Remarques sur le transport optimal entropique et l'algorithme de Sinkhorn dans le cas multi-marges

Guillaume Carlier [a]

SFO, 6 mai 2021 (en ligne). Collaboration avec Maxime Laborde.

[a]CEREMADE, Université Paris Dauphine et MOKAPLAN (Inria-Dauphine).

Plan

① Optimal and entropic transport

② Multi-marginal problems

③ The multi-marginal Schrödinger system is well-posed

④ Multi-marginal Sinkhorn converges linearly

Optimal and entropic optimal transport

Classical (two marginals) OT, the Monge-Kantorovich problem. Given X, Y (e.g. subsets of $\mathbf{R}^d$, or more generally Polish spaces) $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a cost c (continuous, say): $X \times Y \to \mathbf{R}$:

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times Y} c(x,y) \mathrm{d}\gamma(x,y) \tag{1}$$

where $\Pi(\mu,\nu)$ is the set of transport plans between μ and ν i.e. the set of probability measures on $X \times Y$ having μ and ν as marginals. Huge literature, Brenier solved the quadratic case, Gangbo McCann... Books by Villani, Santambrogio.

Kantorovich dual:

$$\sup\left\{\int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu \ : \ \varphi(x)+\psi(y) \leq c(x,y)\right\} \tag{2}$$

Given φ, the largest feasible ψ is the c-transform of ϕ:

$$\psi(y) = \varphi^c(y) = \inf_{x\in X}\{c(x,y)-\varphi(x)\}$$

Complementary slackness: optimal plans γ concentrate on pairs (x,y) for which $\varphi(x)+\varphi^c(y)=c(x,y)$. Brenier: $c(x,y)=\frac{1}{2}|x-y|^2$, γ is supported by the graph of the subdifferential of a convex potential, link with Monge-Ampère etc...

Entropic OT, regularization $\varepsilon > 0$

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times Y} c(x,y) \mathrm{d}\gamma(x,y) + \varepsilon H(\gamma | \mu \otimes \nu)$$

where when p and q are probability measures on Z, we denote

$$H(q|p) = \begin{cases} \int_Z (\log(m) - 1) \mathrm{d}q \text{ if } q = mp \\ +\infty \text{ otherwise} \end{cases}$$

Same as

$$\inf_{\gamma \in \Pi(\mu,\nu)} \varepsilon H(\gamma | e^{-\frac{c}{\varepsilon}} \mu \otimes \nu) \tag{3}$$

At least formally, writing Lagrange multipliers φ and ψ for the marginal constraints, the optimal plan takes the form

$$\gamma_\varepsilon = e^{\frac{\varphi \oplus \psi - c}{\varepsilon}} \mu \otimes \nu$$

(where $(\varphi \oplus \psi)(x, y) = \varphi(x) + \psi(y)$). Setting

$$K = e^{-\frac{c}{\varepsilon}}, a = e^{\frac{\varphi}{\varepsilon}},\ b = e^{\frac{\psi}{\varepsilon}}$$

K: Gibbs kernel, a and b Schrödinger potentials, γ_ε is given by

$$\gamma_\varepsilon = ((a \otimes b)K)\mu \otimes \nu.$$

With the mass conservation constraints: $\gamma_\varepsilon \in \Pi(\mu, \nu)$

Mass conservation gives two nonlinear integral equations for a and b (or φ and ψ)

$$1 = a(x) \int_Y K(x,y) b(y) \mathrm{d}\nu(y), \; 1 = b(y) \int_X K(x,y) a(x) \mathrm{d}\mu(x)$$

or

$$\varphi(x) = -\varepsilon \log \Big(\int_Y e^{\frac{1}{\varepsilon}(\psi(y) - c(x,y))} \mathrm{d}\nu(y) \Big)$$

$$\psi(y) = -\varepsilon \log \Big(\int_X e^{\frac{1}{\varepsilon}(\varphi(x) - c(x,y))} \mathrm{d}\mu(x) \Big)$$

Schrödinger system which is the Euler-Lagrange system for the dual of (3):

$$\sup_{\varphi,\psi} \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu - \varepsilon \int_{X \times Y} e^{\frac{\varphi \oplus \psi - c}{\varepsilon}} \mu \otimes \nu \tag{4}$$

N.B: obvious invariance $(\varphi + \lambda, \psi - \lambda)$ with $\lambda \in \mathbf{R}$ same as $(a/M, Mb)$.

Sinkhorn algorithm (or Gauss-Seidel):

$$a^{t+1}(x) = \frac{1}{\int_Y K(x,y) b^t(y) \mathrm{d}\nu(y)},$$

$$b^{t+1}(y) = \frac{1}{\int_X K(x,y) a^{t+1}(x) \mathrm{d}\mu(x)}$$

if $c \in L^\infty(\mu \otimes \nu)$, K bounded away from 0, linear convergence is well-known. Elegant proof using the so-called Hilbert projective metric and a theorem of Birkhoff (Franklin and Lorenz).

Note that Sinkhorn is also block coordinate descent in the dual (4), indeed fixing φ and maximizing the dual functional in ψ gives

$$\psi(y) = -\varepsilon \log\Big(\int_X e^{\frac{\varphi(x)-c(x,y)}{\varepsilon}} \mathrm{d}\mu(x)\Big)$$

(soft c-transform: this is Laplace method!). Huge literature, appears under different names and in different settings:

- Schrödinger system appears for the first time in the seminal paper of Schrödinger in 1931, raised the interest of Bernstein (1935), Beurling (1960)....
- large deviations, weakly interacting particle systems, stochastic control, Dawson and Gärtner (1987), Föllmer (1988), Mikami (2004), Léonard (2001, 2014)....

- matrix scaling DAD problem: Sinkhorn, Sinkhorn-Knopp (1967), Borwein, Lewis, Nussbaum (1990's)...
- Statistics, IPFP, Csiszàr (1975), Rüschendorf (1995),...
- use for computational OT is more recent: Cuturi's lightspeed paper (2013) had an important impact, Galichon and Salanié (2012), Cuturi and Peyré's book (2018), discrete case: nice stochastic interpretation of the dual (see Galichon's book, 2016).

Multi-marginal problems

These are problems of the form

$$\inf_{\gamma\in\Pi(\mu_1,\ldots,\mu_N)} \int_{X_1\times\ldots\times X_N} c(x_1,\ldots,x_N)\mathrm{d}\gamma(x_1,\ldots,x_N) \qquad (5)$$

Where $\Pi(\mu_1,\ldots,\mu_N)$ is the set of (multi-marginal plans) of probability measures on $\prod_{i=1}^N X_i$ having $\mu_1,\ldots,\mu_N$ as marginals. Motivations in physics (fluid dynamics, electronic correlation structure), economics and machine learning.

Multi-marginal OT has its roots in Brenier's relaxation (1989) of Arnol'd's (1966) interpretation of incompressible Euler as the equation of geodesics in the group of measure preserving diffeomorphisms. Time $t \in [0,1]$, flat torus $\mathbf{T^d}$:

$$\partial_t u + (u \cdot \nabla) u + \nabla p = 0, (t,x) \in [0,1] \times \mathbf{T^d},\ \mathrm{div}(u) = 0$$

flow $\partial_t X_t = u(t, X_t)$, $X_0 = \mathrm{id}$, measure preserving and at least formally is a critical point of $\int \|\dot{X}(t,.)\|_{L^2}^2 \mathrm{d}t$ over paths of measure preserving diffeomorphisms.

Setting $e_t(\omega) = \omega(t)$ evaluation map, flat torus $\mathbf{T^d}$, $\gamma \in \mathcal{P}(\mathbf{T^d} \times \mathbf{T^d})$ bistochastic (i.e. has uniform marginals, think of $\gamma = (\mathrm{id}, X_1)_\# \mathcal{L}^d$), GIF($\gamma$): generalized incompressible flows compatible with the joint distribution γ of particles at times 0 and 1

$$\mathrm{GIF}(\gamma) := \{P \in \mathcal{P}(C([0,1], \mathbf{T^d}) \ : \ e_{t\#}P = \mathcal{L}^d, \ (e_0, e_1)_\# P = \gamma\}$$

Brenier's formulation is an OT problem with infinitely many marginal constraints:

$$\inf_{P \in \mathrm{GIF}(\gamma)} \int_{C([0,1],\mathbf{T^d})} \int_0^1 |\dot{\omega}(t)|^2 \mathrm{d}t \mathrm{d}P(\omega)$$

Discretizing in time (and forgetting about the initial/terminal joint marginal constraint) leads to an optimal transport problem with a quadratic cost and several marginal constraints: Ganbo and Swiech. The Wasserstein barycenter problem fits in this category as well, gained some interest in machine learning. Multi-population matching problems in economics as well.

Symmetric and repulsive costs. Density functional theory is a very important field in computational chemistry, in the strictly correlated (or semi classical limit) regime it leads to

$$\inf_{\gamma\in\Pi(\rho,\dots,\rho)} \int_{\mathbf{R}^{3N}} \sum_{1\le i<j\le N} \frac{1}{|x_i - x_j|} \mathrm{d}\gamma(x_1,\dots,x_N)$$

where ρ is the density of one electron and γ represents the (symmetric) density of N electrons.

Buttazzo, de Pascale, Gori-Giorgi and Friesecke, Cotar and Klüppelberg at about the same time (2012-2013). Very active field, Champion, Colombo, Di Marino, Gerolin, Nenna, Pass.

Not so much is known about the structure of optimal plans in general with the notable exceptions of Pass (2011, 2012), Kim and Pass (2014). Tempting to approximate with entropic OT:

$$\inf_{\gamma\in\Pi(\mu_1,\ldots,\mu_N)} \varepsilon H(\gamma | e^{-\frac{c}{\varepsilon}} \otimes_{i=1}^N \mu_i) \tag{6}$$

look for $\gamma_\varepsilon = e^{\frac{\oplus_{i=1}^N \varphi_i - c}{\varepsilon}} \otimes_{i=1}^N \mu_i$ with potentials solving the multi-marginal Schrödinger system

$$1 = e^{\frac{\varphi_i(x_i)}{\varepsilon}} \int_{\prod_{j\neq i} X_j} e^{\frac{\oplus_{j\neq i}^N \varphi_i - c}{\varepsilon}} \otimes_{j\neq i} \mu_j. \tag{7}$$

Lots of things to say about the zero noise-limit $\varepsilon \to 0^+$ (Conforti, Léonard, Pal, Nutz, Tamanini). But for the rest of the talk $\varepsilon = 1$!

The Schrödinger system is well-posed

Joint work with Maxime Laborde. $N \geq 2$, N probability spaces $(X_i, \mathcal{F}_i, m_i)$, $i = 1, \dots, N$ and set

$$X := \prod_{i=1}^{N} X_i, \mathcal{F} := \bigotimes_{i=1}^{N} \mathcal{F}_i, \; m := \bigotimes_{i=1}^{N} m_i. \tag{8}$$

Given $i \in \{1, \dots, N\}$, denote by $X_{-i} := \prod_{j \neq i}^{N} X_j$, $m_{-i} := \bigotimes_{j \neq i}^{N} m_j$, identify X to $X_i \times X_{-i}$ i.e. will denote $x = (x_1, \dots, x_N) \in X$ as $x = (x_i, x_{-i})$.

$L^\infty_{++}(X_i, \mathcal{F}_i, m_i)$ (respectively $L^\infty_{++}(X, \mathcal{F}, m)$) the interior of the positive cone of $L^\infty(X_i, \mathcal{F}_i, m_i)$ (respectively $L^\infty(X, \mathcal{F}, m)$)

Kernel $K = e^{-c} \in L^\infty_{++}(X, \mathcal{F}, m)$ as well as densities $\mu_i \in L^\infty_{++}(X_i, \mathcal{F}_i, m_i)$ with the same total mass:

$$\int_{X_i} \mu_i \mathrm{d}m_i = \int_{X_j} \mu_j \mathrm{d}m_j,\ i, j \in \{1, \dots, N\}. \tag{9}$$

Aim: show the well-posedness of the multi-marginal Schrödinger system. Multi-marginal Schrödinger system: find potentials φ_i in $L^\infty(X_i, \mathcal{F}_i, m_i)$ such that for every i and m_i-almost every $x_i \in X_i$ one has:

$$\mu_i(x_i) = e^{\varphi_i(x_i)} \int_{X_{-i}} K(x_i, x_{-i}) e^{\sum_{j\neq i} \varphi_j(x_j)} \mathrm{d}m_{-i}(x_{-i}). \tag{10}$$

If $\varphi = (\varphi_1, \ldots, \varphi_N)$ solves (10) so does every family of potentials of the form $(\varphi_1 + \lambda_1, \ldots, \varphi_N + \lambda_N)$ where the λ_i's are constants with zero-sum. There are $N-1$ degrees of freedom, let us add to (10) the additional $N-1$ linear equations:

$$\int_{X_i} \varphi_i \mathrm{d}m_i = 0, \ i = 1, \ldots, N-1. \tag{11}$$

With this normalization, (10) has at most one solution, because (10) implies that φ solves

$$\sup_{\varphi=(\varphi_1,\ldots,\varphi_N)} \sum_{i=1}^{N} \int_{X_i} \varphi_i \mu_i \mathrm{d}m_i - \int_X K(x) e^{\sum_{j=1}^N \varphi_j(x_j)} \mathrm{d}m(x) \tag{12}$$

which is strictly concave in $\oplus_{i=1}^N \varphi_i$.

Define the Banach space

$$E := \left\{ \varphi \in \prod_{i=1}^{N} L^\infty(X_i, \mathcal{F}_i, m_i) : \int_{X_i} \varphi_i \mathrm{d}m_i = 0, \ i = 1, \dots, N-1 \right\}$$

For $\varphi = (\varphi_1, \dots, \varphi_N) \in \prod_{i=1}^{N} L^\infty(X_i, \mathcal{F}_i, m_i)$ define $T(\varphi) = (T_1(\varphi), \dots, T_N(\varphi)) \in \prod_{i=1}^{N} L^\infty(X_i, \mathcal{F}_i, m_i)$ by

$$T_i(\varphi)(x_i) := \int_{X_{-i}} K(x_i, x_{-i}) e^{\sum_{j=1}^{N} \varphi_j(x_j)} \mathrm{d}m_{-i}(x_{-i}). \qquad (13)$$

Note that $T(E) = T(\prod_{i=1}^N L^\infty(X_i, \mathcal{F}_i, m_i)) \subset F_{++}$ where

$$F_{++} := F \cap \prod_{i=1}^N L^\infty_{++}(X_i, \mathcal{F}_i, m_i), \tag{14}$$

and

$$F := \left\{ \mu \in \prod_{i=1}^N L^\infty(X_i, \mathcal{F}_i, m_i) \ : \int_{X_1} \mu_1 dm_1 = \ldots = \int_{X_N} \mu_N dm_N \right\}. \tag{15}$$

With these definitions the Schrödinger system simply writes $\mu = T(\varphi)$.

Well-posedness

Theorem 1 *For every* $\mu \in F_{++}$, *the multi-marginal Schrödinger system (10) admits a unique solution* $\varphi = S(\mu) \in E$, *moreover* $S \in C^\infty(F_{++}, E)$.

Simply based on the inverse function theorem and the Fredholm alternative theorem.

Local invertibility. Convenient to define the map $\widetilde{T} = (\widetilde{T}_1, \dots, \widetilde{T}_N)$ by $\widetilde{T}_i(\varphi) := \log(T_i(\varphi))$ for $\varphi = (\varphi_1, \dots, \varphi_N) \in \prod_{i=1}^N L^\infty(X_i, \mathcal{F}_i, m_i)$ i.e.

$$\widetilde{T}_i(\varphi)(x_i) := \varphi_i(x_i) + \log\left(\int_{X_{-i}} K(x_i, x_{-i}) e^{\sum_{j\neq i}\varphi_j(x_j)} \mathrm{d}m_{-i}(x_{-i})\right). \tag{16}$$

Both $\widetilde{T}$ and T are of class C^∞, more precisely for φ and h in $\prod_{i=1}^N L^\infty(X_i, \mathcal{F}_i, m_i)$, we have

$$\widetilde{T}_i'(\varphi)(h)(x_i) = h_i(x_i) + \frac{\int_{X_{-i}} K(x_i, x_{-i}) e^{\sum_{k\neq i}\varphi_k(x_k)} \sum_{j\neq i} h_j(x_j) \mathrm{d}m_{-i}(x_{-i})}{\int_{X_{-i}} K(x_i, x_{-i}) e^{\sum_{j\neq i}\varphi_j(x_j)} \mathrm{d}m_{-i}(x_{-i})}$$

and

$$T_i'(\varphi)(h)(x_i) = e^{\widetilde{T}_i(\varphi)(x_i)} \widetilde{T}_i'(\varphi)(h)(x_i). \tag{17}$$

Fix $\varphi := (\varphi_1, \ldots, \varphi_N) \in E$, observe that $\widetilde{T}'(\varphi)$ extends to a bounded linear self map of $\prod_{i=1}^N L^2(X_i, \mathcal{F}_i, m_i)$ which is of the form

$$\widetilde{T}'(\varphi) := \mathrm{id} + L \tag{18}$$

with L compact. Convenient to define the (equivalent to m) probability

$$Q_\varphi(\mathrm{d}x) = \frac{K(x) e^{\sum_{j=1}^N \varphi_j(x_j)} m(\mathrm{d}x)}{\int_X K(x) e^{\sum_{j=1}^N \varphi_j(x_j)} \mathrm{d}m(x)}. \tag{19}$$

For $i = 1, \ldots, N$, disintegrate Q_φ with respect to its i-th marginal Q_φ^i:

$$Q_\varphi(\mathrm{d}x_i, \mathrm{d}x_{-i}) = Q_\varphi^{-i}(\mathrm{d}x_{-i}|x_i) \otimes Q_\varphi^i(\mathrm{d}x_i) \tag{20}$$

where $Q_\varphi^{-i}(\mathrm{d}x_{-i}|x_i)$ is the conditional probability of x_{-i} given x_i according to Q_φ. The compact operator L can then conveniently be expressed in terms of the corresponding conditional expectations operators. Indeed, setting $L(h) = (L_1(h), \ldots, L_N(h))$, we have

$$L_i(h)(x_i) = \int_{X_{-i}} \Big(\sum_{j \neq i} h_j(x_j)\Big) Q_\varphi^{-i}(\mathrm{d}x_{-i}|x_i) \text{ for } m_i\text{-a.e. } x_i \in X_i.$$

Assume $\widetilde{T}'(\varphi)(h) = 0$ (equivalently $T'(\varphi)(h) = 0$) then

$$h_i(x_i) = -\int_{X_{-i}} \Big(\sum_{j \neq i} h_j(x_j)\Big) Q_{\varphi}^{-i}(\mathrm{d}x_{-i}|x_i)$$

multiplying by $h_i(x_i)$ and then integrating with respect to Q_{φ}^{i} gives

$$\int_{X_i} h_i^2(x_i)\mathrm{d}Q_{\varphi}^{i}(x_i) = -\sum_{j,\, j \neq i} \int_X h_i(x_i) h_j(x_j) \mathrm{d}Q_{\varphi}(x)$$

Summing over i thus yields

$$\int_X \Big(\sum_{i=1}^N h_i(x_i)\Big)^2 \mathrm{d}Q_\varphi(x) = \sum_{i=1}^N \int_{X_i} h_i^2(x_i)\mathrm{d}Q_\varphi^i(x_i) + \sum_{i,j,\, j\neq i} \int_X h_i(x_i)h_j(x_j)\mathrm{d}Q_\varphi(x) = 0.$$

Since Q_φ is equivalent to m, we deduce that $\sum_{i=1}^N h_i(x_i) = 0$ m-a.e. that is h is constant and its components sum to 0. Hence $\ker(\widetilde{T}'(\varphi))$ has dimension $N-1$ and $\ker(\widetilde{T}'(\varphi)) \cap E = \{0\}$ i.e. $\widetilde{T}'(\varphi)$ is one to one on E.

Using the Fredholm alternative, we deduce that $R(\mathrm{id} + L)$ has codimension $N - 1$ (same as F). It is then easy to deduce that $T'(\varphi)(E) = F$ and local invertibility follows. Therefore $T(E)$ is open, it is also closed in F_{++} hence T is also onto by connectedness of F_{++}.

Sinkhorn algorithm converges linearly

Same L^∞ setting as before $K = e^{-c}$ bounded from above and away from 0, fixed marginals m_i (so take m_i as reference measures, i.e. take $\mu_i = 1$ in the notations of the previous part). We want to solve

$$e^{\varphi_i(x_i)} \int_{X_{-i}} e^{-c(x_1,\ldots,x_N)+\sum_{j\neq i}\varphi_j(x_j)} \mathrm{d}m_{-i}(x_{-i}) = 1, \qquad (21)$$

for every i and m_i-a.e. x_i, which means that $Q_\varphi = Ke^{\oplus\varphi_i} m$ has marginals $(m_1, \ldots, m_N)$. Let us do it in a Gauss-Seidel/Sinkhorn/IPFP way i.e. by fitting one constraint at a time.

The system (21) is well-known to be the Euler-Lagrange optimality condition for the convex minimization problem

$$\inf_{\varphi \in \prod_{i=1}^{N} L^{\infty}(X_i, \mathcal{F}_i, m_i)} F(\varphi) := -\sum_{i=1}^{N} \int_{X_i} \varphi_i \mathrm{d}m_i + \int_X \mathrm{d}Q_\varphi \qquad (22)$$

and if φ solves (21), the measure $Q_\varphi = Ke^{\oplus \varphi_i} m$ solves the multi-marginal entropy minimization:

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

Again, $N-1$ degrees of freedom, so impose

$$\int_{X_1} \varphi_1 \mathrm{d}m_1 = \ldots = \int_{X_{N-1}} \varphi_{N-1} \mathrm{d}m_{N-1} = 0. \tag{23}$$

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

$$\inf_{\varphi \in E} F(\varphi) \;\text{ where } E := \prod_{i=1}^{N-1} L^\infty_\diamond(X_i, \mathcal{F}_i, m_i) \times L^\infty(X_N, \mathcal{F}_N, m_N). \tag{24}$$

Starting from $\varphi^0 \in E$, the updates of the Sinkhorn algorithm, consists, given $\varphi^t = (\varphi_1^t, \ldots, \varphi_N^t) \in E$, in:

$$\varphi_1^{t+1} := \operatorname{argmin}_{\varphi_1 \in L^\infty_\diamond(X_1, \mathcal{F}_1, m_1)} F(\varphi_1, \varphi_2^t, \ldots, \varphi_2^t) \tag{25}$$

i.e.

$$\varphi_1^{t+1}(x_1) := -\log\Big(\int_{X_{-1}} e^{\sum_{j=2}^N \varphi_j^t(x_j)} K(x_1, x_{-1}) \mathrm{d}m_{-1}(x_{-1})\Big) + \lambda_1^t, \tag{26}$$

where

$$\lambda_1^t = \int_{X_1} \Big(\log\Big(\int_{X_{-1}} e^{\oplus_{j=2}^N \varphi_j^t} K(x_1, x_{-1}) \mathrm{d}m_{-1}(x_{-1})\Big)\Big) \mathrm{d}m_1(x_1). \tag{27}$$

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

$$\varphi_i^{t+1} := \text{argmin}_{\varphi_i \in L^\infty_\diamond(X_i, \mathcal{F}_i, m_i)} F(\varphi_1^{t+1}, \ldots, \varphi_{i-1}^{t+1}, \varphi_i, \varphi_{i+1}^t, \ldots, \varphi_N^t) \tag{28}$$

i.e.

$$\varphi_i^{t+1}(x_i) := -\log\left(\int_{X_{-i}} e^{\oplus_{j=1}^{i-1}\varphi_j^{t+1} \ \oplus_{j=i+1}^{N}\varphi_j^t} K(x_i, x_{-i}) \mathrm{d}m_{-i}\right) + \lambda_i^t, \tag{29}$$

where

$$\lambda_i^t = \int_{X_i}\left(\log\left(\int_{X_{-i}} e^{\oplus_{j=1}^{i-1}\varphi_j^{t+1} \ \oplus_{j=i+1}^{N}\varphi_j^t} K(x_i, x_{-i}) \mathrm{d}m_{-i}(x_{-i})\right)\right) \mathrm{d}m_i. \tag{30}$$

Finally, for $i = N$,

$$\varphi_N^{t+1} := \mathrm{argmin}_{\varphi_N \in L^\infty(X_N, \mathcal{F}_N, m_N)} F(\varphi_1^{t+1}, \dots, \varphi_{N-1}^{t+1}, \dots, \varphi_N) \tag{31}$$

i.e.

$$\varphi_N^{t+1}(x_N) := -\log\left(\int_{X_{-N}} e^{\bigoplus_{j=1}^{N-1}\varphi_j^{t+1}} K(x_N, x_{-N}) \mathrm{d}m_{-N}(x_{-N})\right) \tag{32}$$

The convergence of the Sinkhorn iterates to a solution of (21) (hence a minimizer of (22)) was established by Di Marino and Gerolin, we wish to slightly improve this result by showing that this convergence is linear

Theorem 2 *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 (24). Moreover, there holds*

$$F(\varphi^t) - F(\overline{\varphi}) \leq \left(1 - \frac{e^{-(16N-8)\|c\|_\infty}}{N}\right)^t (F(\varphi^0) - F(\overline{\varphi})). \quad (33)$$

Btw, $\varphi^t - \overline{\varphi}$ also converges linearly in L^p for every $p \in (1, \infty)$.

First step: since c is bounded so are the Sinkhorn iterates:

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

$$\|\varphi_i^t\|_{L^\infty(X_i,\mathcal{F}_i,m_i)} \leq 2\|c\|_\infty, i = 1, \dots, N-1, \tag{34}$$

$$\|\varphi_N^t\|_{L^\infty(X_N,\mathcal{F}_N,m_N)} \leq (2N-1)\|c\|_\infty. \tag{35}$$

Next steps: Sinkhorn iterates thus remain in intervals where the exponential in F remains uniformly convex and with a bounded second derivative. One can therefore argue as Luo-Tseng and Beck-Tetruashvilli analysis of block coordinate descent (paying a bit of attention to the fact we are in infinite dimensions)

Trivial fact: given $M > 0$, $\forall (a,b) \in [-M,M]^2$:

$$e^b - e^a - e^a(b-a) \geq \frac{e^{-M}}{2}(b-a)^2, \ |e^b - e^a| \leq e^M |b-a|. \quad (36)$$

so, defining

$$\nu = e^{-(4N-2)\|c\|_\infty}, \quad (37)$$

one has

$$F(\varphi^t) - F(\varphi^{t+1}) \geq \frac{\nu}{2} \sum_{i=1}^N \|\varphi_i^t - \varphi_i^{t+1}\|^2_{L^2(X_i,\mathcal{F}_i,m_i)}. \quad (38)$$

in particular $\varphi^t - \varphi^{t+1}$ tends to 0 strongly in L^2 (and in fact in any L^p). Convergence to $\overline{\varphi}$ solving the Schrödinger system follows (lots of compactness again).

Using (36) again, we get

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

Define

$$\widetilde{\varphi}_i^t := (\varphi_1^{t+1}, \ldots, \varphi_i^{t+1}, \varphi_{i+1}^t, \ldots, \varphi_N^t),\ i = 1, \ldots, N-1,\ \ \widetilde{\varphi}_N^t := \varphi^{t+1}, \tag{39}$$

by construction

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

So

$$F(\overline{\varphi}) - F(\varphi^t) \geq \sum_{i=1}^{N} \int_{X_i} (\partial_i F(\varphi^t) - \partial_i F(\widetilde{\varphi}_i^t))(\overline{\varphi}_i - \varphi_i^t) \mathrm{d}m_i$$

$$+ \frac{\nu}{2} \sum_{i=1}^{N} \|\overline{\varphi}_i - \varphi_i^t\|_{L^2(m_i)}^2$$

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

where we have used Young's inequality in the last line.

We thus have shown that

$$F(\varphi^t) - F(\overline{\varphi}) \leq \frac{1}{2\nu} \sum_{i=1}^{N} \|\partial_i F(\varphi^t) - \partial_i F(\widetilde{\varphi}_i^t)\|^2_{L^2(m_i)}. \tag{40}$$

Using the second inequality in (36) together with the L^∞ bounds on φ^t 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 (40), we thus obtain

$$F(\varphi^t) - F(\overline{\varphi}) \leq \frac{N}{2\nu^3} \sum_{i=1}^{N} \|\varphi_i^t - \varphi_i^{t+1}\|^2_{L^2(m_i)}. \tag{41}$$

Finally, combining (41) with (38), we deduce

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

from which the linear convergence in (33) readily follows.