



# Convergence rate of entropy-regularized multi-marginal optimal transport costs

Luca Nenna, Paul Pegon

## ► To cite this version:

Luca Nenna, Paul Pegon. Convergence rate of entropy-regularized multi-marginal optimal transport costs. Canadian Journal of Mathematics = Journal Canadien de Mathématiques, 2024, 10.4153/S0008414X24000257 . hal-04154453v3

**HAL Id: hal-04154453**

**<https://hal.science/hal-04154453v3>**

Submitted on 9 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Convergence rate of entropy-regularized multi-marginal optimal transport costs

Luca Nenna and Paul Pegon

**Abstract.** We investigate the convergence rate of multi-marginal optimal transport costs that are regularized with the Boltzmann–Shannon entropy, as the noise parameter  $\varepsilon$  tends to 0. We establish lower and upper bounds on the difference with the unregularized cost of the form  $C\varepsilon \log(1/\varepsilon) + O(\varepsilon)$  for some explicit dimensional constants  $C$  depending on the marginals and on the ground cost, but not on the optimal transport plans themselves. Upper bounds are obtained for Lipschitz costs or locally semiconcave costs for a finer estimate, and lower bounds for  $\mathcal{C}^2$  costs satisfying some signature condition on the mixed second derivatives that may include degenerate costs, thus generalizing results previously in the two marginals case and for nondegenerate costs. We obtain in particular matching bounds in some typical situations where the optimal plan is deterministic.

## Notations

In all the article,  $N \in \mathbb{N}^*$  denotes the dimension of the ambient space  $\mathbb{R}^N$  and  $m \in \mathbb{N}$  is an integer such that  $m \geq 2$ .

$B_r^d(x)$  open Euclidean ball of radius  $r$  centered at  $x$  in  $\mathbb{R}^d$ , dropping the superscript  $d$  when  $d = N$ ;

$\omega_d$   $d$ -dimensional volume of  $B_1^d(0)$ ;

$X_i$  a subset of  $\mathbb{R}^N$  for any index  $i \in \{1, \dots, m\}$ ;

$\mathbf{X}$  product  $X_1 \times \dots \times X_m$  whenever the  $X_i$ 's are  $m$  subsets of  $\mathbb{R}^N$ ;

$A_{-i}$  product  $\prod_{1 \leq j \leq m, j \neq i} A_j$  if  $A = A_1 \times \dots \times A_m \subseteq \mathbf{X}$  and  $i \in \{1, \dots, m\}$ ;

$x_i, x, \mathbf{x}$  a point in  $X_i$ , in some  $X_j$ ,  $j \in \{1, \dots, m\}$ , and in  $\mathbf{X}$ , respectively;

$x_q$   $(x_i)_{i \in q}$  if  $q \subseteq \{1, \dots, m\}$  and  $\mathbf{x} \in \mathbf{X}$ ;

$e_i$   $i$ th coordinate map  $e_i : \mathbf{x} = (x_1, \dots, x_m) \mapsto x_i$ ;

$A \Subset B$   $A \subseteq K \subseteq B$  for some compact set  $K$ ;

Received by the editors September 1, 2023; revised February 13, 2024; accepted March 4, 2024.

Published online on Cambridge Core March 15, 2024.

L.N. is partially on academic leave at Inria (team Matherials) for the year 2022–2023 and acknowledges the hospitality of this institution during this period. His work was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH and from H-Code, Université Paris-Saclay. P.P. acknowledges the academic leave provided by Inria Paris (team MOKAPLAN) for the year 2022–2023. Both authors acknowledge the financial support by the ANR project GOTA (ANR-23-CE46-0001).

AMS subject classification: 49Q22, 49N15, 94A17, 49K40.

Keywords: Optimal transport, multi-marginal optimal transport, entropic regularization, Schrödinger problem, convex analysis.



- $|\cdot|$  Euclidean norm on  $\mathbb{R}^N$ ;
- $\|\cdot\|$  norm on  $\mathbb{R}^N \times \cdots \times \mathbb{R}^N$  defined by  $\|\mathbf{x}\| = \max_{1 \leq i \leq m} |x_i|$  if  $\mathbf{x} = (x_1, \dots, x_m)$ ;
- $B_r(\mathbf{x})$  open ball of radius  $r$  centered at  $\mathbf{x} \in (\mathbb{R}^N)^m$  for the above norm;
- $\mathcal{C}_{\text{loc}}^{0,1}(X)$  space of real-valued locally Lipschitz functions on  $X$  which is a sub-manifold of  $\mathbb{R}^N$  or  $(\mathbb{R}^N)^m$ ;
- $[f]_{\mathcal{C}^{0,1}(X)}$  Lipschitz constant of  $f : X \rightarrow \mathbb{R}$  where  $X$  is a subset of  $\mathbb{R}^N$  or  $(\mathbb{R}^N)^m$  for the above norms;
- $\mathcal{C}_{\text{loc}}^{1,1}(X)$  space of differentiable real-valued functions on  $X$ , a sub-manifold of  $\mathbb{R}^N$  or  $(\mathbb{R}^N)^m$ , with locally Lipschitz differential;
- $\mathcal{P}(X)$  space of probability measures on a metric space  $X$ ;
- $\|\cdot\|_{L^p(\mu)}$   $L^p$ -norm induced by a measure  $\mu$ , where  $p \in [1, +\infty]$ ;
- $\text{spt } \mu$  support of the measure  $\mu$ ;
- $\mu \llcorner A$  restriction of the Borel measure  $\mu$  to the Borel set  $A$  defined by  $\mu \llcorner A(E) = \mu(A \cap E)$  for every  $E$ ;
- $\mathcal{H}_X^s$   $s$ -dimensional Hausdorff measure on the metric space  $X$  endowed with the Borel  $\sigma$ -algebra (the subscript  $X$  will often be dropped);
- $M_N(\mathbb{R})$  space of real matrices of size  $N \times N$ , endowed with the Frobenius norm induced by the scalar product  $A \cdot B := \text{Tr}(A^T B)$ , for  $A, B \in M_N(\mathbb{R})$ ;
- $S_N(\mathbb{R})$  subspace of real symmetric matrices of size  $N \times N$ ;
- $\Delta_P$  simplex of  $P$ -uples  $t = (t_p)_{p \in P}$  such that  $\forall p, t_p \geq 0$  and  $\sum_{p \in P} t_p = 1$ .

## 1 Introduction

We consider an  $m$ -uple of probability measures  $\mu_i$  compactly supported on sub-manifolds  $X_i \subseteq \mathbb{R}^N$  of dimension  $d_i$  and a cost function  $c : X_1 \times \cdots \times X_m \rightarrow \mathbb{R}_+$ . The Entropic Multi-Marginal Optimal Transport problem is defined as

$$(\text{MOT}_\varepsilon) \quad \text{MOT}_\varepsilon := \inf \left\{ \int_{X_1 \times \cdots \times X_m} c d\gamma + \varepsilon \text{Ent}(\gamma | \otimes_{i=1}^m \mu_i) \mid \gamma \in \Pi(\mu_1, \dots, \mu_m) \right\},$$

where  $\Pi(\mu_1, \dots, \mu_m)$  denotes the set of all probability measures  $\gamma$  having  $\mu_i$  as  $i$ th marginal, i.e.,  $(e_i)_\# \gamma = \mu_i$ , where  $e_i : (x_1, \dots, x_m) \mapsto x_i$ , for every  $i \in \{1, \dots, m\}$ . The classical multi-marginal optimal transport problem corresponds to the case where  $\varepsilon = 0$ . In the last decade, these two classes of problems (entropic optimal transport (EOT) and multi-marginal optimal transport (MOT)) have witnessed a growing interest and they are now an active research topic.

EOT has found applications and proved to be an efficient way to approximate optimal transport (OT) problems, especially from a computational viewpoint. Indeed, when it comes to solving EOT by alternating Kullback–Leibler projections on the two marginal constraints, by the algebraic properties of the entropy, such iterative projections correspond to the celebrated Sinkhorn’s algorithm [Sin64], applied in this framework in the pioneering works [Ben+15, Cut13]. The simplicity and the good convergence guarantees (see [Car22, FL89, GN22, MG20]) of this method compared to the algorithms used for the OT problems, then determined the success of EOT for

applications in machine learning, statistics, image processing, language processing, and other areas (see the monograph [PC19] or the lecture notes [Nut] and the references therein).

As concerns MOT, it arises naturally in many different areas of applications, including economics [CE10], financial mathematics [BHP13, DS14a, DS14b, Enn+22], statistics [BK18, CCG16], image processing [Rab+11], tomography [Abr+17], machine learning [Haa+21, TJK22], fluid dynamics [Bre89], and quantum physics and chemistry, in the framework of density functional theory [BDG12, CFK13, FGG23]. The structure of solutions to the MOT problem is a notoriously delicate issue, and is still not well understood, despite substantial efforts on the part of many researchers [Car03, CDD15, CN08, CS16, GS98, Hei02, KP14, KP15, MP17, Pas11, Pas12, PV21a, PV21b] (see also the surveys [DGN17, Pas15]). Since  $\text{MOT}_\varepsilon$  can be seen a perturbation of  $\text{MOT}_0$ , it is natural to study the behavior as  $\varepsilon$  vanishes. In this paper, we are mainly interested in investigating the rate of convergence of the entropic cost  $\text{MOT}_\varepsilon$  to  $\text{MOT}_0$  under some mild assumptions on the cost functions and marginals.

In particular, we are going to extend the techniques introduced in [CPT23] for two marginals to the multi-marginal case which will also let us generalize the bounds in [CPT23] to the case of degenerate cost functions. For the two marginals and nondegenerate case, we also refer the reader to a very recent (and elegant) paper [MS23] where the authors push a little further the analysis of the convergence rate by disentangling the roles of  $\int c d\gamma$  and the relative entropy in the total cost and deriving convergence rate for both these terms. Notice that concerning the convergence rate of the entropic MOT, an upper bound has been already established in [EN23], which depends on the number of marginals and the quantization dimension of the optimal solutions to  $(\text{MOT}_\varepsilon)$  with  $\varepsilon = 0$ . Here we provide an improved, smaller, upper bound, which will depend only on the marginals, but not on the OT plans for the unregularized problem, and we also provide a lower bound depending on a signature condition on the mixed second derivatives of the cost function, that was introduced in [Pas12]. The main difficulty consists in adapting the estimates of [CPT23] to the local structure of the optimal plans described in [Pas12].

Our main findings can be summarized as follows: we establish two **upper bounds**, one valid for locally Lipschitz costs and a finer one valid for locally semiconcave costs. The proofs rely, as in [CPT23], on a multi-marginal variant of the block approximation introduced in [Car+17]. Notice that in this case the bound will depend only on the dimension of the support of the marginals. Moreover, for locally semiconcave cost functions, by exploiting Alexandrov-type results as in [CPT23], we improve the upper bound by a  $1/2$  factor, obtaining the following inequality for some  $C^* \in \mathbb{R}_+$ :

$$(1.1) \quad \boxed{\text{MOT}_\varepsilon \leq \text{MOT}_0 + \frac{1}{2} \left( \sum_{i=1}^m d_i - \max_{1 \leq i \leq m} d_i \right) \varepsilon \log(1/\varepsilon) + C^* \varepsilon.}$$

We stress that this upper bound is smaller than or equal to the one provided in [EN23, Theorem 3.8], which is of the form  $\frac{1}{2}(m-1)D\varepsilon \log(1/\varepsilon) + O(\varepsilon)$ , where  $D$  is a quantization dimension of the support of an OT plan. Thus,  $D$  must be greater than or equal to the maximum dimension of the support of the marginals, and of

course  $\sum_{i=1}^m d_i - \max_{1 \leq i \leq m} d_i \leq (m-1) \max_{1 \leq i \leq m} d_i$ . The inequality may be strict, for example, in the two marginals case with unequal dimension, as shown in Section 5.

For the **lower bound**, from the dual formulation of  $(\text{MOT}_\varepsilon)$ , we have

$$\text{MOT}_\varepsilon \geq \text{MOT}_0 - \varepsilon \log \int_{\prod_{i=1}^m X_i} e^{-\frac{E(x_1, \dots, x_m)}{\varepsilon}} d \otimes_{i=1}^m \mu_i(x_i),$$

where  $E(x_1, \dots, x_m) = c(x_1, \dots, x_m) - \oplus_{i=1}^m \phi_i(x_i)$  is the duality gap and  $(\phi_1, \dots, \phi_m)$  are Kantorovich potentials for the unregularized problem  $(\text{MOT}_\varepsilon)$  with  $\varepsilon = 0$ . By using the singular values decomposition of the bilinear form obtained as an average of mixed second derivatives of the cost and a signature condition introduced in [Pas11], we are able to prove that  $E$  detaches quadratically from the set  $\{E = 0\}$  and this allows us to estimate the previous integral in the desired way as in [CPT23] and improve the results in [EN23] where only an upper bound depending on the quantization dimension of the solution to the unregularized problem is provided. Moreover, this slightly more flexible use of Minty's trick compared to [CPT23] allows us to obtain a lower bound also for degenerate cost functions in the two marginals setting. Given a  $\kappa$  depending on a signature condition (see  $(\text{PS}(\kappa))$ ) on the second mixed derivatives of the cost, the lower bound can be summarized as follows:

$$(1.2) \quad \boxed{\text{MOT}_\varepsilon \geq \text{MOT}_0 + \frac{\kappa}{2} \varepsilon \log(1/\varepsilon) - C_* \varepsilon,}$$

for some  $C_* \in \mathbb{R}_+$ .

The paper is organized as follows: in Section 2, we recall the MOT problem, some results concerning the structure of the optimal solution, in particular the ones in [Pas11], and define its entropy regularization. Section 3 is devoted to the upper bounds stated in Theorems 3.1 and 3.5. In Section 4, we establish the lower bound stated in Proposition 4.2. Finally, in Section 5, we provide some examples for which we can get the matching bounds.

## 2 Preliminaries

Given  $m$  probability compactly supported measures  $\mu_i$  on sub-manifolds  $X_i$  of dimension  $d_i$  in  $\mathbb{R}^N$  for  $i \in \{1, \dots, m\}$  and a continuous cost function  $c : X_1 \times X_2 \times \dots \times X_m \rightarrow \mathbb{R}_+$ , the MOT problem consists in solving the following optimization problem:

$$(\text{MOT}) \quad \text{MOT}_0 := \inf_{\gamma \in \Pi(\mu_1, \dots, \mu_m)} \int_{\mathbf{X}} c(x_1, \dots, x_m) d\gamma,$$

where  $\mathbf{X} := X_1 \times X_2 \times \dots \times X_m$  and  $\Pi(\mu_1, \dots, \mu_m)$  denotes the set of probability measures on  $\mathbf{X}$  whose marginals are the  $\mu_i$ . The formulation above is also known as the *Kantorovich problem*, and it amounts to a linear minimization problem over a convex, weakly compact set; it is then not difficult to prove the existence of a solution by the direct method of calculus of variations. Much of the attention in the OT community is rather focused on uniqueness and the structure of the minimizers. In particular, one is mainly interested in determining if the solution is concentrated on the graph of a function  $(T_2, \dots, T_m)$  over the first marginal, where  $(T_i)_\# \mu_1 = \mu_i$  for  $i \in \{1, \dots, m\}$ , in which case this function induces a solution *à la Monge*, that is,  $\gamma = (\text{Id}, T_2, \dots, T_m)_\# \mu_1$ .

In the two marginals setting, the theory is fairly well understood and it is well known that under mild conditions on the cost function (e.g., twist condition) and marginals (e.g., being absolutely continuous with respect to Lebesgue), the solution to (MOT) is unique and is concentrated on the graph of a function; we refer the reader to [San15] to have glimpse of it. The extension to the multi-marginal case is still not well understood, but it has attracted recently a lot of attention due to a diverse variety of applications.

In particular, in his seminal works, Pass [Pas11, Pas12] established some conditions, more restrictive than in the two marginals case, to ensure the existence of a solution concentrated on a graph. In this work, we rely on the following (local) result in [Pas12] giving an upper bound on the dimension of the support of the solution to (MOT). Let  $P$  be the set of partitions of  $\{1, \dots, m\}$  into two nonempty disjoint subsets:  $p = \{p_-, p_+\} \in P$  if  $p_- \cup p_+ = \{1, \dots, m\}$ ,  $p_- \cap p_+ = \emptyset$  and  $p_-, p_+ \neq \emptyset$ . Then, for each  $p \in P$ , we denote by  $g_p$  the bilinear form on the tangent bundle  $TX$

$$g_p := D_{p_- p_+}^2 c + D_{p_+ p_-}^2 c \quad \text{where} \quad D_{pq}^2 c := \sum_{i \in p, j \in q} D_{x_i x_j}^2 c$$

for every  $p, q \subseteq \{1, \dots, m\}$ , and  $D_{x_i x_j}^2 c := \sum_{\alpha_i, \alpha_j} \frac{\partial^2 c}{\partial x_i^{\alpha_i} \partial x_j^{\alpha_j}} dx_i^{\alpha_i} \otimes dx_j^{\alpha_j}$ , defined for every  $i, j$  on the whole tangent bundle  $TX$ . Define

$$(2.1) \quad G_c := \left\{ \sum_{p \in P} t_p g_p \mid (t_p)_{p \in P} \in \Delta_P \right\}$$

to be the convex hull generated by the  $g_p$ , then it is easy to verify that each  $g \in G_c$  is symmetric and therefore its signature, denoted by  $(d^+(g), d^-(g), d^0(g))$ , is well defined. Then, the following result from [Pas12] gives a control on the dimension of the support of the optimizer(s) in terms of these signatures.

**Theorem 2.1** (Part of [Pas12, Theorem 2.3]) *Let  $\gamma$  a solution to (MOT) and suppose that the signature of some  $g \in G_c$  at a point  $\mathbf{x} \in X$  is  $(d^+, d^-, d^0)$ , that is, the number of positive, negative, and zero eigenvalues. Then, there exists a neighborhood  $N_{\mathbf{x}}$  of  $\mathbf{x}$  such that  $N_{\mathbf{x}} \cap \text{spt } \gamma$  is contained in a Lipschitz sub-manifold of  $X$  with dimension no greater than  $\sum_{i=1}^m d_i - d^+$ .*

**Remark 2.2** For the following, it is important to notice that by standard linear algebra arguments, we have for each  $g \in G_c$  that  $d^+(g) \leq \sum_{i=1}^m d_i - \max_i d_i$ . This implies that the smallest bound on the dimension of  $\text{spt } \gamma$  which Theorem 2.1 can provide is  $\max_i d_i$ .

**Remark 2.3** (Two marginals case) When  $m = 2$ , the only  $g \in G_c$  coincides precisely with the pseudo-metric introduced by Kim and McCann in [KM10]. Assuming for simplicity that  $d_1 = d_2 = d$ , they noted that  $g$  has signature  $(d, d, 0)$  whenever  $c$  is nondegenerate, so Theorem 2.1 generalizes their result since it applies even when non-degeneracy fails providing new information in the two marginals case: the signature of  $g$  is  $(r, r, 2d - 2r)$  where  $r$  is the rank of  $D_{x_1 x_2}^2 c$ . Notice that this will help us to generalize the results established in [CPT23, EN23] to the case of a degenerate cost function.

It is well known that under some mild assumptions, the Kantorovich problem (MOT) is dual to the following:

$$(MD) \quad \sup \left\{ \sum_{i=1}^m \int_{X_i} \phi_i(x_i) d\mu_i \mid \phi_i \in \mathcal{C}_b(X_i), \sum_{i=1}^m \phi_i(x^i) \leq c(x_1, \dots, x_m) \right\}.$$

Besides, it admits solutions  $(\phi_i)_{1 \leq i \leq m}$ , called *Kantorovich potentials*, when  $c$  is continuous and all the  $X_i$ 's are compact, and these solutions may be assumed *c-conjugate*, in the sense that for every  $i \in \{1, \dots, m\}$ ,

$$(2.2) \quad \forall x \in X_i, \quad \phi_i(x) = \inf_{(x_j)_{j \neq i} \in X_{-i}} c(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots) - \sum_{1 \leq j \leq m, j \neq i} \phi_j(x_j).$$

We recall the entropic counterpart of (MOT): given  $m$  probability measures  $\mu_i$  on  $X_i$  as before, and a continuous cost function  $c : X \rightarrow \mathbb{R}_+$ , the  $MOT_\varepsilon$  problem is

$$(MOT_\varepsilon) \quad MOT_\varepsilon = \inf \left\{ \int_{X_1 \times \dots \times X_m} c d\gamma + \varepsilon \text{Ent}(\gamma \mid \otimes_{i=1}^m \mu_i) \mid \gamma \in \Pi(\mu_1, \dots, \mu_m) \right\},$$

where  $\text{Ent}(\cdot \mid \otimes_{i=1}^m \mu_i)$  is the Boltzmann–Shannon relative entropy (or Kullback–Leibler divergence) w.r.t. the product measure  $\otimes_{i=1}^m \mu_i$ , defined for general probability measures  $p, q$  as

$$\text{Ent}(p \mid q) = \begin{cases} \int_{\mathbb{R}^d} \rho \log(\rho) dq, & \text{if } p = \rho q, \\ +\infty, & \text{otherwise.} \end{cases}$$

The fact that  $q$  is a probability measure ensures that  $\text{Ent}(p \mid q) \geq 0$ . The dual problem of  $(MOT_\varepsilon)$  reads as

$$(MD_\varepsilon) \quad MOT_\varepsilon = \varepsilon + \sup \left\{ \sum_{i=1}^m \int_{X_i} \phi_i(x_i) d\mu_i - \varepsilon \int_X e^{\frac{\sum_{i=1}^m \phi_i(x_i) - c(x)}{\varepsilon}} d \otimes_{i=1}^m \mu_i \mid \phi_i \in \mathcal{C}_b(X_i) \right\},$$

which is invariant by  $(\phi_1, \dots, \phi_m) \mapsto (\phi_1 + \lambda_1, \dots, \phi_m + \lambda_m)$  where  $(\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$  and  $\sum_{i=1}^m \lambda_i = 0$  (see [Léol4, MG20, NW22] for some recent presentations). It admits an equivalent “log-sum-exp” form:

$$(MD'_\varepsilon) \quad MOT_\varepsilon = \sup \left\{ \sum_{i=1}^m \int_{X_i} \phi_i(x_i) d\mu_i - \varepsilon \log \left( \int_X e^{\frac{\sum_{i=1}^m \phi_i(x_i) - c(x)}{\varepsilon}} d \otimes_{i=1}^m \mu_i \right) \mid \phi_i \in \mathcal{C}_b(X_i) \right\},$$

which is invariant by the same transformations without assuming  $\sum_{i=1}^m \lambda_i = 0$ .

From  $(MOT_\varepsilon)$  and  $(MD_\varepsilon)$ , we recover, as  $\varepsilon \rightarrow 0$ , the unregularized multi-marginal optimal transport (MOT) and its dual (MD) we have introduced above. The link between MOT and its entropic regularization is very strong, and a consequence of the  $\Gamma$ -convergence of  $(MOT_\varepsilon)$  toward (MOT) (one can adapt the proof in [Car+17] or see [BCN19, GKR20] for  $\Gamma$ -convergence in some specific cases) is that

$$\lim_{\varepsilon \rightarrow 0} MOT_\varepsilon = MOT_0.$$

By the direct method in the calculus of variations and strict convexity of the entropy, one can show that  $(\text{MOT}_\varepsilon)$  admits a unique solution  $\gamma_\varepsilon$ , called *optimal entropic plan*. Moreover, there exist  $m$  real-valued Borel functions  $\phi_i^\varepsilon$  such that

$$(2.3) \quad \gamma_\varepsilon = \exp\left(\frac{\oplus_{i=1}^m \phi_i^\varepsilon - c}{\varepsilon}\right) \otimes_{i=1}^m \mu_i,$$

where  $\oplus_{i=1}^m \phi_i^\varepsilon := (x_1, \dots, x_m) \mapsto \sum_{i=1}^m \phi_i^\varepsilon(x_i)$ , and in particular we have that

$$(2.4) \quad \text{MOT}_\varepsilon = \sum_{i=1}^m \int_{X_i} \phi_i^\varepsilon d\mu_i$$

and these functions have continuous representatives and are uniquely determined up a.e. to additive constants. The reader is referred to the analysis of [MG20], to [Nen16] for the extension to the multi-marginal setting, and to [BL92, BLN94, Csi75, FG97, RT98] for earlier references on the two marginals framework.

The functions  $\phi_i^\varepsilon$  in (2.3) are called *Schrödinger potentials*, the terminology being motivated by the fact that they solve the dual problem  $(\text{MD}_\varepsilon)$  and are as such the (unique) solutions to the so-called *Schrödinger system*: for all  $i \in \{1, \dots, m\}$ ,

$$(2.5) \quad \phi_i(x_i) = -\varepsilon \log \int_{X_{-i}} e^{\frac{\oplus_{1 \leq j \leq m, j \neq i} \phi_j^\varepsilon - c(x)}{\varepsilon}} d \otimes_{1 \leq j \leq m, j \neq i} \mu_j \quad \text{for } \mu_i\text{-a.e. } x_i,$$

where  $X_{-i} = \prod_{1 \leq j \leq m, j \neq i} X_j$ . Note that (2.5) is a “softmin” version of the multi-marginal  $c$ -conjugacy relation for Kantorovich potentials.

### 3 Upper bounds

We start by establishing an upper bound, which will depend on the dimension of the marginals, for locally Lipschitz cost functions. We will then improve it for locally semiconcave (in particular  $\mathcal{C}^2$ ) cost functions.

#### 3.1 Upper bound for locally Lipschitz costs

The natural notion of dimension which arises is the *entropy dimension*, also called *information dimension* or *Rényi dimension* [Rén59].

**Definition 3.1** (Rényi dimension (following [You82])) If  $\mu$  is a probability measure over a metric space  $X$ , we set for every  $\delta > 0$ ,

$$H_\delta(\mu) = \inf \left\{ \sum_{n \in \mathbb{N}} \mu(A_n) \log(1/\mu(A_n)) \mid \forall n, \text{diam}(A_n) \leq \delta, \text{ and } X = \bigsqcup_{n \in \mathbb{N}} A_n \right\},$$

where the infimum is taken over countable partitions  $(A_n)_{n \in \mathbb{N}}$  of  $X$  by Borel subsets of diameter less than  $\delta$ , and we define the *lower and upper entropy dimensions* of  $\mu$ , respectively by

$$\underline{\dim}_R(\mu) := \liminf_{\delta \rightarrow 0^+} \frac{H_\delta(\mu)}{\log(1/\delta)}, \quad \overline{\dim}_R(\mu) := \limsup_{\delta \rightarrow 0^+} \frac{H_\delta(\mu)}{\log(1/\delta)}.$$



Notice that if  $\mu$  is compactly supported on a Lipschitz manifold of dimension  $d$ , then  $N_\delta(\text{spt } \mu) \leq d \log(1/\delta) + C$  for some constant  $C > 0$  and  $\delta \in (0, 1]$ , where  $N_\delta(\text{spt } \mu)$  is the box-counting number of  $\text{spt } \mu$ , i.e., the minimal number of sets of diameter  $\delta > 0$  which cover  $\text{spt } \mu$ . In particular, by concavity of  $t \mapsto t \log(1/t)$ , we have

$$(3.1) \quad H_\delta(\mu) \leq \log N_\delta(\text{spt } \mu).$$

We refer to the beginning of [CPT23, Section 3.1] for additional information and references on Rényi dimension.

The following theorem establishes an upper bound for locally Lipschitz costs.

**Theorem 3.1** *Assume that for  $i \in \{1, \dots, m\}$ ,  $\mu_i \in \mathcal{P}(X_i)$  is a compactly supported measure on a Lipschitz sub-manifold  $X_i$  of dimension  $d_i$  and  $c \in \mathcal{C}_{\text{loc}}^{0,1}(X)$ , then*

$$(3.2) \quad \text{MOT}_\varepsilon \leq \text{MOT}_0 + \left( \sum_{i=1}^m d_i - \max_{j \in \{1, \dots, m\}} d_j \right) \varepsilon \log(1/\varepsilon) + O(\varepsilon).$$

**Proof** Given an optimal plan  $\gamma_0$  for  $\text{MOT}_0$ , we use the so-called “block approximation” introduced in [Car+17]. For every  $\delta > 0$  and  $i \in \{1, \dots, m\}$ , consider a partition  $X_i = \bigsqcup_{n \in \mathbb{N}} A_i^n$  of Borel sets such that<sup>1</sup>  $\text{diam}(A_i^n) \leq \delta$  for every  $n \in \mathbb{N}$ , and set

$$\mu_i^n := \begin{cases} \frac{\mu_i \llcorner A_i^n}{\mu_i(A_i^n)}, & \text{if } \mu_i(A_i^n) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

then for every  $m$ -uple  $n = (n_1, \dots, n_m) \in \mathbb{N}^m$ ,

$$(\gamma_0)^n := \gamma_0(A_1^{n_1} \times \dots \times A_m^{n_m}) \mu_1^{n_1} \otimes \dots \otimes \mu_m^{n_m},$$

and finally,

$$\gamma_\delta := \sum_{n \in \mathbb{N}^m} (\gamma_0)^n.$$

By definition,  $\gamma_\delta \ll \otimes_{i=1}^m \mu_i$  and we may check that its marginals are the  $\mu_i$ 's. Besides,  $\gamma_\delta(A) = \gamma_0(A)$  for every  $A = \prod_{i=1}^m A_i^{n_i}$  where  $n \in \mathbb{N}^m$ , and for  $\otimes_{i=1}^m \mu_i$ -almost every  $x = (x_1, \dots, x_m) \in \prod_{i=1}^m A_i^{n_i}$ ,

$$\frac{d\gamma_\delta}{d \otimes_{i=1}^m \mu_i}(x_1, \dots, x_m) := \begin{cases} \frac{\gamma_0(A_1^{n_1} \times \dots \times A_m^{n_m})}{\mu_1(A_1^{n_1}) \dots \mu_m(A_m^{n_m})}, & \text{if } \mu_1(A_1^{n_1}) \dots \mu_m(A_m^{n_m}) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Let us compute its entropy and assume for simplicity that the measure  $\mu_m$  is the one such that  $\overline{\dim}_R(\mu_m) = \max_{i \in \{1, \dots, m\}} \overline{\dim}(\mu_i)$ :

<sup>1</sup>We always consider the Euclidean distance over  $\mathbb{R}^N$ , but since the supports of the measures are compact and the sub-manifolds are Lipschitz, we may equivalently consider the intrinsic metric over the sub-manifolds: they are equivalent distances at small scale, i.e., for  $|y - x| \leq \delta_0$  for some  $\delta_0 > 0$ .

$$\begin{aligned}
 \text{Ent}(\gamma_\delta \mid \otimes_{i=1}^m \mu_i) &= \sum_{n \in \mathbb{N}^m} \int_{\prod_{i=1}^m A_i^{n_i}} \log \left( \frac{\gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m})}{\mu_1(A_1^{n_1}) \cdots \mu_m(A_m^{n_m})} \right) d\gamma_\delta \\
 &= \sum_{n \in \mathbb{N}^m} \gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m}) \log \left( \frac{\gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m})}{\mu_1(A_1^{n_1}) \cdots \mu_m(A_m^{n_m})} \right) \\
 &= \sum_{n \in \mathbb{N}^m} \gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m}) \log \left( \frac{\gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m})}{\mu_m(A_m^{n_m})} \right) \\
 &\quad + \sum_{j=1}^{m-1} \sum_{n \in \mathbb{N}^m} \gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m}) \log(1/\mu_j(A_j^{n_j})) \\
 &= \sum_{n \in \mathbb{N}^m} \gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m}) \log \left( \frac{\gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m})}{\mu_m(A_m^{n_m})} \right) \\
 &\quad + \sum_{j=1}^{m-1} \sum_{n_j \in \mathbb{N}} \gamma_0 \left( \prod_{i=1}^{j-1} X_i \times A_j^{n_j} \times \prod_{i=j+1}^m X_i \right) \mu_j(A_j^{n_j}) \log(1/\mu_j(A_j^{n_j})) \\
 &\leq \sum_{j=1}^{m-1} \sum_{n_j \in \mathbb{N}} \mu_j(A_j^{n_j}) \log(1/\mu_j(A_j^{n_j})),
 \end{aligned}$$

the last inequality coming from the inequality  $\gamma_0(A_1^{n_1} \times \cdots \times A_m^{n_m}) \leq \mu_m(A_m^{n_m})$ . Taking partitions  $(A_j^n)_{n \in \mathbb{N}}$  of diameter smaller than  $\delta$  such that  $\sum_{n_j \in \mathbb{N}} \mu_j(A_j^{n_j}) \log(1/\mu_j(A_j^{n_j})) \leq H_\delta(\mu_j) + \frac{1}{m-1}$ , we get

$$\text{Ent}(\gamma_\delta \mid \otimes_{i=1}^m \mu_i) \leq \sum_{j=1}^{m-1} H_\delta(\mu_j) + 1.$$

Since the  $\mu_i$ 's have compact support and  $c$  is locally Lipschitz, for  $\delta$  small enough, there exists  $L \in (0, +\infty)$  not depending on  $\delta$  such that  $[c]_{\mathcal{C}^{0,1}(A)} \leq L$  for every  $A \in \mathcal{A} := \{\prod_{i=1}^m A_i^{n_i} \mid n \in \mathbb{N}^m, \mu_1(A_1^{n_1}) \cdots \mu_m(A_m^{n_m}) > 0\}$ . Notice that the  $\infty$ -Wasserstein distance (see [San15, Section 3.2]) with respect to the norm  $\|\cdot\|^2$  satisfies  $W_\infty(\gamma_\delta, \gamma_0) \leq \delta$ . Indeed,  $\text{diam } A \leq \delta$  and  $\gamma_0(A) = \gamma_\delta(A)$  for every  $A \in \mathcal{A}$ , so that  $\Gamma := \sum_{A \in \mathcal{A}} \gamma_0(A)^{-1} (\gamma_\delta \llcorner A) \otimes (\gamma_0 \llcorner A)$  is a transport plan from  $\gamma_\delta$  to  $\gamma_0$  satisfying  $\Gamma$ -ess sup  $(x, x') \mapsto \|x' - x\| \leq \delta$ . Thus, taking  $\gamma_\delta$  as competitor in (MOT $_\epsilon$ ), we obtain

$$\begin{aligned}
 \text{MOT}_\epsilon &\leq \int c d\gamma_\delta + \epsilon \sum_{j \leq m-1} H_\delta(\mu_j) + \epsilon \\
 &= \text{MOT}_0 + \sum_{A \in \mathcal{A}} \int_A c d(\gamma_\delta - \gamma_0) + \epsilon \sum_{j \leq m-1} H_\delta(\mu_j) + \epsilon \\
 (3.3) \quad &\leq \text{MOT}_0 + \sum_{A \in \mathcal{A}} L W_\infty(\gamma_\delta \llcorner A, \gamma_0 \llcorner A) \gamma_0(A) + \epsilon \sum_{j \leq m-1} H_\delta(\mu_j) + \epsilon \\
 &\leq \text{MOT}_0 + L\delta + \epsilon \sum_{j \leq m-1} \frac{H_\delta(\mu_j)}{\log(1/\delta)} \log(1/\delta) + \epsilon.
 \end{aligned}$$

<sup>2</sup>The Wasserstein distance of order  $p$  is defined here by  $W_p^p(\mu, \nu) := \inf\{\int \|y - x\|^p d\gamma(x, y) \mid \gamma \in \Pi(\mu, \nu)\}$  for  $p \in [1, +\infty)$  and by  $W_\infty(\mu, \nu) = \inf\{\gamma\text{-ess sup}(x, y) \mapsto \|y - x\| \mid \gamma \in \Pi(\mu, \nu)\}$  for  $p = +\infty$ .

Taking  $\delta = \varepsilon$  and recalling that the  $\mu_j$ 's are concentrated on sub-manifolds of dimension  $d_j$ , which implies that  $H_\delta(\mu_j) \leq d_j \log(1/\delta) + \frac{C^*-1-L}{m-1}$  for some  $C^* \geq L+1$  and for every  $j \in \{1, \dots, m\}$ , we get

$$\text{MOT}_\varepsilon \leq \text{MOT}_0 + \left( \sum_{j \leq m-1} d_j \right) \varepsilon \log(1/\varepsilon) + C^* \varepsilon. \quad \blacksquare$$

**Remark 3.2** If the  $\mu_i$ 's are merely assumed to have compact support (not necessarily supported on a sub-manifold), the above proof actually shows the slightly weaker estimate

(3.4)

$$\text{MOT}_\varepsilon \leq \text{MOT}_0 + \left( \sum_{i=1}^m \overline{\dim}_R(\mu_i) - \max_{j \in \{1, \dots, m\}} \overline{\dim}_R(\mu_j) \right) \varepsilon \log(1/\varepsilon) + o(\varepsilon \log(1/\varepsilon)).$$

Indeed, for every  $i$ , by definition of  $\overline{\dim}_R(\mu_i) = d_i$ , we have  $\frac{H_\delta(\mu_i)}{\log(1/\delta)} \leq \sup_{0 < \delta' \leq \delta} \frac{H_{\delta'}(\mu_i)}{\log(1/\delta')} = d_i + o(1)$  as  $\delta \rightarrow 0$ ; thus, taking  $\delta = \varepsilon$  as above, we have  $\varepsilon \frac{H_\delta(\mu_i)}{\log(1/\delta)} \log(1/\delta) \leq (d_i + o(1)) \varepsilon \log(1/\varepsilon)$ .

Besides, notice that by taking  $m = 2$  and  $d_1 = d_2 = d$ , one easily retrieves [CPT23, Proposition 3.1].

### 3.2 Upper bound for locally semiconcave costs

We provide now a finer upper bound under the additional assumptions that the  $X_i$ 's are  $\mathcal{C}^2$  sub-manifolds of  $\mathbb{R}^N$ ,  $c$  is locally semiconcave as in Definition 3.2 (which is the case when  $c \in \mathcal{C}^2(X, \mathbb{R}_+)$ ), and the  $\mu_i$ 's are measures in  $L^\infty(\mathcal{H}_{X_i}^{d_i})$  with compact support in  $X_i$ .

**Definition 3.2** A function  $f : X \rightarrow \mathbb{R}$  defined on a  $\mathcal{C}^2$  sub-manifold  $X \subseteq \mathbb{R}^N$  of dimension  $d$  is *locally semiconcave* if for every  $x \in X$  there exists a local chart (i.e., a  $\mathcal{C}^2$  diffeomorphism)  $\psi : U \rightarrow \Omega$  where  $U \subseteq X$  is an open neighborhood of  $x$  and  $\Omega$  is an open convex subset of  $\mathbb{R}^d$ , such that  $f \circ \psi^{-1}$  is  $\lambda$ -concave for some  $\lambda \in \mathbb{R}$ , meaning  $f \circ \psi^{-1} - \lambda \frac{|\cdot|^2}{2}$  is concave on  $\Omega$ .

**Lemma 3.3** (Local semiconcavity and covering) *Let  $c : X \rightarrow \mathbb{R}_+$  be a locally semiconcave cost function and  $(\phi_i)_{1 \leq i \leq m} \in \prod_{1 \leq i \leq m} \mathcal{C}(K_i)$  be a system of  $c$ -conjugate functions as in (2.2) defined on compact subsets  $K_i \subseteq X_i$ . We can find  $\lambda \in \mathbb{R}$ ,  $J \in \mathbb{N}^*$  and for every  $i \in \{1, \dots, m\}$  a finite open covering  $(U_i^j)_{1 \leq j \leq J}$  of  $K_i$  together with bi-Lipschitz local charts  $\psi_i^j : U_i^j \rightarrow \Omega_i^j$  satisfying the following properties, having set  $\Omega^j := \prod_{1 \leq i \leq m} \Omega_i^{j_i}$  and  $\psi^j := (\psi_1^{j_1}, \dots, \psi_m^{j_m})$  for every  $j = (j_1, \dots, j_m) \in \{1, \dots, J\}^m$ :*

- (1) *for every  $j \in \{1, \dots, J\}^m$ ,  $c \circ (\psi^j)^{-1}$  is  $\lambda$ -concave on  $\Omega^j$ ,*
- (2) *for every  $(i, j) \in \{1, \dots, m\} \times \{1, \dots, J\}$ ,  $\phi_i \circ (\psi_i^j)^{-1}$  is  $\lambda$ -concave on  $\Omega_i^j$ .*

*In particular, all the  $\phi_i$ 's are locally semiconcave.*

**Proof** For every  $i$ , by compactness of the  $K_i$ 's, we can find a finite open covering  $(U_i^j)_{1 \leq j \leq J}$  of  $K_i$  and bi-Lipschitz local charts  $\psi_i^j: U_i^j \rightarrow \Omega_i^j$  such that for every  $\mathbf{j} = (j_1, \dots, j_m) \in \{1, \dots, J\}^m$ ,  $c \circ (\psi^{\mathbf{j}})^{-1} - \lambda^{\mathbf{j}} \frac{\|\cdot\|^2}{2}$  is concave for some  $\lambda^{\mathbf{j}} \in \mathbb{R}$ . We may assume that  $\lambda^{\mathbf{j}} = \lambda$  for every  $\mathbf{j}$ , by taking  $\lambda := \max\{\lambda^{\mathbf{j}} \mid \mathbf{j} \in \{1, \dots, J\}^m\}$ . Fix  $i \in \{1, \dots, m\}$ ,  $j \in \{1, \dots, J\}$ , then for every  $\mathbf{k} = (k_\ell)_{\ell \neq i} \in \{1, \dots, J\}^{m-1}$ , set  $\hat{\mathbf{k}} = (k_1, \dots, k_{i-1}, j, k_{i+1}, \dots)$ . Notice that for every  $y \in \Omega_i^j$ ,

$$\begin{aligned} & \phi_i \circ (\psi_i^j)^{-1}(y) \\ &= \inf_{(x_\ell)_{\ell \neq i} \in K_{-i}} c(x_1, \dots, x_{i-1}, (\psi_i^j)^{-1}(y), x_{i+1}, \dots) - \sum_{\ell: \ell \neq i} \phi_\ell(x_\ell) \\ &= \min_{\mathbf{k}=(k_\ell)_{\ell \neq i}} \inf_{(y_\ell)_{\ell \neq i} \in \Omega^{\hat{\mathbf{k}}}_{-i}} c \circ (\psi^{\hat{\mathbf{k}}})^{-1}(y_1, \dots, y_{i-1}, y, y_{i+1}, \dots) - \sum_{\ell: \ell \neq i} \phi_\ell \circ (\psi_\ell^{k_\ell})^{-1}(y_\ell), \end{aligned}$$

and we see that it is  $\lambda$ -concave as an infimum of  $\lambda$ -concave functions.  $\blacksquare$

We are going to use an integral variant of Alexandrov's theorem which is proved in [CPT23].

**Lemma 3.4** [CPT23, Lemma 3.6] *Let  $f: \Omega \rightarrow \mathbb{R}$  be a  $\lambda$ -concave function defined on a convex open set  $\Omega \subseteq \mathbb{R}^d$ , for some  $\lambda \geq 0$ . There exists a constant  $C \geq 0$  depending only on  $d$  such that*

$$\begin{aligned} (3.5) \quad & \int_{\Omega} \sup_{y \in B_r(x) \cap \Omega} |f(y) - (f(x) + \nabla f(x) \cdot (y - x))| dx \\ & \leq Cr^2 \mathcal{H}^{d-1}(\partial\Omega)([f]_{\mathcal{C}^{0,1}(\Omega)} + \lambda \operatorname{diam}(\Omega)). \end{aligned}$$

We may now state the main result of this section.

**Theorem 3.5** *Let  $c \in \mathcal{C}^2(X)$  and assume that for every  $i \in \{1, \dots, m\}$ ,  $X_i \subseteq \mathbb{R}^N$  is a  $\mathcal{C}^2$  sub-manifold of dimension  $d_i$  and  $\mu_i \in L^\infty(\mathcal{H}_{X_i}^{d_i})$  is a probability measure compactly supported in  $X_i$ . Then there exist constants  $\varepsilon_0, C^* \geq 0$  such that for  $\varepsilon \in (0, \varepsilon_0]$ ,*

$$(3.6) \quad \boxed{\operatorname{MOT}_\varepsilon \leq \operatorname{MOT}_0 + \frac{1}{2} \left( \sum_{i=1}^m d_i - \max_{1 \leq i \leq m} d_i \right) \varepsilon \log(1/\varepsilon) + C^* \varepsilon.}$$

**Proof** The measures  $\mu_i$  being compactly supported in  $X_i$ , take for every  $i \in \{1, \dots, m\}$  an open subset  $U_i$  of  $X_i$  such that  $\operatorname{spt} \mu_i \subseteq U_i \Subset X_i$  and define the compact set  $K_i := \bar{U}_i$ . Take  $(\phi_i)_{1 \leq i \leq m} \in \prod_{1 \leq i \leq m} \mathcal{C}(K_i)$  an  $m$ -uple of  $c$ -conjugate Kantorovich potentials and a transport plan  $\gamma_0 \in \Pi(\mu_1, \dots, \mu_m)$  which are optimal for the unregularized problems (MD) and (MOT), respectively. In particular,

$$\begin{aligned} (3.7) \quad & E := c - \oplus_{i=1}^m \phi_i \geq 0 \quad \text{on} \quad U = \prod_{1 \leq i \leq m} U_i, \\ & E = 0 \quad \text{on} \quad \operatorname{spt} \gamma_0 \subseteq U. \end{aligned}$$

For every  $i \in \{1, \dots, m\}$ , we consider the coverings  $(U_i^j)_{1 \leq j \leq J}$  and bi-Lipschitz local charts  $\psi_i^j: U_i^j \rightarrow \Omega_i^j$  for  $j \in \{1, \dots, J\}$  provided by Lemma 3.3 and we notice by compactness that there exist open subsets  $\tilde{U}_i^j \Subset U_i^j$  such that for a small  $\delta_0 > 0$ ,

the  $\delta_0$ -neighborhood of  $\tilde{\Omega}_i^j := \psi_i^j(\tilde{U}_i^j)$  is included in  $\Omega_i^j$  for every  $j$ , and  $(\tilde{U}_i^j)_{1 \leq j \leq J}$  is still an open covering of  $K_i$ . For  $\delta \in (0, \delta_0)$ , we consider the block approximation  $\gamma_\delta$  of  $\gamma_0$  built in the proof of Theorem 3.1, as well as some  $\kappa_\delta \in \Pi(\gamma_0, \gamma_\delta)$  such that  $\sup_{(\mathbf{x}_0, \mathbf{x}) \in \text{spt } \kappa_\delta} \|\mathbf{x}_0 - \mathbf{x}\| \leq \delta$ . For every  $\mathbf{j} = (j_1, \dots, j_m) \in \{1, \dots, J\}^m$ , we set  $E^{\mathbf{j}} := E \circ (\psi^{\mathbf{j}})^{-1}$ ,  $U^{\mathbf{j}} := \prod_{i=1}^m U_i^{j_i}$ , and  $\tilde{U}^{\mathbf{j}} := \prod_{i=1}^m \tilde{U}_i^{j_i}$ , and we write

$$\begin{aligned} \int_{\mathbf{X}} c d\gamma_\delta - \int_{\mathbf{X}} c d\gamma_0 &= \int_U E d\gamma_\delta \\ &= \int_{U \times U} E(\mathbf{x}) d\kappa_\delta(\mathbf{x}_0, \mathbf{x}) \\ &\leq \sum_{\mathbf{j} \in \{1, \dots, J\}^m} \int_{(\mathbf{x}_0, \mathbf{x}) \in \tilde{U}^{\mathbf{j}} \times U} E(\mathbf{x}) d\kappa_\delta(\mathbf{x}_0, \mathbf{x}) \\ &\leq \sum_{\mathbf{j} \in \{1, \dots, J\}^m} \int_{(\mathbf{x}_0, \mathbf{x}) \in (U^{\mathbf{j}})^2} E^{\mathbf{j}}(\psi^{\mathbf{j}}(\mathbf{x})) d\kappa_\delta(\mathbf{x}_0, \mathbf{x}). \end{aligned}$$

Notice that for every  $\mathbf{j} \in \{1, \dots, J\}^m$  and  $\gamma_0$ -a.e.  $\mathbf{x}_0 \in U^{\mathbf{j}}$ ,  $E^{\mathbf{j}}$  is differentiable at  $\psi^{\mathbf{j}}(\mathbf{x}_0)$ , or equivalently  $E$  is differentiable at  $\mathbf{x}_0$ . Indeed,  $c$  is differentiable everywhere, and for every  $i \in \{1, \dots, m\}$  and  $j \in \{1, \dots, J\}$ ,  $\phi_i \circ (\psi_i^j)^{-1}$  is semiconcave thus differentiable  $\mathcal{L}^{d_i}$ -a.e.; hence,  $\phi_i$  is differentiable  $\mu_i$ -a.e. on  $U_i^j$  because  $\mu_i \ll \mathcal{H}^{d_i}$  and  $\psi_i^j$  is bi-Lipschitz, which in turn implies that  $\oplus_{i=1}^m \phi_i$  is differentiable  $\gamma_0$ -a.e. on  $U^j$  because  $\gamma_0 \in \Pi(\mu_1, \dots, \mu_m)$ . Moreover, by (3.7), we have  $T_{\psi^j(\mathbf{x}_0)} E^{\mathbf{j}} \equiv 0$  for  $\gamma_0$ -a.e.  $\mathbf{x}_0 \in U^{\mathbf{j}}$ , where  $T_{\gamma_0} f$  designates the first-order Taylor expansion  $y \mapsto f(\gamma_0) + \nabla f(\gamma_0) \cdot (y - \gamma_0)$  for any function  $f$  which is differentiable at  $\gamma_0$ . We may then compute

$$\begin{aligned} (3.8) \quad &\int_{\mathbf{X}} c d\gamma_\delta - \int_{\mathbf{X}} c d\gamma_0 \\ &\leq \sum_{\mathbf{j} \in \{1, \dots, J\}^m} \int_{(\mathbf{x}_0, \mathbf{x}) \in (U^{\mathbf{j}})^2} \left( E^{\mathbf{j}}(\psi^{\mathbf{j}}(\mathbf{x})) - T_{\psi^{\mathbf{j}}(\mathbf{x}_0)} E^{\mathbf{j}}(\psi^{\mathbf{j}}(\mathbf{x}) - \psi^{\mathbf{j}}(\mathbf{x}_0)) \right) d\kappa_\delta(\mathbf{x}_0, \mathbf{x}) \\ &= \sum_{\mathbf{j} = (j_1, \dots, j_m)} \left( \int_{(\mathbf{x}_0, \mathbf{x}) \in (U^{\mathbf{j}})^2} \left( c^{\mathbf{j}}(\psi^{\mathbf{j}}(\mathbf{x})) - T_{\psi^{\mathbf{j}}(\mathbf{x}_0)} c^{\mathbf{j}}(\psi^{\mathbf{j}}(\mathbf{x}) - \psi^{\mathbf{j}}(\mathbf{x}_0)) \right) d\kappa_\delta(\mathbf{x}_0, \mathbf{x}) \right. \\ &\quad \left. - \sum_{i=1}^m \int_{(\mathbf{x}_0, \mathbf{x}) \in (U_i^{j_i})^2} \left( \phi_i^{j_i}(\psi_i^{j_i}(\mathbf{x})) - T_{\psi_i^{j_i}(\mathbf{x}_0)} \phi_i^{j_i}(\psi_i^{j_i}(\mathbf{x}) - \psi_i^{j_i}(\mathbf{x}_0)) \right) d(e_i, e_i)_\# \kappa_\delta(\mathbf{x}_0, \mathbf{x}) \right). \end{aligned}$$

Now, since  $c^{\mathbf{j}}$  is  $\lambda$ -concave on each  $\Omega^{\mathbf{j}} := \psi^{\mathbf{j}}(U^{\mathbf{j}})$ , whenever  $\|\mathbf{x}_0 - \mathbf{x}\| \leq \delta$ , we have

$$(3.9) \quad c^{\mathbf{j}}(\psi^{\mathbf{j}}(\mathbf{x})) - T_{\psi^{\mathbf{j}}(\mathbf{x}_0)} c^{\mathbf{j}}(\psi^{\mathbf{j}}(\mathbf{x}) - \psi^{\mathbf{j}}(\mathbf{x}_0)) \leq \lambda \frac{|\psi^{\mathbf{j}}(\mathbf{x}) - \psi^{\mathbf{j}}(\mathbf{x}_0)|^2}{2} \leq \frac{m\lambda L^{\mathbf{j}}}{2} \delta^2,$$

where  $L^{\mathbf{j}} := \max\{\text{Lip}(\psi^{\mathbf{j}}), \text{Lip}((\psi^{\mathbf{j}})^{-1})\}$ . Besides, we may apply Lemma 3.4 to each  $\phi_i^{j_i}$  over  $\Omega_i^{j_i}$  to get

$$\begin{aligned}
 (3.10) \quad & \left| \int_{(x_0, x) \in (U_i^{j_i})^2} \left( \phi_i^{j_i}(\psi_i^{j_i}(x)) - T_{\psi_i^{j_i}(x_0)} \phi_i^{j_i}(\psi_i^{j_i}(x) - \psi_i^{j_i}(x_0)) \right) d(e_i, e_i)_{\#} \kappa_{\delta}(x_0, x) \right| \\
 & \leq \int_{x_0 \in U_i^{j_i}} \sup_{y \in B_{L^j \delta}(\psi_i^{j_i}(x_0)) \cap \Omega_i^{j_i}} \left| \phi_i^{j_i}(y) - T_{\psi_i^{j_i}(x_0)} \phi_i^{j_i}(y - \psi_i^{j_i}(x_0)) \right| d(e_i, e_i)_{\#} \kappa_{\delta}(x_0, x) \\
 & = \int_{U_i^{j_i}} \sup_{y \in B_{L^j \delta}(\psi_i^{j_i}(x_0)) \cap \Omega_i^{j_i}} \left| \phi_i^{j_i}(y) - T_{\psi_i^{j_i}(x_0)} \phi_i^{j_i}(y - \psi_i^{j_i}(x_0)) \right| d\mu_i(x_0) \\
 & \leq \int_{\Omega_i^{j_i}} \sup_{y \in B_{L^j \delta}(y_0) \cap \Omega_i^{j_i}} \left| \phi_i^{j_i}(y) - T_{y_0} \phi_i^{j_i}(y - y_0) \right| d(\psi_i^{j_i})_{\#} \mu_i(y_0) \\
 & \leq \|\mu_i\|_{L^\infty(\mathcal{H}^{d_i})} L^j C (L^j \delta)^2 \mathcal{H}^{d_i-1}(\partial \Omega_i^{j_i}) \left( [\phi_i^{j_i}]_{\mathcal{C}^{0,1}(\Omega_i^{j_i})} + \lambda \text{diam}(\Omega_i^{j_i}) \right) \\
 & \leq C^j \delta^2,
 \end{aligned}$$

for some constant  $C^j \in (0, +\infty)$  which does not depend on  $\delta$ . Reporting (3.9) and (3.10) in (3.8) yields

$$\int_{\mathbf{X}} c d\gamma_{\delta} - \int_{\mathbf{X}} c d\gamma_0 \leq \sum_{j \in \{1, \dots, J\}^m} \left( \frac{m\lambda L^j}{2} + C^j \right) \delta^2 =: C' \delta^2.$$

Finally, we proceed as in the end of the proof of Theorem 3.1, taking  $\gamma_{\delta}$  as competitor in the primal formulation (MOT $_{\varepsilon}$ ), so as to obtain

$$\begin{aligned}
 \text{MOT}_{\varepsilon} - \text{MOT}_0 & \leq \int_{\mathbf{X}} c d\gamma_{\delta} - \int_{\mathbf{X}} c d\gamma_0 + \varepsilon \sum_{i \leq m-1} H_{\delta}(\mu_i) + \varepsilon \\
 & \leq C' \delta^2 + \varepsilon \sum_{i \leq m-1} (d_i \log(1/\delta) + C''),
 \end{aligned}$$

where  $C'' \in (0, +\infty)$  is a constant such that  $H_{\delta}(\mu_i) \leq d_i \log(1/\delta) + C'' - 1$ . Taking  $\delta = \sqrt{\varepsilon}$  for  $\varepsilon \leq \delta_0^2$  yields

$$\text{MOT}_{\varepsilon} - \text{MOT}_0 \leq \frac{1}{2} \left( \sum_{i=1}^{m-1} d_i \right) \varepsilon \log(1/\varepsilon) + (C' + (m-1)C'')\varepsilon,$$

and we obtain the desired estimate recalling that the index  $i = m$  was chosen merely to simplify notations. ■

## 4 Lower bound for $\mathcal{C}^2$ costs with a signature condition

In this section, we consider a cost  $c \in \mathcal{C}^2(\mathbf{X}, \mathbb{R}_+)$  where  $\mathbf{X} = X_1 \times \dots \times X_m$  and we will assume that for every  $i \in \{1, \dots, m\}$ , the measure  $\mu_i$  is compactly supported on a  $\mathcal{C}^2$  sub-manifold  $X_i \subseteq \mathbb{R}^N$  of dimension  $d_i$ . We are going to establish a lower bound in the same form as the fine upper bound of Theorem 3.5, the dimensional constant being this time related to the signature of some bilinear forms, following ideas from [Pas12].

**Lemma 4.1** *Let  $c \in \mathcal{C}^2(\mathbf{X}, \mathbb{R}_+)$  and  $(\phi_1, \dots, \phi_m) \in \mathcal{C}(K_1) \times \dots \times \mathcal{C}(K_m)$  be a system of  $c$ -conjugate functions on subsets  $K_i \subseteq X_i$  for every  $i$ . We set  $E := c - \phi_1 \oplus \dots \oplus$*

$\phi_m$  on  $\mathbf{K} := K_1 \times \cdots \times K_m$  and we take  $\tilde{\mathbf{x}} \in \mathbf{K}$  as well as some  $g_{\tilde{\mathbf{x}}} \in \{g(\tilde{\mathbf{x}}) \mid g \in G_c\}$  of signature  $(d^+, d^-, d^0)$ ,  $G_c$  being defined in (2.1). Then there exists local coordinates around  $\tilde{\mathbf{x}}$ , i.e.,  $\mathcal{C}^2$  diffeomorphisms

$$u = (u^0, u^-, u^+) : U \subseteq \mathbf{X} \rightarrow B_\rho^{d^+}(0) \times B_\rho^{d^-}(0) \times B_\rho^{d^0}(0),$$

$U$  being an open neighborhood of  $\tilde{\mathbf{x}}$ , such that if  $\mathbf{x}, \mathbf{x}' \in B_r(\tilde{\mathbf{x}}) \subseteq U$ ,

$$(4.1) \quad \frac{E(\mathbf{x}') + E(\mathbf{x})}{2} \geq |u^+(\mathbf{x}') - u^+(\mathbf{x})|^2 - |u^-(\mathbf{x}') - u^-(\mathbf{x})|^2 - \eta(r) |u(\mathbf{x}') - u(\mathbf{x})|^2,$$

where  $\eta(r) \geq 0$  tends to 0 as  $r \rightarrow 0$ .

**Proof** Let  $p = \{p_-, p_+\} \in P$ . For  $y \in \prod_{i \in p_\pm} K_i$ , we set

$$\phi_{p_\pm}(y) := \sum_{i \in p_\pm} \phi_i(y_i).$$

We identify any  $\mathbf{x} \in \mathbf{K}$  with  $(x_{p_-}, x_{p_+})$ . Since the  $\phi_i$ 's are  $c$ -conjugate, for  $\mathbf{x}, \mathbf{x}' \in \mathbf{K}$ , it holds

$$\begin{aligned} E(\mathbf{x}') &= c(x'_{p_-}, x'_{p_+}) - \phi_{p_-}(x'_{p_-}) - \phi_{p_+}(x'_{p_+}) \\ &\geq c(x'_{p_-}, x'_{p_+}) - (c(x'_{p_-}, x_{p_+}) - \phi_{p_+}(x_{p_+})) - (c(x_{p_-}, x'_{p_+}) - \phi_{p_-}(x_{p_-})) \\ &= c(x'_{p_-}, x'_{p_+}) - c(x'_{p_-}, x_{p_+}) - c(x_{p_-}, x'_{p_+}) + c(x_{p_-}, x_{p_+}) - E(\mathbf{x}). \end{aligned}$$

Now we do computations in local charts  $\psi_i : U_i \subseteq X_i \rightarrow \psi_i(U_i) \subseteq \mathbb{R}^{d_i}$  which are  $\mathcal{C}^2$  diffeomorphisms such that  $B_R(\tilde{x}_i) \subseteq U_i$  for some  $R > 0$  and  $\psi_i(U_i)$  are balls centered at 0 for every  $i \in \{1, \dots, m\}$ . With a slight abuse, we use the same notation for points and functions written in these charts, and use Taylor's integral formula:<sup>3</sup>

$$E(\mathbf{x}') + E(\mathbf{x}) \geq \int_0^1 \int_0^1 D_{p_- p_+}^2 c(\mathbf{x}_{s,t})(x'_{p_-} - x_{p_-}, x'_{p_+} - x_{p_+}) ds dt,$$

where  $\mathbf{x}_{s,t} := (x_{p_-} + (1-s)x'_{p_-}, x_{p_+} + (1-t)x'_{p_+})$  for  $s, t \in [0, 1]$ . Since  $|D_{p_- p_+}^2 c(\mathbf{x}_{s,t}) - D_{p_- p_+}^2 c(\tilde{\mathbf{x}})| \leq \eta(r)$  where  $\eta$  is the maximum for  $p \in P$  of the moduli of continuity of  $D_{p_- p_+}^2 c$  at  $\tilde{\mathbf{x}}$ . Since  $\eta$  is independent from  $p$  and tends to 0 as  $r \rightarrow 0$  because  $c$  is  $\mathcal{C}^2$ , and by definition  $D^2 c(\tilde{\mathbf{x}})(x'_{p_-} - x_{p_-}, x'_{p_+} - x_{p_+}) = \frac{1}{2} g_p(\tilde{\mathbf{x}})(\mathbf{x}' - \mathbf{x}, \mathbf{x}' - \mathbf{x})$ , it holds

$$E(\mathbf{x}) + E(\mathbf{x}') \geq \frac{1}{2} g_p(\tilde{\mathbf{x}})(\mathbf{x}' - \mathbf{x}, \mathbf{x}' - \mathbf{x}) - \eta(r) \|\mathbf{x}' - \mathbf{x}\|^2.$$

Taking  $g_{\tilde{\mathbf{x}}} = \sum_{p \in P} t_p g_p(\tilde{\mathbf{x}})$  for some  $(t_p)_{p \in P} \in \Delta_P$  and averaging the previous inequality yields

$$(4.2) \quad E(\mathbf{x}) + E(\mathbf{x}') \geq \frac{1}{2} g_{\tilde{\mathbf{x}}}(\mathbf{x}' - \mathbf{x}, \mathbf{x}' - \mathbf{x}) - \eta(r) \|\mathbf{x}' - \mathbf{x}\|^2.$$

<sup>3</sup> Any linear combination  $az_i + by_i$  will designate  $\psi_i^{-1}(a\psi_i(z_i) + b\psi_i(y_i))$ .

Finally, we can find a linear isomorphism  $Q \in GL(\sum_{i=1}^m d_i, \mathbb{R})$  which diagonalizes  $g_{\bar{x}}$ , such that after setting  $u := Q \circ (\psi_1, \dots, \psi_m)$  and denoting  $u = (u^+, u^-, u^0) : \prod_{i=1}^m U_i \rightarrow \mathbb{R}^{d^+} \times \mathbb{R}^{d^-} \times \mathbb{R}^{d^0}$ , where  $(d^+, d^-, d^0)$  is the signature of  $g_{\bar{x}}$ , it holds

$$\frac{1}{4} g_{\bar{x}}(\mathbf{x}' - \mathbf{x}, \mathbf{x}' - \mathbf{x}) = |u^+(\mathbf{x}') - u^+(\mathbf{x})|^2 - |u^-(\mathbf{x}') - u^-(\mathbf{x})|^2.$$

Reporting this in (4.2), we get the result by replacing  $\eta$  with  $\|Q\|^{-1} \eta$  and restricting  $u$  to  $U := u^{-1}(B_{\rho}^{d^+}(0) \times B_{\rho}^{d^-}(0) \times B_{\rho}^{d^0}(0))$  for some small  $\rho > 0$ . ■

We will use the following positive signature condition:

(PS( $\kappa$ ))

for every  $\mathbf{x} \in X$ ,  $d_c^+(\mathbf{x}) \geq \kappa$  where  $d_c^{\pm}(\mathbf{x}) := \max \{d^{\pm}(g)(\mathbf{x}) \mid g \in G_c\}$ .

**Proposition 4.2** Let  $c \in \mathcal{C}^2(X)$  and assume that for every  $i \in \{1, \dots, m\}$ ,  $X_i \subseteq \mathbb{R}^N$  is a  $\mathcal{C}^2$  sub-manifold of dimension  $d_i$  and  $\mu_i \in L^{\infty}(\mathcal{H}_{X_i}^{d_i})$  is a probability measure compactly supported in  $X_i$ . If (PS( $\kappa$ )) is satisfied, then there exists a constant  $C_* \in [0, \infty)$  such that for every  $\varepsilon > 0$ ,

$$(4.3) \quad \text{MOT}_{\varepsilon} \geq \text{MOT}_0 + \frac{\kappa}{2} \varepsilon \log(1/\varepsilon) - C_* \varepsilon.$$

**Proof** The measures  $\mu_i$  being supported on some compact subsets  $K_i \subseteq X_i$ , consider a family  $(\phi_i)_{1 \leq i \leq m} \in \prod_{i=1}^m \mathcal{C}(K_i)$  of  $c$ -conjugate Kantorovich potentials. Taking  $(\phi_i)_{1 \leq i \leq m}$  as competitor in (MD' $_{\varepsilon}$ ), we get the lower bound

$$\begin{aligned} \text{MOT}_{\varepsilon} &\geq \sum_{i=1}^m \int_{K_i} \phi_i d\mu_i - \varepsilon \log \left( \int_K e^{-\frac{E}{\varepsilon}} d \otimes_{i=1}^m \mu_i \right) \\ &= \text{MOT}_0 - \varepsilon \log \left( \int_K e^{-\frac{E}{\varepsilon}} d \otimes_{i=1}^m \mu_i \right), \end{aligned}$$

where  $E := c - \oplus_{i=1}^m \phi_i$  on  $K = \prod_{i=1}^m K_i$  as in Lemma 4.1. We are going to show that for some constant  $C > 0$  and for every  $\varepsilon > 0$ ,

$$\int_K e^{-E/\varepsilon} d \otimes_{i=1}^m \mu_i \leq C \varepsilon^{\kappa/2},$$

which yields (4.3) with  $C_* = \log(C)$ .

For every  $\bar{x} \in K$ , we consider a quadratic form  $g_{\bar{x}} \in \{g(\bar{x}) \mid g \in G_c\}$  of signature  $(\kappa, d^-, d^0)$ , which is possible thanks to (PS( $\kappa$ )), and take a local chart<sup>4</sup>

$$u_{\bar{x}} : U \subseteq X \rightarrow B_R^{\kappa}(0) \times B_R^{d^-}(0) \times B_R^{d^0}(0)$$

as given by Lemma 4.1, such that (4.1) holds with  $\eta(r) \leq 1/2$  for every  $r$  such that  $B_r(\bar{x}) \subseteq U$ . Notice that  $u_{\bar{x}}$  is bi-Lipschitz with some constant  $L_{\bar{x}}$  on  $V_{\bar{x}} := u_{\bar{x}}^{-1}(B_{R/2}^{\kappa}(0) \times B_{R/2}^{d^-}(0) \times B_{R/2}^{d^0}(0))$ .

For every  $i \in \{1, \dots, m\}$ , we may write  $\mu_i = f_i \mathcal{H}_{X_i}^{d_i}$  for some density  $f_i : X_i \rightarrow \mathbb{R}_+$ . By applying several times the co-area formula [Fed96, Theorem 3.2.22] to the

<sup>4</sup> Although  $U, R, d^-$ , and  $d^0$  depend on  $\bar{x}$ , we do not index them with  $\bar{x}$  so as to ease notations.



projection maps onto  $X_i$ , we may justify that

$$\mathcal{H}_X^d = \otimes_{i=1}^m \mathcal{H}_{X_i}^{d_i} \quad \text{where} \quad d := \sum_i d_i.$$

We set  $E_{\bar{x}} := E \circ u_{\bar{x}}^{-1} : B_R^\kappa(0) \times B_R^{d^-}(0) \times B_R^{d^0}(0) \rightarrow [0, +\infty]$ , and we apply the area formula

$$\begin{aligned} \int_{V_{\bar{x}}} e^{-E/\varepsilon} d \otimes_{i=1}^m \mu_i &= \int_{V_{\bar{x}}} e^{-E/\varepsilon} \otimes_{i=1}^m f_i d \mathcal{H}_X^d \\ &= \int_{B_{R/2}^\kappa(0) \times B_{R/2}^{d^-}(0) \times B_{R/2}^{d^0}(0)} e^{-E_{\bar{x}}/\varepsilon} \otimes_{i=1}^m f_i J u_{\bar{x}}^{-1} d \mathcal{H}^\kappa \otimes \mathcal{H}^{d^-} \otimes \mathcal{H}^{d^0} \\ &\leq L_{\bar{x}} \prod_{i=1}^m \|\mu_i\|_{L^\infty(\mathcal{H}_{X_i}^{d_i})} \int_{B_{R/2}^\kappa(0) \times B_{R/2}^{d^-}(0) \times B_{R/2}^{d^0}(0)} e^{-E_{\bar{x}}(u^+, u^-, u^0)/\varepsilon} d(u^+, u^-, u^0). \end{aligned}$$

Now, for every  $(u^-, u^0) \in B_{R/2}^{d^-}(0) \times B_{R/2}^{d^0}(0)$ , consider a minimizer of  $E_{\bar{x}}(\cdot, u^-, u^0)$  over  $\tilde{B}_{R/2}^\kappa(0)$  denoted by  $f^+(u^-, u^0)$ . By (4.1) of Lemma 4.1, for every  $(u^+, u^-, u^0) \in B_{R/2}^\kappa(0) \times B_{R/2}^{d^-}(0) \times B_{R/2}^{d^0}(0)$ ,

$$\begin{aligned} E_{\bar{x}}(u^+, u^-, u^0) &\geq \frac{1}{2} (E_{\bar{x}}(f^+(u^-, u^0), u^-, u^0) + E_{\bar{x}}(u^+, u^-, u^0)) \\ &\geq (1 - 1/2) |u^+ - f^+(u^-, u^0)|^2 = \frac{1}{2} |u^+ - f^+(u^-, u^0)|^2. \end{aligned}$$

As a consequence, we obtain

$$\begin{aligned} &\int_{B_{R/2}^\kappa(0) \times B_{R/2}^{d^-}(0) \times B_{R/2}^{d^0}(0)} e^{-E_{\bar{x}}(u^+, u^-, u^0)/\varepsilon} d(u^+, u^-, u^0) \\ &\leq \int_{B_{R/2}^{d^-}(0) \times B_{R/2}^{d^0}(0)} \int_{B_{R/2}^\kappa(0)} e^{-\frac{|u^+ - f^+(u^-, u^0)|^2}{2\varepsilon}} du^+ d(u^-, u^0) \\ &\leq \varepsilon^{\kappa/2} \omega_{d^-} \omega_{d^0} R^{d^- + d^0} \int_{\mathbb{R}^\kappa} e^{-|u|^2/2} du = C_{\bar{x}} \varepsilon^{\kappa/2} \end{aligned}$$

for some constant  $C_{\bar{x}} > 0$  (which depends on  $\bar{x}$  through  $R$ ,  $d^-$ , and  $d^0$ ). The sets  $\{V_{\bar{x}}\}_{\bar{x} \in \Sigma}$  form an open covering of the compact set  $\Sigma := \{x \in K \mid E(x) = 0\}$ ; hence, we may extract a finite covering  $V_{\bar{x}_1}, \dots, V_{\bar{x}_L}$  and for every  $\varepsilon > 0$ ,

$$\int_{\bigcup_{\ell=1}^L V_{\bar{x}_\ell}} e^{-E/\varepsilon} d \otimes_{i=1}^m \mu_i \leq \varepsilon^{\kappa/2} \left( \sum_{\ell=1}^L L_{\bar{x}_\ell} C_{\bar{x}_\ell} \right) \left( \prod_{i=1}^m \|\mu_i\|_{L^\infty(\mathcal{H}_{X_i}^{d_i})} \right) = C_1 \varepsilon^{\kappa/2},$$

for some constant  $C_1 \in (0, +\infty)$ . Finally, since  $E$  is continuous and does not vanish on the compact set  $K' := K \setminus \bigcup_{\ell=1}^L V_{\bar{x}_\ell}$ , it is bounded from below on  $K'$  by some constant  $C_2 > 0$ . Therefore, for every  $\varepsilon > 0$ ,

$$\int_K e^{-E/\varepsilon} d \otimes_{1 \leq i \leq m} \mu_i \leq C_1 \varepsilon^{\kappa/2} + e^{-C_2/\varepsilon} \leq C \varepsilon^{\kappa/2},$$

for some constant  $C > 0$ . This concludes the proof.  $\blacksquare$

## 5 Examples and matching bound

We devote this section to applying the results we have stated above to several cost functions. For simplicity, we can assume that the dimensions of the  $X_i$  are all equal to some common  $d$  and the cost function  $c$  is  $\mathcal{C}^2$ . As in [Pas12], we consider, for the lower bound, the metric  $\bar{g}$  such that  $t_p = \frac{1}{2^{m-1}-1}$  for all  $p \in P$ , we remind that  $P$  is the set of partition of  $\{1, \dots, m\}$  into two nonempty disjoint subsets.

**Example 5.1 (Two marginals case)** In previous works [CPT23, EN23] concerning the rate of convergence for the two marginals problem, it was assumed that the cost function must satisfy a nondegeneracy condition, that is,  $D_{x_1, x_2}^2 c$  must be of full rank. A direct consequence of our analysis is that we can provide a lower bound (the upper bound does not depend on such a condition) for costs for which the nondegeneracy condition fails. Let  $r$  be the rank of  $D_{x_1, x_2}^2 c$  at the point where the nondegeneracy condition fails, then the signature of  $\bar{g}$  at this point is given by  $(r, r, 2d - 2r)$  meaning that locally the support of the optimal  $\gamma_0$  is at most  $(2d - r)$ -dimensional. Thus, the bounds become

$$\frac{r}{2} \varepsilon \log(1/\varepsilon) - C_* \varepsilon \leq \text{OT}_\varepsilon - \text{OT}_0 \leq \frac{d}{2} \varepsilon \log(1/\varepsilon) + C^* \varepsilon,$$

for some constants  $C_*, C^* > 0$ . Notice that if  $D_{x_1, x_2}^2 c$  has full rank, then  $r = d$  and we retrieve the matching bound results of [CPT23, EN23].

**Example 5.2 (Two marginals case and unequal dimension)** Consider now the two marginals case but unequal dimensional, that is, for example,  $d_1 > d_2$ . Then, if  $D_{x_1, x_2}^2 c$  has full rank, that is,  $r = d_2$ , we obtain a matching bound depending only on the lower-dimensional marginal

$$\frac{d_2}{2} \varepsilon \log(1/\varepsilon) - C_* \varepsilon \leq \text{OT}_\varepsilon - \text{OT}_0 \leq \frac{d_2}{2} \varepsilon \log(1/\varepsilon) + C^* \varepsilon,$$

for some constants  $C_*, C^* > 0$ . If  $\mu_1$  is absolutely continuous with respect to  $\mathcal{H}^{d_1}$  on some smooth sub-manifold of dimension  $d_1$ , then any OT plan would be concentrated on a set of Hausdorff dimension no less than  $d_1$ , and thus the upper bound given in [EN23, Theorem 3.8] would be  $\frac{d_1}{2} \varepsilon \log(1/\varepsilon) + O(\varepsilon)$ , which is strictly worse than our estimate.

**Example 5.3 (Negative harmonic cost)** Consider the cost  $c(x_1, \dots, x_m) = h(\sum_{i=1}^m x_i)$  where  $h$  is  $\mathcal{C}^2$  and  $D^2 h > 0$ . Assuming that the marginals have finite second moments, when  $h(x) = |x|^2$ , this kind of cost is equivalent to the harmonic negative cost that is  $c(x_1, \dots, x_m) = -\sum_{i < j} |x_i - x_j|^2$  (here  $|\cdot|$  denotes the standard euclidean norm) (see [DGN17] for more details). It follows now that the signature of the metric  $\bar{g}$  is  $(d, (m-1)d, 0)$ ; thus, the bounds between  $\text{MOT}_\varepsilon$  and  $\text{MOT}_0$  that we obtain are

$$\frac{d}{2} \varepsilon \log(1/\varepsilon) - C_* \varepsilon \leq \text{MOT}_\varepsilon - \text{MOT}_0 \leq \frac{1}{2} ((m-1)d) \varepsilon \log(1/\varepsilon) + C^* \varepsilon,$$

for some constants  $C_*, C^* > 0$ . We remark that it is known from [DGN17, Pas12] that a transport plan  $\gamma_0$  is optimal if and only if it is supported on the set  $\{(x_1, \dots, x_m) \mid \sum_{i=1}^m x_i = l\}$ , where  $l \in \mathbb{R}^d$  is any constant and there exist solutions whose support has dimension exactly  $(m-1)d$ .

**Example 5.4 (Gangbo–Świąch cost and Wasserstein barycenter)** Suppose that  $c(x_1, \dots, x_m) = \sum_{i < j} |x_i - x_j|^2$ , known as the Gangbo–Świąch cost [GŚ98]. Notice that the cost is equivalent to  $c(x_1, \dots, x_m) = h(\sum_{i=1}^m x_i)$  where  $h$  is  $\mathcal{C}^2$  and  $D^2 h < 0$ , then the signature of  $\bar{g}$  is  $((m-1)d, d, 0)$  and we have a matching bound

$$\frac{1}{2} \left( (m-1)d \right) \varepsilon \log(1/\varepsilon) - C_* \varepsilon \leq \text{MOT}_\varepsilon - \text{MOT}_0 \leq \frac{1}{2} \left( (m-1)d \right) \varepsilon \log(1/\varepsilon) + C^* \varepsilon.$$

Notice now that considering the  $\text{MOT}_0$  problem with a cost  $c(x_1, \dots, x_m) = \sum_i |x_i - T(x_1, \dots, x_m)|^2$ , where  $T(x_1, \dots, x_m) = \sum_{i=1}^m \lambda_i x_i$  is the Euclidean barycenter, is equivalent to the  $\text{MOT}_0$  with the Gangbo–Świąch cost and the matching bound above still holds. Moreover, the multi-marginal problem with this particular cost has been shown [AC11] to be equivalent to the Wasserstein barycenter, that is,  $T_\# \gamma_0 = \nu$  is the barycenter of  $\mu_1, \dots, \mu_m$ .

## References

- [Abr+17] I. Abraham, R. Abraham, M. Bergounioux, and G. Carlier, *Tomographic reconstruction from a few views: a multi-marginal optimal transport approach*. Appl. Math. Optim. 75(2017), no. 1, 55–73.
- [AC11] M. Agueh and G. Carlier, *Barycenters in the Wasserstein space*. SIAM J. Math. Anal. 43(2011), no. 2, 904–924. <https://doi.org/10.1137/100805741>
- [BHP13] M. Beiglböck, P. Henry-Labordere, and F. Penkner, *Model-independent bounds for option prices – a mass transport approach*. Finance Stochast. 17(2013), no. 3, 477–501.
- [Ben+15] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, *Iterative Bregman projections for regularized transportation problems*. SIAM J. Sci. Comput. 37(2015), no. 2, A1111–A1138. <https://doi.org/10.1137/141000439>
- [BCN19] J.-D. Benamou, G. Carlier, and L. Nenna, *Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm*. Numer. Math. 142(2019), no. 1, 33–54. <https://doi.org/10.1007/s00211-018-0995-x>
- [BK18] J. Bigot and T. Klein, *Characterization of Barycenters in the Wasserstein space by averaging optimal transport maps*. ESAIM Probab. Stat. 22(2018), 35–57.
- [BL92] J. M. Borwein and A. S. Lewis, *Decomposition of multivariate functions*. Canad. J. Math. 44(1992), no. 3, 463–482. <https://doi.org/10.4153/CJM-1992-030-9>
- [BLN94] J. M. Borwein, A. S. Lewis, and R. D. Nussbaum, *Entropy minimization, DAD problems, and doubly stochastic kernels*. J. Funct. Anal. 123(1994), no. 2, 264–307. <https://doi.org/10.1006/jfan.1994.1089>
- [Bre89] Y. Brenier, *The least action principle and the related concept of generalized flows for incompressible perfect fluids*. J. Amer. Math. Soc. 2(1989), no. 2, 225–255.
- [BDG12] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi, *Optimal-transport formulation of electronic density-functional theory*. Phys. Rev. A 85(2012), no. 6, 062502.
- [Car03] G. Carlier, *On a class of multidimensional optimal transportation problems*. J. Convex Anal. 10(2003), no. 2, 517–529.
- [Car22] G. Carlier, *On the linear convergence of the multimarginal Sinkhorn algorithm*. SIAM J. Optim. 32(2022), 786–794. <https://doi.org/10.1137/21M1410634>
- [CCG16] G. Carlier, V. Chernozhukov, and A. Galichon, *Vector quantile regression: an optimal transport approach*. Ann. Stat. 44(2016), no. 3, 1165–1192.

- [Car+17] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*. SIAM J. Math. Anal. 49(2017), no. 2, 1385–1418. <https://doi.org/10.1137/15M1050264>
- [CE10] G. Carlier and I. Ekeland, *Matching for teams*. Econom. Theory 42(2010), no. 2, 397–418.
- [CN08] G. Carlier and B. Nazaret, *Optimal transportation for the determinant*. ESAIM Control Optim. Calc. Var. 14(2008), no. 4, 678–698. <https://doi.org/10.1051/cocv:2008006>
- [CPT23] G. Carlier, P. Pegon, and L. Tamanini, *Convergence rate of general entropic optimal transport costs*. Calc. Var. 62(2023), no. 4, 116. <https://doi.org/10.1007/s00526-023-02455-0>
- [CDD15] M. Colombo, L. De Pascale, and S. Di Marino, *Multimarginal optimal transport maps for one-dimensional repulsive costs*. Canad. J. Math. 67(2015), no. 2, 350–368. <https://doi.org/10.4153/CJM-2014-011-x>
- [CS16] M. Colombo and F. Stra, *Counterexamples in multimarginal optimal transport with coulomb cost and spherically symmetric data*. Math. Models Methods Appl. Sci. 26(2016), no. 6, 1025–1049. <https://doi.org/10.1142/S021820251650024X>
- [CFK13] C. Cotar, G. Friesecke, and C. Klüppelberg, *Density functional theory and optimal transportation with coulomb cost*. Commun. Pure Appl. Math. 66(2013), no. 4, 548–599. <https://doi.org/10.1002/cpa.21437>
- [Csi75] I. Csiszar, *I-divergence geometry of probability distributions and minimization problems*. Ann. Probab. 3(1975), no. 1, 146–158. <https://doi.org/10.1214/aop/1176996454>
- [Cut13] M. Cuturi, *Sinkhorn distances: lightspeed computation of optimal transport*, Advances in Neural Information Processing Systems, 26, Curran Associates, Red Hook, NY, 2013.
- [DGN17] S. Di Marino, A. Gerolin, and L. Nenna, *Optimal transportation theory with repulsive costs*. In: Topological optimization and optimal transport, Radon Series on Computational and Applied Mathematics, De Gruyter, Berlin–Boston, 2017, pp. 204–256, Chapter 9. Comment: Survey for the special volume for RICAM (Special Semester on New Trends in Calculus of Variations). <https://www.degruyter.com/view/books/9783110430417/9783110430417-010/9783110430417-010.xml>
- [DS14a] Y. Dolinsky and H. M. Soner, *Martingale optimal transport and robust hedging in continuous time*. Probab. Theory Relat. Fields 160(2014), no. 1, 391–427.
- [DS14b] Y. Dolinsky and H. M. Soner, *Robust hedging with proportional transaction costs*. Finance Stochast. 18(2014), no. 2, 327–347.
- [EN23] S. Eckstein and M. Nutz, *Convergence rates for regularized optimal transport via quantization*. Math. Oper. Res. (2023). Comment: Fixed a typo in Theorem 3.8. <https://doi.org/10.1287/moor.2022.0245>
- [Enn+22] H. Ennaji, Q. Mérigot, L. Nenna, and B. Pass, *Robust risk management via multi-marginal optimal transport* (2022). <https://doi.org/10.48550/ARXIV.2211.07694>
- [Fed96] H. Federer, *Geometric measure theory*. In: B. Eckmann and B. L. Waerden (eds.), Classics in mathematics, Springer, Berlin–Heidelberg, 1996. <https://doi.org/10.1007/978-3-642-62010-2>
- [FG97] H. Föllmer and N. Gantert, *Entropy minimization and Schrödinger processes in infinite dimensions*. Ann. Probab. 25(1997), no. 2, 901–926. <https://doi.org/10.1214/aop/1024404423>
- [FL89] J. Franklin and J. Lorenz, *On the scaling of multidimensional matrices*. Linear Algebra Appl. 114–115(1989), 717–735. Special Issue Dedicated to Alan J. Hoffman. [https://doi.org/10.1016/0024-3795\(89\)90490-4](https://doi.org/10.1016/0024-3795(89)90490-4)
- [FGG23] G. Friesecke, A. Gerolin, and P. Gori-Giorgi, *The strong-interaction limit of density functional theory*. In: E. Cancès and G. Friesecke (eds.), Density functional theory: modeling, mathematical analysis, computational methods, and applications, Mathematics and Molecular Modeling, Springer, Cham, 2023, pp. 183–266. [https://doi.org/10.1007/978-3-031-22340-2\\_4](https://doi.org/10.1007/978-3-031-22340-2_4)
- [GŚ98] W. Gangbo and A. Świąch, *Optimal maps for the multidimensional Monge–Kantorovich problem*. Commun. Pure Appl. Math. 51(1998), no. 1, 23–45. [https://doi.org/10.1002/\(SICI\)1097-0312\(199801\)51:1\\$<\\$23::AID-CPA2\\$>\\$3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-0312(199801)51:1$<$23::AID-CPA2$>$3.0.CO;2-H)
- [GKR20] A. Gerolin, A. Kausamo, and T. Rajala, *Multi-marginal entropy-transport with repulsive cost*. Calc. Var. 59(2020), no. 3, 90. <https://doi.org/10.1007/s00526-020-01735-3>
- [GN22] P. Ghosal and M. Nutz, *On the convergence rate of Sinkhorn’s algorithm*. Preprint, 2022. <https://doi.org/10.48550/arXiv.2212.06000>

- [Haa+21] I. Haasler, R. Singh, Q. Zhang, J. Karlsson, and Y. Chen, *Multi-marginal optimal transport and probabilistic graphical models*. IEEE Trans. Inf. Theory 67(2021), no. 7, 4647–4668. <https://doi.org/10.1109/TIT.2021.3077465>
- [Hei02] H. Heinrich, *Problème de Monge pour  $n$  Probabilités*. C. R. Math. Acad. Sci. Paris 334(2002), no. 9, 793–795. [https://doi.org/10.1016/S1631-073X\(02\)02341-5](https://doi.org/10.1016/S1631-073X(02)02341-5)
- [KM10] Y.-H. Kim and R. J. McCann, *Continuity, curvature, and the general covariance of optimal transportation*. J. Eur. Math. Soc. 12(2010), no. 4, 1009–1040. <https://doi.org/10.4171/jems/221>
- [KP14] Y.-H. Kim and B. Pass, *A general condition for Monge solutions in the multi-marginal optimal transport problem*. SIAM J. Math. Anal. 46(2014), no. 2, 1538–1550. <https://doi.org/10.1137/130930443>
- [KP15] Y.-H. Kim and B. Pass, *Multi-marginal optimal transport on Riemannian manifolds*. Amer. J. Math. 137(2015), no. 4, 1045–1060. <https://doi.org/10.1353/ajm.2015.0024>
- [Léo14] C. Léonard, *A survey of the Schrödinger problem and some of its connections with optimal transport*. Discrete Contin. Dyn. Syst. 34(2014), no. 4, 1533. <https://doi.org/10.3934/dcds.2014.34.1533>
- [MS23] H. Malamut and M. Sylvestre, *Convergence rates of the regularized optimal transport: disentangling suboptimality and entropy*. Preprint, 2023. <https://doi.org/10.48550/arXiv.2306.06940>
- [MG20] S. D. Marino and A. Gerolin, *An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm*. J. Sci. Comput. 85(2020), no. 2, 27. <https://doi.org/10.1007/s10915-020-01325-7>
- [MP17] A. Moameni and B. Pass, *Solutions to multi-marginal optimal transport problems concentrated on several graphs*. ESAIM Control Optim. Calc. Var. 23(2017), no. 2, 551–567. <https://doi.org/10.1051/cocv/2016003>
- [Nen16] L. Nenna, *Numerical methods for multi-marginal optimal transportation*. Ph.D. thesis, Université Paris sciences et lettres, 2016.
- [Nut] M. Nutz, *Introduction to entropic optimal transport*.
- [NW22] M. Nutz and J. Wiesel, *Stability of Schrödinger potentials and convergence of Sinkhorn's algorithm*. Preprint, 2022. <https://doi.org/10.48550/arXiv.2201.10059>
- [Pas11] B. Pass, *Uniqueness and Monge solutions in the multimarginal optimal transportation problem*. SIAM J. Math. Anal. 43(2011), no. 6, 2758–2775. <https://doi.org/10.1137/100804917>
- [Pas12] B. Pass, *On the local structure of optimal measures in the multi-marginal optimal transportation problem*. Calc. Var. 43(2012), no. 3, 529–536. <https://doi.org/10.1007/s00526-011-0421-z>
- [Pas15] B. Pass, *Multi-marginal optimal transport: Theory and applications*. ESAIM Math. Model. Numer. Anal. 49(2015), no. 6, 1771–1790. <https://doi.org/10.1051/m2an/2015020>
- [PV21a] B. Pass and A. Vargas-Jiménez, *Monge solutions and uniqueness in multi-marginal optimal transport via graph theory*. Preprint, 2021. <https://arxiv.org/abs/2104.09488>
- [PV21b] B. W. Pass and A. Vargas-Jiménez, *Multi-marginal optimal transportation problem for cyclic costs*. SIAM J. Math. Anal. 53(2021), no. 4, 4386–4400. <https://doi.org/10.1137/19M130889X>
- [PC19] G. Peyré and M. Cuturi, *Computational optimal transport: with applications to data science*. Found. Trends Mach. Learn. 11(2019), nos. 5–6, 355–607. <https://doi.org/10.1561/22000000073>
- [Rab+11] J. Rabin, G. Peyré, J. Delon, and M. Bernot, *Wasserstein Barycenter and its application to texture mixing*. In: International conference on scale space and variational methods in computer vision, Springer, Berlin–Heidelberg, 2011, pp. 435–446.
- [Rén59] A. Rényi, *On the dimension and entropy of probability distributions*. Acta Math. Acad. Sci. Hungarica 10(1959), no. 1, 193–215. <https://doi.org/10.1007/BF02063299>
- [RT98] L. Rüschendorf and W. Thomsen, *Closedness of sum spaces and the generalized Schrödinger problem*. Theory Probab. Appl. 42(1998), no. 3, 483–494. <https://doi.org/10.1137/S0040585X97976301>
- [San15] F. Santambrogio, *Optimal transport for applied mathematicians: calculus of variations, PDEs, and modeling*, Progress in Nonlinear Differential Equations and Their Applications, 87, Springer, Cham, 2015. <https://doi.org/10.1007/978-3-319-20828-2>
- [Sin64] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*. Ann. Math. Stat. 35(1964), no. 2, 876–879. <https://doi.org/10.1214/aoms/1177703591>

- [TJK22] N. G. Trillos, M. Jacobs, and J. Kim, *The multimarginal optimal transport formulation of adversarial multiclass classification*. Preprint, 2022. [10.48550/ARXIV.2204.12676](https://arxiv.org/abs/2204.12676).
- [You82] L.-S. Young, *Dimension, entropy and Lyapunov exponents*. *Ergodic Theory Dynam. Systems* 2(1982), no. 1, 109–124. <https://doi.org/10.1017/S0143385700009615>

*LMO and Inria Saclay, ParMA, Université Paris-Saclay, Orsay, France*

*e-mail:* [luca.nenna@universite-paris-saclay.fr](mailto:luca.nenna@universite-paris-saclay.fr)

*CEREMADE and Inria Paris, MOKAPLAN, Université Paris-Dauphine, Paris, France*

*e-mail:* [pegon@ceremade.dauphine.fr](mailto:pegon@ceremade.dauphine.fr)