

First-order Γ -expansion for entropic optimal transport with non-degenerate costs

Luca Nenna^{*†}, Brendan Pass[‡], Paul Pegon[§] and Louis Tocquec[¶]

October 7, 2026

Entropy-regularized optimal transport (EOT) has become a key tool in optimal transport, providing both theoretical insight and efficient computation via the Sinkhorn algorithm. We study the first-order Γ -expansion of EOT as the regularization parameter ε vanishes, by rewriting EOT as the minimization of a relative entropy with respect to a Gibbs measure $\mathbb{P}_\varepsilon$ associated with the duality gap energy. Using a mild extension of the Laplace method, we prove the convergence of the Gibbs measures and derive this Γ -expansion for non-degenerate ground costs, under suitable assumptions on the duality gap, the contact set and the marginals. This yields at the same time an entropic selection principle and an expansion of the EOT cost up to $o(\varepsilon)$ for this class of ground costs, generalizing existing results.

Keywords: optimal transport, entropic regularization, Γ -convergence, Γ -expansion, entropic selection, non-degenerate costs, Laplace method, Gibbs measures, multi-marginal optimal transport.

MSC 2020: 49Q22 (primary); 49J45, 41A60, 60B10, 94A17.

1 Introduction

Optimal transport (OT) has emerged as a central research topic at the intersection of probability theory, partial differential equations, and optimization, with applications in

^{*}Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, ParMA, Inria Saclay, 91405, Orsay, France. email: luca.nenna@universite-paris-saclay.fr

[†]Institut Universitaire de France, Paris, France.

[‡]Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1. email: pass@ualberta.ca

[§]CEREMADE, CNRS UMR 7534, Université Paris-Dauphine – PSL, Paris, France & MOKAPLAN, Inria Paris, Paris, France. Email: pegon@ceremade.dauphine.fr

[¶]Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay & MOKAPLAN, Inria Paris, 75016 Paris, France. Email: louis.tocquec@universite-paris-saclay.fr

fields such as economics, machine learning, and image processing. The monographs of Villani [Vil08] and Santambrogio [San15] provide comprehensive accounts of its mathematical foundations. Applications to economics are developed in [Gal16], while [PC19] surveys computational methods and applications to data science. Large discrete transport problems can nevertheless be expensive to solve, and high-dimensional statistical estimation presents additional difficulties.

These computational challenges have renewed interest in entropy-regularized optimal transport (EOT). Entropic regularization leads to a strictly convex problem whose discrete solution can be computed by iterative matrix scaling, following the work of Sinkhorn [Sin64; Sin67]. Its use for large-scale transport was popularized by Cuturi [Cut13] and developed through iterative Bregman projections in [Ben+15]. Convergence results for the Sinkhorn algorithm range from the matrix-scaling analysis of Franklin and Lorenz [FL89] to continuous and multi-marginal settings [DG20; EN22], and complexity guarantees for approximate discrete OT are established in [ANR17]. Statistical benefits at fixed regularization have also been quantified for Sinkhorn divergences [Gen+19] and entropic transport costs [MN19], with bounds that depend on the regularization parameter. We refer to [PC19; Nut21] for broader perspectives on EOT and its applications.

EOT also has a probabilistic interpretation through the Schrödinger problem and variational problems for stochastic processes [Mik90; Léo14], where ε plays the role of a noise or temperature parameter. It is therefore natural to investigate the behavior of the optimal plans and values as ε tends to zero. The zero-noise connection with quadratic OT was established by Mikami [Mik04], and variational convergence was subsequently developed in [Léo12; Car+17] in the two-marginal case, and in [GKR20] in the multi-marginal case, including singular repulsive costs. Convergence of the associated potentials is studied in [NW22], and a complementary geometric approach is given by Bernton, Ghosal, and Nutz [BGN22], who describe concentration of the entropic optimizers through large-deviation estimates and relate the rate function to Kantorovich potentials.

Finer asymptotics of the optimal value have been studied in several settings. For the quadratic cost, a next-order variational convergence of the entropic cost, viewed as a function of one marginal, arises in the connection between large deviations and Wasserstein gradient flows [Ada+11; DLR13; EMR15]. Pal [Pal24] identifies the next asymptotic term for smooth strictly convex displacement costs under suitable regularity and elliptic assumptions on the potentials, while Conforti and Tamanini [CT21] obtain differentiation formulas and a second-order expansion in the Schrödinger setting. For general C^2 non-degenerate ground costs c (i.e. such that the mixed Hessian $D_{xy}^2 c$ is invertible), Carlier, Pegon, and Tamanini [CPT23] establish the leading correction $\frac{d}{2}\varepsilon \log(1/\varepsilon) + O(\varepsilon)$, which is extended to the multi-marginal setting by Nenna and Pegon [NP25]. A different approach based on quantization of measures yields further convergence bounds in [EN24]. The separate behavior of the transport and entropy terms appearing in the total EOT cost, together with quantitative convergence of the plans, is studied by Malamut and Sylvestre [MS25].

When the unregularized problem admits several minimizers, whether the entropic plans converge and, if so, which optimal plan is selected is an open problem in general.

For discrete problems, the work of Cominetti and San Martín [CS94] on exponential penalties in linear programming establishes convergence to the entropy-minimizing optimal solution, with exponential convergence rates, and non-asymptotic quantitative bounds were later obtained by Weed [Wee18]. In the continuous setting, a direct entropy selection argument still applies when an optimal plan has finite relative entropy [Nut21], but the main difficulty arises when every optimal plan is singular with respect to the product of the marginals, so that the relative entropy is infinite on the entire optimal set.

For the Euclidean distance cost on the real line, which is neither C^1 nor non-degenerate, the set of admissible destinations of a typical source x is a one-dimensional segment, called a transport ray. In this case, Di Marino and Louet [DL18] identify an entropic selection through a renormalized Γ -limit under suitable integrability assumptions. Ley [Ley25] obtains complementary convergence results for less regular marginals under a componentwise mutual-singularity condition. In higher dimensions, selection along transport rays is studied by Aryan and Ghosal [AG25] and more recently by Nutz and Zhong [NZ26], who establish a variational characterization of the selected plan as a constrained relative-entropy minimizer with geometry-dependent reference measures on transport rays.

Main contributions and strategy. The main purpose of this paper is to derive an entropic selection principle, through a Γ -convergence approach, when the ground cost c is non-degenerate but not necessarily twisted. In this situation, the set of admissible destinations of a typical source x is a locally finite set but need not be a singleton. We first consider the case of two marginals (μ_1, μ_2) on the flat torus $\mathbb{T}^d$, to avoid boundary issues, then explain how to handle the case of $\mathbb{R}^d$. We assume that the contact set Σ , defined as the zero set of the duality gap energy $E \stackrel{\text{def.}}{=} c - \bar{\varphi}_1 \oplus \bar{\varphi}_2$, $(\bar{\varphi}_1, \bar{\varphi}_2)$ being a pair of Kantorovich potentials, is a d -dimensional $C^{1,1}$ submanifold. This is a strong assumption: under non-degeneracy alone, McCann, Pass and Warren [MPW12] only guarantee that Σ is contained in a d -dimensional locally Lipschitz submanifold. We impose it to keep technicalities under control, since the construction of a recovery sequence is already delicate in the $C^{1,1}$ setting, leaving the general case for future work.

Under our assumptions, the normal Hessian $\text{Hess}^\perp E$ of E , defined on Σ as the Hessian of E restricted to the normal space of Σ , is positive definite. Roughly speaking, the Gibbs measure $\mathbb{P}_\varepsilon = Z_\varepsilon^{-1} e^{-E/\varepsilon} \mu_1 \otimes \mu_2$ behaves like a Gaussian with covariance matrix $\varepsilon(\text{Hess}^\perp E)^{-1}$ in the normal fibers to the manifold, and concentrates on Σ as $\varepsilon \rightarrow 0$. We extend the results of Hwang [Hwa80], who proves convergence towards a probability measure $\mathbb{P}_0$ with an explicit density with respect to the volume measure on Σ , from C^2 to $C^{1,1}$ regularity, and to compactly supported marginals on $\mathbb{R}^d$, in which case Σ need not be compact. We then prove that the relative entropy with respect to the concentrating Gibbs probabilities, which coincides with the entropy-regularized functional up to an additive constant and a factor ε (see Proposition 2.1), Γ -converges under the original marginal constraints to the relative entropy with respect to this limiting probability. Since there always exists an optimal plan with finite limiting entropy, the limit problem

has a unique minimizer, which is the selected optimal plan, and the entropic plans converge to it.

The principal difficulty lies in the construction of a recovery sequence for the Γ -lim sup, which must approximate the relevant Gaussian profile in the normal fibers while satisfying the marginal constraints exactly. We first perturb an arbitrary admissible plan with finite limiting entropy by a varying Gaussian kernel in the normal fibers, so as to obtain an approximate recovery sequence that converges in energy but only satisfies the marginal constraints in the limit. The most delicate step is then to estimate how much these constraints are violated along the sequence. Roughly speaking, the centering of the normal kernels makes the first-order contributions cancel, so that the squared Wasserstein distance between the marginals of the perturbed plans and the prescribed ones is $o(\varepsilon)$, although the normal variance is of order ε . Thanks to this estimate, the shadow construction of [EN22] finally allows us to correct the marginals while preserving convergence in energy.

To state the resulting Γ -expansion, in the sense of Braides and Truskinovsky [BT08], we consider the functionals defined on $\mathcal{P}(\Omega)$, where $\Omega \stackrel{\text{def.}}{=} \mathbb{T}^d \times \mathbb{T}^d$, by

$$\begin{aligned}\mathcal{F}_\varepsilon(\gamma) &\stackrel{\text{def.}}{=} \int_{\Omega} c \, d\gamma + \varepsilon \text{Ent}(\gamma \mid \mu_1 \otimes \mu_2) + \chi_{\Pi(\mu_1, \mu_2)}(\gamma) \quad (\varepsilon > 0), \\ \mathcal{F}^{(0)}(\gamma) &\stackrel{\text{def.}}{=} \int_{\Omega} c \, d\gamma + \chi_{\Pi(\mu_1, \mu_2)}(\gamma),\end{aligned}$$

where $\Pi(\mu_1, \mu_2)$ denotes the set of couplings $\gamma \in \mathcal{P}(\Omega)$ having μ_1 and μ_2 as marginals, and χ denotes the indicator function in the sense of convex analysis. Writing $\mu_i = \rho_i \, dx$, we also define the first-order functional

$$\mathcal{F}^{(1)}(\gamma) \stackrel{\text{def.}}{=} \text{Ent}(\gamma \mid \mathbb{P}_0) + \chi_{\Pi(\mu_1, \mu_2)}(\gamma) - \log Z_0 - \frac{d}{2} \log(2\pi),$$

where $\mathbb{P}_0$ is the probability measure on Σ given by

$$d\mathbb{P}_0(x_1, x_2) \stackrel{\text{def.}}{=} \frac{1}{Z_0} \frac{\rho_1(x_1)\rho_2(x_2)}{\sqrt{\det \text{Hess}^\perp E(x_1, x_2)}} d\mathcal{H}_{|\Sigma}^d(x_1, x_2),$$

where $\mathcal{H}_{|\Sigma}^d$ denotes the d -dimensional Hausdorff measure restricted to Σ and Z_0 is the normalizing constant.

Theorem A (Γ -expansion and entropic selection). *Assume that (H1)–(H4) of Section 2.2 hold. In particular, $M_1 = M_2 = \mathbb{T}^d$, c is non-degenerate near Σ , and Σ is a d -dimensional $C^{1,1}$ submanifold. Then the following holds.*

(i) (Γ -expansion) *The functionals $\mathcal{F}_\varepsilon$ admit the first-order Γ -expansion*

$$\boxed{\mathcal{F}_\varepsilon - \frac{d}{2} \varepsilon \log(1/\varepsilon) \stackrel{\Gamma}{=} \mathcal{F}^{(0)} + \varepsilon \mathcal{F}^{(1)} + o(\varepsilon)}, \quad (1.1)$$

which means that $(\mathcal{F}_\varepsilon - \min \mathcal{F}^{(0)})/\varepsilon - \frac{d}{2} \log(1/\varepsilon)$ Γ -converges to $\mathcal{F}^{(1)}$ for the weak convergence of probability measures.

As a consequence:

(ii) (Asymptotics of the cost) The minimum of $\mathcal{F}^{(1)}$ is finite, and

$$\min \mathcal{F}_\varepsilon = \min \mathcal{F}^{(0)} + \frac{d}{2}\varepsilon \log(1/\varepsilon) + \varepsilon \min \mathcal{F}^{(1)} + o(\varepsilon). \quad (1.2)$$

(iii) (Entropic selection) The unique minimizer γ_ε of $\mathcal{F}_\varepsilon$ converges weakly, as $\varepsilon \rightarrow 0$, to the unique minimizer γ_0 of $\text{Ent}(\cdot | \mathbb{P}_0)$ among all optimal transport plans, which is also the unique minimizer of $\mathcal{F}^{(1)}$.

We then treat compactly supported marginals on $\mathbb{R}^d$ and illustrate the selection principle for the non-twisted cost of [MPW12, Example 3.1]. The extension to N marginals is stated separately under an explicit normal-rank condition. Finally, we apply it to a repulsive harmonic problem with three marginals on the circle.

Organization of the paper. Section 2 introduces the transport problems, our assumptions, and the geometry of the contact set. Section 3 deals with the contact set and its tubular neighborhood, and states the geometric Laplace method. Section 4 constructs the recovery sequence, proves the constrained Γ -convergence, and derives the Γ -expansion and its consequences. Section 5 gives the proof of the geometric Laplace method in the case of the flat torus and the Euclidean space, and provides the multi-marginal extension of our results, together with some examples.

2 Entropic Optimal Transport, assumptions and geometry of the contact set

2.1 Preliminaries on Entropic Optimal Transport

We consider two compact metric spaces M_1 and M_2 , endowed with probability measures μ_1 and μ_2 , and a continuous cost function $c : \Omega \rightarrow \mathbb{R}$, where $\Omega \stackrel{\text{def.}}{=} M_1 \times M_2$. Typical examples are closed Riemannian manifolds, such as the flat torus of Section 2.2, or compact subsets of $\mathbb{R}^d$, as in Section 5.1. The Monge–Kantorovich optimal transport problem reads

$$\inf_{\gamma} \left\{ \int_{\Omega} c(\mathbf{x}) \, d\gamma(\mathbf{x}) \mid \gamma \in \Pi(\mu_1, \mu_2) \right\}, \quad (\text{OT})$$

where

$$\Pi(\mu_1, \mu_2) \stackrel{\text{def.}}{=} \left\{ \gamma \in \mathcal{P}(\Omega) \mid (\text{proj}_{M_i})_{\#} \gamma = \mu_i \text{ for every } i = 1, 2 \right\}$$

denotes the set of probability measures on Ω having μ_1 and μ_2 as marginals.

It is well known that the optimal transport problem admits a dual formulation, for which strong duality holds and maximizers exist as soon as M_1, M_2 are compact and c is continuous (see [San15, Chapter 1]). The dual problem is given by

$$\sup_{(\varphi_1, \varphi_2)} \left\{ \int_{M_1} \varphi_1 \, d\mu_1 + \int_{M_2} \varphi_2 \, d\mu_2 \mid \varphi_1 \oplus \varphi_2 \leq c, \varphi_i \in C(M_i) \text{ for } i = 1, 2 \right\}, \quad (\text{Dual})$$

where $(\varphi_1 \oplus \varphi_2)(x_1, x_2) \stackrel{\text{def.}}{=} \varphi_1(x_1) + \varphi_2(x_2)$. Any pair $(\bar{\varphi}_1, \bar{\varphi}_2)$ attaining the supremum in (Dual) is called a pair of Kantorovich potentials. Given such a pair, we define¹ the duality gap energy $E : \Omega \rightarrow \mathbb{R}_+$ and the contact set respectively by

$$E \stackrel{\text{def.}}{=} c - \bar{\varphi}_1 \oplus \bar{\varphi}_2 \quad \text{and} \quad \Sigma \stackrel{\text{def.}}{=} \{\mathbf{x} \in \Omega : E(\mathbf{x}) = 0\}.$$

Since E is continuous and nonnegative, strong duality yields the following characterization of optimality:

$$\bar{\gamma} \in \Pi(\mu_1, \mu_2) \text{ is a solution of (OT)} \iff \text{spt}(\bar{\gamma}) \subseteq \Sigma. \quad (2.1)$$

For any $\varepsilon > 0$, the entropy-regularized two-marginal optimal transport problem is defined by

$$\inf_{\gamma} \left\{ \int_{\Omega} c(\mathbf{x}) d\gamma(\mathbf{x}) + \varepsilon \text{Ent}(\gamma \mid \mu_1 \otimes \mu_2) \mid \gamma \in \Pi(\mu_1, \mu_2) \right\}, \quad (\text{OT}_{\varepsilon})$$

where, for $\gamma \in \mathcal{P}(\Omega)$ and a probability measure ν on Ω , the relative entropy is given by

$$\text{Ent}(\gamma \mid \nu) \stackrel{\text{def.}}{=} \begin{cases} \int_{\Omega} \log \left(\frac{d\gamma}{d\nu}(\mathbf{x}) \right) d\gamma(\mathbf{x}) & \text{if } \gamma \ll \nu, \\ +\infty & \text{otherwise.} \end{cases}$$

We use the same notation when ν is a nonnegative Borel measure which is not a probability (e.g. a Hausdorff measure), whenever the integral is well defined.

We start by making the following observation.

Proposition 2.1. *For $\varepsilon > 0$, let $\mathbb{P}_{\varepsilon}$ be the Gibbs probability measure defined by*

$$\mathbb{P}_{\varepsilon} \stackrel{\text{def.}}{=} \frac{1}{Z_{\varepsilon}} \exp \left(-\frac{E}{\varepsilon} \right) \mu_1 \otimes \mu_2, \quad Z_{\varepsilon} \stackrel{\text{def.}}{=} \int_{\Omega} \exp \left(-\frac{E(\mathbf{x})}{\varepsilon} \right) d(\mu_1 \otimes \mu_2)(\mathbf{x}).$$

Then, for every $\gamma \in \Pi(\mu_1, \mu_2)$,

$$\int_{\Omega} c d\gamma + \varepsilon \text{Ent}(\gamma \mid \mu_1 \otimes \mu_2) = \text{OT}(\mu_1, \mu_2) + \varepsilon \text{Ent}(\gamma \mid \mathbb{P}_{\varepsilon}) - \varepsilon \log Z_{\varepsilon}, \quad (2.2)$$

where $\text{OT}(\mu_1, \mu_2)$ denotes the value of (OT). In particular, problem (OT_ε) has the same minimizers as

$$\inf_{\gamma} \left\{ \text{Ent}(\gamma \mid \mathbb{P}_{\varepsilon}) \mid \gamma \in \Pi(\mu_1, \mu_2) \right\}, \quad (\text{EOT}_{\varepsilon})$$

Proof. Let $\gamma \in \Pi(\mu_1, \mu_2)$. Since γ has marginals μ_1 and μ_2 , strong duality gives $\int_{\Omega} \bar{\varphi}_1 \oplus \bar{\varphi}_2 d\gamma = \text{OT}(\mu_1, \mu_2)$, hence $\int_{\Omega} c d\gamma = \text{OT}(\mu_1, \mu_2) + \int_{\Omega} E d\gamma$. Moreover, since E is bounded, the measures $\mathbb{P}_{\varepsilon}$ and $\mu_1 \otimes \mu_2$ are equivalent, with bounded log-density $\log \frac{d\mathbb{P}_{\varepsilon}}{d(\mu_1 \otimes \mu_2)} = -\frac{E}{\varepsilon} - \log Z_{\varepsilon}$, so that

$$\varepsilon \text{Ent}(\gamma \mid \mu_1 \otimes \mu_2) = \varepsilon \text{Ent}(\gamma \mid \mathbb{P}_{\varepsilon}) - \int_{\Omega} E d\gamma - \varepsilon \log Z_{\varepsilon},$$

both sides being possibly infinite. Summing the two identities gives (2.2). $\square$

¹ E may depend on the choice of Kantorovich potentials, but the Γ -expansion result actually will not.

2.2 Main assumptions and geometry of the contact set

We first consider two marginals on flat tori. A Euclidean extension is given in [Section 5.1](#).

- (H1) For $i = 1, 2$, $M_i = \mathbb{T}^d$, with the flat product metric. The cost c is continuous, and we fix optimal continuous Kantorovich potentials $\bar{\varphi}_1, \bar{\varphi}_2$. The cost is C^2 and satisfies the non-degeneracy (or infinitesimal twist) condition on an open neighborhood $U \subset \Omega$ of the contact set Σ :

$$D_{x_1 x_2}^2 c(z) : T_{x_2} M_2 \longrightarrow T_{x_1} M_1 \quad \text{is invertible for every } z = (x_1, x_2) \in U.$$

- (H2) The duality gap $E(x_1, x_2) = c(x_1, x_2) - \bar{\varphi}_1(x_1) - \bar{\varphi}_2(x_2) \geq 0$ belongs to $C^{2,1}(\Omega)$.

- (H3) The contact set $\Sigma = \{E = 0\}$ is a nonempty $C^{1,1}$ submanifold without boundary, of dimension d . We denote by $k \stackrel{\text{def.}}{=} d \geq 1$ its codimension in Ω .

- (H4) The marginals have densities $\mu_i = \rho_i \, d \text{vol}_{M_i}$, with ρ_i Lipschitz and $0 < \alpha \leq \rho_i \leq \beta < \infty$ on M_i , for $i = 1, 2$.

Remark 2.1 (About assumption (H3)). Under (H1)–(H2), the implicit function theorem shows that $S \stackrel{\text{def.}}{=} \{z \in U : \nabla_{x_1} E(z) = 0\}$ is a d -dimensional $C^{1,1}$ submanifold, since $D_{x_2}(\nabla_{x_1} E) = D_{x_1 x_2}^2 c$ is invertible. As $\Sigma \subseteq S$, (H3) amounts to Σ being relatively open in S (compare with [MPW12], where Σ is shown to lie in a Lipschitz submanifold). Moreover, (H3) implies that $\Sigma = \{\nabla E = 0\}$ in a neighborhood of Σ , by the positive definiteness of the normal Hessian (see [Proposition 2.2](#)). Requiring this equality globally would however not be natural: $\{\nabla E = 0\}$ contains the maximum points of E and need not be a d -dimensional submanifold. For the bilinear cost $c(x_1, x_2) = -\langle x_1, x_2 \rangle$ in the Euclidean setting, $E(x_1, x_2) = u(x_1) + u^*(x_2) - \langle x_1, x_2 \rangle$ with u convex is convex in each variable, so that $S = \Sigma$ and (H2) implies (H3).

We set $\rho(x_1, x_2) = \rho_1(x_1)\rho_2(x_2)$ and $\mu^\otimes = \mu_1 \otimes \mu_2 = \rho \, d \text{vol}_\Omega$. For $z \in \Sigma$, the normal Hessian is

$$H_z E \stackrel{\text{def.}}{=} \text{Hess}^\perp E(z) = \text{Hess } E(z)|_{N_z \Sigma}, \quad N_z \Sigma = (T_z \Sigma)^\perp.$$

Proposition 2.2. *Under (H1)–(H3),*

$$\ker \text{Hess } E(z) = T_z \Sigma \quad (z \in \Sigma),$$

and both coordinate projections $\text{proj}_{M_i}^\Sigma : \Sigma \rightarrow M_i$ are local $C^{1,1}$ diffeomorphisms. Moreover, there exist constants $0 < \lambda \leq \Lambda < \infty$ such that

$$\lambda |v|^2 \leq H_z E(v, v) \leq \Lambda |v|^2 \quad (z \in \Sigma, \, v \in N_z \Sigma). \quad (2.3)$$

Proof. Since $E \geq 0$ and $E = 0$ on Σ , we have $\nabla E = 0$ on Σ , and $\text{Hess } E(z)$ is positive semidefinite there. Differentiating $\nabla E = 0$ along Σ gives $T_z \Sigma \subseteq \ker \text{Hess } E(z)$. The separated form of the potentials implies $D_{x_1 x_2}^2 E = D_{x_1 x_2}^2 c$ near Σ : this identity also follows directly by taking mixed differences, without initially assuming differentiability of

the potentials. Thus the mixed block of $\text{Hess } E(z)$ has rank d , so $\text{rank Hess } E(z) \geq d$. The preceding kernel inclusion and $\dim T_z \Sigma = d$ give the reverse inequality. Consequently, $\ker \text{Hess } E(z) = T_z \Sigma$.

The differential of $\text{proj}_{M_1}^\Sigma$ at z is the restriction of $(u, v) \mapsto u$ to $T_z \Sigma$, whose kernel consists of the vectors $(0, v) \in T_z \Sigma = \ker \text{Hess } E(z)$. For such a vector, the first component of $\text{Hess } E(z)[(0, v)] = 0$ reads $D_{x_1 x_2}^2 c(z)v = 0$, whence $v = 0$. Thus $D \text{proj}_{M_1}^\Sigma(z)$ is injective, hence an isomorphism since $\dim T_z \Sigma = d = \dim M_1$, and the inverse function theorem shows that $\text{proj}_{M_1}^\Sigma$ is a local $C^{1,1}$ diffeomorphism. The same argument applies to $\text{proj}_{M_2}^\Sigma$. Finally, positive semidefiniteness and the kernel identity $\ker \text{Hess } E(z) = T_z \Sigma$ make the restriction $H_z E$ of $\text{Hess } E(z)$ to $N_z \Sigma = (T_z \Sigma)^\perp$ positive definite. Continuity of the Hessian and the normal spaces, together with compactness of Σ , gives (2.3). $\square$

Remark 2.2 (Vertical and horizontal branches for the two-marginal problem). By **Proposition 2.2**, both projections $\text{proj}_{M_1}^\Sigma$ and $\text{proj}_{M_2}^\Sigma$ are local diffeomorphisms. In particular, Σ has neither vertical nor horizontal branches, that is, points $z \in \Sigma$ such that $\text{rank}(D \text{proj}_{M_1}^\Sigma(z)) < d$, respectively $\text{rank}(D \text{proj}_{M_2}^\Sigma(z)) < d$.

3 Convergence of the Gibbs measure

3.1 Tubular coordinates around the contact set

By (H3), Σ is a compact $C^{1,1}$ -submanifold of Ω , and thus has positive reach in the sense of Federer [Fed59]. Consequently, there exists $r_0 > 0$ such that, setting, for every $0 < r \leq r_0$,

$$N_r \Sigma \stackrel{\text{def.}}{=} \{(z, v) : z \in \Sigma, v \in N_z \Sigma, \|v\| < r\}, \quad \Sigma_r \stackrel{\text{def.}}{=} \{\mathbf{x} \in \Omega : \text{dist}(\mathbf{x}, \Sigma) < r\},$$

the nearest-point projection $\text{proj}_\Sigma : \Sigma_{r_0} \rightarrow \Sigma$ is well defined and Lipschitz, and the map $(z, v) \mapsto z + v$ is a bi-Lipschitz bijection from $N_r \Sigma$ onto Σ_r , with inverse $\mathbf{x} \mapsto (\text{proj}_\Sigma(\mathbf{x}), \mathbf{x} - \text{proj}_\Sigma(\mathbf{x}))$. Here the addition is understood componentwise and modulo $\mathbb{Z}^d$ on each torus factor.

Lemma 3.1 (Lipschitz normal frame). *Under (H1)–(H3), there exists a Lipschitz map $n = (n_1, \dots, n_k) : \Sigma \rightarrow \mathbb{R}^{2d \times k}$ such that, for every $z \in \Sigma$, the columns $n_1(z), \dots, n_k(z)$ form an orthonormal basis of $N_z \Sigma$.*

Proof. By **Proposition 2.2**, the projection $\text{proj}_{M_1}^\Sigma$ is a local $C^{1,1}$ diffeomorphism, so that, locally, Σ is the graph of a $C^{1,1}$ map $x_1 \mapsto x_2$. Identifying the tangent spaces of the flat tori with $\mathbb{R}^d$, this means that, for every $z \in \Sigma$, $T_z \Sigma = \{(u, B(z)u) : u \in \mathbb{R}^d\}$ for some matrix $B(z) \in \mathbb{R}^{d \times d}$, where $z \mapsto B(z)$ is locally Lipschitz, hence Lipschitz on Σ by compactness. Hence $N_z \Sigma = \{(-B(z)^\top w, w) : w \in \mathbb{R}^d\}$, that is, the columns of

$$N(z) \stackrel{\text{def.}}{=} \begin{pmatrix} -B(z)^\top \\ I_d \end{pmatrix} \in \mathbb{R}^{2d \times k}$$

form a basis of $N_z\Sigma$. Since their Gram matrix $N(z)^\top N(z) = I_d + B(z)B(z)^\top$ is symmetric positive definite, we obtain an orthonormal basis of $N_z\Sigma$ by setting

$$n(z) \stackrel{\text{def.}}{=} N(z)(I_d + B(z)B(z)^\top)^{-1/2}, \quad \text{so that} \quad n(z)^\top n(z) = I_k.$$

The map n is Lipschitz, as B is. $\square$

We fix such a frame n . For every $z \in \Sigma$ and $s \in \mathbb{R}^k$, we thus have

$$n(z)s = \sum_{p=1}^k s_p n_p(z) \in N_z\Sigma.$$

Through the isometry $n(z) : \mathbb{R}^k \rightarrow N_z\Sigma$, we write $H_z E(s, s)$ for $H_z E(n(z)s, n(z)s)$, so that (2.3) holds for every $s \in \mathbb{R}^k$, and $\det H_z E$ for the determinant of the $k \times k$ matrix with entries $H_z E(n_p(z), n_q(z))$, which does not depend on the choice of the orthonormal frame. Denoting by $\mathbb{B}_r$ the open ball of radius r centered at 0 in $\mathbb{R}^k$, we define the tubular parametrization

$$\Phi(z, s) \stackrel{\text{def.}}{=} z + n(z)s, \quad z \in \Sigma, \quad s \in \mathbb{B}_{r_0},$$

which, by the above, is a bi-Lipschitz bijection from $\Sigma \times \mathbb{B}_r$ onto Σ_r for every $0 < r \leq r_0$. Moreover, since $z = \text{proj}_\Sigma(\Phi(z, s))$, we have

$$\text{dist}(\Phi(z, s), \Sigma) = |s| \quad \text{for every } (z, s) \in \Sigma \times \mathbb{B}_{r_0}, \quad (3.1)$$

which will be used repeatedly in the sequel. Since n is Lipschitz, it is differentiable along Σ at $\mathcal{H}^d$ -a.e. point, by Rademacher's theorem. For such $z \in \Sigma$ and $p \in \{1, \dots, k\}$, we denote by

$$G_p(z) : T_z\Sigma \longrightarrow T_z\Sigma$$

the linear operator determined by the tangential component of $Dn_p(z)$, namely

$$\text{proj}_{T_z\Sigma}(Dn_p(z)[X]) = G_p(z)X, \quad X \in T_z\Sigma,$$

where $\text{proj}_{T_z\Sigma}$ denotes the orthogonal projection of $\mathbb{R}^{2d}$ onto $T_z\Sigma$. For $s \in \mathbb{R}^k$, we set

$$G(z, s) \stackrel{\text{def.}}{=} \sum_{p=1}^k s_p G_p(z).$$

Lemma 3.2. *Under (H1)–(H3), for $\mathcal{H}^d$ -a.e. $z \in \Sigma$ and every $s \in \mathbb{B}_{r_0}$, the map Φ is differentiable at (z, s) , and its Jacobian $\mathcal{J}_\Phi(z, s)$, that is, the absolute value of the determinant of $D\Phi(z, s) : T_z\Sigma \times \mathbb{R}^k \rightarrow \mathbb{R}^{2d}$ in orthonormal bases, is*

$$\mathcal{J}_\Phi(z, s) = \left| \det(\text{Id}_{T_z\Sigma} + G(z, s)) \right|.$$

Moreover, there exists $C > 0$ such that

$$\|G(z, s)\| \leq C|s|$$

for $\mathcal{H}^d$ -a.e. $z \in \Sigma$ and every $s \in \mathbb{R}^k$.

Proof. Fix $z \in \Sigma$ at which n is differentiable along Σ . Since $\Phi(z, s) = z + n(z)s$ is affine in s , the map Φ is differentiable at (z, s) for every $s \in \mathbb{B}_{r_0}$, and, for $X \in T_z \Sigma$ and $t \in \mathbb{R}^k$,

$$D\Phi(z, s)[X, t] = X + \sum_{p=1}^k s_p Dn_p(z)[X] + n(z)t = (X + G(z, s)X) + Q(z, s)X + n(z)t,$$

where $Q(z, s)X \stackrel{\text{def.}}{=} \text{proj}_{N_z \Sigma}(\sum_{p=1}^k s_p Dn_p(z)[X]) \in N_z \Sigma$. Hence, decomposing the target space as $\mathbb{R}^{2d} = T_z \Sigma \oplus N_z \Sigma$, the operator $D\Phi(z, s) : T_z \Sigma \times \mathbb{R}^k \rightarrow T_z \Sigma \oplus N_z \Sigma$ has the block lower triangular form

$$D\Phi(z, s) = \begin{pmatrix} \text{Id}_{T_z \Sigma} + G(z, s) & 0_{\mathbb{R}^k \rightarrow T_z \Sigma} \\ Q(z, s) & n(z) \end{pmatrix}.$$

Let $(e_1, \dots, e_d)$ be an orthonormal basis of $T_z \Sigma$. By [Lemma 3.1](#), the vectors $e_1, \dots, e_d, n_1(z), \dots, n_k(z)$ form an orthonormal basis of $\mathbb{R}^{2d}$, in which the block $n(z)$ has matrix I_k . Computing $\mathcal{J}_\Phi(z, s)$ in this basis and in the orthonormal basis of $T_z \Sigma \times \mathbb{R}^k$ formed by $(e_1, \dots, e_d)$ and the canonical basis of $\mathbb{R}^k$, we obtain $\mathcal{J}_\Phi(z, s) = |\det(\text{Id}_{T_z \Sigma} + G(z, s))|$. Finally, denoting by $\text{Lip}(n)$ the Lipschitz constant of n , we have $\|G_p(z)\| \leq \|Dn_p(z)\| \leq \text{Lip}(n)$ for every p , whence $\|G(z, s)\| \leq \sum_{p=1}^k |s_p| \|G_p(z)\| \leq \sqrt{k} \text{Lip}(n) |s|$. $\square$

Since Φ is a bi-Lipschitz bijection from $\Sigma \times \mathbb{B}_r$ onto Σ_r , the area formula gives, for every $0 < r \leq r_0$ and every nonnegative Borel function f on Σ_r ,

$$\int_{\Sigma_r} f \, d\text{vol}_\Omega = \int_\Sigma \int_{\mathbb{B}_r} f(\Phi(z, s)) \mathcal{J}_\Phi(z, s) \, ds \, d\mathcal{H}^d(z). \quad (3.2)$$

After decreasing r_0 , we may assume that $Cr_0 \leq 1/2$, so that $\|G(z, s)\| \leq 1/2$ for $\mathcal{H}^d$ -a.e. $z \in \Sigma$ and every $s \in \mathbb{B}_{r_0}$. In particular, $\det(\text{Id}_{T_z \Sigma} + G(z, s)) > 0$, and expanding the determinant gives

$$\mathcal{J}_\Phi(z, s) = 1 + L(z, s) + O(|s|^2), \quad L(z, s) \stackrel{\text{def.}}{=} \text{tr } G(z, s), \quad (3.3)$$

where the remainder is uniform in z . The function $L(z, \cdot)$ is linear, hence odd, and $|L(z, s)| \leq dC|s|$. In particular,

$$\mathcal{J}_\Phi(z, \sqrt{\varepsilon}s) = 1 + \sqrt{\varepsilon}L(z, s) + O(\varepsilon|s|^2) \quad \text{whenever } \sqrt{\varepsilon}|s| < r_0. \quad (3.4)$$

3.2 Geometric Laplace's method

Proposition 3.3. *Under [\(H1\)](#)–[\(H3\)](#), there exist constants $r > 0$ and $0 < \underline{C} \leq \overline{C} < +\infty$ such that, for every $\mathbf{x} \in \Omega$ satisfying $\text{dist}(\mathbf{x}, \Sigma) \leq r$, one has*

$$\underline{C} \, \text{dist}(\mathbf{x}, \Sigma)^2 \leq E(\mathbf{x}) \leq \overline{C} \, \text{dist}(\mathbf{x}, \Sigma)^2.$$

Proof. Let $r_0 > 0$ be as in [Section 3.1](#), and let $\mathbf{x} \in \Sigma_{r_0}$. Then $\mathbf{x} = \Phi(z, s)$ with $z = \text{proj}_\Sigma(\mathbf{x})$ and a unique $s \in \mathbb{B}_{r_0}$, and $|s| = \text{dist}(\mathbf{x}, \Sigma)$ by [\(3.1\)](#). Since z is a minimum point of $E \geq 0$, one has $E(z) = 0$ and $\nabla E(z) = 0$, so that a Taylor expansion of E gives

$$E(\mathbf{x}) = \frac{1}{2} H_z E(s, s) + \mathcal{R}(z, s), \quad \mathcal{R}(z, s) = O(|s|^3) \quad \text{as } s \rightarrow 0,$$

uniformly with respect to $z \in \Sigma$, since E belongs to $C^{2,1}(\Omega)$. By [\(2.3\)](#), $\lambda|s|^2 \leq H_z E(s, s) \leq \Lambda|s|^2$. Taking $r < r_0$ small enough, we have $|\mathcal{R}(z, s)| \leq \frac{\lambda}{4}|s|^2$ whenever $z \in \Sigma$ and $|s| \leq r$. Consequently,

$$\frac{\lambda}{4} \text{dist}(\mathbf{x}, \Sigma)^2 \leq E(\mathbf{x}) \leq \frac{3\Lambda}{4} \text{dist}(\mathbf{x}, \Sigma)^2,$$

and the conclusion follows by setting $\underline{C} \stackrel{\text{def.}}{=} \frac{\lambda}{4}$ and $\overline{C} \stackrel{\text{def.}}{=} \frac{3\Lambda}{4}$. $\square$

Theorem 3.4. *Under [\(H1\)–\(H4\)](#), as $\varepsilon \rightarrow 0$, the Gibbs measure $\mathbb{P}_\varepsilon$ converges weakly to a probability measure $\mathbb{P}_0$, given by*

$$\text{d}\mathbb{P}_0(z) = \frac{1}{Z_0} (\det H_z E)^{-1/2} \rho(z) \text{d}\mathcal{H}_{|\Sigma}^d(z), \quad Z_0 \stackrel{\text{def.}}{=} \int_{\Sigma} (\det H_z E)^{-1/2} \rho(z) \text{d}\mathcal{H}_{|\Sigma}^d(z).$$

Moreover, $Z_\varepsilon = (2\pi\varepsilon)^{k/2} (Z_0 + o(1))$.

Remark 3.1. We postpone the proof to [Section 5.1.1](#). It follows the general strategy of [\[Hwa80\]](#), with the various steps of the geometric Laplace method made explicit, in a slightly different setting: the proof is formulated so as to cover the case of compactly supported marginals in $\mathbb{R}^d$, and requires only $C^{2,1}$ -regularity of the duality gap E and $C^{1,1}$ -regularity of the contact set Σ .

Remark 3.2. A connection between the geometric Laplace method and optimal transport has already been established by Léger and Vialard in [\[LV23\]](#), under stronger regularity assumptions and assuming that the contact set is the graph of a C^3 -diffeomorphism. These assumptions allow them to derive a first-order geometric Laplace expansion. In contrast, our purpose is only to identify the leading-order limit, under weaker regularity assumptions and without requiring the contact set to be globally a graph.

4 The Γ -expansion of Entropic Optimal Transport

Throughout this section, we assume [\(H1\)–\(H4\)](#), and we consider the functionals defined on $\mathcal{P}(\Omega)$ by

$$\mathcal{G}_\varepsilon(\gamma) \stackrel{\text{def.}}{=} \text{Ent}(\gamma | \mathbb{P}_\varepsilon) + \chi_{\Pi(\mu_1, \mu_2)}(\gamma), \quad \mathcal{G}_0(\gamma) \stackrel{\text{def.}}{=} \text{Ent}(\gamma | \mathbb{P}_0) + \chi_{\Pi(\mu_1, \mu_2)}(\gamma). \quad (4.1)$$

By [Proposition 2.1](#) and [Theorem 3.4](#), the functionals $\mathcal{F}_\varepsilon$, $\mathcal{F}^{(0)}$ and $\mathcal{F}^{(1)}$ of [Theorem A](#) satisfy

$$\begin{aligned} \frac{\mathcal{F}_\varepsilon - \min \mathcal{F}^{(0)}}{\varepsilon} - \frac{k}{2} \log \frac{1}{\varepsilon} &= \mathcal{G}_\varepsilon - \log Z_0 - \frac{k}{2} \log(2\pi) + o(1), \\ \mathcal{F}^{(1)} &= \mathcal{G}_0 - \log Z_0 - \frac{k}{2} \log(2\pi), \end{aligned} \quad (4.2)$$

where the $o(1)$ does not depend on γ . Hence, the first-order Γ -expansion of [Theorem A](#) amounts to the Γ -convergence of $\mathcal{G}_\varepsilon$ to $\mathcal{G}_0$ as $\varepsilon \rightarrow 0$, for the weak convergence of probability measures, that is, to the following two inequalities:

- (Γ -lim inf inequality) for every $\bar{\gamma} \in \mathcal{P}(\Omega)$ and every family $(\gamma_\varepsilon)_{\varepsilon>0}$ in $\mathcal{P}(\Omega)$ such that $\gamma_\varepsilon \rightharpoonup \bar{\gamma}$ as $\varepsilon \rightarrow 0$,

$$\liminf_{\varepsilon \rightarrow 0} \mathcal{G}_\varepsilon(\gamma_\varepsilon) \geq \mathcal{G}_0(\bar{\gamma}); \quad (4.3)$$

- (Γ -lim sup inequality) for every $\bar{\gamma} \in \mathcal{P}(\Omega)$, there exists a family $(\gamma_\varepsilon)_{\varepsilon>0}$ in $\mathcal{P}(\Omega)$, called a recovery sequence for $\bar{\gamma}$, such that $\gamma_\varepsilon \rightharpoonup \bar{\gamma}$ as $\varepsilon \rightarrow 0$ and

$$\limsup_{\varepsilon \rightarrow 0} \mathcal{G}_\varepsilon(\gamma_\varepsilon) \leq \mathcal{G}_0(\bar{\gamma}). \quad (4.4)$$

The Γ -lim inf inequality (4.3) follows directly from the convergence of the Gibbs measures. The core of the section is the Γ -lim sup inequality (4.4), that is, the construction of a recovery sequence: we first perturb a given plan in the normal directions to Σ ([Section 4.1](#)), and then correct its marginals ([Section 4.2](#)). Both inequalities, and then [Theorem A](#), are proved in [Section 4.3](#). Finally, [Section 4.4](#) contains the proof of the key estimate on the marginals of the perturbed plans.

4.1 The recovery sequence: a normal perturbation approach

To prove the Γ -lim sup inequality (4.4), we fix $\bar{\gamma}$ such that $\mathcal{G}_0(\bar{\gamma}) < \infty$ (otherwise, $\gamma_\varepsilon \stackrel{\text{def.}}{=} \bar{\gamma}$ is a recovery sequence), that is, $\bar{\gamma} \in \Pi(\mu_1, \mu_2)$ with $\text{Ent}(\bar{\gamma} | \mathbb{P}_0) < \infty$, and we look for plans $\gamma_\varepsilon \in \Pi(\mu_1, \mu_2)$ such that $\gamma_\varepsilon \rightharpoonup \bar{\gamma}$ and $\limsup_{\varepsilon \rightarrow 0} \mathcal{G}_\varepsilon(\gamma_\varepsilon) \leq \mathcal{G}_0(\bar{\gamma})$. We proceed in two steps: we first spread $\bar{\gamma}$ in the normal directions according to the Gaussian profile of the Gibbs measure, which gives the right entropy but not the right marginals, and we then correct the marginals in [Section 4.2](#).

Since $\bar{\gamma} \ll \mathbb{P}_0 \ll \mathcal{H}_{|\Sigma}^d$, we write $\bar{\gamma} = \bar{\rho} \mathcal{H}_{|\Sigma}^d$. We fix a truncation radius $R \geq 1$ for the rescaled normal variable s , the distance to Σ being $\sqrt{\varepsilon}|s|$, and we take ε small enough so that $\sqrt{\varepsilon}R < r_0$, with r_0 as in [Section 3.1](#); R will be sent to infinity at the end of the proof. Recall that $H_z E = \text{Hess}^\perp E(z)$ is the normal Hessian of E at $z \in \Sigma$ ([Section 2.2](#)), seen as a quadratic form on $\mathbb{R}^k$ through the normal frame $n(z)$ ([Section 3.1](#)).

For $z \in \Sigma$, let

$$K_R(z, s) \stackrel{\text{def.}}{=} \frac{\mathbf{1}_{\mathbb{B}_R}(s) e^{-H_z E(s, s)/2}}{Z_R(z)}, \quad Z_R(z) \stackrel{\text{def.}}{=} \int_{\mathbb{B}_R} e^{-H_z E(s, s)/2} ds,$$

be the density of the centered Gaussian measure on $\mathbb{R}^k$ with covariance $H_z E^{-1}$, truncated to $\mathbb{B}_R$ and renormalized. By (2.3), $Z_R(z) \geq c \stackrel{\text{def.}}{=} \int_{\mathbb{B}_1} e^{-\Lambda|s|^2/2} ds > 0$ for $R \geq 1$, so that the moments of these kernels are bounded uniformly in $z \in \Sigma$ and $R \geq 1$: for every $p \geq 0$,

$$\int_{\mathbb{B}_R} |s|^p K_R(z, s) ds \leq M_p \stackrel{\text{def.}}{=} \frac{1}{c} \int_{\mathbb{R}^k} |s|^p e^{-\lambda|s|^2/2} ds < \infty. \quad (4.5)$$

Setting $\Phi_\varepsilon(z, s) \stackrel{\text{def.}}{=} \Phi(z, \sqrt{\varepsilon}s)$, we have $\bar{\gamma} = (\Phi_\varepsilon)_\#(\bar{\gamma} \otimes \delta_0)$. Replacing the Dirac mass in the normal variable by this truncated Gaussian, we define the *normal perturbation* of $\bar{\gamma}$ by

$$\tilde{\gamma}_{\varepsilon,R} \stackrel{\text{def.}}{=} (\Phi_\varepsilon)_\#(K_R(z, s) \, ds \, d\bar{\gamma}(z)), \quad (4.6)$$

that is, $\int_\Omega f \, d\tilde{\gamma}_{\varepsilon,R} = \int_\Sigma \int_{\mathbb{B}_R} f(\Phi(z, \sqrt{\varepsilon}s)) K_R(z, s) \, ds \, d\bar{\gamma}(z)$ for every nonnegative Borel function f on Ω . Since $K_R(z, \cdot)$ is a probability density, $\tilde{\gamma}_{\varepsilon,R}$ is a probability measure, concentrated on $\Sigma_{\sqrt{\varepsilon}R}$. In the original normal variable $\eta = \sqrt{\varepsilon}s$, its profile in each normal fiber is the Gaussian with covariance $\varepsilon H_z E^{-1}$, truncated at distance $\sqrt{\varepsilon}R$ from Σ , which mimics the Gibbs measure $\mathbb{P}_\varepsilon$ near Σ , since $E(\Phi(z, \eta)) \simeq \frac{1}{2} H_z E(\eta, \eta)$. By (3.2), $\tilde{\gamma}_{\varepsilon,R}$ is absolutely continuous with respect to vol_Ω , with a density $\tilde{p}_{\varepsilon,R}$ vanishing outside $\Sigma_{\sqrt{\varepsilon}R}$ and given in tubular coordinates by

$$\tilde{p}_{\varepsilon,R}(\Phi(z, \sqrt{\varepsilon}s)) = \frac{\bar{\rho}(z) K_R(z, s)}{\varepsilon^{k/2} \mathcal{J}_\Phi(z, \sqrt{\varepsilon}s)} \quad \text{for } \mathcal{H}^d\text{-a.e. } z \in \Sigma \text{ and every } s \in \mathbb{B}_R. \quad (4.7)$$

Since $\rho_i \geq \alpha > 0$, the marginals $\tilde{\mu}_{i,\varepsilon,R}$, $i = 1, 2$, of $\tilde{\gamma}_{\varepsilon,R}$ are absolutely continuous with respect to μ_i . They differ from μ_i in general, and are corrected in Section 4.2. Finally, the image of $K_R(z, s) \, ds \, d\bar{\gamma}(z)$ under the map $(z, s) \mapsto (z, \Phi_\varepsilon(z, s))$ is a coupling of $\bar{\gamma}$ and $\tilde{\gamma}_{\varepsilon,R}$, and the distance between z and $\Phi_\varepsilon(z, s) = z + \sqrt{\varepsilon} n(z)s$ is at most $\sqrt{\varepsilon}|s| \leq \sqrt{\varepsilon}R$, so that

$$W_2(\tilde{\gamma}_{\varepsilon,R}, \bar{\gamma}) \leq \sqrt{\varepsilon}R \longrightarrow 0 \quad (\varepsilon \rightarrow 0). \quad (4.8)$$

In particular, since the marginals of $\bar{\gamma}$ are μ_i and the projections $\Omega \rightarrow M_i$ are 1-Lipschitz, we also have $W_2(\tilde{\mu}_{i,\varepsilon,R}, \mu_i) \leq \sqrt{\varepsilon}R$ for $i = 1, 2$.

Lemma 4.1 (Entropy of the normal perturbation). *For every fixed $R \geq 1$, as $\varepsilon \rightarrow 0$,*

$$\text{Ent}(\tilde{\gamma}_{\varepsilon,R} \mid \mathbb{P}_\varepsilon) \longrightarrow \text{Ent}(\bar{\gamma} \mid \mathbb{P}_0) + b_R(\bar{\gamma}), \quad b_R(\bar{\gamma}) = \int_\Sigma \log \frac{Z_\infty(z)}{Z_R(z)} \, d\bar{\gamma}(z), \quad (4.9)$$

where $Z_\infty(z) \stackrel{\text{def.}}{=} \int_{\mathbb{R}^k} e^{-H_z E(s,s)/2} \, ds = (2\pi)^{k/2} (\det H_z E)^{-1/2}$. Moreover, $b_R(\bar{\gamma}) \geq 0$ and $\sup_{\bar{\gamma} \in \mathcal{P}(\Sigma)} b_R(\bar{\gamma}) \rightarrow 0$ as $R \rightarrow \infty$.

Proof. By (4.7), and since $\mathbb{P}_\varepsilon$ has density $Z_\varepsilon^{-1} e^{-E/\varepsilon} \rho$ with respect to vol_Ω , where ρ is the density of $\mu_1 \otimes \mu_2$, we have, for $\mathcal{H}^d$ -a.e. $z \in \Sigma$ and every $s \in \mathbb{B}_R$,

$$\begin{aligned} \log \left(\frac{d\tilde{\gamma}_{\varepsilon,R}}{d\mathbb{P}_\varepsilon}(\Phi(z, \sqrt{\varepsilon}s)) \right) &= \log \bar{\rho}(z) + \log \left(\frac{Z_\varepsilon}{\varepsilon^{k/2} Z_R(z)} \right) - \log \rho(\Phi(z, \sqrt{\varepsilon}s)) \\ &\quad + \frac{E(\Phi(z, \sqrt{\varepsilon}s)) - \frac{1}{2} H_z E(\sqrt{\varepsilon}s, \sqrt{\varepsilon}s)}{\varepsilon} - \log \mathcal{J}_\Phi(z, \sqrt{\varepsilon}s). \end{aligned}$$

As in the proof of Proposition 3.3, a Taylor expansion of E gives

$$E(\Phi(z, \eta)) = \frac{1}{2} H_z E(\eta, \eta) + \mathcal{R}(z, \eta), \quad |\mathcal{R}(z, \eta)| \leq C|\eta|^3.$$

Since $\text{Ent}(\tilde{\gamma}_{\varepsilon,R} | \mathbb{P}_\varepsilon) = \int_\Omega \log \frac{d\tilde{\gamma}_{\varepsilon,R}}{d\mathbb{P}_\varepsilon} d\tilde{\gamma}_{\varepsilon,R}$, we may compute this integral in tubular coordinates by (4.6), that is, integrate the above identity against $K_R(z, s) ds d\bar{\gamma}(z)$. We obtain

$$\begin{aligned} \text{Ent}(\tilde{\gamma}_{\varepsilon,R} | \mathbb{P}_\varepsilon) &= \int_\Sigma \log \bar{\rho}(z) \bar{\gamma}(dz) + \int_\Sigma \log \left(\frac{Z_\varepsilon}{\varepsilon^{k/2} Z_R(z)} \right) \bar{\gamma}(dz) \\ &\quad - \int_\Sigma \int_{\mathbb{B}_R} \left(\log \rho(\Phi(z, \sqrt{\varepsilon}s)) - \frac{\mathcal{R}(z, \sqrt{\varepsilon}s)}{\varepsilon} \right. \\ &\quad \left. + \log \mathcal{J}_\Phi(z, \sqrt{\varepsilon}s) \right) K_R(z, s) ds \bar{\gamma}(dz). \end{aligned}$$

These integrals are finite: since $\text{Ent}(\bar{\gamma} | \mathbb{P}_0) < \infty$ and the density of $\mathbb{P}_0$ with respect to $\mathcal{H}_\Sigma^d$ is bounded from above and below by positive constants, we have $\int_\Sigma |\log \bar{\rho}| d\bar{\gamma} < \infty$. Since $|\mathcal{R}(z, \sqrt{\varepsilon}s)| \leq C\varepsilon^{3/2}|s|^3$, since $\log \rho$ is Lipschitz (ρ being Lipschitz and bounded below by a positive constant), and since $|\log \mathcal{J}_\Phi(z, \sqrt{\varepsilon}s)| \leq C\sqrt{\varepsilon}|s|$ by (3.4), the double integral in the last line differs from $\int_\Sigma \log \rho d\bar{\gamma}$ by at most

$$C\sqrt{\varepsilon} \int_\Sigma \int_{\mathbb{B}_R} (|s| + |s|^3) K_R(z, s) ds \bar{\gamma}(dz) \leq C\sqrt{\varepsilon} \longrightarrow 0,$$

by (4.5). Moreover, by Theorem 3.4, $Z_\varepsilon = (2\pi\varepsilon)^{k/2}(Z_0 + o(1))$, so that, uniformly in z ,

$$\log \left(\frac{Z_\varepsilon}{\varepsilon^{k/2} Z_R(z)} \right) = \log \frac{(2\pi)^{k/2} Z_0}{Z_R(z)} + o(1) = \log \left(Z_0 \sqrt{\det H_z E} \right) + \log \frac{Z_\infty(z)}{Z_R(z)} + o(1).$$

Combining these identities yields

$$\begin{aligned} \text{Ent}(\tilde{\gamma}_{\varepsilon,R} | \mathbb{P}_\varepsilon) &\longrightarrow \int_\Sigma \log \frac{\bar{\rho}(z)}{\rho(z)} \bar{\gamma}(dz) + \int_\Sigma \log \left(Z_0 \sqrt{\det H_z E} \right) \bar{\gamma}(dz) \\ &\quad + \int_\Sigma \log \frac{Z_\infty(z)}{Z_R(z)} \bar{\gamma}(dz) \\ &= \text{Ent}(\bar{\gamma} | \mathbb{P}_0) + b_R(\bar{\gamma}). \end{aligned}$$

Finally, the Gaussian tail estimate

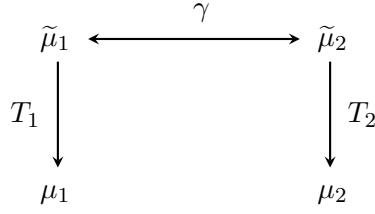
$$0 \leq Z_\infty(z) - Z_R(z) \leq \int_{|s| \geq R} \exp \left(-\frac{\lambda}{2} |s|^2 \right) ds \longrightarrow 0 \quad (R \rightarrow \infty)$$

is uniform in z . Together with $Z_R(z) \geq c > 0$, this shows that $b_R \geq 0$ tends to zero uniformly over probability measures on Σ , and proves the claim. $\square$

4.2 Correction of the marginals

The normal perturbation $\tilde{\gamma}_{\varepsilon,R}$ has the expected entropy (Lemma 4.1), but not the prescribed marginals. To correct them, we use the shadow construction of [EN22]: each perturbed marginal is transported to the prescribed one by an optimal transport map,

and the product of these maps is applied to the plan, as illustrated in the diagram below. In [EN22], the construction is carried out more generally with optimal transport plans; here it is simpler, since the perturbed marginals are absolutely continuous, so that optimal transport maps exist.



Definition 4.1 (Shadow [EN22, Definition 3.1]). Let $\mu_i, \tilde{\mu}_i \in \mathcal{P}(M_i)$ with $\tilde{\mu}_i \ll \text{vol}_{M_i}$, and let T_i denote the optimal transport map from $\tilde{\mu}_i$ to μ_i for the quadratic cost, $i = 1, 2$. Then, for a given $\tilde{\gamma} \in \Pi(\tilde{\mu}_1, \tilde{\mu}_2)$, its *shadow* $\gamma \in \Pi(\mu_1, \mu_2)$ is defined by

$$\gamma \stackrel{\text{def.}}{=} (T_1, T_2)_\# \tilde{\gamma}.$$

The interest of this construction is that it turns a plan with good properties but mismatched marginals into a plan with the prescribed marginals, which remains close to the original one. More precisely, the distance between the shadow and the original plan depends only on the distance between the marginals, and its relative entropy is controlled by that of the original plan, as stated in the following lemma.

Lemma 4.2 (Properties of the shadow [EN22, Lemma 3.2]). *Given $\tilde{\gamma} \in \Pi(\tilde{\mu}_1, \tilde{\mu}_2)$ and its shadow $\gamma \in \Pi(\mu_1, \mu_2)$, one has*

$$W_2(\gamma, \tilde{\gamma})^2 = \sum_{i=1}^2 W_2(\mu_i, \tilde{\mu}_i)^2.$$

Furthermore, the relative entropy of the shadow is bounded by that of the original plan:

$$\text{Ent}(\gamma \mid \mu_1 \otimes \mu_2) \leq \text{Ent}(\tilde{\gamma} \mid \tilde{\mu}_1 \otimes \tilde{\mu}_2).$$

We denote by $\gamma_{\varepsilon, R}$ the shadow of $\tilde{\gamma}_{\varepsilon, R}$. By Lemma 4.2 and (4.8), $W_2(\gamma_{\varepsilon, R}, \tilde{\gamma}_{\varepsilon, R}) \leq \sqrt{2\varepsilon} R$, so that $\gamma_{\varepsilon, R}$ still converges to $\tilde{\gamma}$. However, to remain a recovery sequence, it must also have the same limit energy. It turns out that passing to the shadow adds at most an energy term.

Proposition 4.3 (Entropy of the shadow). *For every $R \geq 1$ and $\varepsilon > 0$ with $\sqrt{\varepsilon} R < r_0$,*

$$\text{Ent}(\gamma_{\varepsilon, R} \mid \mathbb{P}_\varepsilon) \leq \text{Ent}(\tilde{\gamma}_{\varepsilon, R} \mid \mathbb{P}_\varepsilon) + \frac{1}{\varepsilon} \int_{\Omega} E \, d(\gamma_{\varepsilon, R} - \tilde{\gamma}_{\varepsilon, R}). \quad (4.10)$$

Proof. In this proof, we write $\gamma = \gamma_{\varepsilon, R}$, $\tilde{\gamma} = \tilde{\gamma}_{\varepsilon, R}$ and $\tilde{\mu}_i = \tilde{\mu}_{i, \varepsilon, R}$. Recall that $\mathbb{P}_\varepsilon = f_\varepsilon \mu^\otimes$, where $f_\varepsilon \stackrel{\text{def.}}{=} Z_\varepsilon^{-1} e^{-E/\varepsilon}$ (Proposition 2.1). By Lemma 4.1, $\text{Ent}(\tilde{\gamma} \mid \mathbb{P}_\varepsilon) < \infty$, hence

$\text{Ent}(\tilde{\gamma} | \mu^\otimes) \leq \text{Ent}(\tilde{\gamma} | \mathbb{P}_\varepsilon) - \log Z_\varepsilon < \infty$ since $E \geq 0$. Hence the marginal entropies $\text{Ent}(\tilde{\mu}_i | \mu_i)$ are finite by projection, and so are all the quantities below. By a change of reference measure, $\text{Ent}(\cdot | \mathbb{P}_\varepsilon) = \text{Ent}(\cdot | \mu^\otimes) - \int_\Omega \log f_\varepsilon \, d(\cdot)$ on probability measures. Together with [Lemma 4.2](#), this gives

$$\begin{aligned} \text{Ent}(\gamma | \mathbb{P}_\varepsilon) &= \text{Ent}(\gamma | \mu^\otimes) - \int_\Omega \log f_\varepsilon \, d\gamma \\ &\leq \text{Ent}(\tilde{\gamma} | \tilde{\mu}_1 \otimes \tilde{\mu}_2) - \int_\Omega \log f_\varepsilon \, d\gamma \\ &= \text{Ent}(\tilde{\gamma} | \mathbb{P}_\varepsilon) + \int_\Omega \log f_\varepsilon \, d(\tilde{\gamma} - \gamma) \\ &\quad + \text{Ent}(\tilde{\gamma} | \tilde{\mu}_1 \otimes \tilde{\mu}_2) - \text{Ent}(\tilde{\gamma} | \mu^\otimes). \end{aligned}$$

Since $\tilde{\mu}_i \ll \mu_i$ ([Section 4.1](#)), there exist Borel functions $a_i : M_i \rightarrow [0, \infty)$ such that $\tilde{\mu}_i = a_i \mu_i$, so that $\tilde{\mu}_1 \otimes \tilde{\mu}_2 = a_1(x_1)a_2(x_2)\mu^\otimes$. Consequently,

$$\begin{aligned} \text{Ent}(\tilde{\gamma} | \tilde{\mu}_1 \otimes \tilde{\mu}_2) - \text{Ent}(\tilde{\gamma} | \mu^\otimes) &= - \int_\Omega \log(a_1(x_1)a_2(x_2)) \, d\tilde{\gamma}(x_1, x_2) \\ &= - \sum_{i=1}^2 \int_{M_i} \log a_i \, d\tilde{\mu}_i = - \sum_{i=1}^2 \text{Ent}(\tilde{\mu}_i | \mu_i) \leq 0, \end{aligned}$$

where the last inequality follows from the nonnegativity of the relative entropy between probability measures. Finally, since $\log f_\varepsilon = -E/\varepsilon - \log Z_\varepsilon$ and both plans have total mass one,

$$\text{Ent}(\gamma | \mathbb{P}_\varepsilon) \leq \text{Ent}(\tilde{\gamma} | \mathbb{P}_\varepsilon) - \int_\Omega \log f_\varepsilon \, d(\gamma - \tilde{\gamma}) = \text{Ent}(\tilde{\gamma} | \mathbb{P}_\varepsilon) + \frac{1}{\varepsilon} \int_\Omega E \, d(\gamma - \tilde{\gamma}).$$

□

The additional term in [\(4.10\)](#) tends to 0 as soon as $W_2(\tilde{\mu}_{i,\varepsilon,R}, \mu_i) = o(\sqrt{\varepsilon})$ ([Proposition 4.5](#) below). The following estimate is therefore the cornerstone of the construction of the recovery sequence.

Theorem 4.4 (Marginal estimate). *For each fixed $R \geq 1$, the marginals of the normal perturbation [\(4.6\)](#) satisfy*

$$W_2(\tilde{\mu}_{i,\varepsilon,R}, \mu_i) = o_R(\sqrt{\varepsilon}) \quad (i = 1, 2), \quad \text{as } \varepsilon \rightarrow 0, \quad (4.11)$$

where the subscript indicates that the rate of convergence may depend on R .

The proof is given in [Section 4.4](#).

We now verify that the additional term in [\(4.10\)](#) vanishes in the limit.

Proposition 4.5 (Energy of the correction). *For every fixed $R \geq 1$,*

$$\frac{1}{\varepsilon} \left| \int_\Omega E \, d(\gamma_{\varepsilon,R} - \tilde{\gamma}_{\varepsilon,R}) \right| \longrightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Proof. Let $T_{\varepsilon,R} \stackrel{\text{def.}}{=} (T_1, T_2)$, where T_i is the optimal transport map from $\tilde{\mu}_{i,\varepsilon,R}$ to μ_i . By definition of the shadow, $\gamma_{\varepsilon,R} = (T_{\varepsilon,R})_{\#} \tilde{\gamma}_{\varepsilon,R}$. Set

$$\Delta_{\varepsilon,R}^2 \stackrel{\text{def.}}{=} \int_{\Omega} \text{dist}(\mathbf{x}, T_{\varepsilon,R}(\mathbf{x}))^2 \, d\tilde{\gamma}_{\varepsilon,R}(\mathbf{x}) = \sum_{i=1}^2 W_2^2(\tilde{\mu}_{i,\varepsilon,R}, \mu_i).$$

The pushforward identity gives

$$\frac{1}{\varepsilon} \int_{\Omega} E \, d(\gamma_{\varepsilon,R} - \tilde{\gamma}_{\varepsilon,R}) = \frac{1}{\varepsilon} \int_{\Omega} (E(T_{\varepsilon,R}(\mathbf{x})) - E(\mathbf{x})) \, d\tilde{\gamma}_{\varepsilon,R}(\mathbf{x}).$$

We identify Ω with $\mathbb{T}^{2d} = \mathbb{R}^{2d}/\mathbb{Z}^{2d}$, denote by $[y] \in \Omega$ the class of $y \in \mathbb{R}^{2d}$, and by $\tilde{E} \stackrel{\text{def.}}{=} E([\cdot])$ the $\mathbb{Z}^{2d}$ -periodic lift of E , which belongs to $C^{2,1}(\mathbb{R}^{2d})$ by (H2). We set

$$C_E \stackrel{\text{def.}}{=} \sup_{y \in \mathbb{R}^{2d}} \|D^2 \tilde{E}(y)\| < \infty,$$

where $\|\cdot\|$ denotes the operator norm.

We first estimate the gradient by the distance to Σ . Let $y \in \mathbb{R}^{2d}$. Since Σ is compact, there exists $\bar{y} \in \mathbb{R}^{2d}$ such that $[\bar{y}] \in \Sigma$ and $|y - \bar{y}| = \text{dist}([y], \Sigma)$. Since $\nabla \tilde{E}(\bar{y}) = 0$, the fundamental theorem of calculus gives

$$|\nabla \tilde{E}(y)| = \left| \int_0^1 D^2 \tilde{E}(\bar{y} + t(y - \bar{y}))(y - \bar{y}) \, dt \right| \leq C_E \text{dist}([y], \Sigma). \quad (4.12)$$

Now, for each $\mathbf{x} \in \Omega$, choose $\tilde{\mathbf{x}} \in \mathbb{R}^{2d}$ with $[\tilde{\mathbf{x}}] = \mathbf{x}$, and a shortest displacement $v_{\mathbf{x}} \in \mathbb{R}^{2d}$ such that

$$[\tilde{\mathbf{x}} + v_{\mathbf{x}}] = T_{\varepsilon,R}(\mathbf{x}), \quad |v_{\mathbf{x}}| = \text{dist}(\mathbf{x}, T_{\varepsilon,R}(\mathbf{x})).$$

By the fundamental theorem of calculus,

$$E(T_{\varepsilon,R}(\mathbf{x})) - E(\mathbf{x}) = \tilde{E}(\tilde{\mathbf{x}} + v_{\mathbf{x}}) - \tilde{E}(\tilde{\mathbf{x}}) = \int_0^1 \langle \nabla \tilde{E}(\tilde{\mathbf{x}} + \tau v_{\mathbf{x}}), v_{\mathbf{x}} \rangle \, d\tau.$$

For every $\tau \in [0, 1]$, the triangle inequality gives $\text{dist}([\tilde{\mathbf{x}} + \tau v_{\mathbf{x}}], \Sigma) \leq \text{dist}(\mathbf{x}, \Sigma) + \tau |v_{\mathbf{x}}|$. Together with (4.12), we obtain

$$|E(T_{\varepsilon,R}(\mathbf{x})) - E(\mathbf{x})| \leq C_E \text{dist}(\mathbf{x}, T_{\varepsilon,R}(\mathbf{x})) \text{dist}(\mathbf{x}, \Sigma) + C_E \text{dist}(\mathbf{x}, T_{\varepsilon,R}(\mathbf{x}))^2. \quad (4.13)$$

Integrating (4.13) with respect to $\tilde{\gamma}_{\varepsilon,R}$ and applying the Cauchy–Schwarz inequality gives

$$\begin{aligned} \frac{1}{\varepsilon} \left| \int_{\Omega} E \, d(\gamma_{\varepsilon,R} - \tilde{\gamma}_{\varepsilon,R}) \right| &\leq \frac{C_E \Delta_{\varepsilon,R}}{\varepsilon} \left(\int_{\Omega} \text{dist}(\mathbf{x}, \Sigma)^2 \, d\tilde{\gamma}_{\varepsilon,R}(\mathbf{x}) \right)^{1/2} \\ &\quad + \frac{C_E \Delta_{\varepsilon,R}^2}{\varepsilon}. \end{aligned}$$

Recalling the definition (4.6) of $\tilde{\gamma}_{\varepsilon,R}$ and using (3.1), we have

$$\int_{\Omega} \text{dist}(\mathbf{x}, \Sigma)^2 d\tilde{\gamma}_{\varepsilon,R}(\mathbf{x}) = \varepsilon \int_{\Sigma} \int_{\mathbb{B}_R} |s|^2 K_R(z, s) ds \bar{\gamma}(dz) \leq M_2 \varepsilon,$$

by (4.5). Hence

$$\frac{1}{\varepsilon} \left| \int_{\Omega} E d(\gamma_{\varepsilon,R} - \tilde{\gamma}_{\varepsilon,R}) \right| \leq C \left(\frac{\Delta_{\varepsilon,R}}{\sqrt{\varepsilon}} + \frac{\Delta_{\varepsilon,R}^2}{\varepsilon} \right). \quad (4.14)$$

For fixed R , Theorem 4.4 gives $\Delta_{\varepsilon,R} = o_R(\sqrt{\varepsilon})$, so that both terms on the right-hand side tend to zero. $\square$

4.3 Proof of Theorem A

Proof of Theorem A. Step 1. The Γ -lim inf inequality (4.3). Let $\gamma_{\varepsilon} \rightharpoonup \bar{\gamma}$. If $\bar{\gamma} \notin \Pi(\mu_1, \mu_2)$, then, since $\Pi(\mu_1, \mu_2)$ is closed, $\gamma_{\varepsilon} \notin \Pi(\mu_1, \mu_2)$ for ε small enough, so that $\mathcal{G}_{\varepsilon}(\gamma_{\varepsilon}) = +\infty$. Otherwise, the joint lower semicontinuity of the relative entropy, together with the convergence $\mathbb{P}_{\varepsilon} \rightharpoonup \mathbb{P}_0$ given by Theorem 3.4, yields

$$\liminf_{\varepsilon \rightarrow 0} \mathcal{G}_{\varepsilon}(\gamma_{\varepsilon}) \geq \liminf_{\varepsilon \rightarrow 0} \text{Ent}(\gamma_{\varepsilon} | \mathbb{P}_{\varepsilon}) \geq \text{Ent}(\bar{\gamma} | \mathbb{P}_0) = \mathcal{G}_0(\bar{\gamma}).$$

Step 2. The Γ -lim sup inequality (4.4). Let $\bar{\gamma} \in \mathcal{P}(\Omega)$ with $\mathcal{G}_0(\bar{\gamma}) < \infty$ (otherwise there is nothing to prove). For $R \geq 1$ and $\varepsilon > 0$ with $\sqrt{\varepsilon}R < r_0$, let $\tilde{\gamma}_{\varepsilon,R}$ be the normal perturbation of $\bar{\gamma}$ defined in (4.6), and $\gamma_{\varepsilon,R} \in \Pi(\mu_1, \mu_2)$ its shadow. We first fix R . By Lemma 4.1, Proposition 4.3 and Proposition 4.5,

$$\limsup_{\varepsilon \rightarrow 0} \mathcal{G}_{\varepsilon}(\gamma_{\varepsilon,R}) = \limsup_{\varepsilon \rightarrow 0} \text{Ent}(\gamma_{\varepsilon,R} | \mathbb{P}_{\varepsilon}) \leq \mathcal{G}_0(\bar{\gamma}) + b_R(\bar{\gamma}),$$

while Lemma 4.2 and (4.8) give

$$W_2(\gamma_{\varepsilon,R}, \bar{\gamma}) \leq W_2(\gamma_{\varepsilon,R}, \tilde{\gamma}_{\varepsilon,R}) + W_2(\tilde{\gamma}_{\varepsilon,R}, \bar{\gamma}) \leq (\sqrt{2} + 1)\sqrt{\varepsilon}R \rightarrow 0 \quad (\varepsilon \rightarrow 0),$$

so that $\gamma_{\varepsilon,R} \rightharpoonup \bar{\gamma}$, since W_2 metrizes the weak convergence on the compact set Ω . Hence $(\gamma_{\varepsilon,R})_{\varepsilon}$ is a recovery sequence for $\bar{\gamma}$ up to the error $b_R(\bar{\gamma})$, which tends to 0 as $R \rightarrow \infty$ by Lemma 4.1. As is standard in Γ -convergence, this suffices to obtain a recovery sequence by a diagonal argument (see, e.g., [Dal93]), which we recall for completeness. Choose $R_j \uparrow \infty$ and, inductively, $\varepsilon_j \downarrow 0$ such that, for every $0 < \varepsilon \leq \varepsilon_j$, we have $\sqrt{\varepsilon}R_j < r_0$, $W_2(\gamma_{\varepsilon,R_j}, \bar{\gamma}) \leq 1/j$, and $\mathcal{G}_{\varepsilon}(\gamma_{\varepsilon,R_j}) \leq \mathcal{G}_0(\bar{\gamma}) + b_{R_j}(\bar{\gamma}) + 1/j$. Then $\gamma_{\varepsilon} \stackrel{\text{def.}}{=} \gamma_{\varepsilon,R_j}$ for $\varepsilon_{j+1} < \varepsilon \leq \varepsilon_j$ defines a recovery sequence for $\bar{\gamma}$.

Step 3. Conclusion. By (4.2), (1.1) follows from the Γ -convergence of $\mathcal{G}_{\varepsilon}$ to $\mathcal{G}_0$ established in Steps 1 and 2, since adding to $\mathcal{G}_{\varepsilon}$ constants converging as $\varepsilon \rightarrow 0$ does not affect Γ -convergence. The infimum of $\mathcal{F}^{(1)}$ is finite. Indeed, since relative entropies are nonnegative, $\mathcal{F}^{(1)} \geq -\log Z_0 - \frac{k}{2} \log(2\pi)$, and it suffices to show that every optimal transport plan $\bar{\gamma}$ satisfies $\text{Ent}(\bar{\gamma} | \mathbb{P}_0) < \infty$. By (2.1), such a plan satisfies $\bar{\gamma}(\Omega \setminus \Sigma) = 0$. For every Borel set $A \subseteq \Sigma$, since proj_{M_1} is 1-Lipschitz, we have

$$\bar{\gamma}(A) \leq \mu_1(\text{proj}_{M_1}(A)) \leq \beta \mathcal{H}^d(\text{proj}_{M_1}(A)) \leq \beta \mathcal{H}^d(A),$$

so that $\bar{\gamma} \leq \beta \mathcal{H}_{|\Sigma}^d$. Since the density of $\mathbb{P}_0$ with respect to $\mathcal{H}_{|\Sigma}^d$ is bounded from below by a positive constant, by (H4) and the continuity of $\text{Hess}^\perp E$ on the compact set Σ , the density of $\bar{\gamma}$ with respect to $\mathbb{P}_0$ is bounded, hence $\text{Ent}(\bar{\gamma} | \mathbb{P}_0) < \infty$. Compactness and convergence of minima under Γ -convergence then give (1.2).

For (iii), note that, by Proposition 2.1, $\mathcal{F}_\varepsilon$ and $\mathcal{G}_\varepsilon$ have the same minimizers, and so do $\mathcal{F}^{(1)}$ and $\mathcal{G}_0$. Since $\mathcal{P}(\Omega)$ is weakly compact and the functionals $\mathcal{G}_\varepsilon, \mathcal{G}_0$ are lower semicontinuous, strictly convex on their domains and not identically $+\infty$, they admit unique minimizers γ_ε and γ_0 . By the fundamental theorem of Γ -convergence (see, e.g., [Dal93]), every cluster point of $(\gamma_\varepsilon)_{\varepsilon>0}$ as $\varepsilon \rightarrow 0$ minimizes $\mathcal{G}_0$, hence $\gamma_\varepsilon \rightarrow \gamma_0$. Finally, since $\mathcal{G}_0(\gamma_0) < \infty$, we have $\gamma_0 \ll \mathbb{P}_0$, so that $\text{spt}(\gamma_0) \subseteq \Sigma$ and γ_0 is an optimal transport plan by (2.1), hence the unique minimizer of $\text{Ent}(\cdot | \mathbb{P}_0)$ among all optimal transport plans. $\square$

4.4 Proof of Theorem 4.4

Throughout this section, $R \geq 1$ is fixed, and we write $K = K_R$, $\tilde{\gamma}_\varepsilon = \tilde{\gamma}_{\varepsilon,R}$ and $\tilde{\mu}_{i,\varepsilon} = \tilde{\mu}_{i,\varepsilon,R}$. Recall that, by (4.6), the kernel K does not depend on ε .

Let us first explain where the difficulty lies. If Σ is flat and $\text{Hess}^\perp E$ is constant along Σ , the normal directions and the kernel $K(z, \cdot)$ do not depend on z , so that $\tilde{\gamma}_\varepsilon$ is the convolution of $\bar{\gamma}$ with a fixed truncated Gaussian of variance of order ε , supported on the normal space. Each marginal $\tilde{\mu}_{i,\varepsilon}$ is then the convolution of μ_i with the projection η_ε of this Gaussian, which is a centered kernel of variance of order ε . For a general probability measure μ , $W_2(\mu, \mu * \eta_\varepsilon) = O(\sqrt{\varepsilon})$, which is not enough for our purpose, since we need $o(\sqrt{\varepsilon})$. This improves to $O(\varepsilon^{(1+\theta)/2})$ if μ has a $C^{1,\theta}$ density bounded from below, with $\theta \in (0, 1]$, the first-order contributions cancelling out by symmetry. In general, however, Σ is not flat, and several branches of Σ , that is, pieces of Σ on which the projection onto M_i is a diffeomorphism, may lie above the same point of M_i . The first issue is mild, since Σ is locally almost flat. The second one is more serious: $\tilde{\mu}_{i,\varepsilon}$ is the sum of the contributions of the branches, each obtained by regularizing the partial marginal of $\bar{\gamma}$ carried by the branch, and, although these partial marginals sum up to μ_i , their densities need not inherit the regularity of μ_i : they are merely L^1 . We therefore need to handle L^1 densities regularized along a non-flat manifold. To explain the strategy, we first consider the flat case with one-dimensional marginals, and show in Section 4.4.1 that $W_2(\mu, \mu * \eta_\varepsilon) = o(\sqrt{\varepsilon})$ as soon as μ has an L^1 density. The argument relies again on the symmetry of the Gaussian kernel.

4.4.1 A model case: one-dimensional marginals and flat contact set

To illustrate the proof of Theorem 4.4, we first consider one-dimensional marginals and a flat contact set. As explained above, the marginals of the normal perturbation are then convolutions of the original marginals with a fixed truncated Gaussian kernel, and we estimate the Wasserstein distance between a measure μ on $\mathbb{R}$ and its convolution with such a kernel, assuming only that μ is absolutely continuous, that is, has an L^1 density.

A similar approach was introduced in one dimension in [CDS25]. Here we adapt and clarify it before extending it to higher dimensions.

Proposition 4.6 (One-dimensional flat case). *Let $\mu \in \mathcal{P}(\mathbb{R})$ be absolutely continuous with respect to the Lebesgue measure, with finite second moment. For $R > 0$ and $\varepsilon > 0$, let*

$$\eta_\varepsilon(x) \stackrel{\text{def.}}{=} \frac{1}{\sqrt{\varepsilon} c_R} e^{-x^2/(2\varepsilon)} \mathbf{1}_{\{|x| \leq \sqrt{\varepsilon} R\}}, \quad c_R \stackrel{\text{def.}}{=} \int_{-R}^R e^{-z^2/2} dz.$$

Then $\tilde{\mu}_\varepsilon \stackrel{\text{def.}}{=} \mu * \eta_\varepsilon$ satisfies

$$W_2(\mu, \tilde{\mu}_\varepsilon) = \sqrt{R} o(\sqrt{\varepsilon}) \quad (\varepsilon \rightarrow 0),$$

where the $o(\sqrt{\varepsilon})$ depends only on μ , and not on R .

Proof. Throughout the proof, we still denote by μ the density of μ . For $t \in [0, 1]$, let

$$\mu_{t,\varepsilon} \stackrel{\text{def.}}{=} \mu * (x \mapsto t\sqrt{\varepsilon}x)_{\#} \left[\frac{1}{c_R} e^{-x^2/2} \mathbf{1}_{\{|x| \leq R\}} dx \right],$$

where convolution and push-forward are understood as operations on measures, so that $\mu_{0,\varepsilon} = \mu$, $\mu_{1,\varepsilon} = \tilde{\mu}_\varepsilon$, and, for every $t \in [0, 1]$,

$$\mu_{t,\varepsilon}(x) = \frac{1}{c_R} \int_{-R}^R \mu(x - t\sqrt{\varepsilon}z) e^{-z^2/2} dz.$$

Step 1. Computing the flux. For $\phi \in C_c^\infty(\mathbb{R})$ and $t \in [0, 1]$,

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}} \phi d\mu_{t,\varepsilon} &= \frac{d}{dt} \frac{1}{c_R} \int_{-R}^R \int_{\mathbb{R}} \phi(y + t\sqrt{\varepsilon}z) \mu(y) e^{-z^2/2} dy dz \\ &= \frac{1}{c_R} \int_{-R}^R \int_{\mathbb{R}} \phi'(y + t\sqrt{\varepsilon}z) \sqrt{\varepsilon}z \mu(y) e^{-z^2/2} dy dz = \int_{\mathbb{R}} \phi'(x) \mathbf{m}_{t,\varepsilon}(x) dx, \end{aligned}$$

where, by the change of variables $y = x - t\sqrt{\varepsilon}z$,

$$\mathbf{m}_{t,\varepsilon}(x) \stackrel{\text{def.}}{=} \frac{1}{c_R} \int_{-R}^R \sqrt{\varepsilon}z \mu(x - t\sqrt{\varepsilon}z) e^{-z^2/2} dz.$$

Differentiation under the integral sign is justified since ϕ' is bounded, $\mu \in L^1$ and $\int_{-R}^R |z| e^{-z^2/2} dz < \infty$. Thus $\partial_t \mu_{t,\varepsilon} + \partial_x \mathbf{m}_{t,\varepsilon} = 0$ in the sense of distributions. Note that $\mathbf{m}_{0,\varepsilon} = 0$ by symmetry of the Gaussian kernel.

Let us estimate the total variation $\|\mathbf{m}_{t,\varepsilon}\|_{\text{TV}} = \int_{\mathbb{R}} |\mathbf{m}_{t,\varepsilon}| dx$. The flux depends linearly on μ : for $f \in L^1(\mathbb{R})$, let

$$\mathbf{m}_{t,\varepsilon}[f](x) \stackrel{\text{def.}}{=} \frac{1}{c_R} \int_{-R}^R \sqrt{\varepsilon}z f(x - t\sqrt{\varepsilon}z) e^{-z^2/2} dz, \quad \text{so that } \mathbf{m}_{t,\varepsilon} = \mathbf{m}_{t,\varepsilon}[\mu].$$

We will use the following bounds on the moments of the truncated Gaussian, which are uniform in R and follow from an integration by parts and the Cauchy–Schwarz inequality:

$$\frac{1}{c_R} \int_{-R}^R z^2 e^{-z^2/2} dz = 1 - \frac{2R e^{-R^2/2}}{c_R} \leq 1, \quad (4.15)$$

$$\frac{1}{c_R} \int_{-R}^R |z| e^{-z^2/2} dz \leq \left(\frac{1}{c_R} \int_{-R}^R z^2 e^{-z^2/2} dz \right)^{1/2} \leq 1. \quad (4.16)$$

By Fubini's theorem, the translation invariance of the Lebesgue measure and (4.16), we get

$$\|\mathbf{m}_{t,\varepsilon}[f]\|_{L^1} \leq \sqrt{\varepsilon} \|f\|_{L^1} \quad \text{for every } f \in L^1(\mathbb{R}).$$

If moreover $f \in C_c^1(\mathbb{R})$, we write, with $x_s \stackrel{\text{def.}}{=} x - s\sqrt{\varepsilon}z$,

$$f(x - t\sqrt{\varepsilon}z) = f(x) - \int_0^t \sqrt{\varepsilon} z f'(x_s) ds.$$

Since the kernel is centered, that is, $\int_{-R}^R z e^{-z^2/2} dz = 0$, the term containing $f(x)$ cancels, so that, by (4.15),

$$\|\mathbf{m}_{t,\varepsilon}[f]\|_{L^1} = \int_{\mathbb{R}} \left| \frac{\varepsilon}{c_R} \int_{-R}^R z^2 e^{-z^2/2} \int_0^t f'(x_s) ds dz \right| dx \leq t\varepsilon \|f'\|_{L^1} \leq \varepsilon \|f'\|_{L^1}.$$

Now let $(\mu^k)_{k \in \mathbb{N}}$ be a sequence in $C_c^1(\mathbb{R})$ such that $\mu^k \rightarrow \mu$ in L^1 . Writing $\mathbf{m}_{t,\varepsilon} = \mathbf{m}_{t,\varepsilon}[\mu - \mu^k] + \mathbf{m}_{t,\varepsilon}[\mu^k]$, we obtain, for every $k \in \mathbb{N}$ and $t \in [0, 1]$,

$$\frac{\|\mathbf{m}_{t,\varepsilon}\|_{\text{TV}}}{\sqrt{\varepsilon}} \leq \|\mu - \mu^k\|_{L^1} + \sqrt{\varepsilon} \|(\mu^k)'\|_{L^1}.$$

Define, for $r > 0$,

$$\omega(r) \stackrel{\text{def.}}{=} \inf_{k \in \mathbb{N}} \left\{ \|\mu - \mu^k\|_{L^1} + r \|(\mu^k)'\|_{L^1} \right\}.$$

For each fixed k , $0 \leq \limsup_{r \downarrow 0} \omega(r) \leq \|\mu - \mu^k\|_{L^1}$, and letting $k \rightarrow \infty$ shows that $\omega(r) \rightarrow 0$ as $r \rightarrow 0$. Therefore

$$\sup_{t \in [0,1]} \frac{\|\mathbf{m}_{t,\varepsilon}\|_{\text{TV}}}{\sqrt{\varepsilon}} \leq \omega(\sqrt{\varepsilon}) \rightarrow 0 \quad (\varepsilon \rightarrow 0). \quad (4.17)$$

Step 2. Estimate of $W_2(\mu, \tilde{\mu}_\varepsilon)$. By Step 1, the curve $(\mu_{t,\varepsilon})_{t \in [0,1]}$ of probability measures with finite second moments solves, together with $(\mathbf{m}_{t,\varepsilon})_{t \in [0,1]}$, the continuity equation, and it is weakly continuous, with $\mu_{0,\varepsilon} = \mu$ and $\mu_{1,\varepsilon} = \tilde{\mu}_\varepsilon$. Since $|z| \leq R$ in the definition of $\mathbf{m}_{t,\varepsilon}$, we have

$$|\mathbf{m}_{t,\varepsilon}(x)| \leq \frac{\sqrt{\varepsilon}R}{c_R} \int_{-R}^R \mu(x - t\sqrt{\varepsilon}z) e^{-z^2/2} dz = \sqrt{\varepsilon}R \mu_{t,\varepsilon}(x) \quad \text{for a.e. } x \in \mathbb{R}.$$

Thus $\mathbf{m}_{t,\varepsilon} \ll \mu_{t,\varepsilon}$, and the velocity field

$$v_{t,\varepsilon} \stackrel{\text{def.}}{=} \frac{d\mathbf{m}_{t,\varepsilon}}{d\mu_{t,\varepsilon}} \quad \text{satisfies} \quad |v_{t,\varepsilon}| \leq \sqrt{\varepsilon}R \quad \mu_{t,\varepsilon}\text{-a.e.}$$

By the Benamou–Brenier formula (see, e.g., [San15]), with $(\mu_{t,\varepsilon}, v_{t,\varepsilon})_{t \in [0,1]}$ as a competitor, and by (4.17), we get

$$\begin{aligned} W_2^2(\mu, \tilde{\mu}_\varepsilon) &\leq \int_0^1 \int_{\mathbb{R}} |v_{t,\varepsilon}|^2 d\mu_{t,\varepsilon} dt = \int_0^1 \int_{\mathbb{R}} |v_{t,\varepsilon}| d|\mathbf{m}_{t,\varepsilon}| dt \\ &\leq \sqrt{\varepsilon}R \sup_{t \in [0,1]} \|\mathbf{m}_{t,\varepsilon}\|_{\text{TV}} \leq \varepsilon R \omega(\sqrt{\varepsilon}). \end{aligned}$$

Hence $W_2(\mu, \tilde{\mu}_\varepsilon) \leq \sqrt{R} \sqrt{\omega(\sqrt{\varepsilon})} \sqrt{\varepsilon}$, which concludes the proof, since ω depends only on μ . $\square$

4.4.2 General statement

Before turning to the proof, we introduce some notation and a technical lemma. We fix $i \in \{1, 2\}$ and write $M \stackrel{\text{def.}}{=} M_i = \mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$. Every open ball of $\mathbb{T}^d$ with radius less than $1/2$ is the isometric image of a Euclidean ball under the canonical projection $\mathbb{R}^d \rightarrow \mathbb{T}^d$. All the points involved in the computations below lie in such balls, which we identify with the corresponding Euclidean balls, so that sums, segments and convexity are understood in $\mathbb{R}^d$. In this section and in Section 5, C denotes a positive constant that does not depend on ε , whose value may change from line to line, and we write $a \lesssim b$ when $a \leq Cb$.

By Proposition 2.2, $\dim \Sigma = d$ and the projection $\text{proj}_M^\Sigma : \Sigma \rightarrow M$ is a local $C^{1,1}$ -diffeomorphism. More precisely, there exist a finite index set J , an open covering $(U_j)_{j \in J}$ of Σ , and open sets $V_j \subseteq M$, which we may assume convex and contained in balls of radius $1/4$, such that, for every $j \in J$,

$$\text{proj}_M^{U_j} : U_j \longrightarrow V_j$$

is a $C^{1,1}$ -diffeomorphism. We denote its inverse by

$$\alpha_j \stackrel{\text{def.}}{=} \left(\text{proj}_M^{U_j} \right)^{-1} : V_j \longrightarrow U_j.$$

As in Lemma 3.2, the Jacobian $\mathcal{J}_F$ of a map F between d -dimensional manifolds (here, open subsets of $\mathbb{R}^d$ or of Σ) denotes the absolute value of the determinant of DF in orthonormal bases. Let $(\chi_j)_{j \in J}$ be a partition of unity on Σ subordinate to the covering $(U_j)_{j \in J}$, that is, each $\chi_j : \Sigma \rightarrow [0, 1]$ is continuous with compact support in U_j (recall that Σ is compact), and $\sum_{j \in J} \chi_j = 1$ on Σ .

Lemma 4.7. *Let $V \subseteq \mathbb{R}^d$ be an open set and $f : V \rightarrow \mathbb{R}^d$ be 1-Lipschitz. There exists $C_d \geq 2$, depending only on d , such that, for $0 < \tau \leq 1/C_d$, the map $\psi_\tau(x) \stackrel{\text{def.}}{=} x + \tau f(x)$ is a bi-Lipschitz homeomorphism from V onto the open set $\psi_\tau(V)$, and $\kappa_\tau \stackrel{\text{def.}}{=} \psi_\tau^{-1}$ satisfies $\|\mathcal{J}_{\psi_\tau} - 1\|_{L^\infty(V)} + \|\mathcal{J}_{\kappa_\tau} - 1\|_{L^\infty(\psi_\tau(V))} \leq C_d \tau$.*

Proof. Since f is 1-Lipschitz and $\tau < 1$, we have $(1 - \tau)|x - y| \leq |\psi_\tau(x) - \psi_\tau(y)| \leq (1 + \tau)|x - y|$ for $x, y \in V$. Thus ψ_τ is a bi-Lipschitz homeomorphism onto its image, which is open by the invariance of domain theorem. By Rademacher's theorem, f is differentiable a.e. on V , with $\|Df\| \leq 1$, and $D\psi_\tau = \text{Id} + \tau Df$ there. Since the determinant is Lipschitz on the set of matrices B with $\|B\| \leq 1$, there exists C_d such that $|\det(\text{Id} + B) - 1| \leq C_d\|B\|$ for such B . Applying this to $B = \tau Df(x)$ gives $|\mathcal{J}_{\psi_\tau} - 1| \leq C_d\tau$ a.e. on V . Finally, since the Lipschitz maps ψ_τ and κ_τ map negligible sets to negligible sets, for a.e. $y \in \psi_\tau(V)$, κ_τ is differentiable at y and ψ_τ is differentiable at $\kappa_\tau(y)$, so that $\mathcal{J}_{\kappa_\tau}(y) = 1/\mathcal{J}_{\psi_\tau}(\kappa_\tau(y))$ by the chain rule. Hence the same bound holds for $\mathcal{J}_{\kappa_\tau}$, up to enlarging C_d . $\square$

Proof of Theorem 4.4. Let $\varepsilon > 0$ be small enough so that $\sqrt{\varepsilon}R < 1/4$. All the displacements considered below are bounded by $\sqrt{\varepsilon}R$, so that, starting from V_j , they stay in a ball of radius $1/2$, where the identification with $\mathbb{R}^d$ holds. Recall that $\bar{\gamma} = \bar{\rho}\mathcal{H}_\Sigma^d$. We define the $d \times k$ matrix (recall that $k = d$)

$$A(z) \stackrel{\text{def.}}{=} \left(\text{proj}_M(n_1(z)), \dots, \text{proj}_M(n_k(z)) \right).$$

Equivalently, $A(z)s = \sum_{p=1}^k s_p \text{proj}_M(n_p(z))$. Therefore,

$$\text{proj}_M \Phi(z, s) = \text{proj}_M(z + n(z)s) = \text{proj}_M(z) + A(z)s.$$

Since $K(z, \cdot)$ is even, we have, for every $z \in \Sigma$,

$$\int_{\mathbb{B}_R} sK(z, s) ds = 0. \tag{4.18}$$

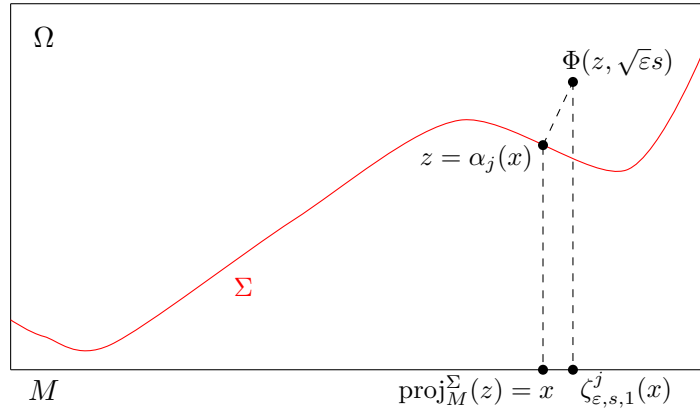


Figure 1: Normal perturbation of $z \in \Sigma$ and its projection onto M .

Step 1. Construction of the interpolation. As in Section 4.4.1, we interpolate between μ and $\tilde{\mu}_\varepsilon$ by letting the variance of the normal perturbation grow from 0 to ε :

for $t \in [0, 1]$, we let $\mu_{t,\varepsilon} \stackrel{\text{def.}}{=} \tilde{\mu}_{t^2\varepsilon}$ be the marginal of the normal perturbation $\tilde{\gamma}_{t^2\varepsilon}$, with the convention $\tilde{\gamma}_0 \stackrel{\text{def.}}{=} \bar{\gamma}$. In other words, by (4.6), for every $\phi \in C^1(M)$,

$$\int_M \phi d\mu_{t,\varepsilon} = \int_{\Sigma} \int_{\mathbb{B}_R} \phi(\text{proj}_M(\Phi(z, t\sqrt{\varepsilon}s))) K(z, s) ds d\bar{\gamma}(z), \quad (4.19)$$

so that $\mu_{0,\varepsilon} = \mu$ and $\mu_{1,\varepsilon} = \tilde{\mu}_\varepsilon$.

To compute $\mu_{t,\varepsilon}$, we use the charts α_j . For $j \in J$ and $x \in V_j$, we write

$$A_j(x) \stackrel{\text{def.}}{=} A(\alpha_j(x)), \quad K_j(x, s) \stackrel{\text{def.}}{=} K(\alpha_j(x), s).$$

Inserting $\sum_{j \in J} \chi_j = 1$ in (4.19), then changing variables $z = \alpha_j(x)$ in each term by the area formula, and using $\text{proj}_M(\Phi(\alpha_j(x), t\sqrt{\varepsilon}s)) = x + t\sqrt{\varepsilon}A_j(x)s$, we get

$$\begin{aligned} \int_M \phi d\mu_{t,\varepsilon} &= \sum_{j \in J} \int_{U_j} \int_{\mathbb{B}_R} \chi_j(z) \phi(\text{proj}_M(\Phi(z, t\sqrt{\varepsilon}s))) K(z, s) ds \bar{\rho}(z) d\mathcal{H}^d(z) \\ &= \sum_{j \in J} \int_{V_j} \int_{\mathbb{B}_R} \phi(x + t\sqrt{\varepsilon}A_j(x)s) K_j(x, s) ds \\ &\quad \times \chi_j(\alpha_j(x)) \bar{\rho}(\alpha_j(x)) \mathcal{J}_{\alpha_j}(x) d\text{vol}_M(x) \\ &= \sum_{j \in J} \int_{V_j} \int_{\mathbb{B}_R} \phi(\zeta_{\varepsilon,s,t}^j(x)) \rho_j(x) K_j(x, s) ds d\text{vol}_M(x), \end{aligned}$$

where, for $s \in \mathbb{B}_R$, $t \in [0, 1]$ and $x \in V_j$, we have set

$$\zeta_{\varepsilon,s,t}^j(x) \stackrel{\text{def.}}{=} x + t\sqrt{\varepsilon}A_j(x)s, \quad \rho_j(x) \stackrel{\text{def.}}{=} \chi_j(\alpha_j(x)) \bar{\rho}(\alpha_j(x)) \mathcal{J}_{\alpha_j}(x).$$

Here $\zeta_{\varepsilon,s,t}^j(x)$ is the projection onto M of the point $\Phi(\alpha_j(x), t\sqrt{\varepsilon}s)$, obtained by moving the point $\alpha_j(x) \in \Sigma$ lying above x in the normal direction (see Figure 1), and ρ_j is the density, with respect to vol_M , of the partial marginal $(\text{proj}_M)_\#(\chi_j \bar{\gamma})$, concentrated on V_j . Hence

$$\mu_{t,\varepsilon} = \sum_{j \in J} \mu_{t,\varepsilon}^j, \quad \mu_{t,\varepsilon}^j \stackrel{\text{def.}}{=} ((s, x) \mapsto \zeta_{\varepsilon,s,t}^j(x))_\# (\rho_j(x) K_j(x, s) ds d\text{vol}_M(x)). \quad (4.20)$$

Equivalently, $\mu_{t,\varepsilon}^j = \int_{\mathbb{B}_R} (\zeta_{\varepsilon,s,t}^j)_\# (\rho_j K_j(\cdot, s) \text{vol}_M) ds$, which is the form used below.

Step 2. Construction of the flux. For every $\phi \in C^1(M)$, by (4.20),

$$\begin{aligned} \frac{d}{dt} \int_M \phi d\mu_{t,\varepsilon} &= \frac{d}{dt} \sum_{j \in J} \int_{V_j} \int_{\mathbb{B}_R} \phi(\zeta_{\varepsilon,s,t}^j(x)) \rho_j(x) K_j(x, s) ds d\text{vol}_M(x) \\ &= \sum_{j \in J} \int_{V_j} \int_{\mathbb{B}_R} \langle \nabla \phi(\zeta_{\varepsilon,s,t}^j(x)), \partial_t \zeta_{\varepsilon,s,t}^j(x) \rangle \rho_j(x) K_j(x, s) ds d\text{vol}_M(x) \\ &= \sum_{j \in J} \int_{V_j} \int_{\mathbb{B}_R} \sqrt{\varepsilon} \langle \nabla \phi(\zeta_{\varepsilon,s,t}^j(x)), A_j(x)s \rangle \rho_j(x) K_j(x, s) ds d\text{vol}_M(x) \\ &= \int_M \langle \nabla \phi, d\mathbf{m}_{t,\varepsilon} \rangle, \end{aligned}$$

where the differentiation under the integral sign is justified since $\nabla\phi$ is bounded and $|\partial_t \zeta_{\varepsilon,s,t}^j(x)| = \sqrt{\varepsilon} |A_j(x)s| \leq \sqrt{\varepsilon} R$, and where the flux $\mathbf{m}_{t,\varepsilon} \in \mathcal{M}(M; TM)$ is the vector measure defined by

$$\mathbf{m}_{t,\varepsilon} \stackrel{\text{def.}}{=} \sqrt{\varepsilon} \sum_{j \in J} \int_{\mathbb{B}_R} (\zeta_{\varepsilon,s,t}^j)_{\#} (\rho_j K_j(\cdot, s) A_j s \text{ vol}_M) \, ds.$$

Integrating in time against a test function in $C_c^1((0, 1))$ yields

$$\partial_t \mu_{t,\varepsilon} + \text{div } \mathbf{m}_{t,\varepsilon} = 0 \quad \text{in } \mathcal{D}'((0, 1) \times M).$$

Step 3. Gaussian symmetry. As in [Section 4.4.1](#), the flux depends linearly on the densities ρ_j . For $j \in J$, $s \in \mathbb{B}_R$ and $f \in L^1(V_j)$, let

$$F_{j,s}[f](x) \stackrel{\text{def.}}{=} \begin{cases} f(x) K_j(x, s) A_j(x) s, & x \in V_j, \\ 0, & x \notin V_j, \end{cases}$$

so that, by definition of the flux,

$$\frac{\mathbf{m}_{t,\varepsilon}}{\sqrt{\varepsilon}} = \sum_{j \in J} \int_{\mathbb{B}_R} (\zeta_{\varepsilon,s,t}^j)_{\#} (F_{j,s}[\rho_j] \text{ vol}_M) \, ds.$$

By [\(4.18\)](#), $\int_{\mathbb{B}_R} F_{j,s}[\rho_j] \, ds = 0$. Subtracting this in each term, we obtain

$$\frac{\mathbf{m}_{t,\varepsilon}}{\sqrt{\varepsilon}} = \sum_{j \in J} \int_{\mathbb{B}_R} \left((\zeta_{\varepsilon,s,t}^j)_{\#} - \text{Id} \right) (F_{j,s}[\rho_j] \text{ vol}_M) \, ds = \sum_{j \in J} \mathbf{I}_{t,\varepsilon}^j[\rho_j], \quad (4.21)$$

where, for $f \in L^1(V_j)$,

$$\mathbf{I}_{t,\varepsilon}^j[f] \stackrel{\text{def.}}{=} \int_{\mathbb{B}_R} \left((\zeta_{\varepsilon,s,t}^j)_{\#} - \text{Id} \right) (F_{j,s}[f] \text{ vol}_M) \, ds.$$

Step 4. Estimate of the displacement term. Fix $j \in J$. We estimate the total variation of $\mathbf{I}_{t,\varepsilon}^j[f]$ in two ways.

(a) For $f \in L^1(V_j)$. Since the push-forward operation does not increase the total variation and $|A(z)s| \leq |n(z)s| = |s|$, [\(4.5\)](#) gives

$$\begin{aligned} \|\mathbf{I}_{t,\varepsilon}^j[f]\|_{\text{TV}} &\leq 2 \int_{\mathbb{B}_R} \|F_{j,s}[f]\|_{L^1(V_j)} \, ds \\ &\leq 2 \int_{V_j} |f(x)| \int_{\mathbb{B}_R} |s| K_j(x, s) \, ds \, d \text{ vol}_M(x) \leq 2M_1 \|f\|_{L^1(V_j)}. \end{aligned}$$

(b) For $f \in C_c^1(V_j)$ supported in a fixed compact set $S \subseteq V_j$. Write $F_s \stackrel{\text{def.}}{=} F_{j,s}[f]$. Fix $t \in [0, 1]$ and $s \in \mathbb{B}_R$, and write $\zeta \stackrel{\text{def.}}{=} \zeta_{\varepsilon,s,t}^j$, $\kappa \stackrel{\text{def.}}{=} \zeta^{-1}$ and $W \stackrel{\text{def.}}{=} \zeta(V_j)$. Since

$|\zeta(x) - x| \leq \sqrt{\varepsilon}R$, we may choose $\varepsilon > 0$ sufficiently small, depending only on S (and not on f , s and t), so that

$$S \Subset W \quad \text{and} \quad \zeta(S) \Subset V_j.$$

By the area formula, the density of $\zeta_{\#}(F_s \text{vol}_M)$ is $F_s \circ \kappa \mathcal{J}_{\kappa} \mathbf{1}_W$, and, since $\text{spt}(F_s) \subseteq S$, the above inclusions show that $F_s \circ \kappa \mathbf{1}_W$ and F_s are both supported in $W \cap V_j$. Hence

$$\begin{aligned} \|\zeta_{\#}(F_s \text{vol}_M) - F_s \text{vol}_M\|_{\text{TV}} &\leq \|\mathcal{J}_{\kappa} - 1\|_{L^{\infty}(W)} \int_W |F_s \circ \kappa| \, d\text{vol}_M \\ &\quad + \int_{W \cap V_j} |F_s \circ \kappa - F_s| \, d\text{vol}_M. \end{aligned}$$

By [Lemma 4.7](#), applied with $\tau \stackrel{\text{def.}}{=} t\sqrt{\varepsilon}|s|L_j$ and $f \stackrel{\text{def.}}{=} A_j s/(|s|L_j)$, where L_j is a Lipschitz constant of A_j on the convex set V_j (recall that n is Lipschitz and α_j is $C^{1,1}$), we have, for ε small enough, $\|\mathcal{J}_{\kappa} - 1\|_{L^{\infty}(W)} + \|\mathcal{J}_{\zeta} - 1\|_{L^{\infty}(V_j)} \leq C\sqrt{\varepsilon}|s|$, and in particular $1/2 \leq \mathcal{J}_{\kappa}, \mathcal{J}_{\zeta} \leq 2$ a.e. Since, by the area formula (change of variables $y = \kappa(x)$),

$$\int_W |F_s \circ \kappa| \, d\text{vol}_M \leq 2 \int_W |F_s \circ \kappa| \mathcal{J}_{\kappa} \, d\text{vol}_M = 2\|F_s\|_{L^1(V_j)},$$

the first term is at most $C\sqrt{\varepsilon}|s| \|F_s\|_{L^1(V_j)}$. For the second term, recall that F_s is defined on M and vanishes outside V_j . By the area formula (change of variables $x = \zeta(y)$),

$$\int_{W \cap V_j} |F_s \circ \kappa - F_s| \, d\text{vol}_M \leq 2 \int_{V_j} |F_s \circ \zeta - F_s| \, d\text{vol}_M \leq 4\sqrt{\varepsilon}|s| \|\nabla F_s\|_{L^1(V_j)}.$$

To prove the last inequality, assume first that F_s is C^1 . Since $u \mapsto \zeta_{\varepsilon,s,ut}^j(y) = y + ut\sqrt{\varepsilon}A_j(y)s$ parametrizes the segment from y to $\zeta(y)$ and $|A_j(y)s| \leq |s|$, we have

$$|F_s(\zeta(y)) - F_s(y)| \leq \sqrt{\varepsilon}|s| \int_0^1 |\nabla F_s(\zeta_{\varepsilon,s,ut}^j(y))| \, du,$$

and we conclude by the area formula (change of variables $x = \zeta_{\varepsilon,s,ut}^j(y)$), since the inverse of $\zeta_{\varepsilon,s,ut}^j$ has Jacobian at most 2 by [Lemma 4.7](#) (applied with ut in place of t). In general, F_s is only Lipschitz on M , since A_j and $K_j(\cdot, s)$ are Lipschitz on V_j (see below) and f has compact support in V_j . We then apply the above to mollifications of F_s , which converge to F_s in $W^{1,1}(M)$, and pass to the limit, using that $\int_{V_j} |g \circ \zeta| \, d\text{vol}_M \leq 2\|g\|_{L^1(M)}$ for $g \in L^1(M)$, since $\mathcal{J}_{\kappa} \leq 2$. Altogether,

$$\|\zeta_{\#}(F_s \text{vol}_M) - F_s \text{vol}_M\|_{\text{TV}} \leq C\sqrt{\varepsilon}|s| \left(\|F_s\|_{L^1(V_j)} + \|\nabla F_s\|_{L^1(V_j)} \right).$$

It remains to bound the L^1 norms of F_s and ∇F_s . By [\(2.3\)](#) and since $Z_R \geq c$, we have $K_j(x, s) \leq c^{-1}e^{-\lambda|s|^2/2}$. Moreover, $x \mapsto H_{\alpha_j(x)}E$ is Lipschitz (since $E \in C^{2,1}$, n is Lipschitz and α_j is $C^{1,1}$), hence so are $x \mapsto Z_R(\alpha_j(x))$ and $K_j(\cdot, s)$, and differentiating a.e. $K_j(x, s) = \mathbf{1}_{\mathbb{B}_R}(s) e^{-H_{\alpha_j(x)}E(s,s)/2} / Z_R(\alpha_j(x))$ with respect to x gives $|\nabla_x K_j(x, s)| \leq$

$C(1 + |s|^2)e^{-\lambda|s|^2/2}$ for a.e. $x \in V_j$. Since $F_s = f K_j(\cdot, s) A_j s$, with $|A_j s| \leq |s|$ and A_j Lipschitz, the product rule yields

$$\|F_s\|_{L^1(V_j)} \leq C e^{-\lambda|s|^2/2} |s| \|f\|_{L^1(V_j)},$$

$$\|\nabla F_s\|_{L^1(V_j)} \leq C e^{-\lambda|s|^2/2} \left[|s| \|\nabla f\|_{L^1(V_j)} + (|s| + |s|^3) \|f\|_{L^1(V_j)} \right].$$

Integrating with respect to $s \in \mathbb{B}_R$, we obtain, uniformly in $t \in [0, 1]$,

$$\|\mathbf{I}_{t,\varepsilon}^j[f]\|_{\text{TV}} \leq C \sqrt{\varepsilon} (M_2 + M_4) \left(\|f\|_{L^1(V_j)} + \|\nabla f\|_{L^1(V_j)} \right),$$

where M_2 and M_4 are defined in (4.5).

Conclusion. Since χ_j has compact support in U_j , ρ_j vanishes outside the compact set $\text{proj}_M(\text{spt } \chi_j) \subseteq V_j$. By mollification, there exist a compact set $S_j \subseteq V_j$ and a sequence $(\rho_j^{(k)})_{k \in \mathbb{N}}$ in $C_c^1(V_j)$, all supported in S_j , such that $\|\rho_j - \rho_j^{(k)}\|_{L^1(V_j)} \rightarrow 0$ as $k \rightarrow \infty$. Applying (b) with $S = S_j$, there exists $\varepsilon_0 > 0$, independent of k , such that, for $0 < \varepsilon \leq \varepsilon_0$ and every $k \in \mathbb{N}$, writing $\mathbf{I}_{t,\varepsilon}^j[\rho_j] = \mathbf{I}_{t,\varepsilon}^j[\rho_j - \rho_j^{(k)}] + \mathbf{I}_{t,\varepsilon}^j[\rho_j^{(k)}]$, (a) and (b) give

$$\sup_{t \in [0,1]} \|\mathbf{I}_{t,\varepsilon}^j[\rho_j]\|_{\text{TV}} \leq 2M_1 \|\rho_j - \rho_j^{(k)}\|_{L^1(V_j)} + C \sqrt{\varepsilon} \left(\|\rho_j^{(k)}\|_{L^1(V_j)} + \|\nabla \rho_j^{(k)}\|_{L^1(V_j)} \right).$$

For $r > 0$, set

$$\omega_j(r) \stackrel{\text{def.}}{=} \inf_{k \geq 1} \left\{ 2M_1 \|\rho_j - \rho_j^{(k)}\|_{L^1(V_j)} + rC \left(\|\rho_j^{(k)}\|_{L^1(V_j)} + \|\nabla \rho_j^{(k)}\|_{L^1(V_j)} \right) \right\},$$

and $\omega(r) \stackrel{\text{def.}}{=} \sum_{j \in J} \omega_j(r)$. Since $\rho_j^{(k)} \rightarrow \rho_j$ in $L^1(V_j)$, we have $\omega(r) \rightarrow 0$ as $r \rightarrow 0$: one first fixes k sufficiently large so that the first term is arbitrarily small, and then lets $r \rightarrow 0$. Combining this with (4.21) yields, for $0 < \varepsilon \leq \varepsilon_0$,

$$\sup_{t \in [0,1]} \frac{\|\mathbf{m}_{t,\varepsilon}\|_{\text{TV}}}{\sqrt{\varepsilon}} \leq \sum_{j \in J} \sup_{t \in [0,1]} \|\mathbf{I}_{t,\varepsilon}^j[\rho_j]\|_{\text{TV}} \leq \omega(\sqrt{\varepsilon}) = o(1).$$

Step 5. Benamou–Brenier estimate. Recall from Step 2 and (4.20) that

$$\begin{aligned} \mathbf{m}_{t,\varepsilon} &= \sqrt{\varepsilon} \sum_{j \in J} \int_{\mathbb{B}_R} (\zeta_{\varepsilon,s,t}^j)_{\#} (\rho_j K_j(\cdot, s) A_j s \text{ vol}_M) \, ds, \\ \mu_{t,\varepsilon} &= \sum_{j \in J} \int_{\mathbb{B}_R} (\zeta_{\varepsilon,s,t}^j)_{\#} (\rho_j K_j(\cdot, s) \text{ vol}_M) \, ds. \end{aligned}$$

Since $|A_j(x)s| \leq |n(\alpha_j(x))s| = |s| \leq R$, we get $|\mathbf{m}_{t,\varepsilon}| \leq \sqrt{\varepsilon} R \mu_{t,\varepsilon}$. Thus $\mathbf{m}_{t,\varepsilon} \ll \mu_{t,\varepsilon}$, and the velocity field

$$v_{t,\varepsilon} \stackrel{\text{def.}}{=} \frac{d\mathbf{m}_{t,\varepsilon}}{d\mu_{t,\varepsilon}} \in L^\infty(\mu_{t,\varepsilon}; TM) \quad \text{satisfies} \quad |v_{t,\varepsilon}| \leq \sqrt{\varepsilon} R \quad \mu_{t,\varepsilon}\text{-a.e.}$$

Since $(\mu_{t,\varepsilon}, v_{t,\varepsilon})$ satisfies the continuity equation, with $\mu_{0,\varepsilon} = \mu$ and $\mu_{1,\varepsilon} = \tilde{\mu}_\varepsilon$, the Benamou–Brenier formula (see, e.g., [San15]) yields

$$\begin{aligned} W_2^2(\mu, \tilde{\mu}_\varepsilon) &\leq \int_0^1 \int_M |v_{t,\varepsilon}|^2 d\mu_{t,\varepsilon} dt = \int_0^1 \int_M |v_{t,\varepsilon}| d|\mathbf{m}_{t,\varepsilon}| dt \\ &\leq \sqrt{\varepsilon} R \sup_{t \in [0,1]} \|\mathbf{m}_{t,\varepsilon}\|_{\text{TV}}. \end{aligned}$$

By Step 4, $\sup_{t \in [0,1]} \|\mathbf{m}_{t,\varepsilon}\|_{\text{TV}} \leq \sqrt{\varepsilon} \omega(\sqrt{\varepsilon})$ for $0 < \varepsilon \leq \varepsilon_0$, hence

$$W_2^2(\mu, \tilde{\mu}_\varepsilon) \leq \varepsilon R \omega(\sqrt{\varepsilon}).$$

Since $\omega(r) \rightarrow 0$ as $r \rightarrow 0$, this gives $W_2(\mu, \tilde{\mu}_\varepsilon) = o_R(\sqrt{\varepsilon})$. $\square$

5 Extension to generalized framework and examples

In this section, we extend [Theorem A](#) to compactly supported marginals on $\mathbb{R}^d$ ([Section 5.1](#)), where we also prove [Theorem 3.4](#), and to $N \geq 2$ marginals ([Section 5.2](#)). We illustrate these extensions with the quadratic cost, the non-twisted cost of [MPW12, Example 3.1], and a repulsive harmonic cost with three marginals on the circle.

5.1 Compactly supported marginals on $\mathbb{R}^d$

In accordance with [Section 2](#), where M_1 and M_2 are compact, we take for M_i compact convex subsets of $\mathbb{R}^d$ containing the supports of the marginals in their interior, and we replace [\(H1\)–\(H4\)](#) by the following assumptions.

- (H1') For $i = 1, 2$, $M_i \subset \mathbb{R}^d$ is a compact convex set, and $\Omega = M_1 \times M_2$. The cost c is continuous on Ω , and we fix continuous Kantorovich potentials $\bar{\varphi}_1, \bar{\varphi}_2$. The cost is C^2 on $\text{int } \Omega$ and non-degenerate on an open neighborhood $U \subset \text{int } \Omega$ of $\Sigma \cap \text{int } \Omega$: $D_{x_1 x_2}^2 c(z)$ is invertible for every $z \in U$.
- (H2') The duality gap $E = c - \bar{\varphi}_1 \oplus \bar{\varphi}_2 \geq 0$ belongs to $C^{2,1}(\text{int } \Omega)$; in particular, $\text{Hess } E$ is bounded on $\text{int } \Omega$.
- (H3') The set $\Sigma \cap \text{int } \Omega$ is a nonempty $C^{1,1}$ submanifold of $\text{int } \Omega$ without boundary, of dimension d . As before, $k \stackrel{\text{def.}}{=} d$ denotes its codimension.
- (H4') For $i = 1, 2$, $\mu_i = \rho_i dx$, where ρ_i vanishes outside a compact set $K_i \subset \text{int } M_i$, is Lipschitz on K_i , and satisfies $0 < \alpha \leq \rho_i \leq \beta$ on K_i . Moreover, setting $\mathcal{K} \stackrel{\text{def.}}{=} K_1 \times K_2$, we have $\mathcal{H}^d(\Sigma \cap \partial \mathcal{K}) = 0$.

We set $\rho(x_1, x_2) \stackrel{\text{def.}}{=} \rho_1(x_1)\rho_2(x_2)$, so that $\mu^\otimes = \mu_1 \otimes \mu_2 = \rho dx = \rho \mathbf{1}_{\mathcal{K}} dx$. Compared with the torus, the new features are that Σ may extend beyond $\mathcal{K}$ and cross its boundary, and that $\mu^\otimes$ is discontinuous across $\partial \mathcal{K}$.

5.1.1 Convergence of the Gibbs measure and the geometric Laplace's method

This section is devoted to the proof of the convergence of the Gibbs measure in the case of compactly supported marginals on $\mathbb{R}^d$. The result slightly extends the setting considered in [Hwa80], since the metric and tubular neighborhoods of the contact set need not coincide. The same proof, with $\mathcal{K} = \omega = \Omega$ (so that $\mathbf{1}_{\mathcal{K}} \equiv 1$, $\Sigma \cap \omega = \Sigma$, and Lemma 5.2 follows from Section 3.1), gives Theorem 3.4 on the torus.

Theorem 5.1 (Convergence of the Gibbs measure). *Under assumptions (H1')–(H4'),*

$$\mathbb{P}_\varepsilon \xrightarrow{*} \mathbb{P}_0, \quad (5.1)$$

where

$$\mathbb{P}_0 = \frac{1}{Z_0} (\det \text{Hess}^\perp E)^{-1/2} \rho \mathbf{1}_{\mathcal{K}} \mathcal{H}^d|_\Sigma \quad \text{and} \quad Z_0 = \int_{\Sigma \cap \mathcal{K}} \frac{\rho_1(z_1) \rho_2(z_2)}{\sqrt{\det \text{Hess}^\perp E(z)}} d\mathcal{H}^d(z).$$

Moreover, $Z_\varepsilon = (2\pi\varepsilon)^{k/2} (Z_0 + o(1))$.

Before turning to the proof, we first establish a preliminary result. We fix an open set ω such that $\mathcal{K} \Subset \omega \Subset \text{int } \Omega$. Since $\Sigma \cap \bar{\omega}$ is a compact subset of the $C^{1,1}$ submanifold $\Sigma \cap \text{int } \Omega$, the constructions of Section 3.1 and the estimates of Propositions 2.2 and 3.3 hold on $\Sigma \cap \bar{\omega}$, with uniform constants. For $r > 0$ small enough, we denote by

$$\mathcal{T}_r \stackrel{\text{def.}}{=} \Phi((\Sigma \cap \omega) \times \mathbb{B}_r) = \{z + v : z \in \Sigma \cap \omega, v \in N_z \Sigma, |v| < r\}$$

the tubular neighborhood of radius r of $\Sigma \cap \omega$, not to be confused with the r -neighborhood $\Sigma_r = \{\mathbf{x} \in \Omega : \text{dist}(\mathbf{x}, \Sigma) < r\}$ of Σ .

Lemma 5.2. *Under assumptions (H1')–(H4'), there exists $r_0 > 0$ such that, for every $0 < r \leq r_0$,*

$$\mathcal{K} \cap \Sigma_r \subset \mathcal{T}_r.$$

Consequently,

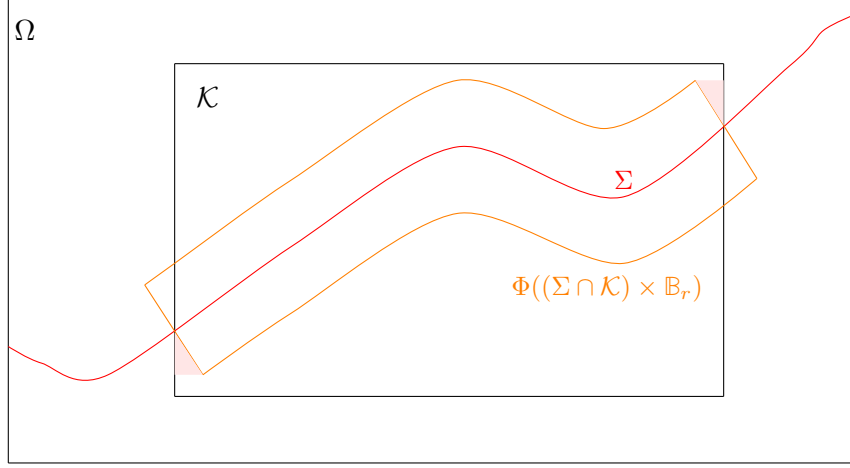
$$x \in \mathcal{K} \setminus \mathcal{T}_r \implies \text{dist}(x, \Sigma) \geq r.$$

Proof. Set $\eta \stackrel{\text{def.}}{=} \text{dist}(\mathcal{K}, \mathbb{R}^{2d} \setminus \omega) > 0$. Since Σ is a $C^{1,1}$ submanifold, it has locally positive reach. By compactness of $\Sigma \cap \bar{\omega}$, there exists $r_0 \in (0, \eta)$ such that the nearest-point projection proj_Σ is uniquely defined on $\mathcal{K} \cap \Sigma_{r_0}$. Let $0 < r \leq r_0$ and $\mathbf{x} \in \mathcal{K} \cap \Sigma_r$. Then $|\mathbf{x} - \text{proj}_\Sigma(\mathbf{x})| = \text{dist}(\mathbf{x}, \Sigma) < r$. Since $\mathbf{x} \in \mathcal{K}$ and $r_0 < \eta$, it follows that $\text{proj}_\Sigma(\mathbf{x}) \in \omega$. Moreover, $\mathbf{x} - \text{proj}_\Sigma(\mathbf{x}) \in N_{\text{proj}_\Sigma(\mathbf{x})} \Sigma$. Hence $\mathbf{x} = \text{proj}_\Sigma(\mathbf{x}) + (\mathbf{x} - \text{proj}_\Sigma(\mathbf{x})) \in \mathcal{T}_r$. $\square$

The following figure illustrates the geometric reason for enlarging the base of the tubular neighborhood from $\Sigma \cap \mathcal{K}$ to $\Sigma \cap \omega$. This enlargement ensures that every point of $\mathcal{K}$ lying outside the tubular neighborhood is uniformly separated from the whole contact set:

$$x \in \mathcal{K} \setminus \mathcal{T}_r \implies \text{dist}(x, \Sigma) \geq r.$$

This is what allows us to decompose the integrals in the proof below.



If the tubular neighborhood were constructed only around $\Sigma \cap \mathcal{K}$, the shaded regions could contain points of $\mathcal{K}$ that remain close to portions of Σ lying outside $\mathcal{K}$. Thus, the distance to Σ would not be controlled solely by lying outside $\Phi((\Sigma \cap \mathcal{K}) \times \mathbb{B}_r)$.

Proof. Let r_0 be as in [Lemma 5.2](#). We fix $r \in (0, r_0]$ such that the bounds of [Proposition 3.3](#) hold on $\mathcal{T}_r$, and $R > 0$ so that $R\sqrt{\varepsilon} < r$ for ε small enough. Let $\phi \in C(\Omega)$ and introduce the unnormalized Gibbs measure

$$d\tilde{\mathbb{P}}_\varepsilon = (2\pi\varepsilon)^{-k/2} \exp\left(-\frac{E}{\varepsilon}\right) d\mu^\otimes,$$

so that $\mathbb{P}_\varepsilon = \tilde{\mathbb{P}}_\varepsilon / \tilde{\mathbb{P}}_\varepsilon(\Omega)$. We decompose Ω as

$$\begin{aligned} \int_\Omega \phi d\tilde{\mathbb{P}}_\varepsilon &= \int_{\Sigma_{R\sqrt{\varepsilon}}} \phi d\tilde{\mathbb{P}}_\varepsilon + \int_{\Sigma_r \setminus \Sigma_{R\sqrt{\varepsilon}}} \phi d\tilde{\mathbb{P}}_\varepsilon + \int_{\Omega \setminus \Sigma_r} \phi d\tilde{\mathbb{P}}_\varepsilon \\ &= \int_{\mathcal{T}_{R\sqrt{\varepsilon}}} \phi d\tilde{\mathbb{P}}_\varepsilon + \int_{\mathcal{T}_r \setminus \mathcal{T}_{R\sqrt{\varepsilon}}} \phi d\tilde{\mathbb{P}}_\varepsilon + \int_{\Omega \setminus \Sigma_r} \phi d\tilde{\mathbb{P}}_\varepsilon \\ &\stackrel{\text{def.}}{=} I_{1,\varepsilon,R}(\phi) + I_{2,\varepsilon,R}(\phi) + I_{3,\varepsilon}(\phi), \end{aligned}$$

where, for the second equality, we used that $\tilde{\mathbb{P}}_\varepsilon$ is supported on $\mathcal{K}$ and that $\mathcal{K} \cap \Sigma_\rho = \mathcal{K} \cap \mathcal{T}_\rho$ for $0 < \rho \leq r$, by [Lemma 5.2](#) and since $\mathcal{T}_\rho \subset \Sigma_\rho$. We analyze these three terms separately.

Step 1. Estimate on $\mathcal{T}_{R\sqrt{\varepsilon}}$. Using the tubular coordinates induced by Φ , we write $E(z, s)$ for $E(\Phi(z, s))$ and define

$$a(z, s) \stackrel{\text{def.}}{=} \phi(\Phi(z, s)) \rho(\Phi(z, s)) \mathcal{J}_\Phi(z, s) \mathbf{1}_{\mathcal{K}}(\Phi(z, s)),$$

for $\mathcal{H}^d \otimes \mathcal{L}^k$ -a.e. $(z, s) \in (\Sigma \cap \omega) \times \mathbb{B}_r$. Note that $|a| \leq \beta^2 \|\phi\|_\infty C_J$, where C_J bounds $\mathcal{J}_\Phi$ on $(\Sigma \cap \bar{\omega}) \times \mathbb{B}_r$. By the area formula [\(3.2\)](#),

$$I_{1,\varepsilon,R}(\phi) = (2\pi\varepsilon)^{-k/2} \int_{\Sigma \cap \omega} \int_{\mathbb{B}_{R\sqrt{\varepsilon}}} e^{-E(z,s)/\varepsilon} a(z, s) ds d\mathcal{H}^d_\Sigma(z).$$

Since $E(z) = 0$ and $\nabla E(z) = 0$ for $z \in \Sigma$, and since $E \in C^{2,1}$, we have

$$E(z, s) = \frac{1}{2} H_z E(s, s) + \mathcal{R}(z, s), \quad |\mathcal{R}(z, s)| \leq C|s|^3 \quad \text{for all } (z, s) \in (\Sigma \cap \bar{\omega}) \times \mathbb{B}_r.$$

Performing the change of variables $s = \sqrt{\varepsilon} \eta$, $ds = \varepsilon^{k/2} d\eta$, and then relabeling the normal variable, we obtain

$$I_{1,\varepsilon,R}(\phi) = (2\pi)^{-k/2} \int_{\Sigma \cap \omega} \int_{\mathbb{B}_R} e^{-\frac{1}{2} H_z E(s,s) - \frac{1}{\varepsilon} \mathcal{R}(z, \sqrt{\varepsilon} s)} a(z, \sqrt{\varepsilon} s) ds d\mathcal{H}_{|\Sigma}^d(z). \quad (5.2)$$

As $\varepsilon \rightarrow 0$, we have $|\mathcal{R}(z, \sqrt{\varepsilon} s)|/\varepsilon \leq C\sqrt{\varepsilon}|s|^3 \rightarrow 0$ and $\mathcal{J}_\Phi(z, \sqrt{\varepsilon} s) \rightarrow \mathcal{J}_\Phi(z, 0) = 1$ by [Lemma 3.2](#). Moreover, $\Phi(z, \sqrt{\varepsilon} s) \rightarrow z$ and $\mathbf{1}_\mathcal{K}$ is locally constant on $\mathbb{R}^{2d} \setminus \partial\mathcal{K}$, so that, ρ being continuous on $\mathcal{K}$, $(\rho \mathbf{1}_\mathcal{K})(\Phi(z, \sqrt{\varepsilon} s)) \rightarrow (\rho \mathbf{1}_\mathcal{K})(z)$ for every $z \in \Sigma \setminus \partial\mathcal{K}$; since $\mathcal{H}^d(\Sigma \cap \partial\mathcal{K}) = 0$, we get $a(z, \sqrt{\varepsilon} s) \rightarrow \phi(z) \rho(z) \mathbf{1}_\mathcal{K}(z)$ for $\mathcal{H}^d \otimes \mathcal{L}^k$ -a.e. $(z, s) \in (\Sigma \cap \omega) \times \mathbb{B}_R$. Since the integrand of (5.2) is bounded by $e^{C\sqrt{\varepsilon} R^3} \beta^2 \|\phi\|_\infty C_J \leq 2\beta^2 \|\phi\|_\infty C_J$ for ε small enough, and $\mathcal{H}^d(\Sigma \cap \omega)$ is finite, the dominated convergence theorem gives

$$\lim_{\varepsilon \rightarrow 0} I_{1,\varepsilon,R}(\phi) = (2\pi)^{-k/2} \int_{\Sigma \cap \mathcal{K}} \phi(z) \rho(z) \int_{\mathbb{B}_R} e^{-H_z E(s,s)/2} ds d\mathcal{H}_{|\Sigma}^d(z).$$

Next, since $(2\pi)^{-k/2} \int_{\mathbb{R}^k} e^{-H_z E(s,s)/2} ds = (\det H_z E)^{-1/2}$, the dominated convergence theorem gives

$$\lim_{R \rightarrow +\infty} \lim_{\varepsilon \rightarrow 0} I_{1,\varepsilon,R}(\phi) = \int_{\Sigma \cap \mathcal{K}} \phi(z) (\det H_z E)^{-1/2} \rho(z) d\mathcal{H}_{|\Sigma}^d(z).$$

Step 2. Estimate on $\mathcal{T}_r \setminus \mathcal{T}_{R\sqrt{\varepsilon}}$. As in Step 1, the area formula and the change of variables $s = \sqrt{\varepsilon} \eta$ give

$$|I_{2,\varepsilon,R}(\phi)| \leq (2\pi)^{-k/2} \beta^2 \|\phi\|_\infty C_J \int_{\Sigma \cap \omega} \int_{R \leq |s| < r/\sqrt{\varepsilon}} e^{-E(z, \sqrt{\varepsilon} s)/\varepsilon} ds d\mathcal{H}_{|\Sigma}^d(z).$$

By [Proposition 3.3](#) and (3.1), $E(z, \sqrt{\varepsilon} s) \geq \underline{C} \operatorname{dist}(\Phi(z, \sqrt{\varepsilon} s), \Sigma)^2 = \underline{C} \varepsilon |s|^2$ for $|s| < r/\sqrt{\varepsilon}$. Hence, uniformly in ε ,

$$|I_{2,\varepsilon,R}(\phi)| \leq (2\pi)^{-k/2} \beta^2 \|\phi\|_\infty C_J \mathcal{H}^d(\Sigma \cap \omega) \int_{\mathbb{R}^k \setminus \mathbb{B}_R} e^{-\underline{C}|s|^2} ds \xrightarrow{R \rightarrow +\infty} 0,$$

as the tail of an integrable Gaussian. This proves that the contribution of $\mathcal{T}_r \setminus \mathcal{T}_{R\sqrt{\varepsilon}}$ is negligible.

Step 3. Estimate on $\Omega \setminus \Sigma_r$. If $\mathcal{K} \setminus \Sigma_r = \emptyset$ there is nothing to do. Otherwise, $\mathcal{K} \setminus \Sigma_r$ is a compact set which does not meet Σ , so that $E > 0$ on it, and we denote

$$\delta \stackrel{\text{def.}}{=} \min_{\mathcal{K} \setminus \Sigma_r} E > 0.$$

Since $\tilde{\mathbb{P}}_\varepsilon$ is supported on $\mathcal{K}$ and $\mu^\otimes(\mathcal{K}) = 1$,

$$|I_{3,\varepsilon}(\phi)| \leq (2\pi\varepsilon)^{-k/2} \|\phi\|_\infty e^{-\delta/\varepsilon} \xrightarrow{\varepsilon \rightarrow 0} 0.$$

Conclusion. Set $L(\phi) \stackrel{\text{def.}}{=} \int_{\Sigma \cap \mathcal{K}} \phi(z) (\det H_z E)^{-1/2} \rho(z) d\mathcal{H}_{|\Sigma}^d(z)$. Since $\int_{\Omega} \phi d\tilde{\mathbb{P}}_{\varepsilon}$ does not depend on R , Steps 1–3 give, for every $R > 0$,

$$\limsup_{\varepsilon \rightarrow 0} \left| \int_{\Omega} \phi d\tilde{\mathbb{P}}_{\varepsilon} - L(\phi) \right| \leq \left| \lim_{\varepsilon \rightarrow 0} I_{1,\varepsilon,R}(\phi) - L(\phi) \right| + \sup_{\varepsilon} |I_{2,\varepsilon,R}(\phi)|,$$

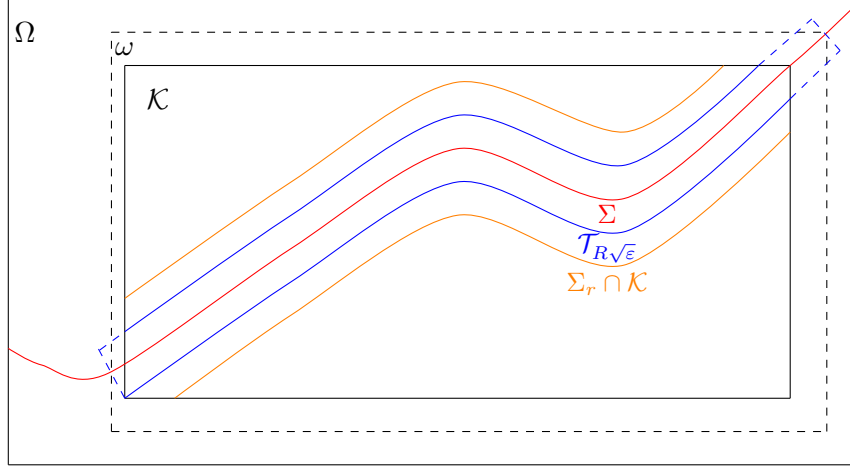
and the right-hand side tends to 0 as $R \rightarrow +\infty$. In other words,

$$\tilde{\mathbb{P}}_{\varepsilon} \xrightarrow{*} (\det \text{Hess}^{\perp} E)^{-1/2} \rho \mathbf{1}_{\mathcal{K}} \mathcal{H}_{|\Sigma}^d.$$

Taking $\phi \equiv 1$, we get $\tilde{\mathbb{P}}_{\varepsilon}(\Omega) = (2\pi\varepsilon)^{-k/2} Z_{\varepsilon} \rightarrow Z_0$, which is the asymptotics of Z_{ε} . Moreover, $Z_0 > 0$: an optimal plan is concentrated on $\Sigma \cap \mathcal{K}$, so that μ_1 is concentrated on $\text{proj}_{M_1}(\Sigma \cap \mathcal{K})$, and since μ_1 is absolutely continuous and proj_{M_1} is 1-Lipschitz, $0 < \mathcal{L}^d(\text{proj}_{M_1}(\Sigma \cap \mathcal{K})) \leq \mathcal{H}^d(\Sigma \cap \mathcal{K})$. Thus

$$\mathbb{P}_{\varepsilon} = \frac{\tilde{\mathbb{P}}_{\varepsilon}}{\tilde{\mathbb{P}}_{\varepsilon}(\Omega)} \xrightarrow{*} \frac{1}{Z_0} (\det \text{Hess}^{\perp} E)^{-1/2} \rho \mathbf{1}_{\mathcal{K}} \mathcal{H}_{|\Sigma}^d,$$

which proves the claim. $\square$



The previous figure provides a schematic illustration of the different neighborhoods involved in the proof. The compact set $\mathcal{K}$ is contained in the intermediate neighborhood ω , which is itself contained in Ω . The contact set Σ may extend beyond $\mathcal{K}$, and the tubular neighborhoods considered in the proof are therefore constructed around the larger portion $\Sigma \cap \omega$.

5.1.2 First-order entropic selection

Theorem 5.3 (Compactly supported Euclidean marginals). *Assume the hypotheses $(H1')$ – $(H4')$ are satisfied and define*

$$\mathcal{F}^{(0)}(\gamma) \stackrel{\text{def.}}{=} \int_{\Omega} c d\gamma + \chi_{\Pi(\mu_1, \mu_2)}(\gamma), \quad \mathcal{F}_{\varepsilon}(\gamma) = \int_{\Omega} c d\gamma + \varepsilon \text{Ent}(\gamma | \mu_1 \otimes \mu_2) + \chi_{\Pi(\mu_1, \mu_2)}(\gamma).$$

Then, with $\mathbb{P}_0$ and Z_0 given by [Theorem 5.1](#),

$$\mathcal{F}_\varepsilon - \frac{k}{2}\varepsilon \log(1/\varepsilon) \stackrel{\Gamma}{=} \mathcal{F}^{(0)} + \varepsilon \left(\text{Ent}(\cdot | \mathbb{P}_0) + \chi_{\Pi(\mu_1, \mu_2)}(\cdot) - \log Z_0 - \frac{k}{2} \log(2\pi) \right) + o(\varepsilon). \quad (5.3)$$

Moreover, the following asymptotics of the energy holds

$$\min \mathcal{F}_\varepsilon = \min \mathcal{F}^{(0)} + \frac{k}{2}\varepsilon \log(1/\varepsilon) + \varepsilon C_0 + o(\varepsilon), \quad (5.4)$$

where $C_0 \stackrel{\text{def.}}{=} \min_{\gamma \in \Pi(\mu_1, \mu_2)} \text{Ent}(\gamma | \mathbb{P}_0) - \log Z_0 - \frac{k}{2} \log(2\pi)$, and the minimizer of $\mathcal{F}_\varepsilon$ converges weakly to the unique minimizer of $\text{Ent}(\cdot | \mathbb{P}_0)$ on $\Pi(\mu_1, \mu_2)$.

Proof. We follow the proof of [Theorem A](#) given in [Section 4](#), and only indicate the changes. By compactness of $\Sigma \cap \bar{\omega}$, the tubular coordinates, the normal ellipticity and the quadratic detachment hold on $\Sigma \cap \omega$ with uniform constants (see [Section 5.1.1](#)), and [Theorem 5.1](#) replaces [Theorem 3.4](#). In particular, the Γ -lim inf inequality (Step 1 of [Section 4.3](#)) is unchanged, and so is Step 3, since optimal plans are concentrated on $\Sigma \cap \mathcal{K}$, where the density of $\mathbb{P}_0$ with respect to $\mathcal{H}_{|\Sigma}^d$ is bounded from below.

It remains to adapt the Γ -lim sup inequality (Step 2 of [Section 4.3](#)). We fix $\bar{\gamma} \in \Pi(\mu_1, \mu_2)$ with $\text{Ent}(\bar{\gamma} | \mathbb{P}_0) < \infty$, which is concentrated on $\Sigma \cap \mathcal{K} \subset \Sigma \cap \omega$, and $R \geq 1$, and we define the normal perturbation $\tilde{\gamma}_{\varepsilon, R}$, its marginals $\tilde{\mu}_{i, \varepsilon, R}$ and its shadow $\gamma_{\varepsilon, R}$ as in [Sections 4.1](#) and [4.2](#).

Step 1. The marginals. The marginal estimate of [Theorem 4.4](#) still holds: its proof only uses inverse charts of $\text{proj}_{M_i}^\Sigma$, which exist on $\Sigma \cap \omega$ by [Proposition 2.2](#) and finitely many of which cover the compact set $\Sigma \cap \mathcal{K}$, and the integrability of the chart densities of $\bar{\gamma}$. The energy estimate of [Proposition 4.5](#) still holds as well: since Ω is convex, the segments used in its proof stay in $\text{int } \Omega$, and (4.12) holds with Σ replaced by $\Sigma \cap \bar{\omega}$. Indeed, a nearest point $\bar{y}$ of $y \in \text{int } \Omega$ in the compact set $\Sigma \cap \bar{\omega}$ lies in $\text{int } \Omega$, so that $\nabla E(\bar{y}) = 0$, and the segment $[\bar{y}, y]$ lies in $\text{int } \Omega$, whence $|\nabla E(y)| \leq C_E \text{dist}(y, \Sigma \cap \bar{\omega})$ with $C_E \stackrel{\text{def.}}{=} \sup_{\text{int } \Omega} \|D^2 E\| < \infty$ by [\(H2'\)](#). Hence, for fixed R ,

$$W_2(\tilde{\mu}_{i, \varepsilon, R}, \mu_i) = o_R(\sqrt{\varepsilon}) \quad (i = 1, 2) \quad \text{and} \quad \frac{1}{\varepsilon} \int E \, d(\gamma_{\varepsilon, R} - \tilde{\gamma}_{\varepsilon, R}) = o_R(1).$$

Step 2. The entropy. The entropy upper bound is the point to adapt since, as we wrote it, $\text{Ent}(\tilde{\gamma}_{\varepsilon, R} | \mathbb{P}_\varepsilon) = +\infty$ in this case: $\tilde{\gamma}_{\varepsilon, R}$ charges $\Omega \setminus \mathcal{K}$, where $\mathbb{P}_\varepsilon$ vanishes. We therefore consider the entropy with respect to the modified marginals:

$$I_{\varepsilon, R} \stackrel{\text{def.}}{=} \text{Ent}(\tilde{\gamma}_{\varepsilon, R} | \tilde{\mu}_{1, \varepsilon, R} \otimes \tilde{\mu}_{2, \varepsilon, R}).$$

Since $d\mathbb{P}_\varepsilon = Z_\varepsilon^{-1} e^{-E/\varepsilon} d(\mu_1 \otimes \mu_2)$, the shadow inequality ([Lemma 4.2](#)) and Step 1 give, for fixed R ,

$$\begin{aligned} \text{Ent}(\gamma_{\varepsilon, R} | \mathbb{P}_\varepsilon) &= \text{Ent}(\gamma_{\varepsilon, R} | \mu_1 \otimes \mu_2) + \log Z_\varepsilon + \frac{1}{\varepsilon} \int E \, d\gamma_{\varepsilon, R} \\ &\leq I_{\varepsilon, R} + \log Z_\varepsilon + \frac{1}{\varepsilon} \int E \, d(\gamma_{\varepsilon, R} - \tilde{\gamma}_{\varepsilon, R}) + \frac{1}{\varepsilon} \int E \, d\tilde{\gamma}_{\varepsilon, R} \\ &\leq I_{\varepsilon, R} + \frac{1}{\varepsilon} \int E \, d\tilde{\gamma}_{\varepsilon, R} + \log Z_\varepsilon + o_R(1). \end{aligned}$$

It remains to show that

$$\limsup_{\varepsilon \rightarrow 0} \left[I_{\varepsilon,R} + \frac{1}{\varepsilon} \int E \, d\tilde{\gamma}_{\varepsilon,R} + \log Z_{\varepsilon} \right] \leq \text{Ent}(\bar{\gamma} | \mathbb{P}_0) + b_R(\bar{\gamma}). \quad (5.5)$$

Since $|\Phi(z, \sqrt{\varepsilon}s) - z| \leq \sqrt{\varepsilon}R$ for $s \in \mathbb{B}_R$ and $\bar{\gamma}$ is concentrated on $\mathcal{K}$, for $\sqrt{\varepsilon}R \leq 1$ the measures $\tilde{\gamma}_{\varepsilon,R}$ and $\tilde{\mu}_{i,\varepsilon,R}$ are concentrated on the fixed compact sets $\mathcal{K}^1 \stackrel{\text{def.}}{=} \{\mathbf{x} : \text{dist}(\mathbf{x}, \mathcal{K}) \leq 1\}$ and $K_i^1 \stackrel{\text{def.}}{=} \{x : \text{dist}(x, K_i) \leq 1\}$. For an absolutely continuous probability measure ν concentrated on a bounded set, we denote by $\text{Ent}(\nu) \stackrel{\text{def.}}{=} \text{Ent}(\nu | \mathcal{L})$ its entropy with respect to the Lebesgue measure. Since $\tilde{\mu}_{1,\varepsilon,R} \otimes \tilde{\mu}_{2,\varepsilon,R}$ has the product of the marginal densities as density, we have

$$I_{\varepsilon,R} = \text{Ent}(\tilde{\gamma}_{\varepsilon,R}) - \text{Ent}(\tilde{\mu}_{1,\varepsilon,R}) - \text{Ent}(\tilde{\mu}_{2,\varepsilon,R}),$$

all these entropies being finite, since $\tilde{\gamma}_{\varepsilon,R}$ has finite entropy and all the measures are concentrated on fixed bounded sets. Hence

$$I_{\varepsilon,R} + \frac{1}{\varepsilon} \int E \, d\tilde{\gamma}_{\varepsilon,R} + \log Z_{\varepsilon} = \left(\text{Ent}(\tilde{\gamma}_{\varepsilon,R}) + \frac{1}{\varepsilon} \int E \, d\tilde{\gamma}_{\varepsilon,R} + \log Z_{\varepsilon} \right) - \sum_{i=1}^2 \text{Ent}(\tilde{\mu}_{i,\varepsilon,R}), \quad (5.6)$$

and we pass to the limit in the two terms of the right-hand side in the next two steps.

Step 3. The first term of (5.6). Let $\tilde{p}_{\varepsilon,R}$ be the density (4.7) of $\tilde{\gamma}_{\varepsilon,R}$, and recall that $K_R(z, s) = Z_R(z)^{-1} e^{-H_z E(s,s)/2}$ on $\mathbb{B}_R$. Writing $E(\Phi(z, \eta)) = \frac{1}{2} H_z E(\eta, \eta) + \mathcal{R}(z, \eta)$ with $|\mathcal{R}(z, \eta)| \leq C|\eta|^3$, the quadratic terms cancel, and, for $\mathcal{H}^d$ -a.e. $z \in \Sigma \cap \mathcal{K}$ and $s \in \mathbb{B}_R$,

$$\log \left(Z_{\varepsilon} e^{E/\varepsilon} \tilde{p}_{\varepsilon,R} \right) (\Phi(z, \sqrt{\varepsilon}s)) = \log \bar{\rho}(z) + \log \frac{Z_{\varepsilon}}{\varepsilon^{k/2} Z_R(z)} + \frac{\mathcal{R}(z, \sqrt{\varepsilon}s)}{\varepsilon} - \log \mathcal{J}_{\Phi}(z, \sqrt{\varepsilon}s).$$

We integrate this identity against $K_R(z, s) \, ds \, d\bar{\gamma}(z)$, which gives the first term of (5.6) by (4.6). The integral of the last two terms is bounded by $C\sqrt{\varepsilon}(M_1 + M_3) \rightarrow 0$, by (3.4) and (4.5). Moreover, since $Z_{\varepsilon} = (2\pi\varepsilon)^{k/2}(Z_0 + o(1))$ by Theorem 5.1, we have, uniformly in z , with Z_{∞} as in Lemma 4.1,

$$\log \frac{Z_{\varepsilon}}{\varepsilon^{k/2} Z_R(z)} = \log \left(Z_0 \sqrt{\det H_z E} \right) + \log \frac{Z_{\infty}(z)}{Z_R(z)} + o(1).$$

Since $\int_{\Sigma} |\log \bar{\rho}| \, d\bar{\gamma} < \infty$ (because $\text{Ent}(\bar{\gamma} | \mathbb{P}_0) < \infty$) and $\mathbb{P}_0$ has density $Z_0^{-1} (\det H_z E)^{-1/2} \rho(z)$ with respect to $\mathcal{H}_{|\Sigma}^d$ on $\Sigma \cap \mathcal{K}$, we obtain

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \left(\text{Ent}(\tilde{\gamma}_{\varepsilon,R}) + \frac{1}{\varepsilon} \int E \, d\tilde{\gamma}_{\varepsilon,R} + \log Z_{\varepsilon} \right) &= \int_{\Sigma} \log \frac{Z_0 \bar{\rho} \sqrt{\det H_z E}}{\rho} \, d\bar{\gamma} + \int_{\Sigma} \log \rho \, d\bar{\gamma} + b_R(\bar{\gamma}) \\ &= \text{Ent}(\bar{\gamma} | \mathbb{P}_0) + \text{Ent}(\mu_1) + \text{Ent}(\mu_2) + b_R(\bar{\gamma}), \end{aligned}$$

where we used that $\bar{\gamma}$ has marginals μ_1, μ_2 and $\rho = \rho_1 \otimes \rho_2$, so that $\int \log \rho \, d\bar{\gamma} = \text{Ent}(\mu_1) + \text{Ent}(\mu_2)$.

Step 4. The marginal entropies in (5.6). By (4.8), $\tilde{\mu}_{i,\varepsilon,R} \rightarrow \mu_i$ as $\varepsilon \rightarrow 0$, and all these measures are concentrated on the compact set K_i^1 . Since $\text{Ent}(\cdot) = \text{Ent}(\cdot | \mathcal{L}_{|K_i^1})$ on such measures, and the relative entropy with respect to a finite measure is weakly lower semicontinuous, we get

$$\liminf_{\varepsilon \rightarrow 0} \text{Ent}(\tilde{\mu}_{i,\varepsilon,R}) \geq \text{Ent}(\mu_i), \quad i = 1, 2.$$

Conclusion. Plugging Steps 3 and 4 into (5.6) gives (5.5), hence

$$\limsup_{\varepsilon \rightarrow 0} \text{Ent}(\gamma_{\varepsilon,R} | \mathbb{P}_\varepsilon) \leq \text{Ent}(\bar{\gamma} | \mathbb{P}_0) + b_R(\bar{\gamma}).$$

The diagonal argument of Step 2 of Section 4.3 then provides a recovery sequence, and the remaining conclusions follow as in the proof of Theorem A. $\square$

The purpose of the following proposition is to show that, in the quadratic case, we recover the known asymptotic expansion of EOT (see, for instance, Pal [Pal24]).

Proposition 5.4. *Assume (H1')–(H4') hold. If moreover $c(x_1, x_2) = \frac{1}{2}|x_1 - x_2|^2$, then, denoting by γ_ε the minimizer of $\mathcal{F}_\varepsilon$, by $\bar{\gamma}$ the optimal plan, and by $\text{Ent}(\mu_i) \stackrel{\text{def.}}{=} \text{Ent}(\mu_i | \mathcal{L}^d) = \int \rho_i \log \rho_i dx_i$,*

$$\mathcal{F}_\varepsilon(\gamma_\varepsilon) = \mathcal{F}^{(0)}(\bar{\gamma}) + \frac{d}{2}\varepsilon \log(1/\varepsilon) - \frac{\varepsilon}{2} \left(\text{Ent}(\mu_1) + \text{Ent}(\mu_2) + d \log(2\pi) \right) + o(\varepsilon).$$

Proof. By Brenier's theorem, in the quadratic case the optimal plan $\bar{\gamma}$ is unique and given by $\bar{\gamma} = (\text{id}, T)_\# \mu_1$, where T is the Brenier map between μ_1 and μ_2 . Consequently, $\bar{\gamma}$ is supported on the graph $\{(x_1, T(x_1)) | x_1 \in \text{spt } \mu_1\} \subseteq \Sigma$ and the tangent space of Σ is

$$T_{(x_1, T(x_1))} \Sigma = \text{Img} \begin{pmatrix} \text{Id} \\ DT(x_1) \end{pmatrix}.$$

Hence, the Hessian of the duality gap is given by

$$\begin{aligned} \text{Hess } E(x_1, T(x_1)) &= \begin{pmatrix} \text{Id} - D^2 \bar{\varphi}_1(x_1) & -\text{Id} \\ -\text{Id} & \text{Id} - D^2 \bar{\varphi}_2(T(x_1)) \end{pmatrix} \\ &= \begin{pmatrix} DT(x_1) & -\text{Id} \\ -\text{Id} & DT(x_1)^{-1} \end{pmatrix}, \end{aligned}$$

where we used that $T = \text{Id} - \nabla \bar{\varphi}_1$ (from $\nabla_{x_1} E = 0$ on Σ) and that $\text{Id} - D^2 \bar{\varphi}_2(T(x_1)) = DT(x_1)^{-1}$, obtained by differentiating the identity $x_1 = T(x_1) - \nabla \bar{\varphi}_2(T(x_1))$ (from $\nabla_{x_2} E = 0$ on Σ). We define $n(x_1)$, the matrix whose columns form an orthonormal basis of the normal space $N_{(x_1, T(x_1))} \Sigma$, as

$$n(x_1) \stackrel{\text{def.}}{=} \begin{pmatrix} -DT(x_1) \\ \text{Id} \end{pmatrix} (\text{Id} + DT(x_1)^2)^{-1/2}.$$

The reduced Hessian of E in the normal directions is therefore

$$\begin{aligned}
& \text{Hess}^\perp E(x_1, T(x_1)) \\
&= n(x_1)^\top \text{Hess} E(x_1, T(x_1)) n(x_1) \\
&= (DT(x_1)^2 + \text{Id})^{-1/2} (DT(x_1)^3 + 2DT(x_1) + DT(x_1)^{-1}) (DT(x_1)^2 + \text{Id})^{-1/2} \\
&= DT(x_1) + DT(x_1)^{-1}
\end{aligned}$$

For μ_1 -almost every x_1 , with $x_2 = T(x_1)$ the Monge-Ampère equation gives

$$\det DT(x_1) = \frac{\rho_1(x_1)}{\rho_2(T(x_1))} = \frac{\rho_1(x_1)}{\rho_2(x_2)}$$

and therefore

$$\det (\text{Hess}^\perp E(x_1, x_2))^{-1/2} = \left(\frac{\rho_1(x_1)}{\rho_2(x_2)} \right)^{1/2} \det (\text{Id} + DT(x_1)^2)^{-1/2}.$$

Consequently, writing $\bar{\rho}$ for the density of $\bar{\gamma}$ with respect to $\mathcal{H}_{|\Sigma}^d$, we obtain

$$\begin{aligned}
\text{Ent}(\bar{\gamma} | \mathbb{P}_0) &= \int_{\Sigma} \log \left(\frac{\bar{\rho}(x_1, x_2)}{Z_0^{-1} \det (\text{Id} + DT(x_1)^2)^{-1/2} \rho_1(x_1)^{3/2} \rho_2(x_2)^{1/2}} \right) d\bar{\gamma}(x_1, x_2) \\
&= \text{Ent}(\bar{\gamma} | \mathcal{H}_{|\Sigma}^d) - \frac{3}{2} \text{Ent}(\mu_1) - \frac{1}{2} \text{Ent}(\mu_2) \\
&\quad + \frac{1}{2} \int \log \det (\text{Id} + DT(x_1)^2) d\mu_1(x_1) + \log Z_0.
\end{aligned}$$

We now compute $\text{Ent}(\bar{\gamma} | \mathcal{H}_{|\Sigma}^d)$ using the graph structure of Σ . Since Σ is parametrized by $x_1 \mapsto (x_1, T(x_1))$, whose differential $\begin{pmatrix} \text{Id} \\ DT(x_1) \end{pmatrix}$ has Gram matrix $\text{Id} + DT(x_1)^2$, the area formula gives

$$d\mathcal{H}_{|\Sigma}^d = \det (\text{Id} + DT(x_1)^2)^{1/2} dx_1.$$

Hence

$$\bar{\rho}(x_1, T(x_1)) = \rho_1(x_1) \det (\text{Id} + DT(x_1)^2)^{-1/2}.$$

Therefore

$$\begin{aligned}
\text{Ent}(\bar{\gamma} | \mathcal{H}_{|\Sigma}^d) &= \int \log \bar{\rho} d\bar{\gamma} = \int \log (\rho_1(x_1) \det (\text{Id} + DT(x_1)^2)^{-1/2}) d\mu_1(x_1) \\
&= \text{Ent}(\mu_1) - \frac{1}{2} \int \log \det (\text{Id} + DT(x_1)^2) d\mu_1(x_1).
\end{aligned}$$

Plugging this into the previous identity, the $\log \det$ terms cancel and we get

$$\begin{aligned}
\text{Ent}(\bar{\gamma} | \mathbb{P}_0) - \log Z_0 &= \text{Ent}(\mu_1) - \frac{3}{2} \text{Ent}(\mu_1) - \frac{1}{2} \text{Ent}(\mu_2) \\
&= -\frac{1}{2} (\text{Ent}(\mu_1) + \text{Ent}(\mu_2)).
\end{aligned}$$

Finally, we conclude by [Theorem 5.3](#). □

Remark 5.1. Our result is consistent with the one of Pal [Pal24], up to the choice of the reference measure. Pal regularizes the transport cost with $\text{Ent}(\gamma)$, that is, with the entropy relative to the Lebesgue measure, and obtains the limit $+\frac{1}{2}(\text{Ent}(\mu_1) + \text{Ent}(\mu_2))$. Since the marginals of γ are fixed,

$$\text{Ent}(\gamma | \mu_1 \otimes \mu_2) = \text{Ent}(\gamma) - \text{Ent}(\mu_1) - \text{Ent}(\mu_2),$$

so that, up to normalizing constants, the two results differ exactly by

$$-\text{Ent}(\mu_1) - \text{Ent}(\mu_2),$$

which gives the value obtained above.

5.1.3 The McCann–Pass–Warren cost

We illustrate [Theorem 5.3](#) with the non-twisted cost of [MPW12, Example 3.1]. In this paragraph, the points of $\mathbb{R}^2 \times \mathbb{R}^2$ are denoted by (x_1, x_2) , with $x_1 = (u_1, v_1)$ and $x_2 = (u_2, v_2)$, and

$$c(x_1, x_2) \stackrel{\text{def.}}{=} e^{u_1+u_2} \cos(v_1 - v_2) + \frac{e^{2u_1}}{2} + \frac{e^{2u_2}}{2} \geq \frac{1}{2}(e^{u_1} - e^{u_2})^2 \geq 0.$$

Equality holds if and only if $u_1 = u_2$ and $\cos(v_1 - v_2) = -1$, that is, if and only if (x_1, x_2) belongs to one of the planes

$$\Sigma_j \stackrel{\text{def.}}{=} \{(x_1, T_j(x_1)) : x_1 \in \mathbb{R}^2\}, \quad T_j(x_1) \stackrel{\text{def.}}{=} (u_1, v_1 + (2j+1)\pi), \quad j \in \mathbb{Z}.$$

Since $D_{x_1 x_2}^2 c(x_1, x_2) = e^{u_1+u_2} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$ with $\theta = v_1 - v_2$, the cost is non-degenerate everywhere, but it is not twisted, since infinitely many planes Σ_j lie above each x_1 .

Marginals and optimal plans. Let $\mu_1 = \mu_2$ be the uniform probability measure on $A \stackrel{\text{def.}}{=} [0, 1] \times [0, 4\pi]$, that is, $\rho_1 = \rho_2 = \frac{1}{4\pi} \mathbf{1}_A$. The map

$$S(x_1) \stackrel{\text{def.}}{=} \begin{cases} (u_1, v_1 + \pi) & \text{if } v_1 \in [0, \pi) \cup [2\pi, 3\pi), \\ (u_1, v_1 - \pi) & \text{if } v_1 \in [\pi, 2\pi) \cup [3\pi, 4\pi], \end{cases}$$

preserves μ_1 and satisfies $S(x_1) \in \{T_0(x_1), T_{-1}(x_1)\}$, so that the plan $(\text{Id}, S)_{\#} \mu_1$ belongs to $\Pi(\mu_1, \mu_2)$ and is concentrated on $\{c = 0\}$. Hence it is optimal, $\text{OT}(\mu_1, \mu_2) = 0$, and $\bar{\varphi}_1 = \bar{\varphi}_2 = 0$ are Kantorovich potentials, the dual value being 0. We take for $M_1 = M_2$ a closed ball containing a neighborhood of $K_1 = K_2 \stackrel{\text{def.}}{=} A$. Then $E = c$ and $\Sigma = \Omega \cap \bigcup_j \Sigma_j$, so that every element of $\Pi(\mu_1, \mu_2)$ concentrated on $\bigcup_j \Sigma_j$ is optimal. The optimal plan is not unique: for instance, the map $x_1 \mapsto (u_1, v_1 + \pi)$ if $v_1 < 3\pi$, $x_1 \mapsto (u_1, v_1 - 3\pi)$ otherwise, also preserves μ_1 and induces another optimal plan. Indeed, above each $x_1 \in A$ lie exactly two points $T_j(x_1) \in A$ (for almost every x_1), and the optimal plans are the couplings which split the mass of x_1 between them.

Assumptions. Assumptions (H1')–(H4') hold: c is smooth and non-degenerate, $E = c$, Σ is a union of affine planes, ρ_i is constant on $K_i = A$, and $\mathcal{H}^2(\Sigma \cap \partial\mathcal{K}) = 0$, since $\Sigma \cap \partial\mathcal{K}$ is contained in a finite union of lines. Hence Theorem 5.3 applies.

The limit Gibbs measure. At $z = (x_1, x_2) \in \Sigma$, we have $u_1 = u_2$ and $v_1 - v_2 \in \pi + 2\pi\mathbb{Z}$, whence $D_{x_1 x_1}^2 E(z) = D_{x_2 x_2}^2 E(z) = e^{2u_1} I_2$ and $D_{x_1 x_2}^2 E(z) = -e^{2u_1} I_2$. Denoting by (e_1, e_2) the canonical basis of $\mathbb{R}^2$, the vectors $(e_1, -e_1)/\sqrt{2}$ and $(e_2, -e_2)/\sqrt{2}$ form an orthonormal basis of $N_z \Sigma$, in which $H_z E = 2e^{2u_1} I_2$, so that $\det H_z E = 4e^{4u_1}$. Since $x_1 \mapsto (x_1, T_j(x_1))$ has Jacobian 2, we have $\mathcal{H}_{|\Sigma_j}^2 = (\text{Id}, T_j)_\#(2\mathcal{L}^2)$. Moreover, for every $a \in [0, 1]$ and a.e. $b \in \mathbb{R}$,

$$\sum_{j \in \mathbb{Z}} \rho_1(a, b + (2j + 1)\pi) = \sum_{j \in \mathbb{Z}} \rho_1(a, b - (2j + 1)\pi) = \frac{1}{2\pi}, \quad (5.7)$$

since the interval $[0, 4\pi]$ has length 4π , and thus contains exactly two points of $b + \pi + 2\pi\mathbb{Z}$ for a.e. b . Therefore

$$Z_0 = \sum_{j \in \mathbb{Z}} \int_{\mathbb{R}^2} \frac{e^{-2u_1}}{2} \rho_1(x_1) \rho_2(T_j(x_1)) 2 \, dx_1 = \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{-2u_1} \rho_1(x_1) \, dx_1 = \frac{1 - e^{-2}}{4\pi},$$

and

$$d\mathbb{P}_0(x_1, x_2) = \frac{2\pi}{1 - e^{-2}} e^{-2u_1} \rho_1(x_1) \rho_2(x_2) d\mathcal{H}_{|\Sigma}^2(x_1, x_2).$$

By (5.7), the marginals of $\mathbb{P}_0$ have densities $\frac{2}{1 - e^{-2}} e^{-2u_i} \rho_i(x_i)$, $i = 1, 2$, so that $\mathbb{P}_0 \notin \Pi(\mu_1, \mu_2)$.

The selected plan.

Proposition 5.5. *As $\varepsilon \rightarrow 0$, the entropic optimal plans converge weakly to*

$$d\gamma_0(x_1, x_2) \stackrel{\text{def.}}{=} \pi \rho_1(x_1) \rho_2(x_2) d\mathcal{H}_{|\Sigma}^2(x_1, x_2), \quad \text{that is,} \quad \gamma_0 = \frac{1}{16\pi} \mathcal{H}_{|\Sigma \cap (A \times A)}^2.$$

Proof. By (5.7) and $\mathcal{H}_{|\Sigma_j}^2 = (\text{Id}, T_j)_\#(2\mathcal{L}^2)$, the marginals of γ_0 have densities

$$\sum_{j \in \mathbb{Z}} 2\pi \rho_1(x_1) \rho_2(T_j(x_1)) = \rho_1(x_1) \quad \text{and} \quad \sum_{j \in \mathbb{Z}} 2\pi \rho_1(T_j^{-1}(x_2)) \rho_2(x_2) = \rho_2(x_2),$$

so that $\gamma_0 \in \Pi(\mu_1, \mu_2)$. Moreover, γ_0 and $\mathbb{P}_0$ are equivalent, with

$$\frac{d\gamma_0}{d\mathbb{P}_0}(x_1, x_2) = g(u_1) \stackrel{\text{def.}}{=} \frac{1 - e^{-2}}{2} e^{2u_1},$$

which only depends on x_1 . Hence, for every $\gamma \in \Pi(\mu_1, \mu_2)$ such that $\text{Ent}(\gamma | \mathbb{P}_0) < \infty$,

$$\begin{aligned} \text{Ent}(\gamma | \mathbb{P}_0) &= \text{Ent}(\gamma | \gamma_0) + \int \log g(u_1) d\gamma(x_1, x_2) = \text{Ent}(\gamma | \gamma_0) + \int \log g(u_1) d\mu_1(x_1) \\ &= \text{Ent}(\gamma | \gamma_0) + \text{Ent}(\gamma_0 | \mathbb{P}_0). \end{aligned}$$

Thus γ_0 is the unique minimizer of $\text{Ent}(\cdot | \mathbb{P}_0)$ on $\Pi(\mu_1, \mu_2)$, and the conclusion follows from Theorem 5.3. $\square$

Since $d\mathbb{P}_0/d\mathcal{H}_{|\Sigma}^2$ is the product of a function of x_1 and a function of x_2 , γ_0 is equivalently the unique minimizer of $\text{Ent}(\cdot|\mathcal{H}_{|\Sigma}^2)$ on $\Pi(\mu_1, \mu_2)$. In other words, the entropic selection splits the mass of each $x_1 \in A$ equally between the two points $T_j(x_1) \in A$, whereas the limit Gibbs measure $\mathbb{P}_0$ is not a coupling of μ_1 and μ_2 .

5.2 The extension to multi-marginal framework

Let $N \geq 2$ and, for $i = 1, \dots, N$, let $M_i = \mathbb{T}^d$ and $\mu_i = \rho_i \text{vol}_{M_i}$, where ρ_i is Lipschitz and $\alpha \leq \rho_i \leq \beta$. We set $\Omega \stackrel{\text{def.}}{=} \prod_{i=1}^N M_i$, $\rho(z) \stackrel{\text{def.}}{=} \prod_{i=1}^N \rho_i(z_i)$ and $\mu^\otimes \stackrel{\text{def.}}{=} \bigotimes_{i=1}^N \mu_i = \rho \text{vol}_\Omega$, and we denote by $\Pi(\mu_1, \dots, \mu_N)$ the set of probability measures on Ω with marginals $\mu_1, \dots, \mu_N$. The cost c is continuous, and we fix continuous Kantorovich potentials $\bar{\varphi}_1, \dots, \bar{\varphi}_N$ such that the duality gap $E \stackrel{\text{def.}}{=} c - \sum_{i=1}^N \bar{\varphi}_i(x_i) \geq 0$ belongs to $C^{2,1}(\Omega)$, and such that $\Sigma \stackrel{\text{def.}}{=} \{E = 0\}$ is a $C^{1,1}$ submanifold of Ω without boundary, of dimension d , satisfying

$$\text{rank Hess } E(z) = (N-1)d \quad \text{for every } z \in \Sigma. \quad (5.8)$$

The Gibbs measures $\mathbb{P}_\varepsilon$, the constants Z_ε and the functionals $\mathcal{F}_\varepsilon$, $\mathcal{F}^{(0)}$, $\mathcal{F}^{(1)}$ are defined as in [Sections 1 and 2](#), with $\Pi(\mu_1, \dots, \mu_N)$ in place of $\Pi(\mu_1, \mu_2)$, and with the codimension $k \stackrel{\text{def.}}{=} (N-1)d$ of Σ in place of d .

For two marginals, (5.8) follows from the non-degeneracy of c ([Proposition 2.2](#)). For $N \geq 3$, it has to be imposed: since $T_z \Sigma \subseteq \ker \text{Hess } E(z)$ and $\dim T_z \Sigma = d$, it amounts to $\ker \text{Hess } E(z) = T_z \Sigma$, that is, to the positive definiteness of $H_z E$. Without it, some normal directions would be degenerate and would contribute differently to the asymptotics.

Theorem 5.6 (Multi-marginal Γ -expansion). *Under these assumptions, the conclusions of [Theorem 3.4](#) and of [Theorem A](#) hold, with k in place of d in (1.1)–(1.2), and with*

$$d\mathbb{P}_0(z) = \frac{1}{Z_0} (\det H_z E)^{-1/2} \rho(z) d\mathcal{H}_{|\Sigma}^d(z), \quad Z_0 \stackrel{\text{def.}}{=} \int_{\Sigma} (\det H_z E)^{-1/2} \rho(z) d\mathcal{H}^d(z).$$

Proof. The proofs of [Sections 3 and 4](#) and of [Section 5.1.1](#) use the two-marginal structure at three places only.

(i) **Normal ellipticity.** In [Proposition 2.2](#), (2.3) is deduced from the non-degeneracy of c . Here, it follows from (5.8), the positive semidefiniteness of $\text{Hess } E$ on Σ , and the compactness of Σ .

(ii) **Local normal frames.** A $C^{1,1}$ submanifold admits local Lipschitz orthonormal frames of its normal bundle; see the Gram–Schmidt construction in [[Min15](#), Section 2.3.1, arXiv version]. We choose a finite cover $(U_a)_a$ of Σ by relatively compact patches with such frames n_a , and a subordinate Lipschitz partition of unity $(\chi_a)_a$. For $r > 0$ sufficiently small, the maps

$$\Phi_a(z, s) \stackrel{\text{def.}}{=} z + n_a(z)s, \quad z \in U_a, \quad s \in \mathbb{B}_r,$$

are bi-Lipschitz tubular parametrizations, by the positive-reach property used in [Section 3.1](#). Since the frames are Lipschitz, the proof of [Lemma 3.2](#) applies on each patch,

and (3.3) holds almost everywhere, with uniform bounds over the finite cover. The area formula then gives

$$\int_{\Sigma_r} h \, d \operatorname{vol}_\Omega = \sum_a \int_{U_a} \chi_a(z) \int_{\mathbb{B}_r} h(\Phi_a(z, s)) \mathcal{J}_{\Phi_a}(z, s) \, ds \, d\mathcal{H}^d(z)$$

for every nonnegative Borel function h . We apply the three-region Laplace argument to each integral in this sum. The weights χ_a depend only on z , and the resulting limits sum to the desired integral because $\sum_a \chi_a = 1$.

On overlaps, $n_b(z) = n_a(z) O_{ab}(z)$ with $O_{ab}(z) \in O(k)$. The coordinate change $(z, s) \mapsto (z, O_{ab}(z)s)$ has absolute Jacobian one almost everywhere, even though O_{ab} depends on z . It also preserves the normal Gaussian measure and $\mathbb{B}_R$, that is

$$H_b = O_{ab}^\top H_a O_{ab}, \quad K_b(z, s) = K_a(z, O_{ab}s),$$

where $H_a(z)$ is the matrix of the normal Hessian in the local orthonormal frame $n_a(z)$. Thus the local normal perturbations agree on overlaps and define the recovery measure in Section 4.1. In the Euclidean case, we use a finite cover of the compact portion $\Sigma \cap \bar{\omega}$ appearing in the Laplace proof.

With (i) and (ii), the proofs of Theorem 3.4 (Section 5.1.1), of Lemma 4.1, Proposition 4.3, Proposition 4.5, and of Steps 1 and 3 of Section 4.3 apply verbatim, with sums over $i \in \{1, \dots, N\}$; recall that Step 3 only uses that $\operatorname{proj}_{M_1}$ is 1-Lipschitz.

(iii) Marginal estimate. The proof of Theorem 4.4 uses local inverses of $\operatorname{proj}_{M_i}^\Sigma$, which, for $N \geq 3$, may have critical points (vertical branches). Let $\bar{\gamma}$ be the plan fixed in Step 2 of Section 4.3. We fix i and $R \geq 1$, and write $p \stackrel{\text{def.}}{=} \operatorname{proj}_{M_i}^\Sigma$, $\mu \stackrel{\text{def.}}{=} \mu_i$ and $\tilde{\mu}_\varepsilon \stackrel{\text{def.}}{=} \tilde{\mu}_{i, \varepsilon, R}$. Let $C \stackrel{\text{def.}}{=} \{z \in \Sigma : \operatorname{rank} Dp(z) < d\}$. By Sard's theorem, $p(C)$ is Lebesgue-negligible, so that $\bar{\gamma}(C) \leq \mu(p(C)) = 0$. Given $\eta > 0$, we choose an open set $V \supset p(C)$ such that $\mu(V) < \eta$, and we split $\bar{\gamma} = \bar{\gamma}^b + \bar{\gamma}^g$, with $\bar{\gamma}^b \stackrel{\text{def.}}{=} \bar{\gamma} \llcorner p^{-1}(V)$. Accordingly, $\mu = \mu^b + \mu^g$ and $\tilde{\mu}_\varepsilon = \tilde{\mu}_\varepsilon^b + \tilde{\mu}_\varepsilon^g$, where μ^b , μ^g and $\tilde{\mu}_\varepsilon^b$, $\tilde{\mu}_\varepsilon^g$ are the i -th marginals of $\bar{\gamma}^b$, $\bar{\gamma}^g$ and of their normal perturbations, defined by (4.6). On the compact set $\Sigma \setminus p^{-1}(V)$, p is a local diffeomorphism, and the proof of Theorem 4.4, which only uses that the chart densities are integrable, gives $W_2^2(\mu^g, \tilde{\mu}_\varepsilon^g) = o_\eta(\varepsilon)$, where W_2 is extended to measures of equal mass. On the other hand, since $|p(\Phi(z, \sqrt{\varepsilon}s)) - p(z)| \leq \sqrt{\varepsilon}|s|$, the map $(z, s) \mapsto (p(z), p(\Phi(z, \sqrt{\varepsilon}s)))$ pushes $K_R(z, s) \, ds \, d\bar{\gamma}^b(z)$ forward to a coupling of μ^b and $\tilde{\mu}_\varepsilon^b$ with quadratic cost at most

$$\varepsilon \int_{\Sigma} \int_{\mathbb{B}_R} |s|^2 K_R(z, s) \, ds \, d\bar{\gamma}^b(z) \leq M_2 \varepsilon \bar{\gamma}^b(\Sigma) = M_2 \varepsilon \mu(V) \leq M_2 \varepsilon \eta,$$

by (4.5). Gluing the two couplings, we get $W_2^2(\mu, \tilde{\mu}_\varepsilon) \leq o_\eta(\varepsilon) + M_2 \varepsilon \eta$. Letting $\varepsilon \rightarrow 0$, then $\eta \rightarrow 0$, gives $W_2(\mu, \tilde{\mu}_\varepsilon) = o_R(\sqrt{\varepsilon})$, which completes the proof. $\square$

Combining this proof with Section 5.1, Theorem 5.6 extends to N compactly supported marginals on $\mathbb{R}^d$, under the analogue of (H1')–(H4') in which $\mathcal{K} \stackrel{\text{def.}}{=} \prod_{i=1}^N K_i$, and the non-degeneracy of c is replaced by (5.8) on $\Sigma \cap \operatorname{int} \Omega$.

5.2.1 Multi-marginal optimal transport with repulsive harmonic cost

We apply [Theorem 5.6](#) to the special case of three marginals on the circle. Let $N = 3$, $d = 1$, and $M_i = \mathbb{S}^1 \subset \mathbb{R}^2$ for $i = 1, 2, 3$. We use the uniform probability $\mu = (2\pi)^{-1} \mathcal{H}_{|\mathbb{S}^1}^1$ and set $\mu_i = \mu$. Here d is the intrinsic dimension of each marginal space. The contact set computed below is a disjoint union of two circles, so $\dim \Sigma = 1 = d$ and its codimension in $(\mathbb{S}^1)^3$ is $k = 2$. We consider the harmonic repulsive cost

$$c(\mathbf{x}) \stackrel{\text{def.}}{=} - \sum_{1 \leq i < j \leq N} |x_i - x_j|^2.$$

The associated multi-marginal optimal transport problem is

$$\inf_{\gamma \in \Pi(\mu_i)} \int_{(\mathbb{S}^1)^N} c(\mathbf{x}) \, d\gamma(\mathbf{x}). \quad (5.9)$$

Proposition 5.7. *Problem (5.9) is equivalent to*

$$\inf_{\gamma \in \Pi(\mu_1, \mu_2, \mu_3)} \int_{(\mathbb{S}^1)^N} \left| \sum_{i=1}^N x_i \right|^2 d\gamma(\mathbf{x}). \quad (5.10)$$

Proof. We start by expanding the cost in the following way:

$$\begin{aligned} \sum_{1 \leq i < j \leq N} |x_i - x_j|^2 &= \sum_{1 \leq i < j \leq N} (|x_i|^2 + |x_j|^2 - 2\langle x_i, x_j \rangle) \\ &= (N-1) \sum_{i=1}^N |x_i|^2 - 2 \sum_{1 \leq i < j \leq N} \langle x_i, x_j \rangle. \end{aligned}$$

On the other hand, $|\sum_{i=1}^N x_i|^2 = \sum_{i=1}^N |x_i|^2 + 2 \sum_{1 \leq i < j \leq N} \langle x_i, x_j \rangle$. Therefore, $\sum_{1 \leq i < j \leq N} |x_i - x_j|^2 = N \sum_{i=1}^N |x_i|^2 - |\sum_{i=1}^N x_i|^2$. Since $x_i \in \mathbb{S}^1$, one has $|x_i| = 1$, and hence $\sum_{i=1}^N |x_i|^2 = N$. Consequently, $c(\mathbf{x}) = |\sum_{i=1}^N x_i|^2 - N^2$. Thus, for every $\gamma \in \Pi(\mu_i)$,

$$\int_{(\mathbb{S}^1)^N} c(\mathbf{x}) \, d\gamma(\mathbf{x}) = \int_{(\mathbb{S}^1)^N} \left| \sum_{i=1}^N x_i \right|^2 d\gamma(\mathbf{x}) - N^2.$$

Since N^2 is independent of γ , minimizing the original cost is equivalent to minimizing

$$\int_{(\mathbb{S}^1)^N} \left| \sum_{i=1}^N x_i \right|^2 d\gamma(\mathbf{x}).$$

□

We denote by $\bar{c}$ the cost of (5.10).

Geometry of the contact set The dual problem associated with (5.10) is

$$\sup_{(\varphi_i)_{i=1}^3} \left\{ \sum_{i=1}^3 \int_{\mathbb{S}^1} \varphi_i(x_i) d\mu_i(x_i) \mid \sum_{i=1}^3 \varphi_i(x_i) \leq \bar{c}(\mathbf{x}) \quad \forall \mathbf{x} \in (\mathbb{S}^1)^3 \right\}. \quad (5.11)$$

Proposition 5.8. *A family of optimal Kantorovich potentials for the dual problem (5.11) is given by*

$$\bar{\varphi}_1 = \bar{\varphi}_2 = \bar{\varphi}_3 = 0.$$

Proof. There exists a coupling $\gamma \in \Pi(\mu_1, \mu_2, \mu_3)$ concentrated on $\{\mathbf{x} \in (\mathbb{S}^1)^3 \mid x_1 + x_2 + x_3 = 0\}$, for instance $\bar{\gamma} = (F)_{\#}\mu_1$, with $F(x) = (x, R_{2\pi/3}x, R_{-2\pi/3}x)$. One can easily check that $\bar{\gamma} \in \Pi(\mu_1, \mu_2, \mu_3)$ and $\bar{c} = 0$ on its support. Thus the value of (5.10) is null and therefore the admissible dual family $(\bar{\varphi}_1, \bar{\varphi}_2, \bar{\varphi}_3) = (0, 0, 0)$ attains the primal value and is a family of optimal Kantorovich potentials. $\square$

Proposition 5.9. *The contact set is a compact manifold without boundary*

$$\Sigma = \{(x, R_{2\pi/3}x, R_{-2\pi/3}x) \mid x \in \mathbb{S}^1\} \sqcup \{(x, R_{-2\pi/3}x, R_{2\pi/3}x) \mid x \in \mathbb{S}^1\},$$

where R_θ denotes the rotation of angle θ .

Proof. By definition, $\Sigma = \{E = 0\}$. Here $E = \bar{c}$ which is equivalent to $x_1 = -x_2 - x_3$. Moreover, we have $|x_2 + x_3|^2 = |x_2|^2 + |x_3|^2 + 2\langle x_2, x_3 \rangle = 2 + 2\langle x_2, x_3 \rangle$, and thus, using $|x_2 + x_3| = |x_1| = 1$, $\langle x_2, x_3 \rangle = -\frac{1}{2}$. It implies that $x_2 = R_{\pm 2\pi/3}x_3$ and thus

$$\begin{aligned} \Sigma &= \{(x, R_{2\pi/3}x, R_{-2\pi/3}x) \mid x \in \mathbb{S}^1\} \sqcup \{(x, R_{-2\pi/3}x, R_{2\pi/3}x) \mid x \in \mathbb{S}^1\} \\ &\stackrel{\text{def.}}{=} \Sigma_+ \sqcup \Sigma_-. \end{aligned}$$

Finally, the maps $F_{\pm} : x \in \mathbb{S}^1 \mapsto F_{\pm}(x) = (x, R_{\pm 2\pi/3}x, R_{\mp 2\pi/3}x) \in (\mathbb{S}^1)^3$ are smooth embeddings. Therefore Σ_+ and Σ_- are diffeomorphic to $\mathbb{S}^1$. Since $\mathbb{S}^1$ is compact and without boundary, Σ is a compact manifold without boundary. $\square$

Entropic selection principle

Proposition 5.10. *The limiting Gibbs measure has the explicit expression*

$$\mathbb{P}_0 = \frac{1}{2}(\gamma_+ + \gamma_-),$$

where $\gamma_{\pm} = (F_{\pm})_{\#}\mu$ and $F_{\pm}$ are the two parametrizations above. Moreover, $\mathbb{P}_0$ is the selected transport plan.

Proof. We verify the normal-rank hypothesis and apply the Gibbs formula of **Theorem 5.6**:

$$\mathbb{P}_0 = \frac{1}{Z_0} (\det \text{Hess}^{\perp} E)^{-1/2} \prod_{i=1}^N \rho_i \mathcal{H}_{|\Sigma}^1.$$

We compute the normal Hessian and the normalization. For $i = 1, 2, 3$, introduce the geodesics

$$\nu_i(\tau) = \exp_{x_i}(\tau v_i),$$

where $v_i \in T_{x_i} \mathbb{S}^1$. We also introduce $S(\tau) \stackrel{\text{def.}}{=} \nu_1(\tau) + \nu_2(\tau) + \nu_3(\tau)$, then we have

$$\frac{d}{d\tau} E(\nu_1(\tau), \nu_2(\tau), \nu_3(\tau)) = 2\langle \dot{S}(\tau), S(\tau) \rangle,$$

and

$$\frac{d^2}{d\tau^2} E(\nu_1(\tau), \nu_2(\tau), \nu_3(\tau)) = 2\langle \ddot{S}(\tau), S(\tau) \rangle + 2|\dot{S}(\tau)|^2.$$

Thus, for $\tau = 0$, since $S(0) = 0$, we obtain

$$\left. \frac{d^2}{d\tau^2} E(\nu_1(\tau), \nu_2(\tau), \nu_3(\tau)) \right|_{\tau=0} = 2|v_1 + v_2 + v_3|^2,$$

and denoting $V = (v_1, v_2, v_3)$ we have

$$\text{Hess } E(\mathbf{x})[V, V] = 2|v_1 + v_2 + v_3|^2.$$

Let $e_i \stackrel{\text{def.}}{=} Jx_i$, where

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

so that e_i is a unit tangent vector to $\mathbb{S}^1$ at x_i . Then

$$E_1 = (e_1, 0, 0), \quad E_2 = (0, e_2, 0), \quad E_3 = (0, 0, e_3)$$

is an orthonormal basis of $T_{\mathbf{x}}(\mathbb{S}^1)^3$. We obtain

$$\text{Hess } E(\mathbf{x})[E_i, E_i] = 2, \quad \text{Hess } E(\mathbf{x})[E_i, E_j] = 2\langle e_i, e_j \rangle = 2\langle x_i, x_j \rangle = -1,$$

for $i \neq j$, since $\mathbf{x} \in \Sigma$. Hence

$$[\text{Hess } E(\mathbf{x})] = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

Moreover,

$$T_{\mathbf{x}}\Sigma = \text{span}\{E_1 + E_2 + E_3\},$$

and therefore

$$N_{\mathbf{x}}\Sigma = \{aE_1 + bE_2 + cE_3 \mid a + b + c = 0\}.$$

On this normal space, the above matrix acts as 3Id . Thus

$$\text{Hess}^\perp E(\mathbf{x}) = 3\text{Id}_{N_{\mathbf{x}}\Sigma},$$

Thus $\text{rank Hess } E = 2 = (N - 1)d$, as required by [Theorem 5.6](#), and $\det \text{Hess}^\perp E = 9$. The induced arclength must also be included in the normalization: the maps $F_\pm$ have speed $\sqrt{3}$ when the unit circle is parametrized by arclength. Hence

$$\mathcal{H}^1(\Sigma_\pm) = 2\pi\sqrt{3}, \quad Z_0 = (2\pi)^{-3} \frac{4\pi\sqrt{3}}{3} = \frac{1}{2\sqrt{3}\pi^2}.$$

Consequently,

$$\mathbb{P}_0 = \frac{1}{4\pi\sqrt{3}} \mathcal{H}^1_{|\Sigma} = \frac{1}{2}(\gamma_+ + \gamma_-).$$

Since $\text{Ent}(\cdot | \mathbb{P}_0)$ is nonnegative and $\mathbb{P}_0 \in \Pi(\mu_i)$, we conclude that the selected transport plan is $\mathbb{P}_0$. $\square$

5.2.2 Repulsive harmonic cost in $\mathbb{R}^d$

Let $N \geq 3$, $d \geq 1$, and consider the repulsive harmonic cost

$$c(\mathbf{x}) \stackrel{\text{def.}}{=} - \sum_{1 \leq i < j \leq N} |x_i - x_j|^2, \quad \mathbf{x} = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N.$$

For $i = 1, \dots, N$, let $\mu_i = \rho_i dx$, where ρ_i vanishes outside a compact set $K_i \subset \mathbb{R}^d$ with $\mathcal{L}^d(\partial K_i) = 0$, and is bounded and continuous on K_i . We take for M_i a closed ball containing a neighborhood of K_i , and we set $\Omega \stackrel{\text{def.}}{=} \prod_i M_i$, $\rho(\mathbf{x}) \stackrel{\text{def.}}{=} \prod_i \rho_i(x_i)$ and $b \stackrel{\text{def.}}{=} \sum_i \int x d\mu_i(x)$. We assume:

- (M) There exists $i_0 \in \{1, \dots, N\}$ such that μ_{i_0} is the law of $b - \sum_{i \neq i_0} X_i$, where the random variables X_i , $i \neq i_0$, are independent with laws μ_i .

Under (M), the law γ_M of $(X_1, \dots, X_N)$, where $X_{i_0} \stackrel{\text{def.}}{=} b - \sum_{i \neq i_0} X_i$, belongs to $\Pi(\mu_1, \dots, \mu_N)$ and is concentrated on the affine subspace

$$\Sigma_b \stackrel{\text{def.}}{=} \left\{ \mathbf{x} \in (\mathbb{R}^d)^N : \sum_{i=1}^N x_i = b \right\},$$

of dimension $(N - 1)d$ and codimension $k = d$. Since

$$c(\mathbf{x}) = \left| \sum_{i=1}^N x_i - b \right|^2 + \sum_{i=1}^N \bar{\varphi}_i(x_i), \quad \bar{\varphi}_i(x) \stackrel{\text{def.}}{=} -N|x|^2 + 2\langle b, x \rangle - \frac{|b|^2}{N},$$

the functions $\bar{\varphi}_i$ are Kantorovich potentials, γ_M is optimal, the duality gap is $E(\mathbf{x}) = |\sum_i x_i - b|^2$, and $\Sigma = \Sigma_b \cap \Omega$. The normal space to Σ_b is spanned by the orthonormal vectors $(e_p, \dots, e_p)/\sqrt{N}$, $p = 1, \dots, d$, so that

$$\Phi(z, s) \stackrel{\text{def.}}{=} z + \frac{(s, \dots, s)}{\sqrt{N}}, \quad z \in \Sigma_b, \quad s \in \mathbb{R}^d,$$

defines global normal coordinates: Φ is an isometry from $\Sigma_b \times \mathbb{R}^d$ onto $(\mathbb{R}^d)^N$, $\mathcal{J}_\Phi \equiv 1$, $\text{dist}(\Phi(z, s), \Sigma_b) = |s|$ and $E(\Phi(z, s)) = N|s|^2$. In particular, $H_z E = 2N I_d$, and the remainder $\mathcal{R}$ of the Taylor expansion of E vanishes.

Theorem 5.11 (Repulsive harmonic cost in $\mathbb{R}^d$). *Assume (M). Then the conclusions of Theorem 3.4 and of Theorem A hold, with $\Pi(\mu_1, \dots, \mu_N)$ in place of $\Pi(\mu_1, \mu_2)$, $k = d$, and*

$$\mathbb{P}_0 \stackrel{\text{def.}}{=} \frac{\rho \mathcal{H}_{|\Sigma_b}^{(N-1)d}}{\int_{\Sigma_b} \rho d\mathcal{H}^{(N-1)d}}, \quad Z_0 \stackrel{\text{def.}}{=} (2N)^{-d/2} \int_{\Sigma_b} \rho d\mathcal{H}^{(N-1)d}.$$

Moreover, the selected plan is γ_M , and

$$\min \mathcal{F}_\varepsilon = \text{OT}(\mu_1, \dots, \mu_N) + \frac{d}{2} \varepsilon \log(1/\varepsilon) - \varepsilon \left(\text{Ent}(\mu_{i_0}) + \frac{d}{2} \log \pi \right) + o(\varepsilon),$$

where $\text{Ent}(\mu_{i_0}) \stackrel{\text{def.}}{=} \int \rho_{i_0} \log \rho_{i_0} dx$.

Proof. We follow the proofs of Theorems 5.1 and 5.3, with Σ_b in place of the d -dimensional contact set: these proofs do not use the dimension of the contact set, and they are simpler here, since the normal coordinates are global and E is exactly quadratic in them.

Step 1. Gibbs measures and Γ -lim inf. The proof of Theorem 5.1 applies with $\mathcal{H}^{(N-1)d}$ in place of $\mathcal{H}^d$, $\mathcal{R} \equiv 0$, $\mathcal{J}_\Phi \equiv 1$ and $\mathcal{T}_r = \Sigma_r$. Its assumption $\mathcal{H}^{(N-1)d}(\Sigma_b \cap \partial\mathcal{K}) = 0$, where $\mathcal{K} = \prod_i K_i$, follows from $\mathcal{L}^d(\partial K_i) = 0$: for $j \neq i$, parametrizing Σ_b by the coordinates $(x_l)_{l \neq j}$, the set $\Sigma_b \cap \{x_i \in \partial K_i\}$ is the image of a negligible set by a Lipschitz map. Moreover, $Z_0 > 0$, since $\rho > 0$ γ_M -a.e. and $\gamma_M \ll \mathcal{H}_{|\Sigma_b}^{(N-1)d}$ (see Step 2). Since $\det H_z E = (2N)^d$ is constant, this gives $\mathbb{P}_\varepsilon \rightarrow \mathbb{P}_0$ and $Z_\varepsilon = (2\pi\varepsilon)^{d/2}(Z_0 + o(1))$, and the Γ -lim inf inequality follows as in Step 1 of Section 4.3.

Step 2. A finite-entropy optimal plan. Parametrizing Σ_b by $(x_i)_{i \neq i_0}$, whose surface Jacobian is $N^{d/2}$, we get $\gamma_M = N^{-d/2} \prod_{i \neq i_0} \rho_i(x_i) \mathcal{H}_{|\Sigma_b}^{(N-1)d}$. Hence

$$\frac{d\gamma_M}{d\mathbb{P}_0}(\mathbf{x}) = \frac{C_M}{\rho_{i_0}(x_{i_0})}, \quad C_M \stackrel{\text{def.}}{=} N^{-d/2} \int_{\Sigma_b} \rho d\mathcal{H}^{(N-1)d}, \quad (5.12)$$

and $\text{Ent}(\gamma_M | \mathbb{P}_0) = \log C_M - \text{Ent}(\mu_{i_0}) < \infty$, since ρ_{i_0} is bounded with compact support. This replaces Step 3 of Section 4.3, where the bound $\bar{\gamma} \leq \beta \mathcal{H}_{|\Sigma}^d$ used $\dim \Sigma = d$.

Step 3. The marginals. Fix $\bar{\gamma} \in \Pi(\mu_1, \dots, \mu_N)$ with $\text{Ent}(\bar{\gamma} | \mathbb{P}_0) < \infty$, and $R \geq 1$. Since $H_z E = 2N I_d$ does not depend on z , neither does the kernel $K_R(z, s) = K_R(s) \propto e^{-N|s|^2} \mathbf{1}_{\mathbb{B}_R}(s)$. Hence, by (4.6), $\tilde{\gamma}_{\varepsilon, R}$ is the law of $Z + \sqrt{\varepsilon/N} (S, \dots, S)$, where $Z \sim \bar{\gamma}$ and $S \sim K_R$ are independent, and its i -th marginal is

$$\tilde{\mu}_{i, \varepsilon, R} = \mu_i * \eta_\varepsilon, \quad \eta_\varepsilon \stackrel{\text{def.}}{=} \text{law of } \sqrt{\varepsilon/N} S.$$

Writing $\sqrt{\varepsilon/N} S = \sqrt{\varepsilon/(2N^2)} S'$, where $S' = \sqrt{2N} S$ has density proportional to $e^{-|s'|^2/2}$ on $\mathbb{B}_{\sqrt{2N}R}$, we are in the setting of Proposition 4.6, with $\varepsilon/(2N^2)$ and $\sqrt{2N}R$ in place of ε and R . Its proof applies verbatim in $\mathbb{R}^d$, and gives $W_2(\tilde{\mu}_{i, \varepsilon, R}, \mu_i) = o_R(\sqrt{\varepsilon})$.

Step 4. Energy and entropy. The rest of the proof of Theorem 5.3 is unchanged. The energy estimate of Proposition 4.5 holds, since $|\nabla E(\mathbf{y})| = 2N \text{dist}(\mathbf{y}, \Sigma_b)$ for every $\mathbf{y} \in (\mathbb{R}^d)^N$. In Step 3 of the proof of Theorem 5.3, the remainder and Jacobian

terms vanish, and $\int |\log \bar{\rho}| d\bar{\gamma} < \infty$ follows from $\text{Ent}(\bar{\gamma} | \mathbb{P}_0) < \infty$ and $\int |\log \rho| d\bar{\gamma} \leq \sum_i \int \rho_i |\log \rho_i| dx < \infty$.

Step 5. Selection and value. By (5.12), $d\gamma_M/d\mathbb{P}_0$ only depends on x_{i_0} . Hence, for every $\gamma \in \Pi(\mu_1, \dots, \mu_N)$ with $\text{Ent}(\gamma | \mathbb{P}_0) < \infty$, as in the proof of Proposition 5.5,

$$\text{Ent}(\gamma | \mathbb{P}_0) = \text{Ent}(\gamma | \gamma_M) + \int \log \frac{C_M}{\rho_{i_0}} d\mu_{i_0} = \text{Ent}(\gamma | \gamma_M) + \text{Ent}(\gamma_M | \mathbb{P}_0),$$

so that γ_M is the unique minimizer of $\text{Ent}(\cdot | \mathbb{P}_0)$ on $\Pi(\mu_1, \dots, \mu_N)$, that is, the selected plan. Finally, since $C_M = 2^{d/2} Z_0$,

$$\min \mathcal{F}^{(1)} = \text{Ent}(\gamma_M | \mathbb{P}_0) - \log Z_0 - \frac{d}{2} \log(2\pi) = -\text{Ent}(\mu_{i_0}) - \frac{d}{2} \log \pi,$$

and the expansion of $\min \mathcal{F}_\varepsilon$ follows from Theorem A. $\square$

Remark 5.2. Condition (M) only involves the marginals: it means that μ_{i_0} is the image of $\mu_1 * \dots * \mu_{i_0-1} * \mu_{i_0+1} * \dots * \mu_N$ by $x \mapsto b - x$; in particular, ρ_{i_0} is then automatically continuous. Since a convex combination of plans concentrated on Σ_b with bounded densities is again such a plan, all the conclusions of Theorem 5.11, except the explicit selected plan and value, remain valid when $(\mu_1, \dots, \mu_N)$ is a convex combination of N -tuples of marginals satisfying (M) with the same b . For $N = 2$, (M) means that μ_2 is the image of μ_1 by $x \mapsto b - x$, and the cost is equivalent to $|x_1 + x_2|^2$: the expansion above is then consistent with Proposition 5.4, after the change of variables $x_2 \mapsto -x_2$ and the rescaling $\varepsilon \mapsto \varepsilon/2$.

Acknowledgments. L.N., P.P. and L.T. acknowledge the support of the FMJH Program PGMO and the ANR project GOTA (ANR-23-CE46-0001). B.P. is pleased to acknowledge the support of Natural Sciences and Engineering Research Council of Canada Discovery Grant number 04864-2024. The authors used AI-assisted tools to help check proofs and ensure consistency of notation, and take full responsibility for the final manuscript.

References

- [Ada+11] S. Adams et al. “From a large-deviations principle to the Wasserstein gradient flow: A new micro-macro passage”. In: *Communications in Mathematical Physics* 307.3 (2011), pp. 791–815. DOI: [10.1007/s00220-011-1328-4](https://doi.org/10.1007/s00220-011-1328-4).
- [AG25] S. Aryan and P. Ghosal. “Entropic Selection Principle for Monge’s Optimal Transport”. In: *arXiv preprint arXiv:2502.16370* (2025).
- [ANR17] J. Altschuler, J. Niles-Weed, and P. Rigollet. “Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration”. In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.
- [Ben+15] J.-D. Benamou et al. “Iterative Bregman projections for regularized transportation problems”. In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138.

- [BGN22] E. Bernton, P. Ghosal, and M. Nutz. “Entropic optimal transport: Geometry and large deviations”. In: *Duke Mathematical Journal* 171.16 (2022), pp. 3363–3400.
- [BT08] A. Braides and L. Truskinovsky. “Asymptotic expansions by Γ -convergence”. In: *Continuum Mechanics and Thermodynamics* 20.1 (2008), pp. 21–62.
- [Car+17] G. Carlier et al. “Convergence of entropic schemes for optimal transport and gradient flows”. In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418.
- [CDS25] M. Colombo, S. Di Marino, and F. Stra. “First order expansion in the semiclassical limit of the Levy–Lieb functional”. In: *Archive for Rational Mechanics and Analysis* 249.6 (2025), pp. 1–57.
- [CPT23] G. Carlier, P. Pegon, and L. Tamanini. “Convergence rate of general entropic optimal transport costs”. In: *Calculus of Variations and Partial Differential Equations* 62.4 (2023), p. 116.
- [CS94] R. Cominetti and J. San Martín. “Asymptotic analysis of the exponential penalty trajectory in linear programming”. In: *Mathematical Programming* 67 (1994), pp. 169–187.
- [CT21] G. Conforti and L. Tamanini. “A formula for the time derivative of the entropic cost and applications”. In: *Journal of Functional Analysis* 280.11 (2021), p. 108964.
- [Cut13] M. Cuturi. “Sinkhorn distances: Lightspeed computation of optimal transport”. In: *Advances in neural information processing systems* 26 (2013).
- [Dal93] G. Dal Maso. *An Introduction to Γ -Convergence*. Vol. 8. Progress in Nonlinear Differential Equations and their Applications. Boston: Birkhäuser, 1993.
- [DG20] S. Di Marino and A. Gerolin. “An Optimal Transport Approach for the Schrödinger Bridge Problem and Convergence of Sinkhorn Algorithm”. In: *Journal of Scientific Computing* 85.2 (Oct. 19, 2020), p. 27. ISSN: 1573-7691. DOI: [10.1007/s10915-020-01325-7](https://doi.org/10.1007/s10915-020-01325-7).
- [DL18] S. Di Marino and J. Louet. “The entropic regularization of the Monge problem on the real line”. In: *SIAM Journal on Mathematical Analysis* 50.4 (2018), pp. 3451–3477.
- [DLR13] M. H. Duong, V. Laschos, and M. Renger. “Wasserstein gradient flows from large deviations of many-particle limits”. In: *ESAIM: Control, Optimisation and Calculus of Variations* 19.4 (2013), pp. 1166–1188. DOI: [10.1051/cocv/2013049](https://doi.org/10.1051/cocv/2013049).
- [EMR15] M. Erbar, J. Maas, and M. Renger. “From large deviations to Wasserstein gradient flows in multiple dimensions”. In: *Electronic Communications in Probability* 20.89 (2015), pp. 1–12. DOI: [10.1214/ECP.v20-4315](https://doi.org/10.1214/ECP.v20-4315).

- [EN22] S. Eckstein and M. Nutz. “Quantitative stability of regularized optimal transport and convergence of sinkhorn’s algorithm”. In: *SIAM Journal on Mathematical Analysis* 54.6 (2022), pp. 5922–5948.
- [EN24] S. Eckstein and M. Nutz. “Convergence rates for regularized optimal transport via quantization”. In: *Mathematics of Operations Research* 49.2 (2024), pp. 1223–1240.
- [Fed59] H. Federer. “Curvature measures”. In: *Transactions of the American Mathematical Society* 93.3 (1959), pp. 418–491.
- [FL89] J. Franklin and J. Lorenz. “On the scaling of multidimensional matrices”. In: *Linear Algebra and its applications* 114–115 (1989), pp. 717–735. DOI: [10.1016/0024-3795\(89\)90490-4](https://doi.org/10.1016/0024-3795(89)90490-4).
- [Gal16] A. Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 2016. DOI: [10.23943/princeton/9780691172767.001.0001](https://doi.org/10.23943/princeton/9780691172767.001.0001).
- [Gen+19] A. Genevay et al. “Sample Complexity of Sinkhorn Divergences”. In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. Vol. 89. Proceedings of Machine Learning Research. PMLR, 2019, pp. 1574–1583.
- [GKR20] A. Gerolin, A. Kausamo, and T. Rajala. “Multi-marginal entropy-transport with repulsive cost”. In: *Calculus of Variations and Partial Differential Equations* 59.3 (2020), p. 90. DOI: [10.1007/s00526-020-01735-3](https://doi.org/10.1007/s00526-020-01735-3).
- [Hwa80] C.-R. Hwang. “Laplace’s method revisited: weak convergence of probability measures”. In: *The Annals of Probability* (1980), pp. 1177–1182.
- [Léo12] C. Léonard. “From the Schrödinger problem to the Monge–Kantorovich problem”. In: *Journal of Functional Analysis* 262.4 (2012), pp. 1879–1920.
- [Léo14] C. Léonard. “A Survey of the Schrödinger Problem and Some of Its Connections with Optimal Transport”. In: *Discrete and Continuous Dynamical Systems* 34.4 (2014), p. 1533. DOI: [10.3934/dcds.2014.34.1533](https://doi.org/10.3934/dcds.2014.34.1533).
- [Ley25] A. Ley. “Entropic selection for optimal transport on the line with distance cost”. In: *arXiv preprint arXiv:2512.05282* (2025).
- [LV23] F. Léger and F.-X. Vialard. “A geometric Laplace method”. In: *Pure and Applied Analysis* 5.4 (2023), pp. 1041–1080.
- [Mik04] T. Mikami. “Monge’s problem with a quadratic cost by the zero-noise limit of h -path processes”. In: *Probability Theory and Related Fields* 129.2 (2004), pp. 245–260. DOI: [10.1007/s00440-004-0340-4](https://doi.org/10.1007/s00440-004-0340-4).
- [Mik90] T. Mikami. “Variational Processes from the Weak Forward Equation”. In: *Communications in mathematical physics* 135.1 (1990), pp. 19–40.
- [Min15] E. Minguzzi. “Convex neighborhoods for Lipschitz connections and sprays”. In: *Monatshefte für Mathematik* 177.4 (2015), pp. 569–625. DOI: [10.1007/s00605-014-0699-y](https://doi.org/10.1007/s00605-014-0699-y). arXiv: [1308.6675](https://arxiv.org/abs/1308.6675) [math.DG].

- [MN19] G. Mena and J. Niles-Weed. “Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem”. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019.
- [MPW12] R. J. McCann, B. Pass, and M. Warren. “Rectifiability of optimal transportation plans”. In: *Canadian Journal of Mathematics* 64.4 (2012), pp. 924–934.
- [MS25] H. Malamut and M. Sylvestre. “Convergence rates of the regularized optimal transport: Disentangling suboptimality and entropy”. In: *SIAM Journal on Mathematical Analysis* 57.3 (2025), pp. 2533–2558.
- [NP25] L. Nenna and P. Pegon. “Convergence rate of entropy-regularized multi-marginal optimal transport costs”. In: *Canadian Journal of Mathematics* 77.3 (2025), pp. 1072–1092.
- [Nut21] M. Nutz. “Introduction to entropic optimal transport”. In: *Lecture notes, Columbia University* (2021).
- [NW22] M. Nutz and J. Wiesel. “Entropic optimal transport: Convergence of potentials”. In: *Probability Theory and Related Fields* 184.1 (2022), pp. 401–424.
- [NZ26] M. Nutz and C. Zhong. “Entropic regularization of Monge’s problem”. In: *arXiv preprint arXiv:2604.21578* (2026).
- [Pal24] S. Pal. “On the difference between entropic cost and the optimal transport cost”. In: *The Annals of Applied Probability* 34.1B (2024), pp. 1003–1028. DOI: [10.1214/23-AAP1983](https://doi.org/10.1214/23-AAP1983).
- [PC19] G. Peyré and M. Cuturi. *Computational optimal transport: With applications to data science*. Now Foundations and Trends, 2019.
- [San15] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Vol. 87. Progress in Nonlinear Differential Equations and Their Applications. Calculus of variations, PDEs, and modeling. Cham: Springer International Publishing, 2015. ISBN: 978-3-319-20828-2. DOI: [10.1007/978-3-319-20828-2](https://doi.org/10.1007/978-3-319-20828-2).
- [Sin64] R. Sinkhorn. “A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices”. In: *The Annals of Mathematical Statistics* 35.2 (June 1964), pp. 876–879. ISSN: 0003-4851, 2168-8990. DOI: [10.1214/aoms/1177703591](https://doi.org/10.1214/aoms/1177703591).
- [Sin67] R. Sinkhorn. “Diagonal Equivalence to Matrices with Prescribed Row and Column Sums”. In: *The American Mathematical Monthly* 74.4 (1967), pp. 402–405.
- [Vil08] C. Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2008.
- [Wee18] J. Weed. “An explicit analysis of the entropic penalty in linear programming”. In: *Proceedings of the 31st Conference On Learning Theory*. Vol. 75. Proceedings of Machine Learning Research. PMLR, 2018, pp. 1841–1855.