ALEA, Lat. Am. J. Probab. Math. Stat. **19**, 1–22 (2022)
DOI: 10.30757/ALEA.v19-01

# Quantization and martingale couplings

**Benjamin Jourdain and Gilles Pagès**

CERMICS, École des Ponts, INRIA, Marne-la-Vallée, France.
*E-mail address*: benjamin.jourdain@enpc.fr
*URL*: http://cermics.enpc.fr/∼jourdain/home.html

Laboratoire de Probabilités, Statistique et Modélisation, UMR 8001, Sorbonne Université, Campus Pierre et Marie Curie, case 158, 4, pl. Jussieu, F-75252 Paris Cedex 5, France.
*E-mail address*: gilles.pages@sorbonne-universite.fr
*URL*: http://www.lpsm.paris/dw/doku.php?id=users:pages:index

**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 quadratic optimal) 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 $\infty$.

## 1. Introduction

For $d \in \mathbb{Z}_+^*$ 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

$$\forall \varphi : \mathbb{R}^d \to \mathbb{R} \text{ convex }, \int_{\mathbb{R}^d} \varphi(x)\mu(dx) \leq \int_{\mathbb{R}^d} \varphi(y)\nu(dy), \tag{1.1}$$

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 $p \geq 1$,

*Received by the editors December 20th, 2020; accepted September 21st, 2021.*
2010 *Mathematics Subject Classification.* 60E15, 65C50, 65D32, 60J22, 60G42.
*Key words and phrases.* Quantization, Martingale Coupling, Weak Martingale Optimal Transport, Adapted Wasserstein distance.
This research benefited from the support of the "Chaire Risques Financiers", Fondation du Risque.

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 Strassen (1965), $\mu \leq_{cvx} \nu$ 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: $\forall x \in \mathbb{R}^d$, $\pi_x \in \mathcal{P}_1(\mathbb{R}^d)$ 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 $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. When the interest rate is zero, for an exotic option written on $d$ assets with payoff given by the function $c$ of their prices at successive 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 Beiglböck et al. (2013), this MOT problem has received recently a great attention in the financial mathematics literature. In particular, the structure of martingale optimal transport couplings has been investigated by Beiglböck and Juillet (2016); Campi et al. (2017); De March and Touzi (2019); Ghoussoub et al. (2019); Henry-Labordère and Touzi (2016), continuous time formulations by Dolinsky and Soner (2014); Galichon et al. (2014); Henry-Labordère et al. (2016), links with the Skorokhod embedding problem by Beiglböck et al. (2017), numerical methods by Alfonsi et al. (2020, 2019); De March (2018); Guo et al. (2019); Henry-Labordère (2019) and stability properties by Backhoff-Veraguas and Pammer (2019); Jourdain and Margheriti (2020); Wiesel (2012). The MOT problem is a particular instance where the measurable cost function $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ 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

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

Backhoff-Veraguas and Pammer (2019) introduced this WMOT problem by adding the martingale constraint to the Weak Optimal Transport problem introduced by Gozlan et al. (2017).

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. Baker (2015) proposes a quantile-based 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 $N, K \in \mathbb{Z}_+^*$ with $N/K \in \mathbb{Z}_+^*$, 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 (2012a) and further studied in Pagès and Wilbertz (2012b,c, 2018) yields another way to preserve the convex order in dimension

$d = 1$ when using the same grid to quantize both probability measures (see the remark after Proposition 10 in Pagès and Wilbertz (2012b)).

In two recent papers, Alfonsi et al. (2019, 2020) 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 (1.1). According to Alfonsi et al. (2019), the convex order restoration 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 that is 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 $d$, Alfonsi et al. (2020) suggest 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 (1.2) 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 $\nu$ in the convex order. Therefore the convex order between $\mu$ and $\nu$ is preserved by these combined approximations. 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 probability measures with bounded support. 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

$$W_p(\mu, \nu)^p = \inf_{\pi \in \mathcal{P}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy). \tag{1.2}$$

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

$$M_p(\eta, \nu) = \inf_{\pi \in \mathcal{M}(\eta,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \pi(dx, dy) \text{ for } (\eta, \nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d). \tag{1.3}$$

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 probability measure $\mu$ with bounded support. The dual quantization is obtained by minimizing $M_p(\mu, \eta)$ over

$$\mathcal{P}_{\geq \mu}(\mathbb{R}^d, N) = \left\{ \eta \in \mathcal{P}(\mathbb{R}^d, N) : \eta \geq_{cvx} \mu \right\}.$$

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 $N, K \to \infty$, $\bar{\pi}^{N,K}$ still converges to $\pi$ for the much finer adapted Wasserstein distance defined in (1.5) below which captures the temporal structure of probability distributions with two time marginals. Numerous financial applications of this adapted Wasserstein distance have been investigated by Backhoff-Veraguas et al. (2020a). According to Backhoff-Veraguas et al. (2020b), 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 only holds in dimension $d = 1$ (see Beiglböck et al., 2021b for a proof of stability when $d = 1$ and Brückerhoff and Juillet, 2021 for a nice counterexample to the stability in dimension $d \geq 2$), 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 $d \in \mathbb{Z}_+^*$.

- $|\cdot|$ denotes the canonical Euclidean norm on $\mathbb{R}^d$.
- $\mathrm{conv}(A)$ denotes the (closed) convex hull of $A \subset \mathbb{R}^d$ and $\mathrm{card}(A)$ its cardinality.
- Let $\mathcal{P}(\mathbb{R}^d)$ denote the set of probability measures on $\mathbb{R}^d$ endowed with its Borel $\sigma$-field $\mathcal{B}(\mathbb{R}^d)$. We endow $\mathcal{P}(\mathbb{R}^d)$ with the weak convergence topology.
- 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 and for $\mu \in \mathcal{P}(\mathbb{R}^d)$, we set

$$\mathcal{P}_{\leq \mu}(\mathbb{R}^d, N) = \left\{\eta \in \mathcal{P}(\mathbb{R}^d, N) : \eta \leq_{cvx} \mu\right\} \text{ and } \mathcal{P}_{\geq \mu}(\mathbb{R}^d, N) = \left\{\eta \in \mathcal{P}(\mathbb{R}^d, N) : \eta \geq_{cvx} \mu\right\}.$$

- 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 owing to Strassen's 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 $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$, then for each $\pi \in \mathcal{M}(\mu,\nu)$,

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

so that

$$M_2^2(\mu,\nu) = \int_{\mathbb{R}^d} |y|^2 \nu(dy) - \int_{\mathbb{R}^d} |x|^2 \mu(dx). \tag{1.4}$$

• 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 $\tilde{\pi}(d\tilde{x},d\tilde{y}) = \tilde{\mu}(d\tilde{x})\tilde{\pi}_{\tilde{x}}(d\tilde{y})$:

$$AW_p(\pi,\tilde{\pi}) = \inf_{m \in \mathcal{P}(\mu,\tilde{\mu})} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} (|x-\tilde{x}|^p + W_p^p(\pi_x,\tilde{\pi}_{\tilde{x}})) m(dx,d\tilde{x}) \right)^{1/p}. \tag{1.5}$$

## 2. 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 Graf and Luschgy, 2000; Pagès, 2015, 2018 among others):

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

$$e_p(\Gamma,\mu)^p = \int_{\mathbb{R}^d} |x - \mathrm{Proj}_\Gamma(x)|^p \mu(dx) \tag{2.1}$$

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

$$e_{p,N}(\mu) := \inf\Big\{ e_p(\Gamma,\mu) : \Gamma \subset \mathbb{R}^d,\ \mathrm{card}(\Gamma) \leq N \Big\} = e_p\big(\Gamma_{p,N},\mu\big).$$

When $\mathrm{card}(\mathrm{supp}(\mu)) \leq N$ then $\Gamma_{p,N} = \mathrm{supp}(\mu)$ and when $\mathrm{card}(\mathrm{supp}(\mu)) > N$, then $\Gamma_{p,N}$ has exactly $N$ pairwise distinct elements. Moreover, Theorem 4.2 in Graf and Luschgy (2000) ensures that

$$\mu\Big(\{x \in \mathbb{R}^d : \exists y \neq \tilde{y} \in \Gamma_{p,N},\ |x-y| = |x-\tilde{y}| = \mathrm{dist}(x,\Gamma_{p,N})\}\Big) = 0, \tag{2.2}$$

so that $\mu \circ \mathrm{Proj}_{\Gamma_{p,N}}^{-1}$ does not depend on the choice of the Borel nearest neighbour projection $\mathrm{Proj}_{\Gamma_{p,N}}$ on $\Gamma_{p,N}$. We denote this probability measure by $\hat{\mu}^{\Gamma_{p,N}}$. In the same way, when $X \sim \mu$, $\mathrm{Proj}_{\Gamma_{p,N}}(X)$ a.s. does not depend on the Borel nearest neighbour projection $\mathrm{Proj}_{\Gamma_{p,N}}$ 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 $\hat{X}^{\Gamma_{2,N}} = \mathrm{Proj}_{\Gamma_{2,N}}(X)$ with distribution $\hat{\mu}^{\Gamma_{2,N}}$ satisfy a *stationarity* (or *self-consistency*) property (see e.g. Graf and Luschgy, 2000, Pagès, 2015 or Pagès, 2018, Proposition 5.1 among others) that is

$$\mathbb{E}\big(X \,|\, \hat{X}^{\Gamma_{2,N}}\big) = \hat{X}^{\Gamma_{2,N}} \text{ so that } \hat{\mu}^{\Gamma_{2,N}} \leq_{cvx} \mu. \tag{2.3}$$

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

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

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

$$e_{2,N}(\mu)^2 = \mathbb{E}[|X - \hat{X}^{\Gamma_{2,N}}|^2] = \mathbb{E}[|X|^2] - \mathbb{E}[|\hat{X}^{\Gamma_{2,N}}|^2] = M_2^2(\hat{\mu}^{\Gamma_{2,N}}, \mu). \tag{2.4}$$

**Proposition 2.1.** *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*

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

*and for any Borel nearest neighbour projection $\mathrm{Proj}_\Gamma$ on $\Gamma$, $\mu \circ \mathrm{Proj}_\Gamma^{-1}$ 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 Graf and Luschgy (2000). Before proving the proposition, let us state and check some easy consequence of the necessary and sufficient condition in $(c)$.

**Corollary 2.2.** *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 2.2:* 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 2.1$(c)$, 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 2.1$(c)$, $\Gamma_{2,N}$ is a quadratic optimal $N$-quantizer of $\nu$ and $\hat{\nu}^{\Gamma_{2,N}} = \hat{\mu}^{\Gamma_{2,N}}$. □

*Proof of Proposition 2.1:* $(a)$ 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. $|x-y| \geq \mathrm{dist}(x,\Gamma)$. Therefore

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

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

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

where the equality follows from (2.1). Therefore the inequalities are equalities and $\mu\circ\mathrm{Proj}_\Gamma^{-1}$ 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,

$$W_p(\mu,\mathcal{P}(\Gamma_N)) \leq W_p(\mu,\nu) = W_p(\mu,\mathcal{P}(\mathbb{R}^d,N)) \leq W_p(\mu,\mathcal{P}(\Gamma)). \tag{2.5}$$

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 (2.5) implies that the first inequality is an equality so that, with $(a)$, $W_p^p(\mu,\nu) = \int_{\mathbb{R}^d}\mathrm{dist}(x,\Gamma_N)^p\mu(dx)$. Hence, for any $W_p$-optimal coupling $\pi\in\mathcal{P}(\mu,\nu)$,

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

Since $1 = \nu(\Gamma_N) = \int_{\mathbb{R}^d}\pi_x(\Gamma_N)\mu(dx)$, $\pi(dx,dy)$ a.e., $|y-x|^p \geq \mathrm{dist}(x,\Gamma_N)^p$. Therefore, $\pi(dx,dy)$ a.e. $y\in\Gamma_N$ and $|y-x| = \mathrm{dist}(x,\Gamma_N)$. With (2.2), we conclude that $\mu(dx)$ a.e. there is a unique point $\hat x^{\Gamma_N}\in\Gamma_N$ such that $|\hat x^{\Gamma_N}-x| = \mathrm{dist}(x,\Gamma_N)$ 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 $(b)$, (2.3) and (2.4), 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 $e_{2,N}(\mu)^2 = M_2^2(\nu,\mu)$. 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 (1.4), one has

$$W_2^2(\mu,\eta) \leq M_2^2(\eta,\mu) = \int_{\mathbb{R}^d}|y|^2\mu(dy) - \int_{\mathbb{R}^d}|x|^2\eta(dx). \tag{2.6}$$

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 (2.6) for the last inequality and the last equality, we obtain that

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

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,

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

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 (2.6) written for $\eta=\nu$ ensures that $\nu$ is a $W_2$-projection of $\mu$ on $\mathcal{P}_{\leq\mu}(\mathbb{R}^d,N)$. □

**Remark about uniqueness.** As a consequence of Proposition 2.1$(b)$, 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 Trushkin, 1982 and also Kieffer, 1982) 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 exact rate of convergence of the $L^p$-quantization error stated for instance in Theorem 5.2 Pagès (2018).

**Theorem 2.3** (Pierce Lemma for primal quantization)**.** *Let $p\geq 1$ and $\eta>0$. For every dimension $d\geq 1$, there exists a real constant $\widetilde{C}^{vor}_{d,\eta,p}>0$ such that, for every random vector $X:(\Omega,\mathcal{A},\mathbb{P})\to\mathbb{R}^d$,*

$$e_{p,N}(X)\leq\widetilde{C}^{vor}_{d,\eta,p}N^{-\frac{1}{d}}\sigma_{p+\eta}(X)\tag{2.7}$$

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

## 3. Dual (Delaunay) quantization

We assume throughout this section that $\mu$ is compactly supported. Let $X:(\Omega,\mathcal{A},\mathbb{P})\to\mathbb{R}^d$ be a random vector lying in $L^\infty(\mathbb{P})$ with distribution $\mu$. Optimal dual (or Delaunay) quantization as introduced in Pagès and Wilbertz (2012b) 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 $p\in[1,+\infty)$,

$$\begin{aligned}d_{p,N}(X)\;=\;\inf_{\check{X}}\Big\{\big\|X-\check{X}\big\|_p:\check{X}:(\Omega\times[0,1],\mathcal{A}\otimes\mathcal{B}([0,1]),\mathbb{P}\otimes\lambda)\to\mathbb{R}^d,\\ \operatorname{card}\big(\check{X}(\Omega\times[0,1])\big)\leq N\text{ and }\mathbb{E}(\check{X}|X)=X\Big\}.\end{aligned}$$

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 Kallenberg (1997), 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

$$d_{p,N}(X)^p = \inf_{\nu\in\mathcal{P}_{\geq\mu}(\mathbb{R}^d,N)} \inf_{\pi\in\mathcal{M}(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^p \pi(dx,dy) = \inf_{\nu\in\mathcal{P}_{\geq\mu}(\mathbb{R}^d,N)} M_p^p(\mu,\nu). \tag{3.1}$$

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

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

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

$$\forall\, y \in \mathrm{conv}(\Gamma), \quad \int_0^1 \mathrm{Proj}_\Gamma^{del}(y,u)du = y, \tag{3.2}$$

from which one derives the dual stationarity property

$$\mathbb{E}\left(\mathrm{Proj}_\Gamma^{del}(X,U) \,\middle|\, X\right) = X \text{ when } U \sim \mathcal{U}([0,1]) \text{ is independent of } X. \tag{3.3}$$

The stationarity property remains valid as soon as $X$ is $\mathrm{conv}(\Gamma)$-valued and implies that the distribution of $\mathrm{Proj}_\Gamma^{del}(X,U)$ 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 $\Gamma = \{x_1, x_2, \ldots, x_N\}$ with $x_1 < x_2 < \ldots < x_N$, the probability measure minimizing $\nu \mapsto M_p^p(\mu,\nu)$ over $\mathcal{P}_{\geq\mu}(\Gamma)$ is the distribution $\check{\mu}^\Gamma$ of $\mathrm{Proj}_\Gamma^{del}(X,U)$ for the splitting operator

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

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 $(X, \mathrm{Proj}_\Gamma^{del}(X,U))$. Last, according to the remark after Proposition 10 in Pagès and Wilbertz (2012b), when $\mu \leq_{cvx} \eta$ with $\eta$ compactly supported in $[x_1, x_N]$, then $\check{\mu}^\Gamma \leq_{cvx} \check{\eta}^\Gamma$. 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 $\varphi : \mathbb{R} \to \mathbb{R}$. 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 Alfonsi et al. (2020), this convex order preservation does not generalize to higher dimensions where the minimizers are not so easy to express.

Whatever the dimension $d \in \mathbb{Z}_+^*$, for every level $N \geq d+1$, there exists an $L^p$-optimal dual quantization grid $\Gamma_{p,N}^{del}$ and a splitting operator $\mathrm{Proj}_{\Gamma_{p,N}^{del}}^{del}$ (see Pagès and Wilbertz, 2012b) such that

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

and $\mathrm{Proj}_{\Gamma_{p,N}^{del}}^{del}(X,U)$ takes each value in $\Gamma_{p,N}^{del}$ with positive probability. For more details on this dual projection, see Pagès and Wilbertz (2012a,b,c) 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 2.1($b$), 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 (3.1),

$$e_{p,N}(\mu) = W_p\big(\mu, \mathcal{P}(\mathbb{R}^d, N)\big) \leq W_p\big(\mu, \mathcal{P}_{\geq\mu}(\mathbb{R}^d, N)\big) \leq \inf_{\nu \in \mathcal{P}_{\geq\mu}(\mathbb{R}^d,N)} M_p(\nu,\mu) = d_{p,N}(\mu). \tag{3.4}$$

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 p.97 Villani, 2009), 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 (3.4) 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 2.1($b$) 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 $\mathcal{P}_{\geq\mu}(\mathbb{R}^d, N)$. 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 as proved in the examples below.

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

$$d_{2,N}(\mu)^2 = \inf_{\nu \in \mathcal{P}_{\geq\mu}(\mathbb{R}^d,N)} \int_{\mathbb{R}^d} |y|^2\nu(dy) - \int_{\mathbb{R}^d} |x|^2\mu(dx). \tag{3.5}$$

**Proposition 3.1.** *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* $\mathrm{Proj}_{\Gamma_{2,N}^{del}}^{del}(X,U)$.

*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, (3.5) 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 $\eta$. □

In the next examples, after revisiting the optimal primal and dual quantizers of the uniform law on the interval $[0,1]$ in terms of Wasserstein projections, we check that, in dimension $d=1$, optimal dual quantization of two distributions does not necessarily preserve the convex order whereas dual quantizations using a common grid does. We last illustrate that the distribution of an $L^p$-optimal dual $N$-quantization of $\mu$ and its $W_p$-projection on $\mathcal{P}_{\geq\mu}(\mathbb{R}^d,N)$ may differ.

**Examples. (a) Projections of $\mu=\mathcal{U}[0,1]$.** For two real numbers $a<b$, let $\mathcal{U}[a,b]$ denote the uniform distribution on $[a,b]$ with density $\frac{1_{[a,b]}(x)}{b-a}$ with respect to the Lebesgue measure. We consider the approximation of $\mu$ by distributions 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 $x_1\leq x_2\leq\ldots\leq x_N$ and $(p_1,\ldots,p_N)\in[0,1]^N$ 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 $p\geq1$, let us give a direct proof that $\hat{\mu}_N$ is the $W_p$-projection of $\mathcal{U}[0,1]$ on $\mathcal{P}(\mathbb{R},N)$ (consequence of Proposition 2.1$(b)$) 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 $(0,1)\ni u\mapsto F^{-1}_{\nu_N}(u)-u$ is equal to

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

If $\nu_N=\hat{\mu}_N$ (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 $\{0,1,2,\ldots,N\}$, $(\eta_N-\hat{\eta}_N)^+$ 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 $p\geq1$,

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

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 $k\in\{1,\ldots,N\}$, $\frac{-1}{2N}\leq x_k-(p_1+\ldots+p_k)$ and $x_k-(p_1+\ldots+p_{k-1})\leq\frac{1}{2N}$ so that $p_k\leq\frac{1}{N}$. With the normalisation, we deduce that $p_\ell=\frac{1}{N}$ for each $\ell\in\{1,\ldots,N\}$, which plugged in the two last inequalities implies that $x_k=\frac{2k-1}{2N}$ for each $k\in\{1,\ldots,N\}$ so that $\nu_N=\hat{\mu}_N$. Therefore, for all $p\geq1$, $\hat{\mu}_N$ 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 $x_1\leq0$ and $x_N\geq1$ so that $[x_1-p_1,x_1]\cap[x_N-1,x_N-(p_1+\ldots+p_{N-1})]\subset\{0\}$ and $\eta_N$ has a density with respect to the Lebesgue measure with values in $\{0,1,\ldots,N-1\}$. 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 $k\in\{1,\ldots,N\}$, $\frac{-1}{2(N-1)}\leq x_k-(p_1+\ldots+p_k)$ and $x_k-(p_1+\ldots+p_{k-1})\leq\frac{1}{2(N-1)}$ so that $p_k\leq\frac{1}{N-1}$ and with the reinforced inequalities $x_1\leq0$ and $x_N\geq1$ due to the convex

order, $p_1 \le \frac{1}{2(N-1)}$ and $p_N \le \frac{1}{2(N-1)}$. 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 $x_1 = 1$ so that $\nu_N = \check{\mu}_N$. Therefore, for all $p \ge 1$, $\check{\mu}_N$ is the $W_p$-projection of $\mathcal{U}[0,1]$ on $\mathcal{P}_{\ge \mathcal{U}[0,1]}(\mathbb{R},N)$ and

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

Notice that when $\nu_N \ge_{cvx} \mathcal{U}[0,1]$, 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 $\varphi : \mathbb{R} \to \mathbb{R}$, $\int_{\mathbb{R}} \varphi(x)(\eta_N - \check{\eta}_N)^+(dx) \ge \int_{\mathbb{R}} \varphi(x)(\check{\eta}_N - \eta_N)^+(dx)$ and $\eta_N \ge_{cvx} \check{\eta}_N$.

**(b) Optimal dual quantization does not preserve the convex order.** 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 $\left\{0, \frac{2}{5}, \frac{7}{15}, \frac{8}{15}, \frac{3}{5}, 1\right\}$. Note that $\mu_6$ is not comparable for the convex order with the projection $\check{\mu}_6 = \frac{1}{10}\delta_0 + \frac{1}{5}\left(\delta_{\frac{1}{5}} + \delta_{\frac{2}{5}} + \delta_{\frac{3}{5}} + \delta_{\frac{4}{5}}\right) + \frac{1}{10}\delta_1$ of $\mathcal{U}[0,1]$ on $\mathcal{P}_{\ge\mathcal{U}[0,1]}(\mathbb{R},6)$. Indeed, one checks that

$$\mu_6 = \frac{2}{5}\left(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{2}{5}}\right) + \frac{1}{5}\left(\frac{1}{6}\delta_{\frac{2}{5}} + \frac{1}{3}\delta_{\frac{7}{15}} + \frac{1}{3}\delta_{\frac{8}{15}} + \frac{1}{6}\delta_{\frac{3}{5}}\right) + \frac{2}{5}\left(\frac{1}{2}\delta_{\frac{3}{5}} + \frac{1}{2}\delta_1\right)$$
$$\check{\mu}_6 = \frac{2}{5}\left(\frac{1}{4}\delta_0 + \frac{1}{2}\delta_{\frac{1}{5}} + \frac{1}{4}\delta_{\frac{2}{5}}\right) + \frac{1}{5}\left(\frac{1}{2}\delta_{\frac{2}{5}} + \frac{1}{2}\delta_{\frac{3}{5}}\right) + \frac{2}{5}\left(\frac{1}{4}\delta_{\frac{3}{5}} + \frac{1}{2}\delta_{\frac{4}{5}} + \frac{1}{4}\delta_1\right)$$
$$\text{with} \quad \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{\frac{2}{5}} \ge_{cvx} \frac{1}{4}\delta_0 + \frac{1}{2}\delta_{\frac{1}{5}} + \frac{1}{4}\delta_{\frac{2}{5}} \quad \text{and} \quad \frac{1}{2}\delta_{\frac{3}{5}} + \frac{1}{2}\delta_1 \ge_{cvx} \frac{1}{4}\delta_{\frac{3}{5}} + \frac{1}{2}\delta_{\frac{4}{5}} + \frac{1}{4}\delta_1$$
$$\text{whereas} \quad \frac{1}{6}\delta_{\frac{2}{5}} + \frac{1}{3}\delta_{\frac{7}{15}} + \frac{1}{3}\delta_{\frac{8}{15}} + \frac{1}{6}\delta_{\frac{3}{5}} \le_{cvx} \frac{1}{2}\delta_{\frac{2}{5}} + \frac{1}{2}\delta_{\frac{3}{5}}.$$

Hence $\int_{\mathbb{R}} \varphi(x)\mu_6(dx) > \int_{\mathbb{R}} \varphi(x)\check{\mu}_6(dx)$ (resp. $\int_{\mathbb{R}} \varphi(x)\mu_6(dx) < \int_{\mathbb{R}} \varphi(x)\check{\mu}_6(dx)$) when the convex function $\varphi : \mathbb{R} \to \mathbb{R}$ 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.

**(c) Optimal dual quantization does not coincide with $W_2$-projection.** Let us finally check on the example $\mu(dx) = 2x\mathbf{1}_{[0,1]}(x)dx$ that the optimal quadratic dual quantization of $\mu$ on $N = 3$ points does not coincide with the $W_2$-projection $\tilde{\mu}$ of $\mu$ on $\mathcal{P}_{\ge\mu}(\mathbb{R}^d, 3)$. We look for $\nu \in \mathcal{P}_{\ge\mu}(\mathbb{R}, 3)$ minimizing either $\int_{\mathbb{R}} y^2\nu(dy)$ to compute the law of the optimal quadratic dual quantization 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 $\mathcal{P}_{\ge\mu}(\mathbb{R}, 3)$

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

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,

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

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

$$\forall x \in \mathbb{R},\ \varphi_\nu(x) := \int_{-\infty}^x \nu((-\infty,y])dy \geq \varphi_\mu(x) = 1_{[0,1]}(x)\frac{x^3}{3} + 1_{\{x>1\}}\left(x-\frac{2}{3}\right). \tag{3.6}$$

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 established by Pagès and Wilbertz (2018) on the convergence rate of dual quantization for bounded random vectors.

**Theorem 3.2** (Pierce Lemma for dual quantization)**.** *Let $p \geq 1$ and $\eta > 0$. For every dimension $d \geq 1$, there exists a real constant $\widetilde{C}_{d,\eta,p}^{del} > 0$ such that, for every random vector $X : (\Omega,\mathcal{A},\mathbb{P}) \to \mathbb{R}^d$, $L^\infty(\mathbb{P})$-bounded,*

$$d_{p,N}(X) \leq \widetilde{C}_{d,\eta,p}^{del} N^{-\frac{1}{d}}\sigma_{p+\eta}(X) \tag{3.7}$$

*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 2.3 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 (2.7) and (3.7) hold with factor $N^{-1/d'}$ so that $N^{-1/d}$ is suboptimal.

## 4. 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 $\Gamma^{del}_{p,K}$ and let $\hat{\mu}^N$ 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 (1.2), only martingale couplings are considered in the minimization defining (1.3)) between $\nu$ (resp. $\hat{\mu}_N$) and $\check{\nu}^K$ (resp. $\mu$):

$$d^p_{p,K}(\nu) = M^p_p(\nu, \check{\nu}^K) \text{ and } e^2_{2,N}(\mu) = M^2_2(\hat{\mu}^N, \mu). \tag{4.1}$$

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 (2.3) 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 $\mathrm{Proj}^{del}_{\Gamma^{del}_{p,K}}(y, U)$ when $y \in \mathrm{Conv}(\Gamma^{del}_{p,K})$ and $\delta_y(d\check{y})$ otherwise. When $\mu \leq_{cvx} \nu$, we are now going to exploit (4.1) to approximate any $\pi \in \mathcal{M}(\mu, \nu)$ by

$$\bar{\pi}^{N,K}(d\hat{x}, d\check{y}) = \int_{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d} \delta_{\mathrm{Proj}_{\Gamma_{2,N}}(x)}(d\hat{x})\pi(dx, dy)q_y(d\check{y}). \tag{4.2}$$

For $(X, Y) \sim \pi$ independent from the random variable $U$ uniformly distributed on $[0,1]$, the random vector $(\mathrm{Proj}_{\Gamma_{2,N}}(X), X, Y, \mathrm{Proj}^{del}_{\Gamma^{del}_{p,K}}(Y, U))$ is distributed according to $\delta_{\mathrm{Proj}_{\Gamma_{2,N}}(x)}(d\hat{x})\pi(dx, dy)q_y(d\check{y})$ and therefore $(\mathrm{Proj}_{\Gamma_{2,N}}(X), \mathrm{Proj}^{del}_{\Gamma^{del}_{p,K}}(Y, U))$ is distributed according to $\bar{\pi}^{N,K}$. Therefore, the first marginal of $\bar{\pi}^{N,K}$ is the distribution $\hat{\mu}^N$ of $\mathrm{Proj}_{\Gamma_{2,N}}(X)$ and its second marginal is the law $\check{\nu}^K$ of $\mathrm{Proj}^{del}_{\Gamma^{del}_{p,K}}(Y, U)$. Moreover, using that $U$ is independent from $(X, Y)$ and (3.2) for the second equality, then that $\pi$ is a martingale coupling for the fourth equality and the stationarity property (2.3) for the last one, we obtain

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

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

**Theorem 4.1** (Convergence rate for martingale couplings)**.** *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 $\nu$. 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 (4.2) are such that*

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

*Moreover, when $p \geq 2$,*

$$W^2_2(\bar{\pi}^{N,K}, \pi) \leq e^2_{2,N}(\mu) + d^2_{p,K}(\nu).$$

*Last, for any $p \geq 1$,*

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

According to Theorems 2.3 and 3.2, $\sup_{N\geq 1} N^{1/d} e_{2,N}(\mu) < \infty$ and $\sup_{K\geq 1} K^{1/d} d_{p,K}(\nu) < \infty$.

*Proof*: We have

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

When $p\le 2$, $\int_{\mathbb{R}^d}|\mathrm{Proj}_{\Gamma_{2,N}}(x)-x|^p\mu(dx) \le \left(\int_{\mathbb{R}^d}|\mathrm{Proj}_{\Gamma_{2,N}}(x)-x|^2\mu(dx)\right)^{p/2} = e^p_{2,N}(\mu)$. 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 in Pagès and Sagna (2018) to obtain that $\sup_{N\ge1} N^{p/d}\int_{\mathbb{R}^d}|\mathrm{Proj}_{\Gamma_{2,N}}(x) - x|^p\mu(dx)<\infty$, which completes the proof of the first inequality. In a similar way, when $p\ge 2$

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

Now, let

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

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 $\check\pi^K_x(d\check y)$ and $\pi_x(dy)$ in the definition of the usual Wasserstein distance, we obtain that

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

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

To do so, we denote by $(x_i)_{1\le i\le\gamma_{2,N}}$ the points in the grid $\Gamma_{2,N}$ with cardinality $\gamma_{2,N}\le N$ and by $(y_j)_{1\le j\le\gamma^{del}_{p,K}}$ the points in the grid $\Gamma^{del}_{p,K}$ with cardinality $\gamma^{del}_{p,K}\le K$ (as soon as the support of $\mu$ (resp. $\nu$) is not restricted to less than $N$ (resp. $K$) points, $\gamma_{2,N}=N$ (resp. $\gamma^{del}_{p,K}=K$)). Let $C_i=\{x\in\mathbb{R}^d : |x-x_i|<\min_{1\le k\le\gamma_{2,N},k\ne i}|x-x_k|\}$, $i\in\{1,\dots,\gamma_{2,N}\}$, denote the open Voronoi cells induced by $\Gamma_{2,N}$. Using (2.2) for the equality, we have

$$\begin{aligned}
AW_p^p(\bar\pi^{N,K},\check\pi^K) &\le \int_{\mathbb{R}^d\times\mathbb{R}^d}(|\hat x - x|^p + W_p^p(\bar\pi^{N,K}_{\hat x},\check\pi^K_x))\delta_{\mathrm{Proj}_{\Gamma_{2,N}}(x)}(d\hat x)\mu(dx)\\
&= \sum_{i=1}^{\gamma_{2,N}}\int_{C_i}(|x_i - x|^p + W_p^p(\bar\pi^{N,K}_{x_i},\check\pi^K_x))\mu(dx). \qquad (4.3)
\end{aligned}$$

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

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

As a consequence, we have

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

With Theorem 6.15 p.103 Villani (2009), we deduce that

$$\begin{aligned} W_p^p(\bar{\pi}^{N,K}_{x_i}, \check{\pi}^K_x) &= 2^{p-1} \max_{1\le j\le \gamma^{del}_{p,K}} |y_j|^p \mathrm{TV}(\hat{\pi}^{N,K}_{x_i}, \hat{\pi}^K_x) \\ &\le 2^{p-1} \max_{1\le j\le \gamma^{del}_{p,K}} |y_j|^p \sum_{j=1}^{\gamma^{del}_{p,K}} \left| q_j(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j(\xi)\mu(d\xi) \right|. \end{aligned}$$

Plugging this inequality in (4.3), we deduce that

$$\begin{aligned} AW_p^p(\bar{\pi}^{N,K}, \check{\pi}^K) \le& \sum_{i=1}^{\gamma_{2,N}} \int_{C_i} |x_i - x|^p \mu(dx) \\ &+ 2^{p-1} \max_{1\le j\le \gamma^{del}_{p,K}} |y_j|^p \sum_{j=1}^{\gamma^{del}_{p,K}} \sum_{i=1}^{\gamma_{2,N}} \int_{C_i} \left| q_j(x) - \frac{1}{\mu(C_i)} \int_{C_i} q_j(\xi)\mu(d\xi) \right| \mu(dx). \end{aligned} \tag{4.4}$$

Using Jensen's inequality in the case $p \in [1,2]$, the $L^2 - L^p$- distortion mismatch Theorem 4.3 Pagès and Sagna (2018) when $2 < p < 2+d$ and Hölder's inequality when $p \ge 2+d$, we get

$$\sum_{i=1}^{\gamma_{2,N}} \int_{C_i} |x_i - x|^p \mu(dx) = \int_{\mathbb{R}^d} |\mathrm{Proj}_{\Gamma_{2,N}}(x) - x|^p \mu(dx) \le \begin{cases} e^p_{2,N}(\mu) \text{ if } p \in [1,2] \\ CN^{-p/d} \text{ if } 2 < p < 2+d \\ 2^{(p-3)\vee 0} \max\limits_{1\le j\le \gamma^{del}_{p,K}} |y_j|^{p-2} e^2_{2,N}(\mu) \text{ if } p \ge 2+d \end{cases},$$

where the constant $C$ does not depend on $N$. With Theorem 2.3, we deduce that the first term in the right-hand side of (4.4) goes to 0 as $N \to \infty$. Since $\int_{\mathbb{R}^d} |q_j(x)|\mu(dx) \le 1$, by Theorem 3.14 p.69 Rudin (1987), there exists a sequence $(q_j^n)_{n\in\mathbb{Z}_+}$ of continuous and compactly supported functions on $\mathbb{R}^d$ such that $\lim_n \int_{\mathbb{R}^d} |q_j^n(x) - q_j(x)|\mu(dx) = 0$. 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

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

where the right-hand side goes to 0 as $n \to +\infty$. We deduce that to prove that the second term in the right-hand side of (4.4) goes to 0 as $N \to +\infty$, 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 $n \in \mathbb{Z}_+$. For $X \sim \mu$ and $\hat{X}^N = \mathrm{Proj}_{\Gamma_{2,N}}(X)$,

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

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 2.3, we deduce that the left-hand side goes to 0 as $N \to +\infty$ and conclude that so does $AW_p(\bar{\pi}^{N,K}, \check{\pi}^K)$. □

## 5. Application to weak martingale optimal transport problems

We endow $\mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d) = \left\{(\mu,\nu) : \mu,\nu \in \mathcal{P}_1(\mathbb{R}^d) \text{ and } \mu \leq_{cvx} \nu\right\}$ 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 $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ Borel measurable, the Weak Martingale Optimal Transport problem introduced in Backhoff-Veraguas and Pammer (2019) consists in computing for $(\mu,\nu) \in \mathcal{P}_{\leq} \times \mathcal{P}_1(\mathbb{R}^d)$

$$V(\mu,\nu) = \inf_{\pi \in \mathcal{M}(\mu,\nu)} \int_{\mathbb{R}^d} C(x,\pi_x)\mu(dx) \tag{5.1}$$

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

$$\tilde{C}(x,\eta) = \begin{cases} C(x,\eta) \text{ if } \int_{\mathbb{R}^d} y\eta(dy) = x \\ +\infty \text{ otherwise} \end{cases}, \tag{5.2}$$

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

$$\tilde{V}(\mu,\nu) = \inf_{\pi \in \mathcal{P}(\mu,\nu)} \int_{\mathbb{R}^d} \tilde{C}(x,\pi_x)\mu(dx) \text{ for } (\mu,\nu) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \tag{5.3}$$

introduced by Gozlan et al. (2017) and studied by Backhoff-Veraguas et al. (2019). 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 $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ with at most linear growth in its second variable. We will make the following assumption on the cost function $C$:

**(C)**The cost function $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ 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

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

The existence of minimal couplings in the WMOT problem (5.1) and the lower semi-continuity of the value function $V$ are deduced from Theorem 2.6 and Proposition A.12$(b)$ Beiglböck et al. (2021b). Theorem 2.6 Beiglböck et al. (2021b) also ensures convergence of the optimal couplings under convergence of the value function, a property which holds in dimension $d = 1$ when $C$ is continuous.

**Proposition 5.1.** *Assume* ***(C)****.*

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

*Let $((\mu_k,\nu_k))_{k\in\mathbb{Z}_+}$ be a $\mathcal{P}_\leq \times \mathcal{P}_1(\mathbb{R}^d)$-valued sequence converging to $(\mu,\nu)$ in this space.*

**(ii):** *Let for each $k \in \mathbb{Z}_+$, $\pi^\star_k$ be an optimal coupling for $V(\mu_k,\nu_k)$. If $V(\mu,\nu) = \lim_k V(\mu_k,\nu_k)$, then all the accumulation points of $(\pi^\star_k)_{k\in\mathbb{Z}_+}$ for the weak convergence topology are minimizers for $V(\mu,\nu)$. If $C$ is moreover strictly convex in the measure argument, then the sequence $(\pi^\star_k)_{k\in\mathbb{Z}_+}$ converges in $AW_1$ to the unique optimal coupling $\pi^\star$ between $\mu$ and $\nu$.*

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

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 4.1 ensures that it is possible to approximate in $AW_1$ distance any optimal martingale coupling $\pi^\star$ 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 5.2.** *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 $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ 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*

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

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

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

With Proposition 5.1, we easily deduce the following corollary.

**Corollary 5.3.** *Assume* ***(C)*** *with the cost function $C$ moreover continuous and let $\mu,\nu,\hat\mu^N,\check\nu^K$ be as in Lemma 5.2. Then*

$$\lim_{N,K\to\infty} V(\hat\mu^N,\check\nu^K) = V(\mu,\nu).$$

*Moreover, all accumulation points as $N,K \to \infty$ of sequences $(\pi^\star_{N,K})_{N,K}$ 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*

$$\lim_{N,K\to\infty} AW_1(\pi^\star_{N,K},\pi^\star) = 0,$$

*where $\pi^\star$ is the unique optimal coupling between $\mu$ and $\nu$.*

*Proof of Lemma 5.2:* We first suppose that $C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ 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 $\mathcal{P}_1(\mathbb{R}^d)$ endowed with the Wasserstein distance $W_1$ 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 $\bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d)$, there exists $\alpha > 0$ such that

$$\forall (x, \eta, \tilde{x}, \tilde{\eta}) \in \bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d) \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \text{ s.t. } |x - \tilde{x}| + W_1(\eta, \tilde{\eta}) \leq \alpha,\ |C(x, \eta) - C(\tilde{x}, \tilde{\eta})| \leq \varepsilon$$

$$\text{and } \forall (x, \eta, \tilde{x}, \tilde{\eta}) \in \bar{B} \times \mathcal{P}_{\bar{B}}(\mathbb{R}^d) \times \bar{B} \times \mathcal{P}_1(\mathbb{R}^d),\ |C(x, \eta)| + |C(\tilde{x}, \tilde{\eta})| \leq \frac{1}{\alpha}\left(1 + \int_{\mathbb{R}^d} |y|\tilde{\eta}(dy)\right).$$

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 4.1, there exists $\bar{\pi}^{N,K} \in \mathcal{M}(\hat{\mu}^N, \check{\nu}^K)$ such that $\limsup_{N\to\infty} AW_1(\bar{\pi}^{N,K}, \pi) \leq d_{p,K}(\nu)$ and $\lim_{N,K\to\infty} AW_1(\bar{\pi}^{N,K}, \pi) = 0$. 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

$$\begin{aligned}
&\left|\int_{\mathbb{R}^d} C(x, \pi_x)\mu(dx) - \int_{\mathbb{R}^d} C(\tilde{x}, \bar{\pi}^{N,K}_{\tilde{x}})\hat{\mu}^N(d\tilde{x})\right| \leq \int_{\mathbb{R}^d\times\mathbb{R}^d} \left|C(x, \pi_x) - C(\tilde{x}, \bar{\pi}^{N,K}_{\tilde{x}})\right| m^{N,K}(dx, d\tilde{x}) \\
&\leq \varepsilon + \int_{\mathbb{R}^d\times\mathbb{R}^d} \left|C(x, \pi_x) - C(\tilde{x}, \bar{\pi}^{N,K}_{\tilde{x}})\right| 1_{\{|x-\tilde{x}|+W_1(\pi_x,\bar{\pi}^{N,K}_{\tilde{x}})\geq\alpha\}} m^{N,K}(dx, d\tilde{x}) \\
&\leq \varepsilon + \frac{1}{\alpha}\int_{\mathbb{R}^d\times\mathbb{R}^d} \left(1 + \int_{\mathbb{R}^d} |y|\bar{\pi}^{N,K}_{\tilde{x}}(dy)\right) 1_{\{|x-\tilde{x}|+W_1(\pi_x,\bar{\pi}^{N,K}_{\tilde{x}})\geq\alpha\}} m^{N,K}(dx, d\tilde{x}) \qquad (5.4)
\end{aligned}$$

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

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

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

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

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

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

Since $\lim_{N,K\to\infty} AW_1(\pi, \bar{\pi}^{N,K}) = 0 = \lim_{K\to\infty} W_1(\nu, \check{\nu}^K)$, with Lemma 3.1 (b) Beiglböck et al. (2021a), we deduce that the second term in the right-hand side of (5.4) tends to 0 as $N, K \to \infty$. Hence

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

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

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

$$\left|\int_{\mathbb{R}^d} C(x,\pi_x)\mu(dx) - \int_{\mathbb{R}^d} C(\tilde{x},\bar{\pi}^{N,K}_{\tilde{x}})\hat{\mu}^N(d\tilde{x})\right| \le \int_{\mathbb{R}^d\times\mathbb{R}^d} \left|C(x,\pi_x) - C(\tilde{x},\bar{\pi}^{N,K}_{\tilde{x}})\right| m^{N,K}(dx,d\tilde{x})$$
$$\le \mathrm{Lip}(C)\int_{\mathbb{R}^d\times\mathbb{R}^d} \left(|x-\tilde{x}| + W_1(\pi_x,\bar{\pi}^{N,K}_{\tilde{x}})\right) m(dx,d\tilde{x}) = \mathrm{Lip}(C)AW_1(\pi,\bar{\pi}^{N,K}).$$

Therefore

$$V(\hat{\mu}^N,\check{\nu}^K) \le \int_{\mathbb{R}^d} C(\tilde{x},\bar{\pi}^{N,K}_{\tilde{x}})\hat{\mu}^N(d\tilde{x}) \le \int_{\mathbb{R}^d} C(x,\pi_x)\mu(dx) + \mathrm{Lip}(C)AW_1(\pi,\bar{\pi}^{N,K})$$
$$\le V(\mu,\nu) + \varepsilon + \mathrm{Lip}(C)AW_1(\pi,\bar{\pi}^{N,K}).$$

Since $\varepsilon$ is arbitrary, we deduce that $\limsup_{N\to\infty} V(\hat{\mu}^N,\check{\nu}^K) \le V(\mu,\nu) + \mathrm{Lip}(C)d_{p,K}(\nu)$. □

## References


Alfonsi, A., Corbetta, J., and Jourdain, B. Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *Int. J. Theor. Appl. Finance*, **22** (3), 1950002, 41 (2019). MR3951948.

Alfonsi, A., Corbetta, J., and Jourdain, B. Sampling of probability measures in the convex order by Wasserstein projection. *Ann. Inst. Henri Poincaré Probab. Stat.*, **56** (3), 1706–1729 (2020). MR4116705.

Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., and Eder, M. Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.*, **24** (3), 601–632 (2020a). MR4118990.

Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., and Eder, M. All adapted topologies are equal. *Probab. Theory Related Fields*, **178** (3-4), 1125–1172 (2020b). MR4168395.

Backhoff-Veraguas, J., Beiglböck, M., and Pammer, G. Existence, duality, and cyclical monotonicity for weak transport costs. *Calc. Var. Partial Differential Equations*, **58** (6), Paper No. 203, 28 (2019). MR4029731.

Backhoff-Veraguas, J. and Pammer, G. Stability of martingale optimal transport and weak optimal transport. *ArXiv Mathematics e-prints* (2019). arXiv: 1904.04171.

Baker, D. M. Quantizations of probability measures and preservation of the convex order. *Statist. Probab. Lett.*, **107**, 280–285 (2015). MR3412787.

Beiglböck, M., Cox, A. M. G., and Huesmann, M. Optimal transport and Skorokhod embedding. *Invent. Math.*, **208** (2), 327–400 (2017). MR3639595.

Beiglböck, M., Henry-Labordère, P., and Penkner, F. Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.*, **17** (3), 477–501 (2013). MR3066985.

Beiglböck, M., Jourdain, B., Margheriti, W., and Pammer, G. Approximation of martingale couplings on the line in the weak adapted topology. *ArXiv Mathematics e-prints* (2021a). arXiv: 2101.02517.

Beiglböck, M., Jourdain, B., Margheriti, W., and Pammer, G. Stability of the Weak Martingale Optimal Transport Problem. *ArXiv Mathematics e-prints* (2021b). arXiv: 2109.06322.

Beiglböck, M. and Juillet, N. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, **44** (1), 42–106 (2016). MR3456332.

Brückerhoff, M. and Juillet, N. Instability of Martingale optimal transport in dimension $d \ge 2$. *ArXiv Mathematics e-prints* (2021). arXiv: 2101.06964.

Campi, L., Laachir, I., and Martini, C. Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.*, **21** (2), 471–486 (2017). MR3626622.

De March, H. Entropic approximation for multi-dimensional martingale optimal transport. *ArXiv Mathematics e-prints* (2018). arXiv: 1812.11104.

De March, H. and Touzi, N. Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Ann. Probab.*, **47** (3), 1726–1774 (2019). MR3945758.

Dolinsky, Y. and Soner, H. M. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields*, **160** (1-2), 391–427 (2014). MR3256817.

Galichon, A., Henry-Labordère, P., and Touzi, N. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, **24** (1), 312–336 (2014). MR3161649.

Ghoussoub, N., Kim, Y.-H., and Lim, T. Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.*, **47** (1), 109–164 (2019). MR3909967.

Gozlan, N., Roberto, C., Samson, P.-M., and Tetali, P. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, **273** (11), 3327–3405 (2017). MR3706606.

Graf, S. and Luschgy, H. *Foundations of quantization for probability distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin (2000). ISBN 3-540-67394-6. MR1764176.

Guo, G., Obłój, J., and Obłój, J. Computational methods for martingale optimal transport problems. *Ann. Appl. Probab.*, **29** (6), 3311–3347 (2019). MR4047982.

Henry-Labordère, P. (Martingale) Optimal Transport and Anomaly Detection with Neural Networks: a Primal-Dual Algorithm. *ArXiv Mathematics e-prints* (2019). arXiv: 1904.04546.

Henry-Labordère, P., Tan, X., and Touzi, N. An explicit martingale version of the one-dimensional Brenier's theorem with full marginals constraint. *Stochastic Process. Appl.*, **126** (9), 2800–2834 (2016). MR3522302.

Henry-Labordère, P. and Touzi, N. An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, **20** (3), 635–668 (2016). MR3519164.

Jourdain, B. and Margheriti, W. A new family of one dimensional martingale couplings. *Electron. J. Probab.*, **25**, Paper No. 136, 50 (2020). MR4171389.

Kallenberg, O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York (1997). ISBN 0-387-94957-7. MR1464694.

Kieffer, J. C. Exponential rate of convergence for Lloyd's method. I. *IEEE Trans. Inform. Theory*, **28** (2), 205–210 (1982). MR651815.

Pagès, G. Introduction to vector quantization and its applications for numerics. In *CEMRACS 2013—modelling and simulation of complex systems: stochastic and deterministic approaches*, volume 48 of *ESAIM Proc. Surveys*, pp. 29–79. EDP Sci., Les Ulis (2015). MR3415387.

Pagès, G. *Numerical probability: An introduction with applications to finance*. Universitext. Springer, Cham (2018). ISBN 978-3-319-90274-6; 978-3-319-90276-0. MR3823785.

Pagès, G. and Sagna, A. Improved error bounds for quantization based numerical schemes for BSDE and nonlinear filtering. *Stochastic Process. Appl.*, **128** (3), 847–883 (2018). MR3758340.

Pagès, G. and Wilbertz, B. Intrinsic stationarity for vector quantization: foundation of dual quantization. *SIAM J. Numer. Anal.*, **50** (2), 747–780 (2012a). MR2914285.

Pagès, G. and Wilbertz, B. Intrinsic stationarity for vector quantization: foundation of dual quantization. *SIAM J. Numer. Anal.*, **50** (2), 747–780 (2012b). MR2914285.

Pagès, G. and Wilbertz, B. Optimal Delaunay and Voronoi quantization schemes for pricing American style options. In *Numerical methods in finance*, volume 12 of *Springer Proc. Math.*, pp. 171–213. Springer, Heidelberg (2012c). MR3287772.

Pagès, G. and Wilbertz, B. Sharp rate for the dual quantization problem. In *Séminaire de Probabilités XLIX*, volume 2215 of *Lecture Notes in Math.*, pp. 405–454. Springer, Cham (2018). MR3838870.

Rudin, W. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition (1987). ISBN 0-07-054234-1. MR924157.

Strassen, V. The existence of probability measures with given marginals. *Ann. Math. Statist.*, **36**, 423–439 (1965). MR177430.

Trushkin, A. V. Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions. *IEEE Trans. Inform. Theory*, **28** (2), 187–198 (1982). MR651813.

Villani, C. *Optimal transport: Old and new*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin (2009). ISBN 978-3-540-71049-3. MR2459454.

Wiesel, J. Continuity of the martingale optimal transport problem on the real line. *ArXiv Mathematics e-prints* (2012). arXiv: 1905.04574.