Report for CSE 5339 2018 — (OTMLSA)
Optimal Transport in Machine Learning and Shape Analysis

Γ-convergence of Entropy Regularized Wasserstein Distance

Samir Chowdhury

1 Introduction

In this paper, we are interested in the convergence of the solution to the entropy-regularized optimal transport (OT) problem to the non-regularized OT problem, as described in [CDPS17]. All these results will be in the case of Euclidean space.

1.1 Background and notation

We write $\mathcal{P}(\mathbb{R}^m)$ to denote the set of Borel probability measures on $\mathbb{R}^n$. Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^m)$, we write $\Pi(\mu,\nu)$ to denote the set of couplings between μ and ν. We use the *narrow/weak topology* on $\mathcal{P}(\mathbb{R}^m)$: for any sequence of measures $(\mu_n)_n$, we have

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

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

We write $\mathcal{P}^r(\mathbb{R}^n)$ denote the elements of $\mathcal{P}(\mathbb{R}^n)$ that are absolutely continuous with respect to 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 (negative) differential entropy as follows:

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

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 1. In the discrete setting, the discrete entropy analogue of 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 $(1/a)\log(1/a) > 0$.

The Euclidean norm is denoted $\|\cdot\|$. 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.

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 $M_2(\gamma)$.

2 Statement of the main result

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$. Suppose also that they are absolutely continuous with respect to Lebesgue measure, have finite second moments, and have 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

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

The OT problem is to find:

$$\underset{\gamma \in \Pi(\mu,\nu)}{\arg\min} \langle c, \gamma \rangle . \tag{1}$$

Here we are guaranteed the existence of a minimizer by compactness of $\Pi(\mu, \nu)$ and by lower semicontinuity of the map $\gamma \mapsto \langle c, \gamma \rangle$. Compactness of $\Pi(\mu, \nu)$ follows from Prokhorov' theorem and by verifying that $\Pi(\mu, \nu)$ is closed.

Let $\varepsilon > 0$. The *ε-entropy regularized OT* problem is to find

$$\underset{\gamma \in \Pi(\mu,\nu)}{\arg\min} \langle c, \gamma \rangle + \varepsilon H(\gamma). \tag{2}$$

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

The motivation for studying the ε-entropy regularized problem 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 of Equation 2 corresponding to each $k \in \mathbb{N}$. We wish to show that these minimizers converge to a minimizer for Equation 1. This property can be seen as a consequence of the Γ-convergence of certain functionals that we define next.

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 1. We say $\mathcal{F}_k$ Γ-converges to $\mathcal{F}$ (denoted $\mathcal{F}_k \xrightarrow{\Gamma} \mathcal{F}$) with respect to the 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
$$\mathcal{F}(\gamma) \leqslant \liminf_{k \to \infty} \mathcal{F}_k(\gamma_k).$$
- (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
$$\mathcal{F}(\gamma) \geqslant \limsup_{k \to \infty} \mathcal{F}_k(\gamma_k).$$

To motivate the main result, we first list some facts involving coercivity of functions. A function $f : X \to \mathbb{R}$ is *coercive* if $\overline{\{f \leqslant t\}}$ 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 $\{f_k \leqslant t\} \subseteq K_t$ for all $k \in \mathbb{N}$. $(\mathcal{F}_k)_k$ is *equi-coercive* by compactness of $\Pi(\mu, \nu)$. The next remark connects the notions of Γ-convergence and equicoercivity.

Remark 2. Since $(\mathcal{F}_k)_k$ is equi-coercive and Γ-converges to $\mathcal{F}$, then $\lim_k \inf \mathcal{F}_k = \inf \mathcal{F}$ and any cluster point of the sequence of minimizers μ_k of $\mathcal{F}_k$ is a minimizer for $\mathcal{F}$. In particular, since there is a unique minimizer of $\mathcal{F}$ via a result of Brenier, the sequence $(\mu_k)_k$ converges narrowly to the unique minimizer for $\mathcal{F}$.

Finally, we state the main result described in this paper.

Theorem 3. *The sequence $(\mathcal{F}_k)_k$ Γ-converges to $\mathcal{F}$ with respect to the narrow topology.*

3 Proof of the main result

We begin with the longer part of the proof, which is to build the recovery sequence.

Definition 2 (Block Approximation)**.** Let $\mu,\ \nu \in \mathcal{P}_2^r(\mathbb{R}^n)$, $H_1(\mu),\ H_1(\nu) < \infty$, $\gamma \in \Pi(\mu, \nu)$. 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 $\sigma \subset \mathbb{R}^n$,

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

An application of the $\pi - \lambda$ theorem along with verification on measurable rectangles shows that γ_ℓ is a Borel probability measure with the following density:

$$f_{\gamma_\ell}(x, y) = \begin{cases} \gamma(Q_j^\ell \times Q_k^\ell)\frac{f_\mu(x)\, f_\nu(y)}{\mu(Q_j^\ell)\,\nu(Q_k^\ell)} & \text{if } \mu(Q_j^\ell) > 0 \text{ and } \nu(Q_k^\ell) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

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

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

The next proposition shows that the block approximation of a coupling is still a coupling.

Proposition 4. *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

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

□

The next lemma gives a quantitative bound on the quality of the block approximation.

Lemma 5 (A geometric perturbation lemma). *Let $\{Q_i\}_{i\in I}$ be a countable partition of $\mathbb{R}^n$ into Borel sets with $\sup_{i\in I} \operatorname{diam}(Q_i) \leqslant C < \infty$, i.e. $\|x-y\|^2 \leqslant C^2$ 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 $W^2(\mu,\nu) \leqslant C^2$.

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

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

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

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

□

The preceding lemma is used to give the following specific bound on the quality of the block approximation. Note the dependence on the dimension of the ambient space.

Corollary 6. *Let μ, $\nu \in \mathcal{P}_2^r(\mathbb{R}^n)$. For $\gamma \in \Pi(\mu,\nu)$ and its block approximation γ_ℓ,*

$$W^2(\gamma,\gamma_\ell) \leqslant 2\,n\,\ell^2 \tag{4}$$

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

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

Corollary 7. *The transport cost of the block approximation converges:*

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

Proof. The previous result shows that $W^2(\gamma_k, \gamma) \to 0$ as $\ell \to 0$. Recall that convergence in the 2-Wasserstein space implies convergence in integration against all test functions growing at most as $x \mapsto \|x\|^2$ [Vil08, Definition 6.8]. So for any φ such that $|\varphi(z)| \leqslant \|z\|^2$, we have

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

Now observe that by the parallelogram law, the map $(x, y) \mapsto \|x - y\|^2$ is bounded above by $2\|x\|^2 + 2\|y\|^2 = 2\|(x, y)\|^2$. Plugging this into the integral above gives the result. $\square$

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

Lemma 8. *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*

$$H_{\mathbb{R}^n}(\mu) \geqslant -C\left(M_{\mathbb{R}^n}(\mu) + 1\right)^\alpha, \qquad H_{neg,\mathbb{R}^n}(\mu) \leqslant C\left(M_{\mathbb{R}^n}(\mu) + 1\right)^\alpha .$$

Proof. See [JKO98, Proposition 4.1]. $\square$

Lemma 9. *There exist constants $C > 0$, $\alpha \in (0, 1)$ such that for $\mu \in \mathcal{P}(\mathbb{R}^n)$,*

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

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

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

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

Now $\operatorname{diam}(Q_j^\ell) = \sqrt{n}\,\ell$ for all $j \in \mathbb{Z}^n$, so a previous result shows that $W^2(\mu, \mu_\ell) \leqslant n\,\ell^2$. Since $\|y\|^2 \leqslant 2\|x\|^2 + 2\|x - y\|^2$, we have the moment bound $M_1(\nu) \leqslant 2M_1(\mu) + 2W^2(\mu, \nu)$. 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 $C > 0$. $\square$

The next proposition bounds the entropy of the block approximation.

Proposition 10 (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*

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

Proof.

$$\begin{aligned} H_2(\gamma_\ell) &= \sum_{\substack{(j,k)\in(\mathbb{Z}^n)^2:\\ \mu(Q_j^\ell)>0,\,\nu(Q_k^\ell)>0}} \int_{Q_j^\ell\times Q_k^\ell} \gamma(Q_j^\ell \times Q_k^\ell)\frac{f_\mu(x)\,f_\nu(y)}{\mu(Q_j^\ell)\,\nu(Q_k^\ell)} \log\left(\gamma(Q_j^\ell \times Q_k^\ell)\frac{f_\mu(x)\,f_\nu(y)}{\mu(Q_j^\ell)\,\nu(Q_k^\ell)}\right) dx\,dy \\ &= \sum_{\substack{(j,k)\in(\mathbb{Z}^n)^2:\\ \mu(Q_j^\ell)>0,\,\nu(Q_k^\ell)>0}} \gamma(Q_j^\ell \times Q_k^\ell)\left[\underbrace{\log\left(\gamma(Q_j^\ell \times Q_k^\ell)\right)}_{\leqslant 0} + \int_{Q_j^\ell}\frac{f_\mu(x)}{\mu(Q_j^\ell)}\log\left(\frac{f_\mu(x)}{\mu(Q_j^\ell)}\right)dx\right. \\ &\qquad\qquad \left. + \int_{Q_k^\ell}\frac{f_\nu(y)}{\nu(Q_k^\ell)}\log\left(\frac{f_\nu(y)}{\nu(Q_k^\ell)}\right)dy\right] \\ &\leqslant H_1(\mu) + H_1(\nu) - \sum_{j\in\mathbb{Z}^n}\mu(Q_j^\ell)\log\left(\mu(Q_j^\ell)\right) - \sum_{k\in\mathbb{Z}^n}\nu(Q_k^\ell)\log\left(\nu(Q_k^\ell)\right). \end{aligned}$$

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 11 (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*

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

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

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

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

$$\limsup_{k\to\infty}\mathcal{F}_k(\gamma_{\ell_k}) = \limsup_{k\to\infty}\langle c, \gamma_{\ell_k}\rangle + \varepsilon_k H_2(\gamma_{\ell_k}) \leqslant \lim_{k\to\infty}\langle c, \gamma_{\ell_k}\rangle = \langle c, \gamma\rangle = \mathcal{F}(\gamma).$$ ☐

The lower bound condition for Γ-convergence is much easier.

Proposition 12 (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*

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

Proof sketch. By a previous lemma, there is a finite constant $C < \infty$ such that $H_{neg,2}(\gamma_k) \leqslant C$. Using this, we verify $\liminf_{k\to\infty}\varepsilon_k\,H_2(\gamma_k) \geqslant 0$. By narrow lower semicontinuity of $\gamma \mapsto (c,\gamma)$ [Vil08], we have:

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

The main result now follows by combining the previous two propositions.

4 Discussion

The bulk of the work above lay in proving $\limsup_{k\to+\infty} \varepsilon_k\, H_2(\gamma_{\ell_k}) \leqslant 0$. The block approximation is an interesting and suitable way to approximate any coupling in the Euclidean case.

References

[CDPS17] Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.