



**HAL**  
open science

# Uniform decomposition of probability measures: quantization, clustering and rate of convergence

Julien Chevallier

► **To cite this version:**

Julien Chevallier. Uniform decomposition of probability measures: quantization, clustering and rate of convergence. *Journal of Applied Probability*, 2018, 55 (4), pp.1037-1045. 10.1017/jpr.2018.69 . hal-01677877v2

**HAL Id: hal-01677877**

**<https://hal.science/hal-01677877v2>**

Submitted on 13 Nov 2018

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

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

# Uniform decomposition of probability measures: quantization, clustering and rate of convergence.

Julien Chevallier\*

Univ. Grenoble Alpes, CNRS, Grenoble INP<sup>†</sup> LJK, 38000 Grenoble, France

## Abstract

The study of finite approximations of probability measures has a long history. In (Xu and Berger, 2017), the authors focus on constrained finite approximations and, in particular, uniform ones in dimension  $d = 1$ . The present paper gives an elementary construction of a uniform decomposition of probability measures in dimension  $d \geq 1$ . This decomposition is then used to give upper-bounds on the rate of convergence of the optimal uniform approximation error. These bounds appear to be the generalization of the ones obtained in (Xu and Berger, 2017) and to be sharp for generic probability measures.

*Keywords:* Uniform approximation, Wasserstein distance, rate of convergence, quantization, clustering.

*Mathematical Subject Classification:* 60E15, 62E17, 60B10, 60F99.

## I Introduction

Finding a good finite decomposition of a given probability measure  $\rho$  on  $\mathbb{R}^d$  is an extensively studied problem. *Quantization* is concerned with the best finitely supported approximation of a probability measure (empirical measures being especially studied for *clustering*). The origins come from signal processing (optimal signal transmission through discretization) [1] but the range of application widened since then (pattern recognition [8], numerical analysis [15], economics [16]). The goodness of the approximation is usually measured in terms of an  $L^p$ -Wasserstein distance  $W_p$  and numerous results are concerned with the rate of convergence of  $e_{p,n}(\rho) := \inf W_p(\rho^{(n)}, \rho)$  to 0 where the infimum is taken with respect to the set of measures  $\rho^{(n)}$  supported by at most  $n$  atoms [11].

*Random empirical quantization* has recently attracted much attention [2, 3, 9] in particular for its application to mean-field interacting particle systems. In that case, the approximating measure is  $R^{(n)} = n^{-1} \sum_{k=1}^n \delta_{X_k}$  where the  $X_k$ 's are i.i.d. random variables distributed according to  $\rho$  and the main results are concerned with rate of convergence of  $\mathbb{E} [W_p(R^{(n)}, \rho)]$  or concentration inequalities of the random variable  $W_p(R^{(n)}, \rho)$ .

In that context, when the approximating measure is  $\mu^{(n)} = n^{-1} \sum_{k=1}^n \delta_{x_k}$  with deterministic  $x_k$ 's, we use the term *deterministic empirical quantization*. This kind of approximation is used for instance when considering mean-field limits with spatial covariates used to weight the interactions between particles [6] (the present paper is mainly inspired from a technical lemma in [6]). In that paper, the mean-field approximation is decomposed into two terms: a variance/fluctuation term (the one appearing in mean-field limits without spatial covariates) and a bias term coming from the approximation of the limit ( $n = +\infty$ ) distribution of particles by the empirical distribution of particles ( $n < +\infty$ ). The control of the bias term falls into deterministic empirical quantization theory. The case of dimension  $d = 1$  is extensively addressed in [17] (the study highly relies on the connection between Wasserstein distances and the quantile function which is specific to  $d = 1$ ). The aim of the present paper is to generalize some of the results stated in [17] to the general case  $d \geq 1$  (notice that the case of a Hilbert space with  $p = 2$  is treated in [10]). The main result gives sharp bounds on the rate of convergence of  $\tilde{e}_{p,n}(\rho) := \inf W_p(\mu^{(n)}, \rho)$  to 0 where the infimum

---

\*e-mail: [julien.chevallier1@univ-grenoble-alpes.fr](mailto:julien.chevallier1@univ-grenoble-alpes.fr)

<sup>†</sup>Institute of Engineering Univ. Grenoble Alpes

is taken with respect to the set of deterministic empirical measures  $\mu^{(n)}$  supported by  $n$  atoms. The rate of convergence depends on the dimension  $d$  and the order  $p$  and shows a transition: it is either the same as for standard quantization (when Lebesgue measure is harder to approximate) or strictly worse (when disconnected measures are harder to approximate).

Let us mention here the related field of quasi-Monte Carlo methods. Mainly devoted to numerical integration with respect to the Lebesgue measure [7], some recent papers deal with any continuous probability distribution by providing a representative point set which minimizes some prescribed energy function [5, 13].

When the measure  $\rho$  is an empirical measure (of some given sample for instance), the result is related to *balanced clustering* in the sense of [14] which is a particular case of constrained clustering [4]. This balance property is advantageous when distributing the workload between travelling salesmen for instance.

The paper is organized as follows. Definitions and notation are given in Section II with a list of previous results found in the literature. Then, Section III contains an elementary uniform decomposition of probability measures (Theorem III.2) which is used to obtain upper-bounds on deterministic empirical quantization rates (Theorem III.3) and balanced clustering rates (Corollary III.6).

## II Notation and previous results

The space  $\mathbb{R}^d$  is equipped with the maximum norm  $\|\cdot\|$  and the balls centered at 0 are denoted by  $\mathcal{B}_r := B(0, r) = [-r, r]^d$  for all  $r \geq 0$ . The diameter of a subset  $A$  of  $\mathbb{R}^d$  is denoted by  $\text{Diam}(A) := \sup_{x, y \in A} \|x - y\|$ . The space of every Borel measures (*resp.* probability measures) on  $\mathbb{R}^d$  is denoted by  $\mathcal{M}(\mathbb{R}^d)$  (*resp.*  $\mathcal{P}(\mathbb{R}^d)$ ). For  $\nu$  in  $\mathcal{M}(\mathbb{R}^d)$ ,  $\text{Supp}(\nu)$  and  $|\nu| := \nu(\mathbb{R}^d)$  respectively denote the support and the mass of the measure  $\nu$ . For a collection of  $n$  positions  $x_1, \dots, x_n$  in  $\mathbb{R}^d$ , we denote its associated empirical measure by  $\mu^{(n)} := n^{-1} \sum_{k=1}^n \delta_{x_k}$ .

For every  $p \geq 1$ , the set of probability measures  $\rho$  such that  $\int \|x\|^p \rho(dx) < +\infty$  is denoted by  $\mathcal{P}_p$ . Then the Wasserstein distance of order  $p$  is denoted by  $W_p$  and defined by, for all  $\rho$  and  $\mu$  in  $\mathcal{P}_p$ ,

$$W_p(\rho, \mu) := \left( \inf_{\pi} \int_{(\mathbb{R}^d)^2} \|x - y\|^p \pi(dx, dy) \right)^{1/p},$$

where the infimum is taken with respect to every couplings  $\pi$  of the two measures  $\rho$  and  $\mu$ .

**Definition II.1.** Let  $\rho$  be in  $\mathcal{P}(\mathbb{R}^d)$ ,  $n \in \mathbb{N}^*$  and  $\rho_1, \dots, \rho_n$  be in  $\mathcal{M}(\mathbb{R}^d)$ . If  $\rho = \sum_{k=1}^n \rho_k$  and  $|\rho_k| = \frac{1}{n}$  for all  $k$  then we call  $(\rho_1, \dots, \rho_n)$  a uniform decomposition of  $\rho$  (of size  $n$ ).

### II.1 State of the art

Given  $\rho$  in  $\mathcal{P}_p$  the *optimal quantization error* of order  $p$  is defined as

$$e_{p,n}(\rho) := \inf_{\rho^{(n)}} W_p(\rho^{(n)}, \rho),$$

where the infimum is taken with respect to the set of measures  $\rho^{(n)}$  supported by at most  $n$  atoms. The literature dealing with the rate of convergence of  $e_{p,n}(\rho)$  to 0 is extensive [11, 12, 18, 19]. One of the most celebrated result is due to Zador [11, Theorem 6.2]. A consequence says that if  $\rho$  is in  $\mathcal{P}_q$  for some  $q > p$  and admits a non trivial absolutely continuous part then  $e_{p,n}(\rho)$  goes to 0 as  $n^{-1/d}$ .

Given  $\rho$  in  $\mathcal{P}_p$  the *random empirical quantization error* of order  $p$  is given by

$$E_{p,n}(\rho) := W_p(R^{(n)}, \rho),$$

where  $R^{(n)} = n^{-1} \sum_{k=1}^n \delta_{X_k}$  is the empirical measure associated with the i.i.d. random variables  $X_k$  which are distributed according to  $\rho$ . Let us mention here a result stated in [9, Theorem 1]: if  $\rho$  is in  $\mathcal{P}_q$  for some  $q$  large enough then  $\mathbb{E}[E_{p,n}(\rho)^p]^{1/p}$  goes to 0 as  $n^{-1/2p}$  or  $n^{-1/d}$  depending on the values of  $p$  and  $d$  (to be precise, an additional logarithmic term appears at the transition  $p = d/2$ ). The rate  $n^{-1/2p}$  comes from the fluctuations in the law of large numbers and the rate  $n^{-1/d}$  comes from standard quantization as stated above.

Given  $\rho$  in  $\mathcal{P}_p$  the *optimal deterministic empirical quantization error* of order  $p$  is given by

$$\tilde{e}_{p,n}(\rho) := \inf_{\mu^{(n)}} W_p(\mu^{(n)}, \rho),$$

where the infimum is taken with respect to the set of deterministic empirical measures  $\mu^{(n)}$  supported by  $n$  atoms. Up to our knowledge, the rate of convergence of  $\tilde{e}_{p,n}(\rho)$  is known in dimension  $d = 1$  only and reads as follows.

**Theorem II.2** ([17, Theorem 5.20 and Remark 5.21]). *Let  $p \geq 1$  and  $d = 1$ .*

(i) *If  $\rho \in \mathcal{P}_q$  with  $q > p$  then  $\tilde{e}_{p,n}(\rho) = o(n^{1/q-1/p})$ .*

(ii) *If  $\text{Supp}(\rho)$  is bounded then the rate of convergence of  $\tilde{e}_{p,n}(\rho)$  is upper-bounded by  $n^{-1/p}$ . Furthermore, if the support of  $\rho$  is disconnected then the rate  $n^{-1/p}$  is sharp.*

Combining the results of the standard quantization and deterministic empirical quantization we expect that for some generic  $\rho$  with bounded support in dimension  $d \geq 1$ , the rate of  $\tilde{e}_{p,n}(\rho)$  is given by  $\max(n^{-1/d}, n^{-1/p})$  (which is sharp when  $\text{Supp}(\rho)$  is disconnected). This is what is shown in Theorem III.3 below (up to a logarithmic term at the transition  $p = d$ ). Moreover, the generalization of Theorem II.2.(i) to  $d \geq 1$  is obtained in Corollary III.5.

### III Main results

This section begins with a technical lemma which is used to control diameters in our construction of a uniform decomposition of probability measures (which is then given in Theorem III.2).

**Lemma III.1.** *Let  $r \geq 0$ ,  $n \geq 1$  and  $\nu$  be in  $\mathcal{M}(\mathbb{R}^d)$  with support included in  $\mathcal{B}_r$  and total mass  $|\nu| \geq 1/n$ . There exists a subset  $A$  of  $\mathcal{B}_r$  such that  $\nu(A) \geq 1/n$  and  $\text{Diam}(A) \leq 4r(n|\nu|)^{-1/d}$ .*

*Proof.* Consider for any  $r' \geq 0$  the maximal mass over balls of radius  $r'$ , namely

$$m(r') = \sup \{ \nu(B(x, r')) : x \in \mathbb{R}^d \}.$$

We prove by contradiction that

$$m(r \lfloor (n|\nu|)^{1/d} \rfloor^{-1}) \geq 1/n. \quad (1)$$

Assume that the  $\nu$ -mass of any ball of radius equal to  $r \lfloor (n|\nu|)^{1/d} \rfloor^{-1}$  is less than  $1/n$ . Yet there exists a covering of the ball  $[-r, r]^d$  into  $\lfloor (n|\nu|)^{1/d} \rfloor^d$  disjoint smaller balls, each one of radius equal to  $r \lfloor (n|\nu|)^{1/d} \rfloor^{-1}$  (the balls are cubes). This implies

$$|\nu| < \lfloor (n|\nu|)^{1/d} \rfloor^d n^{-1} \leq ((n|\nu|)^{1/d})^d n^{-1} = |\nu|$$

yielding a contradiction.

Hence we have proved that we can find a subset  $A$  such that  $\nu(A) \geq 1/n$  and  $\text{Diam}(A) \leq 2r \lfloor (n|\nu|)^{1/d} \rfloor^{-1}$ . The stated result then follows from

$$r \lfloor (n|\nu|)^{1/d} \rfloor^{-1} \leq 2r(n|\nu|)^{-1/d}$$

(treat separately the two cases  $(n|\nu|)^{1/d} \geq 2$  and  $(n|\nu|)^{1/d} < 2$ ).  $\square$

**Theorem III.2.** *Let  $r \geq 0$  and  $\rho$  be in  $\mathcal{P}(\mathbb{R}^d)$  with support included in  $\mathcal{B}_r$ . For all  $n \geq 1$ , there exists a uniform decomposition  $(\rho_1, \dots, \rho_n)$  of  $\rho$  such that*

$$\forall k = 1, \dots, n, \quad \text{Diam}(\text{Supp}(\rho_k)) \leq 4rk^{-1/d}.$$

The proof is based on an iterative construction: each iteration relies on Lemma III.1.

*Proof.* Applying Lemma III.1 to  $\rho$  gives the existence of a subset  $A_n$  such that  $\rho(A_n) \geq 1/n$  and  $\text{Diam}(A_n) \leq 4rn^{-1/d}$ . Then, we define the measure

$$\rho_n := \frac{n^{-1}}{\rho(A_n)} \rho \mathbf{1}_{A_n}.$$

In particular,  $|\rho_n| = 1/n$  and  $\text{Supp}(\rho_n) \subset A_n$ . Applying Lemma III.1 to  $\tilde{\rho}_n = \rho - \rho_n$  (its total mass is  $(n-1)/n$ ) gives a subset  $A_{n-1}$  such that  $\tilde{\rho}_n(A_{n-1}) \geq 1/n$  and  $\text{Diam}(A_{n-1}) \leq 4r(n-1)^{-1/d}$ . Similarly we define  $\rho_{n-1} := \frac{n^{-1}}{\tilde{\rho}_n(A_{n-1})} \tilde{\rho}_n \mathbf{1}_{A_{n-1}}$  and  $\tilde{\rho}_{n-1} = \tilde{\rho}_n - \rho_{n-1}$ . Finally, applying in turn the iterative step to the (decreasing) sequence of measures  $\tilde{\rho}_n, \tilde{\rho}_{n-1}, \dots$  ends the proof.  $\square$

The decomposition stated above is then used to control the rate of convergence of the optimal deterministic empirical quantization error  $\tilde{e}_{p,n}(\rho)$  by exhibiting a particular empirical measure with controlled approximation error. The bounded case is treated in Theorem III.3, the unbounded case in Corollary III.5 and finally an application to the clustering issue (when  $\rho$  is an empirical measure) is given in Corollary III.6.

**Theorem III.3.** *Let  $r \geq 0$  and  $\rho$  be in  $\mathcal{P}(\mathbb{R}^d)$  with support included in  $\mathcal{B}_r$ . For all  $n \geq 1$ , there exist  $x_1, \dots, x_n$  in  $\mathbb{R}^d$ , with associated empirical measure  $\mu^{(n)} = n^{-1} \sum_{k=1}^n \delta_{x_k}$ , such that for all  $p \geq 1$ ,*

$$W_p(\mu^{(n)}, \rho) \leq 4r f_{p,d}(n),$$

where

(i) if  $p < d$ , then  $f_{p,d}(n) := (\frac{d}{d-p})^{1/p} n^{-1/d}$ ;

(ii) if  $p = d$ , then  $f_{p,d}(n) := (\frac{1+\ln n}{n})^{1/d}$ ;

(iii) if  $p > d$ , then  $f_{p,d}(n) := \zeta(p/d) n^{-1/p}$ , where  $\zeta$  is the Zeta Riemann function.

**Remark III.4.** *The rates show a transition between the rate for the approximation of a density  $n^{-1/d}$  (standard quantization) and the rate for approximation of measures with disconnected support  $n^{-1/p}$  (the simplest example being the sum of two Dirac masses, the interested reader is referred to [17, Remark 5.21.(ii)]). At the transition, our construction gives a rate with an additional logarithmic term. This may be an artefact of our too simple construction : this logarithmic term does not appear in dimension 1 for measures with bounded support - see [17, Theorem 5.20.(ii)]. However let us mention that such additional logarithmic term may appear for unbounded measures as highlighted in [17, Example 5.8].*

*Proof.* Let  $\rho_1, \dots, \rho_n$  and  $A_1, \dots, A_n$  be respectively the measures and the subsets of  $\mathcal{B}_r$  given by the decomposition of Theorem III.2. For each  $k$ , let  $x_k$  denote the center of  $A_k$  (which is well-defined since  $A_k$  is a cube) and let  $\mu^{(n)}$  denote the associated empirical measure. We use the canonical coupling associated with the decomposition of  $\rho$  into the  $\rho_k$ 's to control the Wasserstein distance. Namely,

$$W_p(\mu^{(n)}, \rho)^p \leq \sum_{k=1}^n W_p(n^{-1} \delta_{x_k}, \rho_k)^p \leq n^{-1} \sum_{k=1}^n \text{Diam}(A_k)^p \leq \frac{(4r)^p}{n} \sum_{k=1}^n k^{-p/d}.$$

If  $p > d$  then the sum is bounded by  $\zeta(p/d) < +\infty$  and we obtain (iii). If  $p < d$ , then the sum is bounded by  $\int_0^n t^{-p/d} dt = n^{1-p/d}/(1-p/d)$  which gives (i). If  $p = d$ , then the sum is bounded by  $1 + \ln n$  yielding (ii).  $\square$

**Corollary III.5.** *Let  $q \geq 1$  and  $\rho \in \mathcal{P}_q$ . For all  $n \geq 1$ , there exist  $x_1, \dots, x_n$  in  $\mathbb{R}^d$ , with associated empirical measure  $\mu^{(n)}$ , such that for all  $p < q$ ,*

$$W_p(\mu^{(n)}, \rho) = o(f_{p,d}(n)^{1-p/q})$$

where  $f_{p,d}(n)$  is defined in Theorem III.3.

*Proof.* We use a truncation argument to reduce to the case where  $\rho$  is compactly supported. Let  $r > 0$  be a truncation level to be chosen later and define the measure  $\rho^{(r)}$  by

$$\rho^{(r)}(dx) := \rho(dx) \mathbf{1}_{\mathcal{B}_r}(x) + (1 - \rho(\mathcal{B}_r)) \delta_0(dx).$$

By the canonical coupling, we have

$$W_p(\rho, \rho^{(r)})^p \leq \int_{\|x\|>r} \|x\|^p \rho(dx).$$

Yet,  $\int_{\|x\|>r} \|x\|^p \rho(dx) \leq C_q(r) r^{p-q}$  with  $C_q(r) := \int_{\|x\|>r} \|x\|^q \rho(dx)$  which goes to 0 at  $r \rightarrow +\infty$  by assumption. Without loss of generality one can replace  $C_q(r)$  by some  $C(r)$ , satisfying  $C(r) \geq 1/r$  and  $\lim_{r \rightarrow +\infty} C(r) = 0$ , and write the upper-bound

$$W_p(\rho, \rho^{(r)}) \leq C(r)^{1/p} r^{1-q/p}. \tag{2}$$

By Theorem III.3, for all  $r \geq 0$ , there exist empirical measures  $\mu^{(n,r)}$  such that

$$W_p(\mu^{(n,r)}, \rho^{(r)}) \leq 4r f_{p,d}(n).$$

By the triangular inequality,

$$W_p(\mu^{(n,r)}, \rho) \leq g(r) := C(r)r^{1-q/p} + 4r f_{p,d}(n).$$

To optimize  $g(r)$ , let us choose  $\tilde{r} = \tilde{r}(n) := f_{p,d}(n)^{-p/q}$  since it satisfies  $\tilde{r}^{1-q/p} = \tilde{r} f_{p,d}(n)$  and then consider  $r(n) := C(\tilde{r})\tilde{r}$  to compute

$$g(r(n)) \leq C(C(\tilde{r})\tilde{r})C(\tilde{r})^{1-q/p} f_{p,d}(n)^{1-p/q} + 4C(\tilde{r})f_{p,d}(n)^{1-p/q}.$$

Finally, since  $\lim_{n \rightarrow +\infty} \tilde{r}(n) = +\infty$  and  $\lim_{r \rightarrow +\infty} C(r) = 0$ , we easily end the proof.  $\square$

**Corollary III.6.** *Assume that  $N = cn$  with  $c, n$  in  $\mathbb{N}$ . For any  $x_1, \dots, x_N$  in  $\mathbb{R}^d$ , there exist  $C_1, \dots, C_n$  disjoint subsets of indices of  $\{1, \dots, N\}$  such that*

- they form a balanced clustering of  $x_1, \dots, x_N$ , namely the cardinal  $\text{Card}(C_k) = c$  for all  $k = 1, \dots, n$ ;
- each class is controlled, namely for all  $k = 1, \dots, n$ ,

$$\text{Diam}(x_{(C_k)}) \leq 4rk^{-1/d},$$

where  $x_{(C_k)} := \{x_i, i \in C_k\}$  and  $r = \max_{i=1, \dots, N} |x_i|$ .

In particular, there exist  $\bar{x}_1, \dots, \bar{x}_n$  in  $\mathbb{R}^d$  such that

$$\frac{1}{N} \sum_{i=1}^N |x_i - \bar{x}_{k(i)}| \leq 4r f_{1,d}(n), \quad (3)$$

where  $k(i) \in \{1, \dots, n\}$  is such that  $x_i \in C_{k(i)}$  and  $f_{1,d}$  is given by Theorem III.3.

*Proof.* The proof of the existence of the balanced clustering  $C_1, \dots, C_n$  is based on an iterative application of Lemma III.1 similar to the one developed in the proof of Theorem III.2 and is therefore omitted.

The proof of (3) is similar to the end of the proof of Theorem III.3.  $\square$

## IV Algorithms and simulations

In this section, we present two algorithms based on the ideas developed above. They are stated in the framework of balanced clustering: the goal is to construct  $n$  representative points from a sample of size  $N = cn$ . In particular, the results of the two algorithms satisfy the rates of Corollary III.6.

Assume that a subset  $\mathcal{S}$  of  $[a, b]^d \subset \mathbb{R}^d$  with cardinal  $N = cn$  is given. Of course, it can be a  $N$ -sample of some probability distribution (in which case  $\mathcal{S}$  may be a multiset). The two algorithms (fully described in the appendix below) follow the same structure : a loop with  $n$  iterations. At each iteration,  $c$  points are selected and a new point is constructed to represent them.

The first algorithm, called “naive”, is directly inspired by the main proof of the paper. In particular, the fact that there is at least one cube to select at step 3 is ensured by the proof of Lemma III.1.

The result of the naive algorithm may show some geometrical structure which does not come from the underlying distribution (see Figure 1). This is induced by the partitioning of  $[a, b]^d$  into a relatively small number of cubes. In order to correct this side effect, we come up with a refined version of the algorithm where the partition is made at a finer scale (depending on a parameter  $\delta \in \mathbb{N}^*$ ) once and for all: in comparison with the naive version, the interval  $[a, b]$  is partitioned into  $\delta \lfloor n^{1/d} \rfloor$  small intervals (associated with small cubes of length  $\ell = (b - a)/(\delta \lfloor n^{1/d} \rfloor)$ ) instead of  $\lfloor n^{1/d} \rfloor$  medium intervals. Then, for a given scale integer  $h$ , a cube of length  $h\ell$  which is made of  $h^d$  small cubes is called a *cube of scale  $h$*  (e.g. small cubes are of scale 1). Let us denote by  $N_{\text{cube}}(h)$  the number of cubes of scale  $h$ .

The idea of the algorithm is to start from a small scale  $h$  and repeat the following procedure: 1/ scan all the cubes of scale  $h$  and search for those with more than  $c$  points, 2/ update to a larger scale  $h$ .

The choice of the updating scale  $h$  in the algorithm is made so that there is at least one cube of this new scale which contains at least  $c$  points (the argument is the same as for the naive algorithm). Hence the scale  $h$  is increasing and the algorithm is ensured to end because when  $h = \delta \lfloor n^{1/d} \rfloor$  there is only one cube of scale  $h$  which is  $[a, b]^d$  and contains all the remaining points.

**Acknowledgments** This research was supported by the project Labex MME-DII (ANR11-LBX-0023-01) and mainly conducted during the stay of the author at Université de Cergy-Pontoise.

## Appendix

---

### Algorithm 1 Naive version

---

- **set**  $S = \mathcal{S}$
  - **for**  $k = 1, \dots, n$  **do**
    1. consider the canonical partition of  $[a, b]^d$  into  $\lfloor (n - k + 1)^{1/d} \rfloor^d$  cubes (partition the interval  $[a, b]$  into  $\lfloor n^{1/d} \rfloor$  intervals and consider the associated cubes)
    2. compute the number of points of  $S$  which falls into each cube
    3. select one cube  $C$  with more than  $c$  points
    4. select  $c$  points in  $C$ , named as  $\tilde{C}$ , and set  $\mathbf{centers}(k) = \frac{\max(\tilde{C}) + \min(\tilde{C})}{2}$
    5. **update**  $S \leftarrow S \setminus \tilde{C}$
  - **return centers**
- 

---

### Algorithm 2 Refined scanning version

---

- **set**  $h = \delta$  and  $m = 1$
  - consider the canonical partition of  $[a, b]^d$  into  $(\delta \lfloor n^{1/d} \rfloor)^d$  cubes
  - compute the repartition  $\mathcal{R}$  of the points in  $\mathcal{S}$  into the partition of  $[a, b]^d$
  - **for**  $k = 1, \dots, n$  **do**
    - **while** the  $m^{\text{th}}$  cube has less than  $c$  points **do**
      - \* **if**  $m < N_{\text{cube}}(h)$  **set**  $m = m + 1$
      - \* **else set**  $h = \lceil \frac{\delta \lfloor n^{1/d} \rfloor}{\lfloor (n - k + 1)^{1/d} \rfloor} \rceil$  and  $m = 1$
    - select  $c$  points in the  $m^{\text{th}}$  cube of scale  $h$ , named as  $\tilde{C}$ , and set  $\mathbf{centers}(k) = \frac{\max(\tilde{C}) + \min(\tilde{C})}{2}$
    - **update**  $\mathcal{R}$  by discarding the points in  $\tilde{C}$
  - **return centers**
-

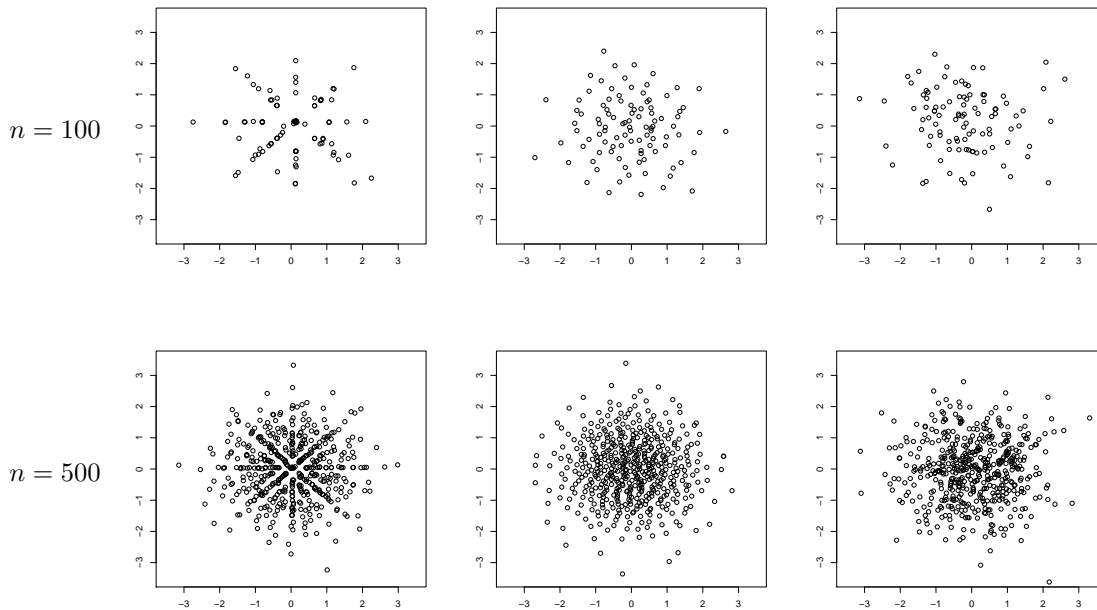


Figure 1: Representative points for the 2D standard Gaussian distribution produced by three methods. From left to right: the naive algorithm, the refined algorithm and standard Monte Carlo. The number  $n$  of representative points changes from one line to the other whereas the size  $N$  of the sample is unchanged and equal to 10000.

## References

- [1] W. R. Bennett. Spectra of quantized signals. *Bell Labs Technical Journal*, 27(3):446–472, 1948.
- [2] E. Boissard, T. Le Gouic, et al. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 50, pages 539–563. Institut Henri Poincaré, 2014.
- [3] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3):541–593, 2007.
- [4] P. Bradley, K. Bennett, and A. Demiriz. Constrained k-means clustering. *Microsoft Research, Redmond*, pages 1–8, 2000.
- [5] Y. Chen, M. Welling, and A. Smola. Super-samples from kernel herding. *arXiv preprint arXiv:1203.3472*, 2012.
- [6] J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost. Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stochastic Processes and their Applications*, 2018.
- [7] J. Dick, F. Y. Kuo, and I. H. Sloan. High-dimensional integration: the quasi-monte carlo way. *Acta Numerica*, 22:133–288, 2013.
- [8] M. Faundez-Zanuy and J. M. Pascual-Gaspar. Efficient on-line signature recognition based on multi-section vector quantization. *Pattern Analysis and Applications*, 14(1):37–45, 2011.
- [9] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, pages 1–32, 2014.
- [10] M. B. Giles, M. Hefter, L. Mayer, and K. Ritter. Random bit quadrature and approximation of distributions on hilbert spaces. *Foundations of Computational Mathematics*, pages 1–34, 2018.



- [11] S. Graf and H. Luschgy. *Foundations of quantization for probability distributions*. Springer, 2007.
- [12] B. Kloeckner. Approximation by finitely supported measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 18(2):343–359, 2012.
- [13] S. Mak and V. R. Joseph. Support points. *arXiv preprint arXiv:1609.01811*, 2016.
- [14] M. I. Malinen and P. Fränti. Balanced k-means for clustering. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 32–41. Springer, 2014.
- [15] G. Pagès. Introduction to vector quantization and its applications for numerics. *ESAIM: Proceedings and Surveys*, 48:29–79, 2015.
- [16] G. Pages and B. Wilbertz. Optimal Delaunay et Voronoi quantization methods for pricing American options. In *Numerical methods in Finance. Workshop Numerical methods in Finance, 2010.*, pages 171–217. Springer, 2010.
- [17] C. Xu and A. Berger. Best finite constrained approximations of one-dimensional probabilities. *arXiv preprint arXiv:1704.07871*, 2017.
- [18] P. Zador. Development and evaluation of procedures for quantizing multivariate distributions. Technical report, STANFORD UNIV CALIF, 1963.
- [19] P. Zador. Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Transactions on Information Theory*, 28(2):139–149, 1982.