Γ-convergence of entropically regularized optimal transport

Samir Chowdhury

March 29, 2018

The Ohio State University

Based on the 2017 paper "Convergence of Entropic Schemes for Optimal Transport" by Carlier, Duval, Peyre, Schmitzer

Review of entropically regularized optimal transport

Let $\mathcal{P}(\mathbb{R}^n)$ denote the set of Borel probability measures on $\mathbb{R}^n$.

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ (absolutely continuous w.r.t Lebesgue measure, finite second moments, finite entropy).

Let $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the squared Euclidean distance:

$$c(x, y) := \|x - y\|^2.$$

For any $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with density f_γ, write

$$\langle c, \gamma \rangle := \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) f_\gamma(x, y) \, dx \, dy.$$

Also write $H(\gamma)$ to denote the entropy of γ, and $\Pi(\mu, \nu)$ to denote the set of couplings between μ and ν.

The **entropically regularized OT** problem is to find

$$\underset{\gamma \in \Pi(\mu, \nu)}{\arg\min} \, \langle c, \gamma \rangle + \varepsilon H(\gamma).$$

The **entropically regularized OT** problem is to find

$$\operatorname*{arg\,min}_{\gamma \in \Pi(\mu,\nu)} \langle c, \gamma \rangle + \varepsilon H(\gamma).$$

Also define the 2-Wasserstein distance $W^2(\mu, \nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \langle c, \gamma \rangle$ and its entropy regularized variant $W_\varepsilon^2(\mu, \nu) := \inf_{\gamma \in \Pi(\mu,\nu)} (\langle c, \gamma \rangle + \varepsilon H_2(\gamma))$.

The motivation is that as $\varepsilon \to 0$, we will obtain a quantity that approximates the 2-Wasserstein distance between μ and ν.

More specifically, let $(\varepsilon_k)_{k \in \mathbb{N}}$ be a nonnegative sequence converging to 0, and let γ_k denote the minimizer (assume for now that such a minimizer always exists) corresponding to each $k \in \mathbb{N}$. Even if $\lim_k \gamma_k$ exists, does it minimize anything at all?

This question can be phrased as a question of convergence of functionals, and the appropriate notion of convergence turns out to be Γ-convergence.

Γ-convergence formulation

We use the **narrow/weak topology** on $\mathcal{P}(\mathbb{R}^m)$:

$$\mu_n \to \mu \text{ iff } \int f \, d\mu_n \to \int f \, d\mu$$

for all continuous, bounded test functions $f : \mathbb{R}^m \to \mathbb{R}$.

Today's talk will always involve differential entropy, which is the continuous analogue of discrete entropy. This is defined next.

Let $\mathcal{P}^r(\mathbb{R}^n)$ denote the elements of $\mathcal{P}(\mathbb{R}^n)$ that are absolutely continuous w.r.t Lebesgue measure. By the Radon-Nikodym theorem, for each $\mu \in \mathcal{P}^r(\mathbb{R}^n)$ there exists a density function f_μ such that for any Borel set $\sigma \subseteq \mathbb{R}^n$, we have:

$$\mu(\sigma) = \int_\sigma f_\mu(x) \, dx$$

Then for any $\mu \in \mathcal{P}(\mathbb{R}^m)$, we define the differential entropy as follows:

$$H_{\mathbb{R}^m}(\mu) := \begin{cases} \int_{\mathbb{R}^m} f_\mu(x) \log f_\mu(x)\, dx & : \mu \in \mathcal{P}^r(\mathbb{R}^m) \\ +\infty & : \text{otherwise.} \end{cases}$$

Notation: For any $\mu \in \mathcal{P}(\mathbb{R}^n)$, we denote entropy by $H_1(\mu)$, and for any $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$, we denote entropy by $H_2(\gamma)$.

Remark

In the discrete setting, the term above would always be negative. However, the differential entropy can have both positive and negative parts. For example, the uniform distribution on an interval $[0, a]$ *has density* $1/a$*, and if* $a < 1$*, then* $(1/a)\log(1/a) > 0$*.*

For any $\mu \in \mathcal{P}(\mathbb{R}^m)$, the second moment is defined as:

$$M_{\mathbb{R}^m}(\mu) := \int_{\mathbb{R}^m} \|x\|^2 \, d\mu(x)$$

We let $\mathcal{P}_2(\mathbb{R}^m)$ denote the elements of $\mathcal{P}(\mathbb{R}^m)$ with finite second moments.

Notation: For any $\mu \in \mathcal{P}(\mathbb{R}^n)$, we denote $M_{\mathbb{R}^n}$ by $M_1(\mu)$, and for any $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$, we denote $M_{\mathbb{R}^n \times \mathbb{R}^n}$ by $M_2(\gamma)$.

Next we set up the relevant functionals.

Let $\mu, \nu \in \mathcal{P}_2^r(\mathbb{R}^n)$, with $H_1(\mu) < \infty$, $H_1(\nu) < \infty$. Let $(\varepsilon_k)_k \in \mathbb{R}_+^{\mathbb{N}}$, $\varepsilon_k \to 0$.

Define $\mathcal{F}_k : \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{R} \cup \{\infty\}$ by:

$$\gamma \mapsto \begin{cases} \langle c, \gamma \rangle + \varepsilon_k H_2(\gamma) & : \gamma \in \Pi(\mu, \nu) \\ +\infty & : \text{otherwise.} \end{cases}$$

Also define $\mathcal{F} : \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) \to \mathbb{R} \cup \{\infty\}$ by:

$$\gamma \mapsto \begin{cases} \langle c, \gamma \rangle & : \gamma \in \Pi(\mu, \nu) \\ +\infty & : \text{otherwise.} \end{cases}$$

Definition

We say $\mathcal{F}_k$ gamma-converges to $\mathcal{F}$ (denoted $\mathcal{F}_k \xrightarrow{\Gamma} \mathcal{F}$) w.r.t. the narrow topology if for all $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$, the following are satisfied:

- (Liminf condition/lower bound) For any sequence $(\gamma_k)_k$ in $\mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ such that $\gamma_k \xrightarrow{narrow} \gamma$, we have

$$\mathcal{F}(\gamma) \leq \liminf_{k\to\infty} \mathcal{F}_k(\gamma_k).$$

- (Limsup condition/recovery sequence) There exists a sequence $(\gamma_k)_k$ in $\mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ such that $\gamma_k \xrightarrow{narrow} \gamma$ and

$$\mathcal{F}(\gamma) \geq \limsup_{k\to\infty} \mathcal{F}_k(\gamma_k).$$

Theorem

The sequence $(\mathcal{F}_k)_k$ Γ-converges to $\mathcal{F}$ with respect to the narrow topology.

A function $f : X \to \mathbb{R}$ is **coercive** if $\overline{\{f \leq t\}}$ is countably compact for each $t \in \mathbb{R}$.

Note that $\mathcal{F}$ is coercive, by precompactness of $\Pi(\mu, \nu)$.

A sequence $(f_k)_k$ of functions $X \to \mathbb{R}$ is **equi-coercive** if for each $t \in \mathbb{R}$, there exists a compact set $K_t \subseteq X$ such that $\{f_k \leq t\} \subseteq K_t$ for all $k \in \mathbb{N}$.

$(\mathcal{F}_k)_k$ is **equi-coercive** by precompactness of $\Pi(\mu, \nu)$.

Fact: since $(\mathcal{F}_k)_k$ is equi-coercive and Γ-converges to $\mathcal{F}$, the minimizers μ_k of $\mathcal{F}_k$ converge narrowly to the minimizer for $\mathcal{F}$.

Proof of Γ-convergence: Limsup condition

The longer part of the proof is to build the recovery sequence. The following technique is the crux of the argument.

Definition (Block Approximation)

Let μ, $\nu \in \mathcal{P}_2^r(\mathbb{R}^n)$, $H_1(\mu)$, $H_1(\nu) < \infty$, $\gamma \in \Pi(\mu, \nu)$. Given $k = (k_1, \ldots, k_n) \in \mathbb{Z}^n$, define $Q_k := [k_1, k_1 + 1) \times \ldots \times [k_n, k_n + 1) \subset \mathbb{R}^n$, and for $\ell > 0$, write $Q_k^\ell := \ell \cdot Q_k$. Define the *block approximation* of γ at scale ℓ by

$$\gamma_\ell := \sum_{(j,k)\in(\mathbb{Z}^n)^2} \gamma\left(Q_j^\ell \times Q_k^\ell\right)(\mu_{j,\ell} \otimes \nu_{k,\ell}),$$

where for every Borel set $\sigma \subset \mathbb{R}^n$,

$$\mu_{j,\ell}(\sigma) := \begin{cases} \frac{\mu(\sigma\cap Q_j^\ell)}{\mu(Q_j^\ell)} & \text{if } \mu(Q_j^\ell) > 0, \\ 0 & \text{otherwise.} \end{cases} \qquad \nu_{k,\ell}(\sigma) := \begin{cases} \frac{\nu(\sigma\cap Q_j^\ell)}{\nu(Q_k^\ell)} & \text{if } \nu(Q_k^\ell) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Fact: γ_ℓ is a Borel probability measure with density

$$\gamma_\ell(x, y) = \begin{cases} \gamma(Q_j^\ell \times Q_k^\ell) \frac{f_\mu(x)\, f_\nu(y)}{\mu(Q_j^\ell)\, \nu(Q_k^\ell)} & \text{if } \mu(Q_j^\ell) > 0 \text{ and } \nu(Q_k^\ell) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $(j, k) \in (\mathbb{Z}^n)^2$ is uniquely determined by $(x, y) \in Q_j^\ell \times Q_k^\ell$.

Intuition: the block approximation is a "separation of variables" trick, allowing us to write γ_ℓ in the form $f_\mu \cdot f_\nu$. This allows us to apply tools such as Fubini's theorem, and also to apply logarithm rules to separate terms.

Proposition

For $\gamma \in \Pi(\mu, \nu)$ the block approximation γ_ℓ is also in $\Pi(\mu, \nu)$.

Proof.

For any Borel set $\sigma \subset \mathbb{R}^n$, $(\mu_{j,\ell} \otimes \nu_{k,\ell})(\mathbb{R}^n \times \sigma) = \nu_{k,\ell}(\sigma)$ if $\mu(Q_j^\ell) > 0$, and 0 otherwise.

Also, $\gamma\left(Q_j^\ell \times Q_k^\ell\right) = 0$ if $\mu(Q_j^\ell) = 0$. Thus

$$\begin{aligned}
\gamma_\ell(\mathbb{R}^n \times \sigma) &= \sum_{(j,k)\in(\mathbb{Z}^n)^2} \gamma\left(Q_j^\ell \times Q_k^\ell\right)(\mu_{j,\ell} \otimes \nu_{k,\ell})(\mathbb{R}^n \times \sigma) \\
&= \sum_{\substack{(j,k)\in(\mathbb{Z}^n)^2: \\ \mu(Q_j^\ell)>0}} \gamma\left(Q_j^\ell \times Q_k^\ell\right)\nu_{k,\ell}(\sigma) \\
&= \sum_{\substack{(j,k)\in(\mathbb{Z}^n)^2: \\ \mu(Q_j^\ell)>0}} \gamma\left(Q_j^\ell \times Q_k^\ell\right)\nu_{k,\ell}(\sigma) + \sum_{\substack{(j,k)\in(\mathbb{Z}^n)^2: \\ \mu(Q_j^\ell)=0}} \gamma\left(Q_j^\ell \times Q_k^\ell\right)\nu_{k,\ell}(\sigma) \\
&= \sum_{k\in\mathbb{Z}^n} \nu_{k,\ell}(\sigma) \underbrace{\sum_{j\in\mathbb{Z}^n} \gamma\left(Q_j^\ell \times Q_k^\ell\right)}_{=\nu(Q_k^\ell)} = \nu(\sigma). \qquad \square
\end{aligned}$$

Lemma (A geometric perturbation lemma)

Let $\{Q_i\}_{i\in I}$ be a countable partition of $\mathbb{R}^n$ into Borel sets with $\sup_{i\in I}\operatorname{diam}(Q_i) \le C < \infty$, i.e. $\|x-y\|^2 \le C^2$ for x, $y \in Q_i$ for any $i \in I$. Let μ, $\nu \in \mathcal{P}(\mathbb{R}^n)$ be such that $\mu(Q_i) = \nu(Q_i)$ for all $i \in I$.

Then $W^2(\mu,\nu) \le C^2$.

Proof

Denote by $\hat{I}$ the subset of I such that $\mu(Q_i) = \nu(Q_i) > 0$ for $i \in \hat{I}$.

For $i \in \hat{I}$ and every Borel $\sigma \subset \mathbb{R}^n$ define

$$\mu_i(\sigma) := \frac{\mu(\sigma \cap Q_i)}{\mu(Q_i)}$$

Define ν_i analogously.

Then μ_i, $\nu_i \in \mathcal{P}(\mathbb{R}^n)$, with support contained in $\overline{Q_i}$.

For every $i \in \hat{I}$ let $\gamma_i \in \Pi(\mu_i,\nu_i)$. Then $\operatorname{supp}\gamma_i \subset \overline{Q_i}^2$ and so

$$\langle c,\gamma_i\rangle = \int_{R^n\times\mathbb{R}^n} \|x-y\|^2\, d\gamma_i(x,y) = \int_{Q_i\times Q_i} \|x-y\|^2\, d\gamma_i(x,y) \le C^2\,.$$

Define $\gamma := \sum_{i\in\hat{I}} \mu(Q_i)\,\gamma_i$.

Then $\gamma \in \Pi(\mu,\nu)$. Here we use the fact that $\{Q_i\}_i$ is a partition of $\mathbb{R}^n$.

We then have

$$W^2(\mu,\nu) \le \langle c,\gamma\rangle = \sum_{i\in\hat{I}} \mu(Q_i)\,\langle c,\gamma_i\rangle \le \sum_{i\in\hat{I}} \mu(Q_i)\,C^2 = C^2\,. \qquad \square$$

The intuition behind this lemma is that it gives a quantitative bound on the quality of the block approximation. More specifically, we get the following result:

Corollary

Let μ, $\nu \in \mathcal{P}_2^r(\mathbb{R}^n)$. For $\gamma \in \Pi(\mu,\nu)$ and its block approximation γ_ℓ,

$$W^2(\gamma,\gamma_\ell) \le 2\,n\,\ell^2 \tag{2}$$

and $\gamma_\ell \xrightarrow{\text{narrow}} \gamma$ as $\ell \to 0^+$.

Proof sketch: An m-cube with side length ℓ has $\mathrm{diam}^2 = m\ell^2$. We are in the case of $\mathbb{R}^n \times \mathbb{R}^n$, so $m = 2n$.

Convergence in the 2-Wasserstein space implies convergence in integration against all test functions growing at most as $x \mapsto \|x\|^2$ (see Villani OT book, Def 6.8). So for any φ such that $|\varphi(z)| \leq \|z\|^2$, we have

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \varphi(z) \, d\gamma_\ell(x, y) \to \int_{\mathbb{R}^n \times \mathbb{R}^n} \varphi(z) \, d\gamma.$$

Corollary

The transport cost of the block approximation converges:

$$\lim_{\ell \to 0^+} \langle c, \gamma_\ell \rangle = \langle c, \gamma \rangle \,.$$

Proof sketch: By the parallelogram law, the map $(x, y) \mapsto \|x - y\|^2$ is bounded above by $2\|x\|^2 + 2\|y\|^2 = 2\|(x, y)\|^2$.

It remains to control the entropy of the block approximation. The following lemmas will turn out to be useful:

Lemma

There exists a constant $C > 0$ *and an exponent* $0 < \alpha < 1$ *(both depending on* n*) such that for every* $\mu \in \mathcal{P}_2^r(\mathbb{R}^n)$ *one has*

$$H_{\mathbb{R}^n}(\mu) \geq -C\,(M_{\mathbb{R}^n}(\mu)+1)^{\alpha}\,, \qquad H_{neg,\mathbb{R}^n}(\mu) \leq C\,(M_{\mathbb{R}^n}(\mu)+1)^{\alpha}\,.$$

(We omit the proof.)

Lemma

There exist constants $C > 0$, $\alpha \in (0,1)$ *such that for* $\mu \in \mathcal{P}(\mathbb{R}^n)$,

$$\sum_{j\in\mathbb{Z}^n} \mu(Q_j^\ell) \log\left(\mu(Q_j^\ell)\right) \geq -C(M_1(\mu) + n\,\ell^2 + 1)^\alpha + n\,\log(\ell)\,.$$

Proof

Consider the function

$$f_{\mu_\ell}(x) := \begin{cases} \frac{1}{\ell^n}\mu(Q_j^\ell) & : x \in Q_j^\ell \\ 0 & : \text{ otherwise.} \end{cases}$$

Then f_{μ_ℓ} is the density of a probability measure μ_ℓ. In particular, $\mu_\ell(Q_j^\ell) = \mu(Q_j^\ell)$.

We have

$$\begin{aligned}
\sum_{j\in\mathbb{Z}^n} \mu(Q_j^\ell)\log\left(\mu(Q_j^\ell)\right) &= \sum_{j\in\mathbb{Z}^n}\int_{Q_j^\ell} f_{\mu_\ell}(x)\log(\mu(Q_j^\ell))\,dx \\
&= \sum_{j\in\mathbb{Z}^n}\int_{Q_j^\ell} f_{\mu_\ell}(x)\log\left(\mu_\ell(x)\cdot\ell^n\right)\,dx \\
&= \int_{\mathbb{R}^n} f_{\mu_\ell}(x)\log\left(f_{\mu_\ell}(x)\cdot\ell^n\right) \\
&= H_1(\mu_\ell) + n\,\log(\ell) \geq -K(M(\mu_\ell)+1)^\alpha + n\,\log(\ell)
\end{aligned}$$

for suitable constants $K > 0$, $\alpha \in (0,1)$ by the previous lemma.

Now $\mathrm{diam}(Q_j^\ell) = \sqrt{n}\,\ell$ for all $j \in \mathbb{Z}^n$, so a previous result shows that $W^2(\mu,\mu_\ell) \leq n\,\ell^2$.

Fact: Since $\|y\|^2 \leq 2\|x\|^2 + 2\|x-y\|^2$, we have the moment bound $M_1(\nu) \leq 2M_1(\mu) + 2W^2(\mu,\nu)$.

Using the moment bound above, we replace the $M_1(\mu_\ell)$ term by a $M_1(\mu) + n\ell^2$ term and get the desired result for some adjusted constant $C > 0$. □

The next proposition bounds the entropy of the block approximation.

Proposition (Bounding Entropy of Block Approximation)

There are constants $C > 0$ and $\alpha \in (0, 1)$ such that the entropy of the block approximation γ_ℓ of $\gamma \in \Pi(\mu, \nu)$ at scale $\ell > 0$ is bounded by

$$H_2(\gamma_\ell) \leq H_1(\mu) + H_1(\nu) + C \left((M(\mu) + n\,\ell^2 + 1)^\alpha + (M(\nu) + n\,\ell^2 + 1)^\alpha - 2\,n\,\log(\ell) \right) .$$

Proof.

$$H_2(\gamma_\ell) = \sum_{\substack{(j,k)\in(\mathbb{Z}^n)^2:\\ \mu(Q_j^\ell)>0,\, \nu(Q_k^\ell)>0}} \int_{Q_j^\ell \times Q_k^\ell} \gamma(Q_j^\ell \times Q_k^\ell) \frac{f_\mu(x)\, f_\nu(y)}{\mu(Q_j^\ell)\,\nu(Q_k^\ell)} \log\left(\gamma(Q_j^\ell \times Q_k^\ell)\frac{f_\mu(x)\, f_\nu(y)}{\mu(Q_j^\ell)\,\nu(Q_k^\ell)}\right) dx\, dy$$

$$= \sum_{\substack{(j,k)\in(\mathbb{Z}^n)^2:\\ \mu(Q_j^\ell)>0,\, \nu(Q_k^\ell)>0}} \gamma(Q_j^\ell \times Q_k^\ell) \Bigg[\underbrace{\log\left(\gamma(Q_j^\ell \times Q_k^\ell)\right)}_{\leq 0} + \int_{Q_j^\ell} \frac{f_\mu(x)}{\mu(Q_j^\ell)} \log\left(\frac{f_\mu(x)}{\mu(Q_j^\ell)}\right) dx$$

$$+ \int_{Q_k^\ell} \frac{f_\nu(y)}{\nu(Q_k^\ell)} \log\left(\frac{f_\nu(y)}{\nu(Q_k^\ell)}\right) dy\Bigg]$$

$$\leq H_1(\mu) + H_1(\nu) - \sum_{j\in\mathbb{Z}^n} \mu(Q_j^\ell) \log\left(\mu(Q_j^\ell)\right) - \sum_{k\in\mathbb{Z}^n} \nu(Q_k^\ell) \log\left(\nu(Q_k^\ell)\right).$$

The second equality follows from properties of logarithms and couplings.

The inequality follows by using properties of couplings and log on the latter two terms. The first term is bounded above by 0, and disappears.

From the last step, the result follows by using one of the previous lemmas on the two negative terms.

□

Finally, we have:

Proposition (Limsup Condition)

For every $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ there is a non-negative sequence $(\ell_k)_{k\in\mathbb{N}}$ converging to zero, such that

$$\mathcal{F}(\gamma) \geq \limsup_{k\to\infty} \mathcal{F}_k(\gamma_{\ell_k})\,.$$

Proof sketch.

Let $\gamma \in \Pi(\mu,\nu)$. We saw previously that $\gamma_{\ell_k} \in \Pi(\mu,\nu)$.

By the previous proposition, we have

$$H_2(\gamma_{\ell_k}) \leq H_1(\mu) + H_1(\nu) + C\left((M(\mu) + n\,\ell_k^2 + 1)^\alpha + (M(\nu) + n\,\ell_k^2 + 1)^\alpha - 2\,n\,\log(\ell_k)\right).$$

For $\ell_k = \varepsilon_k$, we have $\limsup_{k\to+\infty} \varepsilon_k\, H_2(\gamma_{\ell_k}) \leq 0$. Here we used the fact that $\varepsilon_k\,\log(\ell_k) \to 0$ as $k \to \infty$. Thus

$$\limsup_{k\to\infty} \mathcal{F}_k(\gamma_{\ell_k}) = \limsup_{k\to\infty} \left\langle c, \gamma_{\ell_k}\right\rangle + \varepsilon_k H_2(\gamma_{\ell_k}) \leq \lim_{k\to\infty} \left\langle c, \gamma_{\ell_k}\right\rangle = \langle c, \gamma\rangle = \mathcal{F}(\gamma). \quad \square$$

Liminf condition

Proposition (Liminf Condition)

Let $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ *and a sequence* $(\gamma_k)_{k\in\mathbb{N}}$ *in* $\mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ *such that* $\gamma_k \to \gamma$ *narrowly. Then*

$$\mathcal{F}(\gamma) \leq \liminf_{k\to\infty} \mathcal{F}_k(\gamma_k)\,.$$

Proof sketch.

By a previous lemma, there is a finite constant $C < \infty$ such that $H_{neg,2}(\gamma_k) \leq C$. Using this, we verify $\liminf_{k\to\infty} \varepsilon_k\, H_2(\gamma_k) \geq 0$. By narrow lower semicontinuity of $\gamma \mapsto (c, \gamma)$ (see Villani OT book), we have:

$$\mathcal{F}(\gamma) \leq \liminf_{k\to\infty} \mathcal{F}(\gamma_k) \leq \liminf_{k\to\infty} \mathcal{F}(\gamma_k) + \varepsilon_k H_2(\gamma_k) = \liminf_{k\to\infty} \mathcal{F}_k(\gamma_k).$$

□

Summary

- The Limsup and Liminf conditions show that $\mathcal{F}_k$ Γ-converges to $\mathcal{F}$ with respect to the narrow topology.
- Minimizers of $\mathcal{F}_k$ converge narrowly to the minimizer for $\mathcal{F}$.
- This gives a theoretical justification for computationally replacing the OT problem by its entropically regularized variant.
- Writing "The" minimizer above because the OT plan in our context is unique (**Brenier's Polar Factorization Theorem**, we'll see this later).
- Bulk of proof lay in proving properties of the block approximation, which is a natural way to approximate any coupling.
- Extension: for a natural extension to Polish spaces, see Léonard's 2011 paper.