Now for any set of n pairs of point $(x_1, y_1), \ldots, (x_n, y_n)$, if there exists $\sigma \in \text{Perm}(\{1, \ldots, n\})$ s.t.

$$\sum_{i=1}^{n} c(x_i, y_i) - \sum_{i=1}^{n} c(x_i, y_{\sigma(i)}) > 0, \text{ then}$$

$$\prod_{i=1}^{n} \frac{d\pi_\varepsilon^*}{d\mu \otimes \nu}(x_i, y_i) \longrightarrow 0 \quad \text{as } \varepsilon \to 0.$$

So if $\pi_\varepsilon^* \xrightarrow{\text{weakly}} \pi^*$, then, $\pi^* \in \Pi(\mu, \nu)$, support (π^*) is c-cyclicaly monotone.

Hence $\pi^* \in \underset{\pi \in \Pi(\mu,\nu)}{\arg\inf} \left\{ \int c \, d\pi \right\}$

Theorem: If MK problem w.r.t the cost function c has an unique solution π^*, then,

$$\pi_\varepsilon^* \xrightarrow{\text{weakly}} \pi^*.$$

Sinkhorn Algorithm.

Sinkhorn algorithm is an iterative procedure to solve entropic optimal transport problem.

In sinkhorn, there are two steps in each iteration.

Start with some function $\phi_0 = 0$.

(a) Define ψ_t at any $y \in \mathcal{Y}$ as

$$\psi_t(y) = -\log\left(\int e^{-c(x,y)+\phi_t(x)} d\mu(x)\right)$$

(b) Define ϕ_{t+1} at any $x \in \mathcal{X}$ as.

$$\phi_{t+1}(x) = -\log\left(\int e^{-c(x,y)+\psi_t(y)} d\nu(y)\right)$$

Why such updates are chosen?

It turns out above two steps are co-ordinate descent schemes for the concave maximization problem dual EOT, i.e.,

$$\sup_{\substack{\phi \in L^1(\mu) \\ \psi \in L^1(\nu)}} \Big\{ \underbrace{\int \phi \, d\mu + \int \psi \, d\nu - \iint \left(e^{\phi+\psi-c} - 1\right) d\mu \, d\nu}_{= \, G(\phi, \psi)} \Big\}$$

(a) corresponds to setting $\psi_t := \underset{\psi \in L^1(\nu)}{\operatorname{argmax}} \, G(\phi_t, \psi)$.

(b) corresponds to

$$\phi_{t+1} := \underset{\phi \in L^1(\mu)}{\operatorname{argmax}} \, G(\phi, \psi_t)$$

So we have the relation

$$G(\phi_{t-1}, \psi_{t-1}) \leq G(\phi_{t-1}, \psi_t) \leq G(\phi_t, \psi_t)$$

If we define Borel measures π_{2t-1} and π_{2t}

$\pi_{2t-1} \ll \mu \otimes \nu$ and $\pi_{2t} \ll \mu \otimes \nu$ as.

$$\frac{d\pi_{2t-1}}{d\mu\otimes\nu}(x,y) = e^{-c(x,y) + \phi_t(x) + \psi_{t-1}(y)}$$

$$\frac{d\pi_{2t}}{d\mu\otimes\nu}(x,y) = e^{-c(x,y) + \phi_t(x) + \psi_t(y)}$$

Then $$\pi_{2t} := \underset{\pi \in \Pi(*,\nu)}{\operatorname{argmin}} KL(\pi \| \pi_{2t-1})$$

$$\pi_{2t+1} := \underset{\pi \in \Pi(\mu,*)}{\operatorname{argmin}} KL(\pi \| \pi_{2t}).$$

Theorem: (Eckstein '2022) [Linear rate of convergence]

Suppose c is a lower semi-continuous cost function such that $|c(x,y)| \leq A\|x\|^p + B\|y\|^p$ for some $A, B, p > 0$.

and $\int e^{\|x\|^{p+\delta}} d\mu(x) < \infty$, $\int e^{\|y\|^{p+\delta}} d\nu(y) < \infty$.

Then. $$\| \pi_{2t} - \pi^*_{(\varepsilon=1)} \|_{TV} \leq D_1 \kappa^t$$

$$| G(\phi_t, \psi_t) - G(\phi^*, \psi^*) | \leq D_2 \kappa^t$$

for some $D_1, D_2 > 0$, $\kappa \in (0,1)$.

Remark:

Notice that the above theorem does not cover the case when $c(x,y) = \|x-y\|^2$ and μ, ν are Gaussian distribution. Chiarini et al. (2024) proved similar result for those case separately. They have optimal dependence of rate on ε.

Theorem: (Bounded cost case)

Suppose c is lower semi-continuous cost function, such that $|c(x,y)| < B$ for $B>0$ for all x, y,

Then
$$\|\phi_t - \phi^*_{(\varepsilon=1)}\|^2_{L^2(\mu)} + \|\psi_t - \psi^*_{(\varepsilon=1)}\|^2_{L^2(\nu)} \leq D_1 \kappa^t$$
$$|G(\phi_t, \psi_t) - G(\phi^*, \psi^*)| \leq D_2 \kappa^t$$
for some $D_1, D_2 > 0$ and $\kappa \in (0,1)$.

Proof: See section 6.5 of Marcel Nutz's survey Introduction to Entropic Optimal Transport.

We will try to understand why (ϕ_t, ψ_t) converges?

Following Lemma would be useful

Lemma: Fix $x \in \mathcal{X}$, $y \in \mathcal{Y}$

(i) For any $t > 0$.

$$\inf_{z \in \mathcal{Y}} \left[c(x,z) - \psi_t(z) \right] \leq \phi_{t+1}(x) \leq \int c(x,z)\, d\nu(z)$$

$$\inf_{z \in \mathcal{X}} \left[c(z,x) - \phi_t(x) \right] \leq \psi_t(y) \leq \int c(z,y)\, d\mu(z)$$

Fact:
$\int \phi_t \, d\mu \geq 0$
$\int \psi_t \, d\nu \geq 0$

Proof: Look at Schrödinger's equation

$$\phi_{t+1}(x) = -\log\left(\int e^{-c(x,z)+\psi_t(z)} d\nu(z) \right)$$

Since log is concave

$$\phi_{t+1}(x) \leq \int c(x,z)\, d\nu(z) - \int \psi_t(z)\, d\nu(z)$$

$$\leq \int c(x,z)\, d\nu(z)$$

On the other hand,

$$\int e^{-c(x,z)+\psi_t(z)} d\nu(z)$$

$$\leq e^{-\inf_{z \in \mathcal{Y}} \left[c(x,z) - \psi_t(z) \right]}$$

Hence $\phi_{t+1}(x) \geq \inf_{z \in \mathcal{Y}} \left[c(x,z) - \psi_t(z) \right]$

(ii) If c is bounded, then

$$-2\|c\|_\infty \leq \phi_{t+1}(x) \leq \|c\|_\infty$$

(iii) If $\int e^{\alpha c(x,y)} d\mu(x) d\nu(y) < \infty$ for some $\alpha > 0$, then

$$\int e^{\alpha \phi_{t+1}(x)} d\mu(x) \leq \underbrace{\int e^{\alpha c(x,y)} d\mu(x) d\nu(y)}_{< \infty}$$

$$\int e^{\alpha \psi_t(y)} d\nu(y) \leq \int e^{\alpha c(x,y)} d\mu(x) d\nu(y)$$

So when c is bdd and continuous, $(\phi_{t+1}, \phi_t, \psi_t)$ is an uniformly bounded sequece of continuous function. By Arzela - Ascoli, there is a convergent subsequence. Along that subsequence.

$$\lim_{t\to\infty} \psi_t(y) = -\log\left(\int e^{-c(z,y) + \lim_{t\to\infty} \phi_t(z)} d\mu(z)\right)$$

and

$$\lim_{t\to\infty} \phi_t(x) = -\log\left(\int e^{-c(x,z) + \lim_{t\to\infty}\psi_t(z)} d\nu(z)\right)$$

which are the equation satisfied by the schrödinger's potentials.

Hence all subsequence converges to same fixed point. Hence the sequence converges and in the limit satisfies Schrödinger's equation.

Linear Convergence for Bounded Cost Function.

Theorem: Let c be bounded cost function. and $(\varphi_t, \psi_t)_{t \geq 0}$ be the iterates of Sinkhorn algorithm. Let (φ_*, ψ_*) be the unique EOT potential s.t. $\lim_{t \to \infty} \int \varphi_t \, d\mu = \int \varphi_* \, d\mu$. Then.

1. $$G(\varphi_*, \psi_*) - G(\varphi_t, \psi_t) \leq \beta^t \left(G(\varphi_*, \psi_*) - G(\varphi_0, \psi_0) \right)$$

2. $$\| \varphi_* - \varphi_t \|^2_{L^2(\mu)} + \| \psi_* - \psi_t \|^2_{L^2(\nu)} \leq \beta_0 \beta^t \left(G(\varphi_*, \psi_*) - G(\varphi_0, \psi_0) \right).$$

where $\beta := 1 - e^{-24 \|c\|_\infty} \in (0,1)$.

$\beta_0 := 2 e^{6 \|c\|_\infty}$

(**Note:** If c is continuous in both variables and μ, ν are supported on bounded domain, then c is bounded.)

Step 1: (Use strong concavity of G).

Suppose $\phi, \phi' \in L^2(\mu)$, $\psi, \psi' \in L^2(\nu)$.

and $\phi \oplus \psi - c \geq -\alpha$ and $\phi' \oplus \psi' - c \geq -\alpha$

for some $\alpha \in \mathbb{R}$; then

Claim 1:

$$\boxed{\begin{aligned} & G(\phi', \psi') - G(\phi, \psi) \\ \geq & \int \partial_1 G(\phi', \psi')(x) [\phi'(x) - \phi(x)] \mu(dx) \\ & + \int \partial_2 G(\phi', \psi')(y) [\psi'(y) - \psi(y)] \mu(dy) \\ & + \frac{e^{-\alpha}}{2} \left(\|(\phi - \phi')\|^2_{L^2(\mu)} + \|\psi - \psi'\|^2_{L^2(\nu)} \right) \end{aligned}}$$

where

$$\partial_1 G(\phi, \psi) = 1 - \int e^{\phi(x) + \psi(y) - c(x,y)} \nu(dy)$$

$$\partial_2 G(\phi, \psi) = 1 - \int e^{\phi(x) + \psi(y) - c(x,y)} \mu(dx).$$

Note: Proof of the above claim follows from the observation.

$$e^b - e^a \geq (b-a) e^a + e^{-\alpha} |b-a|^2$$

$$a, b \in [-\alpha, \infty).$$

Step 2: If c is bdd, then.

$$\phi_t \oplus \psi_t - c \geq -6 \|c\|_\infty$$

$$\phi_t \oplus \psi_{t+1} - c \geq -6 \|c\|_\infty.$$

and

Step 3:

$$\boxed{\begin{aligned}&\text{Claim 2: Set } \sigma = e^{-6\|c\|_\infty},\\ &\text{Then, } G(\phi_t,\psi_t) - G(\phi_*,\psi_*)\\ &\geq -\frac{1}{2\sigma}\|\partial_2 G(\phi_t,\psi_t) - \partial_2 G(\phi_t,\psi_{t+1})\|^2\\ &\quad + \frac{\sigma}{2}\|\phi_t - \phi_*\|^2_{L^2(\mu)}\end{aligned}}$$

To see claim 2, apply claim 1 of Step 1.

$$\begin{aligned}&G(\phi_t,\psi_t) - G(\phi_*,\psi_*)\\ &\geq \int \partial_1 G(\phi_t,\psi_t)(x)\,[\phi_t(x) - \phi_*(x)]\,\mu(dx)\\ &\quad + \int \partial_2 G(\phi_t,\psi_t)(y)\,[\psi_t(y) - \psi_*(y)]\,\nu(dy)\\ &\quad + \frac{\sigma}{2}\left(\|\phi_t - \phi_*\|^2_{L^2(\mu)} + \|\psi_t - \psi_*\|^2_{L^2(\nu)}\right)\end{aligned}$$

Notice that

$$\partial_1 G(\phi_t,\psi_t) = 0 \quad \text{and} \quad \partial_2 G(\phi_t,\psi_{t+1}) = 0.$$

Hence

$$\begin{aligned}&G(\phi_t,\psi_t) - G(\phi_*,\psi_*)\\ &\geq \int \left(\partial_2 G(\phi_t,\psi_t)(y) - \partial_2 G(\phi_t,\psi_{t+1})(y)\right)\cdot\\ &\qquad\qquad \left(\psi_t(y) - \psi_*(y)\right)\nu(dy)\\ &\quad + \frac{\sigma}{2}\left(\|\phi_t - \phi_*\|^2_{L^2(\mu)} + \|\psi_t - \psi_*\|^2_{L^2(\nu)}\right)\end{aligned}$$

By Hölder's and Young's inequality.

$$\int \left(\partial_2 G(\phi_t,\psi_t)(y) - \partial_2 G(\phi_t,\psi_{t+1})(y)\right) \left(\psi_t(y) - \psi_*(y)\right) \nu(dy)$$

$$\geq -\frac{1}{2\sigma} \|\partial_2 G(\phi_t,\psi_t) - \partial_2 G(\phi_t,\psi_{t+1})\|^2_{L^2(\mu)} - \frac{\sigma}{2} \|\psi_t - \psi_*\|^2_{L^2(\mu)}$$

Combining we get Claim 2.

Step 4:

$$\boxed{\begin{array}{l}\text{Claim 3:}\\ \|\partial_2 G(\phi_t,\psi_t) - \partial_2 G(\phi_t,\psi_{t+1})\|^2_{L^2(\mu)} \leq \frac{1}{\sigma^2} \|\psi_t - \psi_{t+1}\|^2_{L^2(\nu)}\end{array}}$$

Expand $\partial_2 G(\phi_t,\psi_t) - \partial_2 G(\phi_t,\psi_{t+1})$ and notice that

$$\int \left| e^{\phi_t \oplus \psi_t - c} - e^{\phi_t \oplus \psi_{t+1} - c} \right| \mu(dx)$$

$$\leq e^{\sigma \|c\|_\infty} |\psi_t(y) - \psi_{t+1}(y)|$$

$$= \frac{1}{\sigma} |\psi_t(y) - \psi_{t+1}(y)|$$

Step 5:

$$\boxed{\begin{aligned}&\textbf{Claim 4:}\\ &G(\phi_{t+1},\psi_{t+1}) - G(\phi_t,\psi_t)\\ &\geq \frac{\sigma}{2}\Big(\|\phi_{t+1}-\phi_t\|^2_{L^2(\mu)} + \|\psi_{t+1}-\psi_t\|^2_{L^2(\nu)}\Big)\end{aligned}}$$

Use Claim 1 of Step 1.

Step 6:

Combining Claim 2, Claim 3 and Claim 4.

$$G(\phi_*,\psi_*) - G(\phi_t,\psi_t)$$

$$\leq \frac{1}{\sigma^4}\Big(G(\phi_{t+1},\psi_{t+1}) - G(\phi_t,\psi_t)\Big)$$

Set $\Delta_t = G(\phi_*,\psi_*) - G(\phi_t,\psi_t)$.

Then

$$\Delta_t \leq \frac{1}{\sigma^4}(\Delta_{t+1} - \Delta_t)$$

$$\Rightarrow \quad \Delta_{t+1} \leq (1-\sigma^4)\Delta_t \leq \cdots \leq (1-\sigma^4)^{t+1}\Delta_0$$

and hence,

$$\|\phi_* - \phi_t\|^2_{L^2(\mu)} + \|\psi_* - \psi_t\|^2_{L^2(\nu)} \leq \frac{1}{2\sigma}(1-\sigma^4)^t \Delta_0$$

by Claim 4 (with some tweak).