On the linear convergence of the multi-marginal Sinkhorn algorithm

Guillaume Carlier*

March 22, 2021

Abstract

The aim of this short note is to give an elementary proof of linear convergence of the Sinkhorn algorithm for the entropic regularization of multi-marginal optimal transport. The proof simply relies on: i) the fact that Sinkhorn iterates are bounded, ii) strong convexity of the exponential on bounded intervals and iii) the convergence analysis of the coordinate descent (Gauss-Seidel) method of Beck and Tetruashvili [1].

Keywords: multi-marginal entropic optimal transport, Sinkhorn algorithm, linear convergence, block coordinate descent.

MS Classification: 45G15, 49M05.

1 Introduction

Eventhough the Sinkhorn algorithm[1] is more than 50 years old [16], it has attracted a considerable attention in the last years. It is now at the heart of efficient solvers for the entropic regularization of optimal transport problems, a field on which Cuturi's paper [6] had a tremendous impact (see Cuturi and Doucet [5], Cuturi and Peyré [13], Benamou et al. [2]...). The Sinkhorn algorithm remains fascinating by its simplicity and its connections with the Schrödinger bridge problem first addressed by Schrödinger in [15] and large deviations theory, see Dawson and Gärtner [7], Föllmer [9], Léonard [12, 11].

*CEREMADE, Université Paris Dauphine, PSL and INRIA-Paris, MOKAPLAN, carlier@ceremade.dauphine.fr

[1] also known as *iterative proportional fitting procedure* (IPFP) in the probability and statistics literature, see Rüschendorf, [14] and the references therein.

The linear convergence of the Sinhkorn algorithm for two marginals is well-known. A very elegant proof consists in using a celebrated theorem of Birkhoff to show that the Sinkhorn algorithm consists in iterating a contraction for the Hilbert projective metric, see Franklin and Lorenz [10], and more recently, Chen, Georgiou and Pavon, [4].

To the best of our knowledge, the elegant Hilbert metric proof does not carry over to the multi-marginal case for which an annoying $N-1$ factor (N being the number of marginals) appears in the Lipschitz constant for the Hilbert metric. Convergence of the multi-marginal Sinkhorn algorithm was recently obtained by Di Marino and Gerolin [8] and the well-posedness (existence, uniqueness and smooth dependence on the data) of the Schrödinger system (see (2.3) below) was addressed by completely different arguments (local and global inversion theorems) by the author and Laborde in [3]. In the analysis of [8], a key ingredient is that Sinkhorn iterates are coordinate descent updates for a convex minimization problem (dual to an entropy minimization subject to multi-marginal constraints), see the definition of F in (2.4) below. In this note, we slightly improve the results of Di Marino and Gerolin, by showing linear convergence. The proof relies on the convergence analysis of the coordinate descent method of Beck and Tetruashvili [1] which can easily be used here, since Sinkhorn iterates are bounded in L^∞ so remain in a set where the functional F is uniformly convex.

2 Multi-marginal Sinkhorn algorithm

We are given an integer $N \geq 2$, N probability spaces $(X_i, \mathcal{F}_i, m_i)$, $i = 1, \dots, N$ and set

$$X := \prod_{i=1}^N X_i, \mathcal{F} := \bigotimes_{i=1}^N \mathcal{F}_i, \ m := \bigotimes_{i=1}^N m_i. \tag{2.1}$$

Given $i \in \{1, \dots, N\}$, we will denote by $X_{-i} := \prod_{j\neq i}^N X_j$, $m_{-i} := \bigotimes_{j\neq i}^N m_j$ and will always identify X to $X_i \times X_{-i}$ i.e. will denote $x = (x_1, \dots, x_N) \in X$ as $x = (x_i, x_{-i})$. The set of measures on $(X, \mathcal{F})$ having $m_1, \dots, m_N$ as marginals will be denoted $\Pi(m_1, \dots, m_N)$. Given $p \in [1, \infty]$ and $(\varphi_1, \dots, \varphi_N) \in \prod_{i=1}^N L^p(X_i, \mathcal{F}_i, m_i)$ we will use the notations

$$\oplus_{i=1}^N \varphi_i \ : \ (x_1, \dots, x_N) \in X \mapsto \sum_{i=1}^N \varphi_i(x_i),$$

and

$$\otimes_{i=1}^N \varphi_i \; : \; (x_1, \dots, x_N) \in X \mapsto \prod_{i=1}^N \varphi_i(x_i).$$

Given a cost $c \in L^\infty(X, \mathcal{F}, m)$, we set

$$\|c\|_\infty := \|c\|_{L^\infty(X,\mathcal{F},m)}. \tag{2.2}$$

The associated Gibbs kernel is

$$K := e^{-c},$$

so that $e^{-\|c\|_\infty} \leq K \leq e^{\|c\|_\infty}$, m-almost everywhere. We look for potentials $\varphi = (\varphi_1, \dots, \varphi_N) \in \prod_{i=1}^N L^\infty(X_i, \mathcal{F}_i, m_i)$ such that the measure

$$Q_\varphi := K e^{\oplus_{i=1}^N \varphi_i} m$$

belongs to $\Pi(m_1, \dots, m_N)$ i.e. solve the Schrödinger system:

$$e^{\varphi_i(x_i)} \int_{X_{-i}} e^{-c(x_1,\dots,x_N) + \sum_{j\neq i} \varphi_j(x_j)} \mathrm{d}m_{-i}(x_{-i}) = 1, \tag{2.3}$$

for every i and m_i-a.e. x_i. The system (2.3) is well-known to be the Euler-Lagrange optimality condition for the convex minimization problem

$$\inf_{\varphi \in \prod_{i=1}^N L^\infty(X_i,\mathcal{F}_i,m_i)} F(\varphi) := -\sum_{i=1}^N \int_{X_i} \varphi_i \mathrm{d}m_i + \int_X \mathrm{d}Q_\varphi \tag{2.4}$$

and if φ solves (2.3), the measure Q_φ solves the multi-marginal entropy minimization:

$$\inf_{Q \in \Pi(m_1,\dots,m_N)} H(Q | e^{-c} m).$$

Let us observe that whenever $\lambda_1, \dots, \lambda_N$ are constants which sum to 0, then

$$F(\varphi_1 + \lambda_1, \dots, \varphi_N + \lambda_N) = F(\varphi_1, \dots, \varphi_N)$$

so that one can impose the $N-1$ normalizing constraints:

$$\int_{X_1} \varphi_1 \mathrm{d}m_1 = \ldots = \int_{X_{N-1}} \varphi_{N-1} \mathrm{d}m_{N-1} = 0. \tag{2.5}$$

Denoting by $L^p_\diamond(X_i, \mathcal{F}_i, m_i)$ the space of zero-mean L^p potentials:

$$L^p_\diamond(X_i, \mathcal{F}_i, m_i) := \{\varphi_i \in L^p(X_i, \mathcal{F}_i, m_i) \; : \; \int_{X_i} \varphi_i \mathrm{d}m_i = 0\}$$

we thus consider

$$\inf_{\varphi\in E} F(\varphi) \text{ where } E := \prod_{i=1}^{N-1} L^\infty_\diamond(X_i,\mathcal{F}_i,m_i)\times L^\infty(X_N,\mathcal{F}_N,m_N). \tag{2.6}$$

The Sinkhorn algorithm is nothing but block coordinate descent for the minimization of F over E. Starting from $\varphi^0\in E$, the updates of the Sinkhorn algorithm, consists, given $\varphi^t=(\varphi_1^t,\dots,\varphi_N^t)\in E$, in:

$$\varphi_1^{t+1} := \operatorname{argmin}_{\varphi_1\in L^\infty_\diamond(X_1,\mathcal{F}_1,m_1)} F(\varphi_1,\varphi_2^t,\dots,\varphi_2^t) \tag{2.7}$$

i.e.

$$\varphi_1^{t+1}(x_1) := -\log\Big(\int_{X_{-1}} e^{\sum_{j=2}^N \varphi_j^t(x_j)} K(x_1,x_{-1})\mathrm{d}m_{-1}(x_{-1})\Big)+\lambda_1^t,\ \forall x_1\in X_1 \tag{2.8}$$

where

$$\lambda_1^t = \int_{X_1}\Big(\log\Big(\int_{X_{-1}} e^{\oplus_{j=2}^N\varphi_j^t} K(x_1,x_{-1})\mathrm{d}m_{-1}(x_{-1})\Big)\Big)\mathrm{d}m_1(x_1). \tag{2.9}$$

Then, for $i=2,\dots,N-1$,

$$\varphi_i^{t+1} := \operatorname{argmin}_{\varphi_i\in L^\infty_\diamond(X_i,\mathcal{F}_i,m_i)} F(\varphi_1^{t+1},\dots,\varphi_{i-1}^{t+1},\varphi_i,\varphi_{i+1}^t,\dots,\varphi_N^t) \tag{2.10}$$

i.e.

$$\varphi_i^{t+1}(x_i) := -\log\Big(\int_{X_{-i}} e^{\oplus_{j=1}^{i-1}\varphi_j^{t+1}\ \oplus_{j=i+1}^N\varphi_j^t} K(x_i,x_{-i})\mathrm{d}m_{-i}(x_{-i})\Big)+\lambda_i^t,\ \forall x_i\in X_i \tag{2.11}$$

where

$$\lambda_i^t = \int_{X_i}\Big(\log\Big(\int_{X_{-i}} e^{\oplus_{j=1}^{i-1}\varphi_j^{t+1}\ \oplus_{j=i+1}^N\varphi_j^t} K(x_i,x_{-i})\mathrm{d}m_{-i}(x_{-i})\Big)\Big)\mathrm{d}m_i(x_i). \tag{2.12}$$

Finally, for $i=N$,

$$\varphi_N^{t+1} := \operatorname{argmin}_{\varphi_N\in L^\infty(X_N,\mathcal{F}_N,m_N)} F(\varphi_1^{t+1},\dots,\varphi_{N-1}^{t+1},\dots,\varphi_N) \tag{2.13}$$

i.e.

$$\varphi_N^{t+1}(x_N) := -\log\Big(\int_{X_{-N}} e^{\oplus_{j=1}^{N-1}\varphi_j^{t+1}} K(x_N,x_{-N})\mathrm{d}m_{-N}(x_{-N})\Big),\ \forall x_N\in X_N. \tag{2.14}$$

The convergence of the Sinkhorn iterates to a solution of (2.3) (hence a minimizer of (2.4)) was established by Di Marino and Gerolin [8]. The aim of the next paragraph is to slightly improve this result by showing that this convergence is linear.

3 Linear convergence

Thanks to the normalization (2.5), arguing as in [8], we have uniform bounds on the Sinkhorn iterates:

Lemma 3.1. *For every $t \geq 1$, the Sinkhorn iterates φ^t satisfy the bounds:*

$$\|\varphi_i^t\|_{L^\infty(X_i,\mathcal{F}_i,m_i)} \leq 2\|c\|_\infty, i = 1, \dots, N-1, \tag{3.1}$$

$$\|\varphi_N^t\|_{L^\infty(X_N,\mathcal{F}_N,m_N)} \leq (2N-1)\|c\|_\infty. \tag{3.2}$$

Proof. Since for $m_i \otimes m_i$-a.e. (x_i, y_i) and m_{-i}-a.e. $x_{-i} \in X_{-i}$, one has

$$c(y_i, x_{-i}) \geq c(x_i, x_{-i}) - 2\|c\|_\infty$$

we deduce from (2.8) and (2.11) that for $i = 1, \dots, N-1$, $\varphi_i^t(x_i) - \varphi_i^t(y_i) \leq 2\|c\|_\infty$, using the fact that $\varphi_i^t \in L^\infty_\diamond(X_i, \mathcal{F}_i, m_i)$ and integrating the previous inequality, we immediately deduce (3.1). Once we have these bounds on φ_i^t, for $i = 1, \dots, N-1$, using (2.14), together with $e^{-\|c\|_\infty} \leq K \leq e^{\|c\|_\infty}$, we deduce (3.2). □

Now that we have uniform, bounds on φ^t, we can take advantage of the strong convexity and Lipschitz continuity of the exponential function on bounded intervals, to use the analysis of Beck and Tetruashvili [1]. Indeed, given $M > 0$, one obviously has, $\forall (a, b) \in [-M, M]^2$:

$$e^b - e^a - e^a(b-a) \geq \frac{e^{-M}}{2}(b-a)^2, \; |e^b - e^a| \leq e^M |b-a|. \tag{3.3}$$

Lemma 3.2. *Defining*

$$\nu = e^{-(4N-2)\|c\|_\infty}, \tag{3.4}$$

one has

$$F(\varphi^t) - F(\varphi^{t+1}) \geq \frac{\nu}{2} \sum_{i=1}^N \|\varphi_i^t - \varphi_i^{t+1}\|^2_{L^2(X_i,\mathcal{F}_i,m_i)}. \tag{3.5}$$

Proof. Define

$$\widetilde{\varphi}_i^t := (\varphi_1^{t+1}, \dots, \varphi_i^{t+1}, \varphi_{i+1}^t, \dots, \varphi_N^t), \; i = 1, \dots, N-1, \; \widetilde{\varphi}_N^t := \varphi^{t+1}, \tag{3.6}$$

and write in a telescopic fashion

$$F(\varphi^t) - F(\varphi^{t+1}) = F(\varphi^t) - F(\widetilde{\varphi}_1^t) + \sum_{i=1}^{N-1} (F(\widetilde{\varphi}_i^t) - F(\widetilde{\varphi}_{i+1}^t)).$$

Using successively, the first basic inequality in (3.3), (2.8), the fact that $\varphi_1^t - \varphi_1^{t+1}$ has zero mean against m_1, and the bounds from lemma 3.1, we get

$$\begin{aligned}
F(\varphi^t) - F(\widetilde{\varphi}_1^t) &= \int_X (e^{\varphi_1^t(x_1)} - e^{\varphi_1^{t+1}(x_1)}) \prod_{j=2}^N e^{\varphi_j^t(x_j)} e^{-c(x)} \mathrm{d}m(x) \\
&\geq \int_X (\varphi_1^t(x_1) - \varphi_1^{t+1}(x_1)) e^{\varphi_1^{t+1}(x_1)} \prod_{j=2}^N e^{\varphi_j^t(x_j)} e^{-c(x)} \mathrm{d}m(x) \\
&+ \frac{e^{-2\|c\|_\infty}}{2} \int_X (\varphi_1^t(x_1) - \varphi_1^{t+1}(x_1))^2 \prod_{j=2}^N e^{\varphi_j^t(x_j)} e^{-c(x)} \mathrm{d}m(x) \\
&\geq e^{\lambda_1^t} \int_{X_1} (\varphi_1^t(x_1) - \varphi_1^{t+1}(x_1)) \mathrm{d}m_1(x_1) \\
&+ \frac{e^{-(4N-2)\|c\|_\infty}}{2} \int_{X_1} (\varphi_1^t(x_1) - \varphi_1^{t+1}(x_1))^2 \mathrm{d}m_1(x_1) \\
&= \frac{e^{-(4N-2)\|c\|_\infty}}{2} \int_{X_1} (\varphi_1^t - \varphi_1^{t+1})^2 \mathrm{d}m_1.
\end{aligned}$$

Similarly, for $i = 1, \ldots, N-1$, we have

$$F(\widetilde{\varphi}_i^t) - F(\widetilde{\varphi}_{i+1}^t) \geq \frac{e^{-(4N-2)\|c\|_\infty}}{2} \int_{X_{i+1}} (\varphi_{i+1}^t - \varphi_{i+1}^{t+1})^2 \mathrm{d}m_{i+1}$$

which shows (3.5). □

Since F is bounded from below on E, the left-hand side of (3.5) converges to 0. Note also that since φ^t and φ^{t+1} belong to E, one has the identity

$$\| \oplus_{i=1}^N (\varphi^{t+1} - \varphi^t) \|^2_{L^2(X,\mathcal{F},m)} = \sum_{i=1}^N \|\varphi_i^t - \varphi_i^{t+1}\|^2_{L^2(X_i,\mathcal{F}_i,m_i)} \tag{3.7}$$

and we deduce from (3.5)

$$\lim_{t\to+\infty} \| \oplus_{i=1}^N (\varphi^{t+1} - \varphi^t) \|^2_{L^2(X,\mathcal{F},m)} = 0. \tag{3.8}$$

Together with the uniform bounds from lemma 3.1, we deduce that $\varphi_i^t - \varphi_i^{t+1}$ as well as $e^{\varphi_i^t} - e^{\varphi_i^{t+1}}$ converge strongly to 0 in $L^p(m_i) = L^p(X_i, \mathcal{F}_i, m_i)$ for every $p \in [1, +\infty)$.

Theorem 3.3. *The sequence of Sinkhorn iterates φ^t converges strongly in $L^p_\diamond(X_1, \mathcal{F}_1, m_1) \times \ldots \times L^p_\diamond(X_{N-1}, \mathcal{F}_{N-1}, m_{N-1}) \times L^p(X_N, \mathcal{F}_N, m_N)$ for every $p \in [1, +\infty)$, to the unique solution $\overline{\varphi}$ of (2.6). Moreover, there holds*

$$F(\varphi^t) - F(\overline{\varphi}) \leq \Big(1 - \frac{e^{-(16N-8)\|c\|_\infty}}{N}\Big)^t (F(\varphi^0) - F(\overline{\varphi})). \tag{3.9}$$

Proof. The convergence of φ^t in every L^p was obtained by Di Marino and Gerolin [8], we include a short proof for the sake of completeness. Setting $a_i^t := e^{\varphi_i^t}$, passing to a subsequence if necessary, we may assume the constants λ_i^t in (2.9)-(2.11) converge and that a_i^t, a_i^{t+1} converges weakly to some a_i in $L^2(m_i)$. Hence, for every i, $\otimes_{j<i} a_j^{t+1} \otimes_{j>i} a_j^t$ weakly converges in $L^2(m_{-i})$ to $\otimes_{j<i} a_j \otimes_{j>i} a_j$. By construction of the Sinkhorn iterates, $\frac{e^{\lambda_i^t}}{a_i^{t+1}}$ is expressed as a Hilbert-Schmidt hence compact integral functional of $\otimes_{j<i} a_j^{t+1} \otimes_{j>i} a_j^t$, hence $\frac{1}{a_i^{t+1}}$ converges strongly in $L^2(m_i)$. Since a_i^t is uniformly bounded and uniformly bounded away from 0, a_i^{t+1} converges strongly in $L^2(m_i)$ as well as in in $L^p(m_i)$ for any $p \in [1, +\infty)$ by the bounds from lemma 3.1. Using again that a_i^t is uniformly bounded and uniformly bounded away from 0, φ_i^t also strongly converges in $L^p(m_i)$ to $\overline{\varphi}_i := e^{a_i}$ and of course $\overline{\varphi} \in E$. Observing that, by construction, $\otimes_{j\leq i} a_j^{t+1} \otimes_{j>i} a_j^t Km$ admits $e^{\lambda_i^t} m_i$ as i-th marginal for $i = 1, \ldots, N-1$ and m_N as N-th marginal, one easily checks that $e^{\oplus_{i=1}^N \overline{\varphi}_i} Km = \otimes_{i=1}^N a_i Km$ admits $m_1, \ldots, m_N$ as marginals (and all the constants λ_i^t, $i = 1, \ldots, N-1$, tend to 0). Thus $\overline{\varphi}$ solves the system (2.3) hence minimizes F over E, but since F is strictly convex over E, this minimizer is unique and in fact the whole sequence φ^t strongly converges in L^p to $\overline{\varphi}$ for every $p \in [1, +\infty)$.

Since $\overline{\varphi}$ satisfies the bounds of lemma 3.1, using (3.3) as we did in the proof of lemma 3.2, we arrive at

$$F(\overline{\varphi}) - F(\varphi^t) \geq \sum_{i=1}^N \int_{X_i} \partial_i F(\varphi^t)(x_i)(\overline{\varphi}_i(x_i) - \varphi_i^t(x_i)) \mathrm{d}m_i(x_i) + \frac{\nu}{2} \sum_{i=1}^N \|\overline{\varphi}_i - \varphi_i^t\|^2_{L^2(m_i)}$$

where ν is the constant in (3.4) and

$$\partial_i F(\varphi)(x_i) = -1 + e^{\varphi_i(x_i)} \int_{X_{-i}} e^{\oplus_{j\neq i} \varphi_j(x_j)} e^{-c(x_i, x_{-i})} \mathrm{d}m_{-i}(x_{-i}).$$

Defining $\widetilde{\varphi}_i^t$ by (3.6), by construction of the Sinkhorn iterates, for $i = 1, \dots, N$, we have

$$\int_{X_i} \partial_i F(\widetilde{\varphi}_i^t)(\overline{\varphi}_i - \varphi_i^t) \mathrm{d}m_i = 0$$

hence

$$\begin{aligned} F(\overline{\varphi}) - F(\varphi^t) &\geq \sum_{i=1}^N \int_{X_i} (\partial_i F(\varphi^t) - \partial_i F(\widetilde{\varphi}_i^t))(\overline{\varphi}_i - \varphi_i^t) \mathrm{d}m_i \\ &+ \frac{\nu}{2} \sum_{i=1}^N \|\overline{\varphi}_i - \varphi_i^t\|^2_{L^2(m_i)} \\ &\geq -\frac{1}{2\nu} \sum_{i=1}^N \|\partial_i F(\varphi^t) - \partial_i F(\widetilde{\varphi}_i^t)\|^2_{L^2(m_i)} \end{aligned}$$

where we have used Young's inequality in the last line. We thus have shown that

$$F(\varphi^t) - F(\varphi) \leq \frac{1}{2\nu} \sum_{i=1}^N \|\partial_i F(\varphi^t) - \partial_i F(\widetilde{\varphi}_i^t)\|^2_{L^2(m_i)}. \tag{3.10}$$

Using the second inequality in (3.3) together with the L^∞ bounds on φ^t from lemma 3.1 and Jensen's inequality yield

$$(\partial_i F(\varphi^t)(x_i) - \partial_i F(\widetilde{\varphi}_i^t(x_i))^2 \leq \frac{1}{\nu^2} \int_{X_{-i}} (\oplus_{j=1}^N \varphi_j^t - \oplus_{j=1}^N (\widetilde{\varphi}_i^t)_j)^2 m_{-i}$$

so that

$$\begin{aligned} \|\partial_i F(\varphi^t) - \partial_i F(\widetilde{\varphi}_i^t)\|^2_{L^2(m_i)} &\leq \frac{1}{\nu^2} \sum_{j=1}^N \|\varphi_j^t - (\widetilde{\varphi}_i^t)_j\|^2_{L^2(m_j)} \\ &\leq \frac{1}{\nu^2} \sum_{j=1}^N \|\varphi_j^t - \varphi_j^{t+1}\|^2_{L^2(m_j)}, \end{aligned}$$

together with (3.10), we thus obtain

$$F(\varphi^t) - F(\overline{\varphi}) \leq \frac{N}{2\nu^3} \sum_{i=1}^N \|\varphi_i^t - \varphi_i^{t+1}\|^2_{L^2(m_i)}. \tag{3.11}$$

Finally, combining (3.11) with (3.5), we deduce

$$F(\varphi^t) - F(\overline{\varphi}) \leq \frac{N}{\nu^4}(F(\varphi^t) - F(\varphi^{t+1})) = \frac{N}{\nu^4}((F(\varphi^t) - F(\overline{\varphi})) - (F(\varphi^{t+1}) - F(\overline{\varphi})))$$

from which the linear convergence in (3.9) readily follows.

□

Remark 3.4. We also have linear convergence of φ^t to $\overline{\varphi}$ in L^2 and every L^p, $p \in [1, +\infty)$.

Acknowledgments: G.C. is grateful to the Agence Nationale de la Recherche for its support through the project MAGA (ANR-16-CE40-0014).

References

[1] Amir Beck and Luba Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM J. Optim.*, 23(4):2037–2060, 2013.

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

[3] Guillaume Carlier and Maxime Laborde. A differential approach to the multi-marginal Schrödinger system. *SIAM J. Math. Anal.*, 52(1):709–717, 2020.

[4] Yongxin Chen, Tryphon Georgiou, and Michele Pavon. Entropic and displacement interpolation: a computational approach using the Hilbert metric. *SIAM J. Appl. Math.*, 76(6):2375–2396, 2016.

[5] M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *Proceedings of the 31st International Conference on Machine Learning (ICML), JMLR W&CP*, volume 32, 2014.

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

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

[8] Simone Di Marino and Augusto Gerolin. An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *J. Sci. Comput.*, 85(2):Paper No. 27, 28, 2020.

[9] Hans Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, volume 1362 of *Lecture Notes in Math.*, pages 101–203. Springer, Berlin, 1988.

[10] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and its applications*, 114:717–735, 1989.

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

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

[13] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

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

[15] Erwin Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. In *Annales de l'institut Henri Poincaré*, volume 2, pages 269–310, 1932.

[16] R. Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *Amer. Math. Monthly*, 74:402–405, 1967.