

A new class of costs for optimal transport planning

J.-J. ALIBERT, G. BOUCHITTÉ and T. CHAMPION

Laboratoire IMATH, Université de Toulon, 83957 La Garde Cedex, France
e-mails: alibert@univ-tln.fr; bouchitte@univ-tln.fr; champion@univ-tln.fr

(Received 22 March 2018; revised 17 October 2018; accepted 23 October 2018; first published online 29 November 2018)

We study a class of optimal transport planning problems where the reference cost involves a non-linear function $G(x, p)$ representing the transport cost between the Dirac measure δ_x and a target probability p . This allows to consider interesting models which favour multi-valued transport maps in contrast with the classical linear case ($G(x, p) = \int c(x, y) dp$) where finding single-valued optimal transport is a key issue. We present an existence result and a general duality principle which apply to many examples. Moreover, under a suitable subadditivity condition, we derive a Kantorovich–Rubinstein version of the dual problem allowing to show existence in some regular cases. We also consider the well studied case of Martingale transport and present some new perspectives for the existence of dual solutions in connection with Γ -convergence theory.

Key words: Optimal transport, Kantorovich–Rubinstein duality, Martingale constraint

2010 Mathematics Subject Classification: 28A33; 49N05; 49M29; 90C46

1 Introduction

In classical optimal transport theory, the primal problem is written in the Monge–Kantorovich form

$$\inf \left\{ \int_{X \times Y} c d\gamma : \gamma \in \Pi(\mu, \nu) \right\}, \quad (1.1)$$

where μ, ν are given probability measures on X and Y , and $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ is a cost function. Here the competitors are probability measures γ on $X \times Y$ with marginals μ and ν , respectively, called transport plans. The particular case where the optimal transport plan $\gamma \in \Pi(\mu, \nu)$ is carried by the graph of a map $T : X \rightarrow Y$ has been extensively studied since the transport map T then solves the original Monge problem. We refer to the books [16, 18, 19] for a detailed presentation of the classical theory.

In this work, we present a different point of view motivated by scenarios where the optimal strategy favours multi-valued transport maps. As a first example, let us describe a very simple toy model where, in a prescribed region Ω of the Euclidean space \mathbb{R}^d (a town), several competing agents (for instance web suppliers) operating in given locations $\{x_1, x_2, \dots, x_N\}$ can reach a prescribed ratio c_i of potential customers. We have then a given discrete measure $\mu = \sum c_i \delta_{x_i}$ and a target probability measure $\nu = f dx$ where $f = f_1 + f_2 + \dots + f_N$ and f_i represents the local

density of customers supplied by x_i ($c_i = \int_{\Omega} f_i dx$ and $\sum_i c_i = 1$). Each agent x_i aims to optimize its own commercial impact by choosing f_i to be spread as much as possible over Ω , namely by maximising the variance $\text{var}(v_i)$ of the probability v_i of density $\frac{f_i}{c_i}$. If the global criterium to be maximised is $\sum_i c_i \text{var}(v_i)$, then, as the variance is a concave function, the optimal choice would be to take $v_i = v$. In fact this trivial strategy is ruled out if we add a classical transport cost involving the distance between the agent and the customer, say $\int_{\Omega} c(x, y) v_i(dy)$, with for instance $c(x, y) = |x - y|^{\alpha}$ with $\alpha \geq 1$. In that case, we are led to a minimisation problem of the kind

$$\inf \left\{ \sum_i c_i G(x_i, v_i) : \sum_i c_i v_i = v \right\} \quad \text{where} \quad G(x, p) := \int_X c(x, y) p(dy) - \text{var}(p).$$

To deal with a more general framework, it is convenient to associate with every transport plan $\gamma \in \Pi(\mu, \nu)$ the family of conditional probabilities γ^x such that

$$\langle \gamma, \varphi \rangle = \int_X \left(\int_Y \varphi(x, y) \gamma^x(dy) \right) \mu(dx),$$

and to incorporate in problem (1.1) a general cost over γ^x as follows;

$$\inf \left\{ \int_X G(x, \gamma^x) d\mu := \int_{X \times Y} c d\gamma + \int_X H(x, \gamma^x) d\mu : \gamma \in \Pi(\mu, \nu) \right\}, \quad (1.2)$$

where the function $G(x, p) : X \times \mathcal{P}(X) \rightarrow \mathbb{R} \cup \{+\infty\}$ is the sum of the usual linear cost $(x, p) \mapsto \int_Y c(x, \cdot) dp$ which appears in (1.1) and a given non-linear function H .

In fact the formulation (1.2) allows to make a connection with another direction in probability theory which received an increasing attention in the recent years, in the context of martingale optimal transport. This corresponds to the case where $H(x, p) = \chi_{[p]=x}$ is the indicator function of the linear constraint $[p] = x$ being $[p]$ the barycentre of p (this is detailed in Example 2.1 below). A classical result due to Strassen [17] states that the infimum in (1.2) is finite under the assumption that μ and ν are in convex order (see Remark 4.6). This problem was introduced in the one-dimensional case $X, Y \subset \mathbb{R}$ in [12] for the special cost $c(x, y) = -|x - y|$ and for more general costs in [4] in the context of mathematical finance, to obtain robust model-independent bounds on option prices: in both works the authors obtain existence for the primal problem (1.2), give a dual problem whose formulation incorporates a Lagrange multiplier for the martingale barycentre constraint, prove there is no duality gap and provide an example for which there is no dual optimal solution (see Example 4.13 below). In [5], under some specific regularity hypotheses on the cost c , the authors use the natural order on the real line to prove that the optimal solutions, so-called left-curtain coupling, have a special monotone structure: this result is obtained through a variational characterisation of the optimal solutions (Lemma 2.1 therein); they also provide a decomposition of the couple (μ, ν) in irreducible components for which they obtain the existence of dual maximisers (see Section 8 therein). Those seminal works have then been extended and precised in several ways. The variational characterisation for left-curtain couplings was used to prove their stability with respect to the marginals (see [13]) and was extended for more general constraints and spaces X, Y (see [3, 20]). Also the precise formulation of the associated dual problem and the existence of dual maximisers have received a particular attention. In [7], the authors propose a quasi-sure formulation for the dual problem for which they prove the existence of a maximiser, and also provide several examples and counter-examples. Still on the real line, [6] provides regularity hypotheses on the cost c which ensure the existence

of point-wise (as opposed to quasi-sure) minimisers for the usual formulation of the martingale dual problem. The d -dimensional case for the cost $c(x, y) = \pm |x - y|$ is addressed in the remarkable paper [10], where the existence of a dual maximiser and the structure of the optimal martingale plans are described under the hypothesis that the measures μ and ν are in subharmonic order. An alternative approach relying on Γ -convergence is proposed in Section 7 of the present work.

The paper is organised as follows. In Section 2 we state our main hypotheses and present the examples we shall consider in the following. In Section 3, we give an existence proof for the primal problem (see Theorem 3.1) **which relies on the lower semicontinuity result presented in Lemma 3.5**. Section 4 is devoted to the dual problem obtained via Fenchel conjugation, and the statement of related optimality conditions: in particular we address the case of entropies depending on the barycentre and propose a possible relaxation in the set of upper semicontinuous functions for the dual problem. The last three sections deal with the existence issue for dual maximisers. In Section 5 we propose a regularisation by penalisation and we extensively study the variance case. In Section 6 we tackle the generalisation of the Kantorovich–Rubinstein sub-additive cost to our setting, and then deduce a general existence result for dual maximisers in the case where the cost G is regular. Finally in Section 7 we provide a new approach for the relaxation of the dual problem in the martingale case that we hope may be fruitful for further work.

Before concluding this introduction, we point out that, while completing the present study, we became aware of the recently published paper [11] (and of its arXiv 2014 version) where the authors consider problem (1.2) in the case of general spaces $X = Y$, with the motivation of obtaining weak versions of Talagrand’s transport–entropy inequality. The examples of costs they consider are costs of Marton type (and its barycentric version) and Samson type that we detail below in Example 2.2. Note that such generalised transport costs have been firstly introduced by Marton in [14]. In [11], a general formulation for the dual problem is given and the absence of duality gap is established as well as a way of recovering Strassen’s result (see Remark 4.6 below). This work is the closest related to the present article, since it gives a first insight on problems of the form (1.2) under general hypotheses. The results obtained therein overlap with those of Sections 3 and 4 of the present paper, although the techniques of proofs slightly differ.

2 Problem setting

In this paper, X and Y are metrisable compact sets. Some of the proofs and results below may hold for general Polish spaces, but we prefer to avoid additional technical difficulties since the main examples we have in mind hold when $X = Y$ is the closure of a bounded open convex subset of \mathbb{R}^d . The measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ are Borel probabilities over X and Y , respectively. In the following $\mathcal{P}(\cdot)$ will be endowed with the weak star topology which, since X and Y are compact, is equivalent to the topology of the tight convergence. As well known this space $\mathcal{P}(\cdot)$ is compact metrisable.

We consider optimal transport problems of the form

$$F(\mu, \nu) = \inf \left\{ \int_{X \times Y} c(x, y) \gamma(dx, dy) + \int_X H(x, \gamma^x) \mu(dx) : \gamma \in \Pi(\mu, \nu) \right\}, \tag{P}$$

where

$$\Pi(\mu, \nu) = \left\{ \gamma \in \mathcal{P}(X \times Y) : \pi_x^* \gamma = \mu, \pi_y^* \gamma = \nu \right\}$$

is the set of transport plans γ from μ to ν (i.e. γ has marginals μ and ν), and $\gamma = \gamma^x \otimes \mu$ is the disintegration of γ with respect to its first marginal μ . This last notation means that

$$\forall f \in \mathcal{C}(X \times Y), \quad \langle \gamma, f \rangle = \int_X \left(\int_Y f(x, y) \gamma^x(dy) \right) \mu(dx).$$

In classical optimal transport theory, $H = 0$ and a particular interest is given to transport plans γ induced by a transport map T , i.e. of the form $\gamma = (id \times T)_\# \mu$. In our context, this can also be written as $\gamma = \delta_{T(x)} \otimes \mu$, i.e. $\gamma^x = \delta_{T(x)}$ for μ almost every x .

We shall make the following assumptions on the costs c and H :

- (A1) the classical cost $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous,
- (A2) the entropy (or perturbation) cost $H : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ satisfies
 - H is lower semicontinuous on $X \times \mathcal{P}(Y)$.
 - for every $x \in X$, $p \mapsto H(x, p)$ is convex.

Here the entropy cost H is meant as a nonlinear perturbation of the classical mass transport cost associated with c . In fact it is also convenient to rewrite (P) by putting c and H in the same global cost defined by

$$G : (x, p) \mapsto G(x, p) = \int_Y c(x, y) p(dy) + H(x, p). \tag{2.1}$$

Then our generalised transport problem reads

$$F(\mu, \nu) = \inf \left\{ \int_X G(x, \gamma^x) \mu(dx) : \gamma \in \Pi(\mu, \nu) \right\}. \tag{2.2}$$

Notice that Assumptions (A1) and (A2) imply that

$$G \text{ is lower semicontinuous in } (x, p) \text{ and convex in } p. \tag{A3}$$

In the following, in order to simplify the presentation, we will prefer using this more concise form, but if needed we will come back to the original formulation with c and H as in the examples below.

Example 2.1 *We shall in particular study the following cases:*

- the Monge cost $c(x, y) = |y - x|$ and the quadratic cost $c(x, y) = |y - x|^2$ where $|\cdot|$ is the Euclidean norm on \mathbb{R}^d ,
- the barycentre constraint case:

$$H(x, p) = H_{bar}(x, p) := \chi_{[p]=x} = \begin{cases} 0 & \text{if } [p] = x \\ +\infty & \text{otherwise,} \end{cases}$$

where $[p] = \int_Y y p(dy)$ denotes the barycentre of p and χ_A is the characteristic function of A . Note that in this case we assume that $X = Y$.

- the variance case, in the case Y is a compact subset of \mathbb{R}^d :

$$H(x, p) = H_{var}(x, p) := -\text{var}(p) = [p]^2 - \int_Y |y|^2 p(dy),$$

where we stress the fact that H_{var} is indeed convex with respect to the variable p since $p \mapsto [p]$ and $p \mapsto \int_Y |\cdot|^2 dp$ are linear in p .

Example 2.2 In [11], in the case $X = Y$, the authors consider the following costs on $X \times \mathcal{P}(X)$

$$G_M(x, p) = \alpha \left(\int_X d(x, y)p(dy) \right) \quad \text{and} \quad G_S(x, p) = \int_X \alpha \left(d(x, y) \frac{dp}{d\rho} \right) \rho(dy)$$

introduced by Marton [14] and Samson [15], respectively, where $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex, d is the distance over X and $\rho \in \mathcal{P}(X)$ is a reference probability. In the case $X = Y = \mathbb{R}^d$, Morton also considered the cost

$$\tilde{G}_M(x, p) = \theta(x - [p])$$

being $\theta : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ a convex function. Section 4.3 is devoted to more general forms of entropies depending on the barycentre.

2.1 Around the barycentre constraint

As noted in [5], when one selects the barycentre constraint $H = H_{bar}$ then the quadratic cost $c(x, y) = |y - x|^2$ turns (P) to an easy problem since the functional to be minimised is constant over the admissible transport plans. Indeed one then has

$$\begin{aligned} \int_{X \times Y} |y - x|^2 \gamma(dx, dy) &= \int_Y |y|^2 \nu(dy) - 2 \int_X [\gamma^x] \cdot x \mu(dx) + \int_X |x|^2 \mu(dx) \\ &= \int_Y |y|^2 \nu(dy) - \int_X |x|^2 \mu(dx) \end{aligned}$$

for any $\gamma \in \Pi(\mu, \nu)$ such that $[\gamma^x] = x$ for μ -a.e. x . Therefore, as far as the barycentre constraint $H = H_{bar}$ is considered, we shall further restrict ourselves to the study of examples involving sub-quadratic costs like the Monge cost $c(x, y) = |y - x|$.

2.2 Around the variance case

Selecting the variance cost $H = H_{var} = -\text{var}$ in problem (P) favours the spreading of the measures γ^x so as to increase their variance $\text{var}(\gamma^x)$. For example if one takes $c = 0$, then the problem (P) amounts to maximising $\int \text{var}(\gamma^x) d\mu$ among transport plans $\gamma \in \Pi(\mu, \nu)$, but then one may compute for any admissible plan γ that

$$\begin{aligned} \int_X \text{var}(\gamma^x) d\mu &= \int_X |[\gamma^x]|^2 d\mu - \int_X \left(\int_Y |y|^2 \gamma^x(dy) \right) d\mu \\ &\leq \left| \int_X [\gamma^x] d\mu \right|^2 - \int_Y |y|^2 \nu(dy) = \text{var}(\nu) \end{aligned}$$

so that the only optimal solution is $\gamma = \mu \times \nu$, i.e. $\gamma^x = \nu$ for a.e. x . Note that the linear part $\int_Y |y|^2 d\gamma^x$ of the variance has no impact on the global cost since it yields to the second moment $\int_Y |y|^2 d\nu$ when integrated with respect to μ . We shall make an extensive study of this case in Section 5.2.

3 Existence of an optimal solution for the primal problem

We first investigate the existence of an optimal solution for (P). Our main result is the following (see also [11, Corollary 9.12]).

Theorem 3.1 *Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be such that $F(\mu, \nu) < +\infty$. Then under Assumptions (A1) and (A2), the problem*

$$F(\mu, \nu) = \inf \left\{ \int_{X \times Y} c(x, y) \gamma(dx, dy) + \int_X H(x, \gamma^x) \mu(dx) : \gamma \in \Pi(\mu, \nu) \right\} \quad (P)$$

admits at least one minimiser.

In the above statement, the regularity hypothesis (Assumption (A1)) on c is rather standard in optimal transport theory. The measurability and lower semicontinuity hypotheses (Assumption (A2)) on H are also quite natural in the context of variational functionals over probability spaces. The condition $F(\mu, \nu) < +\infty$ means that the class of competitors is non-empty: it will be discussed further in Section 4 in particular in the case where $H(x, \cdot)$ takes infinite values. The necessity of the convexity hypothesis (Assumption (A2)) on $H(x, \cdot)$ is illustrated in the next example.

Example 3.2 *Let $X = Y := [-1, 1] \times [0, 1]$ and consider the problem (P) which corresponds to the choice*

$$\begin{aligned} \mu &= \mathcal{H}^1 \llcorner \{0\} \times [0, 1] \quad \text{and} \quad \nu = \frac{1}{2} \mathcal{H}^1 \llcorner \{-1, 1\} \times [0, 1] \\ c(x, y) &:= |y - x|^2 \quad \text{and} \quad H(x, p) = \#(\text{support}(p)) \end{aligned}$$

where $\#(\text{support}(p))$ denotes the cardinal of $\text{support}(p)$ whenever it is finite and $+\infty$ otherwise. We shall denote

$$E(\gamma) := \int_{X \times Y} |y - x|^2 \gamma(dx, dy) + \int_X \#(\text{support}(\gamma^x)) \mu(dx).$$

It is known (see [16]) that the only optimal solution to the classical transport problem

$$m = \inf \left\{ \int_{X \times Y} |y - x|^2 \gamma(dx, dy) : \gamma \in \Pi(\mu, \nu) \right\} \quad (3.1)$$

is the transport plan γ_{opt} for which $\gamma_{opt}^{(0,t)} = \frac{1}{2}(\delta_{(-1,t)} + \delta_{(1,t)})$ for a.e. $t \in [0, 1]$, and that there exists a minimising sequence $(T_n)_n$ of transport maps from μ to ν such that

$$m = \lim_{n \rightarrow +\infty} \int_X |y - x|^2 \gamma_n(dx, dy) \quad \text{with} \quad \gamma_n := \delta_{T_n(x)} \otimes \mu \text{ for all } n.$$

Then one has $E(\gamma_{opt}) = m + 2$, $E(\gamma) > m + 1$ for any admissible $\gamma \neq \gamma_{opt}$ and $\lim_{n \rightarrow +\infty} E(\gamma_n) = m + 1$. As a consequence, $F(\mu, \nu) = m + 1$ and the infimum is not attained.

Example 3.3 *The same argument as in the previous example also yields to a non-existence result when the entropy/perturbation cost $H = \#(\text{support})$ is replaced with the more regular cost $H(x, p) = \text{var}(p)$. In this case, one obtains $F(\mu, \nu) = m$ (defined in (3.1) above) and the infimum is not attained either.*

In view of the definition of the global cost G (see (2.1)), Theorem 3.1 is a consequence of Lemmas 3.4 and 3.5 below, which are stated in some more generality for further use. We denote by $\mathcal{M}(X \times Y)$ the set of bounded non-negative Borel measures on $X \times Y$, and for $\gamma \in \mathcal{M}(X \times Y)$ we still denote by $\gamma = \gamma^x \otimes \mu$ the disintegration of γ with respect to its first marginal μ : note that here μ need not be a probability, but the measures γ^x are indeed probabilities on Y .

In the sequel, for every $x \in X$, we denote by $G^*(x, \cdot)$ the Fenchel conjugate of the convex functional $G(x, \cdot)$ in the duality between $\mathcal{M}(Y)$ and $\mathcal{C}(Y)$, i.e.

$$\forall \psi \in \mathcal{C}(Y), \quad G^*(x, \cdot) : \psi \mapsto \sup \left\{ \int_Y \psi dp - G(x, p) : p \in \mathcal{P}(Y) \right\}.$$

It is convenient to introduce the real number

$$m_G := \min\{G(x, p) : (x, p) \in X \times \mathcal{P}(Y)\}, \tag{3.2}$$

where we note that $m_G > -\infty$ by the lower semicontinuity property of G and the compactness of X and Y . As a preliminary result, we have the following.

Lemma 3.4 *The following properties of G^* hold:*

(i) *Let $\psi \in \mathcal{C}(Y)$ and $\lambda \in \mathbb{R}$. Then for every $x \in X$ it holds*

$$G^*(x, \psi + \lambda) = G^*(x, \psi) + \lambda, \tag{3.3}$$

$$G^*(x, \psi) + m_G \leq \sup_Y \psi. \tag{3.4}$$

(ii) *Let $\psi \in \mathcal{C}(Y)$ and $(x_n)_n$ be a sequence such that $x_n \rightarrow x$ in X , then*

$$\limsup_{n \rightarrow +\infty} G^*(x_n, \psi) \leq G^*(x, \psi).$$

(iii) *For every $x \in X$, $\psi_1 \in \mathcal{C}(Y)$ and $\psi_2 \in \mathcal{C}(Y)$*

$$|G^*(x, \psi_1) - G^*(x, \psi_2)| \leq \sup_Y |\psi_1 - \psi_2|.$$

Proof (i) Since $\int_Y \lambda dp = \lambda$ for every $p \in \mathcal{P}(Y)$, (3.3) follows. Now for every $p \in \mathcal{P}(Y)$ it holds

$$\int_Y \psi dp - G(x, p) \leq \sup_Y \psi - m_G,$$

so (3.4) follows by taking the sup in p .

(ii) From (i), it comes that $\sup_n G^*(x_n, \psi) < +\infty$. Since $\mathcal{P}(Y)$ is weak* compact and G is lower semicontinuous on $X \times \mathcal{P}(Y)$, there exists a sequence (p_n) in $\mathcal{P}(Y)$ such that

$$\forall n, \quad G^*(x_n, \psi) = \int_Y \psi dp_n - G(x_n, p_n).$$

Moreover, there exists $p \in \mathcal{P}(Y)$ and an increasing sequence of integers (n_k) such that

$$\limsup_{n \rightarrow +\infty} G^*(x_n, \psi) = \lim_{k \rightarrow +\infty} G^*(x_{n_k}, \psi) \quad \text{and} \quad p_{n_k} \rightharpoonup p \text{ weak* in } \mathcal{P}(Y).$$

Hence

$$\begin{aligned} \limsup_{n \rightarrow +\infty} G^*(x_n, \psi) &= \limsup_{k \rightarrow +\infty} \left(\int_Y \psi dp_{n_k} - G(x_{n_k}, p_{n_k}) \right) \\ &\leq \lim_{k \rightarrow +\infty} \int_Y \psi dp_{n_k} - \liminf_{k \rightarrow +\infty} G(x_{n_k}, p_{n_k}) \\ &\leq \int_Y \psi dp - G(x, p) \leq G^*(x, \psi). \end{aligned}$$

(iii) Let $p_2 \in \mathcal{P}(Y)$ be such that $G^*(x, \psi_2) + G(x, p_2) = \int_Y \psi_2 dp_2$. Then

$$\begin{aligned} G^*(x, \psi_2) - G^*(x, \psi_1) &\leq \left(\int_Y \psi_2 dp_2 - G(x, p_2) \right) - \left(\int_Y \psi_1 dp_2 - G(x, p_2) \right) \\ &\leq \sup_Y (\psi_2 - \psi_1). \end{aligned} \quad \square$$

Lemma 3.5 *Assume that the sequence $(\gamma_n)_n = (\gamma_n^x \otimes \mu_n)_n$ weakly converges in $\mathcal{M}(X \times Y)$ to $\gamma = \gamma^x \otimes \mu$, then the following holds*

$$\liminf_{n \rightarrow +\infty} \int_X G(x, \gamma_n^x) d\mu_n \geq \int_X G(x, \gamma^x) d\mu.$$

Proof We recall from (3.2) that G is bounded from below, and without loss of generality we shall assume that G is non-negative.

Let us fix a dense sequence $(\psi_k)_{k \geq 0}$ in $\mathcal{C}(Y)$. Then, as $G(x, \cdot)$ is convex weakly lower semicontinuous, one has for every (x, p)

$$G(x, p) = \sup \left\{ \int_Y \psi dp - G^*(x, \psi), \psi \in \mathcal{C}(Y) \right\} = \sup_k \left\{ \int_Y \psi_k dp - G^*(x, \psi_k) \right\} \quad (3.5)$$

where for the last equality we used the Lipschitz property of $G^*(x, \cdot)$ established in Lemma 3.4. Accordingly, we associate with the probability family γ^x the following sequences of Borel functions on X :

$$g_k(x) := \int_Y \psi_k d\gamma^x - G^*(x, \psi_k), \quad \widehat{g}_k(x) = \max\{g_0(x), \dots, g_k(x)\}.$$

It can be checked from (3.2) and (3.4) that each g_k is bounded from below. Moreover, by (3.5), the sequence $(\widehat{g}_k(x))_{k \geq 0}$ converges increasingly to $G(x, \gamma^x)$ for μ -almost all x , and it is also uniformly bounded from below by $\min_X g_0$.

Let now the integer $m \geq 0$ be fixed, we denote by $(B_k)_{k \leq m}$ a Borel partition of X such that $\widehat{g}_m = g_k$ on B_k for all $k \leq m$. For any $k \leq m$ we denote by $(K_{k,p})_{p \geq 0}$ a non-decreasing sequence of compact subsets of B_k such that

$$\lim_{p \rightarrow +\infty} \mu(B_k \setminus K_{k,p}) = 0.$$

For any $p \geq 0$ let also $(\Omega_{k,p})_{k \leq m}$ be a family of disjoint open sets such that $K_{k,p} \subset \Omega_{k,p}$ for all $k \leq m$. Since $G \geq 0$, we have

$$\begin{aligned} \int_X G(x, \gamma_n^x) d\mu_n(x) &\geq \sum_{k=0}^m \int_{\Omega_{k,p}} G(x, \gamma_n^x) d\mu_n(x) \\ &\geq \sum_{k=0}^m \int_{\Omega_{k,p}} \left[\int_Y \psi_k(y) d\gamma_n^x(y) - G^*(x, \psi_k) \right] d\mu_n(x) \\ &= \sum_{k=0}^m \left[\int_{\Omega_{k,p} \times Y} \psi_k(y) d\gamma_n(x, y) + \int_{\Omega_{k,p}} -G^*(x, \psi_k) d\mu_n(x) \right] \end{aligned}$$

for any n, p . By Lemma 3.4 the function $-G^*(\cdot, \psi_k)$ is bounded from below and lower semicontinuous for fixed k , and since the sets $\Omega_{k,p}$ are open we can take the lower limit as $n \rightarrow \infty$ and compute

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \int_X G(x, \gamma_n^x) d\mu_n(x) &\geq \sum_{k=0}^m \left[\int_{\Omega_{k,p} \times Y} \psi_k(y) d\gamma(x, y) + \int_{\Omega_{k,p}} -G^*(x, \psi_k) d\mu(x) \right] \\ &= \sum_{k=0}^m \int_{\Omega_{k,p}} g_k(x) d\mu(x) \\ &= \sum_{k=0}^m \left[\int_{K_{k,p}} \widehat{g}_m(x) d\mu(x) + \int_{\Omega_{k,p} \setminus K_{k,p}} g_k(x) d\mu(x) \right] \\ &\geq \sum_{k=0}^m \left[\int_{K_{k,p}} \widehat{g}_m(x) d\mu(x) + \mu(\Omega_{k,p} \setminus K_{k,p}) \min_X g_k \right]. \end{aligned}$$

For each $k \leq m$ one has

$$\Omega_{k,p} \setminus K_{k,p} = \Omega_{k,p} \setminus \bigcup_{0 \leq l \leq m} K_{l,p} \subset X \setminus \bigcup_{0 \leq l \leq m} K_{l,p} = \bigcup_{0 \leq l \leq m} (B_l \setminus K_{l,p}).$$

We thus get

$$\forall k \leq m, \quad \lim_{p \rightarrow +\infty} \mu(\Omega_{k,p} \setminus K_{k,p}) = 0.$$

As a consequence, passing to the limit in p we obtain

$$\liminf_{n \rightarrow +\infty} \int_X G(x, \gamma_n^x) d\mu_n(x) \geq \sum_{k=0}^m \int_{B_k} \widehat{g}_m(x) d\mu(x) = \int_X \widehat{g}_m(x) d\mu(x),$$

from which the claim follows by the monotone convergence theorem. □

Proof of Theorem 3.1 For fixed probabilities μ, ν , the set of transport plans $\Pi(\mu, \nu)$ is compact for the tight convergence in $\mathcal{P}(X \times Y)$. Then if $(\gamma_n)_n$ is a minimising sequence for (P), we can extract a subsequence which weakly converges in $\mathcal{P}(X \times Y)$ to some $\gamma \in \Pi(\mu, \nu)$, and it results from Lemma 3.5 that γ is an optimal solution of (P). □

Example 3.6 (Variance case with quadratic cost) *We now illustrate the above existence result in the following simple framework: we take $X = Y = [0, 1]$, $\mu = \frac{1}{2}(\delta_0 + \delta_1)$, $\nu = \mathcal{L}^1_{[0,1]}$, for $\lambda \geq 0$ we set $c_\lambda(x, y) = \lambda |y - x|^2$ and take $H(p) = -\text{var}(p)$, so that (P) reads*

$$F_\lambda(\mu, \nu) = \inf \left\{ \lambda \int_X \int_Y |y - x|^2 \gamma^x(dy) d\mu - \int_X \text{var}(\gamma^x) d\mu : \gamma \in \Pi(\mu, \nu) \right\}. \tag{3.6}$$

Since an admissible transport plan γ is such that γ^0 and γ^1 are absolutely continuous with respect to $\mathcal{L}^1_{[0,1]}$ and satisfy the identity $\gamma^0 + \gamma^1 = 2\mathcal{L}^1_{[0,1]}$, one may rewrite the above problem as

$$F_\lambda(\mu, \nu) = \inf \left\{ [\gamma^0]^2 + (\lambda - 1)[\gamma^0] + \frac{1}{6}(1 - \lambda) : 0 \leq \gamma^0 \leq 2\mathcal{L}^1_{[0,1]} \text{ and } \int_0^1 d\gamma^0 = 1 \right\}.$$

For $\lambda \geq \frac{1}{2}$ the solution is uniquely attained by $\gamma^0 = 2\mathcal{L}^1_{[0, \frac{1}{2}]}$ (for which $[\gamma^0] = \frac{1}{4}$) which corresponds to the usual monotone transport plan γ between μ and ν (which spreads the mass at 0 onto $[0, \frac{1}{2}]$ and that at 1 onto $[\frac{1}{2}, 1]$): in this case, the cost c_λ appears to be dominant over the perturbation H . On the other hand, for $\lambda \in [0, \frac{1}{2}[$ any admissible γ^0 such that $[\gamma^0] = \frac{1 - \lambda}{2}$ yields to an optimal solution, and then $F_\lambda(\mu, \nu)$ admits infinitely many minimisers. Notice that the disintegrations γ^0 and γ^1 of such optimal solutions γ may be supported on the whole segment $[0, 1]$ in this case.

4 Duality and optimality conditions

4.1 The duality principle

Here we propose a general framework which rests upon the convexity of the minimal value function $F(\mu, \nu)$. Such a duality principle has been obtained in some particular cases (see [4, 5, 20]) and in a more general setting in [11]. First we extend the definition of F to the set of bounded measures on $X \times Y$ as follows:

$$F(\mu, \nu) = \begin{cases} \inf \left\{ \int_X G(x, \gamma^x) \mu(dx) : \gamma \in \Pi(\mu, \nu) \right\} \\ \text{if } \mu \text{ and } \nu \text{ are non-negative with } \int \mu = \int \nu \\ +\infty \text{ otherwise.} \end{cases}$$

Recall that, for every $x \in X$, $G^*(x, \cdot)$ denotes the Fenchel conjugate of the convex function $G(x, \cdot)$ in the duality between $\mathcal{M}(Y)$ and $\mathcal{C}(Y)$.

Lemma 4.1 *The functional F defined above is positively one homogeneous and convex. Its Fenchel conjugate in the duality between $\mathcal{M}(X) \times \mathcal{M}(Y)$ and $\mathcal{C}(X) \times \mathcal{C}(Y)$ is the indicator function of the following subset K of $\mathcal{C}(X) \times \mathcal{C}(Y)$:*

$$K := \{(\varphi, \psi) \in \mathcal{C}(X) \times \mathcal{C}(Y) : \varphi(x) + G^*(x, \psi) \leq 0 \quad \forall x \in X\}.$$

Proof The homogeneity property is obvious by construction. In order to show the convexity of F , it is enough to establish that the functional $E(\gamma) = E(\gamma^x \otimes \mu) := \int G(x, \gamma^x) \mu(dx)$ is convex as a functional on $\mathcal{P}(X \times Y)$. Let $\gamma_i = \gamma_i^x \otimes \mu_i$ for $i = 1, 2$ two elements of $\mathcal{P}(X \times Y)$. Then, we have

$$\frac{\gamma_1 + \gamma_2}{2} = \left(\frac{d\mu_1}{d(\mu_1 + \mu_2)} \gamma_1^x + \frac{d\mu_2}{d(\mu_1 + \mu_2)} \gamma_2^x \right) \otimes \frac{\mu_1 + \mu_2}{2}$$

Thus $E(\frac{\gamma_1 + \gamma_2}{2}) \leq \frac{E(\gamma_1) + E(\gamma_2)}{2}$ by the convexity of $G(x, \cdot)$. The convexity of E then follows owing to the lower semicontinuity property of E (see Lemma 3.5).

Now the one homogeneity of F implies that the Fenchel conjugate of F coincides with the indicator of the convex subset

$$\tilde{K} := \left\{ (\varphi, \psi) \in \mathcal{C}(X) \times \mathcal{C}(Y) : \int_X \varphi \, d\mu + \int_Y \psi \, d\nu \leq F(\mu, \nu), \forall (\mu, \nu) \in \mathcal{M}(X) \times \mathcal{M}(Y) \right\}.$$

We need only to check that $K = \tilde{K}$. Let (μ, ν) such that $F(\mu, \nu) < +\infty$. Then by using Fenchel inequality, we observe that for every $\psi \in \mathcal{C}(Y)$ and every admissible $\gamma = \gamma^x \otimes \mu$, one has

$$\int_X G(x, \gamma^x) \mu(dx) \geq \int_X \left(\int_Y \psi \, d\gamma^x - G^*(x, \psi) \right) \mu(dx) = \int_Y \psi \, d\nu - \int_X G^*(\cdot, \psi) \, d\mu.$$

By taking the infimum in the left-hand member with respect to all admissible γ^x , we obtain that $(\varphi, \psi) \in \tilde{K}$ whenever $(\varphi, \psi) \in K$. To prove the converse implication, we take μ to be the Dirac mass δ_x at an arbitrary $x \in X$ and ν to be a probability measure on Y such that $G(x, \nu) < +\infty$. Then $(\varphi, \psi) \in \tilde{K}$ implies that:

$$F(\mu, \nu) = G(x, \nu) \geq \varphi(x) + \int_Y \psi \, d\nu.$$

It follows that $G^*(x, \psi) = \sup_{\nu \in \mathcal{P}(Y)} \left\{ \int_Y \psi \, d\nu - G(x, \nu) \right\} \leq -\varphi(x)$, thus $(\varphi, \psi) \in \tilde{K}$. □

We are now in position to introduce the dual problem to (P):

$$\sup \left\{ \int_Y \psi \, d\nu - \int_X G^*(x, \psi) \, d\mu : \psi \in \mathcal{C}(Y) \right\}. \tag{P*}$$

As a direct consequence of previous Lemma 4.1, we obtain the equality $\inf(P) = \sup(P^*)$ and the optimality conditions which characterise an optimal (γ, ψ) . This is an alternative proof of the duality formula in [11, Theorem 9.6] which avoids the measurable selection arguments used therein.

Theorem 4.2 *Under Assumptions (A1) and (A2), we have the following equality*

$$F(\mu, \nu) = \sup \left\{ \int_Y \psi \, d\nu - \int_X G^*(x, \psi) \, d\mu : \psi \in \mathcal{C}(Y) \right\}. \tag{4.1}$$

In particular $F(\mu, \nu)$ is finite if and only if the supremum in the right-hand side above is finite. Furthermore an admissible pair (γ, ψ) is optimal for (P) and (P) if and only if it holds*

$$G(x, \gamma^x) + G^*(x, \psi) = \int_Y \psi \, d\gamma^x \quad \text{for } \mu\text{-almost every } x \in X. \tag{4.2}$$

Note that (4.2) can also be rewritten as

$$\psi \in \partial G(x, \cdot)(\gamma^x) \quad \text{for } \mu\text{-almost every } x \in X,$$

where the subdifferential of $G(x, \cdot)$ is intended in the sense of the duality between $\mathcal{M}(Y)$ and $\mathcal{C}(Y)$.

Remark 4.3 *In the particular case where $H = 0$, one has*

$$-G^*(x, \psi) = \inf \left\{ \int_Y (c(x, \cdot) - \psi) dp : p \in \mathcal{P}(Y) \right\} = \inf \{c(x, y) - \psi(y) : y \in Y\},$$

the latter expression being usually denoted as $\psi^c(x)$ in the literature (e.g. [18, 19]). We then recover the classical dual problem:

$$\sup \left\{ \int_Y \psi dv + \int_X \psi^c d\mu : \psi \in \mathcal{C}(Y) \right\}.$$

Proof of Theorem 4.2 By Lemma 3.5, we can check easily that F is weakly*-lower semicontinuous and proper on $\mathcal{M}(X) \times \mathcal{M}(Y)$. Therefore, since F is also convex, it coincides with its Fenchel biconjugate in the duality between $\mathcal{M}(X) \times \mathcal{M}(Y)$ and $\mathcal{C}(X) \times \mathcal{C}(Y)$. By Lemma 4.1, we infer that it is the support function of the convex set K given therein. That is,

$$F(\mu, \nu) = \sup_{(\varphi, \psi) \in \mathcal{C}(X) \times \mathcal{C}(Y)} \left\{ \int_X \varphi d\mu + \int_Y \psi dv : \varphi(x) + G^*(x, \psi) \leq 0 \text{ in } X \right\}.$$

As, for every $\psi \in \mathcal{C}(Y)$, the function $G^*(x, \psi)$ is bounded upper semicontinuous (see Lemma 3.4), there exists a non-decreasing sequence (φ_n) of admissible functions in $\mathcal{C}(X)$ such that $\sup_n \varphi_n = -G^*(x, \psi)$. Then passing to the limit in the identity above, we are led to (4.1). Furthermore, the optimality of a pair (γ, ψ) for (P) and (4.1) can be checked by testing the equality

$$\begin{aligned} \int_X G(x, \gamma^x) \mu(dx) &= \int_Y \psi dv - \int_X G^*(x, \psi) \mu(dx) \\ &= \int_X \left(\int_Y \psi d\gamma^x \right) \mu(dx) - \int_X G^*(x, \psi) \mu(dx), \end{aligned}$$

which means that the non-negative function $\rho(x) := G(x, \gamma^x) + G^*(x, \psi) - \int_Y \psi d\gamma^x$ vanishes as an element of $L^1_\mu(X)$, thus (4.2). The proof is complete. □

Remark 4.4 *In the recent paper [11], the authors obtain a similar result as Theorem 4.2 in the general case where $X = Y$ is not necessarily bounded, under some mild regularity hypotheses on the cost $(x, p) \mapsto G(x, p)$ (in fact they prove Lemma 5.4 below, which in turn yields Theorem 4.2). As in the above proof, their argument relies on convexity and Fenchel conjugation, as well as on a lower semicontinuity result (Proposition 9.3 therein) which is more restrictive than Lemma 3.5 since it applies to sequences $(\gamma_n)_n$ with fixed first marginal.*

The equality (4.1) allows to derive a necessary and sufficient condition to have $F(\mu, \nu) < +\infty$. To that aim let us introduce the recession function associated with $G^*(x, \cdot)$ (see Theorem 2.5.4 in [2]):

$$(G^*)^\infty(x, \psi) = \lim_{t \rightarrow +\infty} \frac{G^*(x, t\psi)}{t} = \sup \left\{ \int_Y \psi dp : G(x, p) < +\infty \right\}. \tag{4.3}$$

Then $(G^*)^\infty(x, \cdot)$ is convex, lower semicontinuous positively one homogeneous on $\mathcal{C}(Y)$.

Corollary 4.5 A necessary condition to have $F(\mu, \nu) < +\infty$ is that

$$\int_X (G^*)^\infty(x, \psi) \mu(dx) \geq \int_Y \psi \, d\nu \quad \text{for all } \psi \in \mathcal{C}(Y). \tag{4.4}$$

This condition is sufficient if there exists $k \in L^1_\mu(X)$ such that $G(x, p) \leq k(x)$ for every $p \in \text{dom}(G(x, \cdot))$.

Proof Assume that $\alpha := F(\mu, \nu) < +\infty$. Then, by (4.1), for every $\psi \in \mathcal{C}(Y)$ and every $t > 0$, one has

$$\int_X G^*(x, t\psi) \mu(dx) \geq t \int_Y \psi \, d\nu - \alpha.$$

Then, after dividing by t , we may pass to the limit $t \rightarrow +\infty$ taking into account (4.3) and applying monotone convergence theorem. The inequality (4.4) follows. Under the additional assumption on G and by exploiting the second equality in (4.3), we derive that:

$$\begin{aligned} G^*(x, \psi) &= \sup \left\{ \int_Y \psi \, dp - G(x, p) \right\} \geq \sup \left\{ \int_Y \psi \, dp - k(x) : p \in \text{dom } G(x, \cdot) \right\} \\ &= (G^*)^\infty(x, \psi) - k(x). \end{aligned}$$

Therefore, under (4.4), we infer that $F(\mu, \nu) = \sup(P^*) \leq \int_X k \, d\mu$. □

Remark 4.6 In the case where $X = Y$ and H is the martingale constraint, i.e. $H(x, p) = \chi_{[p]=x}$, we derive easily that $(G^*)^\infty(x, \psi) = -(-\psi)^{**}$. Then condition (4.4) amounts to say that for every convex l.s.c. function f ($f = (-\psi)^{**}$), it holds $\int_X f \, d\mu \leq \int_X f \, d\nu$. We recover the convex order condition denoted by $\mu \leq_c \nu$ appearing in Strassen’s theorem [17]: the fact that (4.4) is sufficient for $F(\mu, \nu) < +\infty$ (which amounts here to the existence of a martingale transport plan) was already noted in Section 3 of [11].

4.2 A relaxed version of the dual problem

The existence of a solution ψ in $\mathcal{C}(Y)$ or possibly in a suitable larger class is a difficult issue that we will overcome under some additional regularity assumption on the global cost function $G(x, p)$ (see Sections 5 and 6). Before developing Theorem 4.2 in specific cases, we present now a straightforward extension of dual problem in which we enlarge the class of competitors.

Let us denote by $\mathcal{U}(Y)$ (resp. $\mathcal{U}(X)$) the set of bounded and upper semicontinuous functions on Y (resp. X). Then the map $\psi \in \mathcal{C}(Y) \rightarrow G^*(\cdot, \psi) \in \mathcal{U}(X)$ (see assertion (ii) of Lemma 3.4) can be extended to a map from $\mathcal{U}(Y)$ to $\mathcal{U}(X)$ by setting

$$\forall x \in X, \quad G^*(x, \psi) = \inf \{ G^*(x, \varphi) : \varphi \in \mathcal{C}(Y), \varphi \geq \psi \}.$$

Indeed,

Lemma 4.7 For every $\psi \in \mathcal{U}(Y)$, it holds

$$G^*(x, \psi) = \sup \left\{ \int_X \psi \, dp - G(x, p) : p \in \mathcal{P}(Y) \right\} = \lim_{n \rightarrow +\infty} G^*(x, \psi_n),$$

for any sequence $(\psi_n)_n$ in $\mathcal{C}(Y)$ decreasing to ψ . The map $\psi \in \mathcal{U}(Y) \mapsto G^*(\cdot, \psi) \in \mathcal{U}(X)$ satisfies the properties (i) and (ii) of Lemma 3.4.

Proof Let $x \in X$ and let us show that $G^*(x, \psi) = \lim_{n \rightarrow +\infty} G^*(x, \varphi_n)$ for any sequence $(\varphi_n)_n$ in $\mathcal{C}(Y)$ decreasing to ψ . First we note that $\liminf G^*(x, \varphi_n) \geq G^*(x, \psi)$ since (φ_n) is decreasing to ψ . Then, for $n \geq 1$ consider $p_n \in \mathcal{P}(Y)$ such that

$$G^*(x, \varphi_n) = \int \varphi_n dp_n - G(x, p_n) + \frac{1}{n}.$$

One may assume that $p_n \rightarrow p_\infty$, then for any $n \geq k \geq 1$ we have $G^*(x, \varphi_n) \leq \int \varphi_k dp_n - G(x, p_n) + \frac{1}{n}$, and passing to the limit one gets

$$\limsup_{n \rightarrow +\infty} G^*(x, \varphi_n) \leq \int \varphi_k dp_\infty - G(x, p_\infty).$$

□

Eventually we deduce the following extended version of Theorem 4.2

Proposition 4.8 *Under Assumptions (A1) and (A2), it holds*

$$F(\mu, \nu) = \sup(P^*) = \sup \left\{ \int_Y \psi d\nu - \int_X G^*(x, \psi) \mu(dx) : \psi \in \mathcal{U}(Y) \right\}. \tag{4.5}$$

Moreover the necessary and sufficient optimality condition (4.2) for an optimal pair (γ, ψ) still holds.

Proof We first note that for any $\psi \in \mathcal{U}(Y)$ and $\gamma \in \Pi(\mu, \nu)$ one has

$$\forall x \in X, \quad \int_Y \psi d\gamma^x - G^*(x, \psi) \leq G(x, \gamma^x)$$

and then integrating with respect to μ and taking the infimum in γ it comes to

$$\sup \left\{ \int_Y \psi d\nu - \int_X G^*(x, \psi) \mu(dx) : \psi \in \mathcal{U}(Y) \right\} \leq F(\mu, \nu) = \sup(P^*),$$

where the last equality follows from Theorem 4.2. The reverse inequality is straightforward as well as the validity of (4.2) for characterising an optimal pair (γ, ψ) . □

In the following, we shall still denote by (P^*) the right-hand side of (4.5). Let us remark that this extended version of the dual problem is useful when considering for instance Example 4.13 where, in a precise range of parameter α , a solution exists in $\mathcal{U}([0, 1])$ while no bounded solution exists.

4.3 Case of entropies depending on the barycentre

We are going to particularise Theorem 4.2 in the special case where $X = Y$ is a compact convex subset of \mathbb{R}^d and the convex entropy H involved in the definition of G (see (2.1)) depends only on the barycentre $[p]$ of the probability measure p , typically:

$$H(x, p) = h(x, [p]) \quad \text{where} \quad h = h(x, z) \text{ is l.s.c. in } (x, z) \text{ and convex in } z. \tag{4.6}$$

In this case the optimality conditions characterising an optimal pair (γ, ψ) for the primal–dual problem read as follows:

Theorem 4.9 *Assume $X = Y$ is a compact subset of \mathbb{R}^d and let $\gamma \in \Pi(\mu, \nu)$ and $\psi \in \mathcal{C}(Y)$. Under the assumption (4.6) on H , the pair (γ, ψ) is optimal for (P) and (P^*) if and only if*

$$0 \in \partial(h(x, \cdot) + (c(x, \cdot) - \psi(\cdot))^{**})([\gamma^x]) \quad \text{for } \mu\text{-almost every } x \in X, \tag{4.7}$$

where the biconjugate $(c(x, \cdot) - \psi(\cdot))^{**}$ is the l.s.c. convex envelope for the function $c(x, \cdot) - \psi(\cdot)$ extended to $\mathbb{R}^d \setminus Y$ by $+\infty$.

If moreover $h(x, [\gamma^x]) \in \mathbb{R}$ for μ -almost every $x \in X$ then

$$(c(x, \cdot) - \psi(\cdot))^{**}([\gamma^x]) = \int_Y (c(x, y) - \psi(y))\gamma^x(dy) \quad \mu\text{-a.e.} \tag{4.8}$$

which in particular implies that $c(x, \cdot) - \psi(\cdot)$ is affine on the support of γ^x .

Proof Let $\psi \in \mathcal{C}(Y)$, we compute

$$\begin{aligned} -G^*(x, \psi) &= \inf_{p \in \mathcal{P}(Y)} \left\{ h(x, [p]) + \int_Y (c(x, y) - \psi(y))p(dy) \right\} \\ &= \inf_{z \in \mathbb{R}^d} \left\{ h(x, z) + \inf_{[p]=z} \left\{ \int_Y (c(x, y) - \psi(y))p(dy) \right\} \right\} \\ &= \inf_{z \in \mathbb{R}^d} \left\{ h(x, z) + (c(x, \cdot) - \psi(\cdot))^{**}(z) \right\}. \end{aligned}$$

As a consequence, according to Theorem 4.2, a pair $(\gamma, \psi) \in \Pi(\mu, \nu) \times \mathcal{C}(Y)$ is optimal for (P) and (P^*) if and only for μ -almost every $x \in X$ one has

$$h(x, [\gamma^x]) + \int_Y (c(x, y) - \psi(y))\gamma^x(dy) = \inf_{z \in \mathbb{R}^d} \left\{ h(x, z) + (c(x, \cdot) - \psi(\cdot))^{**}(z) \right\}. \tag{4.9}$$

Taking $z = [\gamma^x]$ in the right-hand term and then using Jensen inequality, we obtain

$$\begin{aligned} h(x, [\gamma^x]) + \int_Y (c(x, y) - \psi(y))\gamma^x(dy) &\leq h(x, [\gamma^x]) + (c(x, \cdot) - \psi(\cdot))^{**}([\gamma^x]) \\ &\leq h(x, [\gamma^x]) + \int_Y (c(x, \cdot) - \psi(\cdot))^{**} d\gamma^x \\ &\leq h(x, [\gamma^x]) + \int_Y (c(x, \cdot) - \psi(\cdot)) d\gamma^x. \end{aligned}$$

Thus the infimum in (4.9) is attained at $[\gamma^x]$ and the proof is thus complete. □

Example 4.10 (Variance case with null cost) *Let us consider the simple example already addressed in Section 2.2: we take the variance cost $H = H_{var} = -\text{var}$ with $c = 0$. In order to apply Theorem 4.9 above, we should in fact take $H(p) = [p]^2 - m_2(\nu)$ where $m_2(\nu)$ is the second-order moment of ν : as noted in Section 2.2, this yields to an equivalent problem. We have already noted that in this case the only optimal solution γ for (P) satisfies $\gamma^x = \nu$ for μ a.e. x . Then by (4.8) an optimal solution ψ of (P^*) should be affine on the support of ν , while (4.7) then implies that $\psi(y) = 2[\nu] \cdot y + \alpha$ on the support of ν for some constant α . Then the solutions of (P^*) are*

the functions of the form $\psi + \phi$ where ϕ is any non-negative continuous function on Y satisfying $\phi = 0$ on the support of ν .

Example 4.11 (Variance case with quadratic cost) *We get back to Example 3.6: as in Example 4.10 we take $H(p) = [p]^2 - m_2(\nu) = [p]^2 - \frac{1}{3}$ instead of $H(p) = -\text{var}(p)$. Then the primal problem (3.6) reads*

$$F_\lambda(\mu, \nu) = \inf \left\{ \int_X \int_Y \lambda |y - x|^2 \gamma^x(dy) \mu(dx) + \int_X (|\gamma^x|^2 - m_2(\nu)) d\mu : \gamma \in \Pi(\mu, \nu) \right\}$$

and setting $G_\lambda(x, p) = \int_Y \lambda |y - x|^2 p(dy) + (|p|^2 - m_2(\nu))$ the dual reads

$$\sup \left\{ \int_Y \psi d\nu - \int_X G_\lambda^*(x, \psi) \mu(dx) : \psi \in C(Y) \right\}. \tag{4.10}$$

Now if the pair (γ, ψ) is optimal for (3.6) and (4.10) then according to (4.8) in Theorem 4.9 the functions $\lambda y^2 - \psi(y)$ and $\lambda(y - 1)^2 - \psi(y)$ should be affine, respectively, on the supports of γ^0 and γ^1 . For $\lambda \geq \frac{1}{2}$ this function should thus be affine on the segments $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$, while for $\lambda \in [0, \frac{1}{2}[$ it should be affine on the whole $[0, 1]$. As a consequence of (4.7) written at the barycentres $[\gamma^0]$ and $[\gamma^1]$, we obtain that the optimal potentials are of the form

$$\psi_\lambda(y) = \lambda y^2 + (1 - \lambda)y + b,$$

whenever $\lambda \in [0, \frac{1}{2}[$, while

$$\psi_\lambda(y) = \begin{cases} \lambda y^2 + \frac{1}{2}y + b & \text{if } y \leq \frac{1}{2} \\ \lambda(y - 1)^2 + \frac{3}{2}y - \frac{1}{2} + b & \text{if } y \geq \frac{1}{2}, \end{cases}$$

if $\lambda \geq \frac{1}{2}$, where $b \in \mathbb{R}$.

Example 4.12 (Monge cost with barycentre constraint) *Let $X = Y := [-1, 1]$ and consider problems (P) and (P*) corresponding to the choice*

$$\begin{aligned} \mu &= \frac{1}{2} dx \llbracket [-1, 1] \text{ and } \nu = \frac{1}{4} \delta_{-1} + \frac{1}{2} \delta_0 + \frac{1}{4} \delta_1, \\ c(x, y) &:= |y - x| \text{ and } H_{bar}(x, p) = h(x, [p]) = \chi_{[p]=x}. \end{aligned}$$

We first solve problem (P). Any $\gamma \in \Pi(\mu, \nu)$ with $\int_X G(x, \gamma^x) \mu(dx) < +\infty$ is such that γ^x must be a convex combination of $\delta_{-1}, \delta_0, \delta_1$ with $[\gamma^x] = x$ for μ -almost every $x \in [-1, 1]$. We are led to write

$$\gamma^x(dy) = \frac{\rho(x) - x}{2} \delta_{-1}(dy) + (1 - \rho(x)) \delta_0(dy) + \frac{\rho(x) + x}{2} \delta_1(dy),$$

where ρ is a non-negative Borel regular function defined μ -almost everywhere and such that $|x| \leq \rho(x) \leq 1$. In addition, the condition $\nu(dy) = \int_X \gamma^x(dy) \mu(dx)$ implies that $\int_{-1}^1 \rho(x) dx = 1$. As a consequence, the unique martingale transport between μ and ν is given as above with $\rho(x) = |x|$.

We now focus on problem (P*). Let us define $\psi \in C[-1, 1]$ by

$$\psi(y) := -(2|y| + \alpha)(1 - |y|) + ay + b.$$

If $\alpha \leq 0, a \in \mathbb{R}, b \in \mathbb{R}$ then a direct computation shows that

$$(| \cdot - x | - \psi(\cdot))^{**}(x) = -\psi(x) \quad \text{and} \quad -\psi(x) = \int_{[-1,1]} (|y - x| - \psi(y)) \gamma^x(dy).$$

According to Theorem 4.9, ψ is optimal for the dual problem (P^*) . Note that the optimal potential ψ is not unique up to an affine function (since one can play with the parameter α).

Example 4.13 (The non-existence example of [4] revisited) Let $X = Y := [0, 1]$ and consider problems (P) and (P^*) corresponding to the choice

$$\begin{aligned} \mu &= \sum_{n=1}^{\infty} |I_n| \delta_{x_n} \quad \text{and} \quad \nu = dy \llcorner [0, 1], \\ c(x, y) &:= -|y - x| \quad \text{and} \quad H(x, p) = h(x, [p]) = \chi_{[p]=x}, \end{aligned}$$

where $\{I_n\}_{n \geq 1}$ is a family of intervals which forms a partition of $[0, 1]$, x_n is the middle point of I_n and it is assumed that

$(x_n)_{n \geq 1}$ is an increasing sequence,

$$\sum_{n=1}^{\infty} |I_n| = 1 \quad \text{and} \quad \sum_{n=1}^{\infty} n |I_n| = +\infty.$$

We first solve problem (P) . Any $\gamma \in \Pi(\mu, \nu)$ with $\int_X G(x, \gamma^x) \mu(dx) < +\infty$ is such that $\sum_{n \geq 1} |I_n| \gamma^{x_n}(dy) = dy \llcorner [0, 1]$ so that $|I_n| \gamma^{x_n}(dy) \leq dy \llcorner [0, 1]$ for all n . We also have

$$\begin{aligned} \frac{|I_1|}{2} = x_1 = [\gamma^{x_1}] &= \int_0^1 y \gamma^{x_1}(dy) = \int_0^1 \gamma^{x_1}([t, 1]) dt \\ &\geq \int_0^{|I_1|} (1 - \gamma^{x_1}([0, t])) dt \geq \int_0^{|I_1|} \left(1 - \frac{t}{|I_1|}\right) dt = \frac{|I_1|}{2}. \end{aligned}$$

Therefore one has $|I_1| \gamma^{x_1}(dy) = dy \llcorner I_1$ and by induction $|I_n| \gamma^{x_n}(dy) = dy \llcorner I_n$ for every $n \geq 1$. As a consequence, the unique martingale transport between μ and ν is such that: for every $n \geq 1$

$$\gamma^{x_n}(dy) = \frac{1_{I_n}(y) dy}{|I_n|}$$

and one has $F(\mu, \nu) = -\frac{1}{4} \sum_{n \geq 1} |I_n|^2$.

We now focus on problem (P^*) . Assume that $\psi \in C[0, 1]$ is optimal for the problem (P^*) . According to Theorem 4.9, for every $n \geq 1$

$$(-| \cdot - x_n | - \psi(\cdot))^{**}(x_n) = \frac{1}{|I_n|} \int_{I_n} (-|y - x_n| - \psi(y)) dy.$$

By Jensen's inequality, this is equivalent to $y \mapsto -|y - x_n| - \psi(y)$ being affine and equal to its convex envelope on I_n . Hence ψ is optimal for (P^*) if and only if for all $y \in [0, 1]$ one has

$$\begin{cases} \psi(y) = \sum_{n \geq 1} (\psi(x_n) + a_n(y - x_n) - |y - x_n|) 1_{I_n}(y) \\ \psi(y) \leq \psi(x_n) + a_n(y - x_n) - |y - x_n| \quad \text{for all } n \geq 1, \end{cases} \tag{4.11}$$

where $a_n |I_n| := \psi(x_n + \frac{|I_n|}{2}) - \psi(x_n - \frac{|I_n|}{2})$. Therefore one has $a_{n+1} + 1 \leq a_n - 1$ which implies $a_n \leq a_1 - 2(n - 1)$. A direct computation gives us

$$\begin{aligned} \psi(x_n) &= \psi(x_n + \frac{|I_n|}{2}) + \frac{1 - a_n}{2} |I_n| \\ &= \psi(0) + \sum_{k=1}^{n-1} a_k |I_k| + \frac{a_n + 1}{2} |I_n| \\ &\leq \psi(0) + \sum_{k=1}^{n-1} (a_1 + 2 - 2k) |I_k| + \frac{1}{2} (a_1 + 3 - 2n) |I_n|. \end{aligned}$$

Passing to the limit as n tends to $+\infty$ we obtain that $\psi(1) - \psi(0) \leq -\infty$. As a consequence, the problem (P^*) has no continuous (thus bounded) solution.

On the other hand, if we assume in addition that

$$|I_n| = \frac{1}{C n^\alpha} \quad \text{with} \quad 1 < \alpha \leq 2 \quad \text{with} \quad C := \sum_{k=1}^{\infty} \frac{1}{k^\alpha} \quad \text{and} \quad a_n = 2(1 - n),$$

then we obtain an upper semicontinuous function ψ_α that satisfies (4.11). We now show that this candidate dual maximiser is in L^1_ν if and only if $\frac{3}{2} < \alpha \leq 2$. Indeed, we compute

$$\begin{aligned} \int_{\bigcup_{k=1}^n I_k} \psi_\alpha(y) \nu(dy) &= \sum_{k=1}^n \int_{I_k} (\psi_\alpha(x_k) + a_k(y - x_k) - |y - x_k|) dy \\ &= \sum_{k=1}^n \left(\psi_\alpha(x_k) |I_k| - \frac{|I_k|^2}{4} \right) \\ &= \sum_{k=1}^n \left(\left(\psi_\alpha(0) + \sum_{i=1}^{k-1} a_i |I_i| + \frac{a_k + 1}{2} |I_k| \right) |I_k| - \frac{|I_k|^2}{4} \right) \\ &= \sum_{k=1}^n \left(\psi_\alpha(0) |I_k| + \left(\frac{5}{4} - k \right) |I_k|^2 + 2 \sum_{i=1}^{k-1} (1 - i) |I_i| |I_k| \right) \\ &= \sum_{k=1}^n \left(\frac{\psi_\alpha(0)}{C} \frac{1}{k^\alpha} - \frac{1}{C^2} \frac{1}{k^{2\alpha-1}} + \frac{5}{4C^2} \frac{1}{k^{2\alpha}} + \frac{2}{C^2} \frac{1}{k^\alpha} \sum_{i=1}^{k-1} \left(\frac{1}{i^\alpha} - \frac{1}{i^{\alpha-1}} \right) \right). \end{aligned}$$

Since ψ_α is bounded from above by $\psi_\alpha(x_1) = \psi_\alpha(0) + \frac{|I_1|}{2}$, we obtain the claim. Also note from the previous computation that any upper semicontinuous function that satisfies (4.11) with values 0 on ∂I_1 is lower than ψ_α , so that when $1 < \alpha \leq \frac{3}{2}$ it does not belong to L^1_ν .

Remark 4.14 In the above example, if the cost function $c(x, y) = -|y - x|$ is replaced by the more usual Monge cost $c(x, y) = |y - x|$, then it can be checked that the dual problem (P^*) admits a solution $\psi \in C[0, 1]$ namely:

$$\psi(y) := \sum_{n=1}^{\infty} \left(|y - x_n| - \frac{|I_n|}{2} \right) 1_{I_n}(y).$$

In fact the unique martingale transport map given in Example 4.13 realises the Monge distance $W_1(\mu, \nu)$ and ψ given above is the associated potential.

5 Existence results for the dual problem

We recall that, in the standard case ($H = 0$), the existence of an optimal solution can be derived under very mild assumptions on the cost c and moreover the optimal pair (ψ, ψ^c) (see Remark 4.3) inherits some regularity from the cost c . In contrast, a general existence result for the dual problem (P^*) in our general framework cannot be expected as illustrated for instance in Example 4.13. In this section, we first show the existence of Lipschitz solutions for a regularised version of (P^*) . Then we focus on the variance case with quadratic cost in which the existence of Lipschitz solutions is provided.

5.1 Existence for a penalised problem

The underlying idea stems from control theory in which a penalisation term is introduced allowing the final state to differ from the target. In this analogy, the penalisation term is taken to be the Monge distance to the target measure ν and the primal problem $F(\mu, \nu)$ is approximated by

$$F_k(\mu, \nu) := \inf \{F(\mu, \tilde{\nu}) + k W_1(\tilde{\nu}, \nu) : \tilde{\nu} \in \mathcal{P}(Y)\} \tag{5.1}$$

for a possibly large scalar $k > 0$. By the lower semicontinuity of the Monge distance with respect to the weak convergence of measures, it can be readily checked that the infimum above is achieved under the standing assumptions on G made in Section 1. Moreover $F_k(\mu, \nu)$ converges increasingly to $F(\mu, \nu)$.

Theorem 5.1 *Assume that Y is a metric space and let $F_k(\mu, \nu)$ be defined by (5.1). Then,*

$$F_k(\mu, \nu) = \max \left\{ \int_Y \psi \, d\nu - \int_X G^*(x, \psi) \, d\mu : \psi \text{ is } k\text{-Lipschitz on } Y \right\}.$$

In particular, if $F(\mu, \nu)$ is finite, then the dual problem (P^) admits a k -Lipschitz solution if and only if F satisfies*

$$F(\mu, \nu) \leq F(\mu, \tilde{\nu}) + k W_1(\tilde{\nu}, \nu) \quad \text{for all } \tilde{\nu} \in \mathcal{P}(Y). \tag{5.2}$$

Remark 5.2 *A straightforward generalisation of Theorem 5.1 can be obtained by substituting the Monge distance with any convex functional $\Phi : \mathcal{P}(Y) \rightarrow \mathbb{R}$ such that $|\Phi(\nu_1) - \Phi(\nu_2)| \leq k W_1(\nu_1, \nu_2)$. In that case, the dual problem involves the Fenchel conjugate Φ^* on $\mathcal{C}(Y)$ whose domain consists of k -Lipschitz functions. Then the following equality holds:*

$$\min \{F(\mu, \tilde{\nu}) + \Phi(\tilde{\nu}) : \tilde{\nu} \in \mathcal{P}(Y)\} = \max \left\{ -\Phi^*(-\psi) - \int_X G^*(x, \psi) \, d\mu : \psi \in \mathcal{C}(Y) \right\},$$

where the right-hand side stands for the dual problem whose solutions exist and are k -Lipschitz.

Remark 5.3 *As will be seen in Section 6, where $X = Y$ and G satisfies a suitable subadditivity condition (see Theorem 6.8), the condition (5.2) is satisfied when $G(x, p)$ is Lipschitz with respect to x . On the other hand, in the specific case considered in Example 4.10 where $G(x, p) = -\text{var}(p)$,*

it can be checked directly that the map $\tilde{v} \mapsto F(\mu, \tilde{v})$ is Lipschitz on $\mathcal{P}(Y)$. Indeed, as $F(\mu, \tilde{v}) = -\text{var}(\tilde{v})$, it holds

$$\begin{aligned} F(\mu, \nu) - F(\mu, \tilde{v}) &= \int_Y y d(\nu - \tilde{v}) \cdot \int_Y y d(\nu + \tilde{v}) - \int_Y |y|^2 d(\nu - \tilde{v}) \\ &= \int_Y y \cdot (\xi - y) d(\nu - \tilde{v}), \end{aligned}$$

where $\xi = \int_Y y d(\nu + \tilde{v})$ is bounded by $2C$ where C is a bound for the compact Y : the function $y \mapsto y \cdot (\xi - y)$ thus has a Lipschitz constant independent of ν and \tilde{v} , so that $F(\mu, \cdot)$ is indeed Lipschitz on $\mathcal{P}(Y)$. The solutions of (P^*) are described in Example 4.10.

The proof of the above result relies on the following result, which is a corollary of Theorem 4.2.

Lemma 5.4 Under Assumptions (A1) and (A2) we have

$$\forall \psi \in \mathcal{C}(Y), \quad F_\mu^*(\psi) = \int_X G^*(x, \psi) \mu(dx), \tag{5.3}$$

where F_μ is defined on $\mathcal{P}(Y)$ by $F_\mu(\nu) := F(\mu, \nu)$.

Proof We infer from Theorem 4.2 that

$$\forall \nu \in \mathcal{P}(Y), \quad F_\mu(\nu) = \sup \left\{ \int_Y \psi d\nu - \int_X G^*(x, \psi) \mu(dx) : \psi \in \mathcal{C}(Y) \right\},$$

so that F_μ is the Fenchel conjugate of the functional $\psi \mapsto \int_X G^*(x, \psi) d\mu$: it follows directly from the definition of $G^*(x, \cdot)$ and from Lemma 3.4(iii) that this functional is convex and l.s.c. on $\mathcal{C}(Y)$, so that it is equal to its biconjugate, which concludes the proof. \square

Proof of Theorem 5.1 We first note that the two extrema in the equality are attained from direct compactness and lower semicontinuity arguments over $\mathcal{C}(Y)$ and $\mathcal{P}(Y)$, respectively. Denoting by χ_{kLip} the characteristic function of the subset of k -Lipschitz functions in $\mathcal{C}(Y)$, we compute

$$\begin{aligned} &\sup \left\{ \int_Y \psi d\nu - \int_X G^*(x, \psi) d\mu : \psi \text{ is } k\text{-Lipschitz on } Y \right\} \\ &= \sup \left\{ \int_Y \psi d\nu - F_\mu^*(\psi) - \chi_{kLip}(\psi) : \psi \in \mathcal{C}(Y) \right\} \\ &= (F_\mu^* + \chi_{kLip})^*(\nu) = (F_\mu \nabla (\chi_{kLip}^*))^{**}(\nu), \end{aligned}$$

where in the last equality the notation $F_\mu \nabla (\chi_{kLip}^*)$ stands for the inf-convolution of F_μ and χ_{kLip}^* over $\mathcal{M}(Y)$, and we have used that χ_{kLip} is convex l.s.c. on $\mathcal{C}(Y)$ so that $\chi_{kLip} = \chi_{kLip}^{**}$. We now compute that for any $\nu, \tilde{\nu}$ in $\mathcal{P}(Y)$ we have

$$\chi_{kLip}^*(\nu - \tilde{\nu}) = \sup \left\{ \int_Y \psi d(\nu - \tilde{\nu}) : \psi \text{ is } k\text{-Lipschitz on } Y \right\} = kW_1(\tilde{\nu}, \nu),$$

where we use a well-known characterisation of the 1-Wasserstein distance (e.g. see [18]). We thus get

$$(F_\mu \nabla(\chi_{kLip}^*)) (v) = \inf \left\{ F(\mu, \tilde{v}) + k W_1(\tilde{v}, v) : \tilde{v} \in \mathcal{P}(Y) \right\}.$$

The right-hand side is convex and continuous (and even k -Lipschitz with respect to W_1) over $\mathcal{P}(Y)$ so it is equal to its Fenchel biconjugate, which concludes the proof. \square

5.2 The variance case with a quadratic cost

Here we assume that $X = Y$ is the closure of a bounded convex subset of \mathbb{R}^d . Let $\lambda > 0$ be a real positive parameter. Then, for every pair $(\mu, \nu) \in (\mathcal{M}(X))^2$, we set

$$F_\lambda(\mu, \nu) := \inf \left\{ \lambda \int_{X^2} |x - y|^2 d\gamma - \int_X \text{var}(\gamma^x) d\mu : \gamma \in \Pi(\mu, \nu) \right\}.$$

As shown in the lemma below, whose proof follows from the discussion in Section 2.2, $F_\lambda(\mu, \nu)$ as a function of ν can be seen as an interpolation between the Wasserstein distance $W_2(\mu, \cdot)$ and the variance.

Lemma 5.5 *It holds $F_\lambda(\mu, \nu) \searrow -\text{var}(\nu)$ as $\lambda \searrow 0$, whereas*

$$W_2^2(\mu, \nu) - \frac{\text{var}(\nu)}{\lambda} \leq \frac{F_\lambda(\mu, \nu)}{\lambda} \leq W_2^2(\mu, \nu).$$

It is convenient to search solutions ψ to the dual problem associated with (\mathcal{P}_λ) in the form $\psi = -\varphi$. After some subtle manipulations which will be detailed in the proof of Theorem 5.6 below, it turns out that we are led to search φ solving

$$\sup \left\{ - \int_X \varphi \nabla \frac{\lambda}{2} |\cdot|^2 d\mu - \int_X \varphi^* d\nu : \varphi \text{ convex continuous} \right\}, \tag{Q_\lambda}$$

where

$$\varphi \nabla \frac{\lambda}{2} |\cdot|^2 (x) := \inf_{z \in X} \left\{ \varphi(z) + \frac{\lambda}{2} |x - z|^2 \right\}$$

denotes the Moreau–Yosida transform of φ (implicitly extended by $+\infty$ for $z \notin X$).

Theorem 5.6 *The supremum of (Q_λ) is achieved and*

$$\frac{F_\lambda(\mu, \nu)}{2\lambda} = \sup(Q_\lambda) + c_\lambda, \quad c_\lambda = \frac{m_2(\mu) + m_2(\nu)}{2} - \frac{1}{2\lambda} m_2(\nu). \tag{5.4}$$

Moreover a pair (γ, φ) is optimal iff it holds for μ -a.e x

$$\text{supp}(\gamma^x) \subset \partial\varphi \left[x - \frac{1}{\lambda} [\gamma^x] \right]. \tag{5.5}$$

Note that, for $\lambda \rightarrow +\infty$, we recover the well-known optimality condition for the W_2 -Wasserstein distance. Indeed if μ is absolutely continuous, then an optimal transport is obtained by taking $T(x) = \nabla\varphi(x)$ (Brenier’s map) which by the convexity of φ is well-defined μ -a.e. For $\lambda = 1$, the

duality formulation (\mathcal{Q}_λ) also appears in [11, Theorem 2.11(3)] for the special cost \overline{T}_2 considered therein (which then amounts to (3.6)).

In order to prove Theorem 5.6, we will need the following technical result.

Lemma 5.7 *Let $\varphi : \mathbb{R}^d \rightarrow (-\infty, +\infty)$ be a convex l.s.c. proper function. Then for every $x \in \mathbb{R}^d$ and $\lambda > 0$, the following identity holds*

$$\frac{\lambda}{2} |x|^2 - \varphi \nabla \frac{\lambda}{2} |\cdot|^2 (x) = \varphi^* \nabla \frac{1}{2\lambda} |\cdot|^2 (\lambda x). \tag{5.6}$$

Proof One has

$$\begin{aligned} \frac{\lambda}{2} |x|^2 - \varphi \nabla \frac{\lambda}{2} |\cdot|^2 (x) &= \sup_{y \in \mathbb{R}^d} \left\{ \frac{\lambda}{2} |x|^2 - \frac{\lambda}{2} |y - x|^2 - \varphi(y) \right\} \\ &= \sup_{y \in \mathbb{R}^d} \left\{ \lambda x \cdot y - \left[\varphi(y) + \frac{\lambda}{2} |y|^2 \right]^2 \right\} \\ &= \left[\varphi + \frac{\lambda}{2} |\cdot|^2 \right]^* (\lambda x). \end{aligned}$$

The equality (5.6) follows by noting that the convex continuous function $\varphi^* \nabla \frac{1}{2\lambda} |\cdot|^2$ admits $\varphi + \frac{\lambda}{2} |\cdot|^2$ as Fenchel conjugate. □

Proof of Theorem 5.6 We compute

$$\begin{aligned} \frac{F_\lambda(\mu, \nu)}{2\lambda} &= \inf_{\gamma \in \Pi(\mu, \nu)} \left\{ \int_{X^2} \frac{|x - y|^2}{2} \gamma(dx dy) + \frac{1}{2\lambda} \left(\int_X |\gamma^x|^2 d\mu - \int_{X^2} |y|^2 \gamma^x(dy) \mu(dx) \right) \right\} \\ &= \frac{m_2(\mu) + m_2(\nu)}{2} - \frac{1}{2\lambda} m_2(\nu) + \inf_{\gamma \in \Pi(\mu, \nu)} \left\{ \int_X \left(-x \cdot [\gamma^x] + \frac{1}{2\lambda} |[\gamma^x]|^2 \right) \mu(dx) \right\} \\ &= c_\lambda + \inf_{\gamma \in \Pi(\mu, \nu)} \left\{ \int_X G_\lambda(x, \gamma^x) \mu(dx) \right\} \\ &= c_\lambda + \sup \left\{ \int_Y \psi d\nu - \int_X G_\lambda^*(x, \psi) \mu(dx) : \psi \in \mathcal{U}(X) \right\}, \end{aligned} \tag{5.7}$$

where in the third line we set $G_\lambda(x, p) := -x \cdot [p] + \frac{1}{2\lambda} |[p]|^2$, for every $p \in \mathcal{P}(X)$ and in the last line, we applied the duality formula (4.5). In this case the Fenchel conjugate G_λ^* can be determined as follows:

$$\begin{aligned} -G_\lambda^*(x, \psi) &= \inf_{p \in \mathcal{P}(X)} \left\{ \int -\psi dp - x \cdot [p] + \frac{1}{2\lambda} |[p]|^2 \right\} \\ &= \inf_{z \in X} \left\{ (-\psi)^{**}(z) + \frac{1}{2\lambda} |\lambda x - z|^2 \right\} - \frac{\lambda}{2} |x|^2 \\ &= \left((-\psi)^{**} \nabla \frac{1}{2\lambda} |\cdot|^2 \right) (\lambda x) - \frac{\lambda}{2} |x|^2 \\ &= - \left((-\psi)^* \nabla \frac{\lambda}{2} |\cdot|^2 \right) (x), \end{aligned} \tag{5.8}$$

where for the last equality we applied the identity (5.6) to the convex l.s.c. function $\varphi = (-\psi)^*$. Noticing that $-\varphi^* = -(-\psi)^{**} \geq \psi$, we infer from (5.7) that

$$\frac{F_\lambda(\mu, \nu)}{2\lambda} \leq c_\lambda + \sup \left\{ - \int_X \varphi \nabla \frac{\lambda}{2} |\cdot|^2 d\mu - \int_X \varphi^* dv : \varphi \text{ convex} \right\}.$$

The converse inequality is straightforward by restricting the supremum in (5.7) to those elements $\psi \in \mathcal{U}(X)$ which are concave functions. So far we have established the duality formula (5.4).

The existence issue can be fixed by very similar arguments to those in [9]. Let $\{\varphi_n\}$ be a sequence of convex continuous functions on X which is maximising for (\mathcal{Q}_λ) . Up to adding a constant we assume that $\inf_X \varphi_n = 0$ so that $\varphi_n^*(0) = 0$ (φ_n is extended by $+\infty$ over $\mathbb{R}^d \setminus X$). Let $R > 0$ such that $X \subset B_R$. Then as $\varphi_n^*(y) = \sup_{x \in X} (x \cdot y - \varphi_n(x))$, it is straightforward that

$$|\varphi_n^*(y_1) - \varphi_n^*(y_2)| \leq R |y_1 - y_2| \quad \forall y_1, y_2 \in \mathbb{R}^d.$$

On the other hand, the non-negative function defined by

$$\varphi_n^R(x) := \inf_{y \in X} \{ \varphi_n(y) + R |y - x| \} \tag{5.9}$$

satisfies the following properties:

$$\varphi_n^R \leq \varphi_n, \quad \varphi_n^R(x) \leq 2R^2, \quad (\varphi_n^R)^* = (\varphi_n)^* \text{ on } B_R, \quad \varphi_n^R(x) \in \text{Lip}_R(X). \tag{5.10}$$

Indeed, the first two inequalities in (5.10) follow from (5.9) by choosing $y = 0$ or such that $\varphi_n(y) = 0$. The third relation holds since $(\varphi_n^R)^* = \varphi_n^* + \chi_{B_R}$ on \mathbb{R}^d whereas the equi-Lipschitz property of φ_n^R is straightforward. It follows from Ascoli's Theorem that both φ_n^R and $(\varphi_n^R)^*$ are relatively compact in $\mathcal{C}(X)$. Thus, possibly after extracting a subsequence, we may assume that $\varphi_n^R \rightarrow \varphi, (\varphi_n^R)^* \rightarrow \varphi^*$ uniformly in X . Here we exploit the classical fact that the two limits are convex conjugate to each other (see for instance [1]). Similarly, as $(\varphi_n^R)^* + \frac{1}{2\lambda} |\cdot|^2 \rightarrow (\varphi)^* + \frac{1}{2\lambda} |\cdot|^2$ uniformly in X , by passing to the conjugates (which are still R -Lipschitz), we deduce that $\varphi_n^R \nabla \frac{\lambda}{2} |\cdot|^2 \rightarrow \varphi \nabla \frac{\lambda}{2} |\cdot|^2$ in $\mathcal{C}(X)$. Therefore, by exploiting (5.10), we can conclude that the convex continuous function φ is optimal. Indeed,

$$\begin{aligned} \int_X \varphi \nabla \frac{\lambda}{2} |\cdot|^2 d\mu + \int_X \varphi dv &= \lim_{n \rightarrow \infty} \left(\int_X \left(\varphi_n^R \nabla \frac{\lambda}{2} |\cdot|^2 \right) d\mu + \int_X (\varphi_n^R)^* dv \right) \\ &\leq \limsup_{n \rightarrow \infty} \left(\int_X \left(\varphi_n \nabla \frac{\lambda}{2} |\cdot|^2 \right) d\mu + \int_X (\varphi_n)^* dv \right) = -\text{sup}(\mathcal{Q}_\lambda). \end{aligned}$$

Now we turn to the optimality condition. Let (γ, φ) be an admissible pair for $(\mathcal{P}_\lambda, \mathcal{Q}_\lambda)$ and let $\psi = -\varphi^*$. Then, in view of (4.2), (γ, φ) is optimal if and only if the inclusion $\psi \in \partial G_\lambda(x, \gamma^x)$ holds for μ -a.e $x \in X$. In view of the definition of G_λ and of (5.8), denoting by $w : X \rightarrow X$ a Borel function such that $\varphi \nabla \frac{\lambda}{2} |\cdot|^2 (x) = \varphi(w(x)) + \frac{\lambda}{2} |x - w(x)|^2$, we have

$$\begin{aligned} G_\lambda(x, \gamma^x) + G_\lambda^*(x, \psi) - \langle \gamma^x, \psi \rangle &= -x \cdot [\gamma^x] + \frac{1}{2\lambda} \|\gamma^x\|^2 + \varphi \nabla \left(\frac{\lambda}{2} |\cdot|^2 \right) (x) + \langle \gamma^x, \varphi \rangle \\ &= \int_X [\varphi(w(x)) + \varphi^*(y) - w(x) \cdot y] \gamma^x(dy) \\ &\quad + \frac{\lambda}{2} \left| x - w(x) - \frac{1}{\lambda} [\gamma^x] \right|^2. \end{aligned}$$

Thus one has $\psi \in \partial G_\lambda(x, \gamma^x)$ if and only if $x - w(x) = \frac{1}{\lambda}[\gamma^x]$ together with the condition that $y \in \partial\varphi(w(x))$ holds γ^x a.e. This is exactly the requirement (5.5). □

Example 5.8 *Let us illustrate Theorem 5.6 in the setting of Example 3.6: if φ_λ is convex continuous on $[0, 1]$ and is an optimal solution for (Q_λ) then it should satisfy $\text{support}(\gamma^x) \subset \partial\varphi_\lambda(x - \frac{1}{\lambda}[\gamma^x])$ for $x = 0$ and $x = 1$, which yields that φ_λ^* is affine with derivative $x - \frac{1}{\lambda}[\gamma^x]$ on $\text{support}(\gamma^x)$ for $x = 0$ and $x = 1$. For $\lambda \in]0, \frac{1}{2}[$ this gives*

$$\varphi_\lambda^*(y) = \frac{\lambda - 1}{2\lambda}y + b$$

while for $\lambda \geq \frac{1}{2}$ one has

$$\varphi_\lambda^*(y) = \begin{cases} -\frac{1}{4\lambda}y + \frac{1}{2} - \frac{1}{4\lambda} + b & \text{if } y \leq \frac{1}{2} \\ (1 - \frac{3}{4\lambda})y + b & \text{if } y \geq \frac{1}{2}. \end{cases}$$

Note that this form is not the same as that found in Example 4.11 where we solve the dual problem (4.10). However we observe that the optimal potential ψ_λ found for the latter one coincides up to an additive constant with $\lambda|\cdot|^2 - 2\lambda\varphi_\lambda^*$.

6 The subadditive case

In this part we shall assume that $X = Y$ and that X is the closure of a bounded convex subset of \mathbb{R}^d . In addition, we will assume that the function $G(x, p)$ satisfies (A3) and

$$G(x, \delta_x) = 0 \quad \text{for every } x \in X. \tag{6.1}$$

This amounts to say that $F(\mu, \nu) = 0$ whenever $\mu = \nu$.

In order to avoid multiple occurrences of minus sign, it is convenient to reformulate the relaxed dual problem appearing in (4.5) with the new unknown $\varphi = -\psi$. Then $G^*(x, \psi) = -\varphi^G(x)$ where we have set

$$\forall x \in X, \quad \varphi^G(x) := \inf \left\{ \int_X \varphi dp + G(x, p) : p \in \mathcal{P}(X) \right\}. \tag{6.2}$$

Accordingly, we can rewrite the dual problem as

$$\sup(\mathcal{P}^*) = \sup \left\{ \int_X \varphi^G d\mu - \int_X \varphi d\nu : \varphi \in \mathcal{S}_b(X) \right\},$$

where $\mathcal{S}_b(X)$ denotes the set of bounded lower semicontinuous functions from X to \mathbb{R} . We will denote by \mathcal{G} the following subclass of invariant functions

$$\mathcal{G} = \{ \varphi \in \mathcal{S}_b(X) : \varphi = \varphi^G \}. \tag{6.3}$$

Lemma 6.1 *For every $\varphi \in \mathcal{S}_b(X)$, the infimum in (6.2) is achieved (for all x) and φ^G belongs to $\mathcal{S}_b(X)$. Moreover $\varphi^G \leq \varphi$ and the invariant set \mathcal{G} is a convex subset of $\mathcal{S}_b(X)$.*

Proof The first statement is straightforward by using the lower semicontinuity of φ and that of G on the compact set of $X \times \mathcal{P}(X)$. The second one follows from (6.1) and of the fact that $(\theta \varphi_1 + (1 - \theta) \varphi_2)^G \geq \theta \varphi_1^G + (1 - \theta) \varphi_2^G$ for every φ_1, φ_2 in $\mathcal{S}_b(X)$ and $\theta \in [0, 1]$. □

We will use the following definition.

Definition 6.2 We say that the cost function G is idempotent if it holds

$$\varphi^{GG} = \varphi^G, \text{ for every } \varphi \in \mathcal{S}_b(X).$$

The next duality result is motivated by the celebrated Kantorovich–Rubinstein formulation for (\mathcal{P}^*) which arises in the case where $H = 0$ and c is subadditive and vanishes on the diagonal (we refer for instance to [18] for more details). Like in the Rubinstein case, we expect that the regularising property of the operator $\varphi \mapsto \varphi^G$ will be a cornerstone to obtain the existence of a minimiser for (\mathcal{P}^*) .

Theorem 6.3 Assume that the transform $\varphi \in \mathcal{S}_b(X) \mapsto \varphi^G \in \mathcal{S}_b(X)$ is idempotent and let \mathcal{G} be defined by (6.3). Then

$$F(\mu, \nu) = \sup(\mathcal{P}^*) = \sup \left\{ \int_X \varphi d\mu - \int_X \varphi d\nu : \varphi \in \mathcal{G} \right\}. \tag{6.4}$$

Proof As \mathcal{G} is a subset of $\mathcal{S}_b(X)$, we clearly have that $\sup(\mathcal{P}^*)$ is larger than the right-hand side of (6.4). On the other hand, as $\varphi^G \leq \varphi$ for all $\varphi \in \mathcal{S}_b(X)$ (see the lemma above), one has

$$\sup(\mathcal{P}^*) \leq \sup \left\{ \int_X \varphi^G d\mu - \int_X \varphi^G d\nu : \varphi \in \mathcal{S}_b(X) \right\} \leq \sup \left\{ \int_X \psi d\mu - \int_X \psi d\nu : \psi \in \mathcal{G} \right\},$$

where in the last inequality, we used that $\{\varphi^G : \varphi \in \mathcal{S}_b(X)\}$ is a subset of \mathcal{G} by the idempotent property. □

The link between the idempotent property of G and the subadditivity of the functional $F(\mu, \nu)$ is precised in the following key result.

Proposition 6.4 Assume that G satisfies (A3) and (6.1). Then the following assertions are equivalent:

- (i) $F(\mu, \nu) \leq F(\mu, p) + F(p, \nu)$, for every $(\mu, \nu, p) \in (\mathcal{P}(X))^3$.
- (ii) For every $(p, \nu) \in (\mathcal{P}(X))^2$ and every p -measurable family $\{\gamma^y\}_{y \in X}$, one has

$$G(x, \nu) \leq G(x, p) + \int_X G(y, \gamma^y) p(dy) \text{ whenever } \nu = \int_X \gamma^y p(dy). \tag{6.5}$$

- (iii) G is idempotent.

Proof (i) \Rightarrow (ii). Apply (i) for $\mu = \delta_x, \nu = \int \gamma^y p(dy)$. Then,

$$G(x, \nu) \leq G(x, p) + F(p, \nu) \leq G(x, p) + \int_X G(y, \gamma^y) p(dy),$$

where, in the last inequality, we use the definition of $F(p, \nu)$ as an infimum.

- (ii) \Rightarrow (iii). As $\varphi^{GG} \leq \varphi^G$, it is enough to check that, for every $(x, p) \in X \times \mathcal{P}(X)$, one has

$$G(x, p) + \int_X \varphi^G(y) p(dy) \geq \varphi^G(x).$$

For every $y \in X$, we chose a minimiser γ^y for $\varphi^G(y)$ (see (6.2)). In fact we can do this in such a way that the map $y \in X \rightarrow \gamma^y$ is p -measurable. Indeed we may use a selection theorem (e.g. Theorem III.30 in [8]) for the multifunction $y \in X \mapsto \{q \in \mathcal{P}(X) : G(y, q) + \int_X \varphi dq \leq \varphi^G(y)\}$ whose graph is a Borel subset of $X \times \mathcal{P}(X)$. Then setting $\nu(dz) = \int_X \gamma^y(dz) p(dy)$, we obtain

$$\begin{aligned} G(x, p) + \int_X \varphi^G(y) p(dy) &= G(x, p) + \int_X G(y, \gamma^y) p(dy) + \int_{X^2} \varphi(z) \gamma^y(dz) p(dy) \\ &\geq G(x, \nu) + \int_X \varphi(z) \nu(dz) \\ &\geq \varphi^G(x), \end{aligned}$$

where in the second line the inequality (ii) is used.

(iii) \Rightarrow (i). For every $\varphi \in \mathcal{G}$, one has

$$\int_X \varphi d\mu - \int_X \varphi d\nu = \left(\int_X \varphi d\mu - \int_X \varphi dp \right) + \left(\int_X \varphi dp - \int_X \varphi d\nu \right) \leq F(\mu, p) + F(p, \nu).$$

Assuming (iii), we can apply Theorem 6.3 so that the supremum of the left-hand member above with respect to $\varphi \in \mathcal{G}$ is equal to $F(\mu, \nu)$. □

Remark 6.5 *By the characterisation given in Proposition 6.4(ii), it follows immediately that if G_1, G_2 are two idempotent functions, so is the sum $G_1 + G_2$. On the other hand, as the condition Proposition 6.4(i) is stable by passing to a supremum, we deduce that $\sup_i G_i$ is idempotent for any family $\{G_i, i \in I\}$ of idempotent functions.*

Remark 6.6 *It follows from Proposition 6.4(i) that $G(x, p) = \int_X c(x, y) p(dy)$ satisfies the idempotent property if and only if the function c is subadditive in the usual sense (i.e. $c(x, z) \leq c(x, y) + c(y, z)$). In that case, the idempotent property can be extended to cost functions of the kind (see Subsection 4.3) $G(x, p) = \int_X c(x, y) p(dy) + h([p] - x)$ provided $h : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex, l.s.c. subadditive function. In particular we may consider $h(z) = \chi_{z=0}$ which corresponds to the martingale constraint. Alternatively the martingale constraint can be penalised by taking $h(z) = \lambda |z|$ where λ is a large positive parameter. A slightly more general case of idempotent entropy $H(x, p)$ is presented in the next Lemma 6.7 showing thus that $G(x, p) = \int_X c(x, y) p(dy) + H(x, p)$ is idempotent as well.*

Lemma 6.7 *Let $\{\psi_k : 1 \leq k \leq N\}$ be a finite subset of $\mathcal{C}(X)$ and let $L : \mathcal{P}(X) \rightarrow \mathbb{R}^N$ be the linear map defined by $L(p) = \left(\int_X \psi_k dp \right)_{1 \leq k \leq N}$. Then the entropy given by $H(x, p) := h(L(p) - L(\delta_x))$ is idempotent for every convex, l.s.c. subadditive function $h : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$.*

Proof We are done if H satisfies the condition (6.5) of Proposition 6.4. Let $\nu, p \in \mathcal{P}(X)$ and $\{\gamma^y\}$ a p -measurable family such that $\nu = \int_X \gamma^y p(dy)$. By the linearity of L we infer that $L(\nu) - L(p) = \int_X (L(\gamma^y) - L(\delta_y)) p(dy)$. Thus by Jensen inequality:

$$h(L(\nu) - L(p)) \leq \int_X h(L(\gamma^y) - L(\delta^y)) p(dy) = \int_X H(y, \gamma^y) p(dy).$$

The required inequality (6.5) then follows thanks to the subadditivity property of h :

$$\begin{aligned} H(x, \nu) &= h(L(\nu) - L(\delta_x)) \leq h(L(p) - L(\delta_x)) + h(L(\nu) - L(p)) \\ &\leq H(x, p) + \int_X H(y, \gamma^y) p(dy). \end{aligned}$$

□

We conclude this section by an existence result for the dual problem which generalises the case considered by Kantorovich and Rubinstein where $G(x, p) = \int_X c(x, y) p(dy)$ being c a continuous metric on X (see [18, Theorem 1.14]). In particular we may apply the next theorem when $G(x, p) = \int_X c(x, y) p(dy) + \lambda |[p] - x|$ where $\lambda > 0$ acts as a Lagrange multiplier for the martingale constraint $[\gamma^x] = x$.

Theorem 6.8 *Let $G(x, p)$ be an idempotent cost such that the family of functions $\mathcal{F}_G := \{G(\cdot, p) : p \in \mathcal{P}(X)\}$ is equi-continuous. Then the dual problem (\mathcal{P}^*) admits a solution in $\mathcal{C}(X)$. If \mathcal{F}_G is equi-Lipschitz, then (\mathcal{P}^*) admits a Lipschitz solution.*

Proof By Theorem 6.3, we are reduced to show that the supremum in the right-hand side of (6.4) is achieved by an element of \mathcal{G} . Let (φ_n) be a maximising sequence in \mathcal{G} . As we can add to every φ_n a constant c_n without changing the value of $\int \varphi_n d\mu - \int \varphi_n d\nu$ and since $(\varphi_n + c_n)^G = \varphi_n^G + c_n = \varphi_n + c_n$, we may assume that $\inf_X \varphi_n = 0$ for every n (note that the latter infimum is actually a minimum since φ_n is l.s.c and X is compact). On the other hand, recalling that $\varphi_n^G(x) = \inf\{\int \varphi_n dp + G(x, p)\}$, it holds for every $(x, y) \in X^2$:

$$|\varphi_n^G(x) - \varphi_n^G(y)| \leq \sup\{G(x, p) - G(y, p) : p \in \mathcal{P}(X)\} := \omega(x, y),$$

where, by the (uniform) equi-continuity assumption, $\omega(x, y) \leq \varepsilon(|x - y|)$ for a suitable continuous function $\varepsilon(t)$ on \mathbb{R}_+ vanishing at $t = 0$. As $\varphi_n = \varphi_n^G$, we deduce that $\{\varphi_n\}$ is an equi-continuous and uniformly bounded family in $\mathcal{C}(X)$. By Arzelà–Ascoli theorem, it follows that a subsequence of (φ_n) (still denoted by the same symbol) converges uniformly to an element $\varphi_\infty \in \mathcal{C}(X)$. Observe that this function φ_∞ is Lipschitz in case the family \mathcal{F}_G is assumed to be equi-Lipschitz. Let us show that φ solves (\mathcal{P}^*) . Since

$$\int \varphi_\infty d\mu - \int \varphi_\infty d\nu = \lim_{n \rightarrow \infty} \left(\int \varphi_n d\mu - \int \varphi_n d\nu \right) = \sup(\mathcal{P}^*),$$

it is enough to check that $\varphi_\infty \in \mathcal{G}$. To that aim, we pass to the limit $n \rightarrow \infty$ in the inequality $\int \varphi_n dp + G(x, p) \geq \varphi_n^G(x) = \varphi_n(x)$. We then obtain the inequality $\int \varphi_\infty dp + G(x, p) \geq \varphi_\infty(x)$ holding for every $(x, p) \in X \times \mathcal{P}(X)$. Thus $\varphi_\infty^G(x) \geq \varphi_\infty(x)$, hence $\varphi_\infty^G = \varphi_\infty$. □

7 Some perspectives for martingale transport

Here we focus on the case where $X = Y$ is a convex compact subset of \mathbb{R}^d and

$$G(x, p) = \int_X c(x, y) p(dy) + \chi_{[p]=x}, \tag{7.1}$$

being $c : X^2 \rightarrow \mathbb{R}$ a continuous function. We denote by $MT(\mu, \nu)$ the non-empty subset of $\Pi(\mu, \nu)$ consisting of martingale transport plans, i.e. such that $[\gamma^x] = x$ μ -a.e. Recall that the

transport cost $F(\mu, \nu) = \inf\{\int_{X^2} c \, d\gamma, \gamma \in MT(\mu, \nu)\}$ is finite iff $\mu \leq_c \nu$ in the sense of convex order (see Remark 4.6). From now on we assume that

$$\mu \leq_c \nu, \quad X = co(supp(\nu)). \tag{7.2}$$

We observe that assuming the second condition is not restrictive since $\mu \leq_c \nu$ implies that $supp(\mu) \subset co(supp(\nu))$ and that any transport plan $\gamma \in MT(\mu, \nu)$ satisfies $supp(\gamma^x) \subset supp(\nu)$ for μ -a.e. $x \in X$.

The existence issue for the dual problem in the case of (7.1) is a major problem. The difficulty arises with the lack of compactness of maximising sequences and even the absence of bounded solutions can be demonstrated for very specific cost functions (see Example 4.13 where in some cases unbounded solutions do exist in L^1_ν). In fact, except in the one-dimensional case (see [5–7]), very few results are known about the existence of continuous solutions and, related to this, we point out the recent contribution [10], where the achievement for the dual problem is proved in \mathbb{R}^d in the case of the cost $c(x, y) = \pm |x - y|$ under the additional assumption:

$$\int \varphi \, d\mu \leq \int \varphi \, d\nu \quad \text{for every subharmonic function } \varphi \text{ on } \mathbb{R}^d. \tag{7.3}$$

The aim of this section is to provide an alternative approach for existence in the case of (7.1) for c being a continuous function. We will make use of the function

$$\omega(x, y) := \sup\{|c(x, z) - c(y, z)| : z \in X\}, \tag{7.4}$$

which is continuous on X^2 and vanishes on the diagonal. Moreover, we will denote shortly by $\hat{\varphi}$ the G transform of a function φ (defined in (6.2)) in the case of G defined by (7.1), i.e.

$$\hat{\varphi}(x) := \inf_{[p]=x} \left\{ \int_X (c(x, y) + \varphi(y)) p(dy) \right\} = (c(x, \cdot) + \varphi)^{**}(x). \tag{7.5}$$

(Note that here the Fenchel conjugates defined on \mathbb{R}^d are computed assuming that $\varphi = +\infty$ in $\mathbb{R}^d \setminus X$). With these notations, we may rewrite the dual problem and the equality $\inf(P) = \sup(P^*)$ of Theorem 4.2 as

$$F(\mu, \nu) = \sup(P^*) = \sup \left\{ \int_X \hat{\varphi} \, d\mu - \int_X \varphi \, d\nu : \varphi \in \mathcal{S}_b(X) \right\}, \tag{7.6}$$

where $\mathcal{S}_b(X)$ is the set of l.s.c. and bounded functions already introduced in Section 6. Moreover, in a similar way as in (4.8), we have the following necessary and sufficient optimality condition for an admissible pair (γ, φ) in $MT(\mu, \nu) \times \mathcal{S}_b(X)$:

$$\hat{\varphi}(x) = \int_Y (c(x, y) + \varphi(y)) \gamma^x(dy) \quad \mu\text{-a.e.} \tag{7.7}$$

The novelty of our approach is that we are going to relax the maximisation problem above using the topology of Γ -convergence. To that aim we need to allow unbounded l.s.c. competitors and accordingly we first enlarge the admissible class $\mathcal{S}_b(X)$ used in (7.6) as follows:

$$\mathcal{S}(X) := \left\{ \varphi : X \mapsto \mathbb{R} \cup \{+\infty\} : \varphi \text{ l.s.c. and } \text{dom } \varphi \neq \emptyset \right\}.$$

We point out that elements of $\mathcal{S}(X)$ reach a finite minimum on the compact set X and that their Fenchel biconjugate φ^{**} are convex proper. On the other hand, we deduce easily from (7.5) that,

for every $\varphi \in \mathcal{S}(X)$, $\hat{\varphi}$ belongs to $\mathcal{S}(X)$ and satisfies

$$\varphi^{**} + m \leq \hat{\varphi} \leq \varphi^{**} + M \quad \text{for every } \varphi \in \mathcal{S}(X), \tag{7.8}$$

where m, M are, respectively, a lower bound and an upper bound for c on X^2 (in particular every element φ in the invariant set \mathcal{G} (see (6.3)) is such that $\varphi - \varphi^{**}$ is bounded).

Let us now recall the definition and the main features of Γ -convergence in our finite dimensional context. For every sequence (φ_n) in $\mathcal{S}(X)$, we define for every $x \in X$

$$\Gamma\text{-}\liminf_{n \rightarrow \infty} \varphi_n(x) := \inf_{x_n \rightarrow x} \left\{ \liminf_{n \rightarrow \infty} \varphi_n(x_n) \right\}, \quad \Gamma\text{-}\limsup_{n \rightarrow \infty} \varphi_n(x) := \inf_{x_n \rightarrow x} \left\{ \limsup_{n \rightarrow \infty} \varphi_n(x_n) \right\}.$$

We say that φ_n Γ -converges to φ (denoted as $\varphi_n \xrightarrow{\Gamma} \varphi$) if $\Gamma\text{-}\liminf_{n \rightarrow \infty} \varphi_n = \Gamma\text{-}\limsup_{n \rightarrow \infty} \varphi_n = \varphi$ holds on X . Notice that in general φ is not proper unless we assume that the infimum of φ_n does not blow up to infinity. If it is the case then $\varphi \in \mathcal{S}(X)$ (a Γ -limit is always l.s.c.) and, as X is compact, it holds $\inf_X \varphi_n \rightarrow \inf_X \varphi$. Another useful property is the following generalisation of Fatou’s Lemma holding for any sequence (φ_n) of non-negative functions in $\mathcal{S}(X)$ and any probability measure $p \in \mathcal{P}(X)$:

$$\varphi_n \geq 0 \quad \text{and} \quad \varphi_n \xrightarrow{\Gamma} \varphi \Rightarrow \liminf_n \int_X \varphi_n \, dp \geq \int_X \varphi \, dp. \tag{7.9}$$

This can be deduced by noting that $\varphi = \lim_n \overline{g_n}$ where $\overline{g_n}$ is the lower semicontinuous envelope of $g_n := \inf_{m \geq n} \varphi_m$.

Eventually we recall that by Kuratowski’s theorem, every sequence (φ_n) admits a Γ -convergent subsequence (we refer for instance to [1] for further details about these notions and the possibility to construct a metrisable topology associated with the Γ -convergence).

The following stability result holds.

Proposition 7.1 *Let (φ_n) be a sequence in $\mathcal{S}(X)$ such that $|\inf_X \varphi_n| \leq C$ and $\varphi_n \xrightarrow{\Gamma} \varphi$. Then,*

- (i) $\varphi \in \mathcal{S}(X)$ and $\varphi_n^{**} \xrightarrow{\Gamma} \varphi^{**}$
- (ii) It holds $\hat{\varphi}_n \xrightarrow{\Gamma} \hat{\varphi}$. In particular if $\varphi_n \in \mathcal{G}$ for every n , then $\varphi = \hat{\varphi}$.

Proof (i) By the upper bound hypothesis, it holds $\sup_n |\varphi_n^*(0)| < +\infty$. On the other hand, if R denotes the diameter of the compact subset X and since $\varphi_n^*(y) := \sup_{x \in X} \{x \cdot y - \varphi_n(x)\}$, we derive that

$$|\varphi_n^*(y) - \varphi_n^*(z)| \leq R, |y - z|, \quad \text{for all } (y, z) \in X^2.$$

Thus the family $\{\varphi_n^*\}$ is relatively compact in $\mathcal{C}(X)$. Let us show that φ^* is the unique cluster point. In fact by the property of Γ -convergence, one has for every $y \in X$

$$-\varphi_n^*(y) = \inf_{x \in X} \{\varphi_n(x) - x \cdot y\} \rightarrow \inf_{x \in X} \{\varphi(x) - x \cdot y\} = -\varphi^*(y),$$

where we used that X is compact and the fact that $\varphi_n - (\cdot|y) \xrightarrow{\Gamma} \varphi - (\cdot|y)$. In particular for $y = 0$, we obtain that $\varphi^*(0) = \lim_n \varphi_n^*(0)$ is finite, thus $\inf \varphi \in \mathbb{R}$ and φ belongs to $\mathcal{S}(X)$. Now it is a consequence of [1, Corollary 3.13, p. 286] that $\varphi_n^* \xrightarrow{\Gamma} \varphi^*$ and $\varphi_n^{**} \xrightarrow{\Gamma} \varphi^{**}$.

Let us now prove (ii). We show first that if (x_n) is a sequence in X converging to x , then $\liminf_n \hat{\varphi}_n(x_n) \geq \hat{\varphi}(x)$. To that aim, it is enough to apply the assertion (i) to the sequence (ψ_n) in $\mathcal{S}(X)$ where $\psi_n := \varphi_n + c(x_n, \cdot)$. Indeed the continuity of c implies that $c(x_n, \cdot) \rightarrow c(x, \cdot)$ uniformly in X so that one deduces easily that $\psi_n \xrightarrow{\Gamma} \psi$ with $\psi := \varphi + c(x, \cdot)$. On the other hand, as $|\inf \psi_n - \inf \varphi_n| \leq M$, the sequence $(\inf \psi_n)$ remains bounded. Recalling (7.5) we conclude that

$$\liminf_n \hat{\varphi}_n(x_n) = \liminf_n \psi_n^{**}(x_n) \geq \psi^{**}(x) = \hat{\varphi}(x).$$

It remains to show that for every $x \in X$, we can find a sequence (x_n) converging to x such that $\limsup_n \hat{\varphi}_n(x_n) \leq \hat{\varphi}(x)$. We may assume that $\hat{\varphi}(x) < +\infty$. In a similar way as above, by applying the assertion (i) to $\phi_n := \varphi_n + c(x, \cdot)$, we obtain that $\phi_n^{**} \xrightarrow{\Gamma} \phi^{**}$ where $\phi := \varphi + c(x, \cdot)$. In particular this implies the existence of a sequence $x_n \rightarrow x$ such that $\phi_n^{**}(x_n) \rightarrow \phi^{**}(x)$. We observe that $\varphi_n + c(x_n, \cdot) \leq \phi_n + \omega(x, x_n)$, thus passing to the biconjugate, we have

$$\hat{\varphi}_n(x_n) = (\varphi_n + c(x_n, \cdot))^{**}(x_n) \leq \phi_n^{**} + \omega(x, x_n).$$

Therefore, it holds $\limsup \hat{\varphi}_n(x_n) \leq \limsup \phi_n^{**}(x_n) = \phi^{**}(x) = \hat{\varphi}(x)$ and the claim follows. □

The regularisation effect induced by the transform $\varphi \rightarrow \hat{\varphi}$ is summarised in the next result.

Lemma 7.2 *Let (φ_n) be a sequence in $\mathcal{S}(X)$ such that $|\inf_X \varphi_n| \leq C$ and $\varphi_n \xrightarrow{\Gamma} \varphi$. Assume that the convex subset $D_\varphi := \{\varphi^{**} < +\infty\}$ has non-empty interior. Then $\hat{\varphi}_n$ is continuous and converges uniformly to $\hat{\varphi}$ on every compact subset $K \subset\subset D_\varphi$. Furthermore if ω given by (7.4) satisfies $\omega(x, y) \leq K|x - y|$, then φ is locally Lipschitz on D_φ .*

Proof By Ascoli–Arzela’s theorem, we are done if we show that, for every x_0 in the interior of D_φ , we can find $r > 0$ such that $\{\hat{\varphi}_n\}$ is equi-continuous (resp. equi-Lipschitz) in the ball $B(x_0, r)$. To show this claim, we consider any simplex $\Sigma = co\{a_0, a_1, \dots, a_d\}$ such that $\Sigma \subset D_\varphi$ and Σ contains x_0 in its interior. Then we choose r so small that $B(x_0, 3r) \subset \Sigma$. By the assertion (i) of Proposition 7.1, we have $\varphi_n^{**} \xrightarrow{\Gamma} \varphi^{**}$, there exists $a_{i,n} \rightarrow a_i$ such that $\limsup_n \varphi_n^{**}(a_{i,n}) \leq \varphi^{**}(a_i) < +\infty$. As $\Sigma_n = co\{a_{0,n}, a_{1,n}, \dots, a_{n,d}\}$ is arbitrary close to Σ for large n , it contains a smaller ball $B(x_0, 2r)$ so that, by the convexity of φ_n^{**} , we have

$$\limsup_{n \rightarrow \infty} \sup_{B(x_0, 2r)} \varphi_n^{**} \leq C := \max\{\varphi^{**}(a_i), 0 \leq i \leq d\}. \tag{7.10}$$

Accordingly $\{\varphi_n^{**}\}$ is equi-Lipschitz on the smaller ball $B(x_0, r)$. Let us now consider for every x , the convex function $f_n^x := (c(x, \cdot) + \varphi_n)^{**}$. For every $x, y \in X$, it holds

$$|f_n^x - f_n^y| \leq \omega(x, y), \quad f_n^x \leq \varphi_n^{**} + M, \quad \hat{\varphi}_n(x) = f_n^x(x).$$

Then we deduce the following inequality:

$$|\hat{\varphi}_n(x) - \hat{\varphi}_n(y)| \leq |f_n^x(x) - f_n^x(y)| + \omega(x, y). \tag{7.11}$$

Furthermore, by the estimate (7.10) and the fact that $|\inf_X \varphi_n| \leq C$, the family of convex functions $\{f_n^x : n \geq 0, x \in B(x_0, 2r)\}$ is uniformly majorised on $B(x_0, 2r)$ (by a constant a little larger than $C + M$). Thus it is also equi-Lipschitz on $B(x_0, r)$. The equi-continuity (resp. equi-Lipschitz) property of the family $\{\hat{\varphi}_n, n \in \mathbb{N}\}$ on $B(x_0, r)$ that we claimed follows then from (7.11). □

Remark 7.3 By a careful reading of the proof, it is possible to show that, even if D_φ has an empty interior, the uniform convergence of $(\hat{\varphi}_n)$ still holds on every compact set $K_R := \{\varphi^{**} \leq R\}$. Indeed the arguments developed there apply on the relative interior of the possibly lower dimensional convex subset D_φ .

We are now in position to define our relaxed version for the dual problem (P^*) . Let us define the functional $E : \mathcal{S}(X) \rightarrow \mathbb{R} \cup \{+\infty\}$ by

$$E(\varphi) = \begin{cases} \int \varphi \, d\nu - \int \hat{\varphi} \, d\mu & \text{if } \varphi \in \mathcal{S}_b(X) \\ +\infty & \text{otherwise.} \end{cases}$$

It can be easily checked that E is a convex lower bounded functional and it holds $\inf\{E(\varphi) : \varphi \in \mathcal{S}(X)\} = -F(\mu, \nu)$. We consider the Γ -lower semicontinuous envelope defined by

$$\bar{E}(\varphi) := \inf \left\{ \liminf_{n \rightarrow \infty} E(\varphi_n) : \varphi_n \xrightarrow{\Gamma} \varphi \right\}$$

and define the relaxed dual problem as:

$$\inf \{ \bar{E}(\varphi) : \varphi \in \mathcal{S}(X) \}. \tag{Q}$$

Thanks to the compactness argument due to Kuratowski, we deduce immediately the following.

Lemma 7.4 *The relaxed problem (Q) admits a (possibly unbounded) solution and we have*

$$-F(\mu, \nu) = \inf \{ E(\varphi) : \varphi \in \mathcal{S}(X) \} = \min \{ \bar{E}(\varphi) : \varphi \in \mathcal{S}(X) \}.$$

Proof Since E thus \bar{E} are invariant by addition of a constant function, we can choose a minimising sequence (φ_n) in $\mathcal{S}(X)$ such that $\inf \varphi_n = -\varphi_n^*(0) = 0$. By Kuratowski compactness theorem (and since X is compact), we have $\varphi_{n_k} \xrightarrow{\Gamma} \bar{\varphi}$ for a suitable subsequence where $\bar{\varphi}$ is an element of $\mathcal{S}(X)$ (such that $\inf \bar{\varphi} = 0$). We infer that $\bar{E}(\bar{\varphi}) = \inf\{E(\varphi) : \varphi \in \mathcal{S}(X)\} = \inf\{\bar{E}(\varphi) : \varphi \in \mathcal{S}(X)\}$. □

In order to make this abstract existence result meaningful, we need now to identify $\bar{E}(\varphi)$ and in particular the subset $\{\varphi \in \mathcal{S}(X) : \bar{E}(\varphi) < +\infty\}$. This central issue is partially achieved only and we present here some recent progress. In the next theorem 7.5, we will show that

$$\bar{E}(\varphi) = \int \varphi \, d\nu - \int \hat{\varphi} \, d\mu \quad \text{for every } \varphi \in \mathcal{S}(X) \cap L^1_{\mu+\nu}, \tag{7.12}$$

thus in particular $\bar{E} = E$ on $\mathcal{S}_b(X)$. Notice that by (7.8), it holds $\hat{\varphi} \in L^1_\mu$ whenever $\varphi \in L^1_\nu$. In order to identify $\bar{E}(\varphi)$ for elements φ which are not in L^1_ν , we set

$$I_\gamma(\varphi) := \int_X \left(\int_X \varphi \, d\gamma^x - \hat{\varphi}(x) \right) \mu(dx),$$

for every $\gamma \in MT(\mu, \nu)$. Notice that by (7.8) and since $[\gamma^x] = x$, we have

$$g_\varphi(x) := \int_X \varphi \, d\gamma^x - \hat{\varphi}(x) \geq -M \quad \mu\text{-a.e.} \tag{7.13}$$

so that $I_\gamma(\varphi) > -\infty$. In view of (7.12), we see that, for $\varphi \in \mathcal{S}(X) \cap L^1_\nu$, it holds $\bar{E}(\varphi) = I_\gamma(\varphi)$ for every $\gamma \in MT(\mu, \nu)$. The issue of the next result is to extend this equality to a suitable subclass of $\mathcal{S}(X)$. We define

$$\Gamma_0(X) = \left\{ \varphi = \sum_{k=0}^{k=\infty} \psi_k, \psi_k \text{ convex continuous on } X \right\}. \tag{7.14}$$

In the one-dimensional case, it is shown in [7, Lemma 4.1] that $\Gamma_0(X)$ coincides with the set of all convex l.s.c. functions on X . We believe that it is still the case when X is a compact convex subset of \mathbb{R}^d . As a consequence of Proposition 7.1 and Lemma 7.2, we have the following.

Theorem 7.5 *Assume (7.2) and let $\varphi \in \mathcal{S}(X)$ be such that $\mu(\{\varphi^{**} = +\infty\}) = 0$. Then it holds*

$$\bar{E}(\varphi) \geq I_\gamma(\varphi), \quad \text{for all } \gamma \in MT(\mu, \nu). \tag{7.15}$$

Assume furthermore that φ is of the form $\varphi = u + f$ with $u \in \mathcal{S}(X) \cap L^1_{\mu+\nu}$ and $f \in \Gamma_0(X)$. Then $\bar{E}(\varphi) = I_\gamma(\varphi)$ (and $I_\gamma(\varphi)$ is independent of the choice of $\gamma \in MT(\mu, \nu)$). In particular, it holds (7.12).

Proof Let $\varphi \in \mathcal{S}(X)$ such that $\mu(\{\varphi^{**} = +\infty\}) = 0$, $\gamma \in MT(\mu, \nu)$ and g_φ as defined in (7.13). We have to prove that $\bar{E}(\varphi) \geq \int_X g_\varphi d\mu$. To that aim we consider an arbitrary sequence (φ_n) in $\mathcal{S}_b(X)$ such that $\varphi_n \xrightarrow{\Gamma} \varphi$ to which we associate the sequence (g_{φ_n}) in L^1_μ where $g_{\varphi_n}(x) := \int_X \varphi_n d\gamma^x - \hat{\varphi}_n(x)$. We are then reduced to show the following claim: $\liminf_n \int_X g_{\varphi_n} d\mu \geq \int_X g_\varphi d\mu$. Without loss of generality, we may assume that φ_n is non-negative. Then by the Γ -convergence of φ_n , we know already (see (7.9)) that $\liminf \int_X \varphi_n d\gamma^x \geq \int_X \varphi d\gamma^x$. On the other hand, by Lemma 7.2 (and remark 7.3), $\hat{\varphi}_n$ is continuous and converges uniformly to $\hat{\varphi}$ on $K_R := \{\varphi^{**} \leq R\}$ for every R . It follows in particular that $\hat{\varphi}_n \rightarrow \hat{\varphi}$ on $\{\varphi^{**} < +\infty\}$ that is μ -a.e. according to our assumption. Thus we deduce that $\liminf_n g_{\varphi_n} \geq g_\varphi$ holds μ -a.e. and the claim follows by applying Fatou’s lemma. Thus we have proved that $\bar{E}(\varphi) \geq \int_X g_\varphi d\mu$.

Let us now show that the reverse inequality holds when $\varphi = u + f$ with $u \in \mathcal{S}(X) \cap L^1_{\mu+\nu}$ and $f \in \Gamma_0(X)$. It is not restrictive to assume that $u \geq 0$ and $\int_X g_\varphi d\mu < +\infty$. We consider a non-decreasing sequence of non-negative (u_n) in $\mathcal{S}_b(X)$ such that $\sup_n u_n = u$. Let (ψ_k) be a sequence associated with f through the definition (7.14) and set $f_n := \sum_{k=0}^{n-1} \psi_k$. Then (f_n) is a non-decreasing sequence of convex functions in $\mathcal{S}_b(X)$ such that $f = \sup f_n$ and $f_{n+1} - f_n = \psi_n$ is convex continuous. Let us define $\varphi_n := u_n + f_n$. Clearly it holds $\varphi_n \xrightarrow{\Gamma} \varphi$ since (φ_n) is monotone non-decreasing. By the assertion (ii) of Proposition 7.1, we infer that $\hat{\varphi}_n \xrightarrow{\Gamma} \hat{\varphi}$. As $(\hat{\varphi}_n)$ is non-decreasing, we deduce that $\hat{\varphi}_n \rightarrow \hat{\varphi}$ point-wise. All in all we have proved that $g_{\varphi_n}(x) \rightarrow g_\varphi(x)$ for μ -almost all $x \in X$. As $\bar{E}(\varphi) \leq \limsup_n \int_X g_{\varphi_n} d\mu$, we are done if we can establish that $g_{\varphi_n} \rightarrow g_\varphi$ in L^1_μ .

By dominated convergence theorem, it is enough that g_{φ_n} (recall $g_{\varphi_n} \geq -M$) satisfies the upper bound $g_{\varphi_n} \leq \beta$ for a suitable function in L^1_μ . In fact, by (7.8), we have $\hat{\varphi}_n \geq (u_n + f_n)^{**} + m \geq f_n + m$ so that

$$g_{\varphi_n}(x) \leq \int_X u_n \gamma^x(dy) + h_n(x) - m \quad \text{where} \quad h_n(x) = \int_X f_n d\gamma^x - f_n(x).$$

Here we use the same trick as in [7], namely that by the convexity of $f_{n+1} - f_n = \psi_n$, it holds

$$h_{n+1}(x) - h_n(x) = \int_X \psi_n d\gamma^x - \psi_n(x) \geq 0.$$

Thus $u_n \leq u$ and $h_n \leq h$ where $h(x) := \int_X f d\gamma^x - f(x)$. We are led to $g_{\varphi_n} \leq \beta$ where $\beta := \int_X u d\gamma^x + h(x)$. Let us show that the non-negative function β belongs to L^1_μ . By (7.8), we have $\hat{\varphi} \leq \varphi^{**} + M \leq \varphi + M$ so that $g_\varphi \geq \beta - u - M$. Thus,

$$\int_X \beta d\mu \leq \int_X g_\varphi d\mu + \int_X u d\mu + M.$$

The conclusion follows since we assumed that g_φ and u belong to L^1_μ . □

Remark 7.6 As $\bar{E}(\varphi)$ coincides with $E(\varphi)$ when φ is bounded, the functional E is lower semi-continuous on $\mathcal{S}_b(X)$ with respect to the Γ -convergence and \bar{E} is actually an extension of E from $\mathcal{S}_b(X)$ to $\mathcal{S}(X)$. Coming back to the original dual problem (P^*) and exploiting the characterisation of \bar{E} on $\mathcal{S}(X) \cap L^1_{\mu+\nu}$, we deduce that

$$\sup(P^*) = \sup \left\{ \int_X \hat{\varphi} d\mu - \int_X \varphi d\nu : \varphi \in \mathcal{S}(X) \cap L^1_{\mu+\nu} \right\}.$$

As a consequence of Theorem 7.5, we obtain two corollaries.

Corollary 7.7 Assume that there exists a maximising sequence (φ_n) for (P^*) such that

$$\left| \inf_X \varphi_n \right| + \|\varphi_n\|_{L^1_{\mu+\nu}} \leq C. \tag{7.16}$$

Then the dual problem admits a solution $\varphi \in L^1_{\mu+\nu}$.

Proof We can assume that $\inf \varphi_n = 0$ and, possibly after extracting a subsequence, that $\varphi_n \xrightarrow{\Gamma} \varphi$ where $\varphi \in \mathcal{S}(X)$. Thus by Lemma 7.4, φ solves the relaxed problem (Q) meaning that: $-\bar{E}(\varphi) = \sup(P^*) = F(\mu, \nu)$. In view of the last assertion of Theorem 7.5 (see Remark 7.6), it remains to show that $\varphi \in L^1_{\mu+\nu}$. As (φ_n) is assumed to be non-negative and bounded in $L^1_{\mu+\nu}$, by applying (7.9), we get

$$\int_X \varphi d(\mu + \nu) \leq \liminf_n \int_X \varphi_n d(\mu + \nu) < +\infty. \tag{7.17} \quad \square$$

Remark 7.8 Unfortunately it was not possible to establish the upper bound estimate (7.16) in the general case. Nevertheless, such an estimate can be obtained under specific additional conditions on the measures μ and ν . For instance if we assume that the Newtonian potentials P_μ, P_ν associated, respectively, with μ, ν are such that $P_\nu - P_\mu$ is non-negative (thus $\mu \leq \nu$ in the sense of subharmonic order, see (7.3)) and remains bounded away from 0 in a neighbourhood of X , then, by adapting the arguments given in the proof of [10, Theorem 6.3], we get that every maximising sequence (φ_n) is uniformly bounded in X . So the condition (7.16) is fulfilled in this case and Corollary 7.7 provides the existence of a solution for dual problem (7.6).

Corollary 7.9 Let $\gamma \in MT(\mu, \nu)$ and let φ be of the form $\varphi = u + f$ with $u \in \mathcal{S}(X) \cap L^1_{\mu+\nu}$ and $f \in \Gamma_0(X)$. Then γ is optimal for (P) and φ is optimal for the relaxed dual problem (Q) if and only if the condition (7.7) is satisfied.

Proof By Lemma 7.4, the pair (γ, φ) is optimal if and only if one has

$$\int_{X^2} c \, d\gamma + \bar{E}(\varphi) = 0.$$

As by Theorem 7.5, it holds $\bar{E}(\varphi) = I_\gamma(\varphi) = \int_X (\int_X \varphi \, d\gamma^x - \hat{\varphi}(x))$, the latter condition can be rewritten as:

$$\int_X \left(\int_X (\varphi + c(x, \cdot)) \, d\gamma^x - \hat{\varphi}(x) \right) \, d\mu = 0,$$

which is equivalent to (7.7) since, by (7.5), the quantity integrated with respect to μ is non-negative. \square

Remark 7.10 Example 4.13 fits very well to illustrate the successive steps we performed to enlarge the class of admissible competitors in the dual problem. We found a solution $\psi_\alpha = -\varphi_\alpha$ where $\varphi_\alpha \in C([0, 1])$ for $\alpha > 2$ and where $\varphi_\alpha \in L^1_\nu \cap \mathcal{S}([0, 1])$ for $\frac{3}{2} < \alpha \leq 2$. For $1 < \alpha \leq \frac{3}{2}$, we have merely that $\varphi_\alpha \in \mathcal{S}([0, 1])$ with $\{\varphi_\alpha < +\infty\} = [0, 1]$ and $\mu(\{1\}) = 0$. By checking the optimality condition (7.7), we then deduce from Corollary 7.9 that φ_α is still a solution of a dual problem, but in its relaxed form (Q) since $\varphi_\alpha \notin L^1_\nu$. Notice that, in that case and adopting the notations of Example 4.13, we have $\varphi_\alpha = u + f$ with u 1-Lipschitz given by $u(y) = \sum_{n \geq 1} (|y - x_n| - \frac{1}{2} |I_n|) 1_{I_n}(y)$ and $f \in \Gamma_0([0, 1])$ given by

$$f(y) = \sum_{n \geq 1} \left(\frac{1}{2} |I_n| - \psi_\alpha(x_n) - a_n(y - x_n) \right) 1_{I_n}(y).$$

We conclude this section with a challenging open issue: what can we say about the domain of $\bar{E}(\varphi)$ and does it exist alternative expressions for representing $\bar{E}(\varphi)$ when for instance $\mu(\{\varphi^{**} = +\infty\}) > 0$?

Acknowledgements

We acknowledge the anonymous referees for a careful reading of the manuscript and for their valuable suggestions.

Conflicts of interest

None.

References

- [1] ATTOUCH, H. (1984) *Variational Convergence for Functions and Operators*. Pitman Applicable Mathematics Series, Advanced Publishing Program, Boston, MA.
- [2] AUSLENDER, A. & TEOULLE, M. (2003) *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. Springer Monographs in Mathematics, Springer-Verlag, New York.

- [3] BEIGLBÖCK, M. & GRIESSLER, C. (2014) An optimality principle with applications in optimal transport and its offspring. arXiv:1404.7054v2.
- [4] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. & PENKNER, F. (2013) Model-independent bounds for option prices: a mass transport approach. *Finance Stoch.* **17**(3), 477–501.
- [5] BEIGLBÖCK, M. & JUILLET, N. (2016) On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44**(1), 42–106.
- [6] BEIGLBÖCK, M., LIM, T. & OBLÓJ, J. (2017). Dual attainment for the martingale transport problem. arXiv:1705.04273v1.
- [7] BEIGLBÖCK, M., NUTZ, M. & TOUZI, N. (2017) Complete duality for martingale optimal transport on the line. *Ann. Probab.* **45**(5), 3038–3074.
- [8] CASTAING, C. & VALADIER, M. (1977) *Convex Analysis and Measurable Multifunctions*. Lecture Notes in Mathematics, Vol. 580. Springer-Verlag, Berlin-New York.
- [9] GANGBO, W. (1994) An elementary proof of the polar factorization of vector-valued functions. *Arch. Ration. Mech. Anal.* **128**(4), 381–399.
- [10] GHOUSSEUB, N., KIM, Y.-H. & LIM, T. (2016). Structure of optimal martingale transport plans in general dimensions. arXiv:1508.01806.
- [11] GOZLAN, N., ROBERTO, C., SAMSON, P.-M. & TETALI, P. (2017) Kantorovich duality for general transport costs and applications. *J. Funct. Anal.* **273**(11), 3327–3405.
- [12] HOBSON, D. & NEUBERGER, A. (2012) Robust bounds for forward start options. *Math. Finance* **22**(1), 31–56.
- [13] JUILLET, N. (2016) Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.* **52**(4), 1823–1843.
- [14] MARTON, K. (1996) A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6**(3), 556–571.
- [15] SAMSON, P.-M. (2007) Infimum-convolution description of concentration properties of product probability measures, with applications. *Ann. Inst. H. Poincaré Probab. Statist.* **43**(3), 321–338.
- [16] SANTAMBROGIO, F. (2015) *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications, Vol. 87, Birkhäuser/Springer, Cham.
- [17] STRASSEN, V. (1965) The existence of probability measures with given marginals. *Ann. Math. Statist.* **36**, 423–439.
- [18] VILLANI, C. (2003) *Topics in Optimal Transportation*. Graduate Studies in Mathematics, Vol. 58, American Mathematical Society, Providence, RI.
- [19] VILLANI, C. (2009) *Optimal Transport*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 338, Springer-Verlag, Berlin. Old and new.
- [20] ZAEV, D. A. (2015) On the Monge-Kantorovich problem with additional linear constraints. *Mat. Zametki* **98**(5), 664–683.