J. Appl. Prob. **55,** 1037–1045 (2018)
*doi:*10.1017/jpr.2018.69
© *Applied Probability Trust* 2018

UNIFORM DECOMPOSITION OF PROBABILITY MEASURES: QUANTIZATION, CLUSTERING AND RATE OF CONVERGENCE

JULIEN CHEVALLIER,* *Université Grenoble Alpes*

Abstract

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

Keywords: Uniform approximation; Wasserstein distance; rate of convergence; quantization; clustering

2010 Mathematics Subject Classification: Primary 60E15
Secondary 62E17; 60B10; 60F99

1. Introduction

Finding a good finite decomposition of a given probability measure ρ 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 applications has widened since then; for example, pattern recognition [8], numerical analysis [15], and 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 are independent and identically distributed (i.i.d.) random variables distributed according to ρ and the main results are concerned with the 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, we use the term *deterministic empirical quantization*. This type 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 was decomposed into two terms: a variance/ fluctuation term (the one appearing in mean-field limits without spatial covariates)

Received 28 November 2017; revision received 12 November 2018.
* Postal address: Université Grenoble Alpes, CNRS, LJK, 38000 Grenoble, France.
Email address: julien.chevallier1@univ-grenoble-alpes.fr

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press

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 $d = 1$ was addressed extensively in [17] (the study relies heavily on the connection between Wasserstein distances and the quantile function which is specific to $d = 1$). Our aim is to generalize some of the results stated in [17] to the general case $d \geq 1$ (note that the case of a Hilbert space with $p = 2$ was treated in [10]). Our main result yields sharp bounds on the rate of convergence of $\tilde{e}_{p,n}(\rho) := \inf W_p(\mu^{(n)}, \rho) \to 0$, where the infimum 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 exhibits a transition: it is either the same as for standard quantization (when the Lebesgue measure is not easy to approximate) or strictly worse (when disconnected measures are not easy 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], the authors of some recent papers have dealt with any continuous probability distribution by providing a representative point set which minimizes some prescribed energy function [5], [13].

When the measure ρ 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 can be found in Section 2 together with a list of previous results from the literature. Then, Section 3 contains an elementary uniform decomposition of probability measures (Theorem 2) which we use to obtain upper bounds on deterministic empirical quantization rates (Theorem 3) and balanced clustering rates (Corollary 2).

2. 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 $\mathrm{diam}(A) := \sup_{x,y\in A} \|x - y\|$. The space of all Borel measures (respectively probability measures) on $\mathbb{R}^d$ is denoted by $\mathcal{M}(\mathbb{R}^d)$ (respectively $\mathcal{P}(\mathbb{R}^d)$). For $\nu \in \mathcal{M}(\mathbb{R}^d)$, $\mathrm{supp}(\nu)$ and $|\nu| := \nu(\mathbb{R}^d)$ denote the respective support and mass of the measure ν. For a collection of n positions $x_1, \ldots, x_n \in \mathbb{R}^d$, 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 ρ such that $\int ||x||^p \rho(\mathrm{d}x) < +\infty$ is denoted by $\mathcal{P}_p$. Then the Wasserstein distance of order p is denoted by W_p and defined by, for all ρ and μ in $\mathcal{P}_p$,

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

where the infimum is taken with respect to all couplings π of the two measures ρ and μ.

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

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press

2.1. 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 ρ 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 ρ 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 ρ. Let us mention here a result stated in [9, Theorem 1]: if ρ 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 ρ 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. To the best of the author's knowledge, the rate of convergence of $\tilde{e}_{p,n}(\rho)$ is known in dimension $d = 1$ only. We state the result in the next theorem.

Theorem 1. (Bradley *et al.* [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* $\mathrm{supp}(\rho)$ *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 ρ 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 ρ 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 $\mathrm{supp}(\rho)$ is disconnected). This is the subject of Theorem 3 below (up to a logarithmic term at the transition $p = d$). Moreover, we obtain the generalization of Theorem 1(i) to $d \geq 1$ in Corollary 1.

3. Main results

We begin this section 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 2).

Lemma 1. *Let $r \geq 0, n \geq 1$, and ν 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* $\mathrm{diam}(A) \leq 4r(n|\nu|)^{-1/d}$.

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press

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

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

Assume that the ν-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 there exists a subset A such that $\nu(A) \geq 1/n$ and $\operatorname{diam}(A) \leq 2r\lfloor(n|\nu|)^{1/d}\rfloor^{-1}$. 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 $(n|\nu|)^{1/d} < 2$). □

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

$$\operatorname{diam}(\operatorname{supp}(\rho_k)) \leq 4rk^{-1/d} \quad \textit{for all } k = 1, \ldots, n.$$

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

Proof of Theorem 2. Applying Lemma 1 to ρ yields the existence of a subset A_n such that $\rho(A_n) \geq 1/n$ and $\operatorname{diam}(A_n) \leq 4rn^{-1/d}$. 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 $\operatorname{supp}(\rho_n) \subset A_n$. Applying Lemma 1 to $\tilde{\rho}_n = \rho - \rho_n$ (its total mass is $(n-1)/n$) yields a subset A_{n-1} such that $\tilde{\rho}_n(A_{n-1}) \geq 1/n$ and $\operatorname{diam}(A_{n-1}) \leq 4r(n-1)^{-1/d}$. Similarly, we define $\rho_{n-1} := (n^{-1}/\tilde{\rho}(A_{n-1}))\tilde{\rho}_n\mathbf{1}_{A_{n-1}}$ and $\tilde{\rho}_{n-1} = \tilde{\rho}_n - \rho_{n-1}$. Finally, applying, in turn, the iterative step to the (decreasing) sequence of measures $\tilde{\rho}_n, \tilde{\rho}_{n-1}, \ldots$ completes the proof. □

We then use the decomposition stated above 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. We treat the bounded case in Theorem 3, the unbounded case in Corollary 1 and, finally, an application to the clustering issue (when ρ is an empirical measure) can be found in Corollary 2.

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

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

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press

where

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

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

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

Remark 1. We see that the rates exhibit 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, we refer the interested reader to [17, Remark 5.21(ii)]). At the transition, our construction yields a rate with an additional logarithmic term. This may be an artefact of our simple construction: this logarithmic term does not appear in dimension 1 for measures with bounded support; see [17, Theorem 5.20(ii)]. However, we mention that such an additional logarithmic term may appear for unbounded measures as highlighted in [17, Example 5.8].

Proof of Theorem 3. Let $\rho_1, \dots, \rho_n$ and $A_1, \dots, A_n$ be the respective measures and subsets of $\mathcal{B}_r$ given by the decomposition of Theorem 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 ρ into the ρ_k to control the Wasserstein distance. Namely,

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

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 $\int_0^n t^{-p/d}\,\mathrm{d}t = n^{1-p/d}/(1-p/d)$ which yields (i). If $p = d$ then the sum is bounded by $1 + \ln n$ yielding (ii). □

Corollary 1. *Let* $q \ge 1$ *and* $\rho \in \mathcal{P}_q$. *For all* $n \ge 1$, *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 3.*

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

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

By the canonical coupling, we have

$$W_p(\rho, \rho^{(r)})^p \le \int_{\|x\|>r} \|x\|^p \rho(\mathrm{d}x).$$

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

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

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press

From Theorem 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 $g(r)$, we 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 $\lim_{n\to+\infty} \tilde{r}(n) = +\infty$ and $\lim_{r\to+\infty} C(r) = 0$, the proof follows. □

Corollary 2. *Assume that $N = cn$ with $c, n \in \mathbb{N}$. For any $x_1, \ldots, x_N \in \mathbb{R}^d$, there exist $C_1, \ldots, C_n$ disjoint subsets of indices of $\{1, \ldots, N\}$ such that:*

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

$$\mathrm{diam}(x_{(C_k)}) \leq 4r k^{-1/d},$$

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

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

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

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

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

The proof of (1) is similar to the final part of the proof of Theorem 3. □

4. 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 2.

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 an N-sample of some probability distribution (in which case $\mathcal{S}$ may be a multiset). The two algorithms 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 the third step is ensured by the proof of Lemma 1.

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press

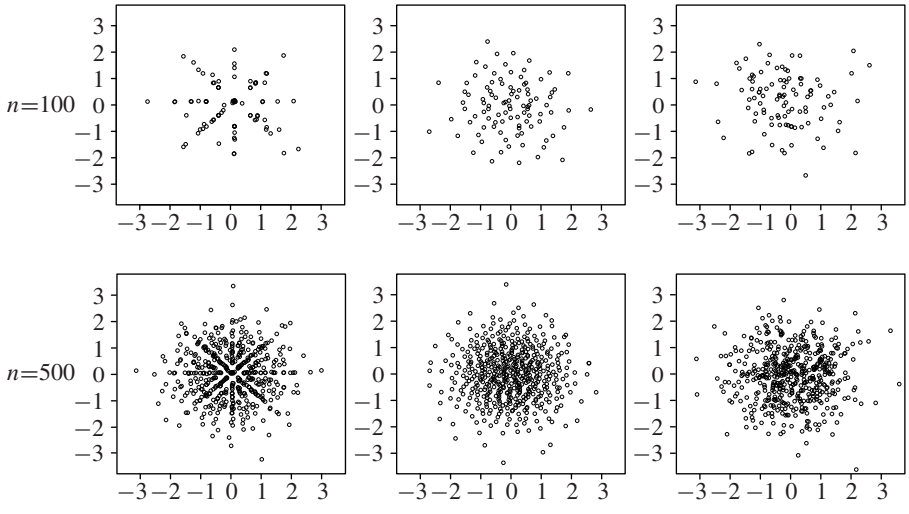


FIGURE 1: Representative points for the two-dimensional standard Gaussian distribution produced by three methods. *Left to right:* the naive algorithm, the refined algorithm, and standard Monte Carlo. The number n of representative points changes from one line to the other whereas the size N of the sample is unchanged and equal to 10,000.

Algorithm 1. (*Naive version.*) The algorithm proceeds as follows.

- **set** $S = \mathcal{S}$
- **for** $k = 1, \ldots, n$ **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 fall 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 $\texttt{centers}(k) = \frac{1}{2}(\max(\tilde{C}) + \min(\tilde{C}))$
  5. **update** $S \leftarrow S \backslash \tilde{C}$
- **return** `centers`

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 develop 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). We denote by $N_{\text{cube}}(h)$ the number of cubes of scale h.

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press

The idea of the algorithm is to start from a small scale h and repeat the following procedure:

- scan all the cubes of scale h and search for those with more than c points;
- update to a larger scale h.

Algorithm 2. (*Refined scanning version.*) The algorithm proceeds as follows.

- **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** $k = 1, \ldots, n$ **do**
  - **while** the mth cube has less than c points **do**
    - **if** $m < N_{\text{cube}}(h)$ **set** $m = m + 1$
    - **else set** $h = \lceil \delta \lfloor n^{1/d} \rfloor / \lfloor (n - k + 1)^{1/d} \rfloor \rceil$ and $m = 1$
  - select c points in the mth cube of scale h, named as $\tilde{C}$, and set $\texttt{centers}(k) = \frac{1}{2}(\max(\tilde{C}) + \min(\tilde{C}))$
  - **update** $\mathcal{R}$ by discarding the points in $\tilde{C}$
- **return** `centers`

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 assured 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.

Acknowledgement

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

References

[1] Bennett, W. R. (1948). Spectra of quantized signals. *Bell Systems Tech. J.* **27,** 446–472.
[2] Boissard, E. and Le Gouic, T. (2014). On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. H. Poincaré Prob. Statist.* **50,** 539–563.
[3] Bolley, F., Guillin, A. and Villani, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. *Prob. Theory Relat. Fields* **137,** 541–593.
[4] Bradley, P., Bennett, K. and Demiriz, A. (2000). Constrained k-means clustering. Tech. Rep. MSR-2000-65, Microsoft Research.
[5] Chen, Y., Welling, M. and Smola, A. (2012). Super-samples from kernel herding. Preprint. Available at http://arxiv.org/abs/1203.3472.
[6] Chevallier, J., Duarte, A., Löcherbach, E. and Ost, G. (2019). Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stoch. Process. Appl.* **124,** 1–27.
[7] Dick, J., Kuo, F. Y. and Sloan, I. H. (2013). High-dimensional integration: the quasi-Monte Carlo way. *Acta Numer.* **22,** 133–288.
[8] Faundez-Zanuy, M. and Pascual-Gaspar, J. M. (2011). Efficient on-line signature recognition based on multi-section vector quantization. *Pattern Anal. Appl.* **14,** 37–45.

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press

[9] Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Prob. Theory Relat. Fields* **162,** 707–734.
[10] Giles, M. B., Hefter, M., Mayer, L. and Ritter, K. (2018). Random bit quadrature and approximation of distributions on Hilbert spaces. *Found. Comput. Math.* Available at https://doi.org/10.1007/s10208-018-9382-3.
[11] Graf, S. and Luschgy, H. (2000). *Foundations of Quantization for Probability Distributions*. Springer, Berlin.
[12] Kloeckner, B. (2012). Approximation by finitely supported measures. *ESAIM Control Optim. Calc. Var.* **18,** 343–359.
[13] Mak, S. and Joseph, V. R. (2018). Support points. *Ann. Statist.* **46,** 2562–2592.
[14] Malinen, M. I. and Fränti, P. (2014). Balanced K-means for clustering. In *Structural, Syntactic, and Pattern Recognition* (Lecture Notes Comput. Sci. **8621**), Springer, Berlin, pp. 32–41.
[15] Pagès, G. (2015). Introduction to vector quantization and its applications for numerics. *ESAIM Proc. Surveys* **48,** 29–79.
[16] Pagès, G. and Wilbertz, B. (2012). Optimal Delaunay and Voronoi quantization methods for pricing American style options. In *Numerical methods in Finance* (Springer Proc. Math. **12**), Springer, Berlin, pp. 171–217.
[17] Xu, C. and Berger, A. (2017). Best finite constrained approximations of one-dimensional probabilities. Preprint. Available at http://arxiv.org/abs/1704.07871.
[18] Zador, P. (1963). Development and evaluation of procedures for quantizing multivariate distributions. Tech. Rep. Stanford University.
[19] Zador, P. (1982). Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory* **28,** 139–149.

https://doi.org/10.1017/jpr.2018.69 Published online by Cambridge University Press