_\varepsilon(u^*, \mathcal{F}^{(c,\varepsilon)}(u^*)) \geq D_\varepsilon(u^*, v^*)$ ; however, by the maximality of  $u^*, v^*$  we have also  $D_\varepsilon(u^*, v^*) \geq D_\varepsilon(u^*, \mathcal{F}^{(c,\varepsilon)}(u^*))$ , and so we conclude that  $D_\varepsilon(u^*, \mathcal{F}^{(c,\varepsilon)}(u^*)) = D_\varepsilon(u^*, v^*)$ . Thanks to (2.7) we then deduce that  $v^* = \mathcal{F}^{(c,\varepsilon)}(u^*)$ . We can follow a similar argument to prove that conversely  $u^* = \mathcal{F}^{(c,\varepsilon)}(v^*)$ .

2.  $\Rightarrow$  3. A simple calculation shows for every  $u \in L_\varepsilon^{\text{exp}}(\rho_1)$  and  $v \in L_\varepsilon^{\text{exp}}(\rho_2)$  we have  $(\pi_1)_\#(\kappa(e^{u/\varepsilon}, e^{v/\varepsilon})) = e^{(u-v^{(c,\varepsilon)})/\varepsilon} \rho_1$  and similarly  $(\pi_2)_\#(\kappa(e^{u/\varepsilon}, e^{v/\varepsilon})) = e^{(v-u^{(c,\varepsilon)})/\varepsilon} \rho_2$ . So if we assume 2. it is trivial to see that in fact  $\gamma^* = \kappa(e^{u^*/\varepsilon}, e^{v^*/\varepsilon}) \in \Pi(\rho_1, \rho_2)$

3.  $\Rightarrow$  4. since  $\gamma^* \in \Pi(\rho_1, \rho_2)$ , from Lemma 2.10 we have

$$\varepsilon \text{KL}(\gamma^*|\mathcal{K}) \geq D_\varepsilon(u, v) + \varepsilon \quad \forall u \in L_\varepsilon^{\text{exp}}(\rho_1), v \in L_\varepsilon^{\text{exp}}(\rho_2) \tag{2.11}$$

$$\varepsilon \text{KL}(\gamma|\mathcal{K}) \geq D_\varepsilon(u^*, v^*) + \varepsilon \quad \forall \gamma \in \Pi(\rho_1, \rho_2). \tag{2.12}$$

Moreover, since by definition  $\gamma^* = \kappa(e^{u^*/\varepsilon}, e^{v^*/\varepsilon})$ , Lemma 2.10 assure us also that

$$\varepsilon \text{KL}(\gamma^*|\mathcal{K}) \geq D_\varepsilon(u^*, v^*) + \varepsilon. \tag{2.13}$$

Putting now (2.11),(2.12) and (2.13) together we obtain

$$\varepsilon \text{KL}(\gamma|\mathcal{K}) \geq D_\varepsilon(u^*, v^*) + \varepsilon = \varepsilon \text{KL}(\gamma^*|\mathcal{K}) \geq D_\varepsilon(u, v) + \varepsilon;$$

in particular we have  $\varepsilon \text{KL}(\gamma|\mathcal{K}) \geq \varepsilon \text{KL}(\gamma^*|\mathcal{K})$  which grants us that  $\gamma^*$  is a minimizer for (1.1) and that in particular  $\text{OT}_\varepsilon(\rho_1, \rho_2) = \varepsilon \text{KL}(\gamma^*|\mathcal{K}) = D_\varepsilon(u^*, v^*) + \varepsilon$ .

4.  $\Rightarrow$  1. Looking at (2.10) and minimizing in  $\gamma$  we find that

$$\text{OT}_\varepsilon(\rho_1, \rho_2) \geq D_\varepsilon(u, v) + \varepsilon \quad \forall u \in L_\varepsilon^{\text{exp}}(\rho_1), v \in L_\varepsilon^{\text{exp}}(\rho_2);$$

using that by hypothesis  $\text{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$ , we get that

$$D_\varepsilon(u^*, v^*) \geq D_\varepsilon(u, v) \quad \forall u \in L_\varepsilon^{\text{exp}}(\rho_1), v \in L_\varepsilon^{\text{exp}}(\rho_2),$$

that is,  $u^*, v^*$  are maximizing potentials for (2.8).

Notice that in proving  $3 \Rightarrow 4$  we incidentally proved that  $\gamma^*$  is the (unique) minimizer.  $\square$

Finally, we conclude this section by giving a short proof of the duality between (1.1) and (2.8).

**Proposition 2.12** (General duality) *Let  $\varepsilon > 0$  be a positive number,  $(X, d_X)$  and  $(Y, d_Y)$  be Polish metric spaces,  $c : X \times Y \rightarrow \mathbb{R}$  be a bounded cost function,  $\rho_1 \in \mathcal{P}(X)$ ,  $\rho_2 \in \mathcal{P}(Y)$  be probability measures. Then duality holds*

$$\text{OT}_\varepsilon(\rho_1, \rho_2) = \max \{ D_\varepsilon(u, v) : u \in L_\varepsilon^{\text{exp}}(\rho_1), v \in L_\varepsilon^{\text{exp}}(\rho_2) \} + \varepsilon.$$

**Proof** From Theorem 2.8 we have the existence of a maximizing pair of potentials  $u^*, v^*$ . In particular we have

$$\max \{ D_\varepsilon(u, v) : u \in L_\varepsilon^{\text{exp}}(\rho_1), v \in L_\varepsilon^{\text{exp}}(\rho_2) \} + \varepsilon = D_\varepsilon(u^*, v^*) + \varepsilon;$$

this, together with point 4 in Proposition 2.11 (which is true since 1 holds true), proves the duality.  $\square$

By a similar argument, one can show that the duality holds also for the functional  $\mathcal{S}_\varepsilon(\rho_1, \rho_2; \mathfrak{m}_1, \mathfrak{m}_2)$ .

**Corollary 2.13** *Let  $\varepsilon > 0$  be a positive number,  $(X, d_X, \mathfrak{m}_1)$  and  $(Y, d_Y, \mathfrak{m}_2)$  be Polish metric measure spaces,  $c : X \times Y \rightarrow \mathbb{R}$  be a bounded cost function,  $\rho_1 \in \mathcal{P}(X)$ ,  $\rho_2 \in \mathcal{P}(Y)$  be probability measures. Then duality holds*

$$\mathcal{S}_\varepsilon(\rho_1, \rho_2; \mathfrak{m}_1, \mathfrak{m}_2) = \max \{ \tilde{D}_\varepsilon(u, v) : u \in L_\varepsilon^{\text{exp}}(\mathfrak{m}_1), v \in L_\varepsilon^{\text{exp}}(\mathfrak{m}_2) \} + \varepsilon.$$

### 3 Convergence of the Sinkhorn/IPFP Algorithm

In this section, we give an alternative proof for the convergence of the Sinkhorn algorithm. The aim of the Iterative Proportional Fitting Procedure (IPFP, also known as Sinkhorn algorithm) is to construct the measure  $\gamma^\varepsilon$  realizing minimum in (1.1) by alternatively matching one marginal distribution to the target marginals  $\rho_1$  and  $\rho_2$ : this leads to the construction of the IPFP sequences  $(a^n)_{n \in \mathbb{N}}$  and  $(b^n)_{n \in \mathbb{N}}$ , defined in (1.4).

We now look at the new variables  $u_n := \varepsilon \ln(a^n)$  and  $v_n := \varepsilon \ln(b^n)$ : we can then rewrite the system (1.4) as

$$\begin{aligned} v_n(y)/\varepsilon &= -\log\left(\int_X k(x, y)e^{\frac{u_{n-1}(x)}{\varepsilon}} d\rho_1\right) \\ u_n(x)/\varepsilon &= -\log\left(\int_Y k(x, y)e^{\frac{v_n(y)}{\varepsilon}} d\rho_2\right) \end{aligned}$$

In other words, using the  $(c, \varepsilon)$ -transform and the expression of  $k(x, y)$  given in (1.2),  $v^n(y) = (u^{(n-1)})^{(c, \varepsilon)}$  and  $u^n(x) = (v^n)^{(c, \varepsilon)}$ .

**Theorem 3.1** *Let  $(X, d_X)$  and  $(Y, d_Y)$  be Polish metric spaces,  $\rho_1 \in \mathcal{P}(X)$  and  $\rho_2 \in \mathcal{P}(Y)$  be probability measures and  $c : X \times Y \rightarrow \mathbb{R}$  be a Borel bounded cost. If  $(a^n)_{n \in \mathbb{N}}$  and  $(b^n)_{n \in \mathbb{N}}$  are the IPFP sequences defined in (1.4), then there exists a sequence of positive real numbers  $(\lambda^n)_{n \in \mathbb{N}}$  such that*

$$a^n/\lambda^n \rightarrow a \text{ in } L^p(\rho_1) \quad \text{and} \quad \lambda^n b^n \rightarrow b \text{ in } L^p(\rho_2), \quad 1 \leq p < +\infty,$$

where  $(a, b)$  solve the Schrödinger problem. In particular, the sequence  $\gamma^n = a^n b^n k$ , where  $k$  is defined in (1.2) converges in  $L^p(\rho_1 \otimes \rho_2)$  to  $\gamma_{opt}^\varepsilon$ , the density of the minimizer of (1.1) with respect to  $\rho_1 \otimes \rho_2$ , for any  $1 \leq p < +\infty$ .

**Proof** Let  $(a^n)_{n \in \mathbb{N}}$  and  $(b^n)_{n \in \mathbb{N}}$  be the IPFP sequence defined in (1.4). Let us write  $a^n = e^{u_n/\varepsilon}$ ,  $b^n = e^{v_n/\varepsilon}$ ; then, in this new variables, we noticed that the iteration can be written with the help of the  $(c, \varepsilon)$ -transform:

$$\begin{cases} v_{2n+1} = (u_{2n})^{(c, \varepsilon)} \\ u_{2n+1} = u_{2n} \end{cases}, \quad \begin{cases} v_{2n+2} = v_{2n+1} \\ u_{2n+2} = (v_{2n+1})^{(c, \varepsilon)} \end{cases}.$$

Notice that, as soon as  $n \geq 2$ , we have  $u_n \in L^\infty(\rho_1)$  and  $v_n \in L^\infty(\rho_2)$  thanks to the regularizing properties of the  $(c, \varepsilon)$ -transform proven in Lemma 2.3 and, moreover, thanks to (2.6) and Proposition 2.12 we have

$$D_\varepsilon(u_n, v_n) \leq D_\varepsilon(u_{n+1}, v_{n+1}) \leq \dots \leq OT_\varepsilon(\rho_1, \rho_2) - \varepsilon.$$

Then, by the same argument used in the proof of Lemma 2.7 it is easy to prove that there for each  $n \geq 2$  there exists  $\ell_n \in \mathbb{R}$  such that  $\|u_n - \ell_n\|_\infty, \|v_n + \ell_n\| \leq \frac{3}{2}\|c\|_\infty$ . Now, thanks to Proposition 2.4 we have that the sequences  $u_n - \ell_n$  and  $v_n + \ell_n$  are precompact in every  $L^p$ , for  $1 \leq p < \infty$ ; in particular let us consider any limit point  $u, v$ . Then we have a subsequence  $u_{n_k}, v_{n_k}$  such that  $u_{n_k} \rightarrow u, v_{n_k} \rightarrow v$  in  $L^\infty$  and  $u_{n_k+1} = (v_{n_k})^{(c, \varepsilon)}$  (or the opposite). Using the continuity in  $L^p$  of the  $(c, \varepsilon)$ -transform, and the fact that an increasing and bounded sequence has vanishing increments, we obtain

$$D_\varepsilon(v^{(c, \varepsilon)}, v) - D_\varepsilon(u, v) = \lim_{n_k \rightarrow \infty} D_\varepsilon(u_{n_k+1}, v_{n_k+1}) - D_\varepsilon(u_{n_k}, v_{n_k}) = 0.$$

In particular, by (2.7), we have  $u = v^{(c, \varepsilon)}$ . Analogously, we obtain that  $v = u^{(c, \varepsilon)}$  by doing the same calculation using the potentials  $(u_{n_k+2}, v_{n_k+2})$  and then

$$D_\varepsilon(u, u^{(c, \varepsilon)}) - D_\varepsilon(u, v) = \lim_{n_k \rightarrow \infty} D_\varepsilon(u_{n_k+2}, v_{n_k+2}) - D_\varepsilon(u_{n_k}, v_{n_k}) = 0.$$

Now we can use Proposition 2.11: the implication  $2 \Rightarrow 1$  proves that  $(u, v)$  is a maximizer.<sup>2</sup> In particular  $a = e^{u/\varepsilon}, b = e^{v/\varepsilon}$  are solutions of the Schrödinger equation and taking  $\lambda^n = e^{\ell_n/\varepsilon}$

<sup>2</sup> in order to prove that there is a unique limit point at this stage, it is sufficient to take  $\ell_n$  that minimizes  $\|u_n - \ell_n - u\|_2$ .

we get the convergence result for  $a^n$  and  $b^n$ , using that the exponential is Lipschitz in bounded domains.

In order to prove also the convergence of the plans, it is sufficient to note that for free we have  $u_n + v_n \rightarrow u + v$  in  $L^p(\rho_1 \otimes \rho_2)$ , since now the translations are cancelled. Again, the fact that the exponential is Lipschitz on bounded domains and the boundedness of  $k$ , will let us conclude that in fact  $\gamma^n \rightarrow \gamma$  in  $L^p(\rho_1 \otimes \rho_2)$  for every  $1 \leq p < \infty$ .  $\square$

**Remark 3.2** Notice that as long as we have more hypothesis on the smoothness of the cost function  $c$  we can use precompactness of the sequences  $u_n - \ell_n$  and  $v_n + \ell_n$  on larger space, obtaining faster convergence. For example if  $c$  is uniformly continuous we will get the uniform convergence instead of strong  $L^p$  convergence.

### 4 Multi-marginal Schrödinger Problem

In this section we generalize the results obtain previously for the Schrödinger problem with more than two marginals, including a proof of convergence of the Sinkhorn algorithm in the several marginals case.

We consider  $(X_1, d_1), \dots, (X_N, d_N)$  Polish spaces,  $\rho_1, \dots, \rho_N$  probability measures respectively in  $X_1, \dots, X_N$  and  $c : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$  a bounded cost. Define  $\rho^N = \rho_1 \otimes \dots \otimes \rho_N$  by the product measure. For every  $\gamma \in \mathcal{M}(X_1 \times \dots \times X_N)$ , the relative entropy of  $\gamma$  with respect to the Gibbs Kernel  $\mathcal{K}(x_1, \dots, x_N) = k^N(x_1, \dots, x_N)\rho^N = e^{-\frac{c(x_1, \dots, x_N)}{\varepsilon}}$   $d\rho_1 \otimes \dots \otimes \rho_N$  is defined by

$$\text{KL}^N(\gamma|\mathcal{K}) = \begin{cases} \int_{X_1 \times \dots \times X_N} \gamma \log\left(\frac{\gamma}{k^N}\right) d\rho^N & \text{if } \gamma \ll \rho^N \\ +\infty & \text{otherwise.} \end{cases} \tag{4.1}$$

An element  $\gamma \in \Pi(\rho_1, \dots, \rho_N)$  is called coupling and is a probability measure on the product space  $X_1 \times \dots \times X_N$  having the *ith*-marginal equal to  $\rho_i$ , i.e  $\gamma \in \mathcal{P}(X_1 \times \dots \times X_N)$  such that  $(e_i)_\# \gamma = \rho_i, \forall i \in \{1, \dots, N\}$ .

The Multi-marginal Schrödinger problem is defined as the infimum of the Kullback–Leibler divergence  $\text{KL}^N(\gamma|\mathcal{K})$  over the couplings  $\gamma \in \Pi(\rho_1, \dots, \rho_N)$

$$\text{OT}_\varepsilon^N(\rho_1, \dots, \rho_N) = \inf_{\gamma \in \Pi(\rho_1, \dots, \rho_N)} \varepsilon \int_{X_1 \times \dots \times X_N} \text{KL}(\gamma|\mathcal{K}) d\gamma. \tag{4.2}$$

Optimal Transport problems with several marginals or its entropic-regularization appears, for instance, in economics Carlier and Ekeland [17], and Chiappori et al. [22]; imaging (e.g. [27,68]); and in theoretical chemistry (e.g. [30,42,46]). The first important instance of such kind of problems is attributed to Brenier’s generalised solutions of the Euler equations for incompressible fluids [9–11].

We point out that the entropic-regularization of the multi-marginal transport problem leads to a problem of multi-dimensional matrix scaling [36,64]. An important example in this setting is the Entropy-Regularized Wasserstein Barycenter introduced by Agueh and Carlier [1]. The Wasserstein Barycenter defines a non-linear interpolation between several probabilities measures generalizing the Euclidian barycenter and turns out to be equivalent to Gangbo-Świąch cost [38], that is  $c(x_1, \dots, x_N) = \frac{1}{2} \|x_j - x_i\|^2$ .

In the next section we extend to the multi-marginal setting the notions and properties of the Entropy  $c$ -transform done in Sect. 2. As a consequence, we generalise the proof of convergence of IPFP.

### 4.1 Entropy-Transform

Analogously to Definitions (2.1) and (2.2) in Sect. 2.1, we define the following Entropy  $c$ -transforms  $\hat{u}_1^{(N,c,\varepsilon)}, \hat{u}_2^{(N,c,\varepsilon)}, \dots, \hat{u}_N^{(N,c,\varepsilon)}$ . Notice that the notation  $\hat{u}_i$  stands for  $\hat{u}_i = (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_N)$ .

**Definition 4.1** (*Entropic  $c$ -transform or  $(c, \varepsilon)$ -transform*) Let  $i \in \{1, \dots, N\}$  and  $\varepsilon > 0$  be a positive number. Consider  $(X_i, d_{X_i})$  Polish spaces,  $\rho_i \in \mathcal{P}(X_i)$  probability measures and let  $c$  a bounded measurable cost on  $X_1 \times \dots \times X_N$ . For every  $i$ , the Entropy  $c$ -transform  $\hat{u}_i^{(N,c,\varepsilon)}$  is defined by the functional  $\mathcal{F}_i^{(N,c,\varepsilon)} : \prod_{j \neq i} L_\varepsilon^{\text{exp}}(\rho_j) \rightarrow L^0(\rho_i)$ ,

$$\hat{u}_i^{(N,c,\varepsilon)}(x_i) = \mathcal{F}_i^{(N,c,\varepsilon)}(\hat{u}_i)(x_i) = -\varepsilon \log \left( \int_{\prod_{j \neq i} X_j} e^{\frac{\sum_{j \neq i} u_j(x_j) - c(x_1, \dots, x_N)}{\varepsilon}} d(\otimes_{j \neq i} \rho_j) \right). \tag{4.3}$$

In particular, we have  $\hat{u}_i^{(N,c,\varepsilon)} \in L_\varepsilon^{\text{exp}}(\rho_i)$ . For  $u_i \in L_\varepsilon^{\text{exp}}(X_i, \rho)$ , we denote the constant  $\lambda_{u_i}$  by

$$\lambda_{u_i} = \varepsilon \log \left( \int_{\prod_{j \neq i} X_j} e^{\frac{\sum_{j \neq i} u_j(x_j)}{\varepsilon}} d(\otimes_{j \neq i} \rho_j) \right).$$

There is also the possibility to reconduce us to the case  $N = 2$ : notice that if one considers the spaces  $X_i$  and  $Y_i = \prod_{j \neq i} X_j$ , then  $c$  is also a natural cost function on  $X_i \times Y_i$ . We can then consider  $\rho_i$  as a measure on  $X_i$  and  $\otimes_{j \neq i} \rho_j$  as a measure on  $Y_i$ . In this way we able to construct an entropic  $c$ -trasform  $\mathcal{F}^{(c,\varepsilon)}$  associated to this 2-marginal problem and it is clear that

$$\mathcal{F}_i^{(N,c,\varepsilon)}(\hat{u}_i) = \mathcal{F}^{(c,\varepsilon)}\left(\sum_{j \neq i} u_j\right).$$

The following lemma extend Lemma 2.3 in the multi-marginal setting. We omit the proof since it follow by similar arguments.

**Lemma 4.2** *For every  $i \in \{1, \dots, N\}$ , the Entropy  $c$ -transform  $\hat{u}_i^{(N,c,\varepsilon)}$  is well defined. Moreover,*

(i)  $\hat{u}_i^{(N,c,\varepsilon)} \in L^\infty(\rho_i)$ . In particular,

$$\begin{aligned} -\|c\|_\infty - \varepsilon \log \left( \int_{\prod_{j \neq i} X_j} e^{\frac{\sum_{j \neq i} u_j(x_j)}{\varepsilon}} d(\otimes_{j \neq i} \rho_j) \right) &\leq \hat{u}_i^{(N,c,\varepsilon)}(x_i) \\ &\leq \|c\|_\infty - \varepsilon \log \left( \int_{\prod_{j \neq i} X_j} e^{\frac{\sum_{j \neq i} u_j(x_j)}{\varepsilon}} d(\otimes_{j \neq i} \rho_j) \right). \end{aligned}$$

(ii)  $\hat{u}_i^{(N,c,\varepsilon)} \in L_\varepsilon^{\text{exp}}(\rho_i)$ .

(iii)

$$|\hat{u}_i^{(N,c,\varepsilon)}(x_i) + \sum_{i \neq j} \lambda_{u_j}| \leq \|c\|_\infty. \tag{4.4}$$

(iv) if  $c$  is  $L$ -Lipschitz (resp.  $\omega$ -continuous), then  $\hat{u}_i^{(N,c,\varepsilon)}$  is  $L$ -Lipschitz (resp.  $\omega$ -continuous);

(v) if  $|c| \leq M$ , then  $\text{osc}(\hat{u}_i^{(N,c,\varepsilon)}) \leq 2M$  and  $\mathcal{F}_i^{(N,c,\varepsilon)} : \prod_{j \neq i} L^\infty(\rho_j) \rightarrow L^p(\rho_i)$  for  $i = 1, \dots, n$  are compact operators for every  $1 \leq p < \infty$ .

### 4.2 Entropy-Kantorovich Duality

We introduce the dual functional  $D_\varepsilon^N : L_\varepsilon^{\text{exp}}(\rho_1) \times \dots \times L_\varepsilon^{\text{exp}}(\rho_N) \rightarrow [0, +\infty]$ ,

$$D_\varepsilon^N(u_1, \dots, u_N) = \sum_{i=1}^N \int_{X_i} u_i d\rho_i - \varepsilon \int_{X_1 \times \dots \times X_N} e^{\frac{\sum_{i=1}^N u_i(x_i) - c(x_1, \dots, x_N)}{\varepsilon}} d(\rho_1 \otimes \dots \otimes \rho_N). \tag{4.5}$$

In the sequel we will use the invariance by translation of the dual problem, and thus we introduce the following projection operator:

**Lemma 4.3** *Let us consider the operator  $P : \prod_{i=1}^N L^\infty(\rho_i) \rightarrow \prod_{i=1}^N L^\infty(\rho_i)$  defined as*

$$P_i(u) = \begin{cases} u_i - \lambda_{u_i} & \text{if } i = 1, \dots, N - 1 \\ u_i + \sum_{j \neq i}^{N-1} \lambda_{u_j} & \text{if } i = N. \end{cases}$$

Then the following properties hold

- (i)  $D_\varepsilon^N(P(u)) = D_\varepsilon^N(u)$ ;
- (ii)  $\|P_i(u)\|_\infty \leq \text{osc}(u_i) + |\sum_{i=1}^N \lambda_{u_i}|$ , for all  $i = 1, \dots, N$ ;
- (iii) let  $v = P(u)$ . Then  $u_i = \mathcal{F}_i^{(N,c,\varepsilon)}(\hat{u}_i)$  if only if  $v_i = \mathcal{F}_i^{(N,c,\varepsilon)}(\hat{v}_i)$ .

**Proof**

- (i) In order to prove  $D_\varepsilon^N(P(u)) = D_\varepsilon^N(u)$  we first observe that

$$\sum_{i=1}^N P_i(u)(x_i) = u_N(x_N) + \sum_{i=1}^{N-1} \lambda_{u_i} + \sum_{i=1}^{N-1} (u_i(x_i) - \lambda_{u_i}) = \sum_{i=1}^N u_i(x_i).$$

In particular we have (here we denote  $X = X_1 \times \dots \times X_N$ )

$$\begin{aligned} D_\varepsilon^N(P(u)) &= \sum_{i=1}^N \int_{X_i} P_i(u) d\rho_i - \varepsilon \int_{X_1 \times \dots \times X_N} e^{\frac{\sum_{i=1}^N P_i(u)(x_i) - c(x_1, \dots, x_N)}{\varepsilon}} d(\rho_1 \otimes \dots \otimes \rho_N) \\ &= \int_X \sum_{i=1}^N P_i(u)(x_i) d\rho^N - \varepsilon \int_X e^{\frac{\sum_{i=1}^N P_i(u)(x_i) - c(x_1, \dots, x_N)}{\varepsilon}} d\rho^N \\ &= \int_X \sum_{i=1}^N u_i(x_i) d\rho^N - \varepsilon \int_X e^{\frac{\sum_{i=1}^N u_i(x_i) - c(x_1, \dots, x_N)}{\varepsilon}} d\rho^N = D_\varepsilon^N(u) \end{aligned}$$

- (ii) The inequality is not trivial only if  $u_i \in L^\infty(\rho_i)$ . In this case obviously we have  $\inf u_i \leq \lambda_{u_i} \leq \sup u_i$  and in particular

$$-\text{osc}(u_i) = \inf u_i - \sup u_i \leq u_i(x_i) - \lambda_{u_i} \leq \sup u_i - \inf u_i = \text{osc}(u_i),$$

that is  $\|u_i - \lambda_{u_i}\|_\infty \leq \text{osc}(u_i)$ . This proves already the bound for  $i < N$ ; for  $i = N$  we have, letting  $\lambda = \sum_{i=1}^N \lambda_{u_i}$

$$\|P_N(u)\|_\infty = \|u_N - \lambda_{u_N} + \sum_{i=1}^N \lambda_{u_i}\|_\infty \leq \|u_N - \lambda_{u_N}\|_\infty + |\lambda| \leq \text{osc}(u_N) + |\lambda|$$

- (iii) This is obvious from the fact that  $\mathcal{F}_i^{(N,c,\varepsilon)}(\widehat{u_i - \lambda_i}) = \mathcal{F}_i^{(N,c,\varepsilon)}(\hat{u}_i) + \sum_{j \neq i} \lambda_j$ .

□

This projection operator allows us to generalize Lemma 2.7:

**Lemma 4.4** *Let us consider  $u_i \in L_\varepsilon^{\text{exp}}(\rho_i)$ , for every  $i = 1, \dots, N$ . Then there exist  $u_i^* \in L_\varepsilon^{\text{exp}}(\rho_i)$  for  $i = 1, \dots, N$  such that*

- $D_\varepsilon^N(u_1, \dots, u_N) \leq D_\varepsilon^N(u_1^*, \dots, u_N^*)$ ;
- $\|u_i^*\|_\infty \leq 3\|c\|_\infty$  for every  $i = 1, \dots, N$ .

**Proof** Let us construct the following sequence of potentials:

$$\begin{cases} u_1^1 = \mathcal{F}_i^{(N,c,\varepsilon)}(\widehat{u}_1) \\ u_2^1 = u_2 \\ u_3^1 = u_3 \\ \dots \\ u_N^1 = u_N \end{cases}, \quad \begin{cases} u_1^2 = u_1^1 \\ u_2^2 = \mathcal{F}_i^{(N,c,\varepsilon)}(\widehat{u}_2) \\ u_3^2 = u_3^1 \\ \dots \\ u_N^2 = u_N^1 \end{cases}, \dots, \quad \begin{cases} u_1^N = u_1^{N-1} \\ u_2^N = u_2^{N-1} \\ u_3^N = u_3^{N-1} \\ \dots \\ u_N^N = \mathcal{F}_i^{(N,c,\varepsilon)}(\widehat{u}_N^{N-1}). \end{cases}$$

Then let us consider  $u^* = P(u^N)$ . First of all we notice that, using the multimarginal analogous of Lemma 2.6 we have

$$D_\varepsilon^N(u_1, \dots, u_N) \leq D_\varepsilon^N(u_1^1, \dots, u_N^1) \leq \dots \leq D_\varepsilon^N(u_1^N, \dots, u_N^N) = D_\varepsilon^N(u_1^*, \dots, u_N^*).$$

Then is clear by construction that for every  $i = 1, \dots, N$  we have  $u_i^N = u_i^i$  and in particular, by Lemma 4.2 (iv) we have  $\text{osc}(u_i^N) \leq 2\|c\|_\infty$ . Moreover, thanks to (4.4) it is easy to see that  $|\sum_i \lambda_{u_i^N}| \leq \|c\|_\infty$ . Now we can use Lemma 4.3 (ii) to conclude that in fact  $\|u_i^*\| \leq 3\|c\|_\infty$ . □

Similarly to Theorem 2.8, Propositions 2.11 and 2.12, the next theorem and the following Proposition state the existence of a maximizer and the Entropic-Kantorovich duality to the multi-marginal case, along with the complementarity conditions. Since the proofs follows the same lines of the case  $N = 2$ , without big changes, we will omit them.

**Theorem 4.5** *For every  $i \in \{1, \dots, N\}$ , let  $(X_i, d_i)$  be Polish metric spaces,  $\rho_i \in \mathcal{P}(X_i)$  be a probability measures and  $c : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$  be a bounded cost function. Then for every  $\varepsilon > 0$ ,*

(i) *The dual function  $D_\varepsilon^N$  is well defined on its definition domain and moreover*

$$D_\varepsilon^N(\widehat{u}_1^{(N,c,\varepsilon)}, \widehat{u}_1) \geq D_\varepsilon^N(u_1, \dots, u_N), \quad \forall u_i \in L_\varepsilon^{\text{exp}}(\rho_i), \tag{4.6}$$

$$D_\varepsilon^N(\widehat{u}_1^{(N,c,\varepsilon)}, \widehat{u}_1) = D_\varepsilon^N(u_1, \dots, u_N) \text{ if and only if } u_1 = \widehat{u}_1^{(N,c,\varepsilon)}. \tag{4.7}$$

(ii) *The supremum is attained, up to trivial transformations, for a unique  $N$ -tuple  $(u_1^0, \dots, u_N^0)$  and in particular we have*

$$u_i^0 \in L^\infty(\rho_i), \quad \forall i \in \{1, \dots, N\}.$$

*Moreover if we consider  $\gamma^{0,N} = e^{(\sum_i u_i^0(x_i) - c)/\varepsilon} \rho^N$  then  $\gamma^{0,N}$  is the minimizer of (4.2)*

(iii) *Duality holds:*

$$\text{OT}_\varepsilon^N(\rho_1, \dots, \rho_N) = \sup \left\{ D_\varepsilon^N(u_1, \dots, u_N) : u_i \in L_\varepsilon^{\text{exp}}(\rho_i), i \in \{1, \dots, N\} \right\} + \varepsilon.$$

Finally, the result extends the main results of the previous section to the multi-marginal case.

**Proposition 4.6** (Equivalence and complementarity condition) *Let  $\varepsilon > 0$  and for every  $i \in \{1, \dots, N\}$ , let  $(X_i, d_i)$  be Polish metric spaces,  $\rho_i \in \mathcal{P}(X_i)$  be a probability measures and  $c : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$  be a bounded cost function. Then given  $u_i^* \in L^\varepsilon^{\text{exp}}(\rho_i)$  for every  $i = 1, \dots, N$ , the following are equivalent:*

1. (Maximizers)  $u_1^*, \dots, u_N^*$  are maximizing potentials for (4.5);
2. (Maximality condition)  $\mathcal{F}_i^{(N,c,\varepsilon)}(u_i^*) = u_i^*$  for every  $i = 1, \dots, N$ ;
3. (Schrödinger system) let  $\gamma^* = e^{(\sum_i u_i^*(x_i) - c)/\varepsilon} \cdot \rho^N$ , then  $\gamma^* \in \Pi(\rho_1, \dots, \rho_N)$ ;
4. (Duality attainment)  $\text{OT}_\varepsilon^N(\rho_1, \dots, \rho_N) = D_\varepsilon^N(u_1^*, \dots, u_N^*) + \varepsilon$ .

Moreover in those cases  $\gamma^*$ , as defined in 3, is also the (unique) minimizer for the problem (4.2).

The part 3. in proposition 4.6 have been already shown in different settings by J.M Borwein, A.S. Lewis and R.D. Nussbaum ([7, Theorem 4.4], see also [8, section 3]) and G. Carlier & M. Laborde [18]. Our approach, being purely variational, allows us to study the convergence of the Sinkhorn algorithm in the several marginal case in a similar way of done in the previous section.

In fact,  $\text{OT}_\varepsilon^N(\rho_1, \dots, \rho_N)$  defines an unique element  $\gamma_{N,opt}^\varepsilon$  - the  $\text{KL}^N$ -projection on  $\Pi(\rho_1, \dots, \rho_N)$ —which has product density  $\prod_{i=1}^N a_i$  with respect to the Gibbs measures  $\mathcal{K}$ , where  $a_i = e^{u_i^*/\varepsilon}$  as defined in proposition 4.6. Also in this case, an equivalent system to (1.3) can be implicitly written:  $\gamma_{N,opt}^\varepsilon$  is a solution of (4.2) if and only if  $\gamma_{N,opt}^\varepsilon = \otimes_{i=1}^N a_i^\varepsilon(x_i)\mathcal{K}$ , where  $a_i^\varepsilon$  solve

$$a_i^\varepsilon(x) \int_Y \otimes_{j \neq i}^N a_j(x_j)k(x_1, \dots, x_N)d\rho_j(x_j) = 1, \quad \forall i = 1, \dots, N. \tag{4.8}$$

Therefore, by using the marginal condition  $\gamma^\varepsilon \in \Pi(\rho_1, \dots, \rho_N)$ , the functions  $a_i$  can be implicitly computed

$$a_i(x_i) = \frac{1}{\int \prod_{j \neq i}^N a_j(x_j)k(x_1, \dots, x_N)d(\otimes_{j \neq i}^N \rho_j)}, \quad \forall i \in \{1, \dots, N\}.$$

### 4.3 Convergence of the IPFP/Sinkhorn Algorithm for Several Marginals

The goal of this subsection is to prove the convergence of the IPFP/Sinkhorn algorithm in the multi-marginal setting. Analogously to (1.4), define recursively the sequences  $(a_j^n)_{n \in \mathbb{N}}$ ,  $j \in \{1, \dots, N\}$  by

$$\begin{aligned} a_1^0(x_1) &= 1, \\ a_j^0(x_j) &= 1, \quad j \in \{2, \dots, N\}, \\ a_j^n(x_j) &= \frac{1}{\int \otimes_{i < j}^N a_i^n(x_i) \otimes_{i > j}^N a_i^{n-1}(x_i)k^N(x_1, \dots, x_N)d(\otimes_{i \neq j}^N \rho_i)}, \quad \forall n \in \mathbb{N}. \end{aligned} \tag{4.9}$$

Also here, by writing  $a_j^n = \exp(u_j^n/\varepsilon)$ , for all  $j \in \{1, \dots, N\}$ , one can rewrite the IPFP sequences (4.9) in terms of Entropic  $(c, \varepsilon)$ -transforms,

$$u_j^n(x_j) = -\varepsilon \log \left( \int_{\prod_{i \neq j} X_i} k^N(x_1, \dots, x_N) \otimes_{i \neq j} e^{u_i^n(x_i)/\varepsilon} d(\otimes_{i \neq j}^N \rho_i) \right)$$

$$= (\widehat{u^n}_j)^{(N,c,\varepsilon)}(x_j).$$

Then, the proof of convergence of the IPFP in the multi-marginal case, follows a method similar to the one used in Theorem 3.1.

**Theorem 4.7** *Let  $(X_1, d_1), \dots, (X_N, d_N)$  be Polish spaces,  $\rho_1, \dots, \rho_N$  be probability measures in  $X_1, \dots, X_N$ ,  $c : X_1 \times \dots \times X_N \rightarrow [0, +\infty]$  be a bounded cost,  $p$  be an integer  $1 \leq p < \infty$ . If  $(a^n_j)_{n \in \mathbb{N}}, j \in \{1, \dots, N\}$  are the IPFP sequence defined in (4.9), then there exist a sequence  $\lambda^n \in \mathbb{R}^N$ , with  $\lambda^n_i > 0$  and  $\prod_{i=1}^N \lambda^n_i = 1$  such that*

$$\forall j \in \{1, \dots, N\}, \quad a^n_j / \lambda^n_j \rightarrow a_j \text{ in } L^p(\rho_j),$$

where  $(a_j)_{j=1}^N$  solve the Schrödinger system. In particular, the sequence  $\gamma^n = \prod_{i=1}^N a^n_i \mathcal{K}$  converges in  $L^p(\rho_1 \otimes \dots \otimes \rho_N)$  to the optimizer  $\gamma^{\varepsilon}_{opt}$  in (4.2).

**Proof** Let  $i \in \{1, \dots, N\}$  and consider Let  $(a^n_i)_{n \in \mathbb{N}}$  the IPFP sequence defined in (4.9). For every  $i$ , we define  $u^n_i := \varepsilon \ln(a^n_i)$  and then iteratively define the following potentials for every  $p \in \mathbb{N}$

$$\left\{ \begin{array}{l} u_1^{pN+1} = (\widehat{u_1^{pN}})^{(c,\varepsilon)} \\ u_2^{pN+1} = u_2^{pN} \\ u_3^{pN+1} = u_3^{pN} \\ \dots \\ u_N^{pN+1} = u_N^{pN} \end{array} \right. , \quad \left\{ \begin{array}{l} u_1^{pN+2} = u_1^{pN+1} \\ u_{pN+2} = (\widehat{u_2^{pN+1}})^{(c,\varepsilon)} \\ u_3^{pN+2} = u_3^{pN+1} \\ \dots \\ u_N^{pN+2} = u_N^{pN+1} \end{array} \right. , \dots , \quad \left\{ \begin{array}{l} u_1^{pN+N} = u_1^{pN+N-1} \\ u_2^{pN+N} = u_2^{pN+N-1} \\ u_3^{pN+N} = u_3^{pN+N-1} \\ \dots \\ u_N^{pN+N} = (\widehat{u_N^{pN+N-1}})^{(c,\varepsilon)} \end{array} \right. .$$

Notice that  $a^n_i = e^{u^n_i / \varepsilon}$  and moreover  $osc(u^n_i)^{pN+i} \leq \|c\|_\infty$  by Lemma 4.2 (iv), and in particular  $osc(u^n_i) \leq \|c\|_\infty$  as long as  $n \geq N$ . Moreover, thanks to (4.4) we also have  $|\sum_i \lambda u^n_i| \leq \|c\|_\infty$ . In particular, defining  $v^n = P(u^n)$ , we have  $\|v^n\|_\infty \leq 3\|c\|_\infty$  thanks to Lemma 4.3 (ii); using (4.6) and Lemma 4.3 we also have

$$D_\varepsilon^N(v_1^n, \dots, v_N^n) \leq D_\varepsilon^N(v_1^{n+1}, \dots, v_N^{n+1}) \leq \dots \leq OT_\varepsilon(\rho_1, \rho_2, \dots, \rho_N).$$

By the boundedness of  $\|v^n\|_\infty$ , by the compactness in Lemma 4.2 (iv), there exists a subsequence  $k_n$  such that  $v^{k_n}_i$  converges in  $L^p$  to some  $v_i$  for every  $i = 1, \dots, N$ ; by pigeon-hole principle we have that at least a class of residue modulo  $N$  is taken infinitely by the sequence  $k_n$  and we will suppose that without loss of generality this residue class is 0. Up to restricting to the infinite subsequence such that  $k_n \equiv 0 \pmod{N}$ , we can assume that  $v^{k_n}_N = (\widehat{v^{k_n}_N})^{(N,c,\varepsilon)}$  and  $v^{k_n+1}_1 = (\widehat{v^{k_n}_1})^{(N,c,\varepsilon)}$

In particular, by the continuity of the  $(N, c, \varepsilon)$ -transform we have

$$\begin{aligned} & D_\varepsilon^N(\widehat{v_1^{(N,c,\varepsilon)}} , \widehat{v_1}) - D_\varepsilon^N(v_1, v_2, \dots, v_N) \\ &= \lim_{n \rightarrow \infty} D_\varepsilon(v_1^{k_n+1}, \dots, v_N^{k_n+1}) - D_\varepsilon(v_1^{k_n}, \dots, v_N^{k_n}) = 0. \end{aligned}$$

In particular, we have  $v_1 = \widehat{v_1^{(N,c,\varepsilon)}}$  by (4.7) and in particular  $u^{k_n+1}_i \rightarrow u_i$  for every  $i = 1, \dots, N$ . Now, doing a similar computation, for every  $i = 2, \dots, N$ , we can inductively prove that, for every  $i$ ,

$$\begin{aligned} & D_\varepsilon(\widehat{v_i^{(N,c,\varepsilon)}} , \widehat{v_i}) - D_\varepsilon(v_1, \dots, v_i, \dots, v_N) \\ &= \lim_{n \rightarrow \infty} D_\varepsilon(v_1^{k_n+i}, \dots, v_N^{k_n+i}) - D_\varepsilon(v_1^{k_n+i-1}, \dots, v_N^{k_n+i-1}) = 0. \end{aligned}$$

Hence,  $v_i = \hat{v}_i^{(N,c,\varepsilon)}$ ,  $\forall i \in \{1, \dots, N\}$ . The result follows by noticing that  $(e^{v_1/\varepsilon}, \dots, e^{v_N/\varepsilon})$  solves the Schrödinger system, by Proposition 4.6.  $\square$

*Remark on the multi-marginal problem*  $\mathcal{S}_\varepsilon(\rho_1, \dots, \rho_N; \mathbf{m}_1, \dots, \mathbf{m}_N)$ : More generally, we could also consider the multi-marginal Schrödinger problem with references measures  $\mathbf{m}_i \in \mathcal{P}(X_i)$ ,  $i = 1, \dots, N$ . For simplicity, we denote  $\bar{\rho} = (\rho_1, \dots, \rho_N)$  and  $\bar{\mathbf{m}} = (\mathbf{m}_1, \dots, \mathbf{m}_N)$ . Then the functional  $\mathcal{S}_\varepsilon(\bar{\rho}; \bar{\mathbf{m}})$  is defined by

$$\mathcal{S}_\varepsilon(\bar{\rho}; \bar{\mathbf{m}}) = \min_{\gamma \in \Pi_N(\rho_1, \dots, \rho_N)} \varepsilon \text{KL}^N(\gamma | \mathbf{m}_1 \otimes \dots \otimes \mathbf{m}_N).$$

Analogously to the 2 marginal case, the duality results, existence and regularity of entropic-potentials as well as the convergence of the Sinkhorn algorithm can be extended to that case. We omit the details here since the proof follows by similar arguments.

**Acknowledgements** S. D. M. is member of *Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni* (GNAMPA) of the *Istituto Nazionale di Alta Matematica* (INdAM). A.G. acknowledges the European Research Council under H2020/MSCA-IF“OTmeetsDFT” [Grant ID: 795942]. This work started when the second author visited the first author, while he was working at the *Scuola Normale Superiore* di Pisa (INdAM unit). The authors want to thank G. Carlier, C. Léonard and L. Tamanini for useful discussions.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## 5 Appendix

**Proposition 5.1** *Let  $(X, d, \mu)$  be a measurable metric space with  $\mu(X) = 1$ . Let us assume that  $\mathcal{F} \subset L^p(X, \mu)$  is a family of functions such that:*

- (a) *there exists  $M > 0$  such that  $\|f\|_\infty \leq M$  for every  $f \in \mathcal{F}$ ;*
- (b) *for every  $\sigma$  there exists a set  $N^\sigma$ , a modulus of continuity  $\omega_\sigma$  and a number  $\beta_\sigma \geq 0$  such that*

$$|f(x) - f(x')| \leq \omega_\sigma(d(x, x')) + \beta_\sigma \quad \forall x, x' \notin N^\sigma$$

*where  $N^\sigma$  and  $\beta_\sigma$  are such that  $\mu(N^\sigma) + \beta_\sigma \rightarrow 0$  as  $\sigma \rightarrow 0$ .*

*Then the family  $\mathcal{F}$  is precompact in  $L^p(X, \mu)$ .*

**Proof** Let us fix  $\varepsilon > 0$  and let us consider a sequence  $\sigma_n \rightarrow 0$  such that  $\sum_{n=1}^\infty \mu(N^{\sigma_n}) \leq \varepsilon$ ; then define  $\omega_n = \omega_{\sigma_n}$  and  $\mathcal{N}^\varepsilon := \bigcup_n N^{\sigma_n}$ ; in particular we have  $\mu(\mathcal{N}^\varepsilon) \leq \varepsilon$  and

$$|f(x) - f(x')| \leq \omega_n(d(x, x')) + \beta_{\sigma_n} \quad \forall x, x' \notin \mathcal{N}^\varepsilon, \forall n \in \mathbb{N}. \tag{5.1}$$

Let us define  $\omega^\varepsilon(t) = \inf_n \{\omega_n(t) + \beta_{\sigma_n}\}$ : by (5.1) we have that  $f$  is  $\omega^\varepsilon$ -continuous outside  $\mathcal{N}^\varepsilon$ .

We can verify that  $\omega^\varepsilon$  is a non degenerate modulus of continuity: it is obvious that it is nondecreasing since it is an infimum of nondecreasing functions. Then for every  $\tilde{\varepsilon} > 0$  we can choose  $n$  big enough such that  $\beta_{\sigma_n} < \tilde{\varepsilon}/2$  and then choose  $t$  small enough such that  $\omega_n(t) < \tilde{\varepsilon}/2$ ; in this way we have  $\omega^\varepsilon(t) \leq \omega_n(t) + \beta_{\sigma_n} < \tilde{\varepsilon}$ . In particular  $\omega^\varepsilon(t) \rightarrow 0$  as  $t \rightarrow 0$ .

Now we conclude by a diagonal argument: let us consider a sequence  $(f_n^0)_{n \in \mathbb{N}} \subseteq \mathcal{F}$  and a sequence  $\varepsilon_k \rightarrow 0$ . We want to find a subsequence that is converging strongly in  $L^p$ . We iteratively extract a subsequence  $(f_n^k)$  of  $(f_n^{k-1})$  that is converging uniformly outside  $\mathcal{N}^{\varepsilon_k}$  (thanks to Ascoli-Arzelà) to some function  $f^k$ , which is defined only outside  $\mathcal{N}^{\varepsilon_k}$ . Then let us consider

$$f(x) = \begin{cases} f^k(x) & \text{if } x \notin \mathcal{N}^{\varepsilon_k} \\ 0 & \text{otherwise.} \end{cases}$$

First of all  $f$  is well defined since if  $x \notin \mathcal{N}^{\varepsilon_k}$  and  $x \notin \mathcal{N}^{\varepsilon_j}$  with  $j > k$  then we have that  $f_n^k(x) \rightarrow f^k(x)$  but since  $f_n^j$  is a subsequence of  $f_n^k$  we have also  $f_n^j(x) \rightarrow f^k(x)$ ; however by definition  $f_n^j(x) \rightarrow f^j(x)$  and so  $f^j(x) = f^k(x)$ . Moreover it is clear that  $\|f\|_\infty \leq M$  since this is true for every  $f_n^0$  thanks to property (a). Now we consider the sequence  $g_n = f_n^n$  which is a subsequence of  $f_n^0$ . Let us fix  $\varepsilon > 0$  and choose  $k$  such that  $\varepsilon_k < \varepsilon^p$ ; then let  $n_0 > k$  such that  $|f_n^k - f| \leq \varepsilon$  on  $X \setminus \mathcal{N}^{\varepsilon_k}$  for every  $n \geq n_0$ . Now we have  $g_n = f_n^k$  for some  $n \geq n_0$  and in particular

$$\begin{aligned} \int_X |g_{n_0}(x) - f(x)|^p d\mu &= \int_X |f_n^k(x) - f(x)|^p d\mu \\ &= \int_{\mathcal{N}^{\varepsilon_k}} |f_n^k(x) - f(x)|^p d\mu + \int_{X \setminus \mathcal{N}^{\varepsilon_k}} |f_n^k(x) - f(x)|^p d\mu \\ &\leq \int_{\mathcal{N}^{\varepsilon_k}} (2M)^p d\mu + \int_{X \setminus \mathcal{N}^{\varepsilon_k}} \varepsilon^p d\mu \\ &\leq \mu(\mathcal{N}^{\varepsilon_k})(2M)^p + \varepsilon^p \\ &\leq \varepsilon_k(2M)^p + \varepsilon^p \mu(X) \leq \varepsilon^p(2^p M^p + 1) \end{aligned}$$

In particular we get  $g_n \rightarrow f$  in  $L^p$  and so we're done. □

## References

1. Agueh, M., Carlier, G.: Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43**, 904–924 (2011)
2. Altschuler, J., Weed, J., Rigollet, P.: Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In: *Advances in Neural Information Processing Systems*, pp. 1964–1974 (2017)
3. Arnaudon, M., Cruzeiro, A. B., Léonard, C., Zambrini, J.-C.: An entropic interpolation problem for incompressible viscid fluids. *arXiv preprint arXiv:1704.02126* (2017)
4. Bacharach, M.: Estimating nonnegative matrices from marginal data. *Int. Econ. Rev.* **6**, 294–310 (1965)
5. Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., Peyré, G.: Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.* **37**, A1111–A1138 (2015)
6. Benamou, J.-D., Carlier, G., Nenna, L.: Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm. *Numerische Mathematik* **142**, 1–22 (2017)
7. Borwein, J.M., Lewis, A.S.: Decomposition of multivariate functions. *Can. J. Math.* **44**, 463–482 (1992)
8. Borwein, J.M., Lewis, A.S., Nussbaum, R.D.: Entropy minimization, DAD problems, and doubly stochastic kernels. *J. Funct. Anal.* **123**, 264–307 (1994)
9. Brenier, Y.: The least action principle and the related concept of generalized flows for incompressible perfect fluids. *J. Am. Math. Soc.* **2**, 225–255 (1989)
10. Brenier, Y.: The dual least action problem for an ideal, incompressible fluid. *Arch. Ration. Mech. Anal.* **122**, 323–351 (1993)
11. Brenier, Y.: Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Commun. Pure Appl. Math. A J. Issued Courant Inst. Math. Sci.* **52**, 411–452 (1999)
12. Caffarelli, L.A.: Monotonicity properties of optimal transportation and the fkg and related inequalities. *Commun. Math. Phys.* **214**, 547–563 (2000)
13. Carlen, E.: Stochastic mechanics: a look back and a look ahead. *Diffus. Quantum Theory Radic. Eleme. Math.* **47**, 117–139 (2014)

14. Carlen, E.A.: Conservative diffusions. *Commun. Math. Phys.* **94**, 293–315 (1984)
15. Carlen, E. A.: Existence and sample path properties of the diffusions in Nelson’s stochastic mechanics. In: *Stochastic Processes: Mathematics and Physics*, Springer, Berlin, pp. 25–51 (1986)
16. Carlier, G., Duval, V., Peyré, G., Schmitzer, B.: Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.* **49**, 1385–1418 (2017)
17. Carlier, G., Ekeland, I.: Matching for teams. *Econom. Theory* **42**, 397–418 (2010)
18. Carlier, G., Laborde, M.: A differential approach to the multi-marginal Schrödinger system. *SIAM J. Math. Anal.* **52**(1), 709–717 (2020)
19. Cattiaux, P., Léonard, C.: Minimization of the Kullback information of diffusion processes. *Ann. l’IHP Probabilités et statistiques* **30**, 83–132 (1994)
20. Chen, Y., Conforti, G., Georgiou, T. T., Ripani, L.: Multi-marginal Schrödinger bridges. In: *International Conference on Geometric Science of Information*, pp. 725–732. Springer (2019)
21. Chen, Y., Georgiou, T., Pavon, M.: Entropic and displacement interpolation: a computational approach using the Hilbert metric. *SIAM J. Appl. Math.* **76**, 2375–2396 (2016)
22. Chiappori, P.-A., McCann, R.J., Nesheim, L.P.: Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Econ. Theor.* **42**, 317–354 (2010)
23. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: Scaling algorithms for unbalanced optimal transport problems. *Math. Comput.* **87**, 2563–2609 (2018)
24. Cruzeiro, A.B., Zambrini, J.-C.: Malliavin calculus and Euclidean quantum mechanics. I. Functional calculus. *J. Funct. Anal.* **96**, 62–95 (1991)
25. Csiszár, I.: I-divergence geometry of probability distributions and minimization problems. In: *The Annals of Probability*, pp. 146–158 (1975)
26. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: *Advances in Neural Information Processing Systems*, pp. 2292–2300 (2013)
27. Cuturi, M., Doucet, A.: Fast computation of Wasserstein barycenters. In: *International Conference on Machine Learning*, pp. 685–693 (2014)
28. Cuturi, M., Peyré, G.: Computational optimal transport, foundations and trends®. *Mach. Learn.* **11**, 355–607 (2019)
29. Deming, W.E., Stephan, F.F.: On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Stat.* **11**, 427–444 (1940)
30. Di Marino, S., Gerolin, A., Giesbertz, K., Nenna, L., Seidl, M., Gori-Giorgi, P.: The strictly-correlated electron functional for spherically symmetric systems revisited (in preparation) (2016)
31. Dudley, R.M.: The speed of mean Glivenko–Cantelli convergence. *Ann. Math. Stat.* **40**, 40–50 (1969)
32. Fathi, M., Gozlan, N., Prodhomme, M.: A proof of the Caffarelli contraction theorem via entropic regularization. *arXiv preprint arXiv:1904.06053* (2019)
33. Fényes, I.: A deduction of Schrödinger equation. *Acta Bolyaina* **1** (1946)
34. Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-I., Trounev, A., Peyré, G.: Interpolating between optimal transport and MMD using Sinkhorn divergences. In: *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2681–2690 (2019)
35. Flamary, R., Courty, N.: *POT Python Optimal Transport library* (2017)
36. Franklin, J., Lorenz, J.: On the scaling of multidimensional matrices. *Linear Algebra Appl.* **114**, 717–735 (1989)
37. Galichon, A., Salanié, B.: Matching with trade-offs: revealed preferences over competing characteristics. CEPR Discussion Paper No. DP7858 (2010)
38. Gangbo, W., Swiech, A.: Optimal maps for the multidimensional Monge–Kantorovich problem. *Commun. Pure Appl. Math.* **51**, 23–45 (1998)
39. Genevay, A., Chizat, L., Bach, F., Cuturi, M., Peyré, G.: Sample complexity of Sinkhorn divergences. In: *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1574–1583 (2019)
40. Genevay, A., Peyré, G., Cuturi, M.: Learning Generative models with Sinkhorn divergences. In: *International Conference on Artificial Intelligence and Statistics*, pp. 1608–1617 (2018)
41. Gentil, I., Léonard, C., Ripani, L., Tamanini, L.: An entropic interpolation proof of the HWI inequality. *Stochas. Process. Appl.* **130**(2), 907–923 (2018)
42. Gerolin, A., Grossi, J., Gori-Giorgi, P.: Kinetic correlation functionals from the entropic regularization of the strictly correlated electrons problem. *J. Chem. Theor. Comput.* **16**(1), 488–498 (2019)
43. Gerolin, A., Kausamo, A., Rajala, T.: Multi-marginal entropy-transport with repulsive cost. *Calc. Var.* **59**, 90 (2020)
44. Gigli, N., Tamanini, L.: Benamou–Brenier and duality formulas for the entropic cost on  $RCD^*(K, N)$  spaces. In: *Probability, Theory Related Fields* (2018)
45. Gigli, N., Tamanini, L.: Second order differentiation formula on  $RCD^*(K, N)$  spaces. *J. Eur. Math. Soc. (JEMS)* (2018)

46. Gori-Giorgi, P., Seidl, M., Vignale, G.: Density-functional theory for strongly interacting electrons. *Phys. Rev. Lett.* **103**, 166402 (2009)
47. Gozlan, N., Léonard, C.: A large deviation approach to some transportation cost inequalities. *Probab. Theory Relat. Fields* **139**, 235–283 (2007)
48. Idel, M.: A review of matrix scaling and sinkhorn's normal form for matrices and positive maps. arXiv preprint [arXiv:1609.06349](https://arxiv.org/abs/1609.06349) (2016)
49. Karlsson, J., Ringh, A.: Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport. *SIAM J. Imaging Sci.* **10**, 1935–1962 (2017)
50. Kruithof, J.: Telefoonverkeersrekening. *De Ingenieur* **52**, 15–25 (1937)
51. Léger, F., Li, W.: Hopf-cole transformation via generalized Schrödinger bridge problem. arXiv preprint [arXiv:1901.09051](https://arxiv.org/abs/1901.09051) (2019)
52. Léonard, C.: From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262**, 1879–1920 (2012)
53. Léonard, C.: A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst. A* **34**, 1533–1574 (2014)
54. Li, W., Lu, J., Wang, L.: Fisher information regularization schemes for Wasserstein gradient flows. *J. Comput. Phys.* **416**, 109449 (2020)
55. Li, W., Yin, P., Osher, S.: Computations of optimal transport distance with Fisher information regularization. *J. Sci. Comput.* **75**, 1581–1595 (2018)
56. Luise, G., Rudi, A., Pontil, M., Ciliberto, C.: Differential properties of Sinkhorn approximation for learning with Wasserstein distance. In: *Advances in Neural Information Processing Systems*, pp. 5859–5870 (2018)
57. Luise, G., Salzo, S., Pontil, M., Ciliberto, C.: Sinkhorn barycenters with free support via Frank-Wolfe algorithm. In: *Advances in Neural Information Processing Systems*, pp. 9322–9333 (2019)
58. Mikami, T.: Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Relat. Fields* **129**, 245–260 (2004)
59. Nelson, E.: Derivation of the Schrödinger equation from Newtonian mechanics. *Phys. Rev.* **150**, 1079 (1966)
60. Nelson, E.: *Dynamical Theories of Brownian Motion*, vol. 3. Princeton University Press, Princeton (1967)
61. Nelson, E.: *Quantum Fluctuations*. Princeton University Press, Princeton (1985)
62. Pavon, M., Tabak, E. G., Trigila, G.: The data-driven Schroedinger bridge. arXiv preprint [arXiv:1806.01364](https://arxiv.org/abs/1806.01364) (2018)
63. Rabin, J., Peyré, G., Delon, J., Berton, M.: Wasserstein Barycenter and its application to texture mixing. In: *International Conference on Scale Space and Variational Methods in Computer Vision*, Springer, pp. 435–446 (2011)
64. Raghavan, T.: On pairs of multidimensional matrices. *Linear Algebra Appl.* **62**, 263–268 (1984)
65. Ruschendorf, L.: Convergence of the iterative proportional fitting procedure. *Ann. Stat.* **23**, 1160–1174 (1995)
66. Schrödinger, E.: *Über die umkehrung der naturgesetze*. Verlag Akademie der wissenschaften in kommission bei Walter de Gruyter u, Company (1931)
67. Sinkhorn, R.: A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Stat.* **35**, 876–879 (1964)
68. Solomon, J., De Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., Guibas, L.: Convolutional Wasserstein distances: efficient optimal transportation on geometric domains. *ACM Trans. Gr. (TOG)* **34**, 66 (2015)
69. Weed, J., Bach, F.: Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25**, 2620–2648 (2019)
70. Yule, G.U.: On the methods of measuring association between two attributes. *J. R. Stat. Soc.* **75**, 579–652 (1912)
71. Zambrini, J.: Stochastic mechanics according to E. Schrödinger *Phys. Rev. A* **33**, 1532 (1986)
72. Zambrini, J.-C.: Variational processes and stochastic versions of mechanics. *J. Math. Phys.* **27**, 2307–2330 (1986)
73. Zambrini, J.-C.: The research program of stochastic deformation (with a view toward geometric mechanics). In: *Stochastic Analysis: A Series of Lectures*, Springer, pp. 359–393 (2015)