Quantitative Stability of the Iterative Fitting Procedure

George Deligiannidis

Joint with V. de Bortoli and A. Doucet

Department of Statistics, Oxford University

AUEB Statistics Seminar, Oct 29 2021

Optimal Transport

- Coupling: A coupling of two probability measures μ, ν, on spaces $X, Y \subset \mathbb{R}^d$ respectively, a *coupling* is a *joint distribution* π, with μ, ν as its marginals,

$$\int f(x)\pi(dx, dy) = \int f(x)\mu(dx), \quad \int g(y)\pi(dx, dy) = \int g(y)\nu(dy).$$

- We write $\mathcal{C}(\mu, \nu)$ for the collection of couplings of μ, ν.
- Optimal Transport: the basic problem of Optimal Transport is to find a coupling $\pi \in \mathcal{C}(\mu, \nu)$ that minimizes

$$\inf_{\pi \in \mathcal{C}(\mu,\nu)} \int \pi(dx, dy) c(x, y); \qquad \mathrm{OT}(\mu, \nu)$$

where $c : X \times Y \to \mathbb{R}$ is any cost function.

Optimal Transport

$$\inf_{\pi \in \mathcal{C}(\mu,\nu)} \int \pi(\mathrm{d}x, \mathrm{d}y) c(x,y);$$

- this formulation goes back to Kantorovich;
- the earlier formulation due Monge focuses on mappings $y = T(x)$, that is $\pi(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x)\delta_{T(x)}(\mathrm{d}y)$ and results in non-convex optimisation.
- Brenier 1987: Assuming $c(x,y) = |x-y|^2$, μ, ν have finite second moments and μ gives measure 0 to all sets of Hausdorff dimension $\leq d-1$, there exists a convex mapping $\varphi : \mathcal{X} \mapsto \mathcal{Y}$ such that optimal coupling takes the form of an optimal transport map $\mu(\mathrm{d}x)\delta_{\nabla\varphi(x)}(\mathrm{d}y)$
- When μ is discrete, π is a proper coupling.
- For discrete measures with N atoms, worst case computational cost is $O(N^{5/2})$ for assignment problem, using Hungarian (auction) algorithm, see e.g. Mérigot and Oudet 2016.
- This can be improved by considering the entropy regularised version.

Entropy Regularised Optimal Transport

- Cuturi 2013 realised that the entropy regularised problem

$$\inf_{\pi \in \mathcal{C}(\mu,\nu)} \int \pi(\mathrm{d}x, \mathrm{d}y)\|x-y\|^2 + \epsilon\,\mathrm{KL}\,(\pi|\mu \otimes \nu)\,; \qquad \mathrm{OT}_\epsilon(\mu,\nu)$$

  can be solved efficiently using the Iterative Proportional Fitting Procedure (IPFP), aka Sinkhorn's algorithm.
- the computational cost is roughly $O(N^2)$, see Altschuler, Weed, and Rigollet 2017.
- Equivalent to (static) Schrödinger bridge problem

$$\inf_{\pi \in \mathcal{C}(\mu,\nu)} \mathrm{KL}\,(\pi|\Gamma_\epsilon)\,, \qquad \Gamma_\epsilon(\mathrm{d}x, \mathrm{d}y) = \exp[-c(x,y)/\epsilon]\mu(\mathrm{d}x)\nu(\mathrm{d}y).$$

Recent applications of Schrödinger bridges

- Idea is to use solution π_ϵ^* of OT_ϵ and treat it as an approximation to the solution π^* of OT.
- A lot of recent progress quantifying $\pi_\epsilon^* \to \pi^*$.
- In recent applications, the Schrödinger bridge is used for its own benefits rather than as a computationally feasible approximation to the standard optimal transport problem.

Differentiable Particle Filtering

- Corenflos et al. 2021 used the solution to Schrödinger bridge to build a particle filtering scheme.
- Suppose $\widehat{\pi}_t = N^{-1}\sum_{j=1}^{N} \delta_{X_j^t}$ is particle approximation of π_t;
- Let

$$\widetilde{X}_j^{t+1} \sim q_{t+1}(\cdot|X_j^t), \quad j \in [N]; \tag{1}$$

$$\omega_j^{t+1} = \frac{f_\theta(\widetilde{X}_j^{t+1}|X_j^t)g_\theta(Y^{t+1}|\widetilde{X}_j^{t+1})}{q_{t+1}(\widetilde{X}_j^{t+1}|X_j^t)} \tag{2}$$

$$\mathbb{P}(I_j^{t+1} = k) = \frac{\omega_k^{t+1}}{\sum_{j=1}^{N}\omega_j^{t+1}}, \tag{3}$$

$$X_j^{t+1} \overset{\text{iid}}{\sim} \sum_{k=1}^{N} \frac{\omega_k^{t+1}}{\sum_{j=1}^{N}\omega_j^{t+1}}\delta_{\widetilde{X}_k^{t+1}} \tag{4}$$

- A standard issue is that the resampling step means any estimators produced are not differentiable wrt θ.
- There are approximate ways to bypass this.

Differentiable Particle Filtering

- Corenflos et al. 2021 solves the ϵ-entropy regularised OT problem between

$$\alpha_N^{(t)} = \frac{1}{N}\sum_{i=1}^{N}\delta_{\tilde{X}_t^i}, \quad \beta_N^{(t)} = \sum_i w_t^i \delta_{\tilde{X}_t^i}$$

  to obtain a matrix $\boldsymbol{P}_\epsilon$;
- instead of resampling we use $\boldsymbol{P}_\epsilon$ to produce an ensemble transform; $\boldsymbol{X}_t = \boldsymbol{P}_\epsilon \tilde{\boldsymbol{X}}_t$
- the mapping $\varphi \mapsto \boldsymbol{P}_\epsilon$ can be differentiated so we can get end to end differentiable estimators.
- In analysing the consistency of the differentiable particle filter we can either compare with the solution of the unregularised OT problem,
- OR we can compare the solution of $\mathrm{OT}_\epsilon(\alpha_N, \beta_N)$ with that of $\mathrm{OT}_\epsilon(\alpha_N, \beta_N)$ where $\alpha = \lim_{N\to\infty}\alpha_N, \beta = \lim_N \beta_N$.

Stability of OT

This brings us to the following question

- Question: suppose $\alpha_n \to \alpha$ and $\beta_n \to \beta$ (say weakly);
- *D*oes the solution of $\mathrm{OT}(\alpha_n, \beta_n)$ converges in some sense to that of $\mathrm{OT}(\alpha, \beta)$?
- In the realm of Brenier's theorem, we could talk about convergence of the transport maps.
- In classical Optimal Transport there is a classical, qualitative solution.

Stability of OT

Theorem (5.20 in Villani 2009)

Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces and let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a a continuous, lower bounded cont function. Let $(\mu_k), (\nu_k)$ be sequences of probability measures on $\mathcal{X}, \mathcal{Y}$ respectively, such that $\mu_k \to \mu$ and $\nu_k \to \nu$ weakly. For each k, let π_k be an optimal transference plan between μ_k and ν_k.
If for all $k \in \mathbb{N}$, $\int c \mathrm{d}\pi_k < +\infty$, then up to extraction of a subsequence, π_k converges weakly to some c-cyclically monotone transference plan $\pi \in \mathcal{C}(\mu, \nu)$.
If moreover $\liminf_k \int c \mathrm{d}\pi_k < \infty$, then the optimal transport cost between μ, ν is finite and π is an optimal transport plan.

Stability of OT

- If the optimal transport plan between μ and ν is unique then there is no need to extract a subsequence.
- If plan is given by map, maps converge in probability (Corollary 5.23 Villani 2009).
- Proof is based on compactness and characterization of optimal transport plans in terms of c-cyclical monotonicity: a set $\Gamma \subset X \times Y$ is called c-cyclically monotone if for all collections $(x_1, y_1), \ldots, (x_N, y_N) \subset \Gamma$ we have

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{i+1}), \quad y_{N+1} := y_1.$$

a transport plan is called c-cyclically monotone if its support is.

Quantitative Stability of OT

- Quantitative results have recently appeared in the literature, see Li and Nochetto 2021 and Mérigot, Delalande, and Chazal 2020.
- Li and Nochetto 2021 proves the following: suppose μ_h, ν_h are approximations of μ, ν respectively, $\gamma_h \in \arg\min \mathrm{OT}(\mu_h, \nu_h)$, and $T = \nabla\varphi$ is the unique optimal transport map between μ, ν. Then

$$\left(\int_{\mathrm{X}\times\mathrm{Y}} |T(x) - y|^2 \mathrm{d}\gamma_h(x,y)\right)^{1/2} \tag{5}$$

$$\leq 2\lambda^{1/2}\Delta_h^{1/2}\left[\mathbf{W}_2(\mu,\nu) + \Delta_h\right]^{1/2} \quad + (\lambda \vee 1)\Delta_h, \tag{6}$$

where

$$\Delta_h := \mathbf{W}_2(\mu,\mu_h) + \mathbf{W}_2(\nu,\nu_h),$$

and λ is the Lipschitz constant of $\nabla\varphi$.

- Proof essentially exploits convexity.

Quantitative Stability of OT

Mérigot, Delalande, and Chazal 2020 prove the following:

- let ρ, μ, ν be three probability measures and suppose T_μ, T_ν are Brenier maps sending ρ to μ, ν respectively.
- Then
$$\mathbf{W}_2(\mu,\nu) \leq \|T_\mu - T_\nu\|_{L^2(\rho)} \leq C\mathbf{W}_p(\mu,\nu)^{2/15},$$
for any $p \geq 1$, where C depends on the dimension and the sets X, Y.
- If T_μ is K-Lipschitz then the exponent can be improved to $1/2$.
- They also show that the Monge embedding $\mu \mapsto T_\mu$ is in general not better than $1/2$-Holder.

Stability of Regularised OT

- Until recently less was known on stability of regularised OT.
- The standard approach via compactness does not trivially extend.
- Also until recently, there was no equivalent of cyclical monotonicity to uniquely identify the optimal coupling.
- This was recently done in Ghosal, Nutz, and Bernton 2021 who proved the qualitative stability of Schödinger bridges (no compactness required).

Stability of Regularised OT

- For compact state spaces, Luise et al. 2019, prove stability of the potentials in total variation;
- $\Pi_\epsilon^{\mu,\nu}$ to $\mathrm{OT}_\epsilon(\mu,\nu)$ can be written in the form

$$\Pi_\epsilon^{\mu,\nu}(\mathrm{d}x,\mathrm{d}y) = \exp(\varphi_\epsilon^{\mu,\nu}(x) + \psi_\epsilon^{\mu,\nu}(y) - c(x,y)/\epsilon)\mu(\mathrm{d}x)\nu(\mathrm{d}y),$$

- $\varphi_\epsilon^{\mu,\nu}, \psi_\epsilon^{\mu,\nu}$ are the potentials.
- Luise et al. 2019 prove that

$$\|\varphi_\epsilon^{\mu,\nu} - \varphi_\epsilon^{\mu',\nu'}\|_\infty \leq C(d,\epsilon,\mathrm{X},\mathrm{Y})\left\{\|\mu-\mu'\|_{\mathrm{TV}} + \|\nu-\nu'\|_{\mathrm{TV}}\right\}.$$

  NOTE: TV is too strong to capture convergence of empirical measures.
- For smooth costs, $c \in C^{s+1}$, $s > d/2$, and ν_n an empirical version of ν they establish that

$$\|\varphi_\epsilon^{\mu,\nu} - \varphi_\epsilon^{\mu,\nu_n}\|_\infty \leq C(d,\epsilon,\mathrm{X},\mathrm{Y})\log(3/\tau)n^{-1/2},$$

  w.prob $> 1 - \tau$.
- This is obtained by considering MMD type metrics which do capture convergence in distribution, but they require smoothness of the metric.

Stability of Regularised OT

- We are interested in the full coupling $\Pi_\epsilon^{\mu,\nu}$ rather than just the potentials.
- There the sample complexity will necessarily depend on the dimension;
- Say μ_n, ν_n are empirical versions of μ, ν. Then trivially we have that

$$
\begin{aligned}
&\mathbf{W}_1\left(\Pi_\epsilon^{\mu_n,\nu_n}, \Pi_\epsilon^{\mu,\nu}\right) \\
&= \sup\left\{\int f(x,y)\left[\Pi_\epsilon^{\mu_n,\nu_n} - \Pi_\epsilon^{\mu,\nu}\right](\mathrm{d}x,\mathrm{d}y) : f \in \mathrm{Lip}\,(\mathrm{X}\times\mathrm{Y})\right\} \\
&\geq \sup\left\{\int f(x)\left[\Pi_\epsilon^{\mu_n,\nu_n} - \Pi_\epsilon^{\mu,\nu}\right](\mathrm{d}x,\mathrm{d}y) : f \in \mathrm{Lip}(\mathrm{X})\right\} \\
&=: \mathbf{W}_1\left(\mu_n, \mu\right),
\end{aligned}
$$

and we know this scales like $n^{1/d}$, see e.g. Fournier and Guillin 2015.

The Iterative Proportional Fitting Procedure (IPFP)

- Before stating our main results let us introduce the IPFP, also known as Sinkhorn's algorithm.
- Given two probability measures μ, ν and $\epsilon > 0$ the IPFP iteratively learns the potentials $\varphi_\epsilon^{\mu,\nu}, \psi_\epsilon^{\mu,\nu}$;
We simply write φ, ψ to ease notation.
- Initialise $\varphi^{(0)}, \psi^{(0)} \equiv 0$.
- Given $\varphi^{(t)}, \psi^{(t)}$, $t \geq 0$ set

$$\Rightarrow \varphi^{(t+1)}(x) := -\log \int \exp\{\psi^{(t)}(y) - c(x,y)/\epsilon\}\nu(\mathrm{d}y)$$

$$\Rightarrow \psi^{(t+1)}(y) := -\log \int \exp\{\varphi^{(t+1)}(x) - c(x,y)/\epsilon\}\mu(\mathrm{d}x)$$

$\Rightarrow$ For any $c \in \mathbb{R}$, $(\varphi^{(t)} - c, \psi^{(t)} + c)$ defines the same measure; we fix this choice following Carlier 2021, so that $\mu[\varphi^{(t+1)}] = 0$.

Main results

Theorem 1

Suppose that X, Y *are compact metric spaces and* $c \in \mathrm{Lip}(\mathsf{X} \times \mathsf{Y})$. *For any* $\pi_0, \widehat{\pi}_0 \in \mathscr{P}(\mathsf{X})$, $\pi_1, \widehat{\pi}_1 \in \mathscr{P}(\mathsf{Y})$ *let* $(\mathbb{P}^n)_{n\in\mathbb{N}}$ *and* $(\widehat{\mathbb{P}}^n)_{n\in\mathbb{N}}$ *the IPFP sequence with marginals* (π_0, π_1) *respectively* $(\widehat{\pi}_0, \widehat{\pi}_1)$. *Then any* $n \in \mathbb{N}$ *we have*

$$\mathbf{W}_1(\mathbb{P}^n, \widehat{\mathbb{P}}^n) \leq C\left\{\mathbf{W}_1(\pi_0, \widehat{\pi}_0) + \mathbf{W}_1(\pi_1, \widehat{\pi}_1)\right\}, \tag{7}$$

with

$$C = \mathrm{e}^{10\|c\|_\infty}\{1 + (2\mathrm{Lip}(c) + 10)(\mathrm{diam}(\mathsf{X}) + \mathrm{diam}(\mathsf{Y}))\}. \tag{8}$$

Main results (ctd)

As an immediate consequence of Theorem 1 and the fact that the IPFP sequence converges, we obtain the quantitative stability of Schrödinger bridge.

Corollary 2

For any $\pi_0, \widehat{\pi}_0 \in \mathscr{P}(X)$, $\pi_1, \widehat{\pi}_1 \in \mathscr{P}(Y)$ let $\mathbb{P}^*$, respectively $\widehat{\mathbb{P}}^*$, be the Schrödinger bridge with marginals (π_0, π_1), respectively $(\widehat{\pi}_0, \widehat{\pi}_1)$. Then we have

$$\mathbf{W}_1(\mathbb{P}^*, \widehat{\mathbb{P}}^*) \leq C\{\mathbf{W}_1(\pi_0, \widehat{\pi}_0) + \mathbf{W}_1(\pi_1, \widehat{\pi}_1)\}, \tag{9}$$

with C as in Theorem 1.

Background on Hilbert projective metric I

Will now sketch the main idea for the Schrödinger bridge, rather than IPFP as it is a little clearer.

- In compact spaces we can employ the machinery of the Birkhoff-Hopf contraction theorem;
- suppose E is a real vector space, K is a cone, that is K is convex, $K \cap (-K) = \{0\}$ and $\lambda K \subset K$ for all $\lambda \geq 0$.
- K induces a partial ordering on E: that is we write $x \geq y$ if $x - y \in K$;
- Let C be a part of the cone, that is for any $x, y \in C$ there exist $\alpha, \beta \geq 0$ such that $\alpha x - y \in K$ and $\beta y - x \in K$, C is convex and $\lambda C \subset C$ for all $\lambda > 0$.
- For any $x, y \in C$ we write

$$M(x,y) := \inf\{\beta \geq 0 : \beta y - x \in K\} \quad (10)$$

$$m(x,y) := \sup\{\alpha \geq 0 : x - \alpha y \in K\}. \quad (11)$$

Background on Hilbert projective metric II

- The Hilbert metric is defined for any $x, y \in C$ as
$$d_H(x,y) = \log M(x,y)/m(x,y).$$
- It is a projective metric in the sense that it measures distances between rays $\{\lambda x : \lambda \geq 0\}$ rather than points.

We are now ready to state the Birkhoff contraction theorem.

Background on Hilbert projective metric III

Theorem (Birhoff Contraction Theorem)

Let $(V, \|\cdot\|)$, $(V', \|\cdot\|')$ be two normed real vector spaces, $K \subset V, K' \subset V'$ two cones and $C' \subset K', C \subset K$ two convex parts, and write d_H, d'_H for the Hilbert metric on C, C' respectively. Let $T : V \to V'$ be a linear mapping such that $T(C) \subset C'$. Then

$$\kappa(T) := \sup_{x,y \in C} \frac{d'_H(T(x), T(y))}{d_H(x,y)} \leq \tanh(\Delta(T)/4), \tag{12}$$

where the projective diameter $\Delta(T)$ of T is defined by

$$\Delta(T) := \sup \{ d_H(T(x), T(y)) : x, y \in C, \|x\| = \|y\| = 1 \}.$$

Since $\Delta(T)$ is finite $\kappa(T) < 1$.

Sketch of proof I

- In our context

$$C = \mathscr{C}(\mathrm{X}; (0,\infty)), \quad C' = \mathscr{C}(\mathrm{Y}; (0,\infty)).$$

- In this setting

$$M(f,g) = \sup f/g, \quad m(f,g) = \inf f/g,$$

- so the Hilbert-Birkhoff metric measures the oscillations on the log-scale

$$d_H(f,g) = \|\log(f/g)\|_{\mathrm{osc}} := \sup \log(f/g) - \inf \log(f/g).$$

- Letting $\mu \in \mathscr{P}(\mathrm{X}), \nu \in \mathscr{P}(\mathrm{Y})$ define two linear maps of interest through

$$(\mathcal{E}_\mu f)(y) := \int f(x) K(x,y) \mu(\mathrm{d}x); \quad \mathcal{E}_\mu : C \to C' \tag{13}$$

$$(\mathcal{E}_\nu g)(x) := \int g(y) K(x,y) \nu(\mathrm{d}y); \quad \mathcal{E}_\nu : C' \to C. \tag{14}$$

Sketch of proof II

- Let also $\mathscr{I} : f \mapsto f^{-1}$, where we overload the operator to act on both C and C'.
- As pointed out in Chen, Georgiou, and Pavon 2016 it is easy to show that $\mathscr{I}$ is an isometry w.r.t. the Hilbert metric;
- Also we can bound the projective diameter of $\mathcal{E}_\mu, \mathcal{E}_\nu$.
- In this notation the IPFP iteration that takes $e^{\varphi^{(t)}} \to e^{\varphi^{(t+1)}}$ can be written as

$$\mathscr{S}_{\mu,\nu} \exp[\varphi^{(t)}] := \left[\mathscr{I} \circ \mathcal{E}_\nu \circ \mathscr{I} \circ \mathcal{E}_\mu\right] [\exp \varphi^{(t)}], \tag{15}$$

$$\mathscr{S}^\dagger_{\mu,\nu} \exp[\psi^{(t)}] := \left[\mathscr{I} \circ \mathcal{E}_\mu \circ \mathscr{I} \circ \mathcal{E}_\nu\right] \exp[\psi^{(t)}]. \tag{16}$$

- The Birkhoff contraction theorem show that both of these maps are contractions in the Hilbert metric.
- The pairs of potentials $(\varphi_\epsilon, \psi_\epsilon), (\widehat{\varphi}_\epsilon, \widehat{\psi}_\epsilon)$ defining the Schrödinger bridges for $(\mu, \nu), (\widehat{\mu}, \widehat{\nu})$ respectively will then be fixed points of $(\mathscr{S}_{\mu,\nu}, \mathscr{S}^\dagger_{\mu,\nu})$ and $(\mathscr{S}_{\widehat{\mu},\widehat{\nu}}, \mathscr{S}^\dagger_{\widehat{\mu},\widehat{\nu}})$ resp.

Sketch of proof:main ideas I

For $h \in \mathrm{Lip}(X \times Y)$, writing

$$F_\epsilon(x,y) = \exp\{\varphi_\epsilon(x) + \psi_\epsilon(y)\}, \quad \widehat{F}_\epsilon(x,y) = \exp\{\widehat{\varphi}_\epsilon(x) + \widehat{\psi}_\epsilon(y)\}$$

we want to write

$$\int h(x,y)\left[\Pi_\epsilon^{\mu,\nu}(\mathrm{d}x,\mathrm{d}y) - \Pi_\epsilon^{\widehat{\mu},\widehat{\nu}}(\mathrm{d}x,\mathrm{d}y)\right]$$
$$= \int h(x,y)F_\epsilon(x,y)K_\epsilon(x,y)\mu(\mathrm{d}x)\nu(\mathrm{d}y)s - \int h(x,y)\widehat{F}_\epsilon(x,y)K_\epsilon(x,y)\widehat{\mu}(\mathrm{d}x)\widehat{\nu}(\mathrm{d}y)$$
$$= \int h\left[F_\epsilon - \widehat{F}_\epsilon\right]K_\epsilon(x,y)\mu(\mathrm{d}x)\nu(\mathrm{d}y) + \int h\widehat{F}_\epsilon K_\epsilon(x,y)\left[\mu\otimes\nu - \widehat{\mu}\otimes\widehat{\nu}\right].$$

- The second term looks like it can be controlled by $\mathbf{W}_1(\mu\otimes\nu, \widehat{\mu}\otimes\widehat{\nu})$.
- Issue is that in general $\widehat{\varphi}_\epsilon, \widehat{\psi}_\epsilon$ are only defined on the supports of $\widehat{\mu}, \widehat{\nu}$.
- How can $\widehat{F}_\epsilon$ be Lipschitz?

Sketch of proof:main ideas II

- Here we use the fact that $\widehat{\varphi}_\epsilon, \widehat{\psi}_\epsilon$ are fixed points of the Sinkhorn iteration, that is

$$\widehat{\varphi}_\epsilon(x) = -\log \int \exp\{\psi_\epsilon(y) - c(x,y)/\epsilon\}\widehat{\nu}(\mathrm{d}y) \tag{17}$$

$$\widehat{\psi}_\epsilon(x) = -\log \int \exp\{\varphi_\epsilon(x) - c(x,y)/\epsilon\}\widehat{\mu}(\mathrm{d}x). \tag{18}$$

- Using the above we can extend $\widehat{\varphi}_\epsilon, \widehat{\psi}_\epsilon$ to Lipschitz continuous functions on all of X, Y (compactness is used heavily here), see also Luise et al. 2019,
- So indeed

$$\int h(x,y)\left[\Pi_\epsilon^{\mu,\nu}(\mathrm{d}x,\mathrm{d}y) - \Pi_\epsilon^{\widehat{\mu},\widehat{\nu}}(\mathrm{d}x,\mathrm{d}y)\right]$$
$$\leq \int h\left[F_\epsilon - \widehat{F}_\epsilon\right] K_\epsilon(x,y)\mu(\mathrm{d}x)\nu(\mathrm{d}y) + \mathbf{W}_1(\mu\otimes\nu, \widehat{\mu}\otimes\widehat{\nu}).$$

Sketch of proof:main ideas I

- Next we use the fact that $F_\epsilon, \widehat{F}_\epsilon$ are fixed points of Sinkhorn iterations to control $F_\epsilon - \widehat{F}_\epsilon$.
- Idea here is

$$
\begin{aligned}
d_H(F_\epsilon, \widehat{F}_\epsilon) &= d_H\left(\mathscr{S}_{\mu,\nu} F_\epsilon, \mathscr{S}_{\widehat{\mu},\widehat{\nu}} \widehat{F}_\epsilon\right) \\
&= \underbrace{d_H\left(\mathscr{S}_{\mu,\nu} F_\epsilon, \mathscr{S}_{\mu,\nu} \widehat{F}_\epsilon\right)}_{\text{contraction of Sinkhorn}} + \underbrace{d_H\left(\mathscr{S}_{\mu,\nu} \widehat{F}_\epsilon, \mathscr{S}_{\widehat{\mu},\widehat{\nu}} \widehat{F}_\epsilon\right)}_{\leq \mathbf{W}_1(\mu\otimes\nu, \widehat{\mu}\otimes\widehat{\nu})} \\
&= \kappa d_H\left(F_\epsilon, \widehat{F}_\epsilon\right) + C\mathbf{W}_1(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu}) \\
d_H(F_\epsilon, \widehat{F}_\epsilon) &\leq \frac{C}{1-\kappa} \mathbf{W}_1(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu}).
\end{aligned}
$$

- Final issue is that $d_H(F_\epsilon, \widehat{F}_\epsilon)$ only controls the oscillations $\| \log F_\epsilon - \log \widehat{F}_\epsilon \|_{\mathrm{osc}}$ rather than the supremum;

Sketch of proof:main ideas II

- To bypass this issue notice that

$$\int \frac{\widehat{F}_\epsilon}{F_\epsilon} F_\epsilon K_\epsilon \mu \otimes \nu = \int \frac{\widehat{F}_\epsilon}{F_\epsilon} F_\epsilon K_\epsilon \widehat{\mu} \otimes \widehat{\nu} + C\mathbf{W}_1(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu}) \tag{19}$$

$$= \int \widehat{F}_\epsilon K_\epsilon \widehat{\mu} \otimes \widehat{\nu} + C\mathbf{W}_1(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu}) \tag{20}$$

$$= 1 + C\mathbf{W}_1(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu}) \tag{21}$$

$$\tag{22}$$

- Recall $F_\epsilon K_\epsilon \mu \otimes \nu$ is a probability measure;
- thus the random variable $\widehat{F}_\epsilon / F_\epsilon(X, Y)$ with $(X, Y) \sim F_\epsilon K_\epsilon \mu \otimes \nu$, must either be a.s. equal to $1 + C\mathbf{W}_1(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu})$, or must take values both above and below $1 + C\mathbf{W}_1(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu})$.

Sketch of proof:main ideas III

- In either case using continuity of $\widehat{F}_\epsilon / F_\epsilon$ we can find x_0, y_0 such that

$$\log \frac{\widehat{F}_\epsilon(x_0, y_0)}{F_\epsilon(x_0, y_0)} = \log\left(1 + C\mathbf{W}_1(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu})\right)$$

and thus

$$\sup_{x,y}\left[\log \widehat{F}_\epsilon(x,y) - \log F_\epsilon(x,y)\right] \leq C\mathbf{W}_1\left(\mu \otimes \nu, \widehat{\mu} \otimes \widehat{\nu}\right) + \|\log \widehat{F}_\epsilon(x,y) - \log F_\epsilon(x,y)\|_{\mathrm{osc}}.$$

- We can similarly control $F_\epsilon / \widehat{F}_\epsilon$.

Recent results by Eckstein and Nutz 2021

- A couple of months after our preprint appeared online, Eckstein and Nutz 2021 posted a very nice paper with some quantitative results for Schrödinger bridge.
- They treat the more general, non-compact case and prove stability in the Wasserstein metric.
- They only treat the Schrödinger bridge, that is the limit of the IPFP algorith.
- They prove that the Schrödinger bridge is Hölder continuous, rather than Lipschitz, in the marginals, using very interesting probabilistic techniques involving some approximate couplings.

An incomplete list of references I

- Jason Altschuler, Jonathan Weed, and Philippe Rigollet. “Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration”. In: *Advances in Neural Information Processing Systems*. 2017.
- Yann Brenier. “Polar decomposition and increasing rearrangement of vector-fields”. In: *COMPTES RENDUS DE L ACADEMIE DES SCIENCES SERIE I-MATHEMATIQUE* 305.19 (1987), pp. 805–808.
- Guillaume Carlier. “On the linear convergence of the multi-marginal Sinkhorn algorithm”. In: *HAL preprint hal-03176512* (2021).
- Yongxin Chen, Tryphon Georgiou, and Michele Pavon. “Entropic and displacement interpolation: a computational approach using the Hilbert metric”. In: *SIAM Journal on Applied Mathematics* 76.6 (2016), pp. 2375–2396.

An incomplete list of references II

- Adrien Corenflos, James Thornton, George Deligiannidis, and Arnaud Doucet. "Differentiable particle filtering via entropy-regularized optimal transport". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 2100–2111.
- Marco Cuturi. "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in Neural Information Processing Systems*. 2013.
- Stephan Eckstein and Marcel Nutz. "Quantitative Stability of Regularized Optimal Transport". In: *arXiv preprint arXiv:2110.06798* (2021).
- Nicolas Fournier and Arnaud Guillin. "On the rate of convergence in Wasserstein distance of the empirical measure". In: *Probability Theory and Related Fields* 162.3 (2015), pp. 707–738.

An incomplete list of references III

Promit Ghosal, Marcel Nutz, and Espen Bernton. "Stability of Entropic Optimal Transport and Schrödinger Bridges". In: *arXiv preprint arXiv:2106.03670* (2021).

Wenbo Li and Ricardo H Nochetto. "Quantitative stability and error estimates for optimal transport plans". In: *IMA Journal of Numerical Analysis* 41 (3 2021), pp. 1941–1965.

Giulia Luise, Saverio Salzo, Massimiliano Pontil, and Carlo Ciliberto. "Sinkhorn Barycenters with Free Support via Frank-Wolfe Algorithm". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper/2019/file/9f96f36b7aae3b1ff847c26ac94c604e-Paper.pdf.

An incomplete list of references IV

- Quentin Mérigot, Alex Delalande, and Frédéric Chazal. "Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 3186–3196.
- Quentin Mérigot and Edouard Oudet. "Discrete optimal transport: complexity, geometry and applications". In: *Discrete & Computational Geometry* 55.2 (2016), pp. 263–283.
- Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2009.