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

Luca Nenna* and Paul Pegon†

September 1, 2023

We investigate the convergence rate of multi-marginal optimal transport costs that are regularized with the Boltzmann-Shannon entropy, as the noise parameter ε 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 semi-concave costs for a finer estimate, and lower bounds for $\mathscr{C}^2$ 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 non-degenerate costs. We obtain in particular matching bounds in some typical situations where the optimal plan is deterministic.

Keywords. optimal transport, multi-marginal optimal transport, entropic regularization, Schrödinger problem, convex analysis, Rényi dimension.
2020 Mathematics Subject Classification. Primary: 49Q22 ; Secondary: 49N15, 94A17, 49K40.

Contents

1 Introduction **3**

2 Preliminaries **5**

*Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France. email: luca.nenna@universite-paris-saclay.fr
†CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, Mokaplan, Inria Paris, 75016 Paris, France. email: pegon@ceremade.dauphine.fr

3 Upper bounds **8**
3.1 Upper bound for locally Lipschitz costs 8
3.2 Upper bound for locally semi-concave costs 11

4 Lower bound for $\mathscr{C}^2$ costs with a signature condition **15**

5 Examples and matching bound **18**

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 supscript d when $d = N$;
ω_d	d-dimensional volume of $B_1^d(0)$;
X_i	a subset of $\mathbb{R}^N$ for any index $i \in \{1, \ldots, m\}$
$\boldsymbol{X}$	product $X_1 \times \ldots \times X_m$ whenever $(X_i)_{1 \leq i \leq m}$ is a family of m subsets of $\mathbb{R}^N$;
$\boldsymbol{A}_{-i}$	product $\prod_{1 \leq j \leq m, j \neq i} A_j$ if $\boldsymbol{A} = A_1 \times \ldots \times A_m \subseteq \boldsymbol{X}$ and $i \in \{1, \ldots, m\}$;
$x_i, x, \boldsymbol{x}$	a point in X_i, in some $X_j, j \in \{1, \ldots, m\}$, and in $\boldsymbol{X}$ respectively;
x_q	$(x_i)_{i \in q}$ if $q \subseteq \{1, \ldots, m\}$ and $\boldsymbol{x} \in \boldsymbol{X}$;
e_i	i-th coordinate map $e_i : \boldsymbol{x} = (x_1, \ldots, x_m) \mapsto x_i$;
$A \Subset B$	$A \subseteq K \subseteq B$ for some compact set K;
$\lvert\cdot\rvert$	Euclidean norm on $\mathbb{R}^N$;
$\lVert\cdot\rVert$	norm on $\mathbb{R}^N \times \ldots \times \mathbb{R}^N$ defined by $\lVert\boldsymbol{x}\rVert = \max_{1 \leq i \leq m} \lvert x_i \rvert$ if $\boldsymbol{x} = (x_1, \ldots, x_m)$;
$B_r(\boldsymbol{x})$	open ball of radius r centered at $\boldsymbol{x} \in (\mathbb{R}^N)^m$ for the above norm;
$\mathscr{C}^{0,1}_{\mathrm{loc}}(X)$	space of real-valued locally Lipschitz functions on X which is a sub-manifold of $\mathbb{R}^N$ or $(\mathbb{R}^N)^m$;
$[f]_{\mathscr{C}^{0,1}(X)}$	Lipschitz constant of $f : X \to \mathbb{R}$ where X is a subset of $\mathbb{R}^N$ or $(\mathbb{R}^N)^m$ for the above norms;
$\mathscr{C}^{1,1}_{\mathrm{loc}}(X)$	space of differentiable real-valued functions on X, a sub-manifold of $\mathbb{R}^N$ or $(\mathbb{R}^N)^m$, with locally Lipschitz differential;
$\mathscr{P}(X)$	space of probability measures on a metric space X;
$\lVert\cdot\rVert_{L^p(\mu)}$	L^p norm induced by a measure μ, where $p \in [1, +\infty]$;
$\operatorname{spt} \mu$	support of the measure μ;
$\mathscr{H}^s_X$	s-dimensional Hausdorff measure on the metric space X endowed with the Borel σ-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 $A \cdot B := \mathrm{Tr}(A^T B)$, for $A, B \in M_N(\mathbb{R})$;
$S_N(\mathbb{R})$	subspace of real symmetric matrices of size $N \times N$;
Δ_P	simplex of P-uples $t = (t_p)_{p \in P}$ such that $t_p \geq 0$ for all $p \in P$ and $\sum_{p \in P} t_p = 1$.

1 Introduction

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

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

where $\Pi(\mu_1, \ldots, \mu_m)$ denotes the set of all probability measures γ having μ_i as i-th marginal, i.e. $(e_i)_\sharp \gamma = \mu_i$ where $e_i : (x_1, \ldots, x_m) \mapsto x_i$, for every $i \in \{1, \ldots, m\}$. 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 and multi-marginal optimal transport) have witnessed a growing interest and they are now an active research topic.

Entropic optimal transport (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 [Cut13; Ben+15]. The simplicity and the good convergence guarantees (see [FL89; 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] and references therein).

As concerns multi-marginal optimal transport (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]. The structure of solutions to the multi-marginal optimal transport problem is a notoriously delicate issue, and is still not well understood, despite substantial efforts on the part of many researchers [GŚ98; Car03; CN08; Hei02; Pas11; Pas12; KP14; KP15; CDD15; CS16; PV21b; MP17; PV21a]; see also the surveys [Pas15] and [DGN15]. Since MOT_ε can be seen a perturbation of MOT_0, it is natural to study the behaviour as ε vanishes. In this paper we are mainly interested in investigating the rate of convergence of the entropic cost MOT_ε to 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 non-degenerate 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 $\int c \, \mathrm{d}\gamma$ 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 multi-marginal optimal transport an upper bound has been already established in [EN22], which depends on the number of marginals and the quantization dimension of the optimal solutions to (MOT_ε) with $\varepsilon = 0$. Here we provide a improved, smaller, upper bound, which will depend only on the marginals, but not on the optimal transport plans for the un-regularized 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].

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 semi-concave 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 semi-concave 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 $C^* \in \mathbb{R}_+$

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

We stress that this upper bound is smaller or equal than the one provided in [EN22, 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 optimal transport plan. Thus D must be greater or equal than 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 (MOT_ε) we have

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

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 un-regularized problem (MOT_ε) 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 [EN22] where only an upper bound depending on the quantization dimension of the solution to the un-regularized 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 κ depending on a signature condition (see ($\mathrm{PS}(\kappa)$)) on the second mixed derivatives of the cost, the lower bound can be summerized as follows

$$\boxed{\mathsf{MOT}_\varepsilon \geq \mathsf{MOT}_0 + \frac{\kappa}{2}\varepsilon\log(1/\varepsilon) - C_*\varepsilon.} \tag{1.2}$$

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

The paper is organized as follows: in Section 2 we recall the multi-marginal optimal transport 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 Theorem 3.2 and Theorem 3.7. 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 μ_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 $c : X_1 \times X_2 \times \dots \times X_m \to \mathbb{R}_+$, the multi-marginal optimal transport problem consists in solving the following optimization problem

$$\mathsf{MOT}_0 := \inf_{\gamma \in \Pi(\mu_1, \dots, \mu_m)} \int_{\boldsymbol{X}} c(x_1, \dots, x_m) \, \mathrm{d}\gamma \tag{MOT}$$

where $\boldsymbol{X} := X_1 \times X_2 \times \dots \times X_m$ and $\Pi(\mu_1, \dots, \mu_m)$ denotes the set of probability measures on $\boldsymbol{X}$ whose marginals are the μ_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 optimal transport community is rather focused on uniqueness ans 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 $(T_i)_\sharp \mu_1 = \mu_i$ for $i \in \{1, \dots, m\}$, in which case this function induces a solution *à la Monge*, that is $\gamma = (\mathrm{Id}, T_2, \dots, T_m)_\sharp \mu_1$.

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 [Pas11; Pas12] Pass 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 non-empty disjoint subsets: $p = \{p_-, p_+\} \in P$ if $p_- \bigcup p_+ = \{1, \dots, m\}$, $p_- \bigcap p_+ = \emptyset$ and $p_-, p_+ \neq \emptyset$. Then for each $p \in P$ we denote by g_p the bilinear form on the tangent bundle $T\boldsymbol{X}$

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

for every $p, q \subseteq \{1, \dots, m\}$, and $D^2_{x_i x_j} c := \sum_{\alpha_i, \alpha_j} \frac{\partial^2 c}{x_i^{\alpha_i} x_k^{\alpha_k}} \, \mathrm{d}x_i^{\alpha_i} \otimes \mathrm{d}x_j^{\alpha_j}$, defined for every i, j on the whole tangent bundle $T\mathbf{X}$. Define

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

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 γ a solution to (MOT) and suppose that the signature of some $g \in G_c$ at a point $\boldsymbol{x} \in \boldsymbol{X}$ is (d^+, d^-, d^0), that is the number of positive, negative and zero eigenvalues. Then, there exists a neighbourhood $N_{\boldsymbol{x}}$ of $\boldsymbol{x}$ such that $N_{\boldsymbol{x}} \bigcap \operatorname{spt} \gamma$ is contained in a Lipschitz sub-manifold of $\boldsymbol{X}$ 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 $\operatorname{spt} \gamma$ 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 non-degenerate 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 $D^2_{x_1 x_2} c$. Notice that this will help us to generalize the results established in [CPT23; EN22] 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

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

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 $i \in \{1, \dots, m\}$

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

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

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

where $\mathrm{Ent}(\cdot|\otimes_{i=1}^m \mu_i)$ 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

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

The fact that q is a probability measure ensures that $\mathrm{Ent}(p\,|\,q) \geq 0$. The dual problem of ($\mathrm{MOT}_\varepsilon$) reads as

$$\mathsf{MOT}_\varepsilon = \varepsilon + \sup\left\{\sum_{i=1}^m \int_{X_i} \phi_i(x_i)\,\mathrm{d}\mu_i - \varepsilon \int_{\boldsymbol{X}} e^{\frac{\sum_{i=1}^m \phi_i(x_i) - c(\boldsymbol{x})}{\varepsilon}}\,\mathrm{d}\otimes_{i=1}^m \mu_i \mid \phi_i \in \mathscr{C}_b(X_i)\right\}, \tag{MD_ε}$$

which is invariant by $(\phi_1, \ldots, \phi_m) \mapsto (\phi_1 + \lambda_1, \ldots, \phi_m + \lambda_m)$ where $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$ and $\sum_{i=1}^m \lambda_i = 0$, see [Léo14; NW22; MG20] for some recent presentations. It admits an equivalent "log-sum-exp" form:

$$\mathsf{MOT}_\varepsilon = \sup\left\{\sum_{i=1}^m \int_{X_i} \phi_i(x_i)\,\mathrm{d}\mu_i - \varepsilon \log\left(\int_{\boldsymbol{X}} e^{\frac{\sum_{i=1}^m \phi_i(x_i) - c(\boldsymbol{x})}{\varepsilon}}\,\mathrm{d}\otimes_{i=1}^m \mu_i\right) \mid \phi_i \in \mathscr{C}_b(X_i)\right\}, \tag{MD'_ε}$$

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

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

$$\lim_{\varepsilon \to 0} \mathsf{MOT}_\varepsilon = \mathsf{MOT}_0.$$

By the direct method in the calculus of variations and strict convexity of the entropy, one can show that (MOT_ε) admits a unique solution γ_ε, called *optimal entropic plan*. Moreover, there exist m real-valued Borel functions ϕ_i^ε such that

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

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

$$\mathsf{MOT}_\varepsilon = \sum_{i=1}^m \int_{X_i} \phi_i^\varepsilon\ \mathrm{d}\mu_i \tag{2.4}$$

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

$$\phi_i(x_i) = -\varepsilon \log \int_{\boldsymbol{X}_{-i}} e^{\frac{\oplus_{1\le j\le m, j\neq i}\phi_j^\varepsilon - c(\boldsymbol{x})}{\varepsilon}} \, \mathrm{d} \oplus_{1\le j\le m, j\neq i} \quad \text{for } \mu_i\text{-a.e. } x_i, \tag{2.5}$$

where $\boldsymbol{X}_{-i} = \prod_{1\le j\le m, j\neq i}^m X_j$. 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 semi-concave (in particular $\mathscr{C}^2$) 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 μ is a probability measure over a metric space X, we set for every $\delta > 0$,

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

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

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

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

$$H_\delta(\mu) \le \log N_\delta(\operatorname{spt}\mu). \tag{3.1}$$

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

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

Theorem 3.2. *Assume that for $i \in \{1, \dots, m\}$, $\mu_i \in \mathscr{P}(X_i)$ is a compactly supported measure on a Lipschitz sub-manifold X_i of dimension d_i and $c \in \mathscr{C}^{0,1}_{\mathrm{loc}}(\boldsymbol{X})$, then*

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

Proof. Given an optimal plan γ_0 for 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[1] $\operatorname{diam}(A_i^n) \leq \delta$ for every $n \in \mathbb{N}$, and set

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

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 μ_i's. Besides, $\gamma_\delta(\boldsymbol{A}) = \gamma_0(\boldsymbol{A})$ for every $\boldsymbol{A} = \prod_{i=1}^m A_i^{n_i}$ where $n \in \mathbb{N}^m$, and for $\otimes_{i=1}^m \mu_i$-almost every $\boldsymbol{x} = (x_1, \dots, x_m) \in \prod_{i=1}^m A_i^{n_i}$,

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

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

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

[1] 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 $\delta_0 > 0$.

$$
\begin{aligned}
&= \sum_{n\in\mathbb{N}^m} \gamma_0(A_1^{n_1}\times\ldots\times A_m^{n_m})\log\left(\frac{\gamma_0(A_1^{n_1}\times\ldots\times A_m^{n_m})}{\mu_m(A_m^{n_m})}\right)\\
&\qquad + \sum_{j=1}^{m-1}\sum_{n_j\in\mathbb{N}} \gamma_0\left(\prod_{i=1}^{j-1} X_i\times A_j^{n_j}\times\prod_{i=j+1}^{m} X_i\right)\mu_j(A_j^{n_j})\log(1/\mu_j(A_j^{n_j}))\\
&\le \sum_{j=1}^{m-1}\sum_{n_j\in\mathbb{N}} \mu_j(A_j^{n_j})\log(1/\mu_j(A_j^{n_j})).
\end{aligned}
$$

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

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

Since the μ_i's have compact support and c is locally Lipschitz, for δ small enough there exists $L\in(0,+\infty)$ not depending on δ such that $[c]_{\mathscr{C}^{0,1}(\boldsymbol{A})}\le L$ for every $\boldsymbol{A}\in\mathcal{A}:=\{\prod_{i=1}^m A_i^{n_i}\mid n\in\mathbb{N}^m,\mu_1(A_1^{n_1})\ldots\mu_m(A_m^{n_m})>0\}$. Notice that the ∞-Wasserstein distance (with respect to the norm $\|\cdot\|$)[2] satisfies $W_\infty(\gamma_\delta,\gamma_0)\le\delta$, thus taking γ_δ as competitor in (MOT$_\varepsilon$) we obtain:

$$
\begin{aligned}
\mathsf{MOT}_\varepsilon &\le \int c\,\mathrm{d}\gamma_\delta+\varepsilon\sum_{j\le m-1} H_\delta(\mu_j)+\varepsilon\\
&= \mathsf{MOT}_0+\sum_{\boldsymbol{A}\in\mathcal{A}}\int_{\boldsymbol{A}} c\,\mathrm{d}(\gamma_\delta-\gamma_0)+\varepsilon\sum_{j\le m-1} H_\delta(\mu_j)+\varepsilon\\
&\le \mathsf{MOT}_0+\sum_{\boldsymbol{A}\in\mathcal{A}} LW_\infty(\gamma_\delta\llcorner\boldsymbol{A},\gamma_0\llcorner\boldsymbol{A})\gamma_0(\boldsymbol{A})+\varepsilon\sum_{j\le m-1} H_\delta(\mu_j)+\varepsilon\\
&\le \mathsf{MOT}_0+L\delta+\varepsilon\sum_{j\le m-1}\frac{H_\delta(\mu_j)}{\log(1/\delta)}\log(1/\delta)+\varepsilon.
\end{aligned}
\tag{3.3}
$$

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

$$
\mathsf{MOT}_\varepsilon\le\mathsf{MOT}_0+\Big(\sum_{j\le m-1} d_j\Big)\varepsilon\log(1/\varepsilon)+C^*\varepsilon. \qquad \square
$$

Remark 3.3. If the μ_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

$$
\mathsf{MOT}_\varepsilon\le\mathsf{MOT}_0+\left(\sum_{i=1}^m\overline{\dim}_R(\mu_i)-\max_{j\in\{1,\ldots,m\}}\overline{\dim}_R(\mu_j)\right)\varepsilon\log(1/\varepsilon)+o(\varepsilon\log(1/\varepsilon)). \tag{3.4}
$$

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

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 semi-concave costs

We provide now a finer upper bound under the additional assumptions that the X_i's are $\mathscr{C}^2$ sub-manifolds of $\mathbb{R}^N$, c is locally semi-concave as in Definition 3.4 (which is the case when $c \in \mathscr{C}^2(\boldsymbol{X}, \mathbb{R}_+)$), and the μ_i's are measures in $L^\infty(\mathscr{H}^{d_i}_{X_i})$ with compact support in X_i.

Definition 3.4. A function $f : X \to \mathbb{R}$ defined on a $\mathscr{C}^2$ 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 $\mathscr{C}^2$ diffeomorphism) $\psi : U \to \Omega$ where $U \subseteq X$ is an open neighborhood of x and Ω is an open convex subset of $\mathbb{R}^d$, such that $f \circ \boldsymbol{\psi}^{-1}$ is λ-concave for some $\lambda \in \mathbb{R}$, meaning $f \circ \boldsymbol{\psi}^{-1} - \lambda \frac{|\cdot|^2}{2}$ is concave on Ω.

Lemma 3.5 (Local semiconcavity and covering)**.** *Let $c : \boldsymbol{X} \to \mathbb{R}_+$ be a locally semiconcave cost function and $(\phi_i)_{1\le i\le m} \in \prod_{\le i\le m} \mathscr{C}(K_i)$ 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 $(U_i^j)_{1\le j\le J}$ of K_i together with bi-Lipschitz local charts $\psi_i^j : U_i^j \to \Omega_i^j$ satisfying the following properties, having set $\boldsymbol{\Omega^j} := \prod_{1\le i\le m} \Omega_i^{j_i}$ and $\boldsymbol{\psi^j} := (\psi_1^{j_1}, \dots, \psi_m^{j_m})$ for every $\boldsymbol{j} = (j_1, \dots, j_m) \in \{1, \dots, J\}^m$:*

(i) for every $\boldsymbol{j} \in \{1, \dots, J\}^m$, $c \circ (\boldsymbol{\psi^j})^{-1}$ is λ-concave on $\boldsymbol{\Omega^j}$,

(ii) for every $(i, j) \in \{1, \dots, m\} \times \{1, \dots, J\}$, $\phi_i \circ (\psi_i^j)^{-1}$ is λ-concave on Ω_i^j.

In particular, all the ϕ_i's are locally semiconcave.

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

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

and we see that it is λ-concave as an infimum of λ-concave functions. □

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

Lemma 3.6 ([CPT23, Lemma 3.6]). *Let $f : \Omega \to \mathbb{R}$ be a λ-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:*

$$\int_\Omega \sup_{y \in B_r(x)\cap\Omega} |f(y) - (f(x) + \nabla f(x)\cdot(y-x))| \,\mathrm{d}x \leq Cr^2 \mathscr{H}^{d-1}(\partial\Omega)([f]_{\mathscr{C}^{0,1}(\Omega)} + \lambda \operatorname{diam}(\Omega)). \tag{3.5}$$

We may now state the main result of this section.

Theorem 3.7. *Let $c \in \mathscr{C}^2(\boldsymbol{X})$ and assume that for every $i \in \{1, \dots, m\}$, $X_i \subseteq \mathbb{R}^N$ is a $\mathscr{C}^2$ sub-manifold of dimension d_i and $\mu_i \in L^\infty(\mathscr{H}^{d_i}_{X_i})$ is a probability measure compactly supported in X_i. Then there exists constants $\varepsilon_0, C^* \geq 0$ such that for $\varepsilon \in (0, \varepsilon_0]$*

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

Proof. The measures μ_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 $(\phi_i)_{1\leq i\leq m} \in \prod_{1\leq i\leq m} \mathscr{C}(K_i)$ a 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,

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

For every $i \in \{1, \dots, m\}$ we consider the coverings $(U_i^j)_{1\leq j\leq J}$ and bi-Lipschitz local charts $\psi_i^j : U_i^j \to \Omega_i^j$ for $j \in \{1, \dots, J\}$ provided by Lemma 3.5 and we notice by compactness that there exists open subsets $\tilde{U}_i^j \Subset U_i^j$ such that for a small $\delta_0 > 0$, the δ_0-neighbourhood of $\tilde{\Omega}_i^j := \psi_i^j(\tilde{U}_i^j)$ is included in Ω_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 γ_δ of γ_0 built in the proof of Theorem 3.2, as well as some $\kappa_\delta \in \Pi(\gamma_0, \gamma_\delta)$ such that $\sup_{(\boldsymbol{x_0},\boldsymbol{x})\in\operatorname{spt}\kappa_\delta} \|\boldsymbol{x_0} - \boldsymbol{x}\| \leq \delta$. For every $\boldsymbol{j} = (j_1, \dots, j_m) \in \{1, \dots, J\}^m$, we set $E^{\boldsymbol{j}} := E \circ (\boldsymbol{\psi^j})^{-1}$, $U^{\boldsymbol{j}} := \prod_{i=1}^m U_i^{j_i}$ and $\tilde{U}^{\boldsymbol{j}} := \prod_{i=1}^m \tilde{U}_i^{j_i}$, and we write:

$$\begin{aligned} \int_{\boldsymbol{X}} c \,\mathrm{d}\gamma_\delta - \int_{\boldsymbol{X}} c \,\mathrm{d}\gamma_0 &= \int_{\boldsymbol{U}} E \,\mathrm{d}\gamma_\delta \\ &= \int_{\boldsymbol{U}\times\boldsymbol{U}} E(\boldsymbol{x}) \,\mathrm{d}\kappa_\delta(\boldsymbol{x_0}, \boldsymbol{x}) \\ &\leq \sum_{\boldsymbol{j}\in\{1,\dots,J\}^m} \int_{(\boldsymbol{x_0},\boldsymbol{x})\in\tilde{U}^{\boldsymbol{j}}\times\boldsymbol{U}} E(\boldsymbol{x}) \,\mathrm{d}\kappa_\delta(\boldsymbol{x_0}, \boldsymbol{x}) \\ &\leq \sum_{\boldsymbol{j}\in\{1,\dots,J\}^m} \int_{(\boldsymbol{x_0},\boldsymbol{x})\in(U^{\boldsymbol{j}})^2} E^{\boldsymbol{j}}(\boldsymbol{\psi^j}(\boldsymbol{x})) \,\mathrm{d}\kappa_\delta(\boldsymbol{x_0}, \boldsymbol{x}). \end{aligned}$$

Notice that for every $\boldsymbol{j} \in \{1,\dots,J\}^m$ and γ_0-a.e. $\boldsymbol{x_0} \in U^{\boldsymbol{j}}$, $E^{\boldsymbol{j}}$ is differentiable at $\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0})$, or equivalently E is differentiable at $\boldsymbol{x_0}$. 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 semi-concave thus differentiable $\mathscr{L}^{d_i}$-a.e. hence ϕ_i is differentiable μ_i-a.e. on U_i^j because $\mu_i \ll \mathscr{H}^{d_i}$ and ψ_i^j is bi-Lipschitz, which in turn implies that $\oplus_{i=1}^m \phi_i$ is differentiable γ_0-a.e. on $U^{\boldsymbol{j}}$ because $\gamma_0 \in \Pi(\mu_1,\dots,\mu_m)$. Moreover, by (3.7) we have $T_{\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0})}E^{\boldsymbol{j}} \equiv 0$ for γ_0-a.e. $\boldsymbol{x_0} \in U^{\boldsymbol{j}}$, where $T_{y_0}f$ designates the first order Taylor expansion $y \mapsto f(y_0) + \nabla f(y_0)\cdot(y-y_0)$ for any function f which is differentiable at y_0. We may then compute:

$$
\begin{aligned}
&\int_{\boldsymbol{X}} c\,\mathrm{d}\gamma_\delta - \int_{\boldsymbol{X}} c\,\mathrm{d}\gamma_0 \\
\leq &\sum_{\boldsymbol{j}\in\{1,\dots,J\}^m} \int_{(\boldsymbol{x_0},\boldsymbol{x})\in(U^{\boldsymbol{j}})^2} \Big(E^{\boldsymbol{j}}(\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x})) - T_{\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0})}E^{\boldsymbol{j}}(\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x}) - \boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0}))\Big)\,\mathrm{d}\kappa_\delta(\boldsymbol{x_0},\boldsymbol{x}) \\
= &\sum_{\boldsymbol{j}=(j_1,\dots,j_m)} \Bigg(\int_{(\boldsymbol{x_0},\boldsymbol{x})\in(U^{\boldsymbol{j}})^2} \Big(c^{\boldsymbol{j}}(\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x})) - T_{\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0})}c^{\boldsymbol{j}}(\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x}) - \boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0}))\Big)\,\mathrm{d}\kappa_\delta(\boldsymbol{x_0},\boldsymbol{x}) \\
&- \sum_{i=1}^m \int_{(x_0,x)\in(U_i^{j_i})^2} \Big(\phi_i^{j_i}(\psi_i^{j_i}(x)) - T_{\psi_i^{j_i}(x_0)}\phi_i^{j_i}(\psi_i^{j_i}(x) - \psi_i^{j_i}(x_0))\Big)\,\mathrm{d}(e_i,e_i)_\sharp\kappa_\delta(x_0,x)\Bigg).
\end{aligned}
\tag{3.8}
$$

Now, since $c^{\boldsymbol{j}}$ is λ-concave on each $\Omega^{\boldsymbol{j}} := \boldsymbol{\psi}^{\boldsymbol{j}}(U^{\boldsymbol{j}})$, whenever $\|\boldsymbol{x_0} - \boldsymbol{x}\| \leq \delta$ we have

$$
c^{\boldsymbol{j}}(\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x})) - T_{\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0})}c^{\boldsymbol{j}}(\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x}) - \boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0})) \leq \lambda\frac{|\boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x}) - \boldsymbol{\psi}^{\boldsymbol{j}}(\boldsymbol{x_0})|^2}{2} \leq \frac{m\lambda L^{\boldsymbol{j}}}{2}\delta^2, \tag{3.9}
$$

where $L^{\boldsymbol{j}} := \max\{\mathrm{Lip}(\boldsymbol{\psi}^{\boldsymbol{j}}), \mathrm{Lip}((\boldsymbol{\psi}^{\boldsymbol{j}})^{-1})\}$. Besides, we may apply Lemma 3.6 to each $\phi_i^{j_i}$ over $\Omega_i^{j_i}$ to get

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

for some constant $C^{\boldsymbol{j}} \in (0,+\infty)$ which does not depend on δ. Reporting (3.9) and (3.10)

in (3.8) yields

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

Finally, we proceed as in the end of the proof of Theorem 3.2, taking γ_δ as competitor in the primal formulation (MOT_ε), so as to obtain

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

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

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

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

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

In this section we consider a cost $c \in \mathscr{C}^2(\boldsymbol{X}, \mathbb{R}_+)$ where $\boldsymbol{X} = X_1 \times \ldots \times X_m$ and we will assume that for every $i \in \{1\ldots,m\}$, the measure μ_i is compactly supported on a $\mathscr{C}^2$ sub-manifold $X_i \subseteq \mathbb{R}^N$ of dimension d_i. We are going to establish a a lower bound in the same form as the fine upper bound of Theorem 3.7, the dimensional constant being this time related to the signature of some bilinear forms, following ideas from [Pas12].

Lemma 4.1. *Let $c \in \mathscr{C}^2(\boldsymbol{X}, \mathbb{R}_+)$ and $(\phi_1,\dots,\phi_m) \in \mathscr{C}(K_1)\times\cdots\mathscr{C}(K_m)$ be a system of c-conjugate functions on subsets $K_i \subseteq X_i$ for every i. We set $E := c - \phi_1 \oplus \ldots \oplus \phi_m$ on $\boldsymbol{K} := K_1 \times \ldots \times K_m$ and we take $\bar{\boldsymbol{x}} \in \boldsymbol{K}$ as well as some $g_{\bar{\boldsymbol{x}}} \in \{g(\bar{\boldsymbol{x}}) \mid g \in G_c\}$ of signature (d^+, d^-, d^0), G_c being defined in (2.1). Then there exists local coordinates around $\bar{\boldsymbol{x}}$, i.e. $\mathscr{C}^2$ diffeomorphisms*

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

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

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

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

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 $\boldsymbol{x} \in \boldsymbol{K}$ with (x_{p_-}, x_{p_+}). Since the ϕ_i's are c-conjugate, for $\boldsymbol{x}, \boldsymbol{x'} \in \mathbf{K}$ it holds:

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

Now we do computations in local charts $\psi_i : U_i \subseteq X_i \to \psi_i(U_i) \subseteq \mathbb{R}^{d_i}$ which are $\mathscr{C}^2$ diffeomorphisms such that $B_R(\bar{x}_i) \subseteq U_i$ 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[3]:

$$E(\boldsymbol{x}') + E(\boldsymbol{x}) \geq \int_0^1 \int_0^1 D^2_{p_- p_+} c(\boldsymbol{x}_{s,t})(x'_{p_-} - x_{p_-}, x'_{p_+} - x_{p_+}) \, \mathrm{d}s \, \mathrm{d}t$$

where $\boldsymbol{x}_{s,t} := (x_{p_-} + (1-s)x'_{p_-}, x_{p_+} + (1-t)x'_{p_+})$ for $s, t \in [0,1]$. Since $|D^2_{p_- p_+} c(\boldsymbol{x}_{s,t}) - D^2_{p_- p_+} c(\bar{\boldsymbol{x}})| \leq \eta(r)$ where $\eta(r)$ is independent from p and tends to 0 as $r \to 0$, and since by definition $D^2 c(\bar{\boldsymbol{x}})(x'_{p_-} - x_{p_-}, x'_{p_+} - x_{p_+}) = \frac{1}{2} g_p(\bar{\boldsymbol{x}})(\boldsymbol{x'} - \boldsymbol{x}, \boldsymbol{x'} - \boldsymbol{x})$, it holds:

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

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

$$E(\boldsymbol{x}) + E(\boldsymbol{x'}) \geq \frac{1}{2} g_{\bar{\boldsymbol{x}}}(\boldsymbol{x'} - \boldsymbol{x}, \boldsymbol{x'} - \boldsymbol{x}) - \eta(r) \|\boldsymbol{x'} - \boldsymbol{x}\|^2. \tag{4.2}$$

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

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

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

We will use the following positive signature condition:

$$\text{for every } \boldsymbol{x} \in \boldsymbol{X}, \quad d^+_c(\boldsymbol{x}) \geq \kappa \quad \text{where} \quad d^\pm_c(\boldsymbol{x}) := \max\left\{d^\pm(g)(\boldsymbol{x}) \mid g \in G_c\right\}. \tag{PS(κ)}$$

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

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

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

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

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

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

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

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

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

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

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

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

$$\mathscr{H}^d_{\boldsymbol{X}} = \otimes_{i=1}^m \mathscr{H}^{d_i}_{X_i} \quad \text{where} \quad d := \sum_i d_i.$$

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

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

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

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

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

As a consequence we obtain:

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

for some constant $C_{\bar{\boldsymbol{x}}} > 0$ (which depends on $\bar{\boldsymbol{x}}$ through R, d^- and d^0). The sets $\{V_{\bar{\boldsymbol{x}}}\}_{\bar{\boldsymbol{x}} \in \boldsymbol{\Sigma}}$ form an open covering of the compact set $\boldsymbol{\Sigma} := \{\boldsymbol{x} \in \boldsymbol{K} \mid E(\boldsymbol{x}) = 0\}$, hence we may extract a finite covering $V_{\bar{\boldsymbol{x}}_{\boldsymbol{1}}}, \dots, V_{\bar{\boldsymbol{x}}_{\boldsymbol{L}}}$ and for every $\varepsilon > 0$:

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

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

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

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

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 $\mathscr{C}^2$. As in [Pas12] we consider, for the lower bound, the metric $\overline{g}$ 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 non empty disjoint subsets.

Example 5.1 (**Two marginals case**). In previous works [CPT23; EN22] concerning the rate of convergence for the two marginals problem, it was assumed that the cost function must satisfy a non degeneracy condition, that is $D^2_{x_1 x_2} c$ 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 non-degeneracy condition fails. Let r be the rank of $D^2_{x_1x_2}c$ at the point where the non-degeneracy condition fails, then the signature of $\overline{g}$ at this point is given by $(r, r, 2d-2r)$ meaning that locally the support of the optimal γ_0 is at most $2d-r$ dimensional. Thus, the bounds become

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

for some constants $C_*, C^* > 0$. Notice that if $D^2_{x_1,x_2}c$ has full rank then $r = d$ and we retrieve the matching bound results of [CPT23; EN22].

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 $D^2_{x_1,x_2}c$ has full rank, that is $r = d_2$, we obtain a matching bound depending only on the lower dimensional marginal

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

for some constants $C_*, C^* > 0$. If μ_1 is absolutely continuous with respect to $\mathscr{H}^{d_1}$ on some smooth sub-manifold of dimension d_1, then any optimal transport plan would be concentrated on a set of Hausdorff dimension no less than d_1, and thus the upper bound given in [EN22, 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 $c(x_1, \ldots, x_m) = h(\sum_{i=1}^m x_i)$ where h is $\mathscr{C}^2$ and $D^2h > 0$. 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 $c(x_1, \ldots, x_m) = -\sum_{i<j}|x_i - x_j|^2$ (here $|\cdot|$ denotes the standard euclidean norm), see [DGN15] for more details. It follows now that the signature of the metric $\overline{g}$ is $(d, (m-1)d, 0)$ thus the bounds between MOT_ε and MOT_0 that we obtain are

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

for some constants $C_*, C^* > 0$. We remark that it is known from [Pas12; DGN15] that a transport plan γ_0 is optimal if and only if it is supported on the set $\{(x_1, \ldots, x_m) \mid \sum_{i=1}^m x_i = l\}$, where $l \in \mathbb{R}^d$ is any constant and there exists solutions whose support has dimension exactly $(m-1)d$.

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

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

Notice now that considering the MOT_0 problem with a cost $c(x_1, \ldots, x_m) = \sum_i |x_i - T(x_1, \ldots, x_m)|^2$, where $T(x_1, \ldots, x_m) = \sum_{i=1}^m \lambda_i x_i$ is the Euclidean barycenter , is equivalent to the MOT_0 with the Gangbo- Święch 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 $T_\sharp \gamma_0 = \nu$ is the barycenter of $\mu_1, \ldots, \mu_m$.

Acknowledgments. L.N. is partially on academic leave at Inria (team Matherials) for the year 2022-2023 and acknowledges the hospitality if 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.

References

[Abr+17] I. Abraham, R. Abraham, M. Bergounioux, and G. Carlier. "Tomographic Reconstruction from a Few Views: A Multi-Marginal Optimal Transport Approach". In: *Applied Mathematics & Optimization* 75.1 (2017), pp. 55–73.

[AC11] M. Agueh and G. Carlier. "Barycenters in the Wasserstein Space". In: *SIAM Journal on Mathematical Analysis* 43.2 (2011), pp. 904–924. ISSN: 0036-1410, 1095-7154. DOI: 10.1137/100805741.

[BCN19] J.-D. Benamou, G. Carlier, and L. Nenna. "Generalized Incompressible Flows, Multi-Marginal Transport and Sinkhorn Algorithm". In: *Numerische Mathematik* 142.1 (2019), pp. 33–54. ISSN: 0945-3245. DOI: 10.1007/s00211-018-0995-x.

[BDG12] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi. "Optimal-Transport Formulation of Electronic Density-Functional Theory". In: *Physical Review A* 85.6 (2012), p. 062502.

[Ben+15] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. "Iterative Bregman Projections for Regularized Transportation Problems". In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138. ISSN: 1064-8275. DOI: 10.1137/141000439.

[BHP13] M. Beiglböck, P. Henry-Labordere, and F. Penkner. "Model-Independent Bounds for Option Prices—a Mass Transport Approach". In: *Finance and Stochastics* 17.3 (2013), pp. 477–501.

[BK18] J. Bigot and T. Klein. "Characterization of Barycenters in the Wasserstein Space by Averaging Optimal Transport Maps". In: *ESAIM: Probability and Statistics* 22 (2018), pp. 35–57.

[BL92] J. M. Borwein and A. S. Lewis. "Decomposition of Multivariate Functions". In: *Canadian Journal of Mathematics* 44.3 (1992), pp. 463–482. ISSN: 0008-414X, 1496-4279. DOI: 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". In: *Journal of Functional Analysis* 123.2 (1994), pp. 264–307. ISSN: 0022-1236. DOI: 10.1006/jfan.1994.1089.

[Bre89] Y. Brenier. "The Least Action Principle and the Related Concept of Generalized Flows for Incompressible Perfect Fluids". In: *Journal of the American Mathematical Society* 2.2 (1989), pp. 225–255.

[Car+17] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. "Convergence of Entropic Schemes for Optimal Transport and Gradient Flows". In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418. ISSN: 0036-1410. DOI: 10.1137/15M1050264.

[Car03] G. Carlier. "On a Class of Multidimensional Optimal Transportation Problems". In: *Journal of Convex Analysis* 10.2 (2003), pp. 517–529. ISSN: 0944-6532.

[CCG16] G. Carlier, V. Chernozhukov, and A. Galichon. "Vector Quantile Regression: An Optimal Transport Approach". In: *The Annals of Statistics* 44.3 (2016), pp. 1165–1192.

[CDD15] M. Colombo, L. De Pascale, and S. Di Marino. "Multimarginal Optimal Transport Maps for One-Dimensional Repulsive Costs". In: *Canadian Journal of Mathematics* 67.2 (2015), pp. 350–368. ISSN: 0008-414X. DOI: 10.4153/CJM-2014-011-x.

[CE10] G. Carlier and I. Ekeland. "Matching for Teams". In: *Economic theory* 42.2 (2010), pp. 397–418.

[CFK13] C. Cotar, G. Friesecke, and C. Klüppelberg. "Density Functional Theory and Optimal Transportation with Coulomb Cost". In: *Communications on Pure and Applied Mathematics* 66.4 (2013), pp. 548–599. ISSN: 1097-0312. DOI: 10.1002/cpa.21437.

[CN08] G. Carlier and B. Nazaret. "Optimal Transportation for the Determinant". In: *ESAIM. Control, Optimisation and Calculus of Variations* 14.4 (2008), pp. 678–698. ISSN: 1292-8119. DOI: 10.1051/cocv:2008006.

[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). Preprint (submitted), p. 116. ISSN: 1432-0835. DOI: 10.1007/s00526-023-02455-0.

[CS16] M. Colombo and F. Stra. "Counterexamples in Multimarginal Optimal Transport with Coulomb Cost and Spherically Symmetric Data". In: *Mathematical Models & Methods in Applied Sciences* 26.6 (2016), pp. 1025–1049. ISSN: 0218-2025. DOI: 10.1142/S021820251650024X.

[Csi75] I. Csiszar. "I-Divergence Geometry of Probability Distributions and Minimization Problems". In: *The Annals of Probability* 3.1 (1975), pp. 146–158. ISSN: 0091-1798, 2168-894X. DOI: 10.1214/aop/1176996454.

[Cut13] M. Cuturi. "Sinkhorn Distances: Lightspeed Computation of Optimal Transport". In: *Advances in Neural Information Processing Systems*. Vol. 26. Curran Associates, Inc., 2013.

[DGN15] S. Di Marino, A. Gerolin, and L. Nenna. *Optimal Transportation Theory with Repulsive Costs*. Comment: Survey for the special volume for RICAM (Special Semester on New Trends in Calculus of Variations). 2015. URL: http://arxiv.org/abs/1506.04565. preprint.

[DS14a] Y. Dolinsky and H. M. Soner. "Martingale Optimal Transport and Robust Hedging in Continuous Time". In: *Probability Theory and Related Fields* 160.1 (2014), pp. 391–427.

[DS14b] Y. Dolinsky and H. M. Soner. "Robust Hedging with Proportional Transaction Costs". In: *Finance and Stochastics* 18.2 (2014), pp. 327–347.

[EN22] S. Eckstein and M. Nutz. *Convergence Rates for Regularized Optimal Transport via Quantization*. Comment: Fixed a typo in Theorem 3.8. 2022. URL: http://arxiv.org/abs/2208.14391. preprint.

[Enn+22] H. Ennaji, Q. Mérigot, L. Nenna, and B. Pass. *Robust Risk Management via Multi-Marginal Optimal Transport*. 2022. DOI: 10.48550/ARXIV.2211.07694. URL: https://arxiv.org/abs/2211.07694.

[Fed96] H. Federer. *Geometric Measure Theory*. Classics in Mathematics. Berlin, Heidelberg: Springer, 1996. ISBN: 978-3-642-62010-2. DOI: 10.1007/978-3-642-62010-2.

[FG97] H. Föllmer and N. Gantert. "Entropy Minimization and Schrödinger Processes in Infinite Dimensions". In: *The Annals of Probability* 25.2 (1997), pp. 901–926. ISSN: 0091-1798, 2168-894X. DOI: 10.1214/aop/1024404423.

[FL89] J. Franklin and J. Lorenz. "On the Scaling of Multidimensional Matrices". In: *Linear Algebra and its Applications*. Special Issue Dedicated to Alan J. Hoffman 114–115 (1989), pp. 717–735. ISSN: 0024-3795. DOI: 10.1016/0024-3795(89)90490-4.

[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. ISSN: 1432-0835. DOI: 10.1007/s00526-020-01735-3.

[GŚ98] W. Gangbo and A. Święch. "Optimal Maps for the Multidimensional Monge-Kantorovich Problem". In: *Communications on Pure and Applied Mathematics* 51.1 (1998), pp. 23–45. ISSN: 1097-0312. DOI: 10.1002/(SICI)1097-0312(199801)51:1<23::AID-CPA2>3.0.CO;2-H.

[Haa+21] I. Haasler, R. Singh, Q. Zhang, J. Karlsson, and Y. Chen. "Multi-Marginal Optimal Transport and Probabilistic Graphical Models". In: *IEEE Transactions on Information Theory* 67.7 (2021), pp. 4647–4668. DOI: 10.1109/TIT.2021.3077465.

[Hei02] H. Heinich. "Problème de Monge Pour n Probabilités". In: *Comptes Rendus Mathematique. Academie des Sciences. Paris* 334.9 (2002), pp. 793–795. ISSN: 1631-073X. DOI: 10.1016/S1631-073X(02)02341-5.

[KM10] Y.-H. Kim and R. J. McCann. "Continuity, Curvature, and the General Covariance of Optimal Transportation". In: *Journal of the European Mathematical Society* 12.4 (2010), pp. 1009–1040. ISSN: 1435-9855. DOI: 10.4171/jems/221.

[KP14] Y.-H. Kim and B. Pass. "A General Condition for Monge Solutions in the Multi-Marginal Optimal Transport Problem". In: *Siam Journal On Mathematical Analysis* 46.2 (2014), pp. 1538–1550. ISSN: 0036-1410. DOI: 10.1137/130930443.

[KP15] Y.-H. Kim and B. Pass. "Multi-Marginal Optimal Transport on Riemannian Manifolds". In: *American Journal of Mathematics* 137.4 (2015), pp. 1045–1060. ISSN: 0002-9327. DOI: 10.1353/ajm.2015.0024.

[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.

[MG20] S. D. 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 (2020), p. 27. ISSN: 1573-7691. DOI: 10.1007/s10915-020-01325-7.

[MP17] A. Moameni and B. Pass. "Solutions to Multi-Marginal Optimal Transport Problems Concentrated on Several Graphs". In: *ESAIM. Control, Optimisation and Calculus of Variations* 23.2 (2017), pp. 551–567. ISSN: 1292-8119. DOI: 10.1051/cocv/2016003.

[MS23] H. Malamut and M. Sylvestre. *Convergence Rates of the Regularized Optimal Transport : Disentangling Suboptimality and Entropy*. 2023. URL: http://arxiv.org/abs/2306.06940. preprint.

[Nen16] L. Nenna. "Numerical Methods for Multi-Marginal Optimal Transportation". PhD thesis. Université Paris sciences et lettres, 2016.

[NW22] M. Nutz and J. Wiesel. "Stability of Schrödinger Potentials and Convergence of Sinkhorn's Algorithm". 2022.

[Pas11] B. Pass. "Uniqueness and Monge Solutions in the Multimarginal Optimal Transportation Problem". In: *SIAM Journal on Mathematical Analysis* 43.6 (2011), pp. 2758–2775. ISSN: 0036-1410. DOI: 10.1137/100804917.

[Pas12] B. Pass. "On the Local Structure of Optimal Measures in the Multi-Marginal Optimal Transportation Problem". In: *Calculus of Variations and Partial Differential Equations* 43.3 (2012), pp. 529–536. ISSN: 1432-0835. DOI: 10.1007/s00526-011-0421-z.

[Pas15] B. Pass. "Multi-Marginal Optimal Transport: Theory and Applications". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 49.6 (6 2015), pp. 1771–1790. ISSN: 0764-583X, 1290-3841. DOI: 10.1051/m2an/2015020.

[PC19] G. Peyré and M. Cuturi. "Computational Optimal Transport: With Applications to Data Science". In: *Foundations and Trends in Machine Learning* 11.5-6 (2019), pp. 355–607. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000073.

[PV21a] B. Pass and A. Vargas-Jiménez. "Monge Solutions and Uniqueness in Multi-Marginal Optimal Transport via Graph Theory". 2021.

[PV21b] B. W. Pass and A. Vargas-Jiménez. "Multi-Marginal Optimal Transportation Problem for Cyclic Costs". In: *Siam Journal On Mathematical Analysis* 53.4 (2021), pp. 4386–4400. ISSN: 0036-1410. DOI: 10.1137/19M130889X.

[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. 2011, pp. 435–446.

[Rén59] A. Rényi. "On the Dimension and Entropy of Probability Distributions". In: *Acta Mathematica Academiae Scientiarum Hungarica* 10.1 (1959), pp. 193–215. ISSN: 1588-2632. DOI: 10.1007/BF02063299.

[RT98] L. Rüschendorf and W. Thomsen. "Closedness of Sum Spaces and the Generalized Schrödinger Problem". In: *Theory of Probability & Its Applications* 42.3 (1998), pp. 483–494. ISSN: 0040-585X. DOI: 10.1137/S0040585X97976301.

[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.

[Sin64] R. Sinkhorn. "A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices". In: *The Annals of Mathematical Statistics* 35.2 (1964), pp. 876–879. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177703591.

[TJK22] N. G. Trillos, M. Jacobs, and J. Kim. *The Multimarginal Optimal Transport Formulation of Adversarial Multiclass Classification*. 2022. DOI: 10.48550/ARXIV.2204.12676. URL: https://arxiv.org/abs/2204.12676.

[You82] L.-S. Young. "Dimension, Entropy and Lyapunov Exponents". In: *Ergodic Theory and Dynamical Systems* 2.1 (1982), pp. 109–124. ISSN: 1469-4417, 0143-3857. DOI: 10.1017/S0143385700009615.