METRIC PROPERTIES OF HOMOGENEOUS AND SPATIALLY INHOMOGENEOUS F-DIVERGENCES

NICOLÒ DE PONTI

Abstract. In this paper we investigate the construction and the properties of spatially inhomogeneous divergences, functionals arising from optimal Entropy-Transport problems that are computed in terms of an entropy function F and a cost function. Starting from the power-like entropy $F(s) = (s^p - p(s-1) - 1)/(p(p-1))$ and a suitable cost depending on a metric d on a space X, our main result ensures that for every $p > 1$ the related inhomogeneous divergence induces a distance on the space of finite measures over X. We also study in detail the pure entropic setting, that can be recovered as a particular case when the transport is forbidden. In this situation, corresponding to the classical theory of F-divergences, we show that the construction naturally produces a symmetric divergence and we highlight the important role played by the class of Matusita's divergences.

1. Introduction

Let $F : [0, +\infty) \to [0, +\infty)$ be a superlinear, convex function such that $F(1) = 0$. A central role in the paper is played by the F-divergences, i.e. functionals $D_F : \mathcal{M}(X) \times \mathcal{M}(X) \to [0, +\infty]$ of the form

$$D_F(\gamma||\mu) := \begin{cases} \int_X F\left(\frac{\mathrm{d}\gamma}{\mathrm{d}\mu}\right)\mathrm{d}\mu & \text{if } \gamma \ll \mu, \\ +\infty & \text{otherwise.} \end{cases} \tag{1}$$

Here $\mathcal{M}(X)$ denotes the set of finite, nonnegative, Borel measures over a metric space (X, d) (say complete and separable).

A classical example of F-divergence is given by the possible choice $F = U_1(s) := s\ln(s) - s + 1$, corresponding to the celebrated Kullback-Leibler divergence (also called relative entropy) [KL51], a functional introduced by Kullback and Leibler in 1951 intimately related to the famous Shannon's entropy (see [Sha48], [Lin91]). We notice that the presence of the linear part in the function U_1 is natural in dealing with measures with possibly different total mass (see Remark 1).

Since their introduction by Csiszár [Csi63], and independently Ali and Silvey [AS66], F-divergences have become a fundamental tool in information theory and statistics. They can be interpreted as a sort of "distance function" on the set of finite measures, even if they do not generally fulfill the symmetric property and the triangle inequality. We refer to Liese and Vajda [LV06], [Vaj89] and references therein for a systematic presentation of these functionals, where also their applicability in statistical test is discussed.

Recently, F-divergences have been considered by Liero, Mielke, Savaré [LMS18] as penalizing functionals in the formulation of optimal Entropy-Transport problems, a generalization of optimal transport problems obtained by relaxing the marginal constraints. In contrast with the classical transport setting, the theory that has been developed allows the description of phenomena where the conservation of mass may not hold (see also [LMS16], [Chi+18]), and for this reason in the literature it is also referred to as "unbalanced optimal transport".

N. De Ponti is with the Department of Mathematics, University of Pavia, Pavia 27100, Italy (e-mail: nicolo.deponti01@universitadipavia.it).

The construction of Entropy-Transport problems works as follow: starting from a cost function $c : X \times X \to [0, +\infty]$, we define the Transport functional as

$$\mathcal{T}(\boldsymbol{\gamma}) := \int_{X \times X} c(x_1, x_2) \mathrm{d}\boldsymbol{\gamma}(x_1, x_2), \qquad \boldsymbol{\gamma} \in \mathcal{M}(X \times X), \tag{2}$$

and the Entropy-Transport functional as

$$\mathcal{ET}(\boldsymbol{\gamma}||\mu_1, \mu_2) := D_F(\gamma_1||\mu_1) + D_F(\gamma_2||\mu_2) + \mathcal{T}(\boldsymbol{\gamma}), \tag{3}$$

where $\gamma_i := (\pi^i)_\sharp \boldsymbol{\gamma}$ are the marginals of the measure $\boldsymbol{\gamma}$ obtained as the push-forward through the projection maps $\pi^i(x_1, x_2) = x_i$, $i = 1, 2$. The Entropy-Transport problem between the measures μ_1 and μ_2 is the minimization problem

$$\mathrm{ET}(\mu_1, \mu_2) := \inf_{\boldsymbol{\gamma} \in \mathcal{M}(X \times X)} \mathcal{ET}(\boldsymbol{\gamma}||\mu_1, \mu_2). \tag{4}$$

The aim of the paper is to investigate the metric properties of the function $\mathrm{ET} : \mathcal{M}(X) \times \mathcal{M}(X) \to [0, +\infty]$ with respect to different possible choices of the function F and the cost c. As a result of our study, we will not only produce new Entropy-Transport distances on the space of finite measures, but also give new insights on the properties of the F-divergences.

The starting point of our analysis is the marginal perspective cost

$$H(x_1, r; x_2, t) := \inf_{\theta > 0} F\Big(\frac{\theta}{r}\Big) r + F\Big(\frac{\theta}{t}\Big) t + \theta c(x_1, x_2) = \mathrm{ET}(r\delta_{x_1}, t\delta_{x_2}), \tag{5}$$

a function corresponding to the solution of the Entropy-Transport problem between two Dirac masses $\mu_1 = r\delta_{x_1}$, $\mu_2 = t\delta_{x_2}$. From the general theory developed in [LMS18, Section 5 and 7] (see also [Chi+18, Theorem 3.7]), it follows that ET is a (power) of a distance on $\mathcal{M}(X)$ if the function H is a (power) of a distance on the cone space over X. The latter is the space $\mathfrak{C}(X) = Y/\sim$, where $Y = X \times [0, +\infty)$ and

$$(x_1, r) \sim (x_2, t) \iff r = t = 0 \text{ or } r = t,\ x_1 = x_2\,.$$

When the starting entropy $F(s)$ has a strict minimum at $s = 1$, and the cost c is a symmetric function such that $c(x_1, x_2) = 0$ if and only if $x_1 = x_2$, we show that the induced marginal perspective cost H is symmetric, nonnegative and $H(x_1, r; x_2, t) = 0$ if and only if $(x_1, r) \sim (x_2, t)$.

The question regarding the validity of the triangle inequality is much more challenging: first of all, in the presence of a non-trivial cost function c, an explicit computation of the induced marginal perspective cost is often unavailable. In addition, the triangle inequality is known to be difficult to prove, even in the easier pure entropic case that has been previously considered in the literature (see below). In the present paper, we address this question for an important class of Entropy-Transport problems, the ones generated by the choices $c(x_1, x_2) = f(\mathsf{d}(x_1, x_2))$ for a suitable function $f : [0, +\infty) \to [0, +\infty]$, and $F = U_p$, where the latter corresponds to the class of power-like entropies defined by

$$U_1(s) := s \ln(s) - s + 1, \quad U_p(s) := \frac{1}{p(p-1)}(s^p - p(s-1) - 1) \ \text{ if } p > 1.$$

In this situation, the induced marginal perspective cost H takes the form

$$H = H_p(x_1, r; x_2, t) := \frac{2}{p}\left[\mathfrak{M}_1(r, t) - \mathfrak{M}_{1-p}(r, t)\left(1 + (1-p)\frac{c(x_1, x_2)}{2}\right)_+^{\frac{p}{p-1}}\right], \quad p > 1,$$

$$H = H_1(x_1, r; x_2, t) := 2\Big[\mathfrak{M}_1(r, t) - \mathfrak{M}_0(r, t) e^{-c(x_1, x_2)/2}\Big],$$

where the expressions are written in the terms of the power means

$$\mathfrak{M}_p(r, t) := \Big(\frac{r^p + t^p}{2}\Big)^{\frac{1}{p}},\ p \neq 0, \qquad \mathfrak{M}_0(r, t) := \sqrt{rt}.$$

In Theorem 5 and Theorem 6, which are our main results, we prove that the square root of the induced marginal perspective cost H_p is a distance on $\mathfrak{C}(X)$ for every $p > 1$, where c is given by one of the two following cost functions:

(1) $c = c_p(x_1, x_2) := \frac{2}{p-1}\left[1 - \big(\cos(\mathsf{d}(x_1, x_2) \wedge \pi/2)\big)^{\frac{p-1}{p}}\right]$,
(2) $c(x_1, x_2) = \mathsf{d}(x_1, x_2)$.

The same result is obtained also for the cost $c(x_1, x_2) = \mathsf{d}^2(x_1, x_2)$ and $1 < p \leq 3$.

Thus, we provide an entire class of Entropy-Transport distances on the space $\mathcal{M}(X)$, besides the Gaussian Hellinger-Kantorovich distance and the related Hellinger-Kantorovich distance studied in [LMS18], that correspond to the case $p = 1$, $c(x_1, x_2) = \mathsf{d}^2(x_1, x_2)$ and $p = 1$,

$$c = c_{\mathrm{HK}}(x_1, x_2) := \begin{cases} -\log\big(\cos^2(\mathsf{d}(x_1, x_2))\big) & \text{if } \mathsf{d}(x_1, x_2) < \pi/2, \\ +\infty & \text{otherwise,} \end{cases}$$

respectively. Notice here that the choice of the "exotic" cost c_p defined above is motivated by its counterpart c_{HK}; in particular, it holds

$$\lim_{p \to 1} c_p(x_1, x_2) = c_{\mathrm{HK}}(x_1, x_2).$$

The class of distances we studied includes, for $p = 2$, a transport variant of the Vincze-Le Cam distance [Vin81], [Cam86].

In contrast with the case $p = 1$, where Liero, Mielke and Savaré were able to take advantage of the expression of the induced marginal perspective cost, closely connected with the "natural" metric $\mathsf{d}_{\mathfrak{C}}$ of the cone space (see, e.g., [BBI01, Section 3.6])

$$\mathsf{d}^2_{\mathfrak{C}}\big((x_1, r), (x_2, t)\big) := r^2 + t^2 - 2rt\cos(\mathsf{d}(x_1, x_2) \wedge \pi),$$

the proof of Theorem 5 is based on a careful case by case inspection, where we adapt different results already present in the literature ([Ost96], [ES03], [Kou14]) as well as employ new techniques in order to compare the function H_p with the "model case" $p = 1$. Once that is done, the proof of Theorem 6 follows by some explicit computations, taking advantage of a well-known technical lemma (Lemma 4).

If $F = U_p$, we also prove that for every cost function c it holds

$$H_p \leq H_1 \leq pH_p \qquad p \geq 1. \tag{6}$$

The explicit bounds (6) allow us to prove that all the Entropy-Transport distances previously considered are complete and separable metrics on the space $\mathcal{M}(X)$ inducing the weak topology.

The rest of the paper is devoted to the study of the problem (4) and the property of the function (5) in the pure entropic case, which correspond to the choice

$$c(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise.} \end{cases} \tag{7}$$

In this situation, the structure of the problem does not allow the spatial movement of the mass and we obtain

$$\mathrm{ET}(\mu_1, \mu_2) = \int_X H_F\Big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda}, \frac{\mathrm{d}\mu_2}{\mathrm{d}\lambda}\Big)\mathrm{d}\lambda, \tag{8}$$

where $\lambda \in \mathcal{M}(X)$ is any dominating measure of μ_1, μ_2, i.e. μ_1 and μ_2 are absolutely continuous with respect to λ. Here H_F is defined by

$$H_F(r, t) := \inf_{\theta > 0} F\Big(\frac{\theta}{r}\Big)r + F\Big(\frac{\theta}{t}\Big)t, \tag{9}$$

and we notice the close connection with the marginal perspective cost (5).

We refer to this setting as "pure entropic" since the expression (8) is an equivalent formulation of the divergence induced by the function $f(s) := H_F(s,1)$ (see Lemma 3). In particular, the minimizing procedure (9) corresponds to the construction

$$D_f(\mu_1||\mu_2) := \inf_{\gamma\in\mathcal{M}(X)} D_F(\gamma||\mu_1) + D_F(\gamma||\mu_2)$$

at the level of divergences, and it can be seen as a simple variational way to generate a new symmetric divergence starting from a given convex function F.

The metric properties of F-divergences, and in particular the validity of the triangle inequality, have been investigated by many authors like Csiszár, Endres, Kafka, Osterreicher, Schindelin, Vincze ([Csi67], [ES03], [KOV91], [Ost96], [OV03]), to cite only a few. However, to the author's knowledge, the construction of the function H_F defined in (9) has never been considered in the literature and we will show that it exhibits interesting structural properties.

In contrast with the spatial inhomogeneous case, the explicit expression of the function H_F is easier to obtain and gives rise to well-known statistical functionals, which include the Hellinger distance [Hel09], the Jensen-Shannon divergence [Lin91] and more generally a class of Arimoto-type divergences that has been studied by Osterreicher and Vajda [Ost96], [OV03] (see Example 4). One can also obtain other important functionals, including the symmetric Kullback-Leibler divergence [KL51] (see Example 5) and the class of Matusita's divergence (see Example 3).

Regarding the power-like entropies U_p, we prove that the induced function H_{U_p} is the square of a distance on $\mathbb{R}_+$ for every $p \geq 1$. This completes a study started by Osterreicher and Vajda (see [Ost96], [OV03]), who considered the class of functions corresponding to the case $p < 1$ (see Example 4 and Proposition 3).

Our analysis does not limit to superlinear function F. This is particularly important because, by directly studying the minimizing procedure (9), we show that the only distances between F-divergences are provided by the family of the total variation divergences related to the function $F(s) = c|s-1|$, $c \in (0,+\infty)$. Some general results are proved also for the more difficult case of the divergences F that induce a distance on $\mathcal{M}(X)$ of the form

$$(\gamma,\mu) \mapsto D_F(\gamma||\mu)^a \quad \text{for a power } a \in (0,1).$$

Here, we will emphasize the central role of the class of Matusita's divergences $F(s) = |s^a - 1|^{\frac{1}{a}}$ [Mat64].

The paper is organized as follows.

In section 2 we introduce the necessary preliminaries, including some concepts of measure theory and convex analysis. We also recall the definition of the power means and their main properties.

Section 3 is devoted to the study of the pure entropic case. We define the F-divergence and provide an equivalent definition based on a perspective formulation (Lemma 3), that is crucial in order to show the connection of the metric properties of the divergences with the related properties of the perspective function (Proposition 2). We also provide a list of examples of admissible entropy functions F and we compute the induced marginal perspective function H_F. We then discuss when some of these functions induce a distance. Finally, we study the convergence properties of the iteration of the minimizing procedure (9) and we will highlight the important role played by the class of Matusita's divergences.

In section 4 we introduce the notion of marginal perspective cost and we discuss its main properties.

Section 5 contains a brief introduction to the theory of optimal Entropy-Transport problems. We motivate the "homogeneous formulation" of these problems via the marginal perspective cost H, which is convenient for studying Entropy-Transport distances (Proposition 6).

In the last section we focus on the marginal perspective cost H_p induced by the power-like entropy U_p. We prove our main results (Theorem 5 and Theorem 6), which ensure that, if $p \geq 1$ and c is a suitable cost function, the function H_p is the square of a distance on the corresponding cone space. This allows us to produce new Entropy-Transport metrics on the space $\mathcal{M}(X)$, that are also complete and separable (see Corollary 1) thanks to the bound obtained in Proposition 7.

2. Preliminary

2.1. **Measure theory.** Let (X, d) be a Polish space, i.e. a separable, completely metrizable, topological space. We denote by $\mathcal{B}(X)$ the σ-algebra of the Borel sets and by $\mathcal{M}(X)$ the set of finite, nonnegative, Borel measures on X.

We endow $\mathcal{M}(X)$ with the weak topology, inducing the following notion of convergence:

$$\mu_n \to \mu \iff \int_X f d\mu_n \to \int_X f d\mu \text{ for any } f \in C_b(X), \tag{10}$$

where $C_b(X)$ denotes the set of continuous and bounded functions $f : X \to \mathbb{R}$.

Suppose $\mu \in \mathcal{M}(X)$, let Y be another Polish space and let $T : X \to Y$ be a Borel map. We define the push-forward measure $T_\sharp\mu \in \mathcal{M}(Y)$ by

$$T_\sharp\mu(B) := \mu(T^{-1}(B)) \qquad \text{for every } B \in \mathcal{B}(Y). \tag{11}$$

We say that a measure $\gamma \in \mathcal{M}(X)$ is absolutely continuous with respect to a measure $\mu \in \mathcal{M}(X)$, and we write $\gamma \ll \mu$, if $\gamma(A) = 0$ for any set $A \in \mathcal{B}(X)$ such that $\mu(A) = 0$. In this situation, it is well known that γ admits a Radon-Nykodym derivative with respect to μ, i.e. there exists a measurable function $f : X \to [0, +\infty]$ such that

$$\gamma(B) = \int_B f \mathrm{d}\mu \qquad \text{for every } B \in \mathcal{B}(X).$$

The function f is denoted by $\dfrac{\mathrm{d}\gamma}{\mathrm{d}\mu}$.

The measure γ is singular with respect to $\mu^\perp \in \mathcal{M}(X)$ if there exists a Borel subset $A \subset X$ such that $\mu(A) = \gamma^\perp(X \setminus A) = 0$. More generally, the following version of the Lebesgue's decomposition Theorem holds (see [LMS18, Lemma 2.3]).

Lemma 1. *For every $\gamma, \mu \in \mathcal{M}(X)$ with $\gamma(X) + \mu(X) > 0$, there exist Borel functions $\sigma, \rho : X \to [0, +\infty)$ and a Borel partition (A, A_γ, A_μ) of X that satisfy the following:*

$$\begin{aligned} &A = \{x \in X : \sigma(x) > 0\} = \{x \in X : \rho(x) > 0\}, \quad \sigma \cdot \rho \equiv 1 \textit{ in } A, \\ &\gamma = \sigma\mu + \gamma^\perp, \quad \sigma \in L^1_+(X, \mu), \quad \gamma^\perp(X \setminus A_\gamma) = \mu(A_\gamma) = 0, \\ &\mu = \rho\gamma + \mu^\perp, \quad \rho \in L^1_+(X, \gamma), \quad \mu^\perp(X \setminus A_\mu) = \mu(A_\mu) = 0. \end{aligned} \tag{12}$$

2.2. **Admissible entropy functions.** A function $F : [0, +\infty) \to [0, +\infty]$ belongs to the class $\Gamma_0(\mathbb{R}_+)$ of the admissible entropy functions if F is convex, lower semicontinuous and $F(1) = 0$. The domain of the function F is the set

$$\mathrm{D}(F) := \big\{s \in [0, +\infty) : F(s) < +\infty\big\}. \tag{13}$$

Let $F \in \Gamma_0(\mathbb{R}_+)$, the recession function $\mathrm{rec}(F)$ and the recession constant F'_∞ are defined by

$$\mathrm{rec}(F)(r) := \lim_{\alpha \to +\infty} \frac{F(1 + \alpha r)}{\alpha}, \qquad F'_\infty := \mathrm{rec}(F)(1). \tag{14}$$

We define the right derivative F'_0 at 0, and the asymptotic affine coefficient $\mathrm{aff}F_\infty$ as

$$F'_0 := \begin{cases} -\infty & \text{if } \ F(0) = +\infty, \\ \lim_{s \downarrow 0} \frac{F(s) - F(0)}{s} & \text{otherwise,} \end{cases} \tag{15}$$

$$\mathrm{aff}F_\infty := \begin{cases} +\infty & \text{if } \ F'_\infty = +\infty, \\ \lim_{s \to \infty} \big(F'_\infty s - F(s)\big) & \text{otherwise} \end{cases} \tag{16}$$

Note that the definitions are well posed thanks to the convexity of F.

The Legendre conjugate function $F^* : \mathbb{R} :\to (-\infty, +\infty]$ is defined by

$$F^*(\phi) := \sup_{s\geq 0}\{s\phi - F(s)\}. \tag{17}$$

F^* is the conjugate of the convex function $\tilde{F} : \mathbb{R} \to [0, +\infty]$ obtained by extending F to $+\infty$ for negative arguments. It is convex and lower semicontinuous. Concerning the behavior of F^*, we have the following Lemma (see [LMS18, Section 2.3]):

Lemma 2. *The function F^* is an increasing homeomorphism between (F_0', F_∞') and $(-F(0), \mathrm{aff}F_\infty)$ with $F^*(0) = 0$.*

The reverse entropy function $R : [0, +\infty) \to [0, +\infty]$ is defined by

$$R(s) := \begin{cases} F(\frac{1}{s})s & \text{if } s > 0, \\ F_\infty' & \text{if } s = 0. \end{cases} \tag{18}$$

In particular, R is convex, lower semicontinuous and the map $F \mapsto R$ is an involution of $\Gamma_0(\mathbb{R}_+)$. We also have

$$R(1) = 0, \quad R(0) = F_\infty', \quad R_\infty' = F(0), \quad R_0' = -\mathrm{aff}F_\infty, \quad \mathrm{aff}R_\infty = -F_0'. \tag{19}$$

We denote by $\Gamma_0^s(\mathbb{R}_+)$ the set of the functions $F \in \Gamma_0(\mathbb{R}_+)$ such that F is equals to its reverse entropy R.

The Legendre conjugates of F and R are related by

$$\psi \leq -F^*(\phi) \iff \phi \leq -R^*(\psi). \tag{20}$$

Finally, the perspective function induced by $F \in \Gamma_0(\mathbb{R}_+)$ is the function $\hat{F} : [0, +\infty)\times[0, +\infty) \to [0, +\infty]$, given by

$$\hat{F}(r, t) := \begin{cases} F\left(\frac{r}{t}\right)t & \text{if } t > 0, \\ \mathrm{rec}(F)(r) & \text{if } t = 0. \end{cases} \tag{21}$$

$\hat{F}$ is jointly convex, lower semicontinuous, positively 1-homogeneous in the sense that $\hat{F}(\lambda r, \lambda t) = \lambda\hat{F}(r, t)$ for every $\lambda \geq 0$, and $\hat{F}(1, 1) = 0$. An easy consequence of the definition is the fact that $\hat{F}(r, t) = \hat{R}(t, r)$. In particular, if $F \in \Gamma_0^s(\mathbb{R}_+)$ then $\hat{F}$ is a symmetric function, i.e.

$$\hat{F}(r, t) = \hat{F}(t, r) \qquad \text{for every } r, t \in [0, +\infty).$$

2.3. **Power means.** In this section we study the power means (also called generalized means), a family of functions that includes the well-known arithmetic, geometric and harmonic means. The property of these functions will be useful later on.

In what follows r, t will denote two nonnegative real numbers and p a real parameter, which we suppose for the present not to be 0. The p-power mean between r and t is given by

$$\mathfrak{M}_p(r, t) := \left(\frac{r^p + t^p}{2}\right)^{\frac{1}{p}}, \tag{22}$$

except when $p < 0$ and r or t is zero. In this case $\mathfrak{M}_p$ is equal to zero:

$$\mathfrak{M}_p(r, t) = 0 \quad (p < 0,\ r = 0 \text{ or } t = 0). \tag{23}$$

In the case $p = 0$ we put

$$\mathfrak{M}_0(r, t) := \sqrt{rt} \tag{24}$$

so that $\lim_{p\to 0} \mathfrak{M}_p(r, t) = \mathfrak{M}_0(r, t)$.

It is easy to see that $\mathfrak{M}_p(r,r) = r$ for every $p \in \mathbb{R}$ and every $r \geq 0$. The function $\mathfrak{M}_p$ is positively 1-homogeneous and symmetric. Moreover, it is not difficult to prove that $M_p(r,s) \leq M_p(r,t)$ for every p, r and $s \leq t$.

$\mathfrak{M}_1$ is the well-known arithmetic mean, $\mathfrak{M}_0$ is the geometric mean and $\mathfrak{M}_{-1}$ is called harmonic mean.

The main result (see [Bul03] for a proof) regarding the power means is the following:

Proposition 1. *If $p_1 < p_2$ then*

$$\mathfrak{M}_{p_1}(r,t) \leq \mathfrak{M}_{p_2}(r,t)$$

with the case of equality given by $r = t$, or $p_2 \leq 0$ and $r \wedge t = 0$.

In particular,

$$r \wedge t = \lim_{p \to -\infty} \mathfrak{M}_p(r,t) \leq \mathfrak{M}_p(r,t) \leq \lim_{p \to +\infty} \mathfrak{M}_p(r,t) = r \vee t, \tag{25}$$

for every $p \in \mathbb{R}$, $r,t \in [0,+\infty)$.

3. Pure entropic case

3.1. **F-divergence and marginal perspective function.** Let X be a Polish space and $F \in \Gamma_0(\mathbb{R}_+)$. The F-divergence (also called Csiszár's divergence) is the functional $D_F : \mathcal{M}(X) \times \mathcal{M}(X) \to [0,+\infty]$ defined by

$$D_F(\gamma||\mu) := \int_X F(\sigma)\mathrm{d}\mu + F'_\infty \gamma^\perp(X), \qquad \gamma = \sigma\mu + \gamma^\perp, \tag{26}$$

where $\gamma = \sigma\mu + \gamma^\perp$ is the Lebesgue's decomposition of the measure γ with respect to μ that follows from Lemma 1. In particular, if F is superlinear, i.e. $F'_\infty = +\infty$, the F-divergence is always equal to $+\infty$ unless γ is absolutely continuous with respect to μ. We notice that the presence of the additional term $F'_\infty \gamma^\perp(X)$ in the definition (26) is crucial for the lower semicontinuity property of these functionals (see [LMS18, Corollary 2.9]).

Recalling the definition of $\hat{F}$ (21), we define the perspective divergence $\hat{D}_F : \mathcal{M}(X) \times \mathcal{M}(X) \to [0,+\infty]$ as

$$\hat{D}_F(\gamma||\mu) := \int_X \hat{F}\Big(\frac{\mathrm{d}\gamma}{\mathrm{d}\lambda}, \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\Big)\mathrm{d}\lambda, \tag{27}$$

where $\lambda \in \mathcal{M}(X)$ is any dominating measure of γ and μ, i.e. $\gamma \ll \lambda$, $\mu \ll \lambda$. It is easy to see that such a measure λ always exists (take $\lambda = \gamma + \mu$) and $\hat{D}_F$ does not depend on λ since $\hat{F}$ is positively 1-homogeneous.

Lemma 3. *For every $\gamma, \mu \in \mathcal{M}(X)$ we have*

$$D_F(\gamma||\mu) = \hat{D}_F(\gamma||\mu).$$

Proof. Let $\gamma = \sigma\mu + \gamma^\perp$ be the Lebesgue's decomposition and define $\lambda := \mu + \gamma^\perp$. We observe that λ dominates μ and γ. Since μ is singular with respect to $\gamma^\perp$, there exist A, B Borel subsets such that

$$A \cup B = X, \qquad A \cap B = \emptyset, \qquad \mu(A) = 0, \qquad \gamma^\perp(B) = 0.$$

Put $\gamma := \rho\lambda$, $\mu := \tau\lambda$. The densities ρ, τ satisfies

$$\rho(x) = \begin{cases} 1 & \text{if } x \in A, \\ \sigma(x) & \text{if } x \in B, \end{cases} \qquad \tau(x) = \begin{cases} 0 & \text{if } x \in A, \\ 1 & \text{if } x \in B. \end{cases} \tag{28}$$

Thus,

$$\begin{aligned} \hat{D}_F(\gamma||\mu) = \int_X \hat{F}(\rho,\tau)\mathrm{d}\lambda = \int_B \hat{F}(\rho,\tau)\mathrm{d}\lambda + \int_A \hat{F}(\rho,\tau)\mathrm{d}\lambda = \int_B \hat{F}(\sigma,1)\mathrm{d}\lambda + \int_A \hat{F}(1,0)\mathrm{d}\lambda \\ = \int_B F(\sigma)\mathrm{d}\lambda + \int_A F'_\infty \mathrm{d}\lambda = \int_X F(\sigma)\mathrm{d}\mu + F'_\infty\gamma^\perp(X) = D_F(\gamma||\mu). \end{aligned} \tag{29}$$

□

Some comments regarding what we have done so far are in order.

Remark 1. *We have defined the F-divergence only for functions F in the class $\Gamma_0(\mathbb{R}_+)$. In principle, one could work with convex and lower semicontinuous functions $F : [0,+\infty) \to \mathbb{R}$ such that $F(1) = 0$. In this case, an easy application of Jensen's inequality ensures that D_F is nonnegative between probability measures, i.e. measures $\mu \in \mathcal{M}(X)$ such that $\mu(X) = 1$. However, if there exists a point $q \in [0,\infty]$ such that $F(q) < 0$, then D_F is not nonnegative between measures with different total mass. To see this, let us consider $\gamma := r\delta_x$ and $\mu := t\delta_x$ such that $r/t = q$. It is apparent that $D_F(\gamma||\mu) = \hat{F}(r,t)t = F(q)t < 0$.*

Let us finally mention that, given a convex and lower semicontinuous function $F : [0,+\infty) \to \mathbb{R}$ such that $F(1) = 0$, the function $\tilde{F}(s) := F(s) - c(s-1)$, where c is a subderivative of the function F at $s = 1$ (which exists since F is convex), is in the class $\Gamma_0(\mathbb{R}_+)$ and we have

$$D_{\tilde{F}}(\gamma||\mu) = D_F(\gamma||\mu) - c(\mu(X) - \gamma(X)).$$

In particular $D_{\tilde{F}}$ and D_F coincide between measures with the same total mass.

Remark 2. *Starting from a function $F \in \Gamma_0(\mathbb{R}_+)$, we have seen that we can construct the perspective function $\hat{F}$ thanks to (21), the F-divergence prescribed in (26) and the perspective divergence defined by (27). Moreover, Lemma 3 tell us that the F-divergence and the induced perspective divergence coincide. If we start instead with a lower semicontinuous, jointly convex and positively 1-homogeneous function $H : [0,+\infty) \times [0,+\infty) \to [0,\infty]$ such that $H(1,1) = 0$, we can define the $\mathcal{H}$-divergence $\mathcal{H} : \mathcal{M}(X) \times \mathcal{M}(X) \to [0,+\infty]$ by the formula*

$$\mathcal{H}(\gamma||\mu) := \int_X H\Big(\frac{d\gamma}{d\lambda}, \frac{d\mu}{d\lambda}\Big) d\lambda, \tag{30}$$

where λ is any dominating measure of γ and μ; H also induces a function $f \in \Gamma_0(\mathbb{R}_+)$ simply by taking $f(s) := H(s,1)$. Thus, in studying Csiszár divergences, we have two different but equivalent points of view and, depending on the circumstances, we may choose to work with functions of 1 or 2 variables.

It is an easy consequence of the perspective formulation that the F-divergence is a symmetric functional if and only if $\hat{F}$ is symmetric. However, the symmetry of $\hat{F}$ is not always satisfied (take for instance $F(s) = s\ln(s) - s + 1$).

In order to replace F with a new "symmetric entropy", a natural procedure is the following: define the marginal perspective function $H_F : [0,+\infty) \times [0,+\infty) \to [0,+\infty]$ as the lower semicontinuous envelope of the function

$$\tilde{H}_F(r,t) := \inf_{\theta>0} \Big(R\big(\frac{r}{\theta}\big) + R\big(\frac{t}{\theta}\big)\Big)\theta. \tag{31}$$

An equivalent definition of $\tilde{H}_F$ can be given in term of the induced perspective functions $\hat{F}$ or $\hat{R}$ by:

$$\tilde{H}_F(r,t) = \inf_{\theta>0} \hat{F}(\theta,r) + \hat{F}(\theta,t) = \inf_{\theta>0} \hat{R}(r,\theta) + \hat{R}(t,\theta) \tag{32}$$

The infimum in the definition of $\tilde{H}_F$ is a minimum and it occurs in the interval $[r,t]$ (without loss of generality we are assuming $r \le t$): to see this, it is enough to notice that the function $\theta \mapsto \hat{F}(\theta,r) + \hat{F}(\theta,t)$ is lower semicontinuous and it is decreasing in $[0,r]$ and increasing in $[t,+\infty)$. We will see in section 4, in a more general context, that the function H_F is lower semicontinuous, jointly convex, positively 1-homogeneous and symmetric. Moreover, if the function $F(s)$ has a strict minimum at $s = 1$, $H_F(r,t) = 0$ if and only if $r = t$.

3.2. **Examples.** We consider now different examples of admissible entropy functions and we compute the expressions of the corresponding marginal perspective functions, at least in the case $r \wedge t > 0$. In this regard, it is useful to recall that the function $\tilde{H}_F$ coincides with the marginal perspective function H_F in the interior of its domain (see below Lemma 12).

Example 1. *(Indicator functions) The indicator function of the closed interval with endpoints a and b, $0 \le a \le 1 \le b \le +\infty$, is defined by*

$$I_{[a,b]}(s) = \begin{cases} 0 & \text{if } s \in [a,b], \\ +\infty & \text{if } s \notin [a,b]. \end{cases} \tag{33}$$

When $F = I_{[a,b]}$ one obtains

$$H_{I_{[a,b]}}(r,t) = \begin{cases} 0 & \text{if } \frac{a}{b} \le \frac{r}{t} \le \frac{b}{a}, \\ +\infty & \text{otherwise}, \end{cases} \tag{34}$$

where $\frac{b}{a} = +\infty$ if $a = 0$ and $\frac{a}{b} = 0$ if $b = +\infty$.

Example 2. *(χ^α divergences) Given a parameter $\alpha \ge 1$, the χ^α divergence is induced by the function*

$$\chi^\alpha(s) = |s-1|^\alpha. \tag{35}$$

$\chi^1 = |s-1|$ is the famous total variation entropy.

The entropy function $F = \chi^\alpha$ gives rise to the marginal perspective function

$$H_{\chi^\alpha}(r,t) = \frac{|r-t|^\alpha}{(r+t)^{\alpha-1}}. \tag{36}$$

We can recognize the expression of the so-called Puri-Vincze divergence.

Example 3. *(Matusita divergences) For $0 < a \le 1$ the Matusita's divergence is given by the function $M_a(s) = |s^a - 1|^{\frac{1}{a}}$. Clearly $\chi^1 = M_1$.*

When $F = M_a$ it is easy to see that

$$H_{M_a}(r,t) = 2^{1-\frac{1}{a}} |r^a - t^a|^{\frac{1}{a}}. \tag{37}$$

It is interesting to note that except for the constant factor $2^{1-\frac{1}{a}}$, the Matusita function M_a remains invariant after the minimizing procedure (31). We will come back to this point in section 3.4.

Example 4. *(Power-like entropies) Let p be any real number. We call power-like entropy of order p the function $U_p : [0,+\infty) \to [0,+\infty]$ characterized by*

$$U_p \in \mathcal{C}^\infty(0,+\infty),\, U_p(1) = U_p'(1) = 0,\, U_p''(s) = s^{p-2},\, U_p(0) := \lim_{s \downarrow 0} U_p(s). \tag{38}$$

The function U_p can be computed explicitly and one gets:

$$\begin{cases} U_p(s) = \frac{1}{p(p-1)}(s^p - p(s-1) - 1) & \text{if } p \ne 0,1, \\ U_1(s) = s\ln(s) - s + 1, \\ U_0(s) = s - 1 - \ln(s), \end{cases} \tag{39}$$

with $U_p(0) = 1/p$ for $p > 0$ and $U_p(0) = +\infty$ for $p \le 0$. This family of functions, also called Dichotomy Class, was introduced by Liese and Vajda [LV87],[Vaj89].

Given $F = U_p$, we obtain the following expression:

$$\begin{cases} H_{U_p}(r,t) = \frac{1}{p}\left[r + t - 2^{\frac{p}{p-1}}(r^{1-p} + t^{1-p})^{\frac{1}{1-p}}\right] & p \neq 0, 1, \\ H_{U_1}(r,t) = r + t - 2\sqrt{rt}, \\ H_{U_0}(r,t) = r\ln r + t\ln t - (r+t)\ln\left(\frac{r+t}{2}\right). \end{cases} \tag{40}$$

We can recognize some well-known statistical functionals: for example in the logarithmic entropy case $p = 1$ it appears the Hellinger distance

$$H_{U_1}(r,t) = (\sqrt{r} - \sqrt{t})^2. \tag{41}$$

We have already noticed that the same function is obtained starting from the entropy $U_{\frac{1}{2}}(s) = 2(\sqrt{s} - 1)^2 = 2M_{\frac{1}{2}}$.

For $p = 0$ we have the Jensen-Shannon divergence, a squared distance between measures derived from the Kullback-Leibler divergence ([ES03]).

The quadratic entropy $U_2(s) = \frac{1}{2}(s-1)^2$ gives rise to the triangular discrimination

$$H_{U_2}(r,t) = \frac{1}{2}H_{\chi^2}(r,t) = \frac{1}{2}\frac{(r-t)^2}{(r+t)}. \tag{42}$$

Example 5. *(Power-logarithmic entropies) Given a real number $p \geq 1$, we call power-logarithmic entropy of order p the function $V_p : [0, +\infty) \to [0, +\infty]$ defined as*

$$V_p(s) := s^p - p\ln(s) - 1, \quad s > 0, \tag{43}$$

and $V_p(0) = +\infty$. It is easy to see that $V_p \in \mathcal{C}^\infty(0, +\infty)$ and $V_p(0) = \lim_{s\downarrow 0} V_p(s)$.

Starting from the power-logarithmic entropy of order p one gets:

$$H_{V_p}(r,t) = (r+t)\ln\left[\frac{rt(r^{p-1} + t^{p-1})}{r+t}\right] - p\big(r\ln(t) + t\ln(r)\big). \tag{44}$$

As expected, $H_{V_1} = H_{U_0}$ since $V_1 = U_0$. When $p = 2$, one obtains the symmetric Kullback-Leibler divergence [KL51]:

$$H_{V_2}(r,t) = (r-t)\ln\left(\frac{r}{t}\right). \tag{45}$$

Example 6. *(Double power entropies) Given two parameters p, q such that $p \geq 1,\ 0 < q \leq 1$ and $p \neq q$, or $p < 0,\ q \geq 1$, the double power entropy of order p, q is given by*

$$W_{p,q}(s) := qs^p - ps^q + p - q, \quad s > 0. \tag{46}$$

$W_{p,q}$ is a strictly convex function, $W_{p,q} \in \mathcal{C}^\infty(0, +\infty)$, and it is extendex in 0 by continuity so that $W_{p,q}(0) = p - q$ when p, q are positive, $W_{p,q}(0) = +\infty$ when $p < 0$.

A direct computation shows that:

$$H_{W_{p,q}}(r,t) = (q-p)rt\left[\frac{(r^{q-1} + t^{q-1})^p}{(r^{p-1} + t^{p-1})^q}\right]^{\frac{1}{p-q}} - (q-p)(r+t). \tag{47}$$

For example, when $p = 3/2, q = 1/2$ one gets

$$H_{W_{\frac{3}{2},\frac{1}{2}}}(r,t) = r + t - (rt)^{\frac{1}{4}}(\sqrt{r} + \sqrt{t}). \tag{48}$$

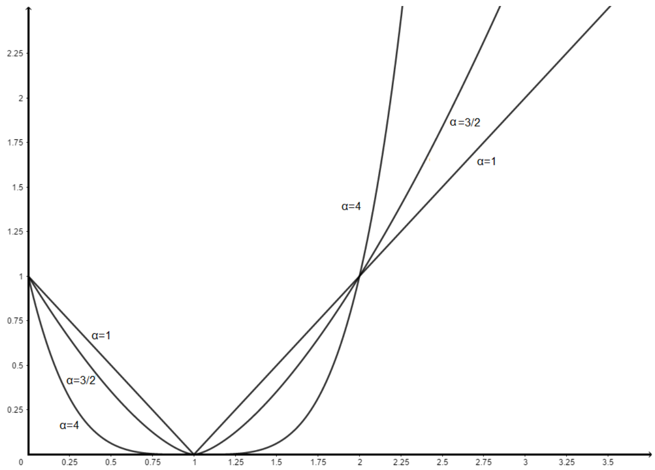

FIGURE 1. χ^α divergences

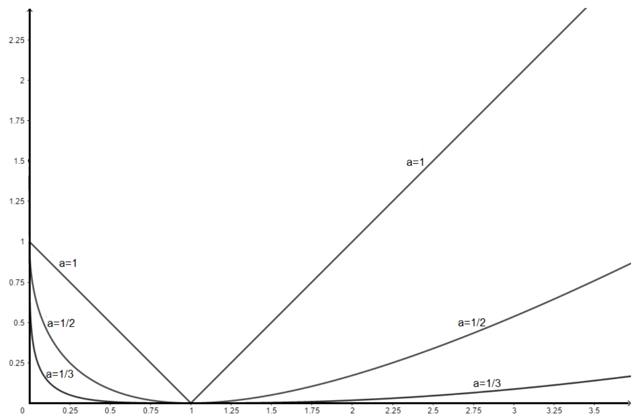

FIGURE 2. Matusita divergences

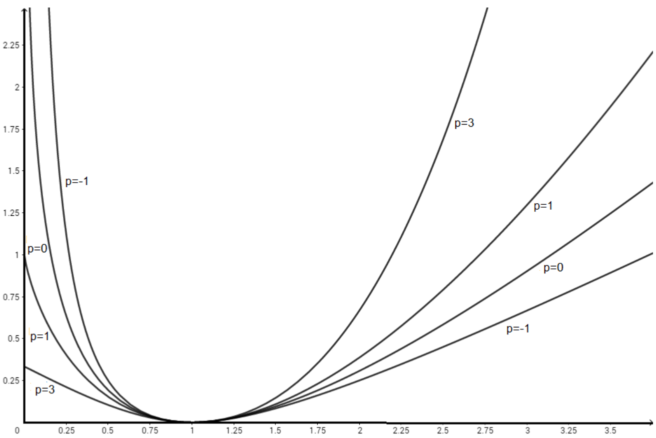

FIGURE 3. Power-like entropies

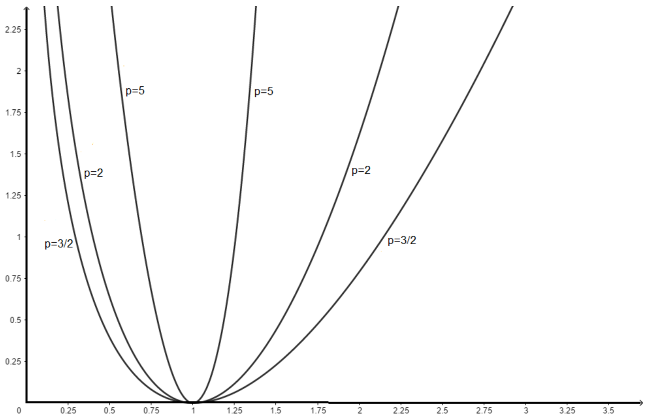

FIGURE 4. Power logarithmic entropies

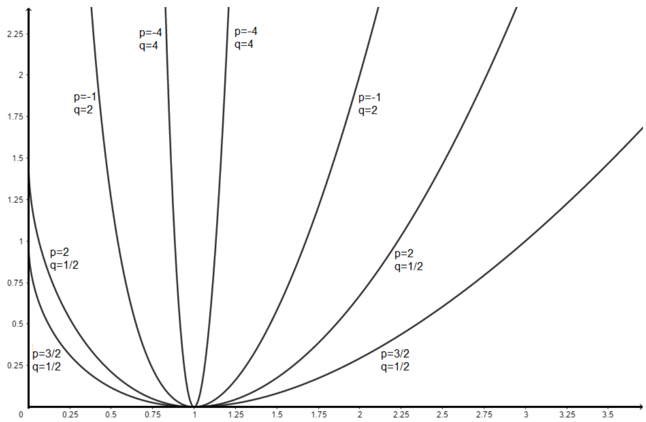

FIGURE 5. Double-power entropies

3.3. **Divergences and triangle inequality.** As we have previously seen, starting from a function $F \in \Gamma_0(\mathbb{R}_+)$ such that $F(s) = 0$ if and only if $s = 1$, the marginal perspective function H_F is nonnegative, symmetric and $H_F(r,t) = 0$ if and only if $r = t$. We begin in this section the discussion regarding the last property that H_F has to fulfill in order to be a metric on $[0,+\infty)$: the triangle inequality.

When we write "d is a metric (or a distance) on X" we mean that $\mathsf{d} : X \times X \to [0,+\infty)$ is a symmetric function, $\mathsf{d}(x,y) = 0$ if and only if $x = y$ and d satisfies the triangle inequality, i.e. $\mathsf{d}(x,z) \leq \mathsf{d}(x,y) + \mathsf{d}(y,z)$ for every $x, y, z \in X$.

Since we will prove that the total variation is the only divergence that is also a metric, it is natural to discuss when the power H_F^a is a distance on $[0,+\infty)$, for $a \in (0,1]$.

The study of the metric properties of the function H_F in order to produce metric divergences is justified by the following Proposition.

Proposition 2. *Let $F \in \Gamma_0(\mathbb{R}_+)$. The functional D_F^a is a distance on $\mathcal{M}(X)$ if and only if $\hat{F}^a$ is a distance on $[0,+\infty)$, $a \in (0,1]$.*

Proof. We recall that, thanks to Lemma 3, $D_F(\mu_1||\mu_2) = \hat{D}_F(\mu_1||\mu_2)$.

Let us suppose that D_F^a is a distance. Then $\hat{F}^a$ is a distance since $D_F^a(r\delta_x||t\delta_x) = \hat{F}^a(r,t)$ for every $r, t \in [0,+\infty)$, where δ_x denotes the Dirac measure at $x \in X$.

For the converse implication, let $\hat{F}^a$ be a distance on the nonnegative real numbers. It is apparent that the (power) of the F-divergence is nonnegative, symmetric and $D_F^a(\mu_1||\mu_2) = 0$ if $\mu_1 = \mu_2$. Moreover, since F^a is a distance, it follows that $F(s) = 0$ if and only if $s = 1$, that also implies $F'_\infty > 0$. Thus, if $D_F^a(\mu_1||\mu_2) = 0$ we easily see using the definition of F-divergence that $\mu_1 = \mu_2$. Let $\mu_1, \mu_2, \mu_3 \in \mathcal{M}(X)$ and consider $\lambda := \mu_1 + \mu_2 + \mu_3$ so that we can write $\mu_i = \tau_i \lambda$, $i = 1, 2, 3$. Then

$$\begin{aligned} D_F^a(\mu_1||\mu_3) &= \Big(\int_X \hat{F}(\tau_1, \tau_3)\mathrm{d}\lambda\Big)^a \leq \Big(\int_X \big(\hat{F}^a(\tau_1, \tau_2) + \hat{F}^a(\tau_2, \tau_3)\big)^{1/a}\mathrm{d}\lambda\Big)^a \\ &\leq \Big(\int_X \big(\hat{F}^a(\tau_1, \tau_2)\big)^{1/a}\mathrm{d}\lambda\Big)^a + \Big(\int_X \big(\hat{F}^a(\tau_2, \tau_3)\big)^{1/a}\mathrm{d}\lambda\Big)^a = D_F^a(\mu_1||\mu_2) + D_F^a(\mu_2||\mu_3), \quad (49) \end{aligned}$$

where we have used the triangle inequality for $\hat{F}^a$ and the Minkowski inequality. □

We recall this simple Lemma:

Lemma 4. *Let (X, d) be a metric space and $f : [0, +\infty) \to [0, +\infty)$ be a concave function such that $f(r) = 0$ if and only if $r = 0$. Then $(X, f(\mathsf{d}))$ is a metric space inducing the same topology.*

Proof. $f(\mathsf{d}(x_1, x_2)) \geq 0$ and $f(\mathsf{d}(x_1, x_2)) = 0$ if and only if $\mathsf{d}(x_1, x_2) = 0$ which implies $x_1 = x_2$. It is clear that

$$(x_1, x_2) \mapsto f(\mathsf{d}(x_1, x_2))$$

is a symmetric function. Since f is concave and $f(r) \geq 0$, f is also increasing and subadditive, thus

$$f\big(\mathsf{d}(x_1, x_3)\big) \leq f\big(\mathsf{d}(x_1, x_2) + \mathsf{d}(x_2, x_3)\big) \leq f\big(\mathsf{d}(x_1, x_2)\big) + f\big(\mathsf{d}(x_2, x_3)\big).$$

The function f is continuous because it is concave and finite valued, so that $f(\mathsf{d})$ is topological equivalent to d. □

An easy consequence of the previous Lemma is that if H_F^a is a metric, then H_F^b is a metric for every $b \in (0, a]$.

The convexity of the function H_F implies that

$$H_F(r, t) \geq H_F(s, t) \text{ and } H_F(r, s) \leq H_F(r, t) \text{ for every } 0 \leq r \leq s \leq t. \quad (50)$$

Using the symmetry, the 1-homogeneity of the function H_F and the property (50), it follows that H_F^a satisfies the triangle inequality if and only if

$$H_F^a(u, 1) \leq H_F^a(u, v) + H_F^a(v, 1) = v^a H_F^a\Big(\frac{u}{v}, 1\Big) + H_F^a(v, 1), \quad \text{for every } 0 \leq u < v < 1. \quad (51)$$

A last useful remark is that

$$\lim_{u \downarrow 0} H_F(u, 1) < +\infty \quad (52)$$

is a necessary condition for the existence of a power a such that H_F^a is a metric.

Regarding the examples previously seen, it was proved by Kafka, Osterreicher and Vincze [KOV91] that $H_{\chi^\alpha}^a$ is a metric when $a = 1/\alpha$.

About the Matusita's divergences, it is apparent that $H_{M_a}^a$ is a distance.

When $p > 1$, $\lim_{u\downarrow 0} H_{V_p}(u, 1) = +\infty$ so that, except for the case $p = 1$, the power-logarithmic entropy is not a metric for every power a.

We now turn the attention to the function $H_p := H_{U_p}$. It has the following expression

$$\begin{cases} H_p(r, t) = \frac{2}{p}\Big[\mathfrak{M}_1(r, t) - \mathfrak{M}_{1-p}(r, t)\Big], & \text{if } p \neq 0, \\ H_0(r, t) = r \ln r + t \ln t - (r + t) \ln\Big(\frac{r+t}{2}\Big), \end{cases} \quad (53)$$

that is also valid when $rt = 0$ with the convention $0 \ln(0) = 0$.

As we have already notice, H_p is the square of a metric on $[0,+\infty)$ for $p=0, p=\frac{1}{2}, p=1$. We investigate now the same question for every real number p. This problem was already considered by Osterreicher and Vajda in the case $p<1$ [Ost96], [OV03]. Following the same approach we prove:

Proposition 3. *The induced marginal perspective function H_p is the square of a metric on the nonnegative real numbers for any $p \in (-\infty, \frac{1}{2}] \cup [1,+\infty)$. $\sqrt{H_p}$ does not satisfy the triangle inequality if $p \in (\frac{1}{2}, 1)$.*

The proof of the previous Proposition is based on the following Lemma. It is the first example in the paper of a fact that will be recurrent: the central role of the class of Matusita's divergences in the study of the metric properties of the marginal perspective function.

Lemma 5. *Let $a \in (0,1]$ and $F \in \Gamma_0^s(\mathbb{R}_+)$. If*

$$h(u) := \frac{(1-u^a)^{\frac{1}{a}}}{F(u)}$$

is decreasing in $[0,1)$, then $\hat{F}^a$ satisfies the triangle inequality.

Proof. Due to the monotonicity of the square root function, one has that

$$h^a(u) = \frac{1-u^a}{F(u)}$$

is decreasing in $[0,1)$, so that $h^a(u) \geq h^a(v)$ and $h^a(u) \geq h^a(\frac{u}{v})$ if $0 \leq u < v < 1$. It follows that

$$\hat{F}^a(u,1) = \frac{1-u^a}{h^a(u)} = \frac{1-v^a}{h^a(u)} + \frac{v^a-u^a}{h^a(u)} \leq \frac{1-v^a}{h^a(v)} + \frac{v^a\left(1-\left(\frac{u}{v}\right)^a\right)}{h^a(\frac{u}{v})} = \hat{F}^a(u,v) + \hat{F}^a(v,1), \tag{54}$$

which is sufficient to prove the triangle inequality since $\hat{F}$ is convex, symmetric and positively 1-homogeneous. □

Proof of Proposition 3. Using now Lemma 5, it remains to show that the function

$$h_p(u) := \frac{(1-\sqrt{u})^2}{f_p(u)}$$

is decreasing in $(0,1)$, where we put $f_p(u) := H_p(u,1)$. The derivative of the function h_p is the following:

$$h_p'(u) = -\frac{2}{p}\left(\frac{1}{\sqrt{u}}-1\right)\frac{1}{f_p^2(u)}\phi_p(u), \tag{55}$$

where we set

$$\phi_p(u) := 2^{-1}(u^{\frac{1}{2}}+1) - 2^{-\frac{1}{1-p}}(u^{1-p}+1)^{\frac{1}{1-p}-1}(u^{\frac{1}{2}-p}+1). \tag{56}$$

Note that $\phi_p(1)=0$ and $\psi_p(u) := \sqrt{u}\phi_p'(u)$ satisfies:

$$\psi_p(u) = \frac{1}{4} - 2^{-\frac{1}{1-p}}(u^{1-p}+1)^{\frac{1}{1-p}-2}u^{-p}\Big(\frac{1+u^{1-p}}{2} - p(1-\sqrt{u})\Big). \tag{57}$$

The function ψ_p is such that $\psi_p(1)=0$ and

$$\psi_p'(u) = 2^{-\frac{1}{1-p}}p(\frac{1}{2}-p)(u^{1-p}+1)^{\frac{1}{1-p}-3}u^{-p-1}(1-\sqrt{u})(1-u^{1-p}). \tag{58}$$

Now let us suppose $p>1$: we have to prove that ϕ_p is positive in $(0,1)$. This is implied by $\psi_p(u)<0$ in $(0,1)$ which is true because $\psi_p'(u)$ is positive in $(0,1)$. Similar considerations can be applied to the case $p<0$ and $p \in (0,\frac{1}{2})$.

For $p \in (\frac{1}{2}, 1)$ one gets $\psi_p'(u) < 0$ in $(0,1)$ so ψ_p is positive in $(0,1)$. This implies that ϕ_p is negative and so h_p is increasing in $(0,1)$. As a consequence, an analysis of the proof of Lemma 5 shows that the triangle inequality is reversed for these values of p. □

Remark 3. *When $p \in (\frac{1}{2}, 1)$, Osterreicher and Vajda ([OV03]) proved that H_p^{1-p} is a metric.*

3.4. **Marginal perspective function and convergence properties.** In subsection 3.1 we have shown that the construction of the marginal perspective function naturally produces a symmetric divergence. In this subsection we show that this is not the only feature of the minimization procedure (31): iterating this process we will highlight the important role played by the class of Matusita's divergences.

We have already seen how to produce a map $T_1 : \Gamma_0(\mathbb{R}_+) \to \Gamma_0^s(\mathbb{R}_+)$: let $F \in \Gamma_0(\mathbb{R}_+)$, we define $T_1(F)(s) := H_F(1,s)$, where H_F is the lower semicontinuous envelope of the function $\tilde{H}_F$ defined by (31). We also denote by $T_a : \Gamma_0(\mathbb{R}_+) \to \Gamma_0^s(\mathbb{R}_+)$ the map $T_a(F) := 2^{\frac{1}{a}-1}T_1(F)$, where $a \in (0,1]$.

It is clear that the two trivial entropies

$$F(s) \equiv 0 \qquad \text{and} \qquad F(s) = I_{\{1\}} = \begin{cases} 0 & \text{if } s = 1, \\ +\infty & \text{otherwise,} \end{cases} \tag{59}$$

are fixed points of the map T_a for any $a \in (0,1]$. Another important property that follows immediately from the definition is that

$$F_1 \geq F_2 \implies T_a(F_1) \geq T_a(F_2). \tag{60}$$

Due to the difference between the case $a = 1$ and the case $0 < a < 1$, we have divided the analysis of the behaviour of the map T_a. Nevertheless, the strategy behind the proofs is in common: we show that, under suitable conditions, the sequence $\{T_a^{(n)}(F)\}$ is monotone and the limit is a fixed point of the map T_a. We then prove that $T_a(F) = F$ implies $F(s) = c|s^a - 1|^{\frac{1}{a}}$, where $c \in [0, +\infty]$ (in the case $c = +\infty$ we mean that $c|s^a - 1|^{\frac{1}{a}} = I_{\{1\}}(s)$).

We start with a simple Lemma that provides a crucial monotonicity property.

Lemma 6. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ and $a \in (0,1]$, if $\hat{F}^a$ satisfies the triangle inequality then $T_a(F) \geq F$.*

Proof. For any $s, t \in \mathbb{R}_+$ the convexity of the function $x \mapsto x^{\frac{1}{a}}$ yields

$$2^{\frac{1}{a}-1}\hat{F}(1,s) + 2^{\frac{1}{a}-1}\hat{F}(s,t) = \frac{1}{2}\big(2\hat{F}^a(1,s)\big)^{\frac{1}{a}} + \frac{1}{2}\big(2\hat{F}^a(s,t)\big)^{\frac{1}{a}} \geq \big(\hat{F}^a(1,s) + \hat{F}^a(s,t)\big)^{\frac{1}{a}} \geq \hat{F}(1,t) = F(t).$$

The result follows by taking the infimum of the left hand side with respect to s. □

Lemma 7. *Given a function $F \in \Gamma_0^s(\mathbb{R}_+)$, the sequence $\{T_1^{(n)}(F)\}$ is decreasing. If we put*

$$\tilde{F}^\infty(s) := \lim_{n\to\infty} T_1^{(n)}(F)(s) \qquad \text{for any } s \in [0, +\infty)$$

and we denote by F^∞ the lower semicontinuous envelope of $\tilde{F}^\infty$, then F^∞ is a fixed point of the map T_1.

Proof. If $F = I_{\{1\}}$ the Lemma trivially holds, so let us suppose that $F \neq I_{\{1\}}$. Since the map $\theta \mapsto \hat{F}(1,\theta) + \hat{F}(\theta, s)$ is equals to $F(1,s)$ when $\theta = 1$ or $\theta = t$, it follows that $T_1(F)(s) \leq F(s)$ for any s. Thus, the sequence $T_1^{(n)}(F)(s)$ is decreasing and it has a limit $\tilde{F}^\infty(s)$ that is clearly convex, nonnegative and satisfies $\tilde{F}^\infty(1) = 0$. The domain of F must contain an interval of the form $[1/b, b]$ for some $b > 1$. By the convexity of the function F, there exists a constant $c \in (0, +\infty)$ such that

$$F \leq G := \begin{cases} c|s-1| & \text{if } \; s \in [1/b, b], \\ +\infty & \text{otherwise.} \end{cases}$$

In particular $T_1^{(n)}(F)(s) \le T_1^{(n)}(G)(s)$ for any s. An easy computation shows that $\mathrm{D}(\tilde{G}^\infty) = (0, +\infty)$ (see below, Lemma 10), which implies $(0, +\infty) \subset \mathrm{D}(\tilde{F}^\infty)$.

Let F^∞ be the lower semicontinuous envelope of the function $\tilde{F}^\infty$. It is immediate that $F^\infty \in \Gamma_0^s(\mathbb{R}_+)$. We have to show that F^∞ is a fixed point of T_1: with the same reasoning as above, one gets $T_1(F^\infty) \le F^\infty$; in order to prove the reverse inequality we notice that for any $1 \le s \le t$ and any $n \in \mathbb{N}$ it holds

$$T_1^{(n)}(F)(s) + sT_1^{(n)}(F)\left(\frac{t}{s}\right) \ge T_1^{(n+1)}(F)(t).$$

The result follows by taking the limit with respect to n, doing the lower semicontinuous envelope and then minimizing with respect to s. □

Theorem 1. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ be a fixed point of T_1. Then $F(s) = c|s-1|$ for a certain $c \in [0, +\infty]$. In particular, an induced marginal perspective function H_F is a metric on $\mathbb{R}_+$ if and only if $H_F = cM_1$, $c \in (0, +\infty)$.*

Proof. It is clear that the function cM_1 is a fixed point of T_1 for any $c \in [0, +\infty]$. We show now that they are the only fixed points: since $\theta \mapsto \hat{F}(1, \theta) + \hat{F}(\theta, s)$ is a convex function that has the same value when $\theta = 1$ and $\theta = s$, $T_1(F) = F$ implies that

$$F(\theta) + \theta F(\frac{s}{\theta}) = F(s) \qquad \text{for every } 1 \le \theta \le s. \tag{61}$$

If $F \ne I_{\{1\}}$, we can also assume that F is finite valued. To see this, let us assume by contradiction that $s > 1$ is such that $F(s) = +\infty$ and $F(\sqrt{s}) < +\infty$. Equation (61) fails with the choice $\theta = \sqrt{s}$.

Take now $s > 2$ and $\theta = 2$, we have

$$F(2) + 2F\Big(\frac{s}{2}\Big) = F(s).$$

If we choose instead $s > 2$ and $\theta = s/2$ we obtain

$$F\Big(\frac{s}{2}\Big) + \frac{s}{2}F(2) = F(s)$$

By taking the difference of the two obtained equations, one gets $F\big(\frac{s}{2}\big) = F(2)\big(\frac{s}{2} - 1\big)$ for any $s > 2$ and we can conclude that $F(s) = c|s-1|$, $c \in [0, +\infty]$, since $F \in \Gamma_0^s(\mathbb{R}_+)$.

To conclude the proof, let us suppose that H_F is a metric. We put $f(s) := H_F(s, 1)$ so that $f \in \Gamma_0^s(\mathbb{R}_+)$. Lemma 6 implies that $T_1(F) \ge F$; Lemma 7 provides the converse inequality. In particular f is a fixed point of T_1, and the only fixed points that induces a metric on $\mathbb{R}_+$ are the functions of the form cM_1 with $c \in (0, +\infty)$. □

In order to deal with the case $0 < a < 1$ we need some preliminary results and some additional assumptions. We start by proving that every metric of the form $\hat{F}^a$, $a \in (0, 1]$, is a complete metric.

Lemma 8. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ and let us suppose that $D := \hat{F}^a$ is a metric for a number $a \in (0, 1]$. Then there exists $c > 0$ such that*

$$F(s) > c|s^a - 1|^{\frac{1}{a}}$$

and $\hat{F}^a$ is a complete metric.

Proof. For any $0 \le u < v < 1$ we rewrite the distance between u and 1 as

$$D(u, 1) = \frac{g(u)}{g(v)}D(v, 1) + \frac{g(u)}{g(\frac{u}{v})}D(u, v) \tag{62}$$

where $g(u) := \dfrac{F^a(u)}{1 - u^a}$. Since the triangle inequality holds, at least one of the numbers $\frac{g(u)}{g(v)}$ and $\frac{g(u)}{g(\frac{u}{v})}$ is less or equal than 1. Choosing $u := v^2$, it follows $g(v^2) \le g(v)$ for any $v < 1$. By contradiction let us suppose it does not exists a positive constant c such that $F(s) > c|s^a - 1|^{\frac{1}{a}}$, so that there exists a sequence $v_n \in (0,1)$ such that $g(v_n) \to 0$. Then, we can find a $\bar{v} \in (0,1)$ such that $D(0,1) = g(0) > g(\bar{v})$. On the other hand, the sequence w_n defined by $w_0 := \bar{v}$, $w_n := w_{n-1}^2$ converges to 0, and by continuity of the function g it follows $g(w_n) \to g(0)$. Since $g(0) > g(w_0)$ and $n \mapsto g(w_n)$ is decreasing, we get a contradiction.

Now it is easy to show that the metric D is complete: since $\hat{F}^a$ is a metric, $\hat{F}$ is symmetric and $D(0,1) = F^a(0) := c_2 < +\infty$. From the convexity of the function F it follows $F^a(s) \le c_2|s-1|^a$ so that

$$c_1 M_a^a \le D \le c_2 M_1^a.$$

The result follows using the fact that M_a^a and M_1^a are two complete metrics inducing the same convergence. □

Let (X, d) be a metric space and put $I = [0,1]$. A curve $\gamma : I \to X$ is a constant speed geodesic if

$$\mathsf{d}\big(\gamma(t), \gamma(t')\big) = \mathsf{d}(\gamma(0), \gamma(1))|t - t'| \quad \text{for every } t, t' \in I. \tag{63}$$

A metric space (X, d) is a geodesic space if for every pair of points $x, y \in X$ there exists a constant speed geodesic between x and y. A well-known fact is that a complete metric space is a geodesic space if and only if for every pair of points $x, y \in X$ there exists $z \in X$ such that $\mathsf{d}(x,z) = \mathsf{d}(z,y) = \frac{1}{2}\mathsf{d}(x,y)$. The point z is called mid-point between x and y.

We now ready to prove the analogous of Theorem 1 in the case $0 < a < 1$, under an additional assumption.

Theorem 2. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ and let us suppose that $\hat{F}^a$ is a distance and $T_a(F) = F$, where $a \in (0,1)$. Then $F(s) = c|s^a - 1|^{\frac{1}{a}}$ for a constant $c \in (0, +\infty)$.*

Proof. We can assume that F is finite valued since $\hat{F}^a$ is a metric. The fixed point property $T_a(F) = F$ implies that for every $s > 1$ it exists θ, $1 \le \theta \le s$, such that

$$2^{\frac{1}{a}-1}\hat{F}(1,\theta) + 2^{\frac{1}{a}-1}\hat{F}(\theta, s) = \hat{F}(1,s), \tag{64}$$

Since $\hat{F}^a$ is a metric and the function $f(x) = x^a$ is concave we can deduce

$$\hat{F}^a(1,s) \le \hat{F}^a(1,\theta) + \hat{F}^a(\theta,s) = \frac{[2^{\frac{1}{a}}\hat{F}(1,\theta)]^a + [2^{\frac{1}{a}}\hat{F}(\theta,s)]^a}{2} \le \Big(2^{\frac{1}{a}-1}\hat{F}(1,\theta) + 2^{\frac{1}{a}-1}\hat{F}(\theta,s)\Big)^a. \tag{65}$$

Equation (64) implies the equality in the inequality (65), which implies $\hat{F}(1,\theta) = \hat{F}(\theta, s)$.

Since $\hat{F}$ is symmetric, positively 1-homogogeneous and $\hat{F}^a$ is a complete metric, Lemma 8 implies that $(\mathbb{R}_+, \hat{F}^a)$ is a one dimensional geodesic space, so it must be isometric to $(\mathbb{R}_+, |\cdot|)$ (for a reference see [BBI01], chapter 2). In particular there exists $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ increasing and continuous such that we can write $\hat{F}^a(r,t) = |\phi(t) - \phi(r)|$. Using again the 1-homogeneity of the function $\hat{F}$, it follows $\hat{F}^a(r,t) = r^a\hat{F}^a(1, \frac{t}{r})$ for $r > 0$, so that

$$\phi(t) - \phi(r) = r^a\Big(\phi\big(\frac{t}{r}\big) - \phi(1)\Big), \quad t \ge r. \tag{66}$$

Evaluating equation (66) for $t = 2r$ we get

$$\phi(2r) - \phi(r) = r^a(\phi(2) - \phi(1)), \quad r \ge 1, \tag{67}$$

whereas the choice $r = 2$ yields

$$\phi(t) - \phi(2) = 2^a\Big(\phi\big(\frac{t}{2}\big) - \phi(1)\Big), \quad t \ge 2. \tag{68}$$

Now consider the previous equation with $t = 2r$, it follows

$$\phi(2r) - \phi(r) = \phi(2) + 2^a(\phi(r) - \phi(1)) - \phi(r), \quad r \ge 1. \tag{69}$$

Using now the identities (67) and (69), it follows

$$r^a(\phi(2) - \phi(1)) = \phi(2) + 2^a(\phi(r) - \phi(1)) - \phi(r) \quad \text{for any } r \ge 1,$$

and we can compute $\phi(r)$ as
$$\phi(r) = \frac{\phi(2) - \phi(1)}{2^a - 1}(r^a - 1) + \phi(1),$$
so that $F^a(r) = (r^a - 1)\dfrac{F^a(2)}{2^a - 1}$ for every $r \geq 1$, which prove the theorem. □

Remark 4. *We do not know if the assumption that $\hat{F}^a$ is a metric can be removed in order to obtain the same characterization as in Theorem 1. The difficulty is that the value of the function $\theta \mapsto 2^{\frac{1}{a}-1}\hat{F}(1,\theta) + 2^{\frac{1}{a}-1}\hat{F}(\theta, s)$ at $\theta = 1$ and $\theta = s$ is strictly greater that $\hat{F}(1, s)$, if $a < 1$.*

In order to obtain that also in the case $0 < a < 1$ the limit function is a fixed point of the map T_a, we need the following Lemma. It is an easy consequence of general results in the theory of Γ-convergence (see e.g. [Mas93]), but we give a direct proof in our simplified setting.

Lemma 9. *Let X be a compact space and let $f_n : X \to [0, +\infty]$ be a sequence of lower semicontinuous functions such that $f_n(x) \leq f_{n+1}(x)$ for every $n \in \mathbb{N}$ and every $x \in X$. Then*
$$\lim_{n\to\infty} \min_{x\in X} f_n(x) = \min_{x\in X} f_\infty(x),$$
where we put $f_\infty(x) := \lim_{n\to\infty} f_n(x)$.

Proof. The functions f_n and f_∞ are lower semicontinuous over a compact set so that they have a minimum. Since $f_n(x) \leq f_\infty(x)$ for every $x \in X$ it is clear that
$$\lim_{n\to\infty} \min_{x\in X} f_n(x) \leq \min_{x\in X} f_\infty(x).$$
Let us suppose now $a < \min_{x\in X} f_\infty(x)$, so that for every $x \in X$ $a < f_\infty(x)$. Since $\lim_n f_n(x) = f_\infty(x)$, there exists $n = n(x)$ such that $a < f_n(x)$. It follows that the family $\{a < f_n\}_{n\in\mathbb{N}}$ is an open cover of X. Let $n_1, ..., n_j$ be a finite collection of indexes such that
$$X \subset \{a < f_{n_1}\} \cup ... \cup \{a < f_{n_j}\}.$$
Let $N := \max\{n_1, ..., n_j\}$, so that $X \subset \{a < f_N\}$ since f_n are increasing. This implies that $a < f_n(x)$ for every $x \in X$ so that $a < \lim_{n\to\infty} \min_{x\in X} f_n(x)$. Since a is an arbitrary number less than $\min_{x\in X} f_\infty(x)$, the Lemma follows. □

We can now state the Theorem about the convergence of the iterations of the map T^a.

Theorem 3. *Let $a \in (0, 1)$. Given a function $F \in \Gamma_0^s(\mathbb{R}_+)$, if $\hat{F}^a$ is a metric then the sequence $\{T_a^{(n)}(F)\}$ converges pointwise to a fixed point of the map T_a. In particular, if the limit function F^∞ is such that $(\hat{F^\infty})^a$ is a metric, then $F^\infty(s) = c|s^a - 1|^{\frac{1}{a}}$ where $c \in (0, +\infty)$.*

Proof. Lemma 6 implies that $T_a(F) \geq F$. By the monotonicity property (60) the sequence $T_a^{(n)}(F)$ is increasing so it converges pointwise to a function F^∞. It is clear that $F^\infty \in \Gamma_0^s(\mathbb{R}_+)$ (recall that the monotone increasing limit of a sequence of lower semicontinuous functions is lower semicontinuous). Since $\hat{F}^a$ is a metric, F is finite finite valued, as well as $T_a^{(n)}(F)$. We want to show that F^∞ is a fixed point of T_a:
$$\begin{aligned} T_a(F^\infty)(s) &= sc^-\Big(2^{\frac{1}{a}-1} \inf_{\theta>0} \big(F^\infty(\theta) + \theta F^\infty(\frac{s}{\theta})\big)\Big) = sc^-\Big(2^{\frac{1}{a}-1} \lim_{n\to\infty} \inf_{\theta>0} \big(T_a^{(n)}(F)(\theta) + \theta T_a^{(n)}(F)(\frac{s}{\theta})\big)\Big) \\ &= sc^-\Big(\lim_{n\to\infty} T_a^{(n+1)}(F)(s)\Big) = F^\infty(s) \end{aligned}$$
where we have denoted by $sc^-(f)$ the lower semicontinuous envelope of the function f and we have used Lemma 9 applied to $f_n(\theta) := T_a^{(n)}(F)(\theta) + \theta T_a^{(n)}(F)(\frac{s}{\theta})$ and $X := [1, s]$. The conclusion follows from Theorem 2. □

Remark 5. *It is not difficult to show that F^∞ can be equal to $I_{\{1\}}$. Take for instance $F(s) = |s-1|$ and consider the sequence $T_a^{(n)}(F)$ with $a \in (0,1)$.*

In the final part of this section we want to study the connection between the behaviour of the function F in a neighborhood of 1 and the limit function F^∞. we start with two lemmas:

Lemma 10. *Let $a \in (0,1]$, $b > 1$, $c \in (0,+\infty)$ and $\bar{F} \in \Gamma_0^s(\mathbb{R}_+)$ be the function defined by*

$$\bar{F}(s) := \begin{cases} c|s^a - 1|^{\frac{1}{a}} & s \in [1/b, b], \\ +\infty & \text{otherwise.} \end{cases}$$

Then

$$\lim_{n\to\infty} T_a^{(n)}(\bar{F})(s) = c|s^a - 1|^{\frac{1}{a}} \qquad s \in (0,+\infty).$$

Proof. It is sufficient to consider the case $s > 1$; it holds

$$T_a(\bar{F})(s) = 2^{\frac{1}{a}-1} \inf_{\theta\in[1,s]} \bar{F}(\theta) + \theta \bar{F}\Big(\frac{s}{\theta}\Big). \tag{70}$$

When $b^2 < s$ it is clear that $T_a(\bar{F})(s) = +\infty$. Moreover, we notice that in the case

$$\mathfrak{M}_a(1,s) \le b, \text{ and } \frac{s}{\mathfrak{M}_a(1,s)} \le b, \tag{71}$$

the expression (70) is minimized by $\theta = \mathfrak{M}_a(1,s)$, so that $T_a(\bar{F})(s) = c|s^a - 1|^{\frac{1}{a}}$ for such an s. Using now the bound given by Proposition 1, we deduce that the inequalities (71) are certainly satisfied when $1 \le s \le 2b-1$. The theorem is now an easy consequence of the fact that the sequence $b_0 := b$, $b_{n+1} := 2b_n - 1$ is strictly increasing and it diverges to $+\infty$. ☐

Lemma 11. *Let $a \in (0,1]$, $b > 1$, $c \in (0,+\infty)$ and $\underline{F} \in \Gamma_0^s(\mathbb{R}_+)$ be the function defined by $\underline{F}(s) := c|s^a - 1|^{\frac{1}{a}}$ when $s \in [1/b, b]$ and extended linearly outside the interval in such a way that the left derivative of $\underline{F}$ at b is the slope of the linear extension in $[b,+\infty)$. Then $\lim_{n\to\infty} T_a^{(n)}(\underline{F})(s) = c|s^a - 1|^{\frac{1}{a}}$.*

Proof. The lemma follows if we prove that

$$\underline{F}^a(t) \le \underline{F}^a(s) + s\underline{F}^a\Big(\frac{t}{s}\Big) \tag{72}$$

for every $1 \le s \le t$. Indeed (72) implies that H^a is a distance, so that, by Theorem 3, $T_a^{(n)}(\underline{F})$ must converge to a function F^∞ that is a fixed point of T_a. Since $T_a^{(n)}(\underline{F})(s) = c|s^a - 1|^{\frac{1}{a}}$ for every n and every $s \in [\frac{1}{b}, b]$, it holds $F^\infty(s) = c|s^a - 1|^{\frac{1}{a}}$ for every $s \in [\frac{1}{b}, b]$ and this implies that $F^\infty(s) = c|s^a - 1|^{\frac{1}{a}}$ for every s. Indeed, let us suppose by contradiction there exists $s_0 > 1$ such that $F^\infty(s_0) \ne c|(s_0)^a - 1|^{\frac{1}{a}}$ and consider the constant $k \ne c$ such that $F^\infty(s_0) = k|(s_0)^a - 1|^{\frac{1}{a}}$. Since F^∞ and $k|(s_0)^a - 1|^{\frac{1}{a}}$ are fixed points of T_a and they coincide in s_0, it must exists another number s_1, $1 < s_1 < s_0$, where they coincide. Iterating the argument it is easy to show that F^∞ and $k|(s_0)^a - 1|^{\frac{1}{a}}$ have to coincide on a sequence of numbers that converges to 1 but this is absurd since $F^\infty(s) = c|s^a - 1|^{\frac{1}{a}}$ for every $s \in [\frac{1}{b}, b]$ and the functions $c|s^a - 1|^{\frac{1}{a}}$ and $k|s^a - 1|^{\frac{1}{a}}$ coincide only at $s = 1$.

It remains to show that (72) holds. We use Lemma 5: we have to prove that the function

$$s \mapsto \frac{|s^a - 1|^{\frac{1}{a}}}{\underline{F}(s)}$$

is increasing in $(1,+\infty)$: this is obvious in the interval $(1,b]$; consider now two numbers r, t such that $b < r < t$. We define $s \mapsto l_r(s)$ to be the affine function that coincide with $\underline{F}$ at b and such that

$l_r(r) = c|r^a - 1|^{\frac{1}{a}}$), and we notice that the convexity of the function $s \mapsto c|s^a - 1|^{\frac{1}{a}}$ implies that the slope of l_r is greater or equal than the positive slope of the function $\underline{F}$ in $(b, +\infty)$. Using again the convexity of the function $c|s^a - 1|^{\frac{1}{a}}$ and the trivial fact that the quotient

$$s \mapsto \frac{l_r(s)}{\underline{F}(s)}$$

is increasing in $(b, +\infty)$, we conclude because

$$\frac{|t^a - 1|^{\frac{1}{a}}}{\underline{F}(t)} \geq \frac{l_r(t)}{\underline{F}(t)} \geq \frac{l_r(r)}{\underline{F}(r)} = \frac{c|r^a - 1|^{\frac{1}{a}}}{\underline{F}(r)}. \tag{73}$$

□

Theorem 4. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ be a function such that*

$$\lim_{s \to 1} \frac{F(s)}{c|s^a - 1|^{\frac{1}{a}}} = 1. \tag{74}$$

Then

$$\lim_{n \to +\infty} T_a^{(n)}(F)(s) = c|s^a - 1|^{\frac{1}{a}} \qquad s \in (0, +\infty). \tag{75}$$

Proof. For every $\epsilon > 0$ there exists a $b > 1$ such that

$$(1 - \epsilon)c|s^a - 1|^{\frac{1}{a}} \leq F(s) \leq (1 + \epsilon)c|s^a - 1|^{\frac{1}{a}}, \quad s \in [1/b, b],$$

so that

$$(1 - \epsilon)\underline{F} \leq F \leq (1 + \epsilon)\bar{F},$$

where $\underline{F}, \bar{F}$ are defined in Lemma 10 and 11. Take now an arbitrary $s \in (0, +\infty)$, from the monotonicity property (60) it follows

$$(1 - \epsilon)T_a^{(n)}(\underline{F}) \leq T_a^{(n)}(F) \leq (1 + \epsilon)T_a^{(n)}(\bar{F}),$$

so that by Lemma 10 and Lemma 11 one gets

$$(1 - \epsilon)c|s^a - 1|^{\frac{1}{a}} \leq \liminf_{n \to \infty} T_a^{(n)}(F)(s) \leq \limsup_{n \to \infty} T_a^{(n)}(F)(s) \leq (1 + \epsilon)c|s^a - 1|^{\frac{1}{a}}.$$

Since ϵ is arbitrary, there exists the limit of $T_a^{(n)}(F)(s)$ and it is equal to $c|s^a - 1|^{\frac{1}{a}}$.

□

4. Marginal perspective cost

4.1. **Marginal perspective function.** In this section we introduce the marginal perspective cost. We modify the definition of the marginal perspective function that we have seen in section 3 in order to take into account the presence of a "spatial" cost function. The construction is motivated by the study of optimal Entropy-Transport problem (see [LMS18, section 5] and the section 5 of the present paper).

First of all, given a number $c \in [0, +\infty)$ and an admissible entropy function F, we define the function $H_c : [0, +\infty) \times [0, +\infty) \to [0, +\infty]$ as the lower semicontinuous envelope of the function

$$\tilde{H}_c(r_1, r_2) := \inf_{\theta > 0} \Big(R\big(\frac{r_1}{\theta}\big) + R\big(\frac{r_2}{\theta}\big) + c\Big)\theta, \tag{76}$$

where R is the reverse entropy function of F. We notice that the function H_0 coincides with the marginal perspective function H_F introduced in section 3. The function $\tilde{H}_c$ can be also computed as

$$\tilde{H}_c(r_1, r_2) = \inf_{\theta > 0} F\Big(\frac{\theta}{r_1}\Big)r_1 + F\Big(\frac{\theta}{r_2}\Big)r_2 + \theta c, \quad \text{if } r_1 \wedge r_2 > 0, \tag{77}$$

or in terms of the perspective function as

$$\tilde{H}_c(r_1, r_2) = \inf_{\theta>0} \hat{F}(\theta, r_1) + \hat{F}(\theta, r_2) + \theta c \tag{78}$$

$$= \inf_{\theta>0} \hat{R}(r_1, \theta) + \hat{R}(r_2, \theta) + \theta c. \tag{79}$$

If $c = +\infty$ we set

$$H_\infty(r_1, r_2) = F(0)r_1 + F(0)r_2. \tag{80}$$

The following lemma, proved in [LMS18] (lemma 5.3), gives a dual characterization of H_c:

Lemma 12. *For every $c \in [0, +\infty]$ the function H_c can be represented as*

$$H_c(r_1, r_2) = \sup\{r_1\psi_1 + r_2\psi_2 : \psi_i \in \mathrm{D}(R^*),\ R^*(\psi_1) + R^*(\psi_2) \le c\}. \tag{81}$$

In particular H_c is lower semicontinuous, convex and positively 1-homogeneous with respect to (r_1, r_2), increasing and concave with respect to c. Moreover, H_c coincides with $\tilde{H}_c$ in the interior of its domain.

4.2. **Induced marginal perspective cost.** Let X be a Polish space and $c = c(x_1, x_2)$ be a cost function $c : X \times X \to [0, +\infty]$. The induced marginal perspective cost is the function

$$H : X \times [0, +\infty) \times X \times [0, +\infty) \to [0, +\infty]$$

defined as

$$H(x_1, r; x_2, t) := H_{c(x_1,x_2)}(r, t). \tag{82}$$

For our purposes, a particularly important case is when the cost c is induced by the metric d on X (see Section 6). In this setting, we are interested in determining when the function H is the power of a distance on the corresponding cone space. The latter is the space $\mathfrak{C}(X) = Y/\sim$, where $Y = X \times [0, +\infty)$ and

$$(x_1, r) \sim (x_2, t) \iff r_1 = r_2 = 0 \text{ or } r = t,\ x_1 = x_2\,. \tag{83}$$

It is important to highlight that the space $\mathfrak{C}(X)$ can be endowed with a "natural" metric $\mathsf{d}_{\mathfrak{C}}$ (see [BBI01, Prop. 3.6.13]):

$$\mathsf{d}_{\mathfrak{C}}^2\big((x_1, r), (x_2, t)\big) = r^2 + t^2 - 2rt\cos(\mathsf{d}(x_1, x_2) \wedge \pi). \tag{84}$$

Proposition 4. *Let $F(s)$ be an admissible entropy function with a strict minimum at $s = 1$ and let $c : X \times X \to [0, \infty]$ be a symmetric function such that $c(x_1, x_2) = 0$ if and only if $x_1 = x_2$. Then the induced marginal perspective cost H is nonnegative, symmetric in the sense that $H(x_1, r; x_2, t) = H(x_2, t; x_1, r)$ for every $x_1, x_2 \in X$, $r, t \in \mathbb{R}_+$, and $H(x_1, r; x_2, t) = 0$ if and only if $(x_1, r) \sim (x_2, t)$.*

Proof. It is clear that $H \ge 0$. When $r = t = 0$ it follows from the dual representation (81) that $H(x_1, r; x_2, t) = 0$. If $(x_1, r) \sim (x_2, t)$ and $r = t > 0$ then $c(x_1, x_2) = 0$ and the fact that the marginal perspective cost is null follows from the possible choice $\theta = r$ in the expression (76). Since c is symmetric it is also apparent that

$$H(x_1, r; x_2, t) = H(x_2, t; x_1, r).$$

It remains to prove that $H = 0$ implies $(x_1, r) \sim (x_2, t)$. Lemma 2 and equation (19) tell us that R^* is an increasing homeomorphism between $(-\mathrm{aff} F_\infty, F(0))$ and $(-F'_\infty, -F'_0)$ with $R^*(0) = 0$. Since $F(s)$ is a convex function with a strict minimum at $s = 1$, it holds $\mathrm{aff} F_\infty > 0,\ F(0) > 0,\ F'_\infty > 0, F'_0 < 0$. In particular, there exists a positive number $k > 0$ such that the function R^* is finite, continuous and strictly increasing in $(-k, k)$. Hence, it follows again from the representation (81) that $H(x_1, r; x_2, t) = 0$ and $c(x_1, x_2) > 0$ implies $r = t = 0$. Moreover, when $c(x_1, x_2) = 0$ we must have $r = t$: suppose by contradiction that $0 = r < t$ (the other case is similar), in the equation (81) we find $-k < \psi_1 < 0 < \psi_2 < k$ such that $R^*(\psi_1)+R^*(\psi_2) \leq 0$, contradicting the fact $H = 0$. Finally, when $H(x_1, r; x_2, t) = 0$, $c(x_1, x_2) = 0$ and r, t are positive we can prove that $r = t$ using the fact that $\tilde{H}_0 = 0$ implies $r = t$ because, using now the expression (77), we know that for every natural n there exists θ_n such that

$$0 \leq F\Big(\frac{\theta_n}{r}\Big)r + F\Big(\frac{\theta_n}{t}\Big)t < \frac{1}{n}.$$

In particular, for n large enough, $\theta_n \in [K_1, K_2]$ for some constants $0 < K_1 < 1 < K_2$, and by extracting a subsequence θ_{n_j} it follows that $\theta_{n_j} \to \bar{\theta}$. The lower semicontinuity of F forces $\frac{\bar{\theta}}{r} = \frac{\bar{\theta}}{t} = 1$ so that $r = t$. □

5. Entropy-Transport problem

Let X be a Polish space and $c : X \times X \to [0, +\infty]$ be a lower semicontinuous cost function. We define the Transport functional as

$$\mathcal{T} : \mathcal{M}(X \times X) \to [0, +\infty], \qquad \mathcal{T}(\boldsymbol{\gamma}) := \int_{X\times X} c(x_1, x_2)\mathrm{d}\boldsymbol{\gamma}. \tag{85}$$

The classical Transport problem between the measures $\mu_1 \in \mathcal{M}(X)$, $\mu_2 \in \mathcal{M}(X)$ is then defined as

$$\mathrm{T}(\mu_1, \mu_2) := \inf_{\boldsymbol{\gamma}\in\mathsf{M}} \mathcal{T}(\boldsymbol{\gamma}), \tag{86}$$

where the infimum runs over the set

$$\mathsf{M} := \{\boldsymbol{\gamma} \in \mathcal{M}(X \times X) : (\pi^1)_\sharp(\boldsymbol{\gamma}) = \mu_1, (\pi^2)_\sharp(\boldsymbol{\gamma}) = \mu_2\}. \tag{87}$$

Here π^i denote the standard projection from $X \times X$ to X, $i = 1, 2$. The condition on the marginals forces the measures μ_1 and μ_2 to have equal mass, otherwise $\mathsf{M} = \emptyset$.

Optimal Entropy-Transport problems arise naturally when one tries to relax the request on the marginals (87). Let $F \in \Gamma_0(\mathbb{R}_+)$ be a given entropy function, we define the Entropy-Transport functional as

$$\mathcal{ET}(\boldsymbol{\gamma}||\mu_1, \mu_2) := D_F(\gamma_1||\mu_1) + D_F(\gamma_2||\mu_2) + \mathcal{T}(\boldsymbol{\gamma}), \qquad \gamma_i := (\pi^i)_\sharp\boldsymbol{\gamma}, \tag{88}$$

and the Entropy-Transport problem between the measures μ_1, μ_2 as the minimization problem

$$\mathrm{ET}(\mu_1, \mu_2) := \inf_{\boldsymbol{\gamma}\in\mathcal{M}(X\times X)} \mathcal{ET}(\boldsymbol{\gamma}||\mu_1, \mu_2). \tag{89}$$

We notice that the presence of the F-divergences in the cost functional $\mathcal{ET}$ allows to minimize with respect to any measure $\boldsymbol{\gamma} \in \mathcal{M}(X \times X)$ without imposing any condition on the marginals, but it penalizes the measures that do not satisfy the constraint (87) (at least when F have a strict minimum at 1).

Let us recall that we can associate to the entropy function F and the cost c the marginal perspective cost H, whose construction was introduced in section 4. We define the integral functional associated to the marginal perspective cost as

$$\mathcal{H}(\mu_1, \mu_2||\boldsymbol{\gamma}) := \int_{X\times X} H(x_1, \rho_1(x_1); x_2, \rho_2(x_2))\mathrm{d}\boldsymbol{\gamma} + \sum_{i=1}^{2} F(0)\mu_i^\perp(X), \tag{90}$$

where $\mu_i = \rho_i\gamma_i + \mu_i^\perp$ is the Lebesgue's decomposition of the measure μ_i with respect to the marginal $\gamma_i := (\pi^i)_\sharp\boldsymbol{\gamma}$. Another equivalent formulation of the Entropy-Transport problem can be given in terms of the $\mathcal{H}$-integral functional as it is proved in the next result ([LMS18, Theorem 5.5]).

Proposition 5. *For every $\mu_1 \in \mathcal{M}(X)$, $\mu_2 \in \mathcal{M}(X)$, it holds*

$$\mathrm{ET}(\mu_1, \mu_2) = \inf_{\boldsymbol{\gamma}\in\mathcal{M}(X\times X)} \mathcal{H}(\mu_1, \mu_2||\boldsymbol{\gamma}). \tag{91}$$

It is interesting to notice that the class of Entropy-Transport problems is a generalization of the concept of F-divergence. In fact, one can recover the usual pure entropic problem with the choices

$$F \in \Gamma_0(\mathbb{R}_+),\ F'_\infty = +\infty \quad \text{and} \quad c(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise,} \end{cases} \tag{92}$$

that produce (see [LMS18, example E.5])

$$\mathrm{ET}(\mu_1 || \mu_2) = \int_X H_F\Big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda}, \frac{\mathrm{d}\mu_2}{\mathrm{d}\lambda}\Big)\mathrm{d}\lambda, \tag{93}$$

where $\lambda \in \mathcal{M}(X)$ is any dominating measure of μ_1, μ_2 and H_F is the marginal perspective function defined in (31). Thus, the entropy-transport cost ET in this case simply corresponds to the $\mathcal{H}$-divergence between μ_1 and μ_2 defined in (30) associated to the function H_F.

In analogy with the above discussion on the pure entropic case, the advantage of the $\mathcal{H}$-formulation (90) is that it is more suitable for the study of the metric properties of Entropy-Transport problems.

Proposition 6. *Let X be a Polish space and let $a \in (0,1]$. Then ET^a is a distance on $\mathcal{M}(X)$ if the function H^a is a distance on the cone $\mathfrak{C}(X)$.*

Proof. The proof follows from the results obtained by Liero, Mielke and Savaré in [LMS18, Sections 5 and 7], where also much more is proved for the Hellinger-Kantarovich distance (see below for the definition). For a direct proof of this fact, see also [Chi+18, Theorem 3.7]. □

6. Triangle inequality in the Entropy-Transport case

In this section we take advantage of Proposition 6 in order to produce new distances in the space of measures coming from Entropy-Transport problems. We consider a Polish space X with a metric d and the class of power-like entropies $F = U_p$, $p \in \mathbb{R}$. Finally, the cost function $c : X \times X \to [0, +\infty]$ will be a suitable function of the given metric d to be specified later, i.e. $c(x_1, x_2) := f(\mathsf{d}(x_1, x_2))$ for a certain $f : [0, +\infty) \to [0, +\infty]$.

With these choices, we denote by H_p the induced marginal perspective cost. If $p \neq 0, 1$ it holds:

$$H_p(x_1, r; x_2, t) = \frac{2}{p}\left[\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)\left(1 + (1-p)\frac{c(x_1,x_2)}{2}\right)_+^{\frac{p}{p-1}}\right]. \tag{94}$$

When $p = 1$ or $p = 0$ one gets:

$$H_1(x_1, r; x_2, t) = 2\Big[\mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)e^{-c(x_1,x_2)/2}\Big], \tag{95}$$

$$H_0(x_1, r; x_2, t) = r\ln r + t\ln t - (r+t)\ln\left(\frac{r+t}{2 + c(x_1,x_2)}\right), \tag{96}$$

with standard meaning when $c = +\infty$.

For a general cost function c, if $p < 1$, we notice that $H_p(x_1, 0; x_2, t)$ depends on x_1 since $\mathfrak{M}_{1-p}(0,t) > 0$. In particular, H_p is not well defined on the cone $\mathfrak{C}(X)$.

When $p \geq 1$ the situation is more interesting.

Theorem 5. *Let d be a metric on a Polish space X and let $h : [0, +\infty) \to [0, 1]$ be a decreasing function such that $h(0) = 1$. Let us suppose that*

$$\bar{H}_1(x_1, r; x_2, t) := \mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)h\left(\mathsf{d}(x_1,x_2)\right) \tag{97}$$

is the square of a distance on the cone $\mathfrak{C}(X)$. Then

$$\bar{H}_p(x_1, r; x_2, t) := \mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)h\left(\mathsf{d}(x_1,x_2)\right) \tag{98}$$

is the square of a distance on the cone $\mathfrak{C}(X)$ for every $p > 1$.

Proof. Recalling the properties of the power means we have seen in subsection 2.3, it is apparent that $\bar{H}_p$ is nonnegative, symmetric and $\bar{H}_p(x_1, r; x_2, t) = 0$ if and only if $(x_1, r) \sim (x_2, t)$.

We have to prove that for every $p > 1$, for every metric d on X and for every $r, s, t \in [0, +\infty)$, $x_1, x_2, x_3 \in X$ it holds:

$$\sqrt{\bar{H}_p(x_1, r; x_3, t)} \leq \sqrt{\bar{H}_p(x_1, r; x_2, s)} + \sqrt{\bar{H}_p(x_2, s; x_3, t)}. \tag{99}$$

Since the function

$$d \mapsto \sqrt{\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)h(d)} \tag{100}$$

is increasing in $[0, +\infty)$ we can assume

$$\mathsf{d}(x_1, x_3) = \mathsf{d}(x_1, x_2) + \mathsf{d}(x_2, x_3).$$

Without loss of generality, we can also assume $r \leq t$ and we have to deal with three cases:

- $t < s$,
- $r < s \leq t$, (101)
- $s \leq r$.

Step 1. *Case $t < s$*

Lemma 13. *For any fixed r, t, x_1, x_2, x_3, the function*

$$s \mapsto \sqrt{\bar{H}_p(x_1, r; x_2, s)} + \sqrt{\bar{H}_p(x_2, s; x_3, t)} \tag{102}$$

is increasing in $[t, +\infty)$.

Proof. The result follows if we prove that for any fixed x_1, x_2 the function

$$f_p(u) = \bar{H}_p(x_1, 1; x_2, u)$$

is increasing in $[1, +\infty)$. This is easy to prove since

$$f_p'(u) = \frac{1}{2} - \frac{u^{-p}}{2}\Big(\frac{1+u^{1-p}}{2}\Big)^{\frac{p}{1-p}} h(\mathsf{d}(x_1,x_2)) \geq \frac{1}{2} - \frac{u^{-p}}{2}\Big(\frac{1+u^{1-p}}{2}\Big)^{\frac{p}{1-p}} > 0, \tag{103}$$

where the last inequality holds because it is equivalent to the following

$$\mathfrak{M}_{1-p}(1, u) < u.$$

□

Thus, it is sufficient to prove the triangle inequality when $s \leq t$.

Step 2. *Case $r < s \leq t$*

We start with a useful lemma:

Lemma 14. *Let A, B, C three nonnegative numbers. Then*

$$\sqrt{C} \leq \sqrt{A} + \sqrt{B} \tag{104}$$

if and only if for every $\alpha, \beta \in (0,1)$ such that $\alpha + \beta = 1$ we have

$$C \leq \frac{A}{\alpha} + \frac{B}{\beta}. \tag{105}$$

Proof. Let us suppose (104). Then

$$C \leq \Big(\alpha\frac{\sqrt{A}}{\alpha} + \beta\frac{\sqrt{B}}{\beta}\Big)^2 \leq \frac{A}{\alpha} + \frac{B}{\beta}$$

where we have used the Jensen inequality for the convex function $f(x) = x^2$. In order to show that (105) $\Rightarrow$ (104) we notice that if $A = 0$ or $B = 0$ the result is clearly true, otherwise we choose α, β such that $\frac{\sqrt{A}}{\alpha} = \frac{\sqrt{B}}{\beta}$. Thus

$$\big(\sqrt{A} + \sqrt{B}\big)^2 = \Big(\alpha\frac{\sqrt{A}}{\alpha} + \beta\frac{\sqrt{B}}{\beta}\Big)^2 = \frac{A}{\alpha} + \frac{B}{\beta} \geq C.$$

□

In order to simplify the notation, from now on we put $\mathsf{d}(x_1, x_3) = d_{13},\ \mathsf{d}(x_1, x_2) = d_{12},\ \mathsf{d}(x_2, x_3) = d_{23}$. Then, we can use Lemma 14 and the triangle inequality in the case $p = 1$ in order to derive a new inequality. Given $\alpha, \beta \in (0, 1)$ such that $\alpha + \beta = 1$, one gets:

$$\begin{aligned}
\bar{H}_p(x_1, r; x_3, t) &= \bar{H}_1(x_1, r; x_3, t) + \Big[\mathfrak{M}_0(r, t) - \mathfrak{M}_{1-p}(r, t)\Big] h(d_{13}) \leq \\
&\frac{\bar{H}_1(x_1, r; x_2, s)}{\alpha} + \frac{\bar{H}_1(x_2, s; x_3, t)}{\beta} + \Big[\mathfrak{M}_0(r, t) - \mathfrak{M}_{1-p}(r, t)\Big] h(d_{13}) \leq \\
\frac{\bar{H}_p(x_1, r; x_2, s)}{\alpha} &- \frac{\Big[\mathfrak{M}_0(r, s) - \mathfrak{M}_{1-p}(r, s)\Big] h(d_{12})}{\alpha} + \frac{\bar{H}_p(x_2, s; x_3, t)}{\beta} - \frac{\Big[\mathfrak{M}_0(s, t) - \mathfrak{M}_{1-p}(s, t)\Big] h(d_{23})}{\beta} \\
&+ \Big[\mathfrak{M}_0(r, t) - \mathfrak{M}_{1-p}(r, t)\Big] h(d_{13}) \leq \frac{\bar{H}_p(x_1, r; x_2, s)}{\alpha} + \frac{\bar{H}_p(x_2, s; x_3, t)}{\beta},
\end{aligned} \tag{106}$$

where the last inequality in (106) is valid if and only if (using again Lemma 14):

$$\sqrt{\Big[\mathfrak{M}_0(r, t) - \mathfrak{M}_{1-p}(r, t)\Big] h(d_{13})} \leq \sqrt{\Big[\mathfrak{M}_0(r, s) - \mathfrak{M}_{1-p}(r, s)\Big] h(d_{12})} + \sqrt{\Big[\mathfrak{M}_0(s, t) - \mathfrak{M}_{1-p}(s, t)\Big] h(d_{23})}. \tag{107}$$

We notice that $h(d_{13}) \leq h(d_{12}) \wedge h(d_{23})$ since h is decreasing. Thus, it is enough to prove (107) in the case $d_{13} = d_{12} = d_{23} = 0$. Now, we adapt the strategy used in the proof of [ES03, Lemma 2] we put $u := \frac{r}{s} \in (0, 1)$, $\beta u := \frac{t}{s} \in (1, +\infty)$, so that β is a real number greater than 1. Thus, $\frac{1}{\beta} < u < 1$ and, denoted by $F(s)$ the function

$$F(s) = \sqrt{\Big[\mathfrak{M}_0(r, s) - \mathfrak{M}_{1-p}(r, s)\Big]} + \sqrt{\Big[\mathfrak{M}_0(s, t) - \mathfrak{M}_{1-p}(s, t)\Big]}, \tag{108}$$

it follows

$$4\sqrt{s}\frac{d}{ds}F(s) = g_p(u) + g_p(\beta u), \tag{109}$$

where

$$g_p(u) := \frac{\mathfrak{M}_0(u, 1) - \frac{2}{u^{1-p}+1}\mathfrak{M}_{1-p}(u, 1)}{\sqrt{\mathfrak{M}_0(u, 1) - \mathfrak{M}_{1-p}(u, 1)}}. \tag{110}$$

Lemma 15. *The function*

$$u \mapsto g_p(u) + g_p(\beta u)$$

is increasing in $(\frac{1}{\beta},1)$ *with only one zero inside the interval, so that* F *is minimized when* $s=r$ *or* $s=t$ *and the inequality* (107) *holds.*

Proof. Since g_p is continuous in $(0,1)$ and $(1,+\infty)$, it is enough to show that g_p is increasing in $(0,1)$ and $(1,+\infty)$, and

$$\lim_{u\to 1^-} g_p(u)=\sqrt{2(p-1)},\quad \lim_{u\to 1^+} g_p(u)=-\sqrt{2(p-1)}.$$

The limits are easy to compute expanding the function near $u=1$. When $u\in(0,1)\cup(1,+\infty)$ it follows:

$$g_p'(u)=\frac{(p-\frac{1}{2})u^{-p}\Big(\frac{u^{1-p}+1}{2}\Big)^{\frac{2p}{1-p}}-pu^{-p+\frac{1}{2}}\Big(\frac{u^{1-p}+1}{2}\Big)^{\frac{2p-1}{1-p}}+\frac{1}{2}}{2\big[\mathfrak{M}_0(u,1)-\mathfrak{M}_{1-p}(u,1)\big]^{\frac{3}{2}}}. \tag{111}$$

The proof is complete if we show that

$$(p-\frac{1}{2})u^{-p}\Big(\frac{u^{1-p}+1}{2}\Big)^{\frac{2p}{1-p}}-pu^{-p+\frac{1}{2}}\Big(\frac{u^{1-p}+1}{2}\Big)^{\frac{2p-1}{1-p}}+\frac{1}{2}>0$$

for any $p>1$ and any positive u. We put $v=\frac{u^{1-p}+1}{2}$, so that we have to prove

$$(p-\frac{1}{2})\Big(\frac{2v-1}{v^2}\Big)^{\frac{-p}{1-p}}-p\Big(\frac{2v-1}{v^2}\Big)^{\frac{1-2p}{2(1-p)}}+\frac{1}{2}>0$$

for any $p>1$ and $v\in(\frac{1}{2},+\infty)$. Finally we put $w=\Big(\frac{2v-1}{v^2}\Big)^{\frac{1}{p-1}}\in(0,1)$ and we prove that

$$z_p(w):=(p-\frac{1}{2})w^p-pw^{p-\frac{1}{2}}+\frac{1}{2}>0,$$

for any $p>1$ and $w\in(0,1)$. To prove the last inequality, we notice that $z_p(1)=0$ and z_p is a decreasing function because

$$z_p'(w)=p(p-\frac{1}{2})w^{p-\frac{3}{2}}(\sqrt{w}-1)<0.$$

□

Step 3. *Case* $s\le r$

The strategy is to use again Lemma 14 and the triangle inequality for the case $p=1$, but we have to derive a different inequality with respect to the previous step.

Lemma 16. *We denote with* $\theta_p:[0,+\infty)\times[0,+\infty)\to[0,+\infty)$ *the function*

$$\theta_p(r,t):=\frac{\mathfrak{M}_{1-p}(r,t)}{\mathfrak{M}_0(r,t)}.$$

Then $\theta_p(s,t)\le\theta_p(r,t)$.

Proof. It is sufficient to prove that $\theta_p(u,1)$ is increasing in $(0,1)$. This is easy to prove, indeed

$$\sqrt{u}\frac{d}{du}\theta_p(u,1)=\theta_p(u,1)\Big(\frac{u^{1-p}}{u^{1-p}+1}-\frac{1}{2}\Big)\ge 0.$$

□

Let α,β be any two numbers in $(0,1)$ such that $\alpha+\beta=1$. Let us suppose, at first, $\theta_p(s,r)\le\theta_p(r,t)$. Then

$$\begin{aligned}\bar{H}_p(x_1,r;x_3,t) &= \bar{H}_1(x_1,r;x_3,t)\theta_p(r,t)+\mathfrak{M}_1(r,t)\big(1-\theta_p(r,t)\big) \le \\ &\frac{\bar{H}_1(x_1,r;x_2,s)}{\alpha}\theta_p(r,t)+\frac{\mathfrak{M}_1(r,s)}{\alpha}\big(1-\theta_p(r,t)\big)+\frac{\bar{H}_1(x_2,s;x_3,t)}{\beta}\theta_p(r,t)+\frac{\mathfrak{M}_1(s,t)}{\beta}\big(1-\theta_p(r,t)\big) \le \\ &\frac{\bar{H}_p(x_1,r;x_2,s)}{\alpha}+\frac{\bar{H}_p(x_2,s;x_3,t)}{\beta}, \end{aligned} \tag{112}$$

where the first inequality of (112) follows by the triangle inequalities satisfied by $\sqrt{\bar{H}_1}$ and by $\sqrt{\mathfrak{M}_1}$ (where the latter is straightforward to prove), while the second inequality follows because

$$\theta_p(s,r)\le\theta_p(r,t),\quad \theta_p(s,t)\le\theta_p(r,t)\quad\text{and}\quad \bar{H}_1\le\mathfrak{M}_1.$$

It remains to investigate the case $\theta_p(s,r)>\theta_p(r,t)$. Let us suppose

$$\sqrt{\mathfrak{M}_1(r,t)\big(1-\theta_p(r,t)\big)}\le\sqrt{\mathfrak{M}_1(s,r)\big(1-\theta_p(s,r)\big)}+\sqrt{\mathfrak{M}_1(s,t)\big(1-\theta_p(r,t)\big)}. \tag{113}$$

Then

$$\begin{aligned}\bar{H}_p(x_1,r;x_3,t) &= \bar{H}_1(x_1,r;x_3,t)\theta_p(r,t)+\mathfrak{M}_1(r,t)\big(1-\theta_p(r,t)\big) \le \\ &\frac{\bar{H}_1(x_1,r;x_2,s)}{\alpha}\theta_p(r,t)+\frac{\mathfrak{M}_1(s,r)}{\alpha}\big(1-\theta_p(s,r)\big)+\frac{\bar{H}_1(x_2,s;x_3,t)}{\beta}\theta_p(r,t)+\frac{\mathfrak{M}_1(s,t)}{\beta}\big(1-\theta_p(r,t)\big) \le \\ &\frac{\bar{H}_1(x_1,r;x_2,s)}{\alpha}\theta_p(s,r)+\frac{\mathfrak{M}_1(s,r)}{\alpha}\big(1-\theta_p(s,r)\big)+\frac{\bar{H}_1(x_2,s;x_3,t)}{\beta}\theta_p(r,t)+\frac{\mathfrak{M}_1(s,t)}{\beta}\big(1-\theta_p(r,t)\big) \le \\ &\frac{\bar{H}_p(x_1,r;x_2,s)}{\alpha}+\frac{\bar{H}_p(x_2,s;x_3,t)}{\beta}, \end{aligned} \tag{114}$$

where in the first inequality we use (113), in the second we use the hypothesis $\theta_p(s,r)>\theta_p(r,t)$, in the third we reason as in the second step of the inequality (112) in order to replace $\theta_p(r,t)$ with $\theta_p(s,t)$.

Finally, the proof is complete if we prove the inequality (113). Since the case $r=s$ is trivial, we put $u:=\frac{s}{r}<1$, $v:=\frac{t}{r}>1$, so that we can rewrite the inequality (113) in the following equivalent way

$$(1+\sqrt{u})^2\frac{\mathfrak{M}_0(u,1)-\mathfrak{M}_{-1}(u,1)}{\mathfrak{M}_0(u,1)-\mathfrak{M}_{1-p}(u,1)}\le\frac{\sqrt{v}\big(\sqrt{u+v}+\sqrt{1+v}\big)^2}{\mathfrak{M}_0(1,v)-\mathfrak{M}_{1-p}(1,v)}. \tag{115}$$

Now we use the estimate

$$\big(\sqrt{u+v}+\sqrt{1+v}\big)^2\ge 1+4v,$$

so that it is sufficient to prove that for any $u\in(0,1)$ and any $v\in(1,+\infty)$

$$(1+\sqrt{u})^2\frac{\mathfrak{M}_0(u,1)-\mathfrak{M}_{-1}(u,1)}{\mathfrak{M}_0(u,1)-\mathfrak{M}_{1-p}(u,1)}\le\frac{\sqrt{v}(1+4v)}{\mathfrak{M}_0(1,v)-\mathfrak{M}_{1-p}(1,v)} \tag{116}$$

It is easy to see that the last inequality is true at least if $p\ge\frac{3}{2}$. For example, one can bound the left hand side with

$$l(u):=(1+\sqrt{u})^2\frac{\mathfrak{M}_0(u,1)-\mathfrak{M}_{-1}(u,1)}{\mathfrak{M}_0(u,1)-\mathfrak{M}_{-\frac{1}{2}}(u,1)},$$

and the right hand side with

$$r(u):=\frac{\sqrt{v}(1+4v)}{\sqrt{v}-1}.$$

Then, standard computations show that:

$$\sup_{u\in(0,1)} l(u) < \inf_{u\in(1,+\infty)} r(u).$$

If $1 < p < \frac{3}{2}$ one needs precise bounds that we have found in [Kou14]. The supremum of the left hand side of (116) is $\frac{4}{p-1}$. For the right hand side of (116) one has:

$$\frac{\sqrt{v}(1+4v)}{\mathfrak{M}_0(v,1)-\mathfrak{M}_{1-p}(v,1)} = \frac{\mathfrak{M}_{p-1}(1,v)(1+4v)}{\mathfrak{M}_{p-1}(1,v)-\mathfrak{M}_0(1,v)} \geq \frac{\sqrt{v}(1+4v)}{\mathfrak{M}_{p-1}(1,v)-\mathfrak{M}_0(1,v)} \geq 4\frac{\mathfrak{M}_1(1,v)-\mathfrak{M}_0(1,v)}{\mathfrak{M}_{p-1}(1,v)-\mathfrak{M}_0(1,v)}, \tag{117}$$

and again using the results in [Kou14] it is proved that the sharp lower bound for the last expression is $\frac{4}{p-1}$. □

In [LMS18], the authors discussed the case $p = 1$ and $c(x_1, x_2) = \mathsf{d}^2(x_1, x_2)$, which produces the Gaussian Hellinger-Kantorovich distance generated by the marginal perspective cost

$$H(x_1, r; x_2, t) := r + t - 2\sqrt{rt}\exp\left(-\mathsf{d}^2(x_1, x_2)/2\right), \tag{118}$$

and the case $p = 1$ and

$$c_{\mathrm{HK}}(x_1, x_2) = \begin{cases} -\log\left(\cos^2(\mathsf{d}(x_1, x_2))\right) & \text{if } \mathsf{d}(x_1, x_2) < \frac{\pi}{2}, \\ +\infty & \text{otherwise,} \end{cases}$$

which produces the Hellinger-Kantorovich distance generated by the marginal perspective cost

$$H(x_1, r; x_2, t) := r + t - 2\sqrt{rt}\cos(\mathsf{d}(x_1, x_2) \wedge \pi/2). \tag{119}$$

It can be easily proved that (118) and (119) are the square of a distance in $\mathfrak{C}(X)$ since the expressions are strictly related to the metric $\mathsf{d}_{\mathfrak{C}}$ defined in (84) (see [LMS18, Section 7]).

In the next Proposition we discuss some possible choices for the cost function c and we produce new classes of metrics in the cone space $\mathfrak{C}(X)$ (thus, also in $\mathcal{M}(X)$ thanks to Proposition 6).

Theorem 6. *Let (X, d) be a Polish space, $F = U_p$ and consider one of the two following cost functions:*

(1) $c_p(x_1, x_2) := \frac{2}{p-1}\left[1 - \left(\cos(\mathsf{d}(x_1, x_2) \wedge \pi/2)\right)^{\frac{p-1}{p}}\right]$.

(2) $c(x_1, x_2) = \mathsf{d}(x_1, x_2)$.

Then, the induced marginal perspective cost is the square of a distance on $\mathfrak{C}(X)$ for every $p > 1$.

Moreover, if

(3) $c(x_1, x_2) = \mathsf{d}^2(x_1, x_2)$

the induced marginal perspective cost is the square of a distance on $\mathfrak{C}(X)$ for every $1 < p \leq 3$.

In particular, for every possible choice of entropy and cost functions mentioned in the theorem the induced Entropy-Transport cost ET *is the square of a distance on $\mathcal{M}(X)$.*

Proof. 1) With the choice $F = U_p$ and c_p defined above, we obtain

$$H(x_1, r; x_2, t) = \frac{2}{p}\left[\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)\cos(\mathsf{d}(x_1, x_2) \wedge \pi/2))\right].$$

The assertion follows from Theorem 5 applied with $h(d) := \cos(d \wedge \pi/2)$ and the fact that

$$\sqrt{\mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)\cos(d \wedge \pi/2)}$$

is a metric on $\mathfrak{C}(X)$.

2) In this situation we obtain

$$H(x_1,r;x_2,t)=\frac{2}{p}\Big[\mathfrak{M}_1(r,t)-\mathfrak{M}_{1-p}(r,t)\Big(1+(1-p)\frac{\mathsf{d}(x_1,x_2)}{2}\Big)_+^{\frac{p}{p-1}}\Big].$$

Again we want to apply Theorem 5 with

$$h(d):=\Big(1+(1-p)\frac{\mathsf{d}(x_1,x_2)}{2}\Big)_+^{\frac{p}{p-1}}.$$

It is clear that $h(0)=1$ and h is nonnegative and decreasing since $1-p<0$ and $p/(p-1)>0$. It remains to show that

$$\mathfrak{M}_1(r,t)-\mathfrak{M}_0(r,t)\Big(1+(1-p)\frac{\mathsf{d}(x_1,x_2)}{2}\Big)_+^{\frac{p}{p-1}} \tag{120}$$

is the square of a metric on $\mathfrak{C}(X)$. We already know, as a consequence of Lemma 12, that the map

$$d\mapsto\Big(1+(1-p)\frac{d}{2}\Big)_+^{\frac{p}{p-1}}$$

is convex and decreasing with values in $[0,1]$. Since $x\mapsto\arccos(x)$ is a concave function in $[0,1]$, it follows that

$$l_p(d):=\arccos\left[\Big(1+(1-p)\frac{d}{2}\Big)_+^{\frac{p}{p-1}}\right]$$

is concave and $l_p(d)=0$ if and only if $d=0$. Now we conclude by taking advantage of Lemma 4, since (119) is the square of a metric on $\mathfrak{C}(X)$ for every metric d on X and

$$\mathfrak{M}_1(r,t)-\mathfrak{M}_0(r,t)\cos(l_p(\mathsf{d}(x_1,x_2))\wedge\pi/2)=\mathfrak{M}_1(r,t)-\mathfrak{M}_0(r,t)\Big(1+(1-p)\frac{\mathsf{d}(x_1,x_2)}{2}\Big)_+^{\frac{p}{p-1}}.$$

3) We reason as in the proof of point 2). Now, we have to show that the function

$$f_p(d)=\arccos\Big[(1-(p-1)\frac{d^2}{2})_+^{\frac{p}{p-1}}\Big]$$

is concave and $f_p(d)=0$ if and only if $d=0$. The second statement is obvious, for the first one we cannot reason as before so we proceed with explicit computations. We notice that it is enough to prove that the function is concave when $d\in\Big(0,\sqrt{\frac{2}{p-1}}\Big)$. Let us compute the second derivative: we put

$$g_p(d)=\Big(1+(1-p)\frac{d^2}{2}\Big)^{\frac{p}{p-1}},$$

so that

$$\begin{aligned}
f_p(d)&=\arccos(g_p(d)),\\
g_p'(d)&=\frac{-pdg_p(d)}{\Big(1+(1-p)\frac{d^2}{2}\Big)},\\
g_p''(d)&=\frac{p\Big((p+1)\frac{d^2}{2}-1\Big)g_p(d)}{\Big(1+(1-p)\frac{d^2}{2}\Big)^2}.
\end{aligned}$$

Thus

$$f_p''(d) = -\frac{(1-g_p(d)^2)g_p''(d) + g_p(d)g_p'(d)^2}{(1-g_p(d)^2)^{\frac{3}{2}}} = -\frac{p\Big((p+1)\frac{d^2}{2}-1\Big)g_p(d)\Big(1-g_p(d)^2\Big) + p^2d^2g_p(d)^3}{\Big(1-g_p(d)^2\Big)^{\frac{3}{2}}\Big(1+(1-p)\frac{d^2}{2}\Big)^2}$$
$$= -\frac{pg_p(d)\Big[(p+1)\frac{d^2}{2} - 1 + g_p(d)^2\big((p-1)\frac{d^2}{2}+1\big)\Big]}{\Big(1-g_p(d)^2\Big)^{\frac{3}{2}}\Big(1+(1-p)\frac{d^2}{2}\Big)^2} \tag{121}$$

and f_p is concave if and only if

$$(p+1)\frac{d^2}{2} - 1 + \Big(1+(1-p)\frac{d^2}{2}\Big)^{\frac{2p}{p-1}}\big((p-1)\frac{d^2}{2}+1\big) \ge 0 \tag{122}$$

for every $p \in (1,3]$ and $d \in \Big(0, \sqrt{\frac{2}{p-1}}\Big)$.

We put $z := 1-(p-1)\frac{d^2}{2} \in (0,1)$ and (122) follows if we prove that

$$h_p(z) := \frac{p+1}{p-1}(1-z) - 1 + z^{\frac{2p}{p-1}}(2-z) \ge 0 \tag{123}$$

for every $z \in (0,1)$ and every $p \in (1,3]$. We have

$$h_p'(z) = -\frac{(3p-1)z^{\frac{2p}{p-1}} - 4pz^{\frac{p+1}{p-1}} + p + 1}{p-1},$$
$$h_p''(z) = -\frac{2p(3p-1)z^{\frac{p+1}{p-1}} - 4p(p+1)z^{\frac{2}{p-1}}}{(p-1)^2},$$

and it is straightforward to deduce that in the interval $(0,1)$, for every $p \in (1,3]$, the function h_p is convex, so that h_p' is increasing and thus nonpositive since $h_p'(1) = 0$. This implies that h_p is decreasing but $h_p(1) = 0$ so inequality (123) is satisfied.

The last assertion follows directly from Proposition 6. □

We conclude by proving explicit bounds of H_p in terms of H_1 for any $p > 1$.

Proposition 7. *Let H_p and H_1 be the functions defined in* (94) *and* (95)*. Then for any $p > 1$, $x_1, x_2 \in X$ and $r, t \in [0,+\infty)$ it holds*

$$H_p(x_1, r; x_2, t) \le H_1(x_1, r; x_2, t) \le pH_p(x_1, r; x_2, t). \tag{124}$$

Proof. In order to prove the left inequality, we have to show that for any $p > 1$, $c \in [0,+\infty]$ and $r, t \in [0,+\infty)$ it holds

$$\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)\Big(1+(1-p)c\Big)_+^{\frac{p}{p-1}} \le p\mathfrak{M}_1(r,t) - p\mathfrak{M}_0(r,t)e^{-c}. \tag{125}$$

It is sufficient to study the case $c = 0$. Indeed the function

$$f(c) := \mathfrak{M}_{1-p}(r,t)\Big(1+(1-p)c\Big)_+^{\frac{p}{p-1}} - p\mathfrak{M}_0(r,t)e^{-c} \tag{126}$$

is increasing in $[0,+\infty]$. This is clear when $c > \frac{1}{p-1}$, when $c \in [0, \frac{1}{p-1}]$ we compute the derivative of the function so that

$$f'(c) = p\mathfrak{M}_0(r,t)e^{-c} - p\mathfrak{M}_{1-p}(r,t)\Big(1+(1-p)c\Big)^{\frac{1}{p-1}}$$

and this is nonnegative since $\mathfrak{M}_0 \geq \mathfrak{M}_{1-p}(r,t)$ and

$$e^{-c} \geq \Big(1+(1-p)c\Big)^{\frac{1}{p-1}}. \tag{127}$$

Since $p-1>0$, the last inequality easily follows from the well known $e^x \geq 1+x$, $x \in \mathbb{R}$. So, we have to prove that

$$\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t) \leq p\mathfrak{M}_1(r,t) - p\mathfrak{M}_0(r,t),$$

which can be rewritten (when $r \neq t$, otherwise it is obvious) as

$$1-p \leq \frac{\mathfrak{M}_{1-p}(r,t) - \mathfrak{M}_0(r,t)}{\mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)}$$

and the result follows again by the bounds proved in [Kou14].

The right inequality in (124) is easier to obtain since it is equivalent to

$$0 \leq \mathfrak{M}_0(r,t)e^{-c} - \mathfrak{M}_{1-p}(r,t)\Big(1+(1-p)c\Big)_+^{\frac{p}{p-1}},$$

and one can conclude using that $\mathfrak{M}_0(r,t) \geq \mathfrak{M}_{1-p}(r,t)$ and $e^{-c} \geq \Big(1+(1-p)c\Big)_+^{\frac{p}{p-1}}$, which follows from the inequality (127) proved above and the monotonicity of the function x^p when $x \in [0,1]$ and $p>1$. □

Corollary 1. *Every Entropy-Transport metric considered in Theorem 6 is a complete and separable distance on $\mathcal{M}(X)$ inducing the weak topology.*

Proof. In [LMS18, Theorem 7.25] it is proved that, for every metric d on X, the Gaussian Hellinger-Kantorovich distance is a complete and separable metric on the space $\mathcal{M}(X)$ inducing the weak topology. If $c(x_1,x_2) = \mathsf{d}^2(x_1,x_2)$, the induced marginal perspective cost is metrically equivalent to the marginal perspective cost (118) thanks to Proposition 7 and thus the result follows. If $c(x_1,x_2) = \mathsf{d}(x_1,x_2)$, we observe that if (X,d) is a metric space then also $(X,\sqrt{\mathsf{d}})$ is a metric space inducing the same topology thanks to Lemma 4, so we can conclude using the previous point. The same argument can be applied also to the case

$$c_p(x_1,x_2) = \frac{2}{p-1}\left[1 - \big(\cos(\mathsf{d}(x_1,x_2)\wedge\pi/2)\big)^{\frac{p-1}{p}}\right]$$

since the function

$$g(d) := \frac{2}{p-1}\left[1 - \big(\cos(d\wedge\pi/2)\big)^{\frac{p-1}{p}}\right]$$

is convex, strictly increasing (it is sufficient to notice that the cosine function is concave strictly decreasing in $[0,\pi/2]$ and $(p-1)/p \in (0,1)$) and $g(0)=0$, so that its inverse function is concave, strictly increasing and $g^{-1}(0)=0$. □

Acknowledgment. The author is supported by the GNAMPA Project 2019 "Trasporto ottimo per dinamiche con interazione".

The author thanks Prof. Giuseppe Savaré for many valuable suggestions.

References

[AS66] S.M. Ali and S.D. Silvey. "A general class of coefficients of divergence of one distribution from another". In: *J. Roy. Stat. Soc. Ser. B* 28 (1966), pp. 131–142.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry.* Vol. 33. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2001, pp. xiv+415.

[Bul03] P.S. Bullen. *Handbook of Means and Their Inequalities.* 3300 AA Dordrecht, the Netherlands: Kluwer Academic Publishers, 2003.

[Cam86] L. Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer, 1986.

[Chi+18] L. Chizat et al. "Unbalanced optimal transport: dynamic and Kantorovich formulations". In: *J. Funct. Anal.* 274.11 (2018), pp. 3090–3123.

[Csi63] I. Csiszar. "Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten". In: *Magyar. Tud. Akad. Mat. Kutato Int. Kozl* 8 (1963), pp. 85–108.

[Csi67] I. Csiszar. "Information-type measures of difference of probability distributions and indirect observation". In: *Studia Scientiarum Mathematicarum Hungarica* 2 (1967), pp. 229–318.

[ES03] D.M. Endres and J.E. Schindelin. "A new metric for probability distributions". In: *IEEE Transactions on Information Theory* 7 (7 2003).

[Hel09] E. Hellinger. "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen". In: *J. Reine Angew. Math* 136 (1909).

[KL51] S. Kullback and R.A. Leibler. "On Information and Sufficiency". In: *Annals of Mathematical Statistics* 22 (1951).

[Kou14] O. Kouba. "Bounds for the ratios of differences of power means in two arguments". In: *Mathematical Inequalities and Applications* 17.3 (2014).

[KOV91] P. Kafka, F. Osterreicher, and I. Vincze. "On powers of f-divergences defining a distance". In: *Studia Sci. Math. Hungar.* 26 (1991).

[Lin91] J. Lin. "Divergence measures based on the Shannon entropy". In: *IEEE Transactions on Information Theory* 37 (1991), pp. 145–151.

[LMS16] M. Liero, A. Mielke, and G. Savaré. "Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves". In: *SIAM J. Math. Anal.* 48.4 (2016), pp. 2869–2911.

[LMS18] M. Liero, A. Mielke, and G. Savaré. "Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures". In: *Invent. Math.* 211.3 (2018), pp. 969–1117.

[LV06] F. Liese and I. Vajda. "On divergences and informations in statistics and information theory". In: *IEEE Transactions on Information Theory* 52 (2006), pp. 4394–4412.

[LV87] F. Liese and I. Vajda. *Convex Statistical Distances*. Teubner, 1987.

[Mas93] G. Dal Maso. *An introduction to* Γ*-convergence*. Vol. 8. Progress in Nonlinear Differential Equations and their Applications. Birkhäuser Boston, 1993, pp. xiv+340.

[Mat64] K. Matusita. "Distances and decision rules". In: *Annals of the Institute of Statistical Mathematics* 16 (1964), pp. 305–320.

[Ost96] F. Osterreicher. "On a class of perimeter-type distances of probability distributions". In: *Kybernetika* 32 (4 1996), pp. 389–393.

[OV03] F. Osterreicher and I. Vajda. "A new metric divergences on probability spaces and its applicability in statistics". In: *Annals of the Institute of Statistical Mathematics* 55 (3 2003), pp. 639–653.

[Sha48] C. E. Shannon. "A mathematical theory of communication". In: *Bell System Tech. J.* 27 (1948), pp. 379–423, 623–656.

[Vaj89] I. Vajda. *Theory of statistical inference and information*. Springer Netherlands, 1989.

[Vin81] I. Vincze. "On the Concept and Measure of Information Contained in an Observation". In: *Contributions to Probability*. Academic Press, 1981, pp. 207–214.