Probability measure characterization by L^p-quantization error function

Yating Liu* and Gilles Pagès*,

*Sorbonne Université, CNRS,
Laboratoire de Probabilités, Statistiques et Modélisations (LPSM),
4 place Jussieu, 75252 PARIS Cedex 05, FRANCE.
yating.liu@lpsm.paris and gilles.pages@sorbonne-universite.fr

Abstract

We establish conditions to characterize probability measures by their L^p-quantization error functions in both $\mathbb{R}^d$ and Hilbert settings. This characterization is two-fold: static (identity of two distributions) and dynamic (convergence for the L^p-Wasserstein distance). We first propose a criterion on the quantization level N, valid for any norm on $\mathbb{R}^d$ and any order p based on a geometrical approach involving the Voronoï diagram. Then, we prove that in the L^2-case on a (separable) Hilbert space, the condition on the level N can be reduced to $N = 2$, which is optimal. More quantization based characterization cases in dimension 1 and a discussion of the completeness of a distance defined by the quantization error function can be found at the end of this paper.

Keywords: Probability distribution characterization, Vector quantization, Voronoï diagram, Wasserstein convergence

1 Introduction

Vector quantization was originally developed as an optimal discretization method for signal transmission and compression by the Bell laboratories in the 1950s. Many seminal and historical contributions on vector quantization and its connections with information theory were gathered and published later in [IEE82]. In the unsupervised learning area, vector quantization has a close connection with the automatic classification (clustering) through the k-means algorithm. More recently, in the 1990s, it became an efficient tool in numerical probability to compute regular and conditional expectations (see [Pag98], [BP03] and [PP03]) with in view the pricing of derivative products. Thus, a quantization based numerical schemes have been developed for American option pricing (see [BPP05]), and for the simulation of Backward Stochastic Differential Equation or nonlinear filtering (see [PS18]). For a first mathematically rigorous monograph of various aspects of vector quantization theory, we refer to [GL00] (and the references therein). For more engineering applications to signal compression, see e.g. [GG12] among an extensive literature.

In all these applications, either of probabilistic or statistical nature, vector quantization is used to produce a kind of skeleton of a probability distribution. To be more precise, let $(\Omega, \mathcal{A}, \mathbb{P})$ denote a probability space and let X be a random variable defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and valued in $(E, |\cdot|_E)$, where E is $\mathbb{R}^d$ or a separable Hilbert space H and $|\cdot|_E$ denotes respectively the norm on $\mathbb{R}^d$ or the norm on H induced by the inner product $(\cdot\,|\,\cdot)_H$. Let μ denote the probability distribution of X, denoted by $\mathbb{P}_X = \mu$ or $\mathrm{Law}(X) = \mu$ and assume that μ has a finite p-th moment, $p \in [1, +\infty)$. The quantization grid (also called *codebook* in signal compression or *cluster center* in machine learning theory) is a finite set of points in E, denoted by $\Gamma = \{x_1, ..., x_N\} \subset E$. Let us define the distance between a point ξ and a set A in E by $d(\xi, A) = \min_{a\in A} |\xi - a|_E$. The L^p-mean quantization error of Γ, defined by $e_p(\mu, \Gamma) := \|d(X, \Gamma)\|_p = \Big[\int_E \min_{a\in\Gamma} |\xi - a|_E^p \, \mu(d\xi)\Big]^{\frac{1}{p}}$, is used to describe the accuracy level of representing the probability measure μ by Γ. Let $N \geq 1$. A quantization grid $\Gamma^{*,(N)}$ satisfying

$$e_p(\mu, \Gamma^{*,(N)}) = \inf_{\substack{\Gamma\subset E,\\ \mathrm{card}(\Gamma)\leq N}} \Big[\mathbb{E}\, d(X, \Gamma)^p\Big]^{\frac{1}{p}} = \inf_{\substack{\Gamma\subset E,\\ \mathrm{card}(\Gamma)\leq N}} \Big[\int_E \min_{a\in\Gamma} |\xi - a|_E^p \, \mu(d\xi)\Big]^{\frac{1}{p}} \tag{1.1}$$

is called an *L^p-optimal quantization grid* (or *optimal grid* in short) at *level N*. We refer to [GL00][Theorem 4.12] for the existence of such an optimal grid on $\mathbb{R}^d$ and to [LP02][Proposition 2.1] or [CM88] on (separable) Hilbert spaces. There is usually no closed form for optimal grids, however, in the quadratic case ($p = 2$), it can be computed by the stochastic optimization methods such as the CLVQ algorithm or the randomized Lloyd algorithm (see [Pag15][Section 3], [Kie82] and [PY16]).

Optimal grids $\Gamma^{*,(N)}$ "carries" the information of the initial measure. For example, let $\mu \in \mathcal{P}_{p+\varepsilon}(\mathbb{R}^d)$ for some $\varepsilon > 0$, where $\mathcal{P}_p(E) := \{\mu$ probability distribution on E s.t. $\int_E |\xi|_E^p \, \mu(d\xi) < +\infty\}$. Let $\mu = h \cdot \lambda_d$ be an absolutely continuous distribution (λ_d denotes Lebesgue measure). If for every level $N \geq 1$, $\Gamma^{*,(N)}$ is an optimal quantization grid of μ at level N, then

$$\frac{1}{N} \sum_{x\in\Gamma^{*,(N)}} \delta_x \overset{(\mathbb{R}^d)}{\Longrightarrow} \widetilde{\mu} = \frac{h^{d/(d+p)}(\xi)}{\int h^{d/(d+p)} d\lambda_d} \lambda_d(d\xi), \quad \text{as} \quad N \to +\infty, \tag{1.2}$$

where, for a Polish space S, $\overset{(S)}{\Longrightarrow}$ denotes the weak convergence of probability measures on S. We refer to [GL00][Theorem 7.5] for a proof of this result. This weak convergence (1.2) emphasizes that, an absolutely continuous probability measure μ is entirely characterized by the sequence of L^p-optimal quantization grids $\Gamma^{*,(N)}$ at levels N, $N \geq 1$.

We consider now the L^p-mean quantization error function as follows.

Definition 1.1 (Quantization error function)**.** *Let $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, $p \in [1, +\infty)$. The L^p-mean quantization error function of μ at level N, denoted by $e_{N,p}(\mu, \cdot)$, is defined by:*

$$\begin{array}{rccc} e_{N,p}(\mu,\cdot): & (\mathbb{R}^d)^N & \longrightarrow & \mathbb{R}_+ \\ & x = (x_1, \ldots, x_N) & \longmapsto & e_{N,p}(\mu, x) = \Big[\displaystyle\int_{\mathbb{R}^d} \min_{1\leq i\leq N} |\xi - x_i|^p \, \mu(d\xi)\Big]^{\frac{1}{p}}. \end{array} \tag{1.3}$$

The definition of $e_{N,p}(\mu, \cdot)$ obviously depends on the associated norm on $\mathbb{R}^d$ and the variable of $e_{N,p}(\mu, \cdot)$ is a priori an N-tuple in $(\mathbb{R}^d)^N$. However, for a finite grid $\Gamma \subset \mathbb{R}^d$, if the level $N \geq \mathrm{card}(\Gamma)$, then for any N-tuple $x^\Gamma = (x_1^\Gamma, \ldots, x_N^\Gamma) \in (\mathbb{R}^d)^N$ such that $\Gamma = \{x_1^\Gamma, \ldots, x_N^\Gamma\}$, we have $e_p(\mu, \Gamma) = e_{N,p}(\mu, x^\Gamma)$. For example, $e_p\big(\mu, \{x_1, x_2\}\big) = e_{2,p}\big(\mu, (x_1, x_2)\big) = e_{3,p}\big(\mu, (x_1, x_1, x_2)\big)$,

etc. Note that $e_{N,p}$ is a symmetric function on $(\mathbb{R}^d)^N$ and that, owing to the above definition,

$$\inf_{\Gamma\subset\mathbb{R}^d,\mathrm{card}(\Gamma)\le N} e_p(\mu,\Gamma) = \inf_{x\in(\mathbb{R}^d)^N} e_{N,p}(\mu,x). \tag{1.4}$$

Therefore, throughout this paper, with a slight abuse of notation, we will also denote the L^p-quantization error at level N for a grid Γ of size at most N by $e_{N,p}(\mu,\Gamma)$.

The equality (1.4) directly shows that the optimal grids are characterized by the L^p-mean quantization error functions. Next, we show that the quantization error function $e_{N,p}(\mu,\cdot)$ is entirely characterized by the probability distribution μ.

Notice that for any $\mu\in\mathcal{P}_p(\mathbb{R}^d)$, the function $e_{N,p}(\mu,\cdot)$ defined in (1.3) is 1-Lipschitz continuous for every $N\ge 1$ since for any $x=(x_1,\dots,x_N), y=(y_1,\dots,y_N)\in(\mathbb{R}^d)^N$,

$$\begin{aligned}
|e_{N,p}(\mu,x)-e_{N,p}(\mu,y)| &= \left|\Big[\int_{\mathbb{R}^d}\min_{1\le i\le N}|\xi-x_i|^p\,\mu(d\xi)\Big]^{\frac1p}-\Big[\int_{\mathbb{R}^d}\min_{1\le j\le N}|\xi-y_j|^p\,\mu(d\xi)\Big]^{\frac1p}\right|\\
&\le \Big[\int_{\mathbb{R}^d}\Big|\min_{1\le i\le N}|\xi-x_i|-\min_{1\le j\le N}|\xi-y_j|\Big|^p\mu(d\xi)\Big]^{\frac1p}\quad\text{(by the Minkowski inequality)}\\
&\le \Big[\int_{\mathbb{R}^d}\max_{1\le i\le N}|x_i-y_i|^p\,\mu(d\xi)\Big]^{\frac1p}=\max_{1\le i\le N}|x_i-y_i|\,.
\end{aligned}\tag{1.5}$$

We recall now the definition of the L^p-Wasserstein distance.

Definition 1.2 (L^p-Wasserstein distance)**.** *Let (S,d) be a Polish space and $\mathcal{S}=Bor(S,d)$ be its Borel σ-field. For $p\in[1,+\infty)$, let $\mathcal{P}_p(S)$ denote the set of probability measures on $(S,\mathcal{S})$ with a finite p^{th}-moment. The L^p-Wasserstein distance $\mathcal{W}_p(\mu,\nu)$ between $\mu,\nu\in\mathcal{P}_p(S)$, denoted by $\mathcal{W}_p(\mu,\nu)$, is defined by*

$$\begin{aligned}
\mathcal{W}_p(\mu,\nu) &= \Big(\inf_{\pi\in\Pi(\mu,\nu)}\int_{S\times S}d(x,y)^p\pi(dx,dy)\Big)^{\frac1p}\\
&= \inf\Big\{\Big[\mathbb{E}\,d(X,Y)^p\Big]^{\frac1p},\ X,Y:(\Omega,\mathcal{A},\mathbb{P})\to(S,\mathcal{S})\ \textit{with}\ \mathbb{P}_X=\mu,\mathbb{P}_Y=\nu\Big\},
\end{aligned}\tag{1.6}$$

where in the first line of (1.6), $\Pi(\mu,\nu)$ denotes the set of all probability measures on $(S^2,\mathcal{S}^{\otimes2})$ with respective marginals μ and ν.

If we consider $e_{N,p}(\mu,x)$ as a function of $\mu\in\mathcal{P}_p(\mathbb{R}^d)$, then $e_{N,p}$ is also 1-Lipschitz in μ. In fact, let X,Y be two random variables with probability distributions $\mathbb{P}_X=\mu$ and $\mathbb{P}_Y=\nu$. For every N-tuple $x=(x_1,\dots,x_N)\in(\mathbb{R}^d)^N$, we have

$$\begin{aligned}
\big|e_{N,p}(\mu,x)-e_{N,p}(\nu,x)\big| &= \Big|\ \Big\|\min_{i=1,\dots,N}|X-x_i|\Big\|_p-\Big\|\min_{i=1,\dots,N}|Y-x_i|\Big\|_p\ \Big|\\
&\le \Big\|\min_{i=1,\dots,N}|X-x_i|-\min_{i=1,\dots,N}|Y-x_i|\Big\|_p\text{(by the Minkowski inequality)}\\
&\le \Big\|\max_{i=1,\dots,N}|\ |X-x_i|-|Y-x_i|\ |\Big\|_p\le\|X-Y\|_p\,.
\end{aligned}\tag{1.7}$$

As this inequality holds for every couple (X,Y) of random variables with marginal distributions

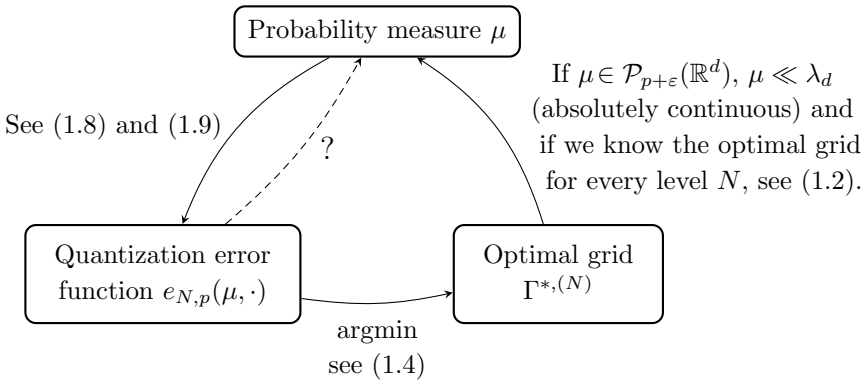
μ and ν, it follows that for every level $N \geq 1$,

$$\|e_{N,p}(\mu,\cdot) - e_{N,p}(\nu,\cdot)\|_{\sup} := \sup_{x\in(\mathbb{R}^d)^N} |e_{N,p}(\mu,x) - e_{N,p}(\nu,x)| \leq \mathcal{W}_p(\mu,\nu). \tag{1.8}$$

Hence, if $(\mu_n)_{n\geq 1}$ is a sequence in $\mathcal{P}_p(\mathbb{R}^d)$ converging for the $\mathcal{W}_p$-distance to $\mu_\infty \in \mathcal{P}_p(\mathbb{R}^d)$, then

$$\|e_{N,p}(\mu_n,\cdot) - e_{N,p}(\mu_\infty,\cdot)\|_{\sup} \leq \mathcal{W}_p(\mu_n,\mu_\infty) \xrightarrow{n\to+\infty} 0. \tag{1.9}$$

Definition 1.1, and the inequalities (1.5), (1.7), (1.8), (1.9) can be directly extended to any separable Hilbert space H. Inequalities (1.8) and (1.9) show that for every $N \geq 1$, and $p\in[1,+\infty)$, the quantization error function $e_{N,p}(\mu,\cdot)$ is characterized by the probability distribution μ. Hence, the characterization relations between a probability measure μ, its L^p-quantization error function and its optimal grids can be synthesized by the following scheme:



The characterization of a probability measure μ by its L^p-optimal quantization grids suggests to consider the "reverse" questions of (1.8) and (1.9): *When is a probability measure $\mu\in\mathcal{P}_p(\mathbb{R}^d)$ characterized by its L^p-quantization error function $e_{N,p}(\mu,\cdot)$? And if so, does the convergence in an appropriate sense of the L^p-quantization error functions characterizes the convergence of their probability distributions for the $\mathcal{W}_p$-distance?*

These questions can be formalized as follows (the first one in a slightly extended sense):

- **Question 1 - Static characterization**:
  If for $\mu,\nu\in\mathcal{P}_p(\mathbb{R}^d)$, $e_{N,p}(\mu,\cdot) = e_{N,p}(\nu,\cdot) + C$ for some real constant C, then do we have $\mu = \nu$ (and $C = 0$)?
- **Question 2 - Characterization of $\mathcal{W}_p$-convergence**:
  If for $\mu_n, n\geq 1$, $\mu_\infty\in\mathcal{P}_p(\mathbb{R}^d)$, $e_{N,p}(\mu_n,\cdot)$ converges pointwise to $e_{N,p}(\mu_\infty,\cdot)$, then do we have $\mathcal{W}_p(\mu_n,\mu_\infty) \xrightarrow{n\to+\infty} 0$?

For any $N_1, N_2\in\mathbb{N}^*$ with $N_1 \leq N_2$, it is clear that $e_{N_2,p}(\mu,\cdot) = e_{N_2,p}(\nu,\cdot)$ $\big($resp. $e_{N_2,p}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{N_2,p}(\mu_\infty,\cdot)\big)$ implies $e_{N_1,p}(\mu,\cdot) = e_{N_1,p}(\nu,\cdot)$ $\big($resp. $e_{N_1,p}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{N_1,p}(\mu_\infty,\cdot)\big)$. Hence, beyond these two above questions, we need to determine an as low as possible level N for which both

answers are positive. For this purpose, we define

$$N_{d,p,|\cdot|} \coloneqq \min\{N \in \mathbb{N}^* \text{ such that answers to } \textit{Questions 1} \text{ and } \textit{2} \text{ for } e_{N,p} \text{ are positive}\}. \quad (1.10)$$

The paper is organized as follows. We first recall in Section 1.1 some properties of the Wasserstein distance $\mathcal{W}_p$. Then in Section 2, we begin to analyze the problem of probability distribution characterization in a general finite dimensional framework by considering any dimension d, any order p and any norm on $\mathbb{R}^d$. We show that a positive answer to *Question 1 and 2* follows from the existence of a bounded open Voronoï cell in a Voronoï diagram of size N, which in turn can be derived from a minimal covering of the unit sphere by unit closed balls centered on the sphere. As a consequence, we define for $N \geq N_{d,p,|\cdot|}$ a quantization based distance $\mathcal{Q}_{N,p} \coloneqq \|e_{N,p}(\mu,\cdot) - e_{N,p}(\nu,\cdot)\|_{\sup}$ which we will prove to be topologically equivalent to the Wasserstein distance $\mathcal{W}_p$. The results in this section are established for $p \geq 1$, but several results can be extended to the case $0 < p < 1$ by the usual adaptations of the proofs.

In Section 3, we consider the quadratic case (*i.e.* the order p=2) and extend the characterization result to probability distributions on a separable Hilbert space H with the norm $|\cdot|_H$ induced by the inner product $(\cdot \mid \cdot)_H$. In this section, we will prove by a purely analytical method that $N_{H,2,|\cdot|_H} = 2$ [1] and the topological equivalence of Wasserstein distance $\mathcal{W}_2$ and the distance $\mathcal{Q}^H_{2,2}(\mu,\nu) \coloneqq \|e_{2,2}(\mu,\cdot) - e_{2,2}(\nu,\cdot)\|_{\sup}$ on $\mathcal{P}_2(H)$.

Section 4 is devoted to the one-dimensional setting. Quantization based characterization not yet covered by the discussion in Section 2 and Section 3 are established. Furthermore, we prove that $\mathcal{Q}_{1,1}$ is a complete distance on $\mathcal{P}_1(\mathbb{R})$ and give a counterexample to show that the distances $\mathcal{Q}_{N,2}$, $N \geq 2$ are not complete on $\mathcal{P}_2(\mathbb{R})$ in Section 4.2.

1.1 Preliminaries on Wasserstein distance

Let (S,d) be a general Polish metric space. The relation between weak convergence and convergence for the Wasserstein distance $\mathcal{W}_p$ (see Definition 1.2) is recalled in Theorem 1.1. We recall below some useful facts about the L^p-Wasserstein distance that will be called upon further on. The first one is that, for every $p \in [1,+\infty)$, $\mathcal{W}_p$ is a distance on $\mathcal{P}_p(S)$ $\big(\mathcal{W}_p^p$ if $p \in (0,1)\big)$, see e.g. [Vil03][Theorem 7.3] for the proof and [BPR15] for a recent reference. Next, the metric space $\big(\mathcal{P}_p(S), \mathcal{W}_p\big)$ is separable and complete, see e.g. [Bol08] for the proof. More generally, we refer to [Vil09][Chapter 6] for an in depth presentation of Wasserstein distance and its properties.

Theorem 1.1. *(see [Vil03][Theorem 7.12]) Let $\mu_n \in \mathcal{P}_p(S)$ for every $n \in \mathbb{N}^* \cup \{\infty\}$. Let $p \in [1,+\infty)$. Then,*

(a) $\mathcal{W}_p(\mu_n,\mu_\infty) \to 0$ *if and only if* $\begin{cases} (\alpha)\ \mu_n \xrightarrow{(S)} \mu_\infty \\ (\beta)\ \exists\, x_0 \in S, \int_S d(x_0,\xi)^p \mu_n(d\xi) \to \int_S d(x_0,\xi)^p \mu_\infty(d\xi) \end{cases}$.

(b) If

$$\exists\, x_0 \in S, \quad \lim_{R\to+\infty} \sup_{n\geq 1} \int_{d(x_0,\xi)^p \geq R} d(x_0,\xi)^p \mu_n(d\xi) = 0, \quad (1.11)$$

then $(\mu_n)_{n\geq 1}$ is relatively compact for the Wasserstein distance $\mathcal{W}_p$.

[1] Since the dimension of the Hilbert space that we discuss in this section can be finite or infinite, we write directly H instead of d in the subscript of $N_{d,p,|\cdot|}$.

2 General quantization based characterizations on $\mathbb{R}^d$

This section is devoted to establishing a general criterion that positively answers to *Questions 1 and 2* in any dimension d, for any order p and any norm on $\mathbb{R}^d$. The idea is to design an approximate identity $(\varphi_\varepsilon)_{\varepsilon>0}$[2] based on the quantization error function $e_{N,p}(\mu,\cdot)$. Our construction of $(\varphi_\varepsilon)_{\varepsilon>0}$ relies on a purely geometrical idea: it is based on a specified Voronoï diagram containing a bounded open Voronoï cell that we introduce in Section 2.1. The static characterization is established in Theorem 2.1. Furthermore, Theorem 2.2 shows that a pointwise convergence of the quantization error functions is enough to imply the $\mathcal{W}_p$-convergence of a $\mathcal{P}_p(\mathbb{R}^d)$-valued sequence.

2.1 A review of Voronoï diagram, existence of bounded cells

Let $\Gamma=\{x_1,\ldots,x_N\}$ be a grid of size N. The *Voronoï cell* generated by $x_i\in\Gamma$ is defined by

$$V_{x_i}(\Gamma)=\big\{\xi\in\mathbb{R}^d:|\xi-x_i|=\min_{1\le j\le N}|\xi-x_j|\,\big\},\tag{2.1}$$

and $\big(V_{x_i}(\Gamma)\big)_{1\le i\le N}$ is called the *Voronoï diagram* of Γ, which is a finite covering of $\mathbb{R}^d$ (see [GL00]). A Borel measure partition $\big(C_{x_i}(\Gamma)\big)_{1\le i\le N}$ is called a *Voronoï partition* of $\mathbb{R}^d$ induced by Γ if for every $i\in\{1,\ldots,N\}$, $C_{x_i}(\Gamma)\subset V_{x_i}(\Gamma)$. We also define the *open Voronoï cell* generated by $x_i\in\Gamma$ by

$$V^o_{x_i}(\Gamma)=\big\{\xi\in\mathbb{R}^d:|\xi-x_i|<\min_{1\le j\le N,j\neq i}|\xi-x_j|\,\big\}.\tag{2.2}$$

If the norm $|\cdot|$ on $\mathbb{R}^d$ is strictly convex, we have $\mathring{V}_{x_i}(\Gamma)=V^o_{x_i}(\Gamma)$ and $\overline{V^o_{x_i}(\Gamma)}=V_{x_i}(\Gamma)$, where $\mathring{A}$ and $\overline{A}$ denote the interior and the closure of A. Examples of strictly convex norms are the isotropic ℓ_r-norms for $1<r<+\infty$ defined by $\big|(a^1,\ldots,a^d)\big|_r=\big(\,|a^1|^r+\cdots+|a^d|^r\,\big)^{1/r}$. However, this is not true for any norm on $\mathbb{R}^d$, typically not for the ℓ^1-norm (see [GL00][Figure 1.2]) or the $\ell^\infty-$norm.

We recall that $A\subset\mathbb{R}^d$ is star-shaped with respect to $a\in A$ if for every $b\in A$ and any $\lambda\in[0,1]$, $a+\lambda(b-a)\in A$.

Proposition 2.1. *(see [GL00][Proposition 1.2]) Let $\Gamma=\{x_1,\ldots,x_N\}$ be a grid of size $N\ge1$. For every $i\in\{1,\ldots,N\}$, $V_{x_i}(\Gamma)$ and $V^o_{x_i}(\Gamma)$ are star-shaped relative to x_i.*

Now we discuss a sufficient condition to obtain a Voronoï diagram containing a bounded open Voronoï cell. The first result in this direction is a rewriting Proposition 1.10 in [GL00] for Euclidean norms (stated here in view of our applications).

Proposition 2.2 ($|\cdot|$ Euclidean norm)**.** *Let $(b_1,\ldots,b_{d+1})$ be an affine basis of $\mathbb{R}^d$ and let $b_0\in\mathring{\mathrm{Conv}}(\{b_1,\ldots,b_{d+1}\})\neq\varnothing$. Set $\Gamma=\{0,b_1-b_0,\ldots,b_{d+1}-b_0\}$. Then, the open Voronoï cell $V_0^o(\Gamma)$ generated by 0 is bounded.*

Let us provide now a geometrical criterion for a general norm $|\cdot|$ on $\mathbb{R}^d$, let $\bar B_{|\cdot|}(x,r)$ denote the closed ball centered at x with radius r and let $S_{|\cdot|}(x,r)$ denote its sphere.

[2] By approximate identity we mean $\varphi_\varepsilon\in L^1(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d),\lambda_d)$, $\varepsilon>0$, such that $\int_{\mathbb{R}^d}\varphi_\varepsilon d\lambda_d=1$, $\sup_{\varepsilon>0}\int_{\mathbb{R}^d}|\varphi_\varepsilon|\,d\lambda_d<+\infty$ and $\lim_{\varepsilon\to0}\int_{\{|\xi|>\eta\}}\varphi_\varepsilon(\xi)\lambda_d(\xi)=0$ for every $\eta>0$.

Proposition 2.3. *Let $a_1, \dots, a_k \in S_{|\cdot|}(0,1)$ such that $S_{|\cdot|}(0,1) \subset \bigcup_{i=1}^k \bar{B}_{|\cdot|}(a_i, 1)$ (such a covering exists since $S_{|\cdot|}(0,1)$ is compact). If we choose $\Gamma = \{0, a_1, \dots, a_k\}$, then the Voronoï open set $V_0^o(\Gamma) \subset \bar{B}_{|\cdot|}(0,1)$ and $\lambda_d\big(V_0^o(\Gamma)\big) > 0$.*

Proof. As $S_{|\cdot|}(0,1) \subset \bigcup_{i=1}^k \bar{B}_{|\cdot|}(a_i, 1)$, for every $\xi \in S_{|\cdot|}(0,1)$, there exists $j \in \{1, \dots, k\}$ such that $|\xi - a_j| \leq 1 = |\xi|$. If $\Gamma = \{0, a_1, \dots, a_k\}$, then

$$\forall \xi \in S_{|\cdot|}(0,1), \quad \exists\, j \in \{1, \dots, k\} \text{ such that } \xi \in V_{a_j}(\Gamma). \tag{2.3}$$

Assume that there exists $\xi \in V_0^o(\Gamma) \setminus \bar{B}_{|\cdot|}(0,1)$. Since $V_0^o(\Gamma)$ is star-shaped relatively to 0 and $\frac{1}{|\xi|} \in (0,1)$, we have $\frac{\xi}{|\xi|} \in S_{|\cdot|}(0,1) \cap V_0^o(\Gamma)$. This contradicts (2.3) since $V_0^o(\Gamma) \cap V_{a_j}(\Gamma) \neq \varnothing$, $j = 1, \dots, k$. Consequently, $V_0^o(\Gamma) \subset \bar{B}_{|\cdot|}(0,1)$. Finally, $V_0^o(\Gamma)$ is an open set containing 0, therefore, $\lambda_d\big(V_0^o(\Gamma)\big) > 0$. □

The idea of the above proposition is to cover the unit sphere centered at the origin by a finite number of unit balls centered on the unit sphere. This leads us to introduce the following definition.

Definition 2.1. *We define the* minimal sphere covering number $c(d, |\cdot|)$ *as follows,*

$$c(d, |\cdot|) \coloneqq \min\Big\{k : \exists \{a_1, \dots, a_k\} \subset S_{|\cdot|}(0,1) \text{ such that } S_{|\cdot|}(0,1) \subset \bigcup_{i=1}^k \bar{B}_{|\cdot|}(a_i, 1)\Big\} < +\infty.$$

The index $c(d, |\cdot|)$ is finite since the unit sphere is a compact set in $\mathbb{R}^d$. Among all the possible norms, we will focus on the isotropic ℓ_r-norms on $\mathbb{R}^d$. We show some examples of the minimal covering number $c(d, |\cdot|_r)$ in the following proposition (whose proof is postponed to Appendix).

Proposition 2.4. *(i)* $c(1, |\cdot|) = 2$, *where $|\cdot|$ denotes the absolute value.*

(ii) $c(2, |\cdot|_1) = 2$ *and* $c(2, |\cdot|_r) = 3$ *for every* $1 < r < +\infty$.

(iii) $c(d, |\cdot|_\infty) = 2$ *for every dimension* d.

(iv) Let $r \geq 1$ *such that* $2^r \geq d$, *then* $c(d, |\cdot|_r) \leq 2d$.

2.2 A general condition for probability measure characterization

Let $\Gamma = \{x_1, \dots, x_N\}$ be a grid in which there exists at least an $x_{i_0} \in \Gamma$ such that the open Voronoï cell $V_{x_{i_0}}^o(\Gamma)$ is bounded and non-empty. Based on such a grid, one can construct an approximate identity as follows. Let $\varphi : \mathbb{R}^d \to \mathbb{R}_+$ be the function defined by $\varphi(\xi) = \min_{a \in \Gamma \setminus \{x_{i_0}\}} |\xi - a|^p - \min_{a \in \Gamma} |\xi - a|^p$. The function φ is clearly nonnegative, continuous and $\{\varphi > 0\} = V_{x_{i_0}}^o(\Gamma)$ so that $\mathrm{supp}(\varphi) = \overline{V_{x_{i_0}}^o(\Gamma)}$ is compact. Hence, $\int \varphi \, d\lambda_d \in (0, +\infty)$ since $\varphi(x_{i_0}) = d\big(x_{i_0}, \Gamma \setminus \{x_{i_0}\}\big) > 0$ and we can normalize φ by setting $\varphi_1(\xi) \coloneqq \frac{\varphi(x_{i_0} + \xi)}{\int \varphi d\lambda_d}$. For every $\varepsilon > 0$, we define $\varphi_\varepsilon(\xi) \coloneqq \frac{1}{\varepsilon^d} \varphi_1\Big(\frac{\xi}{\varepsilon}\Big)$, then $(\varphi_\varepsilon)_{\varepsilon > 0}$ is clearly an approximate identity (see [Gra14][Section 1.2.4]).

The following theorem gives conditions on the L^p-quantization error function to characterize a probability measure.

Theorem 2.1 (Static characterization)**.** *Let $p \in [1,+\infty)$, let $|\cdot|$ be a norm on $\mathbb{R}^d$ and let $N \geq c(d,|\cdot|)+1$, or $N \geq d+2$ if $|\cdot|$ is Euclidean. Then, the answer to Question 1 is positive i.e. if there exists a constant C such that $e^p_{N,p}(\mu,\cdot) = e^p_{N,p}(\nu,\cdot) + C$, $\mu,\nu \in \mathcal{P}_p(\mathbb{R}^d)$, then $\mu = \nu$. The constant C is a posteriori 0.*

Proof. Following Proposition 2.2 and 2.3, we choose a grid $\Gamma = \{0, a_1, \ldots, a_{N-1}\}$ such that $V_0^o(\Gamma)$ is bounded and $\lambda_d\big(V_0^o(\Gamma)\big) > 0$. We define $\varphi : \mathbb{R}^d \to \mathbb{R}_+$, by $\varphi(\xi) = \min_{a\in\Gamma\setminus\{0\}} |\xi - a|^p - \min_{a\in\Gamma} |\xi - a|^p = \big(\min_{a\in\Gamma\setminus\{0\}} |\xi-a|^p - |\xi|^p\big)_+$ and $(\varphi_\varepsilon)_{\varepsilon>0}$ by $\varphi_\varepsilon(\xi) \coloneqq \frac{1}{C_\varphi \varepsilon^d}\varphi\Big(\frac{\xi}{\varepsilon}\Big)$, where $C_\varphi = \int \varphi\, d\lambda_d$. For any $x \in \mathbb{R}^d$,

$$\begin{aligned}
\varphi_\varepsilon * \mu(x) &= \int_{\mathbb{R}^d} \varphi_\varepsilon(x-\xi)\mu(d\xi) = \int_{\mathbb{R}^d} \frac{1}{\varepsilon^d} \frac{\varphi\big(\frac{x-\xi}{\varepsilon}\big)}{\int \varphi d\lambda_d} \mu(d\xi) \\
&= \frac{1}{C_\varphi \varepsilon^d} \int_{\mathbb{R}^d} \Big(\min_{a\in\Gamma\setminus\{0\}} \Big|\frac{x-\xi}{\varepsilon} - a\Big|^p - \min_{a\in\Gamma} \Big|\frac{x-\xi}{\varepsilon} - a\Big|^p \Big) \mu(d\xi) \\
&= \frac{1}{C_\varphi \varepsilon^{d+p}} \Big[\int_{\mathbb{R}^d} \min_{a\in\Gamma\setminus\{0\}} |x - \varepsilon a - \xi|^p \, \mu(d\xi) - \int_{\mathbb{R}^d} \min_{a\in\Gamma} |x-\varepsilon a - \xi|^p \, \mu(d\xi) \Big].
\end{aligned}$$

If we define two N-tuples $\tilde{x}$ and $\tilde{x}_0$ as $\tilde{x} = (x - \varepsilon a_1, x - \varepsilon a_1, x - \varepsilon a_2, \ldots, x - \varepsilon a_{N-1})$ and $\tilde{x}_0 = (x, x - \varepsilon a_1, x - \varepsilon a_2, \ldots, x - \varepsilon a_{N-1})$, then

$$\int_{\mathbb{R}^d} \min_{a\in\Gamma\setminus\{0\}} |x - \varepsilon a - \xi|^p \, \mu(d\xi) = e^p_{N,p}(\mu,\tilde{x}) \text{ and } \int_{\mathbb{R}^d} \min_{a\in\Gamma} |x - \varepsilon a - \xi|^p \, \mu(d\xi) = e^p_{N,p}(\mu,\tilde{x}_0).$$

Hence, $\varphi_\varepsilon * \mu(x) = \frac{1}{C_\varphi \varepsilon^{d+p}} \big(e^p_{N,p}(\mu,\tilde{x}) - e^p_{N,p}(\mu,\tilde{x}_0)\big)$.

The assumption $e^p_{N,p}(\mu,\cdot) = e^p_{N,p}(\nu,\cdot) + C$ implies that $e^p_{N,p}(\mu,\tilde{x}) - e^p_{N,p}(\mu,\tilde{x}_0) = e^p_{N,p}(\nu,\tilde{x}) - e^p_{N,p}(\nu,\tilde{x}_0)$, so that, for every $x \in \mathbb{R}^d$ and every $\varepsilon > 0$, $\varphi_\varepsilon * \mu(x) = \varphi_\varepsilon * \nu(x)$.

One can finally conclude that $\mu = \nu$ by letting $\varepsilon \to 0$ since $(\varphi_\varepsilon)_{\varepsilon>0}$ is an approximate identity (see [Rud91][Theorem 6.32]). Hence $C = 0$. □

The following theorem shows that the pointwise convergence of the L^p-mean quantization error function is a necessary and sufficient condition for $\mathcal{W}_p$-convergence of probability distributions in $\mathcal{P}_p(\mathbb{R}^d)$.

Theorem 2.2 ($\mathcal{W}_p$-convergence characterization)**.** *Let $p \in [1,+\infty)$ and let $|\cdot|$ be any norm on $\mathbb{R}^d$. Let $\mu_n \in \mathcal{P}_p(\mathbb{R}^d)$ for $n \in \mathbb{N}^* \cup \{\infty\}$. The following properties are equivalent:*

(i) $\mathcal{W}_p(\mu_n,\mu_\infty) \xrightarrow{n\to+\infty} 0$,

(ii) $\forall N \geq 1, \quad e_{N,p}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{N,p}(\mu_\infty,\cdot)$ *uniformly on* $\mathbb{R}^d$,

(iii) $\exists\, N \geq c(d,|\cdot|)+1$ *or* $N \geq d+2$ *if* $|\cdot|$ *is Euclidean such that,* $e_{N,p}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{N,p}(\mu_\infty,\cdot)$ *pointwise on* $\mathbb{R}^d$.

Proof of Theorem 2.2. $(i) \Rightarrow (ii)$ is obvious from (1.9).

$(ii) \Rightarrow (iii)$ is obvious.

$(iii) \Rightarrow (i)$ First of all, it follows from the convergence $e_{N,p}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{N,p}(\mu_\infty,\cdot)$ that

$$e^p_{N,p}(\mu_n,\mathbf{0}) \xrightarrow{n\to+\infty} e^p_{N,p}(\mu_\infty,\mathbf{0}) \text{ i.e. } \int_{\mathbb{R}^d}|\xi|^p\,\mu_n(d\xi) \xrightarrow{n\to+\infty} \int_{\mathbb{R}^d}|\xi|^p\,\mu_\infty(d\xi) < +\infty, \tag{2.4}$$

where $\mathbf{0} = (0,\ldots,0)$. In particular, the sequence $\Big(\int_{\mathbb{R}^d}|\xi|^p\,\mu_n(d\xi)\Big)_{n\geq 1}$ is bounded. Hence, the sequence of probability measures $(\mu_n)_{n\geq 1}$ is tight.

Let $\widetilde{\mu}_\infty$ be a weak limiting probability distribution of $(\mu_n)_{n\geq 1}$ i.e. there exists a subsequence $\alpha(n)$ of n such that $\mu_{\alpha(n)} \overset{(\mathbb{R}^d)}{\Longrightarrow} \widetilde{\mu}_\infty$ as $n\to+\infty$.

Let $x=(x_1,\ldots,x_N)$ be any N-tuple in $(\mathbb{R}^d)^N$. We define a continuous function $f_x:\mathbb{R}^d\to\mathbb{R}$ by $f_x(\xi) := \min_{1\leq i\leq N}|\xi-x_i|^p - |\xi|^p$. Hence, owing to the elementary inequality $v^p-u^p \leq pv^{p-1}(v-u)$ for any $0\leq u\leq v<+\infty$, we derive

$$\big|f_x(\xi)\big| \leq \max_{i\in\{1,\ldots,N\}} p\big(\,|\xi|+|x_i|\,\big)^{p-1}\,|x_i| \leq C_{x,p}(1+|\xi|^{p-1}), \tag{2.5}$$

where $C_{x,p}$ is a constant depending on x and p.

Owing to (2.4) and (2.5), the sequence $\big(\int f_x^{\frac{p}{p-1}}d\mu_n\big)_{n\geq 1}$ is bounded, hence f_x is uniformly integrable with respect to $(\mu_n)_{n\geq 1}$ since $\frac{p}{p-1}>1$, so that f_x is uniformly integrable with respect to any subsequence $(\mu_{\alpha(n)})_{n\geq 1}$. It follows that $\int_{\mathbb{R}^d} f_x(\xi)\mu_{\alpha(n)}(d\xi) \to \int_{\mathbb{R}^d} f_x(\xi)\widetilde{\mu}_\infty(d\xi)$, as $n\to+\infty$, where

$$\int_{\mathbb{R}^d} f_x(\xi)\mu_{\alpha(n)}(d\xi) = \int_{\mathbb{R}^d}\big(\min_{i\in\{1,\ldots,N\}}|\xi-x_i|^p - |\xi|^p\big)\mu_{\alpha(n)}(d\xi) = e^p_{N,p}(\mu_{\alpha(n)},x) - e^p_{N,p}(\mu_{\alpha(n)},\mathbf{0}),$$

$$\text{and } \int_{\mathbb{R}^d} f_x(\xi)\widetilde{\mu}_\infty(d\xi) = e^p_{N,p}(\widetilde{\mu}_\infty,x) - e^p_{N,p}(\widetilde{\mu}_\infty,\mathbf{0}).$$

On the other hand, $e^p_{N,p}(\mu_{\alpha(n)},x) - e^p_{N,p}(\mu_{\alpha(n)},\mathbf{0})$ converges to $e^p_{N,p}(\mu_\infty,x) - e^p_{N,p}(\mu_\infty,\mathbf{0})$ owing to the pointwise convergence in (iii) at $\mathbf{0}=(0,\ldots,0)$ and $x=(x_1,\ldots,x_N)$.

Therefore, $e^p_{N,p}(\widetilde{\mu}_\infty,x) - e^p_{N,p}(\widetilde{\mu}_\infty,\mathbf{0}) = e^p_{N,p}(\mu_\infty,x) - e^p_{N,p}(\mu_\infty,\mathbf{0})$, which implies that for every $x\in(\mathbb{R}^d)^N$, $e^p_{N,p}(\widetilde{\mu}_\infty,x) - e^p_{N,p}(\mu_\infty,x) = C$, where $C = e^p_{N,p}(\widetilde{\mu}_\infty,\mathbf{0}) - e^p_{N,p}(\mu_\infty,\mathbf{0})$ is a real constant. It follows from Theorem 2.1 that $\widetilde{\mu}_\infty = \mu_\infty$, which implies that μ_∞ is the only limiting distribution of $(\mu_n)_{n\geq 1}$ for the weak convergence and consequently $\mu_n \overset{(\mathbb{R}^d)}{\Longrightarrow} \mu$. We have already proved that $\int_{\mathbb{R}^d}|\xi|^p\,\mu_n(d\xi) \xrightarrow{n\to+\infty} \int_{\mathbb{R}^d}|\xi|^p\,\mu_\infty(d\xi)$ from (2.4), which finally shows that $\mathcal{W}_p(\mu_n,\mu_\infty) \xrightarrow{n\to+\infty} 0$ owing to Theorem 1.1. □

A careful reading of the proof shows that the following "à la Paul Lévy" characterization result holds for limiting functions of L^p-quantization error functions.

Corollary 2.1. *Let $p\in[1+\infty)$. Let $(\mu_n)_{n\geq 1}$ be a $\mathcal{P}_p(\mathbb{R}^d)$-valued sequence. If*

$$e_{N,p}(\mu_n,\cdot) \xrightarrow{n\to+\infty} f \quad \textit{pointwise for some } N \textit{ such that } \text{static characterization } \textit{holds true}$$

(Question 1), then there exists $\mu_\infty\in\mathcal{P}_p(\mathbb{R}^d)$ such that $\mu_n \overset{(\mathbb{R}^d)}{\Longrightarrow} \mu_\infty$ as $n\to+\infty$ and

$$f^p = e^p_{N,p}(\mu_\infty,\,\cdot\,) + \lim_n \int_{\mathbb{R}^d}|\xi|^p\,\mu_n(d\xi) - \int_{\mathbb{R}^d}|\xi|^p\,\mu_\infty(d\xi).$$

Now we will take advantage of what precedes to introduce a *quantization based distance* on $\mathcal{P}_p(\mathbb{R}^d)$. Let $\mathcal{C}_b\big((\mathbb{R}^d)^N, \mathbb{R}\big)$ denote the space of bounded $\mathbb{R}$-valued continuous functions defined on $(\mathbb{R}^d)^N$ equipped with the sup norm $\|\cdot\|_{\sup}$. Let $p \in [1, +\infty)$. If $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, $e_{N,p}(\mu, \cdot) - e_{N,p}(\delta_0, \cdot) \in \mathcal{C}_b\big((\mathbb{R}^d)^N, \mathbb{R}\big)$ (note that $e_{N,p}\big(\delta_0, (x_1, \ldots, x_N)\big) = \min_{i=1,\ldots,N} |x_i|$) since inequality (1.8) implies that $\|e_{N,p}(\mu, \cdot) - e_{N,p}(\delta_0, \cdot)\|_{\sup} \leq \mathcal{W}_p(\mu, \delta_0) = \Big[\int_{\mathbb{R}^d} |\xi|^p \, \mu(d\xi)\Big]^{1/p} < +\infty$. Then, we define a function $\mathcal{Q}_{N,p}$ on $\mathcal{P}_p(\mathbb{R}^d)$ by

$$\begin{aligned}(\mu, \nu) \longmapsto \mathcal{Q}_{N,p}(\mu, \nu) &:= \big\|\big(e_{N,p}(\mu, \cdot) - e_{N,p}(\delta_0, \cdot)\big) - \big(e_{N,p}(\nu, \cdot) - e_{N,p}(\delta_0, \cdot)\big)\big\|_{\sup} \\ &= \|e_{N,p}(\mu, \cdot) - e_{N,p}(\nu, \cdot)\|_{\sup}. \end{aligned} \tag{2.6}$$

For any $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, inequality (1.8) implies $\mathcal{Q}_{N,p}(\mu, \nu) \leq \mathcal{W}_p(\mu, \nu) < +\infty$ so that $\mathcal{Q}_{N,p}(\mu, \nu) \in [0, +\infty)$. Combining Theorems 2.1 and 2.2 implies the following result.

Corollary 2.2. *Let $p \in [1, +\infty)$.*

(a) $N_{d,p,|\cdot|} \leq c(d, |\cdot|) + 1$ for any norm and $N_{d,p,|\cdot|} \leq d + 2$ if $|\cdot|$ is Euclidean.

(b) If $N \geq c(d, |\cdot|) + 1$ or $N \geq d + 2$ if $|\cdot|$ is Euclidean, then $\mathcal{Q}_{N,p}$ defined by (2.6) is a distance on $\mathcal{P}_p(\mathbb{R}^d)$ and $\mathcal{Q}_{N,p}$ is topologically equivalent to the Wasserstein distance $\mathcal{W}_p$.

Comments on optimality. If we consider only the quadratic case $p = 2$ and a norm $|\cdot|$ induced by an inner product, the result in Corollary 2.2-(a) is in fact not optimal. In the next section, we will prove that in such a setting, $N_{d,2,|\cdot|} = 2$ and this result can also be extended to any separable (possibly infinite-dimensional) Hilbert space.

3 Quadratic quantization based characterization on a separable Hilbert space: $N_{H,2} = 2$

Let H denote a separable Hilbert space with the inner product $(\cdot \mid \cdot)_H$. Let $|\cdot|_H$ denote the norm on H induced by $(\cdot \mid \cdot)_H$. When there is no ambiguity, we drop the index $_H$ and write $(\cdot \mid \cdot)$ and $|\cdot|$. The separable Hilbert space is a very common setup for applications, for example in functional data analysis: one can set $H = L^2\big([0, T], dt\big)$ and $X = (X_t)_{t \in [0,T]}$ a bi-measurable process such that $\int_0^T \mathbb{E} X_t^2 \, dt < +\infty$. For more information about functional data analysis with an L^2-setup, we refer to [HE15] among others.

We first prove in the quadratic case ($p = 2$), that both static (see further Proposition 3.1) and $\mathcal{W}_2$-convergence (see further Theorem 3.1) characterizations can be obtained at level $N = 2$ by an analytical method. Then we will show that $N_{H,2} := N_{H,2,|\cdot|_H} = 2$ and for any $\mu, \nu \in \mathcal{P}_2(H)$, $\mathcal{Q}_{2,2}(\mu, \nu) := \|e_{2,2}(\mu, \cdot) - e_{2,2}(\nu, \cdot)\|_{\sup}$ is a well-defined distance on $\mathcal{P}_2(H)$ which is topologically equivalent to $\mathcal{W}_2$.

Proofs of quadratic quantization based characterizations rely on the following lemma.

Lemma 3.1. *(a) Let $\mu, \nu \in \mathcal{P}_2(H)$. If for every $u \in H, |u| = 1$, $\mu \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1} = \nu \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1}$, then $\mu = \nu$.*
(b) Let $\mu_n \in \mathcal{P}_2(H)$ for every $n \in \mathbb{N}^* \cup \{\infty\}$. If $\int_H |\xi|^2 \, \mu_n(d\xi) \xrightarrow{n \to +\infty} \int_H |\xi|^2 \, \mu_\infty(d\xi)$ and for every $u \in H$, $|u| = 1$, $\mu_n \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1} \overset{(\mathbb{R})}{\Longrightarrow} \mu_\infty \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1}$, then $\mathcal{W}_2(\mu_n, \mu_\infty) \to 0$.

Proof. As $(H, |\cdot|)$ is separable, let $(h_k)_{k\geq 1}$ be a countable orthonormal basis of $(H, |\cdot|)$.

(a) Let X, Y be random variables with respective distributions μ and ν and let $\lambda \in H$. We define for every $m \geq 1$, $X^{(m)} := \sum_{k=1}^{m} (X \,|\, h_k) h_k$, $Y^{(m)} := \sum_{k=1}^{m} (Y \,|\, h_k) h_k$ and $\lambda^{(m)} := \sum_{k=1}^{m} (\lambda \,|\, h_k) h_k$. For $m \geq 1$, let $u^{(m)} = \frac{\lambda^{(m)}}{|\lambda^{(m)}|}$ (convention $\frac{0}{|0|} = 0$), then we have

$$(\lambda \,|\, X^{(m)}) = \sum_{k=1}^{+\infty} (\lambda \,|\, h_k)(X^{(m)} \,|\, h_k) = \sum_{k=1}^{m} (\lambda \,|\, h_k)(X \,|\, h_k) = |\lambda^{(m)}| \big(X \,|\, u^{(m)}\big).$$

Similarly, $(\lambda \,|\, Y^{(m)}) = |\lambda^{(m)}| (Y \,|\, u^{(m)})$. Let i be such that $i^2 = -1$. It follows that

$$\begin{aligned}\mathbb{E}\, e^{i(\lambda | X^{(m)})} = \mathbb{E}\, e^{i |\lambda^{(m)}| (X | u^{(m)})} &= \int_H e^{i\, |\lambda^{(m)}|\, \xi} \mu \circ \big(\xi \mapsto (u^{(m)} \mid \xi)\big)^{-1}(d\xi) \\ &= \int_H e^{i\, |\lambda^{(m)}|\, \xi} \nu \circ \big(\xi \mapsto (u^{(m)} \mid \xi)\big)^{-1}(d\xi) = \mathbb{E}\, e^{i(\lambda | Y^{(m)})}.\end{aligned}$$

Since we can arbitrarily choose λ, we have for every $m \geq 1$, $\mathrm{Law}(X^{(m)}) = \mathrm{Law}(Y^{(m)})$. Let $F : H \to \mathbb{R}$ be a bounded continuous function. Then, for every $m \geq 1$, $\mathbb{E}\, F(X^{(m)}) = \mathbb{E}\, F(Y^{(m)})$ which implies $\mathbb{E}\, F(X) = \mathbb{E}\, F(Y)$ by letting $m \to +\infty$. Hence, $\mu = \nu$.

(b) For every $n \geq 1$, let X_n be random variables with distribution μ_n and let X_∞ be a random variable with distribution μ_∞. We define for every $n \geq 1$ and for every $m \geq 1$, $X_n^{(m)} := \sum_{i=1}^{m} (X_n | h_i) h_i$ and $X_\infty^{(m)} := \sum_{i=1}^{m} (X_\infty | h_i) h_i$. Following the lines of item (a), we get for every $m \geq 1$, $X_n^{(m)} \overset{(H)}{\Longrightarrow} X_\infty^{(m)}$ as $n \to +\infty$, since the convergence of characteristic function implies weak convergence.

Now, let $F : H \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz coefficient $[F]_{\mathrm{Lip}} := \sup_{x,y \in H} \frac{|F(x) - F(y)|}{|x - y|}$. For every (temporarily) fixed $m \geq 1$,

$$\begin{aligned}&\lim_n \big|\mathbb{E}\, F(X_n) - \mathbb{E}\, F(X_\infty)\big| \\ &\quad \leq \lim_n \big|\mathbb{E}\, F(X_n) - \mathbb{E}\, F(X_n^{(m)})\big| + \lim_n \big|\mathbb{E}\, F(X_n^{(m)}) - \mathbb{E}\, F(X_\infty^{(m)})\big| + \big|\mathbb{E}\, F(X_\infty^{(m)}) - \mathbb{E}\, F(X_\infty)\big| \\ &\quad \leq \lim_n \big|\mathbb{E}\, F(X_n) - \mathbb{E}\, F(X_n^{(m)})\big| + 0 + \big|\mathbb{E}\, F(X_\infty^{(m)}) - \mathbb{E}\, F(X_\infty)\big| \quad (\text{since } X_n^{(m)} \overset{(H)}{\Longrightarrow} X_\infty^{(m)}).\end{aligned}$$

Then, for every $n \geq 1$,

$$\big|\mathbb{E}\, F(X_n) - \mathbb{E}\, F(X_n^{(m)})\big| \leq \mathbb{E}\big|F(X_n) - F(X_n^{(m)})\big| \leq [F]_{\mathrm{Lip}} \mathbb{E}\big|X_n - X_n^{(m)}\big| \leq [F]_{\mathrm{Lip}} \big\|X_n - X_n^{(m)}\big\|_2.$$

Similarly, we also have $\big|\mathbb{E}\, F(X_\infty^{(m)}) - \mathbb{E}\, F(X_\infty)\big| \leq [F]_{\mathrm{Lip}} \big\|X_\infty - X_\infty^{(m)}\big\|_2$.

It follows from Fatou's Lemma for the weak convergence and the convergence assumption made on $\mathbb{E}|X_n|^2$ that

$$\begin{aligned}\limsup_n \big\|X_n - X_n^{(m)}\big\|_2^2 = \limsup_n \mathbb{E}\big|X_n - X_n^{(m)}\big|^2 &= \limsup_n \Big[\mathbb{E}\big|X_n\big|^2 - \mathbb{E}\big|X_n^{(m)}\big|^2\Big] \\ = \mathbb{E}\big|X_\infty\big|^2 - \liminf_n \mathbb{E}\big|X_n^{(m)}\big|^2 \leq \mathbb{E}\big|X_\infty\big|^2 - \mathbb{E}\big|X_\infty^{(m)}\big|^2 &= \big\|X_\infty - X_\infty^{(m)}\big\|_2^2. \qquad (3.1)\end{aligned}$$

Hence, for every $m \geq 1$,

$$\lim_n \big|\mathbb{E}\, F(X_n) - \mathbb{E}\, F(X_\infty)\big| \leq \limsup_n [F]_{\mathrm{Lip}} \big\|X_n - X_n^{(m)}\big\|_2 + [F]_{\mathrm{Lip}} \big\|X_\infty - X_\infty^{(m)}\big\|_2$$

$$\leq 2[F]_{\mathrm{Lip}}\big\|X_\infty - X_\infty^{(m)}\big\|_2.$$

Then, $\big\|X_\infty - X_\infty^{(m)}\big\|_2 \to 0$ as $m \to +\infty$ by the Lebesgue dominated convergence theorem since $\big|X_\infty - X_\infty^{(m)}\big| \leq \big|X_\infty\big| \in L^2(\mathbb{P})$ so that $\mathbb{E}\,F(X_n) \to \mathbb{E}\,F(X_\infty)$ as $n \to +\infty$. Thus, $X_n \stackrel{(H)}{\Longrightarrow} X_\infty$ and we can conclude that $\mathcal{W}_p(\mu_n, \mu_\infty) \to 0$ by applying Theorem 1.1. □

Proposition 3.1 (Static characterization)**.** *Let $\mu, \nu \in \mathcal{P}_2(H)$. If $e_{2,2}^2(\mu,\cdot) = e_{2,2}^2(\nu,\cdot) + C$ for some real constant C, then $\mu = \nu$ and $C = 0$.*

Proof. Let $a, b \in H$, then $e_{2,2}^2\big(\mu,(a,b)\big) = \int_H |\xi - a|^2 \wedge |\xi - b|^2 \,\mu(d\xi)$.

As $e_{2,2}^2\big(\mu,(a,b)\big) = e_{2,2}^2\big(\nu,(a,b)\big) + C$ for every $(a,b) \in H^2$, in particular, if $a = b$, $\int_H |\xi - a|^2 \,\mu(d\xi) = \int_H |\xi - a|^2 \,\nu(d\xi) + C$. Hence, using that $(x-y)_+ = x - x \wedge y$, we have

$$\forall a, b \in H, \quad \int_H \big(\,|\xi - a|^2 - |\xi - b|^2\,\big)_+ \mu(d\xi) = \int_H \big(\,|\xi - a|^2 - |\xi - b|^2\,\big)_+ \nu(d\xi). \tag{3.2}$$

Note that $|\xi - a|^2 - |\xi - b|^2 = 2\Big(b - a \;\Big|\; \xi - \frac{a+b}{2}\Big)$. Hence, if we take $a = \lambda u$ and $b = \lambda' u$ with $\lambda, \lambda' \in \mathbb{R},\ \lambda' > \lambda$ for some common $u \in H$ with $|u| = 1$, we obtain

$$\big(\,|\xi - a|^2 - |\xi - b|^2\,\big)_+ = 2(\lambda' - \lambda)\left((\xi \mid u) - \frac{\lambda + \lambda'}{2}\right)_+.$$

As a consequence of (3.2), we derive that

$$\forall \lambda, \lambda' \in \mathbb{R}, \lambda' > \lambda, \quad \int_H \left((\xi \mid u) - \frac{\lambda + \lambda'}{2}\right)_+ \mu(d\xi) = \int_H \left((\xi \mid u) - \frac{\lambda + \lambda'}{2}\right)_+ \nu(d\xi).$$

In turn, this implies, by letting $\lambda' \to \lambda$,

$$\forall u \in H, |u| = 1,\ \forall \lambda \in \mathbb{R}, \qquad \int_H \Big((\xi \mid u) - \lambda\Big)_+ \mu(d\xi) = \int_H \Big((\xi \mid u) - \lambda\Big)_+ \nu(d\xi). \tag{3.3}$$

The function $\lambda \mapsto \big((\xi \mid u) - \lambda\big)_+$ is right differentiable with $\mathbb{1}_{(\xi|u)>\lambda}$ as a right derivative and μ-integrable. Hence, by the Lebesgue differentiation theorem, we can right differentiate the equality (3.3) which yields for every $u \in H, |u| = 1$ and for every $\lambda \in \mathbb{R}$, $\mu\big((\xi \mid u) > \lambda\big) = \nu\big((\xi \mid u) > \lambda\big)$.

Hence, for every $u \in H, |u| = 1$, $\mu \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1} = \nu \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1}$ since they have the same survival function. We conclude by Lemma 3.1 (a) that $\mu = \nu$ and $C = 0$. □

The following theorem shows the equivalence of $\mathcal{W}_2$-convergence of $(\mu_n)_{n\geq 1}$ in $\mathcal{P}_2(H)$ and the pointwise convergence of quadratic quantization error function $\big(e_{2,2}(\mu_n,\cdot)\big)_{n\geq 1}$.

Theorem 3.1 ($\mathcal{W}_2$-convergence characterization)**.** *Let $\mu_n \in \mathcal{P}_2(H)$ for every $n \in \mathbb{N}^* \cup \{\infty\}$. The following properties are equivalent:*

(i) $\mathcal{W}_2(\mu_n, \mu_\infty) \xrightarrow{n\to+\infty} 0$,

(ii) $e_{2,2}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{2,2}(\mu_\infty,\cdot)$ *uniformly,*

(iii) $e_{2,2}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{2,2}(\mu_\infty,\cdot)$ *pointwise.*

Before proving Theorem 3.1, we recall the convergence of left and right derivatives of a converging sequence of convex functions. Let $\partial_- f$ (respectively $\partial_+ f$) denote the left derivative (resp. right derivative) of a convex function f.

Lemma 3.2. *(See e.g. [Lac82][Theorems 2.5]) Let $f_n : \mathbb{R} \to \mathbb{R}, n \in \mathbb{N}^*$, be a sequence of convex functions converging pointwise to a function $f : \mathbb{R} \to \mathbb{R}$. Let $G := \{x \in \mathbb{R} \,|\, \partial_- f(x) \neq \partial_+ f(x)\}$. Then for every point $x \in \mathbb{R} \setminus G$,*

$$\lim_n \partial_+ f_n(x) = \lim_n \partial_- f_n(x) = f'(x).$$

Proof of Theorem 3.1. $(i) \Rightarrow (ii)$ is obvious from (1.9).

$(ii) \Rightarrow (iii)$ is obvious.

$(iii) \Rightarrow (i)$ For every $(a, b) \in H^2$,

$$e_{2,2}^2\big(\mu_n, (a,b)\big) = \int_H |\xi - a|^2 \wedge |\xi - b|^2 \, \mu_n(d\xi) \xrightarrow{n\to+\infty} e_{2,2}^2\big(\mu_\infty, (a,b)\big) = \int_H |\xi - a|^2 \wedge |\xi - b|^2 \mu_\infty(d\xi).$$

In particular, $\forall a \in H$, $\int_H |\xi - a|^2 \, \mu_n(d\xi) \xrightarrow{n\to+\infty} \int_H |\xi - a|^2 \, \mu_\infty(d\xi)$. Hence, using that $(x-y)_+ = x - x \wedge y$, we get

$$\forall a, b \in H, \quad \int_H \big(|\xi - a|^2 - |\xi - b|^2 \big)_+ \mu_n(d\xi) \xrightarrow{n\to+\infty} \int_H \big(|\xi - a|^2 - |\xi - b|^2 \big)_+ \mu_\infty(d\xi).$$

Following the lines of the proof of Proposition 3.1, we get

$$\forall \lambda \in \mathbb{R}, \, \forall u \in H, |u| = 1, \quad \int_H \Big((\xi \mid u) - \lambda \Big)_+ \mu_n(d\xi) \xrightarrow{n\to+\infty} \int_H \Big((\xi \mid u) - \lambda \Big)_+ \mu_\infty(d\xi). \tag{3.4}$$

For $\mu \in \mathcal{P}_2(H)$ and $u \in S_{|\cdot|}(0,1)$, we define the real-valued convex function ϕ_μ by $\phi_\mu : \lambda \mapsto \int \big((\xi \mid u) - \lambda \big)_+ \mu(d\xi)$. It follows from (3.4) that $(\phi_{\mu_n})_{n\geq 0}$ converges pointwise to ϕ_{μ_∞}. Moreover, ϕ_{μ_n}, ϕ_{μ_∞} are right-differentiable and their right derivatives are given by $\partial_+ \phi_{\mu_n}(\lambda) = \mu_n\big((\xi \mid u) > \lambda\big)$ and $\partial_+ \phi_{\mu_\infty}(\lambda) = \mu_\infty\big((\xi \mid u) > \lambda\big)$ respectively. Note that the functions $1 - \partial_+ \phi_{\mu_n}$ and $1 - \partial_+ \phi_{\mu_\infty}$ are the cumulative distribution functions of the probability distributions $\mu_n \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1}$ and $\mu_\infty \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1}$ and that the set of discontinuity points of $1 - \partial_+ \phi_{\mu_\infty}$ and $\partial_+ \phi_{\mu_\infty}$, is $G = \{\lambda : \mu_\infty\big(\{\xi : (\xi \mid u) = \lambda\}\big) > 0\}$.

We know from Lemma 3.2 that for every $\lambda \in \mathbb{R} \setminus G$, $\partial_+ \phi_{\mu_n}(\lambda) \xrightarrow{n\to+\infty} \partial_+ \phi_{\mu_\infty}(\lambda)$ and that $\partial_- \phi_{\mu_\infty}$ is continuous on $\mathbb{R} \setminus G$. Hence

$$\forall u \in H, |u| = 1, \quad \mu_n \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1} \xrightarrow{(\mathbb{R})} \mu_\infty \circ \big(\xi \mapsto (\xi \mid u)\big)^{-1}. \tag{3.5}$$

Moreover, $e_{2,2}\big(\mu_n, (0,0)\big)$ converges to $e_{2,2}\big(\mu_\infty, (0,0)\big)$, which also reads $\int_H |\xi|^2 \, \mu_n(d\xi) \to \int_H |\xi|^2 \, \mu_\infty(d\xi)$. Consequently, it follows from Lemma 3.1-(b) that $\mathcal{W}_2(\mu_n, \mu_\infty) \to 0$ as $n \to +\infty$. □

Remark 3.1. Proposition 3.1 and Theorem 3.1 directly imply that $N_{H,2,|\cdot|_2} \leq 2$. In fact, for every $a \in H$,

$$e_{1,2}^2(\mu, a) = \int_H |\xi - a|_H^2 \, \mu(d\xi) = \int_H |\xi|_H^2 \, \mu(d\xi) - 2\Big(\int_H \xi \mu(d\xi) \,|\, a \Big)_H + |a|_H^2 \, .$$

Thus, if $\mu, \nu \in \mathcal{P}_2(H)$ are such that

$$\int_H |\xi|_H^2 \, \mu(d\xi) = \int_H |\xi|_H^2 \, \nu(d\xi) \text{ and } \int_H \xi \mu(d\xi) = \int_H \xi \nu(d\xi), \tag{3.6}$$

then we have $e_{1,2}(\mu, \cdot) = e_{1,2}(\nu, \cdot)$. But condition (3.6) is clearly not sufficient to have $\mu = \nu$. Consequently, $N_{H,2,|\cdot|_2} = 2$.

Like what we did in Section 2.2, we define a function $\mathcal{Q}_{2,2}^H$ on $\big(\mathcal{P}_2(H)\big)^2$ by $(\mu, \nu) \mapsto \mathcal{Q}_{2,2}^H(\mu, \nu) = \|e_{2,2}(\mu, \cdot) - e_{2,2}(\nu, \cdot)\|_{\sup}$. Then inequality (1.8) implies that $\mathcal{Q}_{2,2}^H(\mu, \nu) \in [0, +\infty)$. Moreover, Proposition 3.1 and Theorem 3.1 lead the following corollary.

Corollary 3.1. *The distances $\mathcal{Q}_{2,2}^H$ and $\mathcal{W}_2$ are topologically equivalent on $\mathcal{P}_2(H)$.*

We conclude this section by an "à la Paul Lévy" characterization of a limit of quantization errors functions.

Theorem 3.2 (À la Paul Lévy characterization)**.** *Let $(H, |\cdot|_H)$ be a separable Hilbert space. Let $(\mu_n)_{n\geq 1}$ be a $\mathcal{P}_2(H)$-valued sequence and let $f : H \to \mathbb{R}_+$ be such that*

$$e_{2,2}(\mu_n, \cdot) \xrightarrow{n\to+\infty} f \quad \textit{pointwise.}$$

Then there exists $\mu_\infty \in \mathcal{P}_2(H)$ such that $\mu_n \overset{(H_w)}{\Longrightarrow} \mu_\infty$ (where (H_w) stands for the weak topology on H) and

$$f^2 = e_{2,2}(\mu_\infty, \cdot)^2 + \lim_n \int_H |\xi|^2 \, \mu_n(d\xi) - \int_H |\xi|^2 \, \mu_\infty(d\xi).$$

Proof. The sequence $e_{2,2}\big(\mu_n, (0,0)\big)^2 = \int_H |\xi|^2 \mu_n(d\xi)$, $n \geq 1$, is bounded, hence the sequence $(\mu_n)_{n\geq 1}$ is tight for the weak topology (H_w) on H, which generates the same Borel σ-field as the strong one. Consequently there exists a subnet $\mu_{\varphi(n)} \overset{(H_w)}{\Longrightarrow} \mu_\infty \in \mathcal{P}_2(H)$ since the mapping $\xi \mapsto |\xi|^2$ is weakly lower semi-continuous and non-negative (see [Top74][Lemma 2.3 and Theorem 3.1] and [Kel75][Chapter 2] for the definition of subnet). Now note that, for a fixed $x = (x_1, x_2) \in H^2$, the mapping $\xi \mapsto \min\big(|\xi - x_1|^2, |\xi - x_2|^2\big) - |\xi|^2 = \min\big(|x_1|^2 - 2(x_1|\xi), |x_2|^2 - 2(x_2|\xi)\big)$ is weakly continuous and $(\mu_n)_{n\geq 1}$-uniformly integrable since it is sublinear. Hence

$$\begin{aligned} e_{2,2}^2(\mu_{\varphi(n)}, x) &\longrightarrow \int_H \min\big(|x_1|^2 - 2(x_1|\xi), |x_2|^2 - 2(x_2|\xi)\big)\mu_\infty(d\xi) + f^2\big((0,0)\big) \text{ as } n \to +\infty \\ &= e_{2,2}^2(\mu_\infty, x) + f^2\big((0,0)\big) - \int_H |\xi|^2 \, \mu_\infty(d\xi). \end{aligned}$$

For two such limiting distributions μ_∞ and μ'_∞ it follows from what precedes that $e_{2,2}^2(\mu_\infty, \cdot) = e_{2,2}^2(\mu'_\infty, \cdot) + C_\infty$ for some real constant C_∞. Hence $\mu_\infty = \mu'_\infty$ by Proposition 3.1, which in turn implies that $\mu_n \overset{(H_w)}{\Longrightarrow} \mu_\infty$. □

4 Further quantization based characterizations on $\mathbb{R}$

Let $|\cdot|$ denote the absolute value on $\mathbb{R}$. Results from Section 2 (Theorem 2.1 and 2.2, Proposition 2.4-(i)) imply that $N_{1,p} := N_{1,p,|\cdot|} \leq 3$ for any $p \geq 1$. Moreover, Proposition 3.1 and

Theorem 3.1 imply that $N_{1,2} = 2$. Other quantization based characterizations are developed in Section 4.1. Then we discuss the completeness of the distance $\mathcal{Q}_{1,1}$ $\big($defined in (2.6)$\big)$ on $\mathcal{P}_1(\mathbb{R})$ and of $\mathcal{Q}_{2,2}$ on $\mathcal{P}_2(\mathbb{R})$ with opposite answers in Section 4.2.

4.1 Quantization based characterization on $\mathbb{R}$

Proposition 4.1 ($p = 1$)**.** (*a*) *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. If $e_{1,1}(\mu, \cdot) = e_{1,1}(\nu, \cdot) + C$ for some real constant C, then $\mu = \nu$ and $C = 0$.*

(*b*) *If $\mu_n \in \mathcal{P}_1(\mathbb{R})$, $n \in \mathbb{N}^* \cup \{\infty\}$, the following properties are equivalent:*

(*i*) $\mathcal{W}_1(\mu_n, \mu_\infty) \xrightarrow{n\to+\infty} 0$,

(*ii*) $e_{1,1}(\mu_n, \cdot) \xrightarrow{n\to+\infty} e_{1,1}(\mu_\infty, \cdot)$ *uniformly,*

(*iii*) $e_{1,1}(\mu_n, \cdot) \xrightarrow{n\to+\infty} e_{1,1}(\mu_\infty, \cdot)$ *pointwise.*

(*c*) *The distance $\mathcal{Q}_{1,1}$ and $\mathcal{W}_1$ are topologically equivalent on $\mathcal{P}_1(\mathbb{R})$ and $N_{1,1} = 1$.*

Proof. (*a*) The function $e_{1,1}(\mu, \cdot)$ reads $x \mapsto \int_{\mathbb{R}} |\xi - x| \, \mu(d\xi)$, hence it is convex and its right derivative is given by $x \mapsto -1 + 2\mu\big(]-\infty, x]\big)$. So if $e_{1,1}(\mu, \cdot) = e_{1,1}(\nu, \cdot) + C$, we have $\mu\big(]-\infty, x]\big) = \nu\big(]-\infty, x]\big)$ for all $x \in \mathbb{R}$, which implies $\mu = \nu$ (and $C = 0$).

(*b*) It is obvious that (*i*) $\Rightarrow$ (*ii*) and (*ii*) $\Rightarrow$ (*iii*). Now we prove (*iii*) $\Rightarrow$ (*i*).

For every $n \geq 1$, $e_{1,1}(\mu_n, \cdot)$ can also be written as $a \mapsto \int_{\mathbb{R}} |\xi - a| \, \mu_n(d\xi)$, which is convex with right derivative at a given by $-1 + 2\mu_n\big(]-\infty, a]\big)$. Consequently, if $e_{1,1}(\mu_n, \cdot)$ converges pointwise to $e_{1,1}(\mu_\infty, \cdot)$ on $\mathbb{R}$, then $\mu_n\big(]-\infty, a]\big)$ converges pointwise to $\mu_\infty\big(]-\infty, a]\big)$ for all $a \in \mathbb{R}$ such that $\mu_\infty(\{a\}) = 0$ by Lemma 3.2. This implies $\mu_n \xRightarrow{(\mathbb{R})} \mu_\infty$. The convergence of the first moment follows from $e_{1,1}(\mu_n, 0) \xrightarrow{n\to+\infty} e_{1,1}(\mu_\infty, 0)$. Hence, we conclude that $\mathcal{W}_1(\mu_n, \mu_\infty) \xrightarrow{n\to+\infty} 0$ by Theorem 1.1.

(*c*) The claim (*c*) is a direct result from (*a*) and (*b*). □

Proposition 4.2 (Even integer $p \geq 2$)**.** *Let p be an even integer, $p \geq 2$.*

(*a*) *Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$ such that $e^p_{2,p}(\mu, \cdot) = e^p_{2,p}(\nu, \cdot) + C$ for some real constant C. Then $\mu = \nu$.*

(*b*) *If $\mu_n \in \mathcal{P}_p(\mathbb{R})$, $n \in \mathbb{N}^* \cup \{\infty\}$, the following properties are equivalent:*

(*i*) $\mathcal{W}_p(\mu_n, \mu_\infty) \xrightarrow{n\to+\infty} 0$,

(*ii*) $e_{2,p}(\mu_n, \cdot) \xrightarrow{n\to+\infty} e_{2,p}(\mu_\infty, \cdot)$ *uniformly,*

(*iii*) $e_{2,p}(\mu_n, \cdot) \xrightarrow{n\to+\infty} e_{2,p}(\mu_\infty, \cdot)$ *pointwise.*

(*c*) *The distances $\mathcal{Q}_{2,p}$ and $\mathcal{W}_p$ are topologically equivalent on $\mathcal{P}_p(\mathbb{R})$ and $N_{1,p} = 2$.*

The proof of Proposition 4.2 is based on the following lemma.

Lemma 4.1. *Let p be an even number, $p \geq 2$. Let $\mu \in \mathcal{P}_p(\mathbb{R})$ be absolutely continuous with density f* i.e. *$\mu(d\xi) = f(\xi)d\xi$. If f is continuous, then for any $a, b \in \mathbb{R}$ with $a < b$,*

$$e^{p-2}_{2,p-2}\big(\mu,(a,b)\big) = \frac{1}{p(p-1)}\left(\frac{\partial^2 e^p_{2,p}}{\partial a^2}\big(\mu,(a,b)\big) + \frac{\partial^2 e^p_{2,p}}{\partial b^2}\big(\mu,(a,b)\big) - 2\frac{\partial^2 e^p_{2,p}}{\partial a \partial b}\big(\mu,(a,b)\big)\right). \quad (4.1)$$

Proof of Lemma 4.1. Assume that $a < b$, then $e^p_{2,p}\big(\mu,(a,b)\big) = \int_{-\infty}^{\frac{a+b}{2}} |\xi - a|^p f(\xi)d\xi + \int_{\frac{a+b}{2}}^{+\infty} |\xi - b|^p f(\xi)d\xi$. Hence, the function $e^p_{2,p}\big(\mu,(a,b)\big)$ is continuously differentiable in a, since, for any even number $p \geq 2$, we have $\frac{\partial |\xi - a|^p f(\xi)}{\partial a} = p(a-\xi)^{p-1} f(\xi)$ and

$$\sup_{a' \in (a-1, a+1)} \big|p(a'-\xi)^{p-1} f(\xi)\big| \leq p 2^{p-1} f(\xi)\big[\,|a+1|^{p-1} \vee |a-1|^{p-1} + |\xi|^{p-1}\,\big] \in L^1(\lambda)$$

since $\int_{\mathbb{R}} |\xi|^p f(\xi)d\xi < +\infty$. Likewise, $e^p_{2,p}\big(\mu,(a,b)\big)$ is continuously differentiable in b with partial derivatives

$$\frac{\partial e^p_{2,p}\big(\mu,(a,b)\big)}{\partial a} = p\int_{-\infty}^{\frac{a+b}{2}} (a-\xi)^{p-1} f(\xi)d\xi \text{ and } \frac{\partial e^p_{2,p}\big(\mu,(a,b)\big)}{\partial b} = p\int_{\frac{a+b}{2}}^{+\infty} (b-\xi)^{p-1} f(\xi)d\xi.$$

Moreover, we have $\frac{\partial (a-\xi)^{p-1} f(\xi)}{\partial a} = (p-1)(a-\xi)^{p-2} f(\xi)$ and

$$\sup_{a' \in (a-1, a+1)} \big|(p-1)(a'-\xi)^{p-2} f(\xi)\big| \leq (p-1) 2^{p-2} f(\xi)\big[\,|a+1|^{p-2} \vee |a-1|^{p-2} + |\xi|^{p-2}\,\big] \in L^1(d\xi)$$

since $\int_{\mathbb{R}} |\xi|^p f(\xi)d\xi < +\infty$. By a similar reasoning, one derives that $e^p_{2,p}\big(\mu,(a,b)\big)$ is continuously twice differentiable with second order partial derivatives

$$\begin{aligned}
\frac{\partial^2 e^p_{2,p}}{\partial a^2}\big(\mu,(a,b)\big) &= p\Big[\int_{-\infty}^{\frac{a+b}{2}} (p-1)(a-\xi)^{p-2} f(\xi)d\xi - \frac{1}{2^p}(b-a)^{p-1} f\big(\frac{a+b}{2}\big)\Big],\\
\frac{\partial^2 e^p_{2,p}}{\partial b^2}\big(\mu,(a,b)\big) &= p\Big[\int_{\frac{a+b}{2}}^{+\infty} (p-1)(b-\xi)^{p-2} f(\xi)d\xi - \frac{1}{2^p}(b-a)^{p-1} f\big(\frac{a+b}{2}\big)\Big],\\
\frac{\partial^2 e^p_{2,p}}{\partial a \partial b}\big(\mu,(a,b)\big) &= \frac{\partial^2 e^p_{2,p}}{\partial b \partial a}\big(\mu,(a,b)\big) = -p\frac{1}{2^p}(b-a)^{p-1} f\Big(\frac{a+b}{2}\Big).
\end{aligned}$$

Hence, for every $(a,b) \in \mathbb{R}^2$ such that $a < b$,

$$\frac{\partial^2 e^p_{2,p}}{\partial a^2}\big(\mu,(a,b)\big) + \frac{\partial^2 e^p_{2,p}}{\partial b^2}\big(\mu,(a,b)\big) - 2\frac{\partial^2 e^p_{2,p}}{\partial a \partial b}\big(\mu,(a,b)\big) = p(p-1)e^{p-2}_{2,p-2}\big(\mu,(a,b)\big).$$

Proof of Proposition 4.2. (*a*) *Step 1: μ and ν are absolutely continuous with continuous density functions.* Note that $e^p_{2,p}(\mu,\cdot) = e^p_{2,p}(\nu,\cdot) + C$ implies either $\mu = \nu$ by Proposition 3.1 if $p = 2$, or, if $p > 2$ $e^{p-2}_{2,p-2}(\mu,\cdot) = e^{p-2}_{2,p-2}(\nu,\cdot)$ (after differentiation) by Lemma 4.1. We can conclude by induction.

Step 2 (General case). Let X,Y be two random variables with the respective distributions μ and ν, such that

$$\forall (a,b) \in \mathbb{R}^2, \quad e^p_{2,p}\big(X,(a,b)\big) = e^p_{2,p}\big(Y,(a,b)\big) + C. \quad (4.2)$$

Let Z be a random variable with probability distribution $\mathbb{P}_Z = \mathcal{N}(0,1)$, independent of X and Y.

For every $\varepsilon > 0$,

$$e_{2,p}^p\big(X+\varepsilon Z,(a,b)\big) = \iint \min_{x\in\{a,b\}} |\xi+\varepsilon z - x|^p \,\mu(d\xi)\mathbb{P}_Z(dz) = \int e_{2,p}^p\big(X,(a,b)-\varepsilon z\big)\mathbb{P}_Z(dz). \quad (4.3)$$

We derive from (4.2) and (4.3) that

$$\forall (a,b)\in\mathbb{R}^2, \quad e_{2,p}^p\big(X+\varepsilon Z,(a,b)\big) = e_{2,p}^p\big(Y+\varepsilon Z,(a,b)\big) + C. \quad (4.4)$$

Moreover, the random variables $X+\varepsilon Z$ and $Y+\varepsilon Z$ have distributions $\mathcal{N}(0,\varepsilon^2)*\mu$ and $\mathcal{N}(0,\varepsilon^2)*\nu$ respectively, both with continuous densities. It follows from *Step 1* that $\mathrm{Law}(X+\varepsilon Z) = \mathrm{Law}(Y+\varepsilon Z)$ for every $\varepsilon > 0$ so that $\mathrm{Law}(X)$=$\mathrm{Law}(Y)$ by letting $\varepsilon \to 0$.

(b) It is obvious that $(i) \Rightarrow (ii)$ and $(ii) \Rightarrow (iii)$. Now we prove $(iii) \Rightarrow (i)$. It follows from Lemma 4.1 that $e_{2,p}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{2,p}(\mu_\infty,\cdot)$ implies $e_{2,p-2}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{2,p-2}(\mu_\infty,\cdot)$ and, by induction, yields $e_{2,2}(\mu_n,\cdot) \xrightarrow{n\to+\infty} e_{2,2}(\mu_\infty,\cdot)$, so that Theorem 3.1 and Theorem 1.1 imply that μ_n converges weakly to μ_∞. The convergence of the p-th moment follows from $e_{2,p}(\mu_n,0) \xrightarrow{n\to+\infty} e_{2,p}(\mu_\infty,0)$. Hence $\mathcal{W}_p(\mu_n,\mu_\infty) \xrightarrow{n\to+\infty} 0$ by Theorem 1.1.

(c) The claim (a) and (b) directly imply that if p is an even integer, $p \geq 2$, the distances $\mathcal{Q}_{2,p}$ and $\mathcal{W}_p$ are topologically equivalent on $\mathcal{P}_p(\mathbb{R})$ and $N_{1,p} \leq 2$. Now we prove that $N_{1,p} = 2$. Note that for every $x \in \mathbb{R}$, $e_{1,p}^p(\mu,x) = \int_{\mathbb{R}} |\xi - x|^p \,\mu(d\xi) = \int_{\mathbb{R}} (\xi^2 - 2\xi x + x^2)^{\frac{p}{2}} \mu(d\xi)$, which is polynomial in x and whose coefficients are the k-th moments of μ, $k \in \{1,...,p\}$. Thus, as soon as two different distributions μ and ν have the same first p moments, $e_{1,p}^p(\mu,\cdot) = e_{1,p}^p(\nu,\cdot)$. This implies $N_{1,p} > 1$. □

4.2 About completeness of $\big(\mathcal{P}_1(\mathbb{R}),\mathcal{Q}_{1,1}\big)$ and $\big(\mathcal{P}_2(\mathbb{R}),\mathcal{Q}_{N,2}\big)$

We know from [Bol08] that for $p \geq 1$, $(\mathcal{P}_p(\mathbb{R}),\mathcal{W}_p)$ is a complete space and we have proved that $\mathcal{Q}_{1,1}$ (respectively $Q_{2,2}$) is topologically equivalent to $\mathcal{W}_1$ (resp. $\mathcal{W}_2$) on $\mathcal{P}_1(\mathbb{R})$ (resp. $\mathcal{P}_2(\mathbb{R})$). Now we discuss whether $\mathcal{Q}_{1,1}$ and $\mathcal{Q}_{2,2}$ are complete distances.

Proposition 4.3. *The metric space* $\big(\mathcal{P}_1(\mathbb{R}),\mathcal{Q}_{1,1}\big)$ *is complete.*

Proof. The inequality (1.8) directly implies that a Cauchy sequence in $\big(\mathcal{P}_1(\mathbb{R}),\mathcal{W}_1\big)$ is also a Cauchy sequence in $\big(\mathcal{P}_1(\mathbb{R}),\mathcal{Q}_{1,1}\big)$. Now let $(\mu_n)_{n\geq 1}$ be a Cauchy sequence in $\big(\mathcal{P}_1(\mathbb{R}),\mathcal{Q}_{1,1}\big)$. It follows from the definition of $\mathcal{Q}_{1,1}$ that $\big(e_{1,1}(\mu_n,\cdot) - e_{1,1}(\delta_0,\cdot)\big)_{n\geq 1}$ is a Cauchy sequence in $\big(\mathcal{C}_b(\mathbb{R},\mathbb{R}),\|\cdot\|_{\sup}\big)$.

As $\big(\mathcal{C}_b(\mathbb{R},\mathbb{R}),\|\cdot\|_{\sup}\big)$ is complete, there exists a function $g\in\mathcal{C}_b(\mathbb{R},\mathbb{R})$ such that

$$\big\|\big(e_{1,1}(\mu_n,\cdot) - e_{1,1}(\delta_0,\cdot)\big) - g\big\|_{\sup} \xrightarrow{n\to+\infty} 0. \quad (4.5)$$

Note that for any $a\in\mathbb{R}$, $e_{1,1}(\delta_0,a) = |a|$. The sequence $e_{1,1}(\mu_n,0) - e_{1,1}(\delta_0,0) = e_{1,1}(\mu_n,0)$ is also a Cauchy sequence in $\mathbb{R}$. Therefore, $\big(e_{1,1}(\mu_n,0)\big)_{n\geq 1} = \big(\int_{\mathbb{R}} |\xi|\,\mu_n(d\xi)\big)_{n\geq 1}$ is bounded, which implies that $(\mu_n)_{n\geq 1}$ is tight. It follows from Prohorov's theorem that there exists a subsequence $(\mu_{\varphi(n)})_{n\geq 1}$ weakly converging to $\widetilde{\mu}_\infty$. Moreover, by Fatou's lemma in distribution, $\widetilde{\mu}_\infty \in \mathcal{P}_1(\mathbb{R})$ since $\int_{\mathbb{R}} |\xi|\,\widetilde{\mu}_\infty(d\xi) \leq \liminf_n \int_{\mathbb{R}} |\xi|\,\mu_{\varphi(n)}(d\xi) < +\infty$.

Now, we prove that $g = e_{1,1}(\widetilde{\mu},\cdot) - e_{1,1}(\delta_0,\cdot)$. First, let us define a function $f_a(\xi) := |\xi - a| - |\xi|$. For every $a \in \mathbb{R}$, f_a is bounded and continuous. Hence, the weak convergence of $(\mu_{\varphi(n)})_{n\geq 1}$ implies that $\displaystyle\int_{\mathbb{R}} f_a(\xi)\mu_{\varphi(n)}(d\xi) \xrightarrow{n\to+\infty} \int_{\mathbb{R}} f_a(\xi)\widetilde{\mu}_\infty(d\xi)$.

Besides, $\int_{\mathbb{R}} f_a(\xi)\mu_{\varphi(n)}(d\xi) = \int_{\mathbb{R}} \big[\,|\xi - a| - |\xi|\,\big]\mu_{\varphi(n)}(d\xi) = e_{1,1}(\mu_{\varphi(n)}, a) - e_{1,1}(\mu_{\varphi(n)}, 0)$, which converges to $\big(g(a) + e_{1,1}(\delta_0, a)\big) - \big(g(0) + e_{1,1}(\delta_0, 0)\big)$ as $n \to +\infty$ by (4.5). Hence, for every $a \in \mathbb{R}$,

$$\big(g(a) + e_{1,1}(\delta_0, a)\big) - \big(g(0) + \underbrace{e_{1,1}(\delta_0, 0)}_{=0}\big) = \int_{\mathbb{R}} f_a(\xi)\widetilde{\mu}_\infty(d\xi) = e_{1,1}(\widetilde{\mu}_\infty, a) - e_{1,1}(\widetilde{\mu}_\infty, 0),$$

i.e. $e_{1,1}(\widetilde{\mu}_\infty, a) - e_{1,1}(\delta_0, a) - g(a) = e_{1,1}(\widetilde{\mu}_\infty, 0) - g(0)$. Setting $C = e_{1,1}(\widetilde{\mu}_\infty, 0) - g(0)$, we derive that for every $a \in \mathbb{R}$,

$$e_{1,1}(\widetilde{\mu}_\infty, a) - e_{1,1}(\delta_0, a) - g(a) = C. \tag{4.6}$$

Now we prove that $C = 0$. Generally, for any $\nu \in \mathcal{P}_1(\mathbb{R})$, one has

$$\begin{aligned}
&\lim_{a\to+\infty} \big(e_{1,1}(\nu, a) - e_{1,1}(\delta_0, a)\big) = \lim_{a\to+\infty} \big(e_{1,1}(\nu, a) - |a|\,\big) = \lim_{a\to+\infty} \big(e_{1,1}(\nu, a) - a\big) \\
&= \lim_{a\to+\infty} \Big(\int_{\mathbb{R}} |\xi - a|\,\nu(d\xi) - a\Big) = \lim_{a\to+\infty} \Big(\int_{\{\xi\geq a\}} (\xi - a)\nu(d\xi) + \int_{\{\xi<a\}} (a - \xi)\nu(d\xi) - a\Big) \\
&= \lim_{a\to+\infty} \Big(\int_{\{\xi\geq a\}} \xi\nu(d\xi) - 2\int_{\{\xi\geq a\}} a\nu(d\xi) + \int_{\{\xi<a\}} (-\xi)\nu(d\xi)\Big).
\end{aligned}$$

As $\nu \in \mathcal{P}_1(\mathbb{R})$ i.e. $\int_{\mathbb{R}} |\xi|\,\nu(d\xi) < +\infty$, we derive that $\lim_{a\to+\infty} \int_{\xi<a}(-\xi)\nu(d\xi) = \int_{\mathbb{R}}(-\xi)\nu(d\xi)$ and $\lim_{a\to+\infty} \int_{\{\xi\geq a\}} \xi\nu(d\xi) = 0$. This implies

$$0 \leq \lim_{a\to+\infty} \int_{\{\xi\geq a\}} a\,\nu(d\xi) \leq \lim_{a\to+\infty} \int_{\{\xi\geq a\}} \xi\,\nu(d\xi) = 0.$$

After a similar calculation with $\lim_{a\to-\infty} \big(e_{1,1}(\nu, a) - e_{1,1}(\delta_0, a)\big)$, we get

$$\lim_{a\to+\infty} \big[e_{1,1}(\nu, a) - e_{1,1}(\delta_0, a)\big] = \int_{\mathbb{R}} (-\xi)\nu(d\xi) \text{ and } \lim_{a\to-\infty} \big[e_{1,1}(\nu, a) - e_{1,1}(\delta_0, a)\big] = \int_{\mathbb{R}} \xi\nu(d\xi). \tag{4.7}$$

Combining (4.6) and (4.7) with $\nu = \widetilde{\mu}_\infty$ shows that

$$\lim_{a\to+\infty} g(a) = -C - \int_{\mathbb{R}} \xi\widetilde{\mu}_\infty(d\xi) \text{ and } \lim_{a\to-\infty} g(a) = -C + \int_{\mathbb{R}} \xi\widetilde{\mu}_\infty(d\xi).$$

On the other hand, for every $n \geq 1$, (4.7) applied to $\nu = \mu_{\varphi(n)}$ implies

$$\lim_{a\to\pm\infty} e_{1,1}(\mu_{\varphi(n)}, a) - e_{1,1}(\delta_0, a) = \mp \int_{\mathbb{R}} \xi\mu_{\varphi(n)}(d\xi).$$

Up to a new extraction of $\mu_{\varphi(n)}$, still denoted by $\mu_{\varphi(n)}$, we may assume that $\int_{\mathbb{R}} \xi\mu_{\varphi(n)}(d\xi) \to \widetilde{C} \in \mathbb{R}$ as $n \to +\infty$ since $\big(e_{1,1}(\mu_n, 0)\big)_{n\geq 1} = \big(\int_{\mathbb{R}} |\xi|\,\mu_n(d\xi)\big)_{n\geq 1}$ is bounded.

Now the uniform convergence (4.5) implies that

$$\lim_n \lim_{a\to\pm\infty} \Big[e_{1,1}(\mu_{\varphi(n)}, a) - e_{1,1}(\delta_0, a) - g(a) \Big] = 0$$

so that $\widetilde{C} = C + \int_{\mathbb{R}} \xi \widetilde{\mu}_\infty(d\xi) = -C + \int_{\mathbb{R}} \xi \widetilde{\mu}_\infty(d\xi)$, which in turn implies $C = 0$, i.e. $g = e_{1,1}(\widetilde{\mu}_\infty, \cdot) - e_{1,1}(\delta_0, \cdot)$. Then it follows from (4.5) that

$$\big\| \big(e_{1,1}(\mu_n, \cdot) - e_{1,1}(\delta_0, \cdot)\big) - \big(e_{1,1}(\widetilde{\mu}_\infty, \cdot) - e_{1,1}(\delta_0, \cdot)\big) \big\| = \|e_{1,1}(\mu_n, \cdot) - e_{1,1}(\widetilde{\mu}_\infty, \cdot)\|_{\sup} \xrightarrow{n\to+\infty} 0$$

Hence, $\mathcal{W}_1(\mu_n, \widetilde{\mu}_\infty) \to 0$ by applying Proposition 4.1, that is, $(\mu_n)_{n\geq 1}$ is a Cauchy sequence in $(\mathcal{P}_1(\mathbb{R}), \mathcal{W}_1)$. The completeness of $(\mathcal{P}_1(\mathbb{R}), \mathcal{W}_1)$ implies immediately that $(\mathcal{P}_1(\mathbb{R}), \mathcal{Q}_{1,1})$ is complete. □

Theorem 4.1. *For any $N \geq 2$, the metric space $\big(\mathcal{P}_2(\mathbb{R}), \mathcal{Q}_{N,2}\big)$ is not complete.*

We will build a sequence on $\mathcal{P}_2(\mathbb{R})$ which is Cauchy for $\mathcal{Q}_{N,2}$ but not for $\mathcal{W}_2$. First, we have the following result.

Lemma 4.2. *Let $(\mu_n)_{n\geq 1}$ be a $\mathcal{P}_2(\mathbb{R}^d)$-valued sequence which converges weakly to μ_∞ and, for $n \in \mathbb{N}^* \cup \{\infty\}$, let X_n denote a μ_n-distributed random variable . Assume that $\lim_n \mathbb{E}\,|X_n|^2$ exists and is finite. Then*

$$\sup_{a\in\mathbb{R}^d} \Big| e_{2,2}\big(\mu_n, (a,a)\big) - \sqrt{e^2_{2,2}\big(\mu_\infty, (a,a)\big) + C_0} \Big| \xrightarrow{n\to+\infty} 0, \tag{4.8}$$

where $C_0 = \lim_n \mathbb{E}\,|X_n|^2 - \mathbb{E}\,|X_\infty|^2 \in [0, +\infty)$.

Proof of Lemma 4.2. An elementary computation shows that

$$e^2_{2,2}\big(\mu_n, (a,a)\big) = \int_{\mathbb{R}^d} |\xi - a|^2 \, \mu_n(d\xi) = \int_{\mathbb{R}^d} |\xi|^2 \, \mu_n(d\xi) - 2\Big(\int_{\mathbb{R}^d} \xi \mu_n(d\xi) \,\big|\, a \Big) + |a|^2 .$$

As $\Big(\int_{\mathbb{R}^d} |\xi|^2 \, \mu_n(d\xi) \Big)_{n\geq 1}$ is bounded and $\mu_n \xrightarrow{(\mathbb{R}^d)} \mu_\infty$, we have $\int_{\mathbb{R}^d} \xi \mu_n(d\xi) \to \int_{\mathbb{R}^d} \xi \mu_\infty(d\xi)$. It follows that

$$\begin{aligned} e^2_{2,2}\big(\mu_n, (a,a)\big) &= \int_{\mathbb{R}^d} |\xi|^2 \, \mu_n(d\xi) - 2\Big(\int_{\mathbb{R}^d} \xi \mu_n(d\xi) \,\big|\, a \Big) + |a|^2 \\ &\xrightarrow{n\to+\infty} \int_{\mathbb{R}^d} |\xi|^2 \, \mu_\infty(d\xi) + C_0 - 2\Big(\int_{\mathbb{R}^d} \xi \mu_\infty(d\xi) \,\big|\, a \Big) + |a|^2 = e^2_{2,2}\big(\mu_\infty, (a,a)\big) + C_0. \end{aligned}$$

Therefore, for every compact set K in $\mathbb{R}^d$, we have

$$\sup_{a\in K} \Big| e_{2,2}\big(\mu_n, (a,a)\big) - \sqrt{e^2_{2,2}\big(\mu_\infty, (a,a)\big) + C_0} \Big| \xrightarrow{n\to+\infty} 0, \tag{4.9}$$

owing to Arzelá-Ascoli theorem, since all functions $e_{N,p}$ are 1-Lipschitz continuous (see (1.5)). On the other hand, we have

$$\begin{aligned} &\Big| e_{2,2}\big(\mu_n, (a,a)\big) - \sqrt{e^2_{2,2}(\mu_\infty, (a,a)) + C_0} \Big| \\ &\qquad = \frac{\Big| e^2_{2,2}\big(\mu_n, (a,a)\big) - \Big(e^2_{2,2}\big(\mu_\infty, (a,a)\big) + C_0 \Big) \Big|}{e_{2,2}\big(\mu_n, (a,a)\big) + \sqrt{e^2_{2,2}\big(\mu_\infty, (a,a)\big) + C_0}} \end{aligned}$$

$$
\begin{aligned}
&\leq \frac{\left|\mathbb{E}\big(\,|X_n|^2 - 2(a\,|\,X_n) + |a|^2\,\big) - \mathbb{E}\big(\,|X_\infty|^2 - 2(a\,|\,X_\infty) + |a|^2\,\big) - C_0\right|}{\|X_n - a\|_2 + \|X_\infty - a\|_2} \\
&\leq \frac{2\,|(a\mid \mathbb{E}X_\infty - \mathbb{E}X_n)| + \left|\mathbb{E}\,|X_n|^2 - \mathbb{E}\,|X_\infty|^2 - C_0\right|}{\|X_n - a\|_2 + \|X_\infty - a\|_2} \\
&\leq \frac{2\,|a|\,|\mathbb{E}X_\infty - \mathbb{E}X_n| + \left|\mathbb{E}\,|X_n|^2 - \mathbb{E}\,|X_\infty|^2 - C_0\right|}{\big|\,\|X_n\|_2 - |a|\,\big| + \big|\,\|X_\infty\|_2 - |a|\,\big|}.
\end{aligned} \tag{4.10}
$$

Let $A \coloneqq 2\sup_{n\in\mathbb{N}\cup\{\infty\}} \mathbb{E}\,|X_n|^2$, then

$$
\begin{aligned}
&\sup_{|a|>A} \left|e_{2,2}\big(\mu_n,(a,a)\big) - \sqrt{e_{2,2}^2(\mu_\infty,(a,a)) + C_0}\right| \\
&\quad\leq \sup_{|a|>A} \frac{2\,|a|\,|\mathbb{E}X_\infty - \mathbb{E}X_n| + \left|\mathbb{E}\,|X_n|^2 - \mathbb{E}\,|X_\infty|^2 - C_0\right|}{|a| - \|X_n\|_2 + |a| - \|X_\infty\|_2} \\
&\quad\leq \sup_{|a|>A} \frac{2\,|a|\,|\mathbb{E}X_\infty - \mathbb{E}X_n| + \left|\mathbb{E}\,|X_n|^2 - \mathbb{E}\,|X_\infty|^2 - C_0\right|}{2\,|a| - A} \\
&\quad\leq \sup_{|a|>A} 2\,|\mathbb{E}X_\infty - \mathbb{E}X_n| + \frac{\left|\mathbb{E}\,|X_n|^2 - \mathbb{E}\,|X_\infty|^2 - C_0\right|}{A} \xrightarrow{n\to+\infty} 0
\end{aligned} \tag{4.11}
$$

Hence, (4.9) and (4.11) imply that

$$
\sup_{a\in\mathbb{R}^d} \left|e_{2,2}\big(\mu_n,(a,a)\big) - \sqrt{e_{2,2}^2\big(\mu_\infty,(a,a)\big) + C_0}\right| \xrightarrow{n\to+\infty} 0. \qquad \square
$$

Let $Z:\Omega\to\mathbb{R}$ be $\mathcal{N}(0,1)$-distributed. We define for every $n\in\mathbb{N}$,

$$
X_n \coloneqq e^{\frac{n}{2}Z - \frac{n^2}{4}}. \tag{4.12}
$$

For $n \geq 1$, let μ_n denote the probability distribution of X_n. It is obvious that X_n converges a.s. to $X_\infty = 0$, so that $\mu_\infty = \delta_0$. Moreover, for every $p > 0$, $\mathbb{E}\,X_n^p = e^{\frac{pn^2}{8}(p-2)}$. Hence, $\mathbb{E}\,X_n = e^{-\frac{n^2}{8}} \longrightarrow 0 = \mathbb{E}\,X_\infty$ as $n \to +\infty$ so that $\mathcal{W}_1(\mu_n,\mu_\infty) \to 0$ whereas $\mathbb{E}\,X_n^2 = 1$ for every $n\in\mathbb{N}$.

Hence $\mathbb{E}X_n^2$ does not converge to $\mathbb{E}\,X_\infty^2 = 0$, which entails that μ_n does not converge to μ_∞ for the Wasserstein distance $\mathcal{W}_2$ and thus μ_n is not a $\mathcal{W}_2$-Cauchy sequence. We first prove $(\mu_n)_{n\geq1}$ is a Cauchy sequence in $\big(\mathcal{P}_2(\mathbb{R}),\mathcal{Q}_{2,2}\big)$. The proof relies on the following three lemmas.

Lemma 4.3. *Let $Z:\Omega\to\mathbb{R}$ be $\mathcal{N}(0,1)$-distributed. Then, $\forall z>0$, $\mathbb{P}(Z\geq z) \leq \frac{e^{-\frac{z^2}{2}}}{z\sqrt{2\pi}}$.*

Proof. $\mathbb{P}(Z\geq z) = \int_z^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}dx \leq \int_z^{+\infty} \frac{x}{z}\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}dx = \frac{e^{-\frac{z^2}{2}}}{z\sqrt{2\pi}}$. $\square$

Lemma 4.4. *Define (X_n) as in (4.12), then $\sup_{K\geq0} K\,\mathbb{E}(X_n-K)_+ \to 0$ as $n\to+\infty$.*

Proof. We have

$$
K\,\mathbb{E}(X_n-K)_+ = K\int_0^\infty \mathbb{P}\Big((X_n-K)_+ \geq u\Big)du = K\int_0^{+\infty} \mathbb{P}(X_n > u+K)du
$$

$$= K\int_K^{+\infty} \mathbb{P}(X_n \geq v)dv = K\int_K^{+\infty} \mathbb{P}\Big(e^{\frac{n}{2}Z-\frac{n^2}{4}} \geq v\Big)dv$$
$$= K\int_K^{+\infty} \mathbb{P}\Big(Z \geq \frac{n}{2} + \frac{2}{n}\ln v\Big)dv = K\int_{\ln K}^{\infty} \mathbb{P}\Big(Z \geq \frac{n}{2} + \frac{2}{n}u\Big)e^u du \ \text{ (setting } u = \ln v).$$

By Lemma 4.3, $\mathbb{P}\Big(Z \geq \frac{n}{2} + \frac{2}{n}u\Big) \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2}(\frac{n}{2}+\frac{2}{n}u)^2}}{\frac{n}{2}+\frac{2}{n}u} = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{n^2}{8}-\frac{2}{n^2}u^2-u}}{\frac{n}{2}+\frac{2}{n}u}$. It follows that,

$$K\,\mathbb{E}(X_n - K)_+ \leq K\int_{\ln K}^{\infty} \frac{e^{-\frac{n^2}{8}-\frac{2}{n^2}u^2-u}}{\frac{n}{2}+\frac{2}{n}u} e^u \frac{du}{\sqrt{2\pi}} \leq \frac{Ke^{-\frac{n^2}{8}}}{\frac{n}{2}+\frac{2}{n}\ln K} \int_{\ln K}^{\infty} e^{-\frac{2}{n^2}u^2} \frac{du}{\sqrt{2\pi}}$$
$$= \frac{Ke^{-\frac{n^2}{8}}}{\frac{n}{2}+\frac{2}{n}\ln K} \int_{\frac{2}{n}\ln K}^{\infty} e^{-\frac{w^2}{2}} \frac{n}{2} \frac{dw}{\sqrt{2\pi}} \ \text{ (by setting } w = \frac{2}{n}u)$$
$$= \frac{Ke^{-\frac{n^2}{8}}}{\frac{n}{2}+\frac{2}{n}\ln K} \frac{n}{2}\,\mathbb{P}\Big(Z \geq \frac{2}{n}\ln K\Big) \leq \frac{nKe^{-\frac{n^2}{8}}}{2(\frac{n}{2}+\frac{2}{n}\ln K)} \frac{e^{-\frac{1}{2}\frac{4}{n^2}(\ln K)^2}}{\sqrt{2\pi}\frac{2}{n}\ln K} \ \text{ (by Lemma 4.3)}$$
$$= \frac{n}{2\sqrt{2\pi}} e^{-\frac{n^2}{8}} \frac{Ke^{-\frac{2}{n^2}(\ln K)^2}}{(1+\frac{4}{n^2}\ln K)\ln K} = \frac{n}{2\sqrt{2\pi}} e^{-\frac{n^2}{8}} \frac{e^{\ln K(1-\frac{2}{n^2}\ln K)}}{(1+\frac{4}{n^2}\ln K)\ln K}. \tag{4.13}$$

Since the function $u \mapsto u(1-\frac{2}{n^2}u)$ attains its maximum at $u = \frac{n^2}{4}$ with maximum value $\frac{n^2}{8}$, we will discuss the value of $K\,\mathbb{E}(X_n - K)_+$ in the following three cases:

$$(i)\ K \geq e^{\frac{n^2}{4}}, \qquad (ii)\ e^{\rho\frac{n^2}{4}} \leq K \leq e^{\frac{n^2}{4}}, \qquad (iii)\ 0 \leq K \leq e^{\rho\frac{n^2}{4}},$$

with the same fixed $\rho \in (0, \frac{1}{2})$ in both (ii) and (iii).

Case (i): $K \geq e^{\frac{n^2}{4}}$, then $\ln K \geq \frac{n^2}{4}$. It follows that

$$K\,\mathbb{E}(X_n - K)_+ \leq \frac{n}{2\sqrt{2\pi}} e^{-\frac{n^2}{8}} \frac{e^{\ln K(1-\frac{2}{n^2}\ln K)}}{(1+\frac{4}{n^2}\ln K)\ln K} \leq \frac{n}{2\sqrt{2\pi}} e^{-\frac{n^2}{8}} \frac{e^{\frac{n^2}{8}}}{(1+\frac{4}{n^2}\times\frac{n^2}{4})\frac{n^2}{4}} = \frac{1}{n\sqrt{2\pi}} \to 0.$$

Case (ii): $e^{\rho\frac{n^2}{4}} \leq K \leq e^{\frac{n^2}{4}}$ with a fixed $\rho \in (0, \frac{1}{2})$, then $\rho\frac{n^2}{4} \leq \ln K \leq \frac{n^2}{4}$. It follows that

$$K\,\mathbb{E}(X_n - K)_+ \leq \frac{ne^{-\frac{n^2}{8}}}{2\sqrt{2\pi}} \frac{e^{\ln K(1-\frac{2}{n^2}\ln K)}}{(1+\frac{4}{n^2}\ln K)\ln K} \leq \frac{ne^{-\frac{n^2}{8}}}{2\sqrt{2\pi}} \frac{e^{\frac{n^2}{8}}}{(1+\frac{4}{n^2}\times\rho\frac{n^2}{4})\rho\frac{n^2}{4}} = \frac{2}{n(1+\rho)\rho\sqrt{2\pi}} \to 0.$$

Case (iii): $0 \leq K \leq e^{\rho\frac{n^2}{4}}$ with the same $\rho \in (0, \frac{1}{2})$ as in the situation (ii), then

$$K\,\mathbb{E}(X_n - K)_+ \leq e^{\frac{\rho}{4}n^2}\mathbb{E}X_n = e^{\frac{\rho}{4}n^2} \cdot e^{-\frac{n^2}{8}} = e^{\frac{1}{4}(\rho-\frac{1}{2})n^2} \xrightarrow{n\to+\infty} 0.$$

Therefore, $\sup_{K>0}\ K\,\mathbb{E}(X_n - K)_+ \xrightarrow{n\to+\infty} 0$. □

By Lemma 4.2, $\sup_{a\in\mathbb{R}^d} \Big|e_{2,2}\big(\mu_n,(a,a)\big) - \sqrt{e^2_{2,2}\big(\mu_\infty,(a,a)\big) + C_0}\Big| \xrightarrow{n\to+\infty} 0$. Consequently, it is reasonable to guess that $e_{N,2}(\mu_n,\cdot) \xrightarrow[n\to+\infty]{\|\cdot\|_{\sup}} \sqrt{e^2_{N,2}(\mu_\infty,\cdot)+1}$ so that $(\mu_n)_{n\in\mathbb{N}}$ is a Cauchy

sequence in $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{Q}_{N,2})$. Let $g_N : \mathbb{R}^N \to \mathbb{R}_+$ be defined by

$$(a_1, \dots, a_N) \mapsto g_N\big((a_1, \dots, a_N)\big) \coloneqq \sqrt{e^2_{N,2}\big(\mu_\infty, (a_1, \dots, a_N)\big) + 1} = \sqrt{\min_{1\le i\le N} |a_i|^2 + 1}.$$

Proposition 4.4. *For every $N \geq 2$,*

$$\sup_{(a_1,\dots,a_N)\in\mathbb{R}^N} \big|e_{N,2}\big(\mu_n, (a_1, \dots, a_N)\big) - g_N\big((a_1, \dots, a_N)\big)\big| \xrightarrow{n\to+\infty} 0.$$

Therefore, $(\mu_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $(\mathcal{P}_2(\mathbb{R}), \mathcal{Q}_{N,2})$ by the definition of $\mathcal{Q}_{N,2}$.

Proof. We proceed by induction.

$\triangleright$ $N = 2$. Since the functions g_2 and $e_{2,2}(\mu_n, \cdot)$ are symmetric, it is only necessary to show that $\sup_{(a,b)\in\mathbb{R}^2,\, |a|\le|b|} \big|e_{2,2}\big(\mu_n, (a, b)\big) - g_2(a, b)\big| \xrightarrow{n\to+\infty} 0$. Note that when $|a| \leq |b|$, $g_2(a, b) = \sqrt{|a|^2 + 1} = g_2(a, a)$. We discuss now the value of $\big|e_{2,2}\big(\mu_n, (a, b)\big) - g_2(a, b)\big|$ in the following four cases,

$$(i)\ 0 \le a \le b, \qquad (ii)\ a \le 0 \le b, \begin{cases} (ii,\alpha) & a \le 0 \le b \text{ with } |a| \le \frac{1}{2}|b| \\ (ii,\beta) & a \le 0 \le b \text{ with } \frac{1}{2}|b| \le |a| \le |b|, \end{cases}$$

$$(iii)\ b \le 0 \le a, \text{ with } |a| \le |b|, \qquad (iv)\ b \le a \le 0.$$

Cases (iii) and (iv): $\mathbf{b < 0}$ and $\mathbf{\frac{a+b}{2} < 0}$. The random variables X_n are positive so that $|x - a| \leq |x - b|$. Hence $e_{2,2}\big(\mu_n, (a, b)\big) = e_{2,2}\big(\mu_n, (a, a)\big)$. With a slight abuse of notation, we will write in what follows $(a, b) \in (iii)$ for $(a, b) \in \{(a, b) \in \mathbb{R}^2 \mid b \leq 0 \leq a, \text{and } |a| \leq |b|\}$. We will adopt the same notation for other cases too. Then for the case (iii) and (iv), it is obvious by applying Lemma 4.2 that

$$\sup_{(a,b)\in(iii)\cup(iv)} \big|e_{2,2}\big(\mu_n, (a, b)\big) - g_2(a, b)\big| = \sup_{(a,b)\in(iii)\cup(iv)} \big|e_{2,2}\big(\mu_n, (a, a)\big) - g_2(a, a)\big| \xrightarrow{n\to+\infty} 0.$$

Case (i): $\mathbf{0 \leq a \leq b}$. We have

$$\begin{aligned}
&\sup_{(a,b)\in(i)} \big|e_{2,2}\big(\mu_n, (a, b)\big) - g_2(a, b)\big| \\
&\quad\le \sup_{(a,b)\in(i)} \big|e_{2,2}\big(\mu_n, (a, b)\big) - e_{2,2}\big(\mu_n, (a, a)\big)\big| + \big|e_{2,2}\big(\mu_n, (a, a)\big) - g_2(a, a)\big| \\
&\quad\le \sup_{(a,b)\in(i)} \left|\sqrt{\int_{\mathbb{R}} |\xi - a|^2 \wedge |\xi - b|^2 \,\mu_n(d\xi)} - \sqrt{\int_{\mathbb{R}} |\xi - a|^2 \,\mu_n(d\xi)}\right| + \big|e_{2,2}\big(\mu_n, (a, a)\big) - g_2(a, a)\big| \\
&\quad\le \sup_{(a,b)\in(i)} \sqrt{\int_{\mathbb{R}} \Big[\, |\xi - a|^2 - \big(\, |\xi - a|^2 \wedge |\xi - b|^2 \,\big)\Big] \mu_n(d\xi)} + \big|e_{2,2}\big(\mu_n, (a, a)\big) - g_2(a, a)\big| \\
&\quad\quad \text{(since } \left|\sqrt{\alpha} - \sqrt{\beta}\right| \le \sqrt{\beta - \alpha} \text{ for } \beta > \alpha > 0) \\
&\quad\le \sup_{(a,b)\in(i)} \sqrt{\int_{\mathbb{R}} \Big(\, |\xi - a|^2 - |\xi - b|^2 \Big)_+ \mu_n(d\xi)} + \big|e_{2,2}\big(\mu_n, (a, a)\big) - g_2(a, a)\big| \\
&\quad\le \sup_{(a,b)\in(i)} \sqrt{\int_{\mathbb{R}} 2(b - a)\Big(\xi - \frac{b + a}{2}\Big)_+ \mu_n(d\xi)} + \big|e_{2,2}\big(\mu_n, (a, a)\big) - g_2(a, a)\big|
\end{aligned}$$

$$\leq \sup_{(a,b)\in(i)} 2\sqrt{\int_{\mathbb{R}} \frac{b}{2}\Big(\xi-\frac{b}{2}\Big)_+ \mu_n(d\xi)} + \big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|$$
$$\leq 2\sqrt{\sup_{K\geq 0} K\,\mathbb{E}(X_n-K)_+} + \sup_{a\in\mathbb{R}}\big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big| \xrightarrow{n\to+\infty} 0.$$

Case (ii,α): $\mathbf{a\leq 0\leq b}$, with $\mathbf{|a|\leq \frac{1}{2}\,|b|}$. We have

$$\sup_{(a,b)\in(ii,\alpha)}\big|e_{2,2}\big(\mu_n,(a,b)\big)-g_2(a,b)\big|$$
$$\leq \sup_{(a,b)\in(ii,\alpha)}\big|e_{2,2}\big(\mu_n,(a,b)\big)-e_{2,2}\big(\mu_n,(a,a)\big)\big| + \big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|$$
$$\leq \sup_{(a,b)\in(ii,\alpha)}\sqrt{\int_{\mathbb{R}} 2(b-a)\Big(\xi-\frac{b+a}{2}\Big)_+\mu_n(d\xi)} + \big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|$$
$$\leq \sup_{(a,b)\in(ii,\alpha)}\sqrt{\int_{\mathbb{R}} 3\cdot b\Big(\xi-\frac{b}{4}\Big)_+\mu_n(d\xi)} + \big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|$$
$$\leq 2\sqrt{3}\cdot\sqrt{\sup_{K\geq 0} K\,\mathbb{E}(X_n-K)_+} + \sup_{a\in\mathbb{R}}\big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big| \xrightarrow{n\to+\infty} 0.$$

Case (ii,β): $\mathbf{a\leq 0\leq b}$, with $\mathbf{\frac{1}{2}\,|b|\leq |a|\leq |b|}$. One has

$$\sup_{(a,b)\in(ii,\beta)}\big|e_{2,2}\big(\mu_n,(a,b)\big)-g_2(a,b)\big|$$
$$\leq \sup_{(a,b)\in(ii,\beta)}\big|e_{2,2}\big(\mu_n,(a,b)\big)-e_{2,2}\big(\mu_n,(a,a)\big)\big| + \big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|$$
$$\leq \sup_{(a,b)\in(ii,\beta)}\frac{\big|e_{2,2}^2\big(\mu_n,(a,b)\big)-e_{2,2}^2\big(\mu_n,(a,a)\big)\big|}{e_{2,2}\big(\mu_n,(a,b)\big)+e_{2,2}\big(\mu_n,(a,a)\big)} + \big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|$$
$$\leq \sup_{(a,b)\in(ii,\beta)}\frac{\int_{\mathbb{R}} 2(b-a)\big(\xi-\frac{b+a}{2}\big)_+\mu_n(d\xi)}{\|X_n-a\|_2} + \sup_{a\in\mathbb{R}}\big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|$$
$$\leq \sup_{(a,b)\in(ii,\beta)}\frac{2(b-a)\,\mathbb{E}\left(X_n-\frac{b+a}{2}\right)_+}{\|X_n-a\|_2} + \sup_{a\in\mathbb{R}}\big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|.$$

As $\|X_n-a\|_2 = \big(\underbrace{\mathbb{E}X_n^2}_{=1}\underbrace{-2a\,\mathbb{E}X_n}_{\geq 0}+|a|^2\big)^{1/2}\geq\sqrt{1+|a|^2}$, we have

$$\sup_{(a,b)\in(ii,\beta)}\big|e_{2,2}\big(\mu_n,(a,b)\big)-g_2(a,b)\big|$$
$$\leq \sup_{(a,b)\in(ii,\beta)}\frac{2(b+|a|)\mathbb{E}\big[X_n-\frac{b+a}{2}\big]_+}{\sqrt{1+|a|^2}} + \sup_{a\in\mathbb{R}}\big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|$$
$$\leq \sup_{(a,b)\in(ii,\beta)}\frac{4b\,\mathbb{E}X_n}{\sqrt{1+\frac{b^2}{4}}} + \sup_{a\in\mathbb{R}}\big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big|.$$
$$\leq 8\,\mathbb{E}X_n + \sup_{a\in\mathbb{R}}\big|e_{2,2}\big(\mu_n,(a,a)\big)-g_2(a,a)\big| \xrightarrow{n\to+\infty} 0.$$

▷ **From N to $N+1$.** Assume now that $\sup_{(a_1,\ldots,a_N)\in\mathbb{R}^N}\big|e_{N,2}\big(\mu_n,(a_1,\ldots,a_N)\big)-g_N(a_1,\ldots,a_N)\big|$ goes 0 as

$n \to +\infty$. Then, for the level $N+1$, we assume without loss of generality that $|a_1| \le |a_2| \le \cdots \le |a_{N+1}|$ since g_{N+1} and $e_{N,2}(\mu_n, \cdot)$ are symmetric. Under this assumption,

$$g_{N+1}(a_1, \ldots, a_{N+1}) = g_2(a_1, a_1) = \sqrt{|a_1|^2 + 1}. \tag{4.14}$$

We discuss now the value of $\sup_{(a_1,\ldots,a_{N+1})\in\mathbb{R}^{N+1}} \big|e_{N+1,2}\big(\mu_n, (a_1, \ldots, a_{N+1})\big) - g_{N+1}(a_1, \ldots, a_{N+1})\big|$ in the following cases:

(i) $\exists\, i \in \{2, \ldots, N+1\}$ such that $a_i < 0$, $\qquad$ (ii) $0 \le a_1 \le a_2 \le \cdots \le a_{N+1}$,

$$(iii)\ a_1 \le 0 \le a_2 \le \cdots \le a_{N+1}, \begin{cases} (iii,\alpha)\ a_1 \le 0 \le a_2 \le \cdots \le a_{N+1}, \text{ with } |a_1| \le \frac{1}{2}\,|a_{N+1}| \\ (iii,\beta)\ a_1 \le 0 \le a_2 \le \cdots \le a_{N+1}, \text{ with } \frac{1}{2}\,|a_{N+1}| \le |a_1| \le |a_{N+1}| \end{cases}.$$

Case (i): $\exists\, \mathbf{i} \in \{2, \ldots, N+1\}$ such that $\mathbf{a_i < 0}$. For every $n \ge 1$, X_n is a.s. positive. Hence, $|X_n - a_1| \le |X_n - a_i|$ a.s. since we assume that $|a_1| \le |a_2| \le \cdots \le |a_{N+1}|$. Therefore,

$$e_{N+1,2}\big(\mu_n, (a_1, \ldots, a_{N+1})\big) = e_{N,2}\big(\mu_n, (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{N+1})\big).$$

It follows from (4.14) that

$$\begin{aligned} &\sup_{(a_1,\ldots,a_{N+1})\in\mathbb{R}^{N+1}} \big|e_{N+1,2}\big(\mu_n, (a_1, \ldots, a_{N+1})\big) - g_{N+1}(a_1, \ldots, a_{N+1})\big| \\ &= \sup_{(a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_{N+1})\in\mathbb{R}^{N}} \big|e_{N,2}\big(\mu_n, (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{N+1})\big) - g_N(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_{N+1})\big|, \end{aligned}$$

which converges to 0 as $n \to +\infty$ owing to the assumption on the level N.

Case (ii): $\mathbf{0 \le a_1 \le a_2 \le \cdots \le a_{N+1}}$.

$$\begin{aligned} &\sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \big|e_{N+1,2}\big(\mu_n, (a_1, \ldots, a_{N+1})\big) - g_{N+1}(a_1, \ldots, a_{N+1})\big| \\ &\qquad \le \sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \big|e_{N+1,2}\big(\mu_n, (a_1, \ldots, a_{N+1})\big) - e_{N,2}\big(\mu_n, (a_1, \ldots, a_N)\big)\big| \\ &\qquad\quad + \sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \big|e_{N,2}\big(\mu_n, (a_1, \ldots, a_N)\big) - g_N(a_1, \ldots, a_N)\big|. \end{aligned} \tag{4.15}$$

The second term on the right hand side of (4.15) converges to 0 as $n \to +\infty$ owing to the assumption on the level N.

For the first term on the right hand side of (4.15), we have

$$\begin{aligned} &\sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \big|e_{N+1,2}\big(\mu_n, (a_1, \ldots, a_{N+1})\big) - e_{N,2}\big(\mu_n, (a_1, \ldots, a_N)\big)\big| \\ &= \sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \sqrt{\int_{\mathbb{R}} \min_{1\le i\le N} |\xi - a_i|^2\, \mu_n(d\xi)} - \sqrt{\int_{\mathbb{R}} \Big[\min_{1\le i\le N} |\xi - a|^2\Big] \wedge |\xi - a_{N+1}|^2\, \mu_n(d\xi)} \\ &\le \sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \sqrt{\int_{\mathbb{R}} \min_{1\le i\le N} |\xi - a_i|^2 - \Big[\min_{1\le i\le N} |\xi - a_i|^2\Big] \wedge |\xi - a_{N+1}|^2\, \mu_n(d\xi)} \\ &= \sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \sqrt{\int_{\mathbb{R}} \big(\min_{1\le i\le N} |\xi - a_i|^2 - |\xi - a_{N+1}|^2\big)_+ \mu_n(d\xi)} \\ &\le \sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \sqrt{\int_{\mathbb{R}} \big(|\xi - a_1|^2 - |\xi - a_{N+1}|^2\big)_+ \mu_n(d\xi)} \end{aligned}$$

$$= \sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \sqrt{\int_{\mathbb{R}} 2(a_{N+1}-a_1)\big(\xi-\frac{a_1+a_{N+1}}{2}\big)_+\mu_n(d\xi)}$$
$$\le \sup_{0\le a_1\le a_2\le\cdots\le a_{N+1}} \sqrt{\int_{\mathbb{R}} 2\cdot a_{N+1}\big(\xi-\frac{a_{N+1}}{2}\big)_+\mu_n(d\xi)} \le 2\cdot\sqrt{\sup_{K\ge 0} K\,\mathbb{E}(X_n-K)_+} \xrightarrow{n\to+\infty} 0.$$

Case (iii,α): $\mathbf{a_1\le 0\le a_2\le\cdots\le a_{N+1}}$ with $\mathbf{|a_1|\le\frac{1}{2}|a_{N+1}|}$.

$$\sup_{(a_1,\ldots,a_{N+1})\in(iii,\alpha)} \big|e_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)-g_{N+1}(a_1,\ldots,a_{N+1})\big|$$
$$\le \sup_{(a_1,\ldots,a_{N+1})\in(iii,\alpha)} \big|e_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)-e_{N,2}\big(\mu_n,(a_1,\ldots,a_N)\big)\big|$$
$$+ \sup_{(a_1,\ldots,a_{N+1})\in(iii,\alpha)} \big|e_{N,2}\big(\mu_n,(a_1,\ldots,a_N)\big)-g_N(a_1,\ldots,a_N)\big|. \tag{4.16}$$

Like in Case (ii), the second term on the right hand side of (4.16) converges to 0 as $n\to+\infty$. For the first term of the right hand side of (4.16), we have

$$\sup_{(a_1,\ldots,a_{N+1})\in(iii,\alpha)} \big|e_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)-e_{N,2}\big(\mu_n,(a_1,\ldots,a_N)\big)\big|$$
$$\le \sup_{(a_1,\ldots,a_{N+1})\in(iii,\alpha)} \sqrt{\int_{\mathbb{R}} 2(a_{N+1}-a_1)\big(\xi-\frac{a_1+a_{N+1}}{2}\big)_+\mu_n(d\xi)}$$
$$\le \sup_{(a_1,\ldots,a_{N+1})\in(iii,\alpha)} \sqrt{\int_{\mathbb{R}} 3\cdot a_{N+1}\big(\xi-\frac{a_{N+1}}{4}\big)_+\mu_n(d\xi)} \le 2\sqrt{3}\cdot\sqrt{\sup_{K\ge 0} K\,\mathbb{E}(X_n-K)_+} \longrightarrow 0.$$

Case (iii,β): $\mathbf{a_1\le 0\le a_2\le\cdots\le a_{N+1}}$ with $\mathbf{\frac{1}{2}|a_{N+1}|\le|a_1|\le|a_{N+1}|}$.

Since we assume $|a_1|\le|a_2|\le\cdots\le|a_{N+1}|$, then for any $i\in\{2,\ldots,N+1\}$, we have $\frac{1}{2}|a_i|\le|a_1|\le|a_i|$. It follows that

$$\sup_{(a_1,\ldots,a_{N+1})\in(iii,\beta)} \big|e_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)-g_{N+1}(a_1,\ldots,a_{N+1})\big|$$
$$\le \sup_{(a_1,\ldots,a_{N+1})\in(iii,\beta)} \big|e_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)-e_{2,2}\big(\mu_n,(a_1,a_1)\big)\big|$$
$$+ \sup_{a_1\in\mathbb{R}} \big|e_{2,2}\big(\mu_n,(a_1,a_1)\big)-g_N(a_1,a_1)\big|. \tag{4.17}$$

The second part of (4.17), $\sup_{a_1\in\mathbb{R}} \big|e_{2,2}\big(\mu_n,(a_1,a_1)\big)-g_N(a_1,a_1)\big|$ converges to 0 as $n\to+\infty$ owing to Lemma 4.2. Then for the first part of (4.17), we have

$$\sup_{(a_1,\ldots,a_{N+1})\in(iii,\beta)} \big|e_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)-e_{2,2}\big(\mu_n,(a_1,a_1)\big)\big|$$
$$= \sup_{(a_1,\ldots,a_{N+1})\in(iii,\beta)} \frac{e^2_{2,2}\big(\mu_n,(a_1,a_1)\big)-e^2_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)}{e_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)+e_{2,2}\big(\mu_n,(a_1,a_1)\big)}$$
$$\le \sup_{(a_1,\ldots,a_{N+1})\in(iii,\beta)} \frac{\int_{\mathbb{R}}|\xi-a_1|^2-\min_{1\le i\le N+1}|\xi-a_i|^2\,\mu_n(d\xi)}{\|X_n-a_1\|_2}$$
$$\le \sup_{(a_1,\ldots,a_{N+1})\in(iii,\beta)} \frac{\int_{\mathbb{R}}\big(|\xi-a_1|^2-\min_{2\le i\le N+1}|\xi-a_2|^2\big)_+\mu_n(d\xi)}{\|X_n-a_1\|_2}$$

$$\leq \sup_{(a_1,\ldots,a_{N+1})\in(iii,\beta)} \frac{1}{\|X_n - a_1\|_2}\Big[\sum_{i=2}^{N+1}\int_{\mathbb{R}}\big(|\xi-a_1|^2-|\xi-a_i|^2\big)_+\mu_n(d\xi)\Big]$$

Since $a_1 < 0$, $\|X_n - a_1\|_2 = \big(\mathbb{E}X_n^2 - 2a_1\mathbb{E}X_n + |a_1|^2\big)^{1/2} \geq \sqrt{1+|a_1|^2}$. Therefore,

$$\begin{aligned}\frac{\int_{\mathbb{R}}\big(|\xi-a_1|^2-|\xi-a_i|^2\big)_+\mu_n(d\xi)}{\|X_n-a_1\|_2} &= \frac{\int_{\mathbb{R}} 2(a_i-a_1)\big(\xi-\frac{a_i+a_1}{2}\big)_+\mu_n(d\xi)}{\|X_n-a_1\|_2}\\ &\leq \frac{4a_i\mathbb{E}X_n}{\sqrt{1+|a_1|^2}} \leq \frac{4a_i\mathbb{E}X_n}{\frac{1}{2}a_i} = 8\,\mathbb{E}X_n.\end{aligned}$$

for $i\in\{2,\ldots,N+1\}$. Consequently,

$$\sup_{(a_1,\ldots,a_{N+1})\in(iii,\beta)}\big|e_{N+1,2}\big(\mu_n,(a_1,\ldots,a_{N+1})\big)-e_{2,2}\big(\mu_n,(a_1,a_1)\big)\big| \leq 8N\,\mathbb{E}X_n = 8\,Ne^{-n^2/8}\longrightarrow 0.$$

This completes the proof. □

Proof of Theorem 4.1. Let μ_n be the probability distribution of X_n defined in (4.12). If for some $N \geq 2$, $\big(\mathcal{P}_2(\mathbb{R}),\mathcal{Q}_{N,2}\big)$ were complete, then there exists a probability measure $\widetilde{\mu}$ in $\mathcal{P}_2(\mathbb{R})$ such that $\mathcal{Q}_{N,2}(\mu_n,\widetilde{\mu}) \longrightarrow 0$. Then, $\mathcal{W}_2(\mu_n,\widetilde{\mu}) \longrightarrow 0$ by applying Proposition 4.2, which creates a contradiction. □

Remark 4.1. The extension of this result to a Hilbert or simply multidimensional setting, although likely, is not straightforward.

Appendix: some examples of $c(d,|\cdot|_r)$

Proof of Proposition 2.4. (i) is obvious.
(ii) $c(2,|\cdot|_1)=2$ is obvious (see Figure 1). Now we prove that $c(2,|\cdot|_r)=3$ for every $r\in(1,+\infty)$.

We choose $a_1=(0,1)$, $a_2=\big((1-2^{-r})^{\frac{1}{r}},-\frac{1}{2}\big)$ and $a_3=\big(-(1-2^{-r})^{\frac{1}{r}},-\frac{1}{2}\big)$. We will first show that $S_{|\cdot|_r}(0,1)\subset\bigcup_{1\leq i\leq 3}\bar{B}_{|\cdot|_r}(a_i,1)$.

Let (x,y) be any point on $S_{|\cdot|_r}(0,1)$, then $|x|^r+|y|^r=1$.

- If $\frac{1}{2}\leq y\leq 1$, then $(1-y)^r\leq y^r$ so that $\big|(x,y)-a_1\big|_r^r=|x|^r+(1-y)^r=1-y^r+(1-y)^r\leq 1$, that is, $(x,y)\in\bar{B}_{|\cdot|_r}(a_1,1)$.
- If $-1\leq y\leq\frac{1}{2}$ and $x\geq 0$, then

$$\begin{aligned}\big|(x,y)-a_2\big|_r^r &= \big|x-(1-2^{-r})^{\frac{1}{r}}\big|^r+\big|y+\frac{1}{2}\big|^r = \big|(1-|y|^r)^{\frac{1}{r}}-(1-2^{-r})^{\frac{1}{r}}\big|^r+\Big|y+\frac{1}{2}\Big|^r\\ &\leq \big||y|^r-2^{-r}\big|+\big|y+\frac{1}{2}\big|^r,\end{aligned}$$

the last inequality is due to the fact that the function $u\mapsto u^{-\frac{1}{r}}$ is $\frac{1}{r}$-Hölder. As $r\geq 1$, the function $y\mapsto\big||y|^r-2^{-r}\big|+\big|y+\frac{1}{2}\big|^r$ is convex over $[-1,\frac{1}{2}]$. Consequently, it attains its

maximum either at -1 or at $\frac{1}{2}$. Hence, $\big|(x,y)-a_2\big|_r^r$ is upper bounded by 1 since

$$\text{if } y=-1,\ \big||y|^r-2^{-r}\big|+\big|y+\tfrac{1}{2}\big|^r=1-2^{-r}+2^{-r}=1,$$
$$\text{if } y=\tfrac{1}{2},\quad \Big||y|^r-2^{-r}\Big|+\big|y+\frac{1}{2}\big|^r=\big|2^{-r}-2^{-r}\big|+1^r=1.$$

This implies that $(x,y)\in \bar{B}_{|\cdot|_r}(a_2,1)$.

- If $-1\le y\le \frac{1}{2}$ and $x\le 0$, then $(x,y)\in \bar{B}_{|\cdot|_r}(a_3,1)$ by the symmetry of the unit sphere.

Next, we will show $c(2,|\cdot|_r)>2$ for every $1<r<+\infty$. Let a_1 and a_2 denote the two centers of balls on the sphere $S_{|\cdot|}(0,1)$. Since the ℓ^r-ball is centrally symmetric with respect to $(0,0)$, we fix $a_1=(x,y)$ such that $x\in[(\frac{1}{2})^{\frac{1}{r}},1]$, $y\in[0,(\frac{1}{2})^{\frac{1}{r}}]$ and $x^r+y^r=1$.

We first prove that if $r>1$, $x\in[(\frac{1}{2})^{\frac{1}{r}},1]$, $y\in(0,(\frac{1}{2})^{\frac{1}{r}}]$ s.t. $x^r+y^r=1$, then $(x+y)^r>1$. Let $q=r-1$, then $q>0$ and

$$\begin{aligned}(x+y)^r&=(x+y)^{1+q}=(x+y)(x+y)^q=x(x+y)^q+y(x+y)^q\\&>xx^q+yy^q=x^r+y^r=1.\end{aligned}$$

- *Case 1.* We choose a_2 such that a_2 is centrally symmetric to a_1 with respect to the center $(0,0)$, i.e. $a_2=(-x,-y)$.

  We prove $z_1=(y,-x)\notin\cup_{i=1,2}\bar{B}_{|\cdot|_r}(a_i,1)$ and $z_2=(-y,x)\notin\cup_{i=1,2}\bar{B}_{|\cdot|_r}(a_i,1)$. In fact, if $y=0$, then $\big|a_1-z_1\big|_r=\big|a_2-z_1\big|_r=2>1$. If $y>0$, then

$$\begin{aligned}&\big|a_1-z_1\big|_r^r=\big|a_2-z_1\big|_r^r=\big|a_1-z_2\big|_r^r=\big|a_2-z_2\big|_r^r=(x+y)^r+(x-y)^r\\&\ge(x+y)^r>1\end{aligned}$$

- *Case 2.* The point a_2 is not centrally symmetric to a_1.

  Let $H_{a_1}:=\{\eta=(\eta_1,\eta_2)\in\mathbb{R}^2 \text{ s.t. } x\cdot\eta_2=y\cdot\eta_1\}$, which is the straight line (with respect to the Euclidean distance) across the origin and a_1. Then between z_1 and z_2, there exists at least one point which is not in the same side of H_{a_1} as a_2, and this point can not be covered by $\cup_{i=1,2}\bar{B}_{|\cdot|_r}(a_i,1)$.

Figure 2 illustrates that $c(2,|\cdot|_r)=3$ when $r=3$.

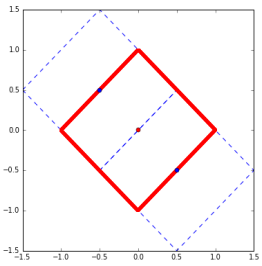


Figure 1: $a_1=(-\frac{1}{2},\frac{1}{2})$, $a_2=(\frac{1}{2},-\frac{1}{2})$, then $S_{|\cdot|_1}(0,1)\subset\bigcup_{i=1,2}\bar{B}_{|\cdot|_1}(a_i,1)$

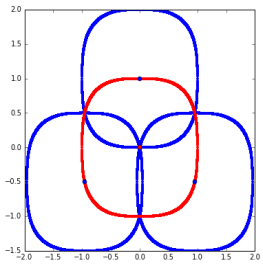


Figure 2: $c(2,|\cdot|_3)=3$

(iii) Let $a_1=(-1,0,\dots,0)$ and $a_2=(1,0,\dots,0)$. We will show that $S_{|\cdot|_\infty}(0,1)\subset\bigcup_{i=1,2}\bar{B}_{|\cdot|_\infty}(a_i,1)$.

Let $x=(x^1,\dots,x^d)\in S_{|\cdot|_\infty}(0,1)$. There exists i_0 such that $\max_{1\le i\le d}|x^i|\le|x^{i_0}|=1$.

- If $i_0 = 1$, and $x^1 = -1$, then $\left|x - a_1\right|_\infty = \left|x^1 + 1\right| \vee \max_{i=\{2,\dots,d\}} \left|x^i\right| \leq 1$, that is, $x \in \bar{B}_{|\cdot|_\infty}(a_1, 1)$.
- If $i_0 = 1$, and $x^1 = 1$, then $\left|x - a_2\right|_\infty = \left|x^1 - 1\right| \vee \max_{i=\{2,\dots,d\}} \left|x^i\right| \leq 1$, that is, $x \in \bar{B}_{|\cdot|_\infty}(a_2, 1)$.
- If $i_0 \geq 2$, and $x^1 \leq 0$, then $\left|x - a_1\right|_\infty = \left|x^1 + 1\right| \vee 1 \leq 1$, that is, $x \in \bar{B}_{|\cdot|_\infty}(a_1, 1)$.
- If $i_0 \geq 2$, and $x^1 \geq 0$, then $\left|x - a_2\right|_\infty = \left|x^1 - 1\right| \vee 1 \leq 1$, that is, $x \in \bar{B}_{|\cdot|_\infty}(a_2, 1)$.

Consequently, we conclude that $S_{|\cdot|_\infty}(0,1) \subset \bigcup_{i=1,2} \bar{B}_{|\cdot|_\infty}(a_i, 1)$ and $c(d, |\cdot|_\infty) > 1$ is obvious.

(iv) Let $a_i = (0, \dots, 1, \dots, 0)$ - the i^{th} coordinate of a_i is equal to 1 and the others equal to 0. We will show that $S_{|\cdot|_r}(0,1) \subset \bigcup_{i=1}^d \left(\bar{B}_{|\cdot|_r}(a_i, 1) \cup \bar{B}_{|\cdot|_r}(-a_i, 1)\right)$.

For any $x = (x^1, \dots, x^d) \in S_{|\cdot|_r}(0,1)$, then there exists $i_0 \in \{1, \dots, d\}$ such that $\left|x^{i_0}\right| \geq \frac{1}{2}$. Otherwise $1 = \sum_{1 \leq i \leq d} \left|x^i\right|^r < d \times 2^{-r} \leq 1$, which yields a contradiction.

• If $x^{i_0} \geq \frac{1}{2}$, then $\left|x - a_{i_0}\right|^r = (1 - x^{i_0})^r + \sum_{i \neq i_0} \left|x^i\right|^r = (1 - x^{i_0})^r + 1 - (x^{i_0})^r$. As $x^{i_0} \leq \frac{1}{2}$, we have $(1 - x^{i_0})^r - (x^{i_0})^r \leq 0$, so that $\left|x - a_{i_0}\right|^r \leq 1$, which implies that $x \in \bar{B}_{|\cdot|_r}(a_{i_0}, 1)$.

• If $x^{i_0} \leq -\frac{1}{2}$, one can similarly prove that $x \in \bar{B}_{|\cdot|_r}(-a_{i_0}, 1)$.

Consequently, we can conclude that $S_{|\cdot|_r}(0,1) \subset \bigcup_{i=1}^d \left(\bar{B}_{|\cdot|_r}(a_i, 1) \cup \bar{B}_{|\cdot|_r}(-a_i, 1)\right)$. □

References

[Bol08] François Bolley. Separability and completeness for the Wasserstein distance. In *Séminaire de probabilités XLI*, pages 371–377. Springer, 2008.

[BP03] Vlad Bally and Gilles Pagès. A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6):1003–1049, 2003.

[BPP05] Vlad Bally, Gilles Pagès, and Jacques Printems. A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance*, 15(1):119–168, 2005.

[BPR15] Patrizia Berti, Luca Pratelli, and Pietro Rigo. Gluing lemmas and Skorohod representations. *Electron. Commun. Probab.*, 20:no. 53, 11, 2015.

[CM88] J. A. Cuesta and C. Matrán. The strong law of large numbers for k-means and best possible nets of Banach valued random variables. *Probab. Theory Related Fields*, 78(4):523–534, 1988.

[GG12] Allen Gersho and Robert M Gray. *Vector quantization and signal compression*, volume 159. Springer Science & Business Media, 2012.

[GL00] Siegfried Graf and Harald Luschgy. *Foundations of quantization for probability distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000.

[Gra14] Loukas Grafakos. *Classical Fourier analysis*, volume 249 of *Graduate Texts in Mathematics*. Springer, New York, third edition, 2014.

[HE15] Tailen Hsing and Randall Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons, 2015.

[IEE82] IEEE Transactions on Information Theory. *IEEE Trans. Inform. Theory*, 28(2), 1982.

[Kel75] John L. Kelley. *General topology*. Springer-Verlag, New York-Berlin, 1975. Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 27.

[Kie82] John C. Kieffer. Exponential rate of convergence for Lloyd's method. I. *IEEE Trans. Inform. Theory*, 28(2):205–210, 1982.

[Lac82] Ivan B. Lacković. On the behaviour of sequences of left and right derivatives of a convergent sequence of convex functions. *Univerzitet u Beogradu. Publikacije Elektrotehničkog Fakulteta. Serija Matematika i Fizika*, pages 19–27, 1982.

[LP02] Harald Luschgy and Gilles Pagès. Functional quantization of Gaussian processes. *J. Funct. Anal.*, 196(2):486–531, 2002.

[Pag98] Gilles Pagès. A space quantization method for numerical integration. *J. Comput. Appl. Math.*, 89(1):1–38, 1998.

[Pag15] Gilles Pagès. Introduction to vector quantization and its applications for numerics. *CEMRACS 2013—modelling and simulation of complex systems: stochastic and deterministic approaches*, 48:29–79, 2015.

[PP03] Gilles Pagès and Jacques Printems. Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods Appl.*, 9(2):135–165, 2003.

[PS18] Gilles Pagès and Abass Sagna. Improved error bounds for quantization based numerical schemes for BSDE and nonlinear filtering. *Stochastic Process. Appl.*, 128(3):847–883, 2018.

[PY16] Gilles Pagès and Jun Yu. Pointwise convergence of the Lloyd I algorithm in higher dimension. *SIAM J. Control Optim.*, 54(5):2354–2382, 2016.

[Rud91] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.

[Top74] Flemming Topsoe. Compactness and tightness in a space of measures with the topology of weak convergence. *Math. Scand.*, 34:187–210, 1974.

[Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[Vil09] Cédric Villani. *Optimal transport, Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.