# Lecture 3 | Entropic OT

Recall: in high dimensions,

$$\mathbb{E} \sup_{\nu} |W_p(\mu_n,\nu) - W_p(\mu,\nu)| = \mathbb{E}\, W_p(\mu_n,\mu) \asymp n^{-1/d}.$$

What happens with regularized OT? Striking difference (Genevay et al.)!

Specialize to $\mu, \nu$ lying in $B_R(0) \subseteq \mathbb{R}^d$.

$$\mathbb{E} \sup_{\nu \in \mathcal{P}(B_R)} |S_\varepsilon(\mu_n,\nu) - S_\varepsilon(\mu,\nu)| \lesssim_{d,\varepsilon,R} n^{-1/2}$$

$$\left( \text{here } S_\varepsilon(\mu,\nu) := \inf_{\gamma \in \Gamma(\mu,\nu)} \int \tfrac{1}{2}\|x-y\|^2 \, d\gamma(x,y) + \varepsilon D(\gamma \| \mu \otimes \nu). \right)$$

Goal: discover why this is so. Note: only dual proof known!
primal proof OPEN (AFAIK)

For simplicity, assume $\varepsilon < 1$. ($\varepsilon > 1$ case HW)

Step 1: Reduction to $R = 1$ case.

$$\begin{aligned} S_\varepsilon(\mu,\nu) &= \inf_{\gamma \in \Gamma(\mu,\nu)} \int \tfrac{1}{2}\|x-y\|^2 \, d\gamma(x,y) + \varepsilon D(\gamma \| \mu \otimes \nu) \\ &= R^2 \cdot \left( \inf_{\gamma \in \Gamma(\mu,\nu)} \int \frac{1}{2} \frac{\|x-y\|^2}{R^2} \, d\gamma(x,y) + \frac{\varepsilon}{R^2} D(\gamma \| \mu \otimes \nu) \right) \end{aligned}$$

Let $\tilde{\mu}, \tilde{\nu}, \tilde{\gamma}$ be pushforwards of $\mu, \nu, \gamma$ by $x \mapsto \frac{x}{R}$.

Then $D(\tilde{\gamma} \| \tilde{\mu} \otimes \tilde{\nu}) = D(\gamma \| \mu \otimes \nu)$, so

$$S_\varepsilon(\mu,\nu) = R^2 S_{\varepsilon/R^2}(\tilde{\mu},\tilde{\nu}), \quad \text{where } \tilde{\mu}, \tilde{\nu} \in B_1.$$

**Step 2:** Bound via dual $\quad c := \frac{1}{2}\|\cdot - \cdot\|^2$

$$S(\mu,\nu) = \sup_{f,g} \int f\, d\mu + \int g\, d\nu - \varepsilon \int e^{\frac{f \oplus g - c}{\varepsilon}}\, d\mu\, d\nu + \varepsilon$$

So

$$S(\mu',\nu) - S(\mu,\nu) = \sup_{f',g'} \int f'\, d\mu' + \int g'\, d\nu - \varepsilon \int e^{\frac{f' \oplus g' - c}{\varepsilon}}\, d\mu' d\nu$$

$$- \sup_{f,g} \int f\, d\mu + \int g\, d\nu - \varepsilon \int e^{\frac{f \oplus g - c}{\varepsilon}}\, d\mu\, d\nu$$

$$\leq \sup_{f,g \in \mathcal{E}} \int f\,(d\mu' - d\mu) - \varepsilon \int e^{\frac{f \oplus g - c}{\varepsilon}}\,(d\mu' - d\mu)\, d\nu$$

as long as $\mathcal{E}$ is a set guaranteed to contain optimal solutions for both problems.

Recall opt conditions: can take $f(x) = -\varepsilon \log \int e^{\frac{g(y) - c(x,y)}{\varepsilon}}\, d\nu(y)$.

So for now let

$$\mathcal{E} := \left\{ f, g : \ f(x) = -\varepsilon \log \int e^{\frac{g(y) - c(x,y)}{\varepsilon}}\, d\nu(y) \right\}.$$

Then for all $(f,g) \in \mathcal{E}$,

$$\int e^{\frac{f(x) + g(y) - c(x,y)}{\varepsilon}}\, d\nu(y) = e^{\frac{f(x)}{\varepsilon}} \int e^{\frac{g(y) - c(x,y)}{\varepsilon}}\, d\nu(y) = 1.$$

So we obtain

$$\left| S(\mu_n,\nu) - S(\mu,\nu) \right| \leq \left| \sup_{f,g \in \mathcal{E}} \int f\,(d\mu_n - d\mu) \right|$$

Same as regular OT case. We want $\overline{\mathcal{E}} \supsetneq \mathcal{E}$ containing optimal duals.
Insight: can be <u>much</u> <u>smaller</u> than OT case.

**Step 3**: Regularity

For notational convenience, define

$$\varphi(x) = \frac{1}{2}\|x\|^2 - f(x)$$
$$\psi(y) = \frac{1}{2}\|y\|^2 - g(y)$$

$$f(x) + g(y) - \frac{1}{2}\|x-y\|^2 = \langle x, y\rangle - \varphi(x) - \psi(y)$$

$$\mathcal{E}' = \left\{ \varphi, \psi : \quad \varphi(x) = \varepsilon \log \int e^{\frac{\langle x,y\rangle - \psi(y)}{\varepsilon}} d\nu(y) \right\}$$

**Key claim**: For all $\varphi \in \mathcal{E}'$, $\varphi \in C^\infty$.

$$\|D^\alpha \varphi\|_\infty \lesssim \varepsilon^{-|\alpha|+1} \quad \forall \alpha, \ |\alpha| \geq 1.$$

**Warm up**:

$$\nabla \varphi(x) = \cancel{\varepsilon} \frac{\int \frac{y}{\cancel{\varepsilon}} e^{\frac{\langle x,y\rangle - \varphi(x) - \psi(y)}{\varepsilon}} d\nu(y)}{\int e^{\frac{\langle x,y\rangle - \psi(y)}{\varepsilon}} d\nu(y)}$$

$\to$ prob. measure on $B_1$.

$$\|\nabla \varphi(x)\| \leq 1 \quad \text{for all } x \in \mathbb{R}^d.$$

In general,

$$\left|D^\alpha\left(\exp \varphi(x)\right)\right| = \left|D^\alpha\left(\varepsilon \cdot \int e^{\frac{\langle x,y\rangle - \psi(y)}{\varepsilon}} d\nu(y)\right)\right|$$

$$= \varepsilon^{-|\alpha|+1} \left|\int y^\alpha e^{\frac{\langle x,y\rangle - \psi(y)}{\varepsilon}} d\nu(y)\right|$$

$$\leq \varepsilon^{-|\alpha|+1} \sup_{y \in B_1} |y^\alpha| \cdot \exp \varphi(x)$$

$$\leq \varepsilon^{-|\alpha|+1} \cdot \exp \varphi(x)$$

To get expression for $D^\alpha \varphi$, we use "Multivariate Faà di Bruno"

For $h : \mathbb{R} \to \mathbb{R}$, $j : \mathbb{R}^d \to \mathbb{R}$,

$$D^\alpha h(j(x)) = \sum_{v=1}^{|\alpha|} \frac{1}{v!} h^{(v)}(j(x)) \sum_{\substack{\beta_1 + \dots + \beta_v = \alpha \\ \beta \neq 0}} \prod_{\ell=1}^{v} D^{\beta_\ell} j(x)$$

Specializing to $h = \log$,

$$D^\alpha \log(j(x)) = \sum_{v=1}^{|\alpha|} \frac{(-1)^{v-1}}{v} \sum_{\beta_1 + \dots + \beta_v = \alpha} \prod_{\ell=1}^{v} \frac{D^{\beta_\ell} j(x)}{j(x)}$$

(proof on HW)

<u>Upshot:</u>

$$D^\alpha \varphi(x) = D^\alpha \log(\exp(\varphi(x)))$$

$$= \sum_{v=1}^{|\alpha|} \frac{(-1)^{v-1}}{v} \sum_{\beta_1 + \dots + \beta_v = \alpha} \prod_{\ell=1}^{v} \frac{D^{\beta_\ell} \exp(\varphi(x))}{\exp(\varphi(x))}$$

$$|D^\alpha \varphi(x)| \lesssim_\alpha \varepsilon^{-|\alpha|+1} . \qquad \square$$

<u>SO FAR:</u> dual solutions to entropic OT may be chosen to be <u>automatically smooth</u> (in fact, analytic!)

Contrast with dual analysis of $W_1$: functions are Lipschitz (but no better) "almost $C^1$" by Rademacher

$W_2$: Semi-convex (but no better) "almost $C^2$" by Alexandrov

Since $\sup_{(f,g) \in \bar{\mathcal{E}}} \int f (d\mu_n - d\mu)$ depends on size of $\bar{\mathcal{E}}$, we expect this to make a big difference.

One path "empirical process theory"

Our path Fourier analysis $\to$ RKHS.

**Step 4**: "Sobolev extension"

Let $\eta \in C^\infty$ be smooth cut off:

$$\eta \equiv 1 \text{ in } B_1$$
$$\eta \equiv 0 \text{ in } B_2^c.$$

Let $\bar{\varphi} = \varphi \cdot \eta$. Then $\| D^\alpha \bar{\varphi} \|_\infty \lesssim \varepsilon^{-|\alpha|+1}$ $\forall \alpha, |\alpha| \geq 1$.

Since $\bar{\varphi} \in C_c^\infty(\mathbb{R}^d)$, consider Fourier transform

$$\bar{\Phi}(\xi) := \int_{\mathbb{R}^d} \bar{\varphi}(x) e^{-2\pi i \langle x, \xi \rangle} dx$$

**Fact**:

$$(-1)^{|\alpha|} \xi^\alpha \bar{\Phi}(\xi) = \int_{\mathbb{R}^d} D^\alpha \bar{\varphi}(x) e^{-2\pi i \langle x, \xi \rangle} d\xi \qquad \text{(IBP)}$$

$$\int \xi^{2\alpha} |\bar{\Phi}(\xi)|^2 d\xi = \int |D^\alpha \bar{\varphi}(x)|^2 dx \qquad \text{(Parseval)}$$

$$\bar{\varphi}(x) = \int_{\mathbb{R}^d} \bar{\Phi}(\xi) e^{2\pi i \langle x, \xi \rangle} d\xi \qquad \text{(Inversion)}$$

Therefore

$$\int_{\mathbb{R}^d} (1 + \|\xi\|^2)^{|\alpha|} |\bar{\Phi}(\xi)|^2 d\xi \lesssim 1 + \int_{\mathbb{R}^d} |D^\alpha \bar{\varphi}(x)|^2 dx$$

$$\lesssim \varepsilon^{-2|\alpha|+2}.$$

"Sobolev extension: we had a function on $B$, with bounded derivatives — constructed an extension in Sobolev space $H^{|\alpha|}$.

HW

By interpolation, we obtain for any $s \geq 1$,

$$\int_{\mathbb{R}^d} (1+\|\xi\|^2)^s |\hat{\Phi}(\xi)|^2 \, d\xi \lesssim_s \varepsilon^{-2s+2}.$$

Write $B^s := \{ \varphi : \int_{\mathbb{R}^d} (1+\|\xi\|^2)^s |\hat{\Phi}(\xi)|^2 \, d\xi \}$.

We have shown that

$$\sup_{f,g \in \mathcal{E}} \int f \, d(\mu_n - \mu) = \int \tfrac{1}{2}\|\cdot\|^2 \, d(\mu_n - \mu) + \sup_{\varphi \in \mathcal{E}'} \int \varphi \, (d\mu - d\mu_n)$$

$$= \int \tfrac{1}{2}\|\cdot\|^2 \, d(\mu_n - \mu) + \sup_{\overline{\varphi} \in B^s} \int \overline{\varphi} \, (d\mu - d\mu_n)$$

$$= \underbrace{\int \tfrac{1}{2}\|\cdot\|^2 \, d(\mu_n - d\mu)}_{\mathbb{E}|\cdot| \lesssim n^{-1/2}} + \sup_{\overline{\varphi} \in B^s} \iint \hat{\Phi}(\xi) e^{2\pi i \langle x, \xi \rangle} \, d\xi \, d(\mu - \mu_n)(x)$$

So it suffices to show

$$\mathbb{E} \left| \sup_{\overline{\varphi} \in B^s} \iint \hat{\Phi}(\xi) e^{2\pi i \langle x, \xi \rangle} \, d\xi \, d(\mu - \mu_n)(x) \right| \lesssim_\varepsilon n^{-1/2}.$$

$$\mathbb{E}\left|\int \overline{\Phi}(\xi)\left(\int e^{2\pi i\langle x,\xi\rangle}\, d(\mu-\mu_n)(x)\right) d\xi\right|$$

$$\lesssim \mathbb{E}\left[\left(\int (1+\|\xi\|^2)^s|\overline{\Phi}(\xi)|^2\, d\xi\right)^{1/2}\left(\int (1+\|\xi\|^2)^{-s}\left|\int e^{2\pi i\langle x,\xi\rangle}\, d(\mu-\mu_n)(x)\right|^2 d\xi\right)^{1/2}\right]$$

$$\lesssim_s \varepsilon^{-s+1}\left(\int (1+\|\xi\|^2)^{-s}\,\underbrace{\mathbb{E}\left|\int e^{2\pi i\langle x,\xi\rangle}\, d(\mu-\mu_n)(x)\right|^2}_{\lesssim n^{-1}.} d\xi\right)^{1/2}$$

So, for any $s>0$,

$$\mathbb{E}\left|\sup_{\varphi\in B^s}\iint \overline{\Phi}(\xi)\, e^{2\pi i\langle x,\xi\rangle}\, d\xi\, d(\mu-\mu_n)(x)\right| \lesssim$$

$$\varepsilon^{-s+1}\cdot n^{-1/2}\cdot \underbrace{\left(\int (1+\|\xi\|^2)^{-s}\, d\xi\right)^{1/2}}_{<\infty \text{ for all } s>d/2.}$$

For any $s>d/2$, there exists $c_{s,d}>0$ such that

$$\mathbb{E}\sup_{\nu\in B_1}|S(\mu_n,\nu)-S(\mu,\nu)| \le c_{s,d}\,\varepsilon^{-s+1} n^{-1/2}. \quad \square$$

What is going on? Found we could choose dual solutions to lie in the Hilbert space $H^s$ for $s>d/2$:

this is space of functions $\varphi$ equipped with norm

$$\|\varphi\|_{H^s} := \left( \int (1+\|\xi\|^2)^s \, |\Phi(\xi)|^2 \, d\xi \right)^{1/2}$$

Key property:

$$|\varphi(x)| = \int \Phi(\xi) e^{2\pi i \langle x, \xi \rangle} d\xi$$

$$\lesssim \|\varphi\|_{H^s} \cdot \left( \int (1+\|\xi\|^2)^{-s} \right)^{1/2}$$

$$\lesssim \|\varphi\|_{H^s} \quad \text{if} \quad s > d/2.$$

So point evaluations are uniformly continuous. Statistical properties explored on HW.