



**HAL**  
open science

## Quantization and martingale couplings

Benjamin Jourdain, Gilles Pagès

► **To cite this version:**

Benjamin Jourdain, Gilles Pagès. Quantization and martingale couplings. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 2022, 19, 10.30757/alea.v19-01 . hal-03083022

**HAL Id: hal-03083022**

**<https://hal.science/hal-03083022>**

Submitted on 18 Dec 2020

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

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

# Quantization and martingale couplings

BENJAMIN JOURDAIN<sup>‡</sup>

GILLES PAGÈS<sup>‡</sup>

## Abstract

Quantization provides a very natural way to preserve the convex order when approximating two ordered probability measures by two finitely supported ones. Indeed, when the convex order dominating original probability measure is compactly supported, it is smaller than any of its dual quantizations while the dominated original measure is greater than any of its stationary (and therefore any of its optimal) quadratic primal quantization. Moreover, the quantization errors then correspond to martingale couplings between each original probability measure and its quantization. This permits to prove that any martingale coupling between the original probability measures can be approximated by a martingale coupling between their quantizations in Wassertein distance with a rate given by the quantization errors but also in the much finer adapted Wassertein distance. As a consequence, while the stability of (Weak) Martingale Optimal Transport problems with respect to the marginal distributions has only been established in dimension 1 so far, their value function computed numerically for the quantized marginals converges in any dimension to the value for the original probability measures as the numbers of quantization points go to  $\infty$ .

**AMS Subject Classification (2010):** 60E15, 65C50, 65D32, 60J22, 60G42.

## Introduction

For  $d \in \mathbb{N}^*$  and  $\mu, \nu$  in the set  $\mathcal{P}(\mathbb{R}^d)$  of probability measures on  $\mathbb{R}^d$ , we say that  $\mu$  is smaller than  $\nu$  in the convex order and denote  $\mu \leq_{cvx} \nu$  if

$$\forall \varphi : \mathbb{R}^d \rightarrow \mathbb{R} \text{ convex}, \int_{\mathbb{R}^d} \varphi(x) \mu(dx) \leq \int_{\mathbb{R}^d} \varphi(y) \nu(dy), \quad (0.1)$$

when the integrals make sense (since any real valued convex function is bounded from below by an affine function  $\int_{\mathbb{R}^d} \varphi(x) \mu(dx)$  makes sense in  $\mathbb{R} \cup \{+\infty\}$  as soon as  $\int_{\mathbb{R}^d} |x| \mu(dx) < +\infty$ ). For  $p \geq 1$ , we denote by  $\mathcal{P}_p(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^p \mu(dx) < +\infty\}$  the Wasserstein space with index  $p$  over  $\mathbb{R}^d$ . When  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ , according to the Strassen theorem [50],  $\mu \leq_{cvx} \nu$  if and only if there exists a martingale coupling between  $\mu$  and  $\nu$  that is a probability measure  $\pi(dx, dy)$  on  $\mathbb{R}^d \times \mathbb{R}^d$  with marginals  $\int_{y \in \mathbb{R}^d} \pi(dx, dy)$  and  $\int_{x \in \mathbb{R}^d} \pi(dx, dy)$  equal to  $\mu(dx)$  and  $\nu(dy)$  respectively such that  $\pi(dx, dy) = \mu(dx) \pi_x(dy)$  for some Markov kernel  $\pi_x(dy)$  with the martingale property:

---

\*Université Paris-Est, Cermics (ENPC), INRIA, F-77455 Marne-la-Vallée, France. E-mail: benjamin.jourdain@enpc.fr

†Laboratoire de Probabilités, Statistique et Modélisation, UMR 8001, Campus Pierre et Marie Curie, Sorbonne Université case 158, 4, pl. Jussieu, F-75252 Paris Cedex 5, France. E-mail: gilles.pages@upmc.fr

‡This research benefited from the support of the “Chaire Risques Financiers”, Fondation du Risque

$\forall x \in \mathbb{R}^d$ ,  $\pi_x \in \mathcal{P}_1(\mathbb{R}^d)$  and  $\int_{\mathbb{R}^d} y \pi_x(dy) = x$ . We denote by  $\mathcal{P}(\mu, \nu)$  the set of probability measures on  $\mathbb{R}^d \times \mathbb{R}^d$  with respective marginals  $\mu$  and  $\nu$  and by  $\mathcal{M}(\mu, \nu)$  the subset of  $\mathcal{P}(\mu, \nu)$  consisting of martingale couplings.

Let  $(\mu, \nu)$  belong to the set  $\mathcal{P}_{\leq} \times \mathcal{P}_p(\mathbb{R}^d)$  of couples of elements of  $\mathcal{P}_p(\mathbb{R}^d)$  with the first one smaller than the second in the convex order. In its simplest form, the Martingale Optimal Transport problem consists in computing

$$V_c(\mu, \nu) = \inf_{\pi \in \mathcal{M}(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \pi(dx, dy)$$

and the optimal martingale couplings achieving this infimum for some measurable cost function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ . When the interest rate is zero, for an exotic option written on  $d$  traded assets with payoff given by the function  $c$  of their prices at times  $s$  and  $t$ , then  $V_c(\mu, \nu)$  (resp.  $-V_{-c}(\mu, \nu)$ ) provides a robust lower (resp. upper) price bound when  $\mu$  and  $\nu$  are the respective joint laws of these  $d$  assets at times  $s$  and  $t$  (for instance obtained by calibration of a model to vanilla option prices). Since its introduction in [9], this MOT problem has received recently a great attention in the financial mathematics literature. In particular, the structure of martingale optimal transport couplings [11, 13, 15, 22, 28], continuous time formulations [16, 21, 27], links with the Skorokhod embedding problem [8], numerical methods [2, 1, 14, 25, 26] and stability properties [6, 33, 52] have been investigated. The MOT problem is a particular instance where the measurable cost function  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  is linear in the measure component ( $C(x, \eta) = \int_{\mathbb{R}^d} c(x, y) \eta(dy)$ ) of the Weak Martingale Optimal Transport problem

$$\inf_{\pi \in \mathcal{M}(\mu, \nu)} \int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx)$$

introduced in [6] by adding the martingale constraint to the Weak Optimal Transport problem introduced in [23].

To devise a numerical procedure devoted to the computation of the value and of the optimal couplings in the MOT and WMOT problems, a first natural step consists in approximating  $\mu$  and  $\nu$  by finitely supported probability measures which are still in the convex order. To our best knowledge, few studies consider the problem of preserving the convex order while approximating a sequence of probability measures. We mention the thesis of Baker [7] who proposes the following construction in dimension  $d = 1$ . Let for  $\eta \in \mathcal{P}(\mathbb{R})$  and  $u \in (0, 1)$ ,  $F_{\eta}^{-1}(u) = \inf\{x \in \mathbb{R} : \eta((-\infty, x]) \geq u\}$  be the quantile of  $\eta$  of order  $u$ . For  $(\mu, \nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R})$  and  $N, K \in \mathbb{N}^*$  with  $N/K \in \mathbb{N}^*$ , one has

$$\frac{1}{N} \sum_{i=1}^N \delta_{N \int_{\frac{i-1}{N}}^{\frac{i}{N}} F_{\mu}^{-1}(u) du} \leq_{cvx} \frac{1}{K} \sum_{i=1}^K \delta_{K \int_{\frac{i-1}{K}}^{\frac{i}{K}} F_{\nu}^{-1}(u) du}.$$

*Dual (or Delaunay) quantization* introduced by Pagès and Wilbertz [43] and further studied in [44, 45, 46] gives another way to preserve the convex order in dimension  $d = 1$  when using the same grid to quantize both probability measures.

In two recent papers [1, 2], Alfonsi, Corbetta and Jourdain propose to restore for  $(\mu, \nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$  the convex ordering from any finitely supported approximations  $\tilde{\mu}$  and  $\tilde{\nu}$  of  $\mu$  and  $\nu$ . In dimension  $d = 1$ , one may define the increasing (resp. decreasing) convex order by adding

the constraint that the test function  $\varphi$  is non-decreasing (resp. non-increasing) in (0.1). According to [1], the convex order restauration can be achieved by keeping  $\tilde{\mu}$  (resp.  $\tilde{\nu}$ ) and replacing  $\tilde{\nu}$  (resp.  $\tilde{\mu}$ ) by the supremum (resp. infimum) between  $\tilde{\mu}$  and  $\tilde{\nu}$  for the increasing convex order when  $\int_{\mathbb{R}} x\tilde{\nu}(dx) \leq \int_{\mathbb{R}} x\tilde{\mu}(dx)$  and the decreasing convex order when  $\int_{\mathbb{R}} x\tilde{\nu}(dx) \geq \int_{\mathbb{R}} x\tilde{\mu}(dx)$ . The convex, increasing convex and decreasing convex orders are nicely characterized in terms of the potential function obtained as the anti-derivative of the cumulative distribution function (or of the quantile function). The supremum and infimum of two probability measures for one of these orders can be computed using their potential functions. For a general dimension  $d$ , [2] suggests to keep  $\tilde{\nu}$  and replace  $\tilde{\mu}$  by its projection on the set of probability measures dominated by  $\tilde{\nu}$  for the quadratic Wasserstein distance  $W_2$  (see (0.2) below for the definition of this distance). This projection can be computed by solving a quadratic optimization problem with linear constraints.

In the present paper, when  $\nu$  is compactly supported, we investigate the combined approximation of  $\mu$  by some quadratic-optimal primal quantization and of  $\nu$  by some dual quantization. By construction, any quadratic-optimal primal quantization of  $\mu$  satisfies a stationarity property which implies that it is smaller than  $\mu$  in the convex order. On the other hand, any dual quantization of  $\nu$  is greater than this probability measure in the convex order. Therefore the convex order between  $\mu$  and  $\nu$  is preserved by this combined approximation. Notice that, in contrast with the previous approaches, it cannot be generalized to the convex order preserving approximation of more than two probability measures. Moreover the dual quantization approximation is only possible for compactly supported probability measures. In contrast with these restrictions, we will see that the studied approach proves to provide robust approximations of (Weak) Martingale Optimal Transport problems even in dimension  $d \geq 2$ .

The first section of the paper is devoted to primal (or Voronoi) quantization. For  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$  with  $p \geq 1$ , we show that an element of the set  $\mathcal{P}(\mathbb{R}^d, N)$  of probability measures on  $\mathbb{R}^d$  whose support contains at most  $N$  points is an  $L_p$ -optimal  $N$ -quantization of  $\mu$  iff it is a  $W_p$ -projection of  $\mu$  on  $\mathcal{P}(\mathbb{R}^d, N)$  where the Wasserstein distance  $W_p$  is defined by

$$W_p(\mu, \nu)^p = \inf_{\pi \in \mathcal{P}(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy). \quad (0.2)$$

In the quadratic  $p = 2$  case, any stationary and therefore any optimal  $N$ -quantization  $\hat{\mu}^N$  of the measure  $\mu$  is smaller than  $\mu$  in the convex order. Moreover  $W_2(\hat{\mu}^N, \mu) = M_2(\hat{\mu}^N, \mu)$  where

$$M_p(\eta, \nu) = \inf_{\pi \in \mathcal{M}(\eta, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \text{ for } (\eta, \nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d). \quad (0.3)$$

This enables us to check that any quadratic-optimal primal quantizer for  $\mu$  remains quadratic-optimal for each probability measure smaller than  $\mu$  and greater than its associated quantization for the convex order.

The second section deals with the dual quantization of a compactly supported probability measure  $\mu$  which is obtained by minimizing  $M_p(\mu, \eta)$  over the subset  $\mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)$  of  $\mathcal{P}(\mathbb{R}^d, N)$  consisting in probability measures greater than  $\mu$  in the convex order. Since  $\mathcal{M}(\mu, \eta) \subset \mathcal{P}(\mu, \eta)$ ,  $M_p(\mu, \eta) \geq W_p(\mu, \eta)$  and, in general, the inequality is strict. It turns out, that, even in the quadratic  $p = 2$  case,  $L_p$ -optimal dual quantizations of  $\mu$  are not necessarily  $W_p$ -projections of  $\mu$  on  $\mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)$ . Nevertheless, we check that any quadratic-optimal dual quantizer of  $\mu$  remains quadratic-optimal for each probability measure greater than  $\mu$  and smaller than its associated dual quantization in the convex order.

In the third section, we consider two probability measures  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  such that  $\mu \leq_{cvx} \nu$  with  $\nu$  compactly supported. Since the quantization errors between  $\mu$  and any of its quadratic-optimal primal  $N$ -quantization  $\hat{\mu}^N$  and between  $\nu$  and any of its  $L_p$ -optimal dual  $K$ -quantization  $\check{\nu}^K$  correspond to martingale couplings, we are able to approximate in Wasserstein distance on  $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$  any martingale coupling  $\pi \in \mathcal{M}(\mu, \nu)$  by a martingale coupling  $\bar{\pi}^{N,K} \in \mathcal{M}(\hat{\mu}^N, \check{\nu}^K)$  with a rate given by the quantization errors. We also check that, as  $N, K \rightarrow \infty$ ,  $\bar{\pi}^{N,K}$  converges to  $\pi$  for the much finer adapted Wasserstein distance defined in (0.5) below which captures the temporal structure of probability distributions with two time marginals. Numerous financial applications of this adapted Wasserstein distance have been investigated in [3]. According to [4], the topology induced by this distance is equal to the other adapted topologies which had been introduced in particular in view of financial applications.

The adapted Wasserstein is particularly well suited to deal with Weak Martingale Optimal Transport problems. While their stability with respect to the marginal distributions has only been established in dimension 1 so far [10], this enables us to check in Section 4 that their value function computed numerically for the quantized marginals converge in any dimension to the value for the original probability measures as the numbers of quantization points go to  $\infty$ .

For the reader's convenience, we now list the notations and basic properties that have been introduced in the above text.

#### DEFINITIONS AND NOTATIONS.

Let  $d \in \mathbb{N}^*$

- $|\cdot|$  denotes the canonical Euclidean norm on  $\mathbb{R}^d$ .
- $\text{conv}(A)$  denotes the (closed) convex hull of  $A \subset \mathbb{R}^d$  and  $\text{card}(A)$  or  $|A|$  its cardinality (depending on the context).
- Let  $\mathcal{P}(\mathbb{R}^d)$  denote the set of probability measures on  $\mathbb{R}^d$  endowed with its Borel sigma-field  $\mathcal{B}(\mathbb{R}^d)$ .
- For  $p \geq 1$ , let  $\mathcal{P}_p(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty\}$ .
- For  $p \geq 1$ , let  $\mathcal{P}_{\leq} \times \mathcal{P}_p(\mathbb{R}^d) = \{(\mu, \nu) \in \mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d) : \mu \leq_{cvx} \nu\}$ .
- For every integer  $N \geq 1$ , we denote by  $\mathcal{P}(\mathbb{R}^d, N)$  the set of distributions on  $\mathbb{R}^d$  whose support contains at most  $N$  points.
- For  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ , let

$$\mathcal{P}(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \forall A \in \mathcal{B}(\mathbb{R}^d), \pi(A \times \mathbb{R}^d) = \mu(A) \text{ and } \pi(\mathbb{R}^d \times A) = \nu(A) \right\}$$

denote the set of couplings between  $\mu$  and  $\nu$ . For  $\pi \in \mathcal{P}(\mu, \nu)$ , we denote by  $\pi_x(dy)$  the ( $\mu(dx)$  a.e. unique) Markov kernel such that  $\pi(dx, dy) = \mu(dx)\pi_x(dy)$ .

- For  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$  with  $\mu \leq_{cvx} \nu$ , let

$$\mathcal{M}(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mu, \nu) : \mu(dx) \text{ a.e. } \int_{\mathbb{R}^d} y \pi_x(dy) = x \right\}$$

denote the (non empty according to the Strassen theorem) set of martingale couplings between  $\mu$  and  $\nu$ .

- For  $p \geq 1$  and  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ , let

$$W_p(\mu, \nu) = \inf_{\pi \in \mathcal{P}(\mu, \nu)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \right)^{1/p} \leq \infty$$

denote the Wasserstein distance with index  $p$ . This is a complete metric on  $\mathcal{P}_p(\mathbb{R}^d)$ .

- For  $p \geq 1$  and  $(\mu, \nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$ , let

$$M_p(\mu, \nu) = \inf_{\pi \in \mathcal{M}(\mu, \nu)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \right)^{1/p} \leq \infty.$$

Since  $\mathcal{M}(\mu, \nu) \subset \mathcal{P}(\mu, \nu)$ , we clearly have,  $W_p(\mu, \nu) \leq M_p(\mu, \nu)$ . In the quadratic  $p = 2$  case, when  $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ ,

$$\begin{aligned} \forall \pi \in \mathcal{M}(\mu, \nu), \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^2 \pi(dx, dy) &= \int_{\mathbb{R}^d} |y|^2 \nu(dy) - 2 \int_{\mathbb{R}^d} x \cdot \int_{\mathbb{R}^d} y \pi_x(dy) \mu(dx) + \int_{\mathbb{R}^d} |x|^2 \mu(dx) \\ &= \int_{\mathbb{R}^d} |y|^2 \nu(dy) - \int_{\mathbb{R}^d} |x|^2 \mu(dx) \end{aligned}$$

so that

$$M_2^2(\mu, \nu) = \int_{\mathbb{R}^d} |y|^2 \nu(dy) - \int_{\mathbb{R}^d} |x|^2 \mu(dx). \quad (0.4)$$

- For  $\pi \in \mathcal{P}(\mu, \nu)$  and  $\tilde{\pi} \in \mathcal{P}(\tilde{\mu}, \tilde{\nu})$  we consider the adapted Wasserstein distance with index  $p \geq 1$  between  $\pi(dx, dy) = \mu(dx) \pi_x(dy)$  and  $\tilde{\pi}(d\tilde{x}, d\tilde{y}) = \tilde{\mu}(d\tilde{x}) \tilde{\pi}_{\tilde{x}}(d\tilde{y})$  :

$$AW_p(\pi, \tilde{\pi}) = \inf_{m \in \mathcal{P}(\mu, \tilde{\mu})} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} (|x - \tilde{x}|^p + W_p^p(\pi_x, \tilde{\pi}_{\tilde{x}})) m(dx, d\tilde{x}) \right)^{1/p}. \quad (0.5)$$

## 1 Primal (Voronoi) quantization and Wasserstein projection

In this subsection, we make a connection between primal quantization and various projections (in the Wasserstein sense), including, in the quadratic case, with the one mentioned above in the introduction. Let us first recall the following basic facts about the (primal) Voronoi quantization of  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$  with  $p \geq 1$  (see [24, 41, 37] among others):

- Let  $\Gamma = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$  denote a finite subset of size  $N$ . The  $L^p$ -quantization error modulus  $e_p(\Gamma, \mu)$  satisfies

$$e_p(\Gamma, \mu)^p = \int_{\mathbb{R}^d} |x - \text{Proj}_{\Gamma}(x)|^p \mu(dx) \quad (1.6)$$

where  $\text{Proj}_{\Gamma}$  denotes a Borel nearest neighbour projection on  $\Gamma$  satisfying  $|x - \text{Proj}_{\Gamma}(x)| = \text{dist}(x, \Gamma)$ . If  $X \sim \mu$ , the random variable  $\text{Proj}_{\Gamma}(X)$  with law  $\mu \circ \text{Proj}_{\Gamma}^{-1}$  is called  $\Gamma$ -quantization of  $X$ .

- For any level  $N \geq 1$ , there exists an optimal *grid* or *N-quantizer*  $\Gamma_{p,N}$  such that

$$e_{p,N}(\mu) := \inf \left\{ e_p(\Gamma, \mu) : \Gamma \subset \mathbb{R}^d, |\Gamma| \leq N \right\} = e_p(\Gamma_{p,N}, \mu).$$

When  $\text{card}(\text{supp}(\mu)) \leq N$  then  $\Gamma_{p,N} = \text{supp}(\mu)$  and when  $\text{card}(\text{supp}(\mu)) > N$ , then  $\Gamma_{p,N}$  has exactly  $N$  pairwise distinct elements. Moreover, Theorem 4.2 in [24] ensures that

$$\mu \left( \left\{ x \in \mathbb{R}^d : \exists y \neq \tilde{y} \in \Gamma_{p,N}, |x - y| = |x - \tilde{y}| = \text{dist}(x, \Gamma_{p,N}) \right\} \right) = 0, \quad (1.7)$$

so that  $\mu \circ \text{Proj}_{\Gamma_{p,N}}^{-1}$  does not depend on the choice of the Borel nearest neighbour projection  $\text{Proj}_{\Gamma_{p,N}}$  on  $\Gamma_{p,N}$ . We denote this probability measure by  $\hat{\mu}^{\Gamma_{p,N}}$ . In the same way, when  $X \sim \mu$ ,  $\text{Proj}_{\Gamma_{p,N}}(X)$  a.s. does not depend on the Borel nearest neighbour projection  $\text{Proj}_{\Gamma_{p,N}}$  and is denoted  $\hat{X}^{\Gamma_{p,N}}$ .

– In the quadratic case ( $p = 2$ ), any optimal quantization grid  $\Gamma_{2,N}$  (possibly not unique) and its induced quantization  $\hat{X}^{\Gamma_{2,N}} = \text{Proj}_{\Gamma_{2,N}}(X)$  with distribution  $\hat{\mu}^{\Gamma_{2,N}}$  satisfy a *stationarity* (or *self-consistency*) property (see e.g. [24], [41] or [37], Proposition 5.1 among others) that is

$$\mathbb{E}(X | \hat{X}^{\Gamma_{2,N}}) = \hat{X}^{\Gamma_{2,N}} \text{ so that } \hat{\mu}^{\Gamma_{2,N}} \leq_{cvx} \mu. \quad (1.8)$$

Indeed, the support of the distribution of  $\mathbb{E}(X | \hat{X}^{\Gamma_{2,N}})$  is equal to  $\tilde{\Gamma}_{2,N} := \{\mathbb{E}(X | \hat{X}^{\Gamma_{2,N}} = x), x \in \Gamma_{2,N}\}$  (when  $\text{card}(\text{supp}(\mu)) \leq N$ ,  $\hat{X}^{\Gamma_{2,N}} = X$  and  $\tilde{\Gamma}_{2,N} = \Gamma_{2,N} = \text{supp}(\mu)$ ), contains at most  $N$  points and  $e_2(\tilde{\Gamma}_{2,N}, \mu)^2 = \mathbb{E}(\text{dist}(X, \tilde{\Gamma}_{2,N})^2) \leq \mathbb{E}(|X - \mathbb{E}(X | \hat{X}^{\Gamma_{2,N}})|^2)$ . As a consequence,

$$\begin{aligned} e_{2,N}(\mu)^2 &\leq e_2(\tilde{\Gamma}_{2,N}, \mu)^2 \leq \mathbb{E}(|X - \mathbb{E}(X | \hat{X}^{\Gamma_{2,N}})|^2) = \mathbb{E}(|X - \hat{X}^{\Gamma_{2,N}}|^2) - \mathbb{E}(|\hat{X}^{\Gamma_{2,N}} - \mathbb{E}(X | \hat{X}^{\Gamma_{2,N}})|^2) \\ &= e_{2,N}(\mu)^2 - \mathbb{E}(|\hat{X}^{\Gamma_{2,N}} - \mathbb{E}(X | \hat{X}^{\Gamma_{2,N}})|^2), \end{aligned}$$

so that the second term in the right-hand side vanishes. Therefore the distribution of  $(\hat{X}^{\Gamma_{2,N}}, X)$  belongs to  $\mathcal{M}(\hat{\mu}^{\Gamma_{2,N}}, \mu)$  and

$$e_{2,N}^2(\mu) = \mathbb{E}[|X - \hat{X}^{\Gamma_{2,N}}|^2] = \mathbb{E}[|X|^2] - \mathbb{E}[|\hat{X}^{\Gamma_{2,N}}|^2] = M_2^2(\hat{\mu}^{\Gamma_{2,N}}, \mu). \quad (1.9)$$

**Proposition 1.1** *Let  $p \in [1, +\infty)$  and  $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ .*

(a) *Let  $\Gamma \subset \mathbb{R}^d$  be a finite set and  $\mathcal{P}(\Gamma)$  denote the subset of  $\Gamma$ -supported distributions. Then*

$$W_p(\mu, \mathcal{P}(\Gamma)) := \inf_{\nu \in \mathcal{P}(\Gamma)} W_p(\mu, \nu) = e_p(\Gamma, \mu) := \|\text{dist}(\cdot, \Gamma)\|_{L^p(\mu)}$$

*and for any Borel nearest neighbour projection  $\text{Proj}_{\Gamma}$  on  $\Gamma$ ,  $\mu \circ \text{Proj}_{\Gamma}^{-1}$  is a  $W_p$ -projection of  $\mu$  on  $\mathcal{P}(\Gamma)$ .*

(b) *The probability measure  $\nu \in \mathcal{P}(\mathbb{R}^d, N)$  is a  $W_p$ -projection of  $\mu$  on  $\mathcal{P}(\mathbb{R}^d, N)$  iff  $\nu = \hat{\mu}^{\Gamma_N}$  for some  $L_p$ -optimal  $N$ -quantizer  $\Gamma_N$  of  $\mu$ . Moreover,  $W_p(\mu, \mathcal{P}(\mathbb{R}^d, N)) = e_{p,N}(\mu)$ .*

(c) *Quadratic case ( $p = 2$ ). A subset  $\Gamma$  of  $\mathbb{R}^d$  with cardinality at most  $N$  is a quadratic optimal  $N$ -quantizer of  $\mu$  iff there exists a probability measure  $\nu \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$  such that  $\nu(\Gamma) = 1$  and one of the following equivalent conditions is satisfied*

–  *$\nu$  is a  $W_2$ -projection of  $\mu$  on  $\mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$  i.e.*

$$W_2(\mu, \nu) = W_2(\mu, \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)) := \inf_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} W_2(\mu, \eta),$$

–  $\int_{\mathbb{R}^d} |x|^2 \nu(dx) = \sup_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |x|^2 \eta(dx)$ .

*Moreover, we then have  $W_2(\nu, \mu) = M_2(\nu, \mu)$  and  $\nu = \hat{\mu}^{\Gamma}$ .*

Apart from the interpretation in terms of  $W_p$ -projection, the first statement can be found in Lemma 3.4 p.33 [24]. Before proving the proposition, let us state and check some easy consequence of the necessary and sufficient condition in (b).

**Corollary 1.2** *Let  $\Gamma_{2,N}$  be a quadratic optimal  $N$ -quantizer of  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ . Then for any probability measure  $\nu$  such that  $\hat{\mu}^{\Gamma_{2,N}} \leq_{cvx} \nu \leq_{cvx} \mu$ ,  $\Gamma_{2,N}$  is a quadratic optimal  $N$ -quantizer of  $\nu$  and  $\hat{\nu}^{\Gamma_{2,N}} = \hat{\mu}^{\Gamma_{2,N}}$ .*

**Proof of Corollary 1.2.** Since  $\hat{\mu}^{\Gamma_{2,N}} \in \mathcal{P}_{\leq \nu}(\mathbb{R}^d, N) \subset \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$  and  $\int_{\mathbb{R}^d} |x|^2 \hat{\mu}^{\Gamma_{2,N}}(dx) = \sup_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |x|^2 \eta(dx)$  by the necessary condition in Proposition 1.1 (c), one has

$$\int_{\mathbb{R}^d} |x|^2 \hat{\mu}^{\Gamma_{2,N}}(dx) = \sup_{\eta \in \mathcal{P}_{\leq \nu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |x|^2 \eta(dx).$$

Therefore, by the sufficient condition in Proposition 1.1 (c),  $\Gamma_{2,N}$  is a quadratic optimal  $N$ -quantizer of  $\nu$  and  $\hat{\nu}^{\Gamma_{2,N}} = \hat{\mu}^{\Gamma_{2,N}}$ .  $\square$

**Proof of Proposition 1.1.** (a) Let  $\nu \in \mathcal{P}(\Gamma)$  and  $\pi \in \mathcal{P}(\mu, \nu)$ . Then  $\mu(dx)$  a.e.,  $\pi_x(\Gamma) = 1$  so that  $\pi_x(dy)$  a.e.  $|x - y| \geq \text{dist}(x, \Gamma)$ . Therefore

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \geq \int_{\mathbb{R}^d} \text{dist}(x, \Gamma)^p \mu(dx) = e_p(\Gamma, \mu)^p.$$

Taking the infimum over  $\pi \in \mathcal{P}(\mu, \nu)$  and  $\nu \in \mathcal{P}(\Gamma)$ , we deduce that  $W_p(\mu, \mathcal{P}(\Gamma))^p \geq e_p(\Gamma, \mu)^p$ . Now let  $\text{Proj}_\Gamma$  denote a Borel nearest neighbour projection on  $\Gamma$ . Since  $\mu \circ \text{Proj}_\Gamma^{-1} \in \mathcal{P}(\Gamma)$  and  $\mu \circ (I_d, \text{Proj}_\Gamma)^{-1} \in \mathcal{P}(\mu, \mu \circ \text{Proj}_\Gamma^{-1})$ , we have

$$W_p^p(\mu, \mathcal{P}(\Gamma)) \leq W_p^p(\mu, \mu \circ \text{Proj}_\Gamma^{-1}) \leq \int_{\mathbb{R}^d} |x - \text{Proj}_\Gamma(x)|^p \mu(dx) = e_p(\Gamma, \mu)^p,$$

where the equality follows from (1.6). Therefore the inequalities are equalities and  $\mu \circ \text{Proj}_\Gamma^{-1}$  is a  $W_p$ -projection of  $\mu$  on  $\mathcal{P}(\Gamma)$ .

(b) Let  $\Gamma_{p,N}$  be an  $L^p$ -optimal  $N$ -quantizer of  $\mu$ . Then, for any subset  $\Gamma$  of  $\mathbb{R}^d$  with at most  $N$  points,  $e_p(\Gamma_{p,N}, \mu) = e_{p,N}(\mu) \leq e_p(\Gamma, \mu)$ . With (a), we deduce that

$$W_p(\mu, \hat{\mu}^{\Gamma_{p,N}}) \leq W_p(\mu, \mathcal{P}(\Gamma)).$$

By taking the infimum over  $\Gamma$ , we deduce that  $W_p(\mu, \hat{\mu}^{\Gamma_{p,N}}) \leq W_p(\mu, \mathcal{P}(\mathbb{R}^d, N))$  and  $\hat{\mu}^{\Gamma_{p,N}}$  is a  $W_p$ -projection of  $\mu$  on  $\mathcal{P}(\mathbb{R}^d, N)$  so that

$$e_{p,N}(\mu) = e_p(\mu, \Gamma_{p,N}) = W_p(\mu, \hat{\mu}^{\Gamma_{p,N}}) = W_p(\mu, \mathcal{P}(\mathbb{R}^d, N)).$$

Conversely, let  $\nu$  be a projection of  $\mu$  on  $\mathcal{P}(\mathbb{R}^d, N)$  and let  $\Gamma_N = \{x \in \mathbb{R}^d : \nu(\{x\}) > 0\}$ . The cardinality of  $\Gamma_N$  is at most  $N$ . For any subset  $\Gamma$  of  $\mathbb{R}^d$  with at most  $N$  points,

$$W_p(\mu, \mathcal{P}(\Gamma_N)) \leq W_p(\mu, \nu) = W_p(\mu, \mathcal{P}(\mathbb{R}^d, N)) \leq W_p(\mu, \mathcal{P}(\Gamma)). \quad (1.10)$$

Therefore, by (a),  $e_p(\Gamma_N, \mu) \leq e_p(\Gamma, \mu)$  and since  $\Gamma$  is arbitrary, we deduce that  $\Gamma_N$  is an  $L^p$ -optimal  $N$ -quantizer of  $\mu$ . Moreover, the choice  $\Gamma = \Gamma_N$  in (1.10) implies that the first inequality is an equality so that, with (a),  $W_p^p(\mu, \nu) = \int_{\mathbb{R}^d} \text{dist}(x, \Gamma_N)^p \mu(dx)$ . Hence, for any  $W_p$ -optimal coupling  $\pi \in \mathcal{P}(\mu, \nu)$ ,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (|y - x|^p - \text{dist}(x, \Gamma_N)^p) \pi(dx, dy) = 0.$$



Since  $1 = \nu(\Gamma_N) = \int_{\mathbb{R}^d} \pi_x(\Gamma_N) \mu(dx)$ ,  $\pi(dx, dy)$  a.e.,  $|y - x|^p \geq \text{dist}(x, \Gamma_N)^p$ . Therefore,  $\pi(dx, dy)$  a.e.  $y \in \Gamma_N$  and  $|y - x| = \text{dist}(x, \Gamma_N)$ . With (1.7), we conclude that  $\mu(dx)$  a.e. there is a unique point  $\hat{x}^{\Gamma_N} \in \Gamma_N$  such that  $|\hat{x}^{\Gamma_N} - x| = \text{dist}(x, \Gamma_N)$  and  $\pi_x(dy) = \delta_{\hat{x}^{\Gamma_N}}(dy)$ . Therefore the second marginal  $\nu$  of  $\pi$  is equal to  $\hat{\mu}^{\Gamma_N}$ .

(c) In the quadratic case ( $p = 2$ ), by (b), (1.8) and (1.9), any  $W_2$ -projection  $\nu$  of  $\mu$  on  $\mathcal{P}(\mathbb{R}^d, N)$  is characterized by the existence of a quadratic optimal  $N$ -quantizer  $\Gamma_{2,N}$  of  $\mu$  such that  $\nu = \hat{\mu}^{\Gamma_{2,N}}$ , belongs to the smaller set  $\mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$  and satisfies  $e_{2,N}^2(\mu) = M_2^2(\nu, \mu)$ . Therefore the  $W_2$ -projections of  $\mu$  on  $\mathcal{P}(\mathbb{R}^d, N)$  and on  $\mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$  coincide. To conclude the proof, let us check that  $\nu$  is such a projection iff  $\int_{\mathbb{R}^d} |x|^2 \nu(dx) = \sup_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |x|^2 \eta(dx)$ . If  $\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$ , then, by the comparison between  $W_2$  and  $M_2$  given in the introduction and (0.4), one has

$$W_2^2(\mu, \eta) \leq M_2^2(\eta, \mu) = \int_{\mathbb{R}^d} |y|^2 \mu(dy) - \int_{\mathbb{R}^d} |x|^2 \eta(dx). \quad (1.11)$$

Let  $\nu$  be a  $W_2$ -projection of  $\mu$  on  $\mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$ . Using (b) for the third equality then (1.11) for the last inequality and the last equality, we obtain that

$$\begin{aligned} \int_{\mathbb{R}^d} |y|^2 \mu(dy) - \int_{\mathbb{R}^d} |x|^2 \nu(dx) &= M_2^2(\nu, \mu) = e_{2,N}^2(\mu) = W_2^2(\mu, \mathcal{P}(\mathbb{R}^d, N)) \\ &= W_2^2(\mu, \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)) \leq \inf_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} M_2^2(\eta, \mu) \\ &= \int_{\mathbb{R}^d} |y|^2 \mu(dy) - \sup_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |x|^2 \eta(dx). \end{aligned}$$

Since  $\nu \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$ , the two inequalities are equalities. Therefore

$$\int_{\mathbb{R}^d} |x|^2 \nu(dx) = \sup_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |x|^2 \eta(dx)$$

and since  $W_2^2(\mu, \mathcal{P}(\mathbb{R}^d, N)) \leq W_2^2(\mu, \nu) \leq M_2^2(\nu, \mu)$ , these two inequalities are equalities and  $W_2^2(\mu, \nu) = M_2^2(\nu, \mu)$ . Moreover,

$$e_{2,N}^2(\mu) = W_2^2(\mu, \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)) = \int_{\mathbb{R}^d} |y|^2 \mu(dy) - \sup_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |x|^2 \eta(dx).$$

If  $\nu \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$  is such that  $\int_{\mathbb{R}^d} |x|^2 \nu(dx) = \sup_{\eta \in \mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |x|^2 \eta(dx)$ , the last equality combined with (1.11) written for  $\eta = \nu$  ensures that  $\nu$  is a  $W_2$ -projection of  $\mu$  on  $\mathcal{P}_{\leq \mu}(\mathbb{R}^d, N)$ .  $\square$

**Remark about uniqueness.** As a consequence of Proposition 1.1 (b), it turns out that the uniqueness of  $W_p$ -projections of  $\mu$  on  $\mathcal{P}(\mathbb{R}^d, N)$ , that of distributions  $\hat{\mu}^N$  of  $L^p$ -optimal  $N$ -quantizations and that of  $L^p$ -optimal  $N$ -quantizers are equivalent. In dimension  $d = 1$ , for  $p = 2$ , distributions with log-concave densities have a unique optimal  $N$ -quantizer (see Kiefer [35]) hence this projection is unique. In higher dimension, a general result seems difficult to reach: indeed, the  $\mathcal{N}(0; I_d)$  distribution, being invariant under the action of  $\mathcal{O}(d, \mathbb{R})$  (orthogonal transforms), so are the (hence infinite) sets of its optimal quantizers at levels  $N \geq 2$ .

Let us recall the sharp rate of convergence of the  $L^p$ -quantization error stated for instance in Theorem 5.2 [37].

**Theorem 1.3 (Pierce Lemma for primal quantization)** *Let  $p \geq 1$  and  $\eta > 0$ . For every dimension  $d \geq 1$ , there exists a real constant  $\tilde{C}_{d,\eta,p}^{vor} > 0$  such that, for every random vector  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ ,*

$$e_{p,N}(X) \leq \tilde{C}_{d,\eta,p}^{vor} N^{-\frac{1}{d}} \sigma_{p+\eta}(X) \quad (1.12)$$

where, for every  $r > 0$ ,  $\sigma_r(X) = \inf_{a \in \mathbb{R}^d} \|X - a\|_r \leq +\infty$ .

**Example.** Let us explicit the quadratic optimal primal  $N$ -quantizer of  $\mu(dx) = \frac{1_{]0,1[}(x)}{2\sqrt{x}} dx$ . In [20], by checking that the distortion function has a unique critical point, Fort and Pagès prove uniqueness and derive semi-closed forms for optimal quantizers of three families of one-dimensional distributions indexed by  $\rho > 0$  : the exponential distributions  $1_{(0,\infty)}(x)\rho e^{-\rho x} dx$ , the power distributions  $\rho 1_{(0,1)}(x)x^{\rho-1} dx$  on the interval  $(0, 1)$  which include  $\mu$  for  $\rho = \frac{1}{2}$  and the power distributions  $\rho 1_{(1,\infty)}(x)x^{-(1+\rho)} dx$  on the interval  $(1, +\infty)$ . For  $\mu(dx)$  and  $p = 2$ , their semi-closed form relies on the inductive solution of third degree equations. Working with a different parametrization, we will end up with much simpler second degree equations. We have  $F_\mu(x) = 1_{\{x \geq 0\}}\sqrt{x}$  and  $F_\mu^{-1}(u) = u^2$  for  $u \in (0, 1)$ . We parametrize the boundaries of the optimal quadratic Voronoi cells by  $x_{k+\frac{1}{2}} = F_\mu^{-1}(q_k) = q_k^2$  for  $k \in \{0, \dots, N\}$  where  $0 = q_0 < q_1 < q_2 < \dots < q_{N-1} < q_N = 1$ . In particular  $x_{0+\frac{1}{2}} = 0$  and  $x_{N+\frac{1}{2}} = 1$ . By the stationarity property (1.8), for  $k \in \{1, \dots, N\}$ ,  $x_k = \frac{1}{q_k - q_{k-1}} \int_{x_{k-\frac{1}{2}}}^{x_{k+\frac{1}{2}}} x \mu(dx)$  (by (1.7),  $\mu(\{x_{k-\frac{1}{2}}\}) = \mu(\{x_{k+\frac{1}{2}}\}) = 0$  and we do not need to worry about the inclusion of the boundary points in the integration interval). The equality  $x_{k+\frac{1}{2}} = \frac{x_k + x_{k+1}}{2}$  valid for  $k \in \{1, \dots, N-1\}$  also writes

$$F_\mu^{-1}(q_k) = \frac{1}{2} \left( \frac{1}{q_k - q_{k-1}} \int_{q_{k-1}}^{q_k} F_\mu^{-1}(u) du + \frac{1}{q_{k+1} - q_k} \int_{q_k}^{q_{k+1}} F_\mu^{-1}(u) du \right).$$

For the above choice of  $\mu$ , we obtain

$$q_k^2 = \frac{1}{6} (q_k^2 + q_k q_{k-1} + q_{k-1}^2 + q_{k+1}^2 + q_{k+1} q_k + q_k^2) \text{ so that } q_{k+1}^2 + q_{k+1} q_k + q_k q_{k-1} + q_{k-1}^2 - 4q_k^2 = 0.$$

We deduce that  $q_{k+1} = c_{k+1} q_1$ , where  $c_{k+1}^2 + c_{k+1} c_k + c_k c_{k-1} + c_{k-1}^2 - 4c_k^2 = 0$  so that  $c_{k+1} = \frac{\sqrt{17c_k^2 - 4c_k c_{k-1} - 4c_{k-1}^2} - c_k}{2}$ . Starting from  $c_0 = 0$  and  $c_1 = 1$ , we easily compute inductively the factors  $c_k$  (for instance  $c_2 = \frac{\sqrt{17}-1}{2}$ ). To ensure  $q_N = 1$ , we need  $q_1 = \frac{1}{c_N}$ . Therefore  $q_k = \frac{c_k}{c_N}$  for  $k \in \{0, \dots, N\}$  and the optimal quadratic primal grid is given by

$$x_k = \frac{1}{q_k - q_{k-1}} \int_{q_{k-1}}^{q_k} F_\mu^{-1}(u) du = \frac{q_k^2 + q_k q_{k-1} + q_{k-1}^2}{3} = \frac{c_k^2 + c_k c_{k-1} + c_{k-1}^2}{3c_N^2}, \quad k \in \{1, \dots, N\}.$$

Of course for  $a < b$ , when  $\mu(dx) = \frac{1_{]a,b[}(x)}{2\sqrt{(b-a)(x-a)}} dx$ , then  $x_k = a + (b-a) \frac{c_k^2 + c_k c_{k-1} + c_{k-1}^2}{3c_N^2}$  and when  $\mu(dx) = \frac{1_{]a,b[}(x)}{2\sqrt{(b-a)(b-x)}} dx$ ,  $x_{N+1-k} = b - (b-a) \frac{c_k^2 + c_k c_{k-1} + c_{k-1}^2}{3c_N^2}$ .

## 2 Dual (Delaunay) quantization

We assume throughout this section that  $\mu$  is compactly supported. Let  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$  be a random vector lying in  $L^\infty(\mathbb{P})$  with distribution  $\mu$ . Optimal dual (or Delaunay) quantization as introduced in [44] relies on the best approximation which can be achieved by a discrete random vector  $\check{X}$  that satisfies a certain stationarity assumption on the extended probability space  $(\Omega \times [0, 1], \mathcal{A} \otimes \mathcal{B}([0, 1]), \mathbb{P} \otimes \lambda)$  where  $\mathcal{B}([0, 1])$  and  $\lambda$  respectively denote the Borel sigma field and the Lebesgue measure on the interval  $[0, 1]$ . To be more precise, we define, for  $p \in [1, +\infty)$ ,

$$d_{p,N}(X) = \inf_{\check{X}} \left\{ \|X - \check{X}\|_p : \check{X} : (\Omega \times [0, 1], \mathcal{A} \otimes \mathcal{B}([0, 1]), \mathbb{P} \otimes \lambda) \rightarrow \mathbb{R}^d, \right. \\ \left. \text{card}(\check{X}(\Omega \times [0, 1])) \leq N \text{ and } \mathbb{E}(\check{X}|X) = X \right\}.$$

For every level  $N \geq d+1$ , the set of such  $\check{X}$  is not empty. Indeed, one may choose  $d+1$  points whose convex hull has a non empty interior and includes the support of  $\mu$ . Then the unique probability measure supported on these points with the same expectation as  $\mu$  belongs to the set  $\mathcal{P}_{\geq \mu}(\mathbb{R}^d, d+1)$  of distributions dominating  $\mu$  for the convex order and supported by at most  $d+1$  elements. By Lemma 2.22 in [34], we see that for each  $\nu \in \mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)$  and each martingale coupling  $\pi \in \mathcal{M}(\mu, \nu)$ , there exists on  $(\Omega \times [0, 1], \mathcal{A} \otimes \mathcal{B}([0, 1]), \mathbb{P} \otimes \lambda)$  a random vector  $\check{X}$  such that  $(X, \check{X})$  is distributed according to  $\pi$  and therefore satisfies  $\mathbb{E}(\check{X}|X) = X$ . Hence

$$d_{p,N}(X)^p = \inf_{\nu \in \mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)} \inf_{\pi \in \mathcal{M}(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^p \pi(dx, dy) = \inf_{\nu \in \mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)} M_p^p(\mu, \nu). \quad (2.13)$$

As a consequence,  $d_{p,N}(X)$  only depends on the distribution  $\mu$  of  $X$  and can subsequently also be denoted  $d_{p,N}(\mu)$ . Next, one easily checks that  $\mathcal{P}_{\geq \mu}(\mathbb{R}^d, N) = \bigcup_{\Gamma \in \mathcal{G}_N} \mathcal{P}_{\geq \mu}(\Gamma)$  where

$$\mathcal{G}_N = \{\Gamma \subset \mathbb{R}^d \text{ with cardinality } \leq N \text{ and such that } \text{supp}(\mu) \subset \text{conv}(\Gamma)\}, \\ \text{and } \mathcal{P}_{\geq \mu}(\Gamma) = \{\nu \in \mathcal{P}(\mathbb{R}^d) : \mu \leq_{\text{conv}} \nu \text{ and } \nu(\Gamma) = 1\}.$$

For  $\Gamma \in \mathcal{G}_N$ , there exists a dual projection  $\text{Proj}_\Gamma^{\text{del}} : \text{conv}(\Gamma) \times [0, 1] \rightarrow \Gamma$ , also called a *splitting operator*, which satisfies, beyond measurability, the following stationarity property

$$\forall y \in \text{conv}(\Gamma), \quad \int_0^1 \text{Proj}_\Gamma^{\text{del}}(y, u) du = y, \quad (2.14)$$

from which one derives the dual stationarity property

$$\mathbb{E}(\text{Proj}_\Gamma^{\text{del}}(X, U) | X) = X \text{ when } U \sim \mathcal{U}([0, 1]) \text{ is independent of } X. \quad (2.15)$$

The stationarity property remains valid as soon as  $X$  is  $\text{conv}(\Gamma)$ -valued and implies that the distribution of  $\text{Proj}_\Gamma^{\text{del}}(X, U)$  belongs to  $\mathcal{P}_{\geq \mu}(\Gamma)$  which is therefore non empty.

For  $\Gamma \in \mathcal{G}_N$ , let

$$d_p(\mu, \Gamma)^p = \inf_{\nu \in \mathcal{P}_{\geq \mu}(\Gamma)} M_p^p(\mu, \nu),$$

so that  $d_{p,N}(\mu)^p = \inf_{\Gamma \in \mathcal{G}_N} d_p(\mu, \Gamma)^p$ .

In dimension  $d = 1$ , when  $\Gamma = \{x_1, x_2, \dots, x_N\}$  with  $x_1 < x_2 < \dots < x_N$ , the probability measure minimizing  $\nu \mapsto M_p^p(\mu, \nu)$  over  $\mathcal{P}_{\geq \mu}(\Gamma)$  is the distribution  $\check{\mu}^\Gamma$  of  $\text{Proj}_\Gamma^{\text{del}}(X, U)$  for the splitting operator

$$\text{Proj}_\Gamma^{\text{del}}(x, u) = \sum_{i=1}^{N-1} 1_{[x_i, x_{i+1})}(x) \left( 1_{\{u \leq \frac{x_{i+1}-x}{x_{i+1}-x_i}\}} x_i + 1_{\{u > \frac{x_{i+1}-x}{x_{i+1}-x_i}\}} x_{i+1} \right) + 1_{\{x=x_N\}} x_N.$$

Moreover, the coupling minimizing  $\int_{\mathbb{R} \times \mathbb{R}} |x - y|^p \pi(dx, dy)$  over  $\mathcal{M}(\mu, \check{\mu}^\Gamma)$  is the distribution of  $(X, \text{Proj}_\Gamma^{\text{del}}(X, U))$ . Last, according to the remark after Proposition 10 in [44], when  $\mu \leq_{\text{cvx}} \eta$  with  $\eta$  compactly supported in  $[x_1, x_N]$ , then  $\check{\mu}^\Gamma \leq_{\text{cvx}} \check{\eta}^\Gamma$ . This can be seen using the affine interpolation on  $\Gamma$

$$\check{\varphi}^\Gamma(x) := 1_{(-\infty, x_1) \cup [x_N, +\infty)}(x) \varphi(x) + \sum_{i=1}^{N-1} 1_{[x_i, x_{i+1})}(x) \left( \frac{x_{i+1} - x}{x_{i+1} - x_i} \varphi(x_i) + \frac{x - x_i}{x_{i+1} - x_i} \varphi(x_{i+1}) \right)$$

of a convex function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ . Indeed  $\check{\varphi}^\Gamma$  is still convex and one has

$$\int_{\mathbb{R}} \varphi(x) \check{\mu}^\Gamma(dx) = \int_{\mathbb{R}} \check{\varphi}^\Gamma(x) \mu(dx) \leq \int_{\mathbb{R}} \check{\varphi}^\Gamma(x) \eta(dx) = \int_{\mathbb{R}} \varphi(x) \check{\eta}^\Gamma(dx).$$

According to the introduction in [2], this convex order preservation does not generalize to higher dimensions where the minimizers are not so easy to express.

Whatever the dimension  $d \in \mathbb{N}^*$ , for every level  $N \geq d + 1$ , there exists an  $L^p$ -optimal dual quantization grid  $\Gamma_{p,N}^{\text{del}}$  and a splitting operator  $\text{Proj}_{\Gamma_{p,N}^{\text{del}}}^{\text{del}}$  (see [44]) such that

$$d_{p,N}(\mu) = d_p(\mu, \Gamma_{p,N}^{\text{del}}) = \|X - \text{Proj}_{\Gamma_{p,N}^{\text{del}}}^{\text{del}}(X, U)\|_p$$

and  $\text{Proj}_{\Gamma_{p,N}^{\text{del}}}^{\text{del}}(X, U)$  takes each value in  $\Gamma_{p,N}^{\text{del}}$  with positive probability. For more details on this dual projection, see [44, 45] where this notion has been developed and analyzed. We will see in the examples that even in dimension one, the convex order is not preserved by optimal dual quantization.

Notice that by Proposition 1.1 (b), the inequality  $W_p^p(\mu, \nu) \leq M_p^p(\mu, \nu)$  valid for  $\nu \in \mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)$  and (2.13),

$$e_{p,N}(\mu) = W_p(\mu, \mathcal{P}(\mathbb{R}^d, N)) \leq W_p(\mu, \mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)) \leq \inf_{\nu \in \mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)} M_p(\nu, \mu) = d_{p,N}(\mu). \quad (2.16)$$

We may wonder whether the last inequality is an equality. Combining the tightness of any sequence of probability measures in  $\mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)$  minimizing the  $W_p$ -distance to  $\mu$  deduced from the inequality

$$\forall \nu \in \mathcal{P}(\mathbb{R}^d), \int_{\mathbb{R}^d} |x|^p \nu(dx) \leq 2^{p-1} \left( \int_{\mathbb{R}^d} |x|^p \mu(dx) + W_p^p(\mu, \nu) \right),$$

the closedness of  $\mathcal{P}_{\geq \mu}(\mathbb{R}^d, N)$  for the weak convergence topology and the lower semi-continuity of the Wasserstein distance for this topology (see for instance Remark 6.12 p97 [51]), we obtain the

existence of a  $W_p$ -projection  $\tilde{\mu}$  of  $\mu$  on  $\mathcal{P}_{\geq\mu}(\mathbb{R}^d, N)$ . The last inequality in (2.16) is an equality iff the set

$$\left\{ \pi \in \mathcal{P}(\mu, \tilde{\mu}) : W_p^p(\mu, \tilde{\mu}) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^p \pi(dx, dy) \right\}$$

of  $W_p$ -optimal couplings between  $\mu$  and some  $W_p$ -projection  $\tilde{\mu}$  of  $\mu$  on  $\mathcal{P}_{\geq\mu}(\mathbb{R}^d, N)$  intersects  $\mathcal{M}(\mu, \tilde{\mu})$ . Moreover,  $\tilde{\mu}$  is then the distribution of an  $L^p$ -optimal dual  $N$ -quantization of  $\mu$ . But there is no reason why the intersection should be non empty. We also may wonder whether, by a somewhat naive symmetry with the situation described in Proposition 1.1 (b) for the Voronoi quantization, the distribution of an  $L^p$ -optimal dual  $N$ -quantization of  $\mu$  coincides with a  $W_p$ -projection  $\tilde{\mu}$  of  $\mu$  on  $\mathcal{P}_{\geq\mu}(\mathbb{R}^d, N)$ . According to the examples below, this property holds when  $\mu$  is the uniform distribution on the interval  $[0, 1]$  (nevertheless  $W_p(\mathcal{U}[0, 1], \mathcal{P}_{\geq\mathcal{U}[0, 1]}(\mathbb{R}^d, N)) < d_{p,N}(\mathcal{U}[0, 1])$ ) but is not true in general.

Note that, in the quadratic case  $p = 2$ , for  $\nu \in \mathcal{P}_{\geq\mu}(\mathbb{R}^d, N)$ , since  $M_2^2(\mu, \nu) = \int_{\mathbb{R}^d} |y|^2 \nu(dy) - \int_{\mathbb{R}^d} |x|^2 \mu(dx)$ ,

$$d_{2,N}(\mu)^2 = \inf_{\nu \in \mathcal{P}_{\geq\mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |y|^2 \nu(dy) - \int_{\mathbb{R}^d} |x|^2 \mu(dx). \quad (2.17)$$

**Proposition 2.1** *In the quadratic case  $p = 2$ , any optimal dual quantization grid  $\Gamma_{2,N}^{del}$  remains optimal for each probability measure  $\eta$  such that  $\mu \leq_{cvx} \eta \leq_{cvx} \check{\mu}_N$  where  $\check{\mu}_N$  denotes the distribution of  $\text{Proj}_{\Gamma_{2,N}^{del}}^{del}(X, U)$ .*

**Proof.** Let  $\eta$  be such that  $\mu \leq_{cvx} \eta \leq_{cvx} \check{\mu}_N$ . Since  $\inf_{\nu \in \mathcal{P}_{\geq\mu}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |y|^2 \nu(dy)$  is attained for  $\nu = \check{\mu}_N$ , so is  $\inf_{\nu \in \mathcal{P}_{\geq\eta}(\mathbb{R}^d, N)} \int_{\mathbb{R}^d} |y|^2 \nu(dy)$ . Therefore, (2.17) written with  $\eta$  replacing  $\mu$  implies that

$$d_{2,N}(\eta)^2 = \int_{\mathbb{R}^d} |y|^2 \check{\mu}_N(dy) - \int_{\mathbb{R}^d} |x|^2 \eta(dx) = M_2^2(\eta, \check{\mu}_N).$$

With the definition of  $d_2(\eta, \Gamma_{2,N}^{del})$ , we deduce that  $d_{2,N}(\eta)^2 \geq d_2(\eta, \Gamma_{2,N}^{del})^2$ . Therefore  $\Gamma_{2,N}^{del}$  is an optimal dual quadratic quantization grid for  $\eta$ .  $\square$

**Examples.** (a) Let  $\mu = \mathcal{U}[0, 1]$  where, for two real numbers  $a < b$ ,  $\mathcal{U}[a, b]$  denotes the uniform distribution on  $[a, b]$  with density  $\frac{1_{[a,b]}(x)}{b-a}$  with respect to the Lebesgue measure. We consider its approximation by probability measures in  $\mathcal{P}(\mathbb{R}, N)$ . A generic element of  $\mathcal{P}(\mathbb{R}, N)$  writes

$$\nu_N = \sum_{k=1}^N p_k \delta_{x_k}$$

with  $x_1 \leq x_2 \leq \dots \leq x_N$  and  $(p_1, \dots, p_N) \in [0, 1]^N$  satisfying  $\sum_{k=1}^N p_k = 1$ . We will consider the particular choices  $\hat{\mu}_N = \frac{1}{N} \sum_{k=1}^N \delta_{\frac{2k-1}{2N}}$  and  $\check{\mu}_N = \frac{1}{2(N-1)} \delta_0 + \frac{1}{N-1} \sum_{k=2}^{N-1} \delta_{\frac{k-1}{N-1}} + \frac{1}{2(N-1)} \delta_1$  of the respective distributions of the optimal primal and dual quantizations of  $\mu$  on  $N$  points. For  $p \geq 1$ , let us recover that  $\hat{\mu}_N$  is the  $W_p$ -projection of  $\mathcal{U}[0, 1]$  on  $\mathcal{P}(\mathbb{R}, N)$  (consequence of Proposition 1.1 (b)) and check that  $\check{\mu}_N$  is the  $W_p$ -projection of  $\mathcal{U}[0, 1]$  on  $\mathcal{P}_{\geq\mathcal{U}[0, 1]}(\mathbb{R}, N)$ . The image of  $\mu$  by  $(0, 1) \ni u \mapsto F_{\nu_N}^{-1}(u) - u$  is equal to

$$\eta_N := p_1 \mathcal{U}[x_1 - p_1, x_1] + p_2 \mathcal{U}[x_2 - (p_1 + p_2), x_2 - p_1] + \dots + p_N \mathcal{U}[x_N - 1, x_N - (p_1 + \dots + p_{N-1})].$$

When  $\nu_N = \hat{\mu}_N$  (resp.  $\nu_N = \check{\mu}_N$ ) then  $\eta_N = \hat{\eta}_N := \mathcal{U}[-\frac{1}{2N}, \frac{1}{2N}]$  (resp.  $\eta_N = \check{\eta}_N := \mathcal{U}[-\frac{1}{2(N-1)}, \frac{1}{2(N-1)}]$ ). Since  $\hat{\eta}_N = N1_{[-\frac{1}{2N}, \frac{1}{2N}]}(x)dx$  and  $\eta_N$  has a density with respect to the Lebesgue measure with values in  $\{0, 1, 2, \dots, N\}$ ,  $(\eta_N - \hat{\eta}_N)^+$  is supported on the complement of  $[-\frac{1}{2N}, \frac{1}{2N}]$  where  $(\hat{\eta}_N - \eta_N)^+$  is supported. Since both measures share the same mass, we deduce that for  $p \geq 1$ ,

$$\int_{\mathbb{R}} |x|^p (\eta_N - \hat{\eta}_N)^+(dx) \geq \int_{\mathbb{R}} |x|^p (\hat{\eta}_N - \eta_N)^+(dx) \text{ i.e. } \int_{\mathbb{R}} |x|^p \eta_N(dx) \geq \int_{\mathbb{R}} |x|^p \hat{\eta}_N(dx).$$

Using that in dimension  $d = 1$ , the comonotonous coupling is  $W_p$ -optimal, we conclude that

$$\forall \nu_N \in \mathcal{P}(\mathbb{R}, N), W_p^p(\mathcal{U}[0, 1], \nu_N) = \int_{\mathbb{R}} |x|^p \eta_N(dx) \geq \int_{\mathbb{R}} |x|^p \hat{\eta}_N(dx) = W_p^p(\mathcal{U}[0, 1], \hat{\nu}_N),$$

with strict inequality unless  $\eta_N = \hat{\eta}_N$ . When  $\eta_N$  is supported on  $[-\frac{1}{2N}, \frac{1}{2N}]$  (which is clearly equivalent to  $\eta_N = \hat{\eta}_N$ ), then for each  $k \in \{1, \dots, N\}$ ,  $\frac{-1}{2N} \leq x_k - (p_1 + \dots + p_k)$  and  $x_k - (p_1 + \dots + p_{k-1}) \leq \frac{1}{2N}$  so that  $p_k \leq \frac{1}{N}$ . With the normalisation, we deduce that  $p_\ell = \frac{1}{N}$  for each  $\ell \in \{1, \dots, N\}$ , which plugged in the two last inequalities implies that  $x_k = \frac{2k-1}{2N}$  for each  $k \in \{1, \dots, N\}$  so that  $\nu_N = \hat{\mu}_N$ . Therefore, for all  $p \geq 1$ ,  $\hat{\mu}_N$  is the  $W_p$ -projection of  $\mathcal{U}[0, 1]$  on  $\mathcal{P}(\mathbb{R}, N)$  and

$$W_p(\mathcal{U}[0, 1], \mathcal{P}(\mathbb{R}, N)) = W_p(\mathcal{U}[0, 1], \hat{\mu}_N) = \left( \int_{\mathbb{R}} |x|^p \hat{\eta}_N(dx) \right)^{1/p} = \frac{1}{2(p+1)^{1/p} N}.$$

If  $\nu_N \geq_{cvx} \mathcal{U}[0, 1]$ , then  $x_1 \leq 0$  and  $x_N \geq 1$  so that  $[x_1 - p_1, x_1] \cap [x_N - 1, x_N - (p_1 + \dots + p_{N-1})] \subset \{0\}$  and  $\eta_N$  has a density with respect to the Lebesgue measure with values in  $\{0, 1, \dots, N-1\}$ . By repeating the previous argument with  $\hat{\eta}_N$  replaced by  $\check{\eta}_N$ , we obtain that

$$\forall \nu_N \in \mathcal{P}_{\geq \mathcal{U}[0,1]}(\mathbb{R}, N), W_p^p(\mathcal{U}[0, 1], \nu_N) = \int_{\mathbb{R}} |x|^p \eta_N(dx) \geq \int_{\mathbb{R}} |x|^p \check{\eta}_N(dx) = W_p^p(\mathcal{U}[0, 1], \check{\mu}_N),$$

with strict inequality unless  $\eta_N = \check{\eta}_N$ . When  $\nu_N \geq_{cvx} \mathcal{U}[0, 1]$  and  $\eta_N$  is supported on  $[-\frac{1}{2(N-1)}, \frac{1}{2(N-1)}]$  (which is clearly equivalent to  $\eta_N = \check{\eta}_N$ ), then for each  $k \in \{1, \dots, N\}$ ,  $\frac{-1}{2(N-1)} \leq x_k - (p_1 + \dots + p_k)$  and  $x_k - (p_1 + \dots + p_{k-1}) \leq \frac{1}{2(N-1)}$  so that  $p_k \leq \frac{1}{N-1}$  and with the reinforced inequalities  $x_1 \leq 0$  and  $x_N \geq 1$  due to the convex order,  $p_1 \leq \frac{1}{2(N-1)}$  and  $p_N \leq \frac{1}{2(N-1)}$ . With the normalisation, we deduce that  $p_1 = p_N = \frac{1}{2(N-1)}$  and  $p_\ell = \frac{1}{N-1}$  for each  $\ell \in \{2, \dots, N-1\}$ , which plugged in the two (reinforced) inequalities implies that  $x_1 = 0$ ,  $x_k = \frac{k-1}{N-1}$  for each  $k \in \{2, \dots, N-1\}$  and  $x_N = 1$  so that  $\nu_N = \check{\mu}_N$ . Therefore, for all  $p \geq 1$ ,  $\check{\mu}_N$  is the  $W_p$ -projection of  $\mathcal{U}[0, 1]$  on  $\mathcal{P}_{\geq \mathcal{U}[0,1]}(\mathbb{R}, N)$  and

$$\begin{aligned} W_p(\mathcal{U}[0, 1], \mathcal{P}_{\geq \mathcal{U}[0,1]}(\mathbb{R}, N)) &= W_p(\mathcal{U}[0, 1], \check{\mu}_N) = \left( \int_{\mathbb{R}} |x|^p \check{\eta}_N(dx) \right)^{1/p} = \frac{1}{2(p+1)^{1/p} (N-1)} \\ &< \left( \frac{2}{(p+1)(p+2)} \right)^{1/p} \frac{1}{N-1} = d_{2,N}(\mathcal{U}[0, 1]). \end{aligned}$$

Notice that when  $\nu_N \geq_{cvx} \mathcal{U}[0, 1]$ , then the measures  $(\eta_N - \check{\eta}_N)^+$  supported on  $[-\frac{1}{2(N-1)}, \frac{1}{2(N-1)}]^c$  and  $(\check{\eta}_N - \eta_N)^+$  supported on  $[-\frac{1}{2(N-1)}, \frac{1}{2(N-1)}]$  share the same mass and barycenter so that for each convex function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\int_{\mathbb{R}} \varphi(x) (\eta_N - \check{\eta}_N)^+(dx) \geq \int_{\mathbb{R}} \varphi(x) (\check{\eta}_N - \eta_N)^+(dx)$  and  $\eta_N \geq_{cvx} \check{\eta}_N$ .

The probability distribution  $\mathcal{U}[0, 1]$  is smaller in the convex order than its dual quantization  $\mu_6 := \frac{1}{5}\delta_0 + \frac{7}{30}\delta_{\frac{2}{5}} + \frac{1}{15}\delta_{\frac{7}{15}} + \frac{1}{15}\delta_{\frac{8}{15}} + \frac{7}{30}\delta_{\frac{3}{5}} + \frac{1}{5}\delta_1$  on the grid  $\{0, \frac{2}{5}, \frac{7}{15}, \frac{8}{15}, \frac{3}{5}, 1\}$ . Note that  $\mu_6$  is not comparable with  $\check{\mu}_6 = \frac{1}{10}\delta_0 + \frac{1}{5}\left(\delta_{\frac{1}{5}} + \delta_{\frac{2}{5}} + \delta_{\frac{3}{5}} + \delta_{\frac{4}{5}}\right) + \frac{1}{10}\delta_1$  for the convex order. Indeed, we have

$$\begin{aligned}\mu_6 &= \frac{2}{5}\left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{2}{5}}\right) + \frac{1}{5}\left(\frac{1}{6}\delta_{\frac{2}{5}} + \frac{1}{3}\delta_{\frac{7}{15}} + \frac{1}{3}\delta_{\frac{8}{15}} + \frac{1}{6}\delta_{\frac{3}{5}}\right) + \frac{2}{5}\left(\frac{1}{2}\delta_{\frac{3}{5}} + \frac{1}{2}\delta_1\right) \\ \check{\mu}_6 &= \frac{2}{5}\left(\frac{1}{4}\delta_0 + \frac{1}{2}\delta_{\frac{1}{5}} + \frac{1}{4}\delta_{\frac{2}{5}}\right) + \frac{1}{5}\left(\frac{1}{2}\delta_{\frac{2}{5}} + \frac{1}{2}\delta_{\frac{3}{5}}\right) + \frac{2}{5}\left(\frac{1}{4}\delta_{\frac{3}{5}} + \frac{1}{2}\delta_{\frac{4}{5}} + \frac{1}{4}\delta_1\right) \\ \text{with } \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{2}{5}} &\geq_{cvx} \frac{1}{4}\delta_0 + \frac{1}{2}\delta_{\frac{1}{5}} + \frac{1}{4}\delta_{\frac{2}{5}}, \quad \frac{1}{6}\delta_{\frac{2}{5}} + \frac{1}{3}\delta_{\frac{7}{15}} + \frac{1}{3}\delta_{\frac{8}{15}} + \frac{1}{6}\delta_{\frac{3}{5}} \leq_{cvx} \frac{1}{2}\delta_{\frac{2}{5}} + \frac{1}{2}\delta_{\frac{3}{5}} \\ \text{and } \frac{1}{2}\delta_{\frac{3}{5}} + \frac{1}{2}\delta_1 &\geq_{cvx} \frac{1}{4}\delta_{\frac{3}{5}} + \frac{1}{2}\delta_{\frac{4}{5}} + \frac{1}{4}\delta_1\end{aligned}$$

so that  $\int_{\mathbb{R}} \varphi(x)\mu_6(dx) > \int_{\mathbb{R}} \varphi(x)\check{\mu}_6(dx)$  (resp.  $\int_{\mathbb{R}} \varphi(x)\mu_6(dx) < \int_{\mathbb{R}} \varphi(x)\check{\mu}_6(dx)$ ) when the convex function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is strictly convex on  $[0, \frac{2}{5}]$  and affine on  $[\frac{2}{5}, 1]$  (resp. affine on  $[0, \frac{2}{5}]$ , strictly convex on  $[\frac{2}{5}, \frac{3}{5}]$  and affine on  $[\frac{3}{5}, 1]$ ). Since  $\mu_6$  is clearly equal to its  $L^p$ -optimal dual 6-quantization, this shows that the convex order is not preserved by optimal dual quantization.

(b) Let now  $\mu(dx) = 2x\mathbf{1}_{[0,1]}(x)dx$ . We look for  $\nu \in \mathcal{P}_{\geq\mu}(\mathbb{R}, 3)$  minimizing either  $\int_{\mathbb{R}} y^2\nu(dy)$  to compute the law of the optimal quadratic dual quantization of  $\mu$  on  $N = 3$  points or  $W_2^2(\mu, \nu)$  to compute  $\tilde{\mu}$ . Since  $d = 1$ ,  $W_2^2(\mu, \nu)$  is equal to the integral  $\int_0^1 (F_\mu^{-1}(u) - F_\nu^{-1}(u))^2 du$  of the squared difference between the quantile functions of  $\mu$  and  $\nu$ . For the first criterion, we are going to check that it is equivalent to minimize over the following parametric subset of  $\mathcal{P}_{\geq\mu}(\mathbb{R}, 3)$

$$\left\{ \nu_u(dy) = \frac{u}{3}\delta_0(dy) + \frac{1+\sqrt{u}}{3}\delta_{\sqrt{u}}(dy) + \frac{2-\sqrt{u}-u}{3}\delta_1(dy) : u \in (0, 1) \right\}.$$

One has  $\int y^2\nu_u(dy) = \frac{2+u^{3/2}-\sqrt{u}}{3}$  and the infimum is attained for  $u = 1/3$  so that  $\Gamma_{2,3}^{del} = \{0, \frac{1}{\sqrt{3}}, 1\}$ ,  $d_{2,3}(\mu)^2 = d_2(\mu, \Gamma_{2,3}^{del})^2 = \int y^2\nu_{1/3}(dy) - \int x^2\mu(dx) = \frac{1}{6} - \frac{2}{3^{5/2}}$ . On the other hand,

$$\begin{aligned}W_2^2(\mu, \nu_u) &= \int_0^{u/3} (0 - \sqrt{v})^2 dv + \int_{u/3}^{(1+\sqrt{u}+u)/3} (\sqrt{u} - \sqrt{v})^2 dv + \int_{(1+\sqrt{u}+u)/3}^1 (1 - \sqrt{v})^2 dv \\ &= -\frac{1}{6} + \frac{u^{3/2} - \sqrt{u}}{3} + 4 \frac{(1 - \sqrt{u})(1 + \sqrt{u} + u)^{3/2} + u^2}{3^{5/2}}.\end{aligned}$$

One easily checks that  $\frac{d}{du}W_2^2(\mu, \nu_u)|_{u=1/3} > 0$  and that  $W_2^2(\mu, \nu_u)$  is minimal for  $u \simeq 0.326$  so that  $\tilde{\mu} \neq \nu_{1/3}$ . Moreover, since  $\mathcal{P}_{\geq\mu}(\Gamma_{2,3}^{del}) = \{v\nu_{1/3} + (1-v)\nu_0 : v \in [0, 1]\}$  contains  $\nu_{1/3}$ ,

$$W_2^2(\mu, \mathcal{P}_{\geq\mu}(\Gamma_{2,3}^{del})) \leq W_2^2(\mu, \nu_{1/3}) \simeq 0.0199758 < 0.0383666 \simeq d_2(\mu, \Gamma_{2,3}^{del})^2.$$

Let us finally check that  $\inf_{\nu \in \mathcal{P}_{\geq\mu}(\mathbb{R}, 3)} \int_{\mathbb{R}} y^2\nu(dy) \geq \inf_{u \in (0, 1)} \int_{\mathbb{R}} y^2\nu_u(dy)$  and that  $\nu_u \in \mathcal{P}_{\geq\mu}(\mathbb{R}, 3)$  for each  $u \in (0, 1)$ . First note that for each  $u \in (0, 1)$ , the mean  $2/3$  of  $\nu_u$  is equal to the one of  $\mu$ . According to the characterization of the convex order in terms of potential functions, we have  $\nu \geq_{cvx} \mu$  iff  $\int_{\mathbb{R}} x\nu(dx) = \frac{2}{3}$  and

$$\forall x \in \mathbb{R}, \varphi_\nu(x) := \int_{-\infty}^x \nu((-\infty, y])dy \geq \varphi_\mu(x) = \mathbf{1}_{[0,1]}(x) \frac{x^3}{3} + \mathbf{1}_{\{x>1\}} \left(x - \frac{2}{3}\right). \quad (2.18)$$

If  $\nu$  weights at most three points, then the convex function  $\varphi_\nu$  is piecewise affine with at most three changes of slope, the left-most slope being equal to 0 and the right-most equal to 1. Therefore if  $\nu \in \mathcal{P}_{\geq \mu}(\mathbb{R}, 3)$  then  $\nu(\{a\}) > 0$  for some  $a \leq 0$  and  $\nu(\{b\}) > 0$  for some  $b \geq 1$ . If, moreover,  $\nu(\{\sqrt{u}\}) > 0$  for some  $u \in (0, 1)$ , then since the slope of  $\varphi_\nu$  is constant on  $(0, \sqrt{u})$  and on  $(\sqrt{u}, 1)$ , for all  $x \in \mathbb{R}$ ,

$$\varphi_\nu(x) \geq 1_{(0, \sqrt{u})}(x) \frac{u^{3/2}}{3} \times \frac{x}{\sqrt{u}} + 1_{(\sqrt{u}, 1)}(x) \left( \frac{u^{3/2}}{3} \times \frac{1-x}{1-\sqrt{u}} + \frac{1}{3} \times \frac{x-\sqrt{u}}{1-\sqrt{u}} \right) + 1_{\{x > 1\}} \left( x - \frac{2}{3} \right) = \varphi_{\nu_u}(x),$$

so that, by convexity of the square function,  $\int_{\mathbb{R}} y^2 \nu(dy) \geq \int_{\mathbb{R}} y^2 \nu_u(dy)$ . If, on the other hand,  $\nu((0, 1)) = 0$ , then  $\varphi_\nu$  has a constant slope on  $(0, 1)$  and we even have

$$\int_{\mathbb{R}} y^2 \nu(dy) \geq \sup_{u \in (0, 1)} \int_{\mathbb{R}} y^2 \nu_u(dy).$$

Therefore  $\inf_{\nu \in \mathcal{P}_{\geq \mu}(\mathbb{R}, 3)} \int_{\mathbb{R}} y^2 \nu(dy) \geq \inf_{u \in (0, 1)} \int_{\mathbb{R}} y^2 \nu_u(dy)$ . To conclude that this inequality is an equality, it is enough to check that  $\nu_u \in \mathcal{P}_{\geq \mu}(\mathbb{R}, 3)$  for each  $u \in (0, 1)$ . This follows from the inequality  $\varphi_{\nu_u}(x) \geq \varphi_\mu(x)$  valid for all  $x \in \mathbb{R}$  and all  $u \in [0, 1]$  since the graph of the convex function  $\varphi_\mu$  is under its chords.

We finally recall the main result on convergence rate of dual quantization for bounded random vectors established in [46].

**Theorem 2.2 (Pierce Lemma for dual quantization)** *Let  $p \geq 1$  and  $\eta > 0$ . For every dimension  $d \geq 1$ , there exists a real constant  $\tilde{C}_{d, \eta, p}^{del} > 0$  such that, for every random vector  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ ,  $L^\infty(\mathbb{P})$ -bounded,*

$$d_{p, N}(X) \leq \tilde{C}_{d, \eta, p}^{del} N^{-\frac{1}{d}} \sigma_{p+\eta}(X) \quad (2.19)$$

where, for every  $r > 0$ ,  $\sigma_r(X) = \inf_{a \in \mathbb{R}^d} \|X - a\|_r < +\infty$ .

**Remark.** Note that this claim and the one in Theorem 1.3 remain true if the support of  $\mathbb{P}_X$  does not span  $\mathbb{R}^d$  as an affine space, but  $A_\mu$  with dimension  $d' < d$ . However, if such is the case, then (1.12) and (2.19) hold with factor  $N^{-1/d'}$  so that  $N^{-1/d}$  is suboptimal.

### 3 Quantized approximations of martingale couplings

Let  $\nu \in \mathcal{P}(\mathbb{R}^d)$  be compactly supported and  $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ . For  $K, N \geq 1$ , let  $\check{\nu}^K$  be an  $L^p$ -optimal dual  $K$ -quantization of  $\nu$  with grid  $\Gamma_{p, K}^{del}$  and  $\hat{\mu}^N$  be a quadratic optimal primal  $N$ -quantization of  $\mu$  with grid  $\Gamma_{2, N}$ . The two quantization errors correspond to martingale quasi-metrics (in comparison to Wasserstein metrics (0.2), only martingale couplings are considered in the minimization defining (0.3)) between  $\nu$  (resp.  $\hat{\mu}_N$ ) and  $\check{\nu}^K$  (resp.  $\mu$ ) :

$$d_{p, K}^p(\nu) = M_p^p(\nu, \check{\nu}^K) \text{ and } e_{2, N}^2(\mu) = M_2^2(\hat{\mu}^N, \mu). \quad (3.20)$$

In contrast with dual quantization where the martingale quasi-metric appears from the very beginning of the construction, the optimization in primal quantization relies on Wasserstein metrics.



But in the quadratic  $p = 2$  case, the stationarity property (1.8) satisfied at optimality implies that  $W_2(\hat{\mu}_N, \mu) = M_2(\hat{\mu}_N, \mu)$ .

Let  $U \sim \mathcal{U}[0, 1]$  and  $q_y(d\check{y})$  denote the law of  $\text{Proj}_{\Gamma_{p,K}^{del}}(y, U)$  when  $y \in \text{Conv}(\Gamma_{p,K}^{del})$  and  $\delta_y(d\check{y})$  otherwise. When  $\mu \leq_{cvx} \nu$ , we are now going to exploit (3.20) to approximate any  $\pi \in \mathcal{M}(\mu, \nu)$  by

$$\bar{\pi}^{N,K}(d\hat{x}, d\check{y}) = \int_{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d} \delta_{\text{Proj}_{\Gamma_{2,N}}(x)}(d\hat{x}) \pi(dx, dy) q_y(d\check{y}). \quad (3.21)$$

For  $(X, Y) \sim \pi$  independent from the random variable  $U$  uniformly distributed on  $[0, 1]$ , the random vector  $(\text{Proj}_{\Gamma_{2,N}}(X), X, Y, \text{Proj}_{\Gamma_{p,K}^{del}}(Y, U))$  is distributed according to  $\delta_{\text{Proj}_{\Gamma_{2,N}}(x)}(d\hat{x}) \pi(dx, dy) q_y(d\check{y})$  and therefore  $(\text{Proj}_{\Gamma_{2,N}}(X), \text{Proj}_{\Gamma_{p,K}^{del}}(Y, U))$  is distributed according to  $\bar{\pi}^{N,K}$ . Therefore, the first marginal of  $\bar{\pi}^{N,K}$  is the distribution  $\hat{\mu}^N$  of  $\text{Proj}_{\Gamma_{2,N}}(X)$  and its second marginal is the law  $\check{\nu}^K$  of  $\text{Proj}_{\Gamma_{p,K}^{del}}(Y, U)$ . Moreover, using that  $U$  is independent from  $(X, Y)$  and (2.14) for the second equality, then that  $\pi$  is a martingale coupling for the fourth equality and the stationarity property (1.8) for the last one, we obtain

$$\begin{aligned} \mathbb{E}[\text{Proj}_{\Gamma_{p,K}^{del}}(Y, U) | \text{Proj}_{\Gamma_{2,N}}(X)] &= \mathbb{E}[\mathbb{E}[\text{Proj}_{\Gamma_{p,K}^{del}}(Y, U) | (X, Y)] | \text{Proj}_{\Gamma_{2,N}}(X)] = \mathbb{E}[Y | \text{Proj}_{\Gamma_{2,N}}(X)] \\ &= \mathbb{E}[\mathbb{E}[Y | X] | \text{Proj}_{\Gamma_{2,N}}(X)] = \mathbb{E}[X | \text{Proj}_{\Gamma_{2,N}}(X)] = \text{Proj}_{\Gamma_{2,N}}(X). \end{aligned}$$

so that  $\bar{\pi}^{N,K} \in \mathcal{M}(\hat{\mu}^N, \check{\nu}^K)$ .

**Theorem 3.1** *Let  $p \geq 1$ ,  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  be such that  $\mu \leq_{cvx} \nu$  with  $\nu$  compactly supported and for  $N, K \geq 1$ , let  $\hat{\mu}^N$  be a quadratic optimal primal  $N$ -quantization of  $\mu$  and  $\check{\nu}^K$  be an  $L^p$ -optimal dual  $K$ -quantization of  $\nu$ . Then, for each  $\pi \in \mathcal{M}(\mu, \nu)$ , the couplings  $\bar{\pi}^{N,K} \in \mathcal{M}(\hat{\mu}^N, \check{\nu}^K)$  defined in (3.21) are such that*

$$W_p^p(\bar{\pi}^{N,K}, \pi) \leq \begin{cases} e_{2,N}^p(\mu) + d_{p,K}^p(\nu) & \text{if } p \leq 2 \\ CN^{-p/d} + 2^{\frac{(p-2)}{2}} d_{p,K}^p(\nu) & \text{for } C < \infty \text{ not depending on } N, K \text{ if } 2 < p < 2 + d \end{cases}.$$

Moreover, when  $p \geq 2$ ,

$$W_2^2(\bar{\pi}^{N,K}, \pi) \leq e_{2,N}^2(\mu) + d_{p,K}^2(\nu).$$

Last, for any  $p \geq 1$ ,

$$\limsup_{N \rightarrow \infty} AW_p(\bar{\pi}^{N,K}, \pi) \leq d_{p,K}(\nu).$$

According to Theorems 1.3 and 2.2,  $\sup_{N \geq 1} N^{1/d} e_{2,N}(\mu) < \infty$  and  $\sup_{K \geq 1} K^{1/d} d_{p,K}(\nu) < \infty$ .

**Proof.** We have

$$\begin{aligned} W_p^p(\bar{\pi}^{N,K}, \pi) &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} (|\hat{x} - x|^2 + |y - \check{y}|^2)^{p/2} \delta_{\text{Proj}_{\Gamma_{2,N}}(x)}(d\hat{x}) \pi(dx, dy) q_y(d\check{y}) \\ &\leq 2^{\frac{(p-2)^+}{2}} \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} (|\hat{x} - x|^p + |y - \check{y}|^p) \delta_{\text{Proj}_{\Gamma_{2,N}}(x)}(d\hat{x}) \pi(dx, dy) q_y(d\check{y}) \\ &= 2^{\frac{(p-2)^+}{2}} \int_{\mathbb{R}^d} |\text{Proj}_{\Gamma_{2,N}}(x) - x|^p \mu(dx) + 2^{\frac{(p-2)^+}{2}} \mathbb{E} \left[ \left| Y - \text{Proj}_{\Gamma_{p,K}^{del}}(Y, U) \right|^p \right] \\ &= 2^{\frac{(p-2)^+}{2}} \int_{\mathbb{R}^d} |\text{Proj}_{\Gamma_{2,N}}(x) - x|^p \mu(dx) + 2^{\frac{(p-2)^+}{2}} d_{p,K}^p(\nu). \end{aligned}$$

When  $p \leq 2$ ,  $\int_{\mathbb{R}^d} |\text{Proj}_{\Gamma_{2,N}}(x) - x|^p \mu(dx) \leq \left( \int_{\mathbb{R}^d} |\text{Proj}_{\Gamma_{2,N}}(x) - x|^2 \mu(dx) \right)^{p/2} = e_{2,N}^p(\mu)$ . On the other hand, when  $2 < p < 2 + d$ , since  $\mu$  is compactly supported, we may apply the  $L^2 - L^p$ -distortion mismatch Theorem 4.3 [38] to obtain that  $\sup_{N \geq 1} N^{p/d} \int_{\mathbb{R}^d} |\text{Proj}_{\Gamma_{2,N}}(x) - x|^p \mu(dx) < \infty$ , which completes the proof of the first inequality. In a similar way, when  $p \geq 2$

$$\begin{aligned} W_2^2(\bar{\pi}^{N,K}, \pi) &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} (|\hat{x} - x|^2 + |y - \check{y}|^2) \delta_{\text{Proj}_{\Gamma_{2,N}}(x)}(d\hat{x}) \mu(dx) \pi_x(dy) q_y(d\check{y}) \\ &= e_{2,N}^2(\mu) + \mathbb{E} \left[ \left| Y - \text{Proj}_{\Gamma_{p,K}^{\text{del}}}(Y, U) \right|^2 \right] \leq e_{2,N}^2(\mu) + \mathbb{E} \left[ \left| Y - \text{Proj}_{\Gamma_{p,K}^{\text{del}}}(Y, U) \right|^p \right]^{2/p} \\ &= e_{2,N}^2(\mu) + d_{p,N}^2(\nu). \end{aligned}$$

Let now

$$\check{\pi}^K(dx, d\check{y}) = \int_{y \in \mathbb{R}^d} \pi(dx, dy) q_y(d\check{y}) = \int_{y \in \mathbb{R}^d} \mu(dx) \pi_x(dy) q_y(d\check{y}).$$

We have  $\check{\pi}^K \in \mathcal{M}(\mu, \check{\nu}^K)$ . Using the identity coupling  $\mu(dx) \delta_x(d\tilde{x})$  between  $\mu$  and  $\mu$  in the definition of the adapted Wasserstein distance then the coupling  $\pi_x(dy) q_y(d\check{y})$  between  $\check{\pi}_x^K(d\check{y})$  and  $\pi_x(dy)$  in the definition of the usual Wasserstein distance, we obtain that

$$\begin{aligned} AW_p^p(\check{\pi}^K, \pi) &\leq \int_{\mathbb{R}^d} W_p^p(\check{\pi}_x^K, \pi_x) \mu(dx) \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_0^1 |\text{Proj}_{\Gamma_{p,K}^{\text{del}}}(y, u) - y|^p du \pi_x(dy) \mu(dx) \\ &\leq \int_{\mathbb{R}^d} \int_0^1 |y - \text{Proj}_{\Gamma_{p,K}^{\text{del}}}(y, u)|^p du \nu(dy) = \mathbb{E} |Y - \text{Proj}_{\Gamma_{p,K}^{\text{del}}}(Y, U)|^p = d_{p,K}^p(\nu). \end{aligned}$$

With the triangle inequality, we deduce that to check the last statement in the theorem, it is enough to prove that  $\lim_{N \rightarrow \infty} AW_p(\bar{\pi}^{N,K}, \check{\pi}^K) = 0$ .

To do so, we denote by  $(x_i)_{1 \leq i \leq |\Gamma_{2,N}|}$  the  $|\Gamma_{2,N}| \leq N$  points in  $\Gamma_{2,N}$  and by  $(y_j)_{1 \leq j \leq |\Gamma_{p,K}^{\text{del}}|}$  the  $|\Gamma_{p,K}^{\text{del}}| \leq K$  points in  $\Gamma_{p,K}^{\text{del}}$  (as soon as the support of  $\mu$  (resp.  $\nu$ ) is not restricted to less than  $N$  (resp.  $K$ ) points,  $|\Gamma_{2,N}| = N$  (resp.  $|\Gamma_{p,K}^{\text{del}}| = K$ )). Let  $C_i = \{x \in \mathbb{R}^d : |x - x_i| < \min_{1 \leq k \leq |\Gamma_{2,N}|, k \neq i} |x - x_k|\}$ ,  $i \in \{1, \dots, |\Gamma_{2,N}|\}$ , denote the open Voronoi cells induced by  $\Gamma_{2,N}$ . Using (1.7) for the equality, we have

$$\begin{aligned} AW_p^p(\bar{\pi}^{N,K}, \check{\pi}^K) &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} (|\hat{x} - x|^p + W_p^p(\bar{\pi}_{\hat{x}}^{N,K}, \check{\pi}_x^K)) \delta_{\text{Proj}_{\Gamma_{2,N}}(x)}(d\hat{x}) \mu(dx) \\ &= \sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} (|x_i - x|^p + W_p^p(\bar{\pi}_{x_i}^{N,K}, \check{\pi}_x^K)) \mu(dx). \end{aligned} \quad (3.22)$$

Since  $\check{\nu}^K(\Gamma_{p,K}^{\text{del}}) = 1$ ,  $\mu(dx)$  a.e.  $\check{\pi}_x^K(d\check{y}) = \sum_{j=1}^{|\Gamma_{p,K}^{\text{del}}|} q_j(x) \delta_{y_j}(d\check{y})$  for some measurable functions  $q_j$  with values in  $[0, 1]$  and such that  $\sum_{j=1}^{|\Gamma_{p,K}^{\text{del}}|} q_j = 1$ . Moreover, for  $i \in \{1, \dots, |\Gamma_{2,N}|\}$ ,

$$\bar{\pi}_{x_i}^{N,K}(d\check{y}) = \frac{1}{\mu(C_i)} \sum_{j=1}^{|\Gamma_{p,K}^{\text{del}}|} \int_{C_i} q_j(\xi) \mu(d\xi) \delta_{y_j}(d\check{y}).$$

As a consequence, we have

$$\mathrm{TV}(\bar{\pi}_{x_i}^{N,K}, \tilde{\pi}_x^K) = \sum_{j=1}^{|\Gamma_{p,K}^{del}|} \left| q_j(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j(\xi) \mu(d\xi) \right|.$$

With Theorem 6.15 p.103 [51], we deduce that

$$W_p^p(\bar{\pi}_{x_i}^{N,K}, \tilde{\pi}_x^K) = 2^{p-1} \max_{1 \leq j \leq |\Gamma_{p,K}^{del}|} |y_j|^p \mathrm{TV}(\hat{\pi}_{x_i}^{N,K}, \hat{\pi}_x^K) \leq 2^{p-1} \max_{1 \leq j \leq |\Gamma_{p,K}^{del}|} |y_j|^p \sum_{j=1}^{|\Gamma_{p,K}^{del}|} \left| q_j(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j(\xi) \mu(d\xi) \right|.$$

Plugging this inequality in (3.22), we deduce that

$$AW_p^p(\bar{\pi}^{N,K}, \tilde{\pi}^K) \leq \sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} |x_i - x|^p \mu(dx) + 2^{p-1} \max_{1 \leq j \leq |\Gamma_{p,K}^{del}|} |y_j|^p \sum_{j=1}^{|\Gamma_{p,K}^{del}|} \sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} \left| q_j(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j(\xi) \mu(d\xi) \right| \mu(dx). \quad (3.23)$$

Using Jensen's inequality in the case  $p \in [1, 2]$ , the  $L^2 - L^p$ - distortion mismatch Theorem 4.3 [38] when  $2 < p < 2 + d$  and Hölder's inequality when  $p \geq 2 + d$ , we get

$$\sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} |x_i - x|^p \mu(dx) = \int_{\mathbb{R}^d} |\mathrm{Proj}_{\Gamma_{2,N}}(x) - x|^p \mu(dx) \leq \begin{cases} e_{2,N}^p(\mu) & \text{if } p \in [1, 2] \\ CN^{-p/d} & \text{if } 2 < p < 2 + d \\ 2^{(p-3) \vee 0} \max_{1 \leq j \leq |\Gamma_{p,K}^{del}|} |y_j|^{p-2} e_{2,N}^2(\mu) & \text{if } p \geq 2 + d \end{cases},$$

where the constant  $C$  does not depend on  $N$ . With Theorem 1.3, we deduce that the first term in the right-hand side of (3.23) goes to 0 as  $N \rightarrow \infty$ . Since  $\int_{\mathbb{R}^d} |q_j(x)| \mu(dx) \leq 1$ , by Theorem 3.14 p.69 [49], there exists a sequence  $(q_j^n)_{n \in \mathbb{N}}$  of continuous and compactly supported functions on  $\mathbb{R}^d$  such that  $\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} |q_j^n(x) - q_j(x)| \mu(dx) = 0$ . Since  $q_j$  takes its values in the interval  $[0, 1]$ , we suppose that so do the functions  $q_j^n$  up to replacing them by  $0 \vee q_j^n \wedge 1$  which affects neither the continuity and compact support property nor the convergence. We have

$$\begin{aligned} & \left| \sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} \left| q_j(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j(\xi) \mu(d\xi) \right| \mu(dx) - \sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} \left| q_j^n(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j^n(\xi) \mu(d\xi) \right| \mu(dx) \right| \\ & \leq \sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} \left| q_j(x) - q_j^n(x) - \frac{1}{\mu(C_i)} \int_{C_i} (q_j(\xi) - q_j^n(\xi)) \mu(d\xi) \right| \mu(dx) \\ & \leq \sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} |q_j(x) - q_j^n(x)| \mu(dx) + \sum_{i=1}^{|\Gamma_{2,N}|} \left| \int_{C_i} (q_j(\xi) - q_j^n(\xi)) \mu(d\xi) \right| \leq 2 \int_{\mathbb{R}^d} |q_j(x) - q_j^n(x)| \mu(dx), \end{aligned}$$

where the right-hand side goes to 0 as  $n \rightarrow \infty$ . We deduce that to prove that the second term in the right-hand side of (3.23) goes to 0 as  $N \rightarrow \infty$ , it is enough to check that so does  $\sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} \left| q_j(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j^n(\xi) \mu(d\xi) \right| \mu(dx)$  for any fixed  $n \in \mathbb{N}$ . For  $X \sim \mu$  and  $\hat{X}^N = \mathrm{Proj}_{\Gamma_{2,N}}(X)$ ,

$$\begin{aligned} & \sum_{i=1}^{|\Gamma_{2,N}|} \int_{C_i} \left| q_j^n(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j^n(\xi) \mu(d\xi) \right| \mu(dx) = \mathbb{E} \left| q_j^n(X) - \mathbb{E}[q_j^n(X) | \hat{X}^N] \right| \\ & \leq \mathbb{E}^{1/2} \left[ \left( q_j^n(X) - \mathbb{E}[q_j^n(X) | \hat{X}^N] \right)^2 \right] \leq \mathbb{E}^{1/2} \left[ \left( q_j^n(X) - q_j^n(\hat{X}^N) \right)^2 \right] \end{aligned}$$

where, for the last inequality, we used that the conditional expectation given  $\hat{X}^N$  is the best quadratic approximation of a random variable by a measurable function of  $\hat{X}^N$ . Let  $\varepsilon > 0$ . Since  $q_j^n$  is continuous and compactly supported, this function is uniformly continuous. Since it takes its values in the interval  $[0, 1]$ , we deduce that there exists  $\eta > 0$  such that for all  $x, y \in \mathbb{R}^d$ ,  $|q_j^n(x) - q_j^n(y)| \leq \varepsilon 1_{\{|x-y| \leq \eta\}} + 1_{\{|x-y| > \eta\}}$ . Therefore

$$\mathbb{E} \left[ \left( q_j^n(X) - q_j^n(\hat{X}^N) \right)^2 \right] \leq \varepsilon^2 + \mathbb{P}(|X - \hat{X}^N| \geq \eta) \leq \varepsilon^2 + \frac{\mathbb{E}[|X - \hat{X}^N|^2]}{\eta^2} = \varepsilon^2 + \frac{e_2(\mu, N)^2}{\eta^2}.$$

With Theorem 1.3, we deduce that the left-hand side goes to 0 as  $N \rightarrow \infty$  and conclude that so does  $AW_p(\bar{\pi}^{N,K}, \bar{\pi}^K)$ .  $\square$

## 4 Application to weak martingale optimal transport problems

We endow  $\mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d) = \{(\mu, \nu) : \mu, \nu \in \mathcal{P}_1(\mathbb{R}^d) \text{ and } \mu \leq_{cvx} \nu\}$  with the metric  $W_1(\mu, \tilde{\mu}) + W_1(\nu, \tilde{\nu})$  between  $(\mu, \nu)$  and  $(\tilde{\mu}, \tilde{\nu})$  and  $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$  with the metric obtained as the sum of the Euclidean distance on  $\mathbb{R}^d$  and the Wasserstein distance  $W_1$  on  $\mathcal{P}_1(\mathbb{R}^d)$ . For a cost function  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  Borel measurable, the Weak Martingale Optimal Transport problem introduced in [6] consists in computing for  $(\mu, \nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$

$$V(\mu, \nu) = \inf_{\pi \in \mathcal{M}(\mu, \nu)} \int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx) \quad (4.24)$$

and the minimal couplings  $\pi \in \mathcal{M}(\mu, \nu)$ . For the choice

$$\tilde{C}(x, \eta) = \begin{cases} C(x, \eta) & \text{if } \int_{\mathbb{R}^d} y \eta(dy) = x \\ +\infty & \text{otherwise} \end{cases}, \quad (4.25)$$

it can be seen as a particular case of the Weak Optimal Transport problem

$$\tilde{V}(\mu, \nu) = \inf_{\pi \in \mathcal{P}(\mu, \nu)} \int_{\mathbb{R}^d} \tilde{C}(x, \pi_x) \mu(dx) \text{ for } (\mu, \nu) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \quad (4.26)$$

introduced by Gozlan, Roberto, Samson and Tetali in [23] and studied by Backhoff-Veraguas, Beiglöck and Pammer in [5]. Indeed, when  $(\mu, \nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$  then for each  $\pi \in \mathcal{P}(\mu, \nu) \setminus \mathcal{M}(\mu, \nu)$ ,  $\int_{\mathbb{R}^d} \tilde{C}(x, \pi_x) \mu(dx) = +\infty$  and for each  $\pi \in \mathcal{M}(\mu, \nu)$ ,  $\int_{\mathbb{R}^d} \tilde{C}(x, \pi_x) \mu(dx) = \int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx)$ , which implies that  $\tilde{V}(\mu, \nu) = V(\mu, \nu)$ .

The martingale optimal transport problem corresponds to the particular case of the WMOT problem when the cost function is linear in the measure component :  $C(x, \eta) = \int_{\mathbb{R}^d} c(x, y) \eta(dy)$  for a Borel measurable function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  with at most linear growth in its second variable.

The existence of minimal couplings in the WMOT problem (4.24) and the lower semi-continuity of the value function  $V$  are deduced from Theorem 2.6 and Proposition 5.8 (b) [10].

**Proposition 4.1** *Assume that  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  is lower semi-continuous, convex in the measure argument (for all  $x \in \mathbb{R}^d$ ,  $\mathcal{P}_1(\mathbb{R}^d) \ni \eta \mapsto C(x, \eta)$  is convex) and such that*

$$\sup_{(x, \eta) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} \frac{|C(x, \eta)|}{1 + |x| + \int_{\mathbb{R}^d} |y| \eta(dy)} < +\infty.$$

*Then for each  $(\mu, \nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$ , there exists  $\pi^* \in \mathcal{M}(\mu, \nu)$ , unique if  $C$  is strictly convex in the measure argument, such that  $V(\mu, \nu) = \int_{\mathbb{R}^d} C(x, \pi_x^*) \mu(dx)$  and  $(\mu, \nu) \mapsto V(\mu, \nu)$  is lower semi-continuous on  $\mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$ .*

Theorem 2.6 [10] also ensures convergence of the optimal couplings under convergence of the value function, a property which holds in dimension  $d = 1$ .

**Proposition 4.2** *Assume that  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  is lower semi-continuous, convex in the measure argument and such that*

$$\sup_{(x, \eta) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} \frac{|C(x, \eta)|}{1 + |x| + \int_{\mathbb{R}^d} |y| \eta(dy)} < +\infty.$$

*Let for each  $k \in \mathbb{N}$ ,  $(\mu_k, \nu_k) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$  and  $\pi_k^*$  be an optimal coupling for  $V(\mu_k, \nu_k)$ , the existence of which is a consequence of Theorem 4.1. If  $((\mu_k, \nu_k))_{k \in \mathbb{N}}$  converges to  $(\mu, \nu)$  in  $\mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$  as  $k \rightarrow \infty$  and  $V(\mu, \nu) = \lim_{k \rightarrow \infty} V(\mu_k, \nu_k)$ , then all the accumulation points of  $(\pi_k^*)_{k \in \mathbb{N}}$  for the weak convergence topology are minimizers for  $V(\mu, \nu)$ . If  $C$  is moreover strictly convex in the measure argument, then the sequence  $(\pi_k^*)_{k \in \mathbb{N}}$  converges in  $AW_1$  to the unique optimal coupling  $\pi^*$  between  $\mu$  and  $\nu$ .*

*If  $d = 1$  and either  $C$  is continuous or  $C$  is continuous in its second argument and for each Borel subset  $A$  of  $\mathbb{R}^d$ ,  $(\mu_k(A))_{k \in \mathbb{N}}$  converges to  $\mu(A)$  as  $k \rightarrow \infty$ , then  $V(\mu, \nu) = \lim_{k \rightarrow \infty} V(\mu_k, \nu_k)$ .*

When  $(\mu_k, \nu_k) = (\hat{\mu}^{N_k}, \check{\nu}^{K_k})$  with  $\hat{\mu}^{N_k}$  a quadratic optimal primal  $N_k$ -quantization of  $\mu$  and  $\check{\nu}^{K_k}$  an  $L^p$ -optimal dual  $K_k$ -quantization of  $\nu$ , then Theorem 3.1 ensures that it is possible to approximate in  $AW_1$  distance any optimal martingale coupling  $\pi^*$  between  $\mu$  and  $\nu$  by martingale couplings between  $\hat{\mu}^{N_k}$  and  $\check{\nu}^{K_k}$  and we deduce the upper-semicontinuity of the value function along this sequence, whatever the dimension  $d$ .

**Lemma 4.3** *Let  $p \geq 1$ ,  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  be such that  $\mu \leq_{cvx} \nu$  with  $\nu$  compactly supported and for  $N, K \geq 1$ ,  $\hat{\mu}^N$  be a quadratic optimal primal  $N$ -quantization of  $\mu$  and  $\check{\nu}^K$  an  $L^p$ -optimal dual  $K$ -quantization of  $\nu$ . If  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  is continuous, then  $V(\mu, \nu)$  is finite. If moreover,  $x \mapsto \sup_{\eta \in \mathcal{P}_1(\mathbb{R}^d)} \frac{|C(x, \eta)|}{1 + \int_{\mathbb{R}^d} |y| \eta(dy)}$  is locally bounded on  $\mathbb{R}^d$ , then*

$$\limsup_{N, K \rightarrow \infty} V(\hat{\mu}^N, \check{\nu}^K) \leq V(\mu, \nu).$$

*If  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  is Lipschitz continuous with constant  $\text{Lip}(C)$ , then*

$$\limsup_{N \rightarrow \infty} V(\hat{\mu}^N, \check{\nu}^K) \leq V(\mu, \nu) + \text{Lip}(C) d_{p, K}(\nu).$$

With Propositions 4.1 and 4.2, we easily deduce the following corollary.

**Corollary 4.4** Let  $p \geq 1$ ,  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  be such that  $\mu \leq_{cvx} \nu$  with  $\nu$  compactly supported and for  $N, K \geq 1$ ,  $\hat{\mu}^N$  be a quadratic optimal primal  $N$ -quantization of  $\mu$  and  $\check{\nu}^K$  an  $L^p$ -optimal dual  $K$ -quantization of  $\nu$ . If  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  is continuous, convex in the measure argument and such that

$$\sup_{(x, \eta) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} \frac{|C(x, \eta)|}{1 + |x| + \int_{\mathbb{R}^d} |y| \eta(dy)} < +\infty,$$

then  $\lim_{N, K \rightarrow \infty} V(\hat{\mu}^N, \check{\nu}^K) = V(\mu, \nu)$ . Moreover, all accumulation points as  $N, K \rightarrow \infty$  of sequences  $(\pi_{N, K}^*)_{N, K}$  of minimizers for  $V(\hat{\mu}^N, \check{\nu}^K)$  are minimizers for  $V(\mu, \nu)$ . If  $C$  is also strictly convex in the measure argument, then  $\lim_{N, K \rightarrow \infty} AW_1(\pi_{N, K}^*, \pi^*) = 0$ , where  $\pi^*$  is the unique optimal coupling between  $\mu$  and  $\nu$ .

**Proof of Lemma 4.3.** We first suppose that  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  is continuous. Let  $\bar{B}$  be some closed ball centered at the origin with sufficiently large radius so that  $\nu(\bar{B}) = 1$  and  $\mathcal{P}_{\bar{B}}(\mathbb{R}^d) = \{\eta \in \mathcal{P}(\mathbb{R}^d) : \eta(\bar{B}) = 1\}$ . The set  $\mathcal{P}_{\bar{B}}(\mathbb{R}^d)$  is a compact subset of  $\mathcal{P}_1(\mathbb{R}^d)$  and therefore the continuous cost function  $C$  is bounded on the compact subset  $\bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d)$  of  $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ . Since  $\hat{\mu}^N \leq_{cvx} \mu \leq_{cvx} \nu$ ,  $\hat{\mu}^N(\bar{B}) = \mu(\bar{B}) = 1$ . Moreover for  $\pi \in \mathcal{P}(\mu, \nu)$ ,  $\int_{\mathbb{R}^d} \pi_x(\bar{B}) \mu(dx) = \nu(\bar{B}) = 1$  so that  $(x, \pi_x) \in \bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d)$   $\mu(dx)$  a.e.. Therefore

$$\inf_{(x, \eta) \in \bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d)} C(x, \eta) \leq V(\mu, \nu) \leq \sup_{(x, \eta) \in \bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d)} C(x, \eta)$$

and  $V(\mu, \nu)$  is finite. Let  $\varepsilon > 0$ . By the continuity and the growth assumption satisfied by the cost function  $C$  and the compactness of  $\bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d)$ ,

$$\begin{aligned} \exists \alpha > 0, \forall (x, \eta, \tilde{x}, \tilde{\eta}) \in \bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d) \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \text{ s.t. } |x - \tilde{x}| + W_1(\eta, \tilde{\eta}) \leq \alpha, |C(x, \eta) - C(\tilde{x}, \tilde{\eta})| \leq \varepsilon \\ \text{and } \forall (x, \eta, \tilde{x}, \tilde{\eta}) \in \bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d) \times \bar{B} \times \mathcal{P}_1(\mathbb{R}^d), |C(x, \eta)| + |C(\tilde{x}, \tilde{\eta})| \leq \frac{1}{\alpha} \left( 1 + \int_{\mathbb{R}^d} |y| \tilde{\eta}(dy) \right). \end{aligned}$$

Let  $\pi \in \mathcal{M}(\mu, \nu)$  be such that  $\int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx) \leq V(\mu, \nu) + \varepsilon$ . By Theorem 3.1, there exists  $\bar{\pi}^{N, K} \in \mathcal{M}(\hat{\mu}^N, \check{\nu}^K)$  such that  $\limsup_{N \rightarrow \infty} AW_1(\bar{\pi}^{N, K}, \pi) \leq d_{p, K}(\nu)$  and  $\lim_{N, K \rightarrow \infty} AW_1(\bar{\pi}^{N, K}, \pi) = 0$ . Let  $m^{N, K} \in \mathcal{P}(\mu, \hat{\mu}^N)$  be an optimal coupling for  $AW_1(\pi, \bar{\pi}^{N, K})$ . We have

$$\begin{aligned} \left| \int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx) - \int_{\mathbb{R}^d} C(\tilde{x}, \bar{\pi}_{\tilde{x}}^{N, K}) \hat{\mu}^N(d\tilde{x}) \right| &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |C(x, \pi_x) - C(\tilde{x}, \bar{\pi}_{\tilde{x}}^{N, K})| m^{N, K}(dx, d\tilde{x}) \\ &\leq \varepsilon + \int_{\mathbb{R}^d \times \mathbb{R}^d} |C(x, \pi_x) - C(\tilde{x}, \bar{\pi}_{\tilde{x}}^{N, K})| \mathbf{1}_{\{|x - \tilde{x}| + W_1(\pi_x, \bar{\pi}_{\tilde{x}}^{N, K}) \geq \alpha\}} m^{N, K}(dx, d\tilde{x}) \\ &\leq \varepsilon + \frac{1}{\alpha} \int_{\mathbb{R}^d \times \mathbb{R}^d} \left( 1 + \int_{\mathbb{R}^d} |y| \bar{\pi}_{\tilde{x}}^{N, K}(dy) \right) \mathbf{1}_{\{|x - \tilde{x}| + W_1(\pi_x, \bar{\pi}_{\tilde{x}}^{N, K}) \geq \alpha\}} m^{N, K}(dx, d\tilde{x}). \end{aligned} \quad (4.27)$$

By Markov inequality,  $\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{1}_{\{|x - \tilde{x}| + W_1(\pi_x, \bar{\pi}_{\tilde{x}}^{N, K}) \geq \alpha\}} m^{N, K}(dx, d\tilde{x}) \leq \frac{AW_1(\pi, \bar{\pi}^{N, K})}{\alpha}$ . Moreover, since

$$\int_{(x, \tilde{x}) \in \mathbb{R}^d \times \mathbb{R}^d} \bar{\pi}_{\tilde{x}}^{N, K}(dy) m^{N, K}(dx, d\tilde{x}) = \int_{\tilde{x} \in \mathbb{R}^d} \bar{\pi}_{\tilde{x}}^{N, K}(dy) \hat{\mu}^N(d\tilde{x}) = \int_{\tilde{x} \in \mathbb{R}^d} \bar{\pi}^{N, K}(d\tilde{x}, dy) = \check{\nu}^K(dy),$$

there is a Markov kernel  $q_y(dx, d\tilde{x})$  such that  $\bar{\pi}_{\tilde{x}}^{N, K}(dy) m^{N, K}(dx, d\tilde{x}) = \check{\nu}^K(dy) q_y(dx, d\tilde{x})$  and

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \int_{\mathbb{R}^d} |y| \bar{\pi}_{\tilde{x}}^{N, K}(dy) \mathbf{1}_{\{|x - \tilde{x}| + W_1(\pi_x, \bar{\pi}_{\tilde{x}}^{N, K}) \geq \alpha\}} m^{N, K}(dx, d\tilde{x}) = \int_{\mathbb{R}^d} |y| \beta(y) \check{\nu}^K(dy)$$

where the function  $\beta(y) := \int_{\mathbb{R}^d \times \mathbb{R}^d} 1_{\{|x-\tilde{x}|+W_1(\pi_x, \bar{\pi}_x^{N,K}) \geq \alpha\}} q_y(dx, d\tilde{x})$  is  $[0, 1]$ -valued and such that

$$\int_{\mathbb{R}^d} \beta(y) \check{\nu}^K(dy) = \int_{\mathbb{R}^d \times \mathbb{R}^d} 1_{\{|x-\tilde{x}|+W_1(\pi_x, \bar{\pi}_x^{N,K}) \geq \alpha\}} m^{N,K}(dx, d\tilde{x}) \leq \frac{AW_1(\pi, \bar{\pi}^{N,K})}{\alpha}.$$

Since  $\lim_{N,K \rightarrow \infty} AW_1(\pi, \bar{\pi}^{N,K}) = 0 = \lim_{K \rightarrow \infty} W_1(\nu, \check{\nu}^K)$ , with Lemma.. BJMP, we deduce that the second term in the right-hand side of (4.27) tends to 0 as  $N, K \rightarrow \infty$ . Hence

$$\begin{aligned} \limsup_{N,K \rightarrow \infty} V(\hat{\mu}^N, \check{\nu}^K) &\leq \limsup_{N,K \rightarrow \infty} \int_{\mathbb{R}^d} C(\tilde{x}, \bar{\pi}_x^{N,K}) \hat{\mu}^N(d\tilde{x}) \\ &\leq \int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx) + \limsup_{N,K \rightarrow \infty} \left| \int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx) - \int_{\mathbb{R}^d} C(\tilde{x}, \bar{\pi}_x^{N,K}) \hat{\mu}^N(d\tilde{x}) \right| \\ &\leq V(\mu, \nu) + 2\varepsilon. \end{aligned}$$

Since  $\varepsilon$  is arbitrary, we conclude that  $\limsup_{N,K \rightarrow \infty} V(\hat{\mu}^N, \check{\nu}^K) \leq V(\mu, \nu)$ .

We now suppose that  $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$  is Lipschitz continuous with constant  $\text{Lip}(C)$ . We then have

$$\begin{aligned} \left| \int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx) - \int_{\mathbb{R}^d} C(\tilde{x}, \bar{\pi}_x^{N,K}) \hat{\mu}^N(d\tilde{x}) \right| &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \left| C(x, \pi_x) - C(\tilde{x}, \bar{\pi}_x^{N,K}) \right| m^{N,K}(dx, d\tilde{x}) \\ &\leq \text{Lip}(C) \int_{\mathbb{R}^d \times \mathbb{R}^d} \left( |x - \tilde{x}| + W_1(\pi_x, \bar{\pi}_x^{N,K}) \right) m(dx, d\tilde{x}) = \text{Lip}(C) AW_1(\pi, \bar{\pi}^{N,K}). \end{aligned}$$

Therefore

$$\begin{aligned} V(\hat{\mu}^N, \check{\nu}^K) &\leq \int_{\mathbb{R}^d} C(\tilde{x}, \bar{\pi}_x^{N,K}) \hat{\mu}^N(d\tilde{x}) \leq \int_{\mathbb{R}^d} C(x, \pi_x) \mu(dx) + \text{Lip}(C) AW_1(\pi, \bar{\pi}^{N,K}) \\ &\leq V(\mu, \nu) + \varepsilon + \text{Lip}(C) AW_1(\pi, \bar{\pi}^{N,K}). \end{aligned}$$

Since  $\varepsilon$  is arbitrary, we deduce that  $\limsup_{N \rightarrow \infty} V(\hat{\mu}^N, \check{\nu}^K) \leq V(\mu, \nu) + \text{Lip}(C) d_{p,K}(\nu)$ .  $\square$

## References

- [1] ALFONSI, A. CORBETTA J. AND JOURDAIN, B. (2019). Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds, *International Journal of Theoretical and Applied Finance*, **22**(3).
- [2] ALFONSI, A. CORBETTA J. AND JOURDAIN, B. (2020). Sampling of probability measures in the convex order by Wasserstein projection, *Annales de l'Institut Henri Poincaré B, Probabilités et Statistiques*, **56**(3):1706-1729.
- [3] BACKHOFF-VERAGUAS, BARTL, D., J. BEIGLBÖCK, M., EDER, M. (2020). Adapted Wasserstein distances and stability in mathematical finance. *Finance and Stoch.*, **24**:601-632.
- [4] BACKHOFF-VERAGUAS, BARTL, D., J. BEIGLBÖCK, M., EDER, M.(2020). All adapted topologies are equal. *Probab. Theory Relat. Fields*, **178** (3-4):1125–1172.
- [5] BACKHOFF-VERAGUAS, J. BEIGLBÖCK, M. AND PAMMER, G. (2019). Existence, duality and cyclical monotonicity for weak transport costs. *Calc. Var.*, **58**(6): paper 203.
- [6] BACKHOFF-VERAGUAS J. AND PAMMER, G. (2019). Stability of martingale optimal transport and weak optimal transport, arXiv:1904.04171.

- [7] BAKER, D. (2012). Martingales with specified marginals, PhD, Université Pierre et Marie Curie (Sorbonne-Université), Paris, France.
- [8] BEIGLBÖCK, M. COX, A. AND HUESMANN, M. (2017). Optimal transport and Skorokhod embedding. *Invent. Math.*, **208**(2):327-400.
- [9] BEIGLBÖCK, M. HENRY-LABORDÈRE, P. AND PENKNER, F. (2013). Model-independent bounds for option prices - a mass transport approach. *Finance Stoch.*, **17**(3):477–501.
- [10] BEIGLBÖCK, M. JOURDAIN, B. MARGHERITI, W. AND PAMMER, M. (2020). Monotonicity and stability of the Weak Martingale Optimal Transport problem, <https://cermics.enpc.fr/margherw/Documents/stabilityWMOT.pdf>.
- [11] BEIGLBÖCK, M. AND JUILLET, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, **44**(1):42–106.
- [12] BEIGLBÖCK, M. NUTZ, M. AND TOUZI, N. (2017). Complete duality for martingale optimal transport on the line. *Ann. Probab.*, **45**(5):3038-3074.
- [13] CAMPI, L. LAACHIR, I. AND MARTINI, C. (2017). Change of numeraire in the two-marginals martingale optimal transport problem. *Finance Stoch.*, **21**(2):471-486.
- [14] DE MARCH, H. (2018). Entropic approximation for multi-dimensional martingale optimal transport, arXiv 1812.11104.
- [15] DE MARCH, H. AND TOUZI, N. (2019). Irreducible convex paving for decomposition of multi-dimensional martingale transport plans. *Ann. Probab.*, **47**(3):1726–1774.
- [16] DOLINSKY, Y. AND SONER, H.M. Robust hedging and martingale optimal transport in continuous time. *Probab. Theory Relat. Fields*, **160**:391–427.
- [17] DU, Q. EMELIANENKO, M. AND JU, L. (2006). Convergence of the Lloyd algorithm for computing centroidal Voronoi tessellations, *SIAM Journal on Numerical Analysis*, **44**:102-119.
- [18] EMELIANENKO, M. JU, L. AND RAND, A. (2008). Nondegeneracy and Weak Global Convergence of the Lloyd Algorithm in  $\mathbb{R}^d$ , *SIAM Journal on Numerical Analysis*, **46**(3):1423-1441.
- [19] FADILI, A. AND PAGÈS, G. (2018). Functional convex order for stochastic differential equations and their approximation schemes. Technical report.
- [20] FORT, J.-C. AND PAGÈS, G. (2002). Asymptotic of optimal quantizers for some scalar distributions. *J. Comput. Appl. Math.*, **146**(2):253–275.
- [21] GALICHON, A. HENRY-LABORDÈRE, P. AND TOUZI, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, **24**(1):312–336.
- [22] GHOUSSEUB, N. KIM, Y.-H. AND LIM, T. (2019). Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.*, **47**(1):109-164.
- [23] GOZLAN, N. ROBERTO, C. SAMSON, P.-M. AND TETALI, P. (2017). Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, **273**(11):3327-3405.
- [24] GRAF, S. AND LUSCHGY, H. (2000). *Foundations of quantization for probability distributions*, LNM 1730, Springer, Berlin, 230p.
- [25] GUO, G. AND OBLÒJ, J. (2019). Computational Methods for Martingale Optimal Transport problems, *Ann. Appl. Probab.*, **29**(6):3311–3347.
- [26] HENRY-LABORDÈRE, P. (2019). (Martingale) optimal transport and anomaly detection with neural networks : a primal-dual algorithm, arXiv:1904.04546.
- [27] HENRY-LABORDÈRE, P. TAN X. AND TOUZI, N. (2016). An explicit martingale version of the one-dimensional Brenier’s theorem with full marginals constraint. *Stochastic Process. Appl.*, **126**(9):2800–2834.



- [28] HENRY-LABORDÈRE, P. AND TOUZI, N. (2016). An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, **20**(3):635–668.
- [29] HIRSCH, F PROFETA, C ROYNETTE, B. AND YOR, M (2011). Peacocks and associated martingales, with explicit constructions. Bocconi & Springer Series, 3. Springer, Milan; Bocconi University Press, Milan, 2011. xxxii+384 pp.
- [30] HOBSON, D. (1998). Robust hedging of the lookback option. *Finance Stoch.*, **2**:329–347.
- [31] HOBSON, D. AND KLIMMEK, M. (2015). Robust price bounds for the forward starting straddle. *Finance Stoch.*, **19**(1):189–214.
- [32] HOBSON, D. AND NEUBERGER, A. (2012). Robust bounds for forward start options. *Math. Finance*, **22**(1):31–56.
- [33] JOURDAIN, B. AND MARGHERITI, W. (2020). A new family of one-dimensional of martingale couplings. *Electron. J. Probab.*, **25** Paper No. 136.
- [34] KALLENBERG, O. (1997). Foundations of modern probability. Probability and its Applications. Springer-Verlag, New York.
- [35] KIEFFER, J. C. (1982). Exponential rate of convergence for Lloyd’s method. I. *IEEE Trans. Inform. Theory*, **28** (2):205–210.
- [36] MONTES, T. (2020). Quantization based schemes for pricing derivatives: a comparison, PhD, Sorbonne-Université.
- [37] PAGÈS, G. (2018). *Numerical Probability: an introduction with applications to Finance*, Springer-Verlag, xvi +579p.
- [38] PAGÈS, G. AND SAGNA, A. (2018). Improved error bounds for quantization based numerical schemes for BSDE and nonlinear filtering. *Stochastic Process. Appl.* **128**(3):847–883.
- [39] PAGÈS, G. AND SAGNA, A. (2020). Strong and weak error analysis of recursive quantization: a general approach with an application to jump diffusions. *IMA J. Num. Anal.*.
- [40] PAGÈS, G. AND SAGNA, A. (2015). Recursive marginal quantization of the Euler scheme of a diffusion process. *Appl. Math. Finance*, **22**(5):463–498.
- [41] PAGÈS, G. (2015). Introduction to optimal quantization for numerics, *ESAIM Proc. & Surveys*, **48**:29–79.
- [42] PAGÈS, G. (2016). Convex order for path-dependent derivatives: a dynamic programming approach. *Séminaire de Probabilités XLVIII*, C. Donati, A. Lejay, A. Rouault eds, LNM 2168, Springer, Cham, 33–96.
- [43] PAGÈS, G. AND WILBERTZ, B. (2012). Dual Quantization for random walks with application to credit derivatives, *J. Comp. Finance*, **16**(2):33–60.
- [44] PAGÈS, G. AND WILBERTZ, B. (2012). Intrinsic stationarity for vector quantization: foundation of dual quantization. *SIAM J. Numer. Anal.*, **50**(2):747–780.
- [45] PAGÈS, G. AND WILBERTZ, B. (2012). Optimal Delaunay and Voronoi quantization schemes for pricing American style options. *Numerical methods in Finance*, 171–213, Springer Proc. Math., **12**, Springer, Heidelberg.
- [46] PAGÈS, G. AND WILBERTZ, B. (2018). Sharp rate for the dual quantization problem, *Séminaire de Probabilités XLV*, C. Donati, A. Lejay, A. Rouault eds, LNM 2215, Springer, Cham, 119–164.
- [47] PAGÈS, G. AND YU, J. (2016). Pointwise convergence of the Lloyd algorithm in higher dimension, *SIAM J. Control Optim.*, **54**(5): 2354–2382.
- [48] RAJAN, V. T. (1991). Optimality of the Delaunay triangulation in  $\mathbb{R}^d$ . In *SCG’91: Proceedings of the seventh annual symposium on Computational geometry*, 357–363, New York, NY, USA, ACM.
- [49] RUDIN, W. (1987). Real and Complex Analysis. Third Edition. McGraw Hill.

- [50] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.*, **36**:423–439.
- [51] VILLANI, C. (2009). *Optimal Transport Old and New*. Springer.
- [52] WIESEL J. (2019). Continuity of the martingale optimal transport problem on the real line, arXiv:1905.04574.