Calculus of Variations

Multi-marginal entropy-transport with repulsive cost

Augusto Gerolin[1] · Anna Kausamo[2] · Tapio Rajala[2]

Abstract

In this paper we study theoretical properties of the entropy-transport functional with repulsive cost functions. We provide sufficient conditions for the existence of a minimizer in a class of metric spaces and prove the Γ-convergence of the entropy-transport functional to a multi-marginal optimal transport problem with a repulsive cost. We point out that our construction can deal with the case when the space X is a domain in $\mathbb{R}^d$, answering a question raised in Benamou et al. (Numer Math 142:33–54, 2019). Finally, we also prove the entropy-regularized version of the Kantorovich duality.

1 Introduction

We consider the following multi-marginal entropy-transport problem

$$I_\varepsilon[\rho] = \inf_{\gamma \in \Pi_N^{\mathrm{sym}}(\rho)} (C_0[\gamma] + \varepsilon E[\gamma]), \tag{1.1}$$

where $C_0[\gamma] = \int_{X^N} c \, d\gamma$ is the transportation cost related to a cost function c, $E[\gamma]$ is the entropy, and $\varepsilon \geq 0$ is a parameter, see Sect. 2 for details. We consider the setting where $(X, d, \mathfrak{m})$ is a Polish measure space, and $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ is an absolutely continuous

✉ Augusto Gerolin
  augustogerolin@gmail.com

  Anna Kausamo
  anna.m.kausamo@jyu.fi

  Tapio Rajala
  tapio.m.rajala@jyu.fi

[1] Department of Theoretical Chemistry, Vrije Universiteit Amsterdam, FEW, De Boelelaan 1083, 1081HV Amsterdam, The Netherlands

[2] Department of Mathematics and Statistics, University of Jyvaskyla, P.O. Box 35 (MaD), 40014 Jyvaskyla, Finland

probability measure with respect to the reference measure $\mathfrak{m}$. An element $\gamma \in \Pi_N^{\text{sym}}(\rho)$ is called a symmetric coupling (or transport plan), that is, a symmetric probability measure in X^N having all marginals equal to $\rho\mathfrak{m}$.

We are interested in a class of repulsive cost functions $c \colon X^N \to \mathbb{R} \cup \{+\infty\}$ of the form

$$c(x_1, \ldots, x_N) = \sum_{1 \leq i < j \leq N} f(d(x_i, x_j)), \quad \text{for all } (x_1, \ldots, x_N) \in X^N.$$

We assume $f \colon \,]0, \infty[\to \mathbb{R}$ to be a continuous and decreasing function that approaches $+\infty$ if $d(x_i, x_j) \to 0$. Among the examples of such cost functions we have the Coulomb cost $f(z) = 1/|z|$, the Riesz cost $f(z) = 1/|z|^s$, $n \geq s \geq \max\{n - 2, 0\}$ (in $\mathbb{R}^n$) and the logarithmic cost $f(z) = -\log(|z|)$. We observe that when $\varepsilon = 0$, this entropy-transport problem reduces to the classical multi-marginal optimal transport problem with repulsive costs [4,6,7,12].

The motivation of this paper comes from both theory and numerics. For repulsive cost functions, the entropy term in (1.1) plays a role of a *regularizer* to compute numerically a solution γ of the multi-marginal optimal transport problem $I_0[\rho]$, see [2]. Numerical experiments suggest that when the regularization parameter ε goes to 0, the minimizer γ_ε converges to a minimizer of $I_0[\rho]$ having minimal entropy among the minimizer of $I_0[\rho]$.

From a theoretical viewpoint, this type of a functional has direct relevance in Density Functional Theory. By choosing carefully the parameter ε, the functional (1.1) provides a lower-bound for the Hohenberg–Kohn functional in Density Functional Theory [15,24,27]. This is an immediate consequence of the Log-Sobolev Inequality.

The entropy-transport problem has appeared previously in the literature in the attractive case, in particular when $c(x_1, x_2) = d(x_1, x_2)^2$. We mention briefly below some of the connections of the entropy-transport with other fields and point out the relevance in the Coulomb case.

Brief comments on some applications of the entropy-transport

Optimal transport and Sinkhorn algorithm: The entropy-transport (1.1) was introduced by Cuturi [9] in order to compute numerically the optimal transport plan for the distance squared cost in the 2-marginals case via the Sinkorn algorithm. Due to its reasonable computational cost, it has been applied to a wide range of problems in various research areas, including Information Theory, Computer Graphics, Statistical Inference, Machine Learning, and Mean-Field Games. The entropic regularization method was also considered in the (attractive) multi-marginal case in the so-called *barycenter problem* introduced by Agueh and Carlier [1] (see also [5,11]) and in numerical methods in the time discretization of Brenier's relaxed formulation of the incompressible Euler equation [3]. For a thorough presentation of the computational aspects we refer to Cuturi and Peyré's book [25].

Second-order calculus on RCD spaces: Gigli and Tamanini [8,17] studied the entropic-transport problem on a class of metric spaces with (Riemannian) Ricci curvature bounded from below (2-marginals case, $c(x_1, x_2) = d(x_1, x_2)^2$). The entropic regularization procedure was crucial for establishing a second-order differential structure in that setting.

Schrödinger problem: In 1926, E. Schrödinger introduced the (linear) Schrödinger equations describing the non-relativistic evolution of a single particle in an electric field with potential energy and also established an equivalence between such equations and a system of diffusion equations [26]. Roughly speaking, the variational problem (see (1.1) with $X = C([0, 1], \mathbb{R}^d)$ and $N = 2$) arises in the Schrödinger manuscript while studying the limit $k \to \infty$ ($N = 2$) of the empirical measures associated to the evolution of k i.i.d. Brownian motions. We refer the reader to Léonard survey [21] for technical details and historical notes.
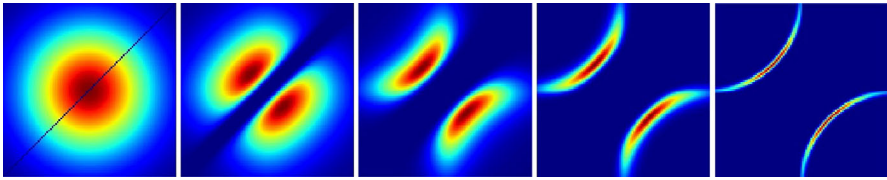
Fig. 1 The dependence of the minimizer of the entropic-transport problem (1.1) on the entropic parameter ε for the one-dimensional Coulomb cost, $N = 2$ and $\rho \sim N(0, 5)$. The pictures show part of the support of the optimal coupling γ_ε around the origin. From the left to right: $\varepsilon = 10^4, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$

Lower bound on the Hohenberg–Kohn functional in density functional theory: This is the particular case where the entropy-transport problem with Coulomb cost comes into play. It has been shown in [24,27] that the functional (1.1) provides a lower bound for computing the ground state energy of the Hohenberg–Kohn functional [4,6,7,12,22]. Below we give a brief description of the result. Notice that in this context $X = \mathbb{R}^d$ and $\mathfrak{m}$ is the Lebesgue measure on $\mathbb{R}^d$.

Assume that $\gamma \in \Pi_N(\rho)$ such that $\sqrt{\gamma} \in H^1(\mathbb{R}^{dN})$. This is the case, for example, when $\gamma(x_1, \ldots, x_N) = |\psi(x_1, \ldots, x_N)|^2$, where $\psi \in H^1(\mathbb{R}^{dN})$ is a ground-state wave function solving the N-electron Schrödinger Equation (see [6,7,12,15,27] for details). Then, we can define the Hohenberg–Kohn functional by

$$\tilde{F}_\hbar^{HK}[\rho]$$
$$= \inf_{\gamma \in \Pi_N(\rho), \sqrt{\gamma} \in H^1(\mathbb{R}^{dN})} \left\{ \frac{\hbar^2}{2} \int_{\mathbb{R}^{dN}} |\nabla \sqrt{\gamma}|^2 dx_1 \ldots dx_N + \int_{\mathbb{R}^{dN}} \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|} d\gamma \right\}.$$

Now, as a consequence of the logarithmic Sobolev inequality for the Lebesgue measure [18], the following result holds: if $\rho \mathcal{L}^d \in \mathcal{P}(\mathbb{R}^d)$ and $\sqrt{\gamma} \in H^1(\mathbb{R}^{dN})$ then

$$C_\varepsilon[\rho] \le \tilde{F}_\hbar^{HK}[\rho], \quad \text{with } \varepsilon = \pi \hbar^2/2.$$

Examples of optimal entropy couplings

Let us present some computational examples of minimizers of $I_\varepsilon[\rho]$ illustrating the role of the parameter ε. Before this, we recall a result on the characterization of minimizers in the one-dimensional case [10]. In particular, according to it the minimizer of $I_0[\rho]$ is concentrated on finitely many graphs and thus singular with respect to the product reference measure.

Theorem 1.1 [10] *Let $\mu \in \mathcal{P}(\mathbb{R})$ be an absolutely continuous probability measure and $f : \mathbb{R} \to \mathbb{R}$ strictly convex, bounded from below and non-increasing function. Then there exists a unique optimal symmetric plan $\gamma \in \Pi_N^{\text{sym}}(\mu)$ that solves*

$$\min_{\gamma \in \Pi_N^{\text{sym}}(\mu)} \int_{\mathbb{R}^N} \sum_{1 \le i < j \le N} f(|x_j - x_i|) \, d\gamma.$$

Moreover, this plan is induced by an optimal cyclical map T, that is, $\gamma_{\text{sym}} = (\gamma_T)^S$, where $\gamma_T = (\text{id}, T, T^{(2)}, \ldots, T^{(N-1)})_\sharp \mu$. An explicit optimal cyclical map is

$$T(x) = \begin{cases} F_\mu^{-1}(F_\mu(x) + 1/N) & \text{if } F_\mu(x) \le (N-1)/N \\ F_\mu^{-1}(F_\mu(x) + 1 - 1/N) & \text{otherwise.} \end{cases}$$

Here $F_\mu(x) = \mu((-\infty, x])$ is the distribution function of μ, and F_μ^{-1} is its lower semicontinuous left inverse.

One-dimensional entropic-transport with Coulomb cost and a Gaussian measure

Let ρ be the normal distribution on the real line with zero mean and standard deviation $\sigma = 5$. We compute numerically the solution of the entropic-transport problem with Coulomb cost in the real line using the Sinkorn algorithm [9]. Notice that by Theorem 1.1, we know that the minimizer of $I_0[\rho]$ is concentrated on a graph. See Fig. 1 for an illustration of the computational results. Our code is based on the Python implementation available at POT library [14].

Organization of the paper

In Sect. 2, we introduce the setting and study sufficient conditions for the existence of minimizers for the entropy-transport problem (1.1). Section 3 is devoted to the $\Gamma-$convergence proof of the entropic-transport functional $C_\varepsilon[\gamma]$ to the multi-marginal optimal transport with repulsive costs $C_0[\gamma]$. In Sect. 4, we study the Kantorovich duality for the entropic-transport problem.

The strategy of the main proof and some technical remarks

The main result of this paper is Theorem 3.1, in which we prove the Γ-convergence of the entropic-regularized functional $C_\varepsilon[\gamma]$ to $C_0[\gamma]$. The technical difficulty on dealing with the Γ-convergence comes from the fact that while for the entropic part $E[\gamma]$ the minimizer γ tends to be *as spread as possible* with respect to $\mathfrak{m}$, for the cost $C_0[\gamma]$ a minimizer can be very singular and have infinite entropy.

We divide the proof in two parts. The part (I), the $\liminf -$inequality, follows basically from the lower-semicontinuity of the costs $C_0[\gamma]$ and $C_\varepsilon[\gamma]$ - which are obtained from the assumption $\rho \log \rho \in L^1_\mathfrak{m}(X)$ on the marginal measure $\rho\mathfrak{m}$, giving a lower bound on the entropy. The part (II), the $\limsup -$inequality, is more involved. In Sect. 3.2, we construct a block approximation γ'_n for a coupling γ with $C_0[\gamma] < +\infty$. Such construction is done in several steps, since we need to construct a competitor γ'_n such that $E[\gamma'_n] < \infty$ and $\gamma'_n \in \Pi^{\mathrm{sym}}_N(\rho)$. The main idea and the rigorous construction are given in Sect. 3.2.

Futhermore, we point out that our construction can deal with the case when the space X is a domain in $\mathbb{R}^d$, answering a question raised in [3]. There the Γ-convergence was proven using convolutions; an approach that does not seem to be easy to implement for domains, or in general metric spaces.

Related works: A proof of the Γ-convergence of (1.1) to the Monge-Kantorovich problem for $c(x, y) = d(x, y)^p$ first appeared in [20,23] via probabilistic methods. In [5], G. Carlier, V. Duval, G. Peyré and B. Schmitzer provided an alternative and more analytical proof carrying out a similar block approximation procedure for the two-marginal squared distance cost in the Euclidean space and the Wasserstein Barycenter.

2 The entropy-regularized repulsive costs

Let (X, d) be a Polish space and $\mathfrak{m}$ be a reference measure on X. We denote by $\mathcal{P}(X)$ the set of Borel probability measures on X, and $\mathcal{P}^{ac}(X)$ the set of Borel probability measures on

X that are absolutely continuous with respect to $\mathfrak{m}$. We denote by $\mathfrak{m}_N$ the product measure $\mathfrak{m} \otimes \mathfrak{m} \otimes \cdots \otimes \mathfrak{m}$. This is the reference measure we use on the product space X^N. On X^N we use the sup-metric, which we denote by d_N.

The class of cost functions $c \colon X^N \to \mathbb{R} \cup \{+\infty\}$ of our interest is given by functions of the form

$$c(x_1, \ldots, x_N) = \sum_{1 \leq i < j \leq N} f(d(x_i, x_j)), \quad \text{for all } (x_1, \ldots, x_N) \in X^N,$$

where $f : [0, \infty[\to \mathbb{R} \cup \{+\infty\}$ satisfies the following conditions

$$f|_{]0,\infty[} \text{ is continuous, decreasing and} \tag{F1}$$

$$\lim_{t \to 0+} f(t) = +\infty. \tag{F2}$$

Above and from now on, we denote by $(x_1, \ldots, x_N)$ points in X^N, so $x_i \in X$ for each i.

We denote by

$$\Pi_N(\rho) = \left\{ \gamma \in \mathcal{P}(X^N) \mid \mathrm{pr}^i_\sharp \gamma = \rho \text{ for all } i \in \{1, \ldots, N\} \right\}$$

the set of *couplings* or *transport plans*, where pr^i is the projection

$$\mathrm{pr}^i(x_1, \ldots, x_i, \ldots, x_N) = x_i \quad \text{for all } (x_1, \ldots, x_i, \ldots, x_N) \in X^N.$$

A measure $\gamma \in \mathcal{P}(X^N)$ is *symmetric* if

$$\int_{X^N} \phi(x_1, \ldots, x_N) \, d\gamma = \int_{X^N} \phi(\overline{\sigma}(x_1, \ldots, x_N)) \, d\gamma, \text{ for all } \phi \in \mathcal{C}(X^N)$$

for all permutations $\overline{\sigma}$ of the N symbols $(x_1, \ldots, x_N)$. We denote by $\mathcal{P}^{\mathrm{sym}}(X^N)$ the set of symmetric probability measures in X^N, and by

$$\Pi_N^{\mathrm{sym}}(\rho) := \Pi_N(\rho) \cap \mathcal{P}^{\mathrm{sym}}(X^N)$$

the set of *symmetric couplings* of ρ.

Let us also introduce the notation for symmetrising measures. If γ is a Borel measure on X^N, we denote by γ^S the symmetrized measure

$$\gamma^S := \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \sigma_\sharp \gamma,$$

where $\mathcal{S}_N$ is the set of permutations of the N coordinates $(x_1, \ldots, x_N)$.

We define the functional $C_0[\gamma]$ to be the cost related to the coupling γ

$$C_0[\gamma] = \int_{X^N} c(x_1, \ldots, x_N) \, d\gamma(x_1, \ldots, x_N).$$

Because of the symmetry of the cost c, we immediately have

Proposition 2.1 *For every $\rho \in (X)$, we have that*

$$\inf_{\gamma \in \Pi_N(\rho)} C_0[\gamma] = \inf_{\gamma \in \Pi_N^{\mathrm{sym}}(\rho)} C_0[\gamma]. \tag{2.1}$$

Moreover, if the infimum is attained on one side of the above equality, then it is attained on both sides.

Given $\varepsilon \geq 0$, we denote by $C_\varepsilon[\gamma]$ the entropy-regularized cost

$$C_\varepsilon[\gamma] = C_0[\gamma] + \varepsilon E[\gamma], \quad \text{for all } \gamma \in \Pi_N^{\text{sym}}(\rho), \tag{2.2}$$

where the entropy $E[\gamma] : \mathcal{P}(X^N) \to \mathbb{R} \cup \{-\infty, +\infty\}$ is defined as

$$E[\gamma] = \begin{cases} \int_{X^N} \rho_\gamma \log \rho_\gamma \, d\mathfrak{m}_N & \text{if } \gamma \ll \mathfrak{m}_N \\ +\infty & \text{otherwise} \end{cases}. \tag{2.3}$$

The notation ρ_γ stands for the Radon-Nikodym derivative of γ with respect to the reference measure $\mathfrak{m}_N$ and $\gamma \ll \mathfrak{m}_N$ means that γ is absolutely continuous with respect to the reference measure $\mathfrak{m}_N$. Let $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$. In this paper we are interested in the following infimum

$$I_\varepsilon[\rho\mathfrak{m}] := \inf_{\gamma \in \Pi_N^{\text{sym}}(\rho)} C_\varepsilon[\gamma]. \tag{2.4}$$

In order to guarantee the lower semicontinuity for $C_\varepsilon[\cdot]$, we will assume $\rho \log \rho \in L_\mathfrak{m}^1(X)$. This will take care of the entropy part $E[\cdot]$ of the cost. In order to establish the lower semicontinuity for the functional $C_0[\cdot]$, we assume that the measure ρ satisfies the following two conditions:

$$\limsup_{r \to 0} \sup_{x \in X} \rho(B(x, r)) < \frac{1}{N(N-1)^2} \quad \text{and} \tag{A}$$

$$\int_{X \setminus B(o, r_0)} f(2d(x, o)) \, d\rho(x) > -\infty \quad \text{for some } o \in X. \tag{B}$$

Above we have, by an abuse of notation, denoted the measure $\rho\mathfrak{m}$ by only the density ρ; we will use the same abbreviation in the rest of the paper if there is no risk of confusion. The Condition (B) is a similar assumption to requiring, in the case of the quadratic cost, that the marginal measures have finite second moments. The Condition (A) guarantees that the cost is finite.

If we endow the spaces $\mathcal{P}(X^N)$ and $\mathcal{P}(X)$ with w^*-topology then, by Prokhorov's theorem, any subset of $\mathcal{P}(X)$ (or $\mathcal{P}(X^N)$) is tight if and only if it is relatively compact.

Remark 2.2 (Entropy-transport seen as a Kullback–Leibler divergence) If μ and ν are measures on a set X, the Kullback–Leibler divergence of μ with respect to ν is defined as

$$\text{KL}[\mu \mid \nu] = \begin{cases} \int_X \log\left(\frac{d\mu}{d\nu}\right) d\mu & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}.$$

Now, if both measures μ and ν are absolutely continuous with respect to some reference measure R of the space X with densities ρ_μ and ρ_ν, respectively, we can write:

$$\text{KL}[\mu \mid \nu] = \begin{cases} \int_X \rho_\mu \log\left(\frac{\rho_\mu}{\rho_\nu}\right) dR & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}.$$

Considering the entropy-regularized MOT problem, we see that the cost functional $C_\varepsilon[\gamma]$ can be alternatively written as the Kullback–Leibler divergence between γ and a *kernel κ* defined below

$$C_\varepsilon[\gamma] = \varepsilon \text{KL}[\gamma \mid \kappa] = \varepsilon \int_{X^N} \rho_\gamma \ln\left(\frac{\rho_\gamma}{\rho_\kappa}\right) d\mathfrak{m}_N,$$

where $\kappa = e^{-c/\varepsilon}\mathfrak{m}_N$.

For the most part, in this paper we have chosen to consider as a reference measure the measure $\mathfrak{m}_N$. However, as the following lemma shows, we could also assume the reference measure to be $(\rho\mathfrak{m})^{\otimes N}$ since the minimizers of the entropy-regularized MOT problem (2.4) do not depend on the choice of the reference measure, at least if there exists a minimizer with finite cost. To state the lemma, let us introduce the notation of relative entropy: for each reference measure R of a Polish space Y, and for each $\gamma \in \mathcal{P}(Y)$, we denote by $E[\gamma \mid R]$ the relative entropy of γ with respect to R, defined as

$$E[\gamma \mid R] = \begin{cases} \int_Y \log\left(\frac{\mathrm{d}\gamma}{\mathrm{d}R}\right) \mathrm{d}\gamma & \text{if } \gamma \ll R \\ +\infty & \text{otherwise} \end{cases}.$$

Now we may consider two, a priori different, entropy-regularized MOT problems: the one introduced in (2.4)

$$I_\epsilon[\rho\mathfrak{m}] = \inf_{\gamma \in \Pi_N^{\text{sym}}(\rho)} C_\varepsilon[\gamma] =: I_\epsilon[\rho\mathfrak{m} \mid \mathfrak{m}], \tag{2.5}$$

and the problem with the reference measure chosen to be $(\rho\mathfrak{m})^{\otimes N}$

$$I_\epsilon[\rho\mathfrak{m} \mid \rho\mathfrak{m}] := \inf_{\gamma \in \Pi_N^{\text{sym}}(\rho)} \left(C_0[\gamma] + E[\gamma \mid (\rho\mathfrak{m})^{\otimes N}]\right). \tag{2.6}$$

The following Lemma 2.3 is used only to go from the compact to the general case in the duality Theorem 4.2. The proof in [11] can be directly applied here to prove Lemma 2.3.

Lemma 2.3 *Let $(X, d, \mathfrak{m})$ be a Polish measure space, $\rho\mathfrak{m} \in \mathcal{P}(X)$ a measure satisfying (A) and (B), and c a cost function satisfying (F1) and (F2). Now for all $\epsilon > 0$ we have*

$$I_\epsilon[\rho\mathfrak{m} \mid \mathfrak{m}] = I_\epsilon[\rho\mathfrak{m} \mid \rho\mathfrak{m}] + N\epsilon KL[\rho\mathfrak{m} \mid \mathfrak{m}] = I_\epsilon[\rho\mathfrak{m} \mid \rho\mathfrak{m}] + N\epsilon \int_X \rho \log \rho \mathrm{d}\mathfrak{m}. \tag{2.7}$$

Moreover, whenever at least one side of the equality above is finite, the problems (2.5) and (2.6) have the same minimizers.

2.1 Some properties of the entropy functional

Let us start by noting that the minimum of the entropy is attained by the product measure and that its value is not $-\infty$.

Proposition 2.4 *Let $(X, d, \mathfrak{m})$ be a Polish metric measure space, and let $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ with $\rho \log \rho \in L_\mathfrak{m}^1(X)$. Then*

$$\min_{\gamma \in \Pi_N^{\text{sym}}(\rho)} E[\gamma] = \int_{X^N} \left(\otimes_{i=1}^N \rho\right) \log\left(\otimes_{i=1}^N \rho\right) \mathrm{d}\mathfrak{m}_N = N \int_X \rho \log \rho \, \mathrm{d}\mathfrak{m} > -\infty.$$

Proof As we will see, the minimality is an immediate consequence of Jensen's inequality. Let $\gamma \in \Pi_N(\rho)$. Then

$$E[\gamma] = \int_{X^N} \rho_\gamma \log(\rho_\gamma) \, d\mathfrak{m}_N = \int_{X^N} \frac{\rho_\gamma}{\otimes_{i=1}^N \rho} \left(\log\left(\frac{\rho_\gamma}{\otimes_{i=1}^N \rho} \right) + \log\left(\otimes_{i=1}^N \rho \right) \right) \otimes_{i=1}^N \rho \, d\mathfrak{m}_N$$

$$\geq \left(\int_{X^N} \rho_\gamma \, d\mathfrak{m}_N \right) \log\left(\int_{X^N} \rho_\gamma \, d\mathfrak{m}_N \right) + \int_{X^N} \rho_\gamma \log\left(\otimes_{i=1}^N \rho \right) \, d\mathfrak{m}_N$$

$$= 0 + E[\otimes_{i=1}^N \rho].$$

$\square$

Using Proposition 2.4 we immediately get the lower semicontinuity of the entropy functional by representing the entropy as relative entropy against the probability measure $\otimes_{i=1}^N (\rho\mathfrak{m})$. See for instance [28, Lemma 4.1] for the lower semicontinuity of the entropy when the reference measure is finite.

Corollary 2.5 *Let $(X, d, \mathfrak{m})$ be a Polish metric measure space, and let $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ with $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$. Then $E[\cdot]$ is lower semicontinuous in the set $\Pi_N^{\mathrm{sym}}(\rho)$.*

Now we are ready to prove the existence of the minimizers for entropy-regularized MOT:

Proposition 2.6 *Let $(X, d, \mathfrak{m})$ be a Polish metric measure space. Assume that the measure $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ satisfies $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$ along with Conditions (A) and (B). Assume that $c \colon X^N \to \mathbb{R} \cup \{+\infty\}$ satisfies the conditions (F1) and (F2). Then, for each $\varepsilon \geq 0$, there exists a minimizer $\gamma \in \Pi_N^{\mathrm{sym}}(\rho)$ for the entropic-regularized cost $C_\varepsilon[\gamma]$.*

Proof We notice that the set $\Pi_N^{\mathrm{sym}}(\rho)$ is compact in the w^*-topology [19]. The functional E is lower semicontinuous by Corollary 2.5, and in our setting the lower semicontinuity of C_0 is proven as a part of the proof of [16, Proposition 3.1]. Since for each $\varepsilon \geq 0$ the functional C_ε is lower semicontinuous, we conclude that it has a minimizer in the set $\Pi_N^{\mathrm{sym}}(\rho)$ $\square$

3 The Γ-convergence of entropic-regularized cost

Now let us turn to the Γ-convergence. From now on, $(\tau_n)_{n\in\mathbb{N}}$ is any sequence of positive real numbers decreasing to zero. Let us introduce the following functionals: for each $n \in \mathbb{N}$

$$\mathcal{C}_n : \mathcal{P}^{\mathrm{sym}}(X^N) \to \mathbb{R} \cup \{+\infty\}, \quad \mathcal{C}_n[\gamma] = \begin{cases} C_{\tau_n}(\gamma) & \text{if } \gamma \in \Pi_N(\rho) \\ +\infty & \text{otherwise} \end{cases}$$

and

$$\mathcal{C} : \mathcal{P}^{\mathrm{sym}}(X^N) \to \mathbb{R} \cup \{+\infty\}, \quad \mathcal{C}[\gamma] = \begin{cases} C_0[\gamma] & \text{if } \gamma \in \Pi_N(\rho) \\ +\infty & \text{otherwise} \end{cases}.$$

The goal of this section is to prove that the sequence $(\mathcal{C}_{n\in\mathbb{N}})$ Γ-converges to $\mathcal{C}$ in the space $\mathcal{P}^{\mathrm{sym}}(X^N)$.

Theorem 3.1 *Let $(X, d, \mathfrak{m})$ be a Polish metric measure space. Let $\rho \in \mathcal{P}^{ac}(X)$ with $\rho \log \rho \in L^1_{\mathfrak{m}}(X)$ satisfying (A) and (B). Then the sequence $(\mathcal{C}_n)$ Γ-converges to $\mathcal{C}$ in the space $\mathcal{P}^{\mathrm{sym}}(X^N)$.*

Let us fix $\gamma \in \mathcal{P}^{\mathrm{sym}}(X^N)$. We need to show that

$$\text{For each sequence } (\gamma_n)_{n \in \mathbb{N}} \text{ that converges to } \gamma \tag{I}$$

we have $\liminf\limits_{n \to \infty} \mathcal{C}_n[\gamma_n] \geq \mathcal{C}[\gamma]$, and

$$\text{There exists a sequence } (\gamma_n)_{n \in \mathbb{N}} \text{ that converges to } \gamma \text{ and} \tag{II}$$

$$\limsup\limits_{n \to \infty} \mathcal{C}_n[\gamma_n] \leq \mathcal{C}[\gamma].$$

The proof of Theorem 3.1 is divided into two parts. The proof of the first part, the liminf-inequality (I), is short and is established in the next subsection. The remainder of this section is then divided into subsections in which the second part, the limsup-inequality (II) is proven.

3.1 Proof of condition (I)

We fix a sequence $(\gamma_n)_{n \in \mathbb{N}}$ that converges to γ. If $\gamma \notin \Pi_N(\rho)$, then since the set $\Pi_N(\rho)$ is compact, for large indices we also have $\gamma_n \notin \Pi_N(\rho)$, so both sides of inequality (I) are $+\infty$, and we are done. Hence we may assume that γ and γ_n's are elements of the set $\Pi_N(\rho)$. Since now $\gamma_n \in \Pi_N(\rho)$, the claim (I) follows from the lower-semicontinuity of $\gamma \mapsto \int c \, d\gamma$ and from the entropy lower bound shown in Proposition 2.4. □

3.2 Constructing an approximation of the coupling γ

First of all, we need to construct an approximation of γ only in the case where $C_0[\gamma] < \infty$: if this is not the case then any sequence (γ_n) converging to γ can be used to prove Condition (II). The idea of the construction is to redefine a large part of γ to be a product measure on finitely many Borel sets with small diameter. In order not to increase the cost by too much, the Borel sets we are using have to be far away from the diagonal compared to the diameter of the sets. We call the part of the measure defined in this way *the core part* of the approximation. For the rest of the measure, we take another finite combination of product measures. However, this time the sets do not need to have small (or even bounded) diameter, but just small measure. This part will be called *the remainder part* of the approximation.

We start the construction by taking out a small part of γ that will later be used to deal with the remainder part of the approximation. For this we take a sequence of radii defined as $r_n = 1/n$. Since $C_0[\gamma] < \infty$, there exists a point $x = (x_1, \ldots, x_N) \in \mathrm{spt}(\gamma)$ with

$$x_i \neq x_j \quad \text{if } 1 \leq i < j \leq N.$$

Moreover, since $\gamma \in \Pi_N(\rho)$ and ρ satisfies (A), we have

$$\gamma(\{(y_1, \ldots, y_N) \in X^N \mid y_i \neq x_j \text{ for all } i, j\})$$

$$\geq 1 - \sum_{i \neq j} \gamma(\{(y_1, \ldots, y_N) \in X^N \mid y_i = x_j\})$$

$$= 1 - \sum_{i \neq j} \rho(\{x_j\}) \geq 1 - N(N-1)\frac{1}{N(N-1)^2} > 0.$$

Thus, using again $C_0[\gamma] < \infty$, there exists another point $x' = (x_{N+1}, \ldots, x_{2N}) \in \mathrm{spt}(\gamma)$, so that

$$x_i \neq x_j \quad \text{if } 1 \leq i < j \leq 2N.$$

From now on, we consider x, x' fixed. Therefore, for $n \in \mathbb{N}$ sufficiently large we have

$$d(x_i, x_j) > r_n \quad \text{if } 1 \le i < j \le 2N. \tag{3.1}$$

Let us denote by

$$B_n := B(x, \tfrac{r_n}{10}) \quad \text{and} \quad B'_n := B(x', \tfrac{r_n}{10})$$

the balls around x and x' with radii $r_n/10$ in the sup-metric of the product space. So,

$$y = (y_1, \ldots, y_N) \in B_n$$

if and only if

$$d(x_i, y_i) < \tfrac{r_n}{10} \quad \text{for all } i \in \{1, \ldots, N\},$$

and analogously for B'_n with the relevant index modifications.

Let us now define

$$\gamma_{B_n} = \left(\frac{\gamma|_{B_n}}{\gamma(B_n)} \right)^S \quad \text{and} \quad \gamma_{B'_n} = \left(\frac{\gamma|_{B'_n}}{\gamma(B'_n)} \right)^S.$$

Observe that γ_{B_n} and $\gamma_{B'_n}$ are symmetric probability measures. Since the marginals of a symmetric measure are the same, we may denote by ρ_{B_n} the marginal of γ_{B_n} and similarly by $\rho_{B'_n}$ the marginal of $\gamma_{B'_n}$. Let us further denote $\tilde{B}_n := \mathrm{spt}\, \gamma_{B_n}$, $\tilde{B}'_n := \mathrm{spt}\, \gamma_{B'_n}$ and

$$\varepsilon_n := \frac{1}{N} \min \left\{ \gamma(\tilde{B}_n), \gamma(\tilde{B}'_n), r_n, \frac{r_n}{f(2r_n/5)} \right\}. \tag{3.2}$$

We then define a measure

$$\gamma_{0,n} := \gamma|_{X^n \setminus (\tilde{B}_n \cup \tilde{B}'_n)} + \frac{\gamma(\tilde{B}_n) - \varepsilon_n}{\gamma(\tilde{B}_n)} \gamma|_{\tilde{B}_n} + \frac{\gamma(\tilde{B}'_n) - \varepsilon_n}{\gamma(\tilde{B}'_n)} \gamma|_{\tilde{B}'_n}.$$

The idea behind the measure $\gamma_{0,n}$ is that we have chopped off a small part of the measure around the points x and x' (symmetrically) for later use. Since we are working with a singular cost, we still need to take out a small neighbourhood of the diagonals before approximating by product measures. We do this now.

We fix a compact $K_n \subset X$ such that

$$\gamma_{0,n}(X^N \setminus K_n^N) < \frac{\varepsilon_n}{2} \tag{3.3}$$

and take a small enough $\delta_n \in (0, r_n)$ so that

$$\gamma_{0,n}(D_{\delta_n}) < \frac{\varepsilon_n}{2},$$
$$D_\alpha := \{(x_1, \ldots, x_N) \in X^N \mid d(x_i, x_j) < \alpha \text{ for some } i \neq j\} \tag{3.4}$$

denotes the α-neighbourhood of the pairwise diagonals. Using K_n and δ_n we then define

$$\gamma_{1,n} := \gamma_{0,n}|_{K_n^N \setminus D_{\delta_n}}. \tag{3.5}$$

The measure $\gamma_{1,n}$ is now the core part of the measure that we approximate. We denote by $\rho_{1,n}$ the marginals of the symmetric measure $\gamma_{1,n}$.

Let us then approximate the measure $\gamma_{1,n}$. We take $\lambda_n \in (0, \delta_n/n)$ so that

$$|f(r) - f(s)| < \varepsilon_n \quad \text{for all } r, s \in [\delta_n/2, 2\,\mathrm{diam}(K_n)] \text{ with } |r - s| \le 2\lambda_n. \tag{3.6}$$

Such λ_n exists by the uniform continuity of f on the compact set $[\delta_n/2, 2 \operatorname{diam}(K_n)]$. Since the set K_n is compact, we may fix a finite Borel partition $\{B_n^i\}_{i=1}^{M_n}$ of the set $\operatorname{spt}(\rho_{1,n})$ such that

$$\operatorname{diam}(B_n^i) < \lambda_n \quad \text{and} \quad \rho_{1,n}(B_n^i) > 0 \quad \text{for all } i \in \{1, \ldots, M_n\}.$$

We are now ready to define the core part approximants $\gamma_{1,n}^a$ as

$$\gamma_{1,n}^a = \sum_{(k_1, \ldots, k_N) \in M_n^N} \frac{\gamma_{1,n}(B_n^{k_1} \times \cdots \times B_n^{k_N})}{\rho_{1,n}(B_n^{k_1}) \cdots \rho_{1,n}(B_n^{k_N})} \rho_{1,n}|_{B_n^{k_1}} \otimes \cdots \otimes \rho_{1,n}|_{B_n^{k_N}}. \tag{3.7}$$

Now let us handle the main part of the remainder of the measure, namely the measure

$$\gamma_{2,n} := \gamma_{0,n}|_{D_{\delta_n} \cup (X^N \setminus K_n^N)}.$$

Because $\gamma_{0,n}$ and the set where we restrict it are symmetric, $\gamma_{2,n}$ is as well. We may thus denote its marginals by $\rho_{2,n}$.

In order to determine which part of the remaining marginal measure should be coupled where, we define a partition $\{A_{i,n}\}_{i=1}^N$ of the space X by setting, for all $i \in \{1, \ldots, N-1\}$

$$A_{i,n} := \{y \in X \mid d(x_i, y) \le \tfrac{r_n}{2}\},$$

and

$$A_{N,n} := X \setminus \bigcup_{i=1}^{N-1} A_{i,n}.$$

Condition (3.1) guarantees that the sets $A_{i,n}$ are pairwise disjoint.

Now we approximate $\gamma_{2,n}$ by the measure

$$\gamma_{2,n}^a := N \left(\sum_{i=1}^N \eta_{n,i} \right)^S,$$

where for all i the measure $\eta_{n,i}$ is the product

$$\eta_{n,i} := \left(\bigotimes_{k=1}^{i-1} \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right) \otimes \rho_{2,n}|_{A_{i,n}} \otimes \left(\bigotimes_{k=i+1}^N \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right).$$

By the definition of the sets $A_{i,n}$, for every $(y_1, \ldots, y_N) \in \operatorname{spt}(\gamma_{2,n}^a)$ we have for each $i \neq j$

$$d(y_i, y_j) \ge |d(y_i, x_j) - d(x_j, y_j)| > \frac{r_n}{2} - \frac{r_n}{10} = \frac{2r_n}{5}, \tag{3.8}$$

where we have assumed (which we can do without loss of generality) that $y_j \in \overline{B}(x_j, r_n/10)$.

What we have done using the measure $\gamma_{2,n}^a$ is that we have coupled the marginals of the measure $\gamma_{2,n}$ with some suitable parts of the marginals of the reserved measure that was taken out around the point x. In this way we have used unevenly the marginals of this reserved part. To handle the rest of the reserved part of the measure around the point x, we now use the reserved measure around the point x'. So, we need to redefine the coupling for the part of the marginal given by

$$\rho_{3,n} := (\operatorname{pr}^1)_\sharp \frac{\varepsilon_n}{\gamma(\tilde{B}_n)} \gamma|_{\tilde{B}_n} + \rho_{2,n} - (\operatorname{pr}^1)_\sharp \gamma_{2,n}^a.$$

We define it as

$$\gamma_{3,n}^a := \left(\sum_{i=1}^N \phi_{n,i} \right)^S,$$

where each $\phi_{n,i}$ is defined as

$$\phi_{n,i} := \left(\bigotimes_{k=N+1}^{N+i-1} \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right) \otimes \rho_{3,n} \otimes \left(\bigotimes_{k=N+i+1}^{2N} \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right).$$

Since $\mathrm{spt}(\rho_{3,n}) \subset \mathrm{spt}(\rho_{B_n})$, we have that for every $(y_1, \ldots, y_N) \in \mathrm{spt}(\gamma_{3,n}^a)$ and each $i \neq j$

$$d(y_i, y_j) \geq |d(x_{k(i)}, x_{k(j)}) - d(y_i, x_{k(i)}) - d(x_{k(j)}, y_j)| > r_n - \frac{r_n}{10} - \frac{r_n}{10} = \frac{4r_n}{5}, \tag{3.9}$$

where $k(i) \neq k(j)$ are the indices for which $y_j \in \overline{B}(x_{k(j)}, r_n/10)$ and $y_i \in \overline{B}(x_{k(i)}, r_n/10)$.

What remains is the part of the measure around x' that was not used for $\gamma_{3,n}^a$. Since $\gamma_{3,n}^a$ used the marginals from this part of the reserved measure evenly, we may simply couple the rest by a measure

$$\gamma_{4,n}^a := b \left(\bigotimes_{k=N+1}^{2N} \frac{\rho_{B_n}|_{B(x_k, r_n/10)}}{\rho_{B_n}(B(x_k, r_n/10))} \right)^S,$$

with b being the correct scaling constant. Similarly as for the previous remainder part, we have that for every $(y_1, \ldots, y_N) \in \mathrm{spt}(\gamma_{4,n}^a)$ and each $i \neq j$ the inequality (3.9) holds.

Now we are ready to define the full approximation as

$$\gamma_n' = \gamma_{1,n}^a + \gamma_{2,n}^a + \gamma_{3,n}^a + \gamma_{4,n}^a.$$

By construction $\gamma_n' \in \Pi_N^{\mathrm{sym}}(\rho)$.

3.3 Narrow convergence of the approximations

Let us now prove that the sequence $(\gamma_n')_n$ narrowly converges to γ. We could argue this by using the Wasserstein distance. However, let us do it here directly using the definition of narrow convergence.

Lemma 3.2 *The sequence $(\gamma_n')_n$ narrowly converges to γ.*

Proof Let $\varphi \in C_b(X^N)$ and $\varepsilon > 0$. We need an index $N_0 \in \mathbb{N}$ such that

$$\left| \int_{X^N} \varphi \, d\gamma - \int_{X^N} \varphi \, d\gamma_n' \right| < \varepsilon \text{ for all } n \geq N_0. \tag{3.10}$$

Let us denote $M := \sup_{x \in X^N} |\varphi(x)|$; we may assume that $M > 0$. Since ρ is inner regular, we can fix a compact set $K \subset X$ such that

$$\rho(X \setminus K) < \frac{\varepsilon}{12NM}.$$

Since $\gamma \in \Pi_N(\rho)$, we now have

$$\gamma(X^N \setminus K^N) < \frac{\varepsilon}{12M}.$$

The function φ, when restricted to K^N, is uniformly continuous. Hence there exists $\delta > 0$ so that

$$|\varphi(x) - \varphi(y)| < \frac{\varepsilon}{12} \quad \text{for all } x, y \in K^N \text{ for which } d_N(x, y) < \delta. \tag{3.11}$$

Now, let $N_0 \in \mathbb{N}$ be so large that

$$\sqrt{N}\lambda_n < \delta \quad \text{and} \quad 6M\varepsilon_n < \frac{\varepsilon}{6} \quad \text{for all } n \geq N_0.$$

Let us show that this choice of N_0 satisfies (3.10). First we note that for all $n \geq N_0$ we have

$$
\left| \int_{X^N} \varphi \, d\gamma - \int_{X^N} \varphi \, d\gamma'_n \right| \leq \left| \int_{X^N} \varphi \, d\gamma - \int_{X^N} \varphi \, d\gamma_{1,n} \right| + \left| \int_{X^N} \varphi \, d\gamma_{1,n} - \int_{X^N} \varphi \, d\gamma'_n \right|
$$

$$
\leq \left| \int_{X^N} \varphi \, d\gamma - \int_{X^N} \varphi \, d\gamma_{1,n} \right| + \left| \int_{X^N} \varphi \, d\gamma_{1,n} - \int_{X^N} \varphi \, d\gamma^a_{1,n} \right|
$$

$$
+ \left| \int_{X^N} \varphi \, d(\gamma^a_{2,n} + \gamma^a_{3,n} + \gamma^a_{4,n}) \right|
$$

$$
< \left| \int_{X^N} \varphi \, d\gamma_{1,n} - \int_{X^N} \varphi \, d\gamma^a_{1,n} \right| + \frac{\varepsilon}{6}, \tag{3.12}
$$

where in the last inequality we have used the following facts: $\gamma(X^N) - \gamma_{1,n}(X^N) < 3\varepsilon_n$ for all n, and for the remainder part of the measure γ'_n we have

$$(\gamma^a_{2,n} + \gamma^a_{3,n} + \gamma^a_{4,n})(X^N) < 3\varepsilon_n \quad \text{for all } n \in \mathbb{N}.$$

It remains to show that for all $n \geq N_0$ we have

$$\left| \int_{X^N} \varphi \, d\gamma_{1,n} - \int_{X^N} \varphi \, d\gamma^a_{1,n} \right| < \frac{5\varepsilon}{6}. \tag{3.13}$$

We first estimate the integrals in the set K^N. Let us fix, for each $(k_1, \ldots, k_N) \in M_n^N$ for which the set

$$(B_n^{k_1} \times \cdots \times B_n^{k_N}) \cap K^N$$

is nonempty, an element

$$z_{k_1,\ldots,k_N} \in (B_n^{k_1} \times \cdots \times B_n^{k_N}) \cap K^N.$$

Now we have, for a fixed $(k_1, \ldots, k_N)$, denoting for simplicity

$$z_0 := z_{k_1,\ldots,k_N}, \ Q = B_n^{k_1} \times \cdots \times B_n^{k_N}, \ \gamma = \gamma_{1,n}, \ \gamma_a = \gamma^a_{1,n},$$

$$
\left| \int_{Q \cap K^N} \varphi(z) \, d\gamma - \int_{Q \cap K^N} \varphi(z) \, d\gamma_a \right|
$$

$$
\leq \left| \int_{Q \cap K^N} \varphi(z) \, d\gamma - \varphi(z_0)\gamma(Q \cap K^N) \right| + \left| \varphi(z_0)\gamma_a(Q \cap K^N) - \int_{Q \cap K^N} \varphi(z) \, d\gamma_a \right|
$$

$$
+ \left| \varphi(z_0)\gamma(Q \cap K^N) - \varphi(z_0)\gamma_a(Q \cap K^N) \right|
$$

$$
\leq \int_{Q \cap K^N} |\varphi(z) - \varphi(z_0)| \, d\gamma + \int_{Q \cap K^N} |\varphi(z_0) - \varphi(z)| \, d\gamma_a
$$

$$
+ M|\gamma(Q \cap K^N) - \gamma_a(Q \cap K^N)|
$$

$$\overset{a)}{<} \gamma(Q \cap K^N) \cdot \frac{\varepsilon}{12} + \gamma_a(Q \cap K^N) \cdot \frac{\varepsilon}{12} + M|\gamma(Q \cap K^N) - \gamma_a(Q \cap K^N)|$$

$$= \gamma(Q \cap K^N) \cdot \frac{\varepsilon}{12} + \gamma_a(Q \cap K^N) \cdot \frac{\varepsilon}{12}$$

$$+ M|\gamma(Q \cap K^N) - \gamma(Q) + \gamma_a(Q) - \gamma_a(Q \cap K^N)|$$

$$\le \gamma(Q \cap K^N) \cdot \frac{\varepsilon}{12} + \gamma_a(Q \cap K^N) \cdot \frac{\varepsilon}{12} + M\gamma(Q \setminus K^N) + M\gamma_a(Q \setminus K^N),$$

where in a) we have used (3.11), and in b) the fact that the total measures of γ and γ_a coincide on 'cubes' Q. Summing the estimate above over all cubes $Q = B_{k_1} \times \cdots \times B_{k_N}$, $(k_1, \ldots, k_N) \in M_n^N$, gives

$$\left| \int_{K^N} \varphi \, d\gamma_{1,n} - \int_{K^N} \varphi \, d\gamma_{1,n}^a \right|$$

$$< \gamma_{1,n}(K^N) \cdot \frac{\varepsilon}{12} + \gamma_a(K^N) \cdot \frac{\varepsilon}{12} + M\gamma(X^N \setminus K^N) + M\gamma_a(X^N \setminus K^N)$$

$$\overset{a)}{<} \frac{\varepsilon}{12} + \frac{\varepsilon}{12} + \frac{\varepsilon}{12} + \frac{\varepsilon}{12} = \frac{\varepsilon}{3}, \tag{3.14}$$

where in inequality a) we have used the fact that $\rho(X \setminus K) < \frac{\epsilon}{12MN}$ and, since the marginals of $\gamma_{1,n}$ and $\gamma_{1,n}^a$ are restrictions of ρ, we can bound both $\gamma_{1,n}(X^N \setminus K^N)$ and $\gamma_{1,n}^a(X^N \setminus K^N)$ by $\frac{\varepsilon}{12M}$. For the same reason, we have

$$\left| \int_{X^N \setminus K^N} \varphi \, d\gamma_{1,n} - \int_{X^N \setminus K^N} \varphi \, d\gamma_{1,n}^a \right| < 2M \cdot \frac{\varepsilon}{12M} = \frac{\varepsilon}{6}. \tag{3.15}$$

Combining estimates (3.14) and (3.15) gives

$$\left| \int_{X^N} \varphi \, d\gamma_{1,n} - \int_{X^N} \varphi \, d\gamma_{1,n}^a \right| < \frac{\varepsilon}{3} + \frac{\varepsilon}{6} = \frac{3\varepsilon}{6} < \frac{5\varepsilon}{6},$$

proving (3.13). □

3.4 Convergence of the cost functional

In order to prove the Γ-limsup inequality (II), we need the cost $C_0[\cdot]$ to converge along the approximating sequence γ_n. We prove this in the following lemma.

Lemma 3.3 *We have $C_0[\gamma_n'] \to C_0[\gamma]$ as $n \to \infty$.*

Proof Let us first consider the remainder part. Recall that for all $n \in \mathbb{N}$ we have

$$(\gamma_n' - \gamma_{1,n}^a)(X^N) = (\gamma_{2,n}^a + \gamma_{3,n}^a + \gamma_{4,n}^a)(X^N) < 3\varepsilon_n.$$

Thus, using the lower bounds (3.8) and (3.9) for distances in the support of the remainder part, and the definition (3.2) of ε_n, we get

$$\int_{X^N} c \, d(\gamma_n' - \gamma_{1,n}^a) \le \frac{N(N-1)}{2} f\left(\frac{2r_n}{5}\right) 3\varepsilon_n \le \frac{3(N-1)}{2} r_n \to 0 \tag{3.16}$$

as $n \to \infty$. By continuity of the integral, we get

$$\int_{X^N} c \, d(\gamma - \gamma_{1,n}) \to 0 \tag{3.17}$$

as $n \to \infty$.

Let us now estimate the core part of the approximation. By the construction (3.7) of $\gamma^a_{1,n}$ and the choice (3.6) of λ_n, we have

$$\left| \int_{X^N} c \, d\gamma^a_{1,n} - \int_{X^N} c \, d\gamma_{1,n} \right| \le \int_{X^N} \frac{N(N-1)}{2} \varepsilon_n \, d\gamma_{1,n} < \frac{N(N-1)}{2} \varepsilon_n. \tag{3.18}$$

Combining (3.16), (3.17) and (3.18) we get

$$\left| C_0[\gamma'_n] - C_0[\gamma] \right| \le \left| \int_{X^N} c \, d(\gamma'_n - \gamma^a_{1,n}) \right| + \left| \int_{X^N} c \, d\gamma^a_{1,n} - \int_{X^N} c \, d\gamma_{1,n} \right|$$
$$+ \left| \int_{X^N} c \, d(\gamma - \gamma_{1,n}) \right| \to 0$$

as $n \to \infty$. $\qquad\square$

3.5 Finiteness of the entropy for the approximations

Next we show that the entropy is finite for the approximating sequence. Notice that, in order to prove (II), we do not need a better estimate on the entropy.

Lemma 3.4 *For each $n \in \mathbb{N}$ we have $E[\gamma'_n] < \infty$.*

Proof In order to see the finiteness of the entropy, it suffices to notice that each γ'_n is a sum of finitely many measures $(\tilde{\gamma}_{n,k})_{k=1}^{N_n}$ each of which is of the form $\tilde{\gamma}_{n,k} = \tilde{\rho}^k_1 \mathfrak{m} \otimes \cdots \otimes \tilde{\rho}^k_N \mathfrak{m}$ with $\tilde{\rho}^k_i \ll \rho$ and $\frac{d\tilde{\rho}^k_i}{d\rho} \le 1$. Indeed, by Proposition 2.4, the entropy is always bounded from below, and so we can make a crude estimate:

$$E[\gamma'_n] = \int_{X^N} \log\left(\sum_{k=1}^{N_n} \frac{d\tilde{\gamma}_{n,k}}{d\mathfrak{m}} \right) d\left(\sum_{k=1}^{N_n} \tilde{\gamma}_{n,k} \right)$$
$$\le \log(N_n) + \sum_{k=1}^{N_n} \int_{X^N} \log\left(\frac{d\tilde{\gamma}_{n,k}}{d\mathfrak{m}} \right) d\tilde{\gamma}_{n,k} < \infty.$$

$\qquad\square$

3.6 Proof of condition (II)

We are now ready to prove the Γ-lim sup inequality (II). By Lemma 3.2 we already know that $(\gamma'_n)_n$ converges to γ. However, $\mathcal{C}_n[\gamma'_n]$ need not converge to $\mathcal{C}[\gamma]$. This can be solved by making the convergence of $(\gamma'_n)_n$ slower by repeating always the same measure for sufficiently (but finitely) many times before moving to the next one. We define $k(n)$ for every $n \in \mathbb{N}$ as

$$k(n) = \min\left(n, \max\left(1, \sup\left\{ k \in \mathbb{N} \,|\, \sqrt{\tau_n} E[\gamma'_j] < 1 \text{ for all } j \le k \right\} \right) \right).$$

By definition, $1 \le k(n) \le n$. Moreover, since for every $j \in \mathbb{N}$ we have $E[\gamma'_j] < \infty$ by Lemma 3.4 and $\tau_n \to 0$ by definition, we have that $k(n) \to \infty$ as $n \to \infty$. Thus, defining $\gamma_n = \gamma'_{k(n)}$, for large enough $n \in \mathbb{N}$ we have

$$\mathcal{C}_n[\gamma_n] = C_0[\gamma'_{k(n)}] + \tau_n E[\gamma'_{k(n)}] < C_0[\gamma'_{k(n)}] + \sqrt{\tau_n}.$$

Recalling that by Lemma 3.3 we have $C_0[\gamma'_{k(n)}] \to C_0[\gamma]$, we conclude the proof. $\qquad\square$

In Proposition 2.6 the existence of a minimizer for the entropy-regularized cost was established. Now that we know that measures γ for which $C_0(\gamma) < \infty$ can be approximated by measures with not only finite costs but also finite entropy, we can say more:

Corollary 3.5 *Let* $(X, d, \mathfrak{m})$ *be a Polish metric measure space. Assume that* $\rho\mathfrak{m} \in \mathcal{P}^{ac}(X)$ *satisfies* $\rho \log \rho \in L^1_\mathfrak{m}(X)$ *and Conditions (A) and (B). Assume that* $c \colon X^N \to \mathbb{R} \cup \{+\infty\}$ *satisfies Conditions (F1) and (F2). Then, for each* $\varepsilon > 0$, *there exists a unique minimizer* $\gamma \in \Pi_N^{\mathrm{sym}}(\rho)$ *for the entropic-regularized cost* $C_\varepsilon[\gamma]$.

Proof Our marginal measure satisfies Conditions (A) and (B), so there exists a measure $\gamma \in \Pi_N^{\mathrm{sym}}(\rho)$ that minimizes C_0 with $C_0[\gamma] < \infty$. It must be noted that this measure can have infinite entropy. However, because of the approximation result presented in the proof of Condition (II) above, we get the existence of a measure $\gamma' \in \Pi_N^{\mathrm{sym}}(\rho)$ such that $C_\epsilon[\gamma'] < \infty$. The uniqueness claim now follows, since the functional $\gamma \mapsto C_\epsilon[\gamma]$ is strictly convex for $\epsilon > 0$. $\qquad\square$

4 Entropic-Kantorovich duality for Coulomb-type costs

We start by recalling the classical Fenchel–Rockafellar Theorem. We refer to the I. Ekeland and R. Témam's book [13, Theorem 4.2] for a more complete presentation and references.

Theorem 4.1 (Fenchel–Rockafellar) *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be Banach spaces and* $A : \mathcal{X} \to \mathcal{Y}$ *be linear and continuous. Let* $F : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ *and* $G : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ *be proper and convex functions. Then*

$$\inf\big\{F[x] + G[Ax] \,\big|\, x \in \mathcal{X}\big\} = \sup\big\{-F^*[-A^*\gamma] - G^*[\gamma] \,\big|\, \gamma \in \mathcal{Y}^*\big\}$$

where $A^* : \mathcal{Y}^* \to \mathcal{X}^*$ *denotes the adjoint operator of* A.

Next we prove the Entropic-Kantorovich duality for the problem (2.4).

Theorem 4.2 (Entropic Duality for repulsive costs) *Let* $(X, d, \mathfrak{m})$ *be a Polish measure space. Suppose* $\rho\mathfrak{m} \in \mathcal{P}(X)$ *such that (A) and (B) hold and* $\rho \log \rho \in L^1_m(X)$, *and* $c \colon X^N \to [0, \infty]$ *is a cost function*

$$c(x_1, \ldots, x_N) = \sum_{1 \le i < j \le N} f(d(x_i, x_j)), \quad \textit{for all } (x_1, \ldots, x_N) \in X^N,$$

where $f \colon [0, +\infty[\to [0, \infty]$ *is a function satisfying (F1) and (F2). Then, for* $\varepsilon > 0$, *the duality holds*

$$\min_{\gamma \in \Pi_N(\rho)} C_\varepsilon[\gamma] = \sup_{u_i \in C_b(X)} \left\{ \sum_{i=1}^N \int_X u_i \mathrm{d}\mathfrak{m} - \varepsilon \int_{X^N} \exp\left(\frac{u_1 \oplus \cdots \oplus u_N - c}{\varepsilon}\right) \mathrm{d}\mathfrak{m}_N \right\} + \varepsilon$$

$$= \sup_{u \in C_b(X)} \left\{ N \int_X u \mathrm{d}\mathfrak{m} - \varepsilon \int_{X^N} \exp\left(\frac{u \oplus \cdots \oplus u - c}{\varepsilon}\right) \mathrm{d}\mathfrak{m}_N \right\} + \varepsilon,$$

where $v_1 \oplus \cdots \oplus v_N$ *denotes the operator* $(v_1 \oplus \cdots \oplus v_N)(x_1, \ldots, x_N) = v_1(x_1) + \cdots + v_N(x_N)$.

Proof First let us assume that X is a compact space. We denote by $\mathcal{X} = (C_b(X))^N$ and $\mathcal{Y} = C_b(X^N)$, where $C_b(X)$ is the space of continuous and bounded functions on X, and

similarly for X^N. By Riesz representation theorem, the space $\mathcal{Y}$ is dual to the space $\mathcal{M}(X^N)$ of signed regular Borel measures on X^N. Thus, we may define the Legendre–Fenchel transform G^* of a functional $G \colon \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ by

$$G^* \colon \mathcal{M}(X^N) \to \mathbb{R} \cup \{+\infty\}, \quad G^*[\pi] = \sup_{\psi \in C_b(X^N)} \left\{ \int_{X^N} \psi \, \mathrm{d}\pi - G[\psi] \right\}.$$

We define the functionals

$$F \colon (C_b(X))^N \to \mathbb{R} \cup \{+\infty\}, \quad (u_1, \ldots, u_N) \mapsto -\sum_{i=1}^{N} \int_X u_i \, \mathrm{d}\rho\mathfrak{m}$$

and

$$G \colon C_b(X^N) \to \mathbb{R} \cup \{+\infty\}, \quad \psi \mapsto \varepsilon \int_{X^N} e^{\frac{1}{\varepsilon}(\psi - c)} \, \mathrm{d}\mathfrak{m}_N,$$

and the operator

$$A \colon (C_b(X))^N \to C_b(X^N), \quad (u_1, \ldots, u_N) \mapsto u_1 \oplus \cdots \oplus u_N.$$

Now, F and G are proper and convex functionals and A is a linear and continuous operator. So, we may apply Fenchel–Rockafellar duality Theorem 4.1 to get

$$\inf\{F[x] + G[Ax] \mid x \in \mathcal{X}\} = \sup\{-G^*[\gamma] - F^*[-A^\dagger \gamma] \mid \gamma \in \mathcal{Y}^*\}.$$

This gives (since for every set S we have $\inf(S) = -\sup(-S)$ and $\sup S = -\inf(-S)$)

$$\inf\{G^*[\gamma] + F^*[-A^\dagger \gamma] \mid \gamma \in \mathcal{Y}^*\} = \sup\{-F[x] - G[Ax] \mid x \in \mathcal{X}\}.$$

It remains to show that the above expression has exactly the form of our duality claim. The claim that the right-hand sides correspond to each other follows immediately from our choices of $\mathcal{X}$, F, and G. So, it remains to show that

$$\inf\{G^*[\gamma] + F^*[-A^\dagger \gamma] \mid \gamma \in \mathcal{Y}^*\} = C_\varepsilon[\gamma] - \varepsilon. \tag{4.1}$$

To prove it, let $\gamma \in \mathcal{M}(X^N)$. Now we have

$$F^*[-A^*\gamma] = \sup \left\{ \int_{X^N} \sum_{i=1}^{N} u_i(x_i) \mathrm{d}\gamma - \sum_{i=1}^{N} \int_X u_i(x_i) \, \mathrm{d}\rho\mathfrak{m}(x_i) \;\middle|\; (u_1, \ldots, u_N) \in C_b(X)^N \right\}$$

$$= \begin{cases} 0 & \text{if } \gamma \in \Pi_N(\rho) \\ +\infty & \text{otherwise} \end{cases}.$$

Let us then compute $G^*[\gamma]$:

$$G^*[\gamma] = \sup_{\psi \in C_b(X^N)} \left\{ \int_{X^N} \psi \, \mathrm{d}\gamma - \varepsilon \int_{X^N} e^{\frac{1}{\varepsilon}(\psi - c)} \, \mathrm{d}\mathfrak{m}_N \right\}.$$

If γ is not absolutely continuous with respect to $\mathfrak{m}_N$, we have $G^*[\gamma] = +\infty$. If $\gamma \ll \mathfrak{m}_N$, then the supremum (that appears in the definition of $G^*[\gamma]$) is realized at $\psi = \varepsilon \log \rho_\gamma + c$; this holds also if the function ρ_γ is not continuous since it can be approximated by a sequence of continuous functions. Thus, we get for $\gamma \ll \mathfrak{m}_N$

$$G^*[\gamma] = \int_{X^N} \left(\rho_\gamma \cdot \psi - \varepsilon e^{\frac{1}{\varepsilon}(\psi - c)} \right) \mathrm{d}\mathfrak{m}_N$$

$$= \int_{X^N} (\varepsilon \rho_\gamma \log \rho_\gamma + c\rho_\gamma - \varepsilon \rho_\gamma) \, \mathrm{d}\mathfrak{m}_N.$$

Hence, if $\gamma \in \Pi_N(\rho)$, we have

$$G^*[\gamma] = C_0[\gamma] + \varepsilon E[\gamma] - \varepsilon.$$

This concludes the duality proof when X is a compact space.

The noncompact case Due to Lemma 2.3, it suffices to prove the claim in the case where the reference measure is $\rho\mathfrak{m}$ instead of $\mathfrak{m}$; the finiteness of the measure $\rho\mathfrak{m}$ now gives access to inner regularity and to the approximability by compact sets. We will for simplicity denote $\rho := \rho\mathfrak{m}$.

The claim is

$$\min_{\gamma \in \Pi_N(\rho)} C_\varepsilon[\gamma] = \sup_{u \in C_b(X)} \left\{ N \int_X u \mathrm{d}\rho - \varepsilon \int_{X^N} \exp\left(\frac{u \oplus \cdots \oplus u - c}{\varepsilon} \right) \mathrm{d}\rho^{\otimes N} \right\} + \varepsilon.$$

For simplicity, let us denote

$$D_\rho(u) := \left\{ N \int_X u \mathrm{d}\rho - \varepsilon \int_{X^N} \exp\left(\frac{u \oplus \cdots \oplus u - c}{\varepsilon} \right) \mathrm{d}\rho^{\otimes N} \right\} + \varepsilon \quad \text{for all } u \in C_b(X).$$

We may assume that $\sup_{u \in C_b(X)} D_\rho(u) > -\infty$; indeed, since we can test with the function $u \equiv 0$, this always holds for cost functions that are bounded from below.

Let us make, in the notation of the primal functional, the dependence on the reference measure explicit by the notation $\gamma \mapsto C_\varepsilon[\gamma \mid \mu]$ when the reference measure on the space X is μ. Thus the original notation $\gamma \mapsto C_\varepsilon[\gamma]$ corresponds to $\gamma \mapsto C_\varepsilon[\gamma \mid \mathfrak{m}]$.

Since the measures ρ and γ are inner regular, there exists a sequence $(K_n)_{n \in \mathbb{N}}$ of compact subsets of X such that

$$\rho(K_n) \to \rho(X) \quad \text{and} \quad \gamma(K_n^N) \to \gamma(X).$$

Let us denote $\gamma_n := \frac{1}{\gamma(K_n^N)} \gamma|_{K_n^N}$ and $\rho_n := \frac{1}{\rho(K_n)} \rho|_{K_n}$. Let us also denote by $\gamma_n^{\min}$ the minimizer of the problem $I_\varepsilon[\rho_n]$. Since γ is the minimizer of the problem $I_\varepsilon(\rho)$ and since (due to the absolute continuity of the integral and continuity of the function $t \mapsto t \log t$)

$$\lim_{n \to \infty} |C_\varepsilon[\gamma_n \mid \rho_n] - C_\varepsilon[\gamma \mid \rho]| = 0, \tag{4.2}$$

we have

$$\lim_{n \to \infty} |C_\varepsilon[\gamma_n \mid \rho_n] - C_\varepsilon[\gamma_n^{\min} \mid \rho_n]| = 0. \tag{4.3}$$

By the duality result proven above for compact spaces, we have for all $n \in \mathbb{N}$

$$\sup_{u \in C_b(K_n)} D_{\rho_n}(u) = C_\varepsilon[\gamma_n^{\min} \mid \rho_n].$$

Again, due to the absolute continuity of the integral, we have

$$\lim_{n \to \infty} \left| \sup_{u \in C_b(K_n)} D_{\rho_n}(u) - \sup_{u \in C_b(X)} D_\rho(u) \right| = 0.$$

Putting these conditions together, we get for all $n \in \mathbb{N}$

$$
\left| C_\epsilon[\gamma \mid \rho] - \sup_{u \in C_b(X)} D_\rho(u) \right|
$$

$$
\leq |C_\epsilon[\gamma \mid \rho] - C_\epsilon[\gamma_n \mid \rho_n]| + \left| C_\epsilon[\gamma_n \mid \rho_n] - C_\epsilon[\gamma_n^{\min} \mid \rho_n] \right|
$$

$$
+ \left| C_\epsilon[\gamma_n^{\min} \mid \rho_n] - \sup_{u \in C_b(K_n)} D_{\rho_n}(u) \right| + \left| \sup_{u \in C_b(K_n)} D_{\rho_n}(u) - \sup_{u \in C_b(X)} D_\rho(u) \right|.
$$

The claim follows by letting $n \to \infty$. $\qquad\square$

References

1. Agueh, M., Carlier, G.: Barycenters in the Wasserstein space. SIAM J. Math. Anal. **43**, 904–924 (2011)
2. Benamou, J.D., Carlier, G., Nenna, L.: A Numerical method to solve multi-marginal optimal transport problems with Coulomb cost. In: Glowinski, R., Osher, S., Yin, W. (eds.) Splitting Methods in Communication, Imaging, Science, and Engineering. Scientific Computation. Springer, Cham (2016)
3. Benamou, J.-D., Carlier, G., Nenna, L.: Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm. Numer. Math. **142**, 33–54 (2019)
4. Buttazzo, G., De Pascale, L., Gori-Giorgi, P.: Optimal-transport formulation of electronic density functional theory. Phys. Rev. A **85**, 062502 (2012)
5. Carlier, G., Duval, V., Peyré, G., Schmitzer, B.: Convergence of entropic schemes for optimal transport and gradient flows. SIAM J. Math. Anal. **49**, 1385–1418 (2017)
6. Cotar, C., Friesecke, G., Klüppelberg, C.: Density functional theory and optimal transportation with coulomb cost. Commun. Pure Appl. Math. **66**, 548–599 (2013)
7. Cotar, C., Friesecke, G., Klüppelberg, C.: Smoothing of transport plans with fixed marginals and rigorous semiclassical limit of the Hohenberg–Kohn functional. Arch. Ration. Mech. Anal. **228**, 891–922 (2018)
8. Cotar, C., Friesecke, G., Klüppelberg, C.: Second order differentiation formula on $RCD^*(K, N)$ spaces. Accepted at JEMS. arXiv:1802.02463 (2018)
9. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems, pp. 2292–2300 (2013)
10. Di Marino, S., De Pascale, L., Colombo, M.: Multimarginal optimal transport maps for 1-dimensional repulsive costs. Can. J. Math. **67**, 350–368 (2015)
11. Di Marino, S., Gerolin, A.: An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. arXiv preprint arXiv:1911.06850 (2019)
12. Di Marino, S., Gerolin, A., Nenna, L.: Optimal transport theory for repulsive costs. In: Topological Optimization and Optimal Transport: In the Applied Sciences, vol. 17 (2017)
13. Ekeland, I., Temam, R.: Analyse convexe et problèmes variationelles. Dunod, Gauthier-Villars, Paris, ix+340 p (1974)
14. Flamary, R., Courty, N.: POT Python Optimal Transport library (2017). https://github.com/rflamary/POT
15. Gerolin, A., Grossi, J., Gori-Giorgi, P.: Kinetic correlation functionals from the entropic regularisation of the strictly-correlated electrons problem. J. Chem. Theory Comput. (2019)
16. Gerolin, A., Kausamo, A., Rajala, T.: Duality theory for multi-marginal optimal transport with repulsive costs in metric spaces. ESAIM Control Optim. Calc. Var. **25**, 62 (2019)
17. Gigli, N., Tamanini, L.: Second order differentiation formula on compact $RCD^*(K, N)$ spaces. arXiv:1701.03932 (2017)

18. Gozlan, N., Léonard, C.: Transport inequalities: a survey. Markov Process. Related Fields **16**, 635–736 (2010)
19. Kellerer, H.G.: Duality theorems for marginal problems. In: Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, vol. 67 (1984)
20. Léonard, C.: From the Schrödinger problem to the Monge–Kantorovich problem. J. Funct. Anal. **262**, 1879–1920 (2012)
21. Léonard, C.: A survey of the Schrödinger problem and some of its connections with optimal transport. Discrete Continuous Dyn. Syst. **34**, 1533–1574 (2014)
22. Lieb, E.H.: Density functionals for Coulomb systems. In: Loss, M., Ruskai, M.B. (eds.) Inequalities. Springer, Berlin, Heidelberg (2002)
23. Mikami, T.: Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. Probab. Theory Related Fields **129**, 245–260 (2004)
24. Nenna, L.: Numerical Methods for Multi-Marginal Opimal Transportation, PhD thesis, Université Paris-Dauphine (2016)
25. Peyré, G., Cuturi, M.: Computational Optimal Transport, vol. 11. Now Publishers Inc, Hanover (2019)
26. Schrödinger, E.: Über die umkehrung der naturgesetze. Verlag Akademie der wissenschaften in kommission bei Walter de Gruyter u, Company (1931)
27. Seidl, M., Di Marino, S., Gerolin, A., Giesbertz, K., Nenna, L., Gori-Giorgi, P.: The strictly-correlated electron functional for spherically symmetric systems revisited. Accepted in Phys. Rev. A. arXiv:1702.05022) (2016)
28. Sturm, K.-T.: On the geometry of metric measure spaces. Acta Math. **196**, 65–131 (2006)