Voronoï tesselation, space quantization algorithms and numerical integration

Gilles PAGÈS *

Abstract

We propose a new method to integrate smooth functions defined on a convex compact set $C \subset \mathbb{R}^d$ using a n-uplet $x^* = (x_1^*, \ldots, x_n^*) \in C^n$ and its Voroni tesselation. x^* is obtained as the minima of a potential function defined on C^n and generalizing the potential of the quantization algorithm (Kohonen with 0 neigbour). We numerically compute x^* and the weights of its tesselation using a stochastic gradient. When specified to the quadratic case, our work provides new results on the Kohonen algorithm with 0 neighbour. Some numerical tests display the asymptotic shape of the equilibrium of the 0-neighbour Kohonen algorithm as the number of units n goes to infinity.

1 Introduction, definitions and notations

High dimensional numerical integration is a major problem of Computational Mathematics and Physics (Financial ingeneering, Neutronics, Boltzmann equation, etc). Usual implemented methods rely on as various fields as Numerical Analysis, Linear Algebra, Number or Ergodic Theory. However, mainly because of the Monte Carlo method, this topic is strongly related with Numerical Probability as well. The aim of this work is to provide a new method for numerical integration derived from a famous stochastic quantization method, the so-called Kohonen algorithm (xith 0 neighbour).

Roughly speaking, two main approaches are actually challenging to integrate a real valued function defined on a high dimensional space, usually $[0, 1]^d$ or, in a more general setting, a (convex) compact subset $C \subset \mathbb{R}^d$.

The first one, relying either on the Strong Law of Large Numbers for (pseudo-)random numbers or on (deterministic) uniformly distributed sequences (cf.[7],[9]), consists in making up infinite $[0, 1]^d$-valued sequences $\xi = (\xi_n)_{n \geq 1}$ having the

---

*Samos, Univ. Paris I, 90, rue de Tolbiac, F-75634 Paris Cedex 13 & Labo. Probabilités et Applications, Univ. P. & M. Curie, Tour 56, F-75252 Paris Cedex 05.

property that

$$\frac{f(\xi_1)+\ldots+f(\xi_n)}{n} \longrightarrow \int_{[0,1]^d} f(u^1,\ldots,u^d)du^1\ldots du^d \text{ as } n \to +\infty$$

When ξ is a sequence of (pseudo-)random numbers – usually obtained on a computer by a *random* or *rand* instruction – the method is known as the Monte Carlo method and the mean asymptotic error goes to 0 as $\frac{1}{\sqrt{n}}$. When ξ is a deterministic uniformly distributed sequence, the algorithm is called a quasi-Monte Carlo method. Most often such deterministic sequences are obtained as some orbits $(T^k(x))_{k\geq 0}$ of an ergodic transform T on $[0,1]^d$. An ergodic transform (on $[0,1]^d$) $T:[0,1]^d \to [0,1]^d$ is a mapping satisfying :

(a) $\forall f$ continuous on $[0,1]^d$, $\int_{[0,1]^d} f(T(x))dx = \int_{[0,1]^d} f(x)dx$

(b) $\forall A$ (Borel set) in $[0,1]^d$, $T(A)=A \implies A=\emptyset$ or $[0,1]^d$ *a.s.*

The most famous T's are (cf.[4]) :

- the rotations of the multidimensional torus $[0,1[^d : T(x)=\{x + \alpha\}$ where $\alpha = (\alpha_1,\ldots,\alpha_d)$ makes up with 1 a family of $d+1$ linearly independant vectors in $\mathbb{R}$ viewed as vector space on $\mathbb{Q}$ ($\{x\}$ denotes the fractional part of x).

- the *p-adic* rotations defined by $T(x) = (x_i \oplus_{p_i} \alpha_i)_{1\leq i\leq d}$ where the p_i's are successive prime numbers, the α_i's are some *p-adic* rationals and $\oplus_{p_i}$ denotes the addition *from the left to the right* with carrying over of regular p_i-adic expansions.

For some of these sequences, called sequences with low discrepancy, the integration error vanishes at a $O(\frac{\ln^d(n)}{n})$ rate when f has finite variation. Unfortunately, in dimension $d\geq 2$, the notion of *finite variation* is rather technical and uneasy to handle. However, we can say that, roughly speaking, such functions are very smooth in the sense that they have d order partial derivatives on $[0,1]^d$. On the other hand, when dealing with some functions only satisfying a usual Lipschitz condition, a far less interesting rate $O((\frac{\ln^d(n)}{n})^{\frac{1}{d+1}})$ holds whenever $d\geq 2$.

When the function f is naturally defined on $[0,1]^d$, the power of both quasi- and true Monte Carlo methods lies in their simple implementation on a computer once the sequence ξ is specified.

However they also have some noticeable drawback for the user. Thus, lengthy enough sequences of true random numbers are not available and the statistical properties of the finite sequences of pseudo-random numbers are always subject to discussion. On the other hand, deterministic uniformly distributed sequence suffers from the lack of accurate numerical bound of the integration error : the theoretical bounds, if asymptotically optimal, are numerically meaningless

with respect to the current implemented the number n of trials. Furthermore, such methods do not fully take into account the smoothness of the integrated functions : in fact, the classical Monte Carlo method is mostly famous for its efficiency when dealing with with erratic functions !

The second class of methods, usually less related with Probability Theory, can be defined in a slightly more general setting : let C be a (convex) compact set in $\mathbb{R}^d$ and $\mu(d\omega^1 \ldots d\omega^d)$ a (finite) non negative measure on C. One usually approximates $\displaystyle\int_{C^*} f(\omega^1, \ldots, \omega^d)\mu(d\omega^1 \ldots d\omega^d)$ by $\displaystyle\sum_{1 \leq i \leq n} \pi(x_i)f(x_i)$ where $x = (x_1, \ldots, x_n)^*$. Usually, there exists, for a given n, an optimal n-uplet x^* that minimizes the error integration bound $E_n(x^*)$ among all n-uplets in C^n for a whole class of functions with a given smoothness. The problem is then to compute this optimal x^* – its related weights $(\pi(x_i^*))_{1 \leq i \leq n}$ if necessary – and an efficient numerical bound for the integration error.

The most famous example of such points is the family of Gauss points on $[0,1]$ that integrate polynomials and analytic functions (one variable) with a remarkable accuracy (cf. [11]).

In the forthcoming sections, we provide a new method to compute high dimensional integrals of functions defined on a convex set that takes into account the regularity of the function in the "Hölder" scale. Let us recall that $f : C \to \mathbb{R}$ is α-hölder, $0 < \alpha < 1$, iff

$$\forall x, y \in \mathbb{R}^d \quad |f(x) - f(y)| < [f]_\alpha |x - y|$$

where $|x| := (\sum_i x_i^2)^{\frac{1}{2}}$ denotes the Euclidian norm. $[f]_\alpha$ will denote the hölder coefficient (or *ratio*) of f. 0-hölder will simply mean *continuous* and we set $[f]_0 := 2|f|_\infty$.

Let $p \in \mathbb{N}$ and $\alpha \in [p, p+1[$. The set of α-smooth functions on C is defined by $C^\alpha(C) := \left\{ f \in C(C) \cap C^p(\overset{\circ}{C}) \text{ such that } f^{(p)} \text{ is } (\alpha - p)\text{-hölder} \right\}$.

The method is based on the study, for every $\alpha \in]0, 2]$, of a potential function defined for every $x \in C$, using the Voronoï tesselation of x. The minimum of this potential function will provide the best n-uplet x^*, the related weights being the hyper-volume of the elements of the tesselation. We will show that all these parameters can be obtained using a stochastic gradient.

Whenever $\alpha = 2$, this stochastic gradient is but the Kohonen algorithm with 0 neighbour. Of course this apparently side remark was actually at the very origin of the initial idea. subsequently, all the below results, read with $\alpha = 2$, actually provide some converging property of the multi-dimensional Kohonen algorithm with 0 neighbour.

We consider

- A convex, compact non empty set $C \subset \mathbb{R}^d$. Most often C will be taken

as the convex hull of the support $supp(\mu)$ of a probability measure μ (see next item). C is the set on which the integrated functions wil be defined.

- A probability measure μ on C that can be practically *simulated* on a computer. Furthermore we will assume that μ is *strongly diffuse* ($\mu(H)=0$ for every hyperplan H). For the sake of simplicity one may simply assume that

$$\mu(d\omega)=1_C(\omega^1,\ldots,\omega^d)g(\omega^1,\ldots,\omega^d)d\omega^1\ldots d\omega^d.$$

- A n-uplet $x=(x_1,\ldots,x_n)\in C^n$ and $C_i(x)$, $1\leq i\leq n$, its Voronoï tesselation defined by

$$\left\{\begin{array}{ll} C_i(x)= & \{u\in C\,/\,x_j\neq x_i\Rightarrow|x_i-u|<|x_j-u|\} \quad\text{if } i:=\min\{k\,/\,x_k=x_i\}\\ C_i(x)= & \emptyset \text{ if there is some } j<i \text{ with } x_j=x_i. \end{array}\right.$$

Finally, the question is : "How to optimally approximate $\int_C f d\mu$ using n values $f(x_1),\ldots,f(x_n)$ of f, but without knowing the possible derivatives $f^{(k)}(x_i)$ of f ?"

Of course, the answer will be in our setting

$$\int_C f(x)\mu(dx)\approx\sum_{i=1}^n\mu(C_i(x))f(x_i) \tag{1}$$

the problem being to yield an accurate error bound and to find the best n-uplet x^*. To this end, we define the μ-magnitude with $\alpha>0$ order at $x\in C^n$ by

$$E_n^{\alpha,\mu}(x)=\int_C\min_{1\leq i\leq n}|x_i-\omega|^\alpha\mu(d\omega).$$

These moduli are obviously continuous as functions of $x\in C^n$ and so reach their minimum. $n\to\min_{x\in C^n}E_n^{\alpha,\mu}(x)$ is decreasing and go to 0 as n goes to infinity (see [10] for further details).

When $\alpha=2$, $E_n^{2,\mu}$ is the potential function of the Kohonen algorithm with 0 neighbour and stimuli distribution μ.

2 Numerical integration: the main result

Theorem 1 *Let $f\in C^\alpha(C)$, $\alpha\in[0,1]$ (resp. $[1,2]$), $x^*\in argminE_n^{\alpha,\mu}$ (resp. $argminE_n^{2,\mu}$), then*

$$\left|\sum_{i=1}^n\mu(C_i(x^*))f(x_i^*)-\int_C f d\mu\right|\leq[f]_\alpha E_n^{\alpha,\mu}(x^*) \quad\text{(resp. } [f']_{\alpha-1}E_n^{\alpha,\mu}(x^*)). \tag{2}$$

When $\alpha \in [0, 1]$ (think to Lipschitz functions when $\alpha = 1$) the result is rather natural and follows from a decomposition of C with respect to the Voronï tesselation. On the opposite, when $\alpha \in [1, 2]$ (think about C^2 functions) some additional terms involving the derivative f' at points x_i^* could reasonnably have been expected. Inequality (2) follows from a technical result that says that x^* does not lie on the boundary of C^n and has no clustered components. At such points,

$E_n^{\alpha, \mu}(x)$ admits a derivative given by $\nabla E_n^{\alpha, \mu}(x) = \alpha \left(\int_{C_i(x)} |x_i - \omega|^{\alpha - 1} \dfrac{x_i - \omega}{|x_i - \omega|} \mu(d\omega) \right)_{1 \le i \le n}$.

Hence $\nabla E_n^{\alpha, \mu}(x^*) = 0$ since x^* is then a local minimum of $E_n^{\alpha, \mu}$ as well. Some extensions (when μ is not strongly diffuse) the differentiability of $E_n^{\alpha, \mu}$ bounded by some counter-examples are studied in full details in [10].

The rather obvious corollary of the above theorem is that if $x^{(n), *}$ denotes an element of $\arg\min E_n^{\alpha, \mu}$ then for every continuous function f,

$$\sum_{i=1}^{n} \mu(C_i(x^{(n), *})) f(x_i^{(n), *}) \longrightarrow \int_C f(x) \mu(dx) \quad \text{as} \quad n \to +\infty$$

which means in less mathematical terms that for large enough n, any – if not unique – element of $\arg\min E_n^{\alpha, \mu}$ provides an excellent skeleton of the probability distribution μ.

The remaining problem, but crucial for any actual application, is then to numerically compute, for a given n, the optimal n-uplets x^* that achieves $\min_{C^n} E_n^{\alpha, \mu}$, the related weights $(\mu(C_i(x^*))_{1 \le i \le n}$ and $\min_{C^n} E_n^{\alpha, \mu}$ itself. Furthermore, in order to know if the method is performing for large values of n, we need some theoretical bounds on $\min_{C^n} E_n^{\alpha, \mu}$ when C has a usual shape, say an hypercube $[0, 1]^d$, and μ is the d-dimensional Lebesgue measure $d\omega^1 \ldots d\omega^d$ on C.

3 Application to $C = [0, 1]^d$ and $\mu = dx$-case

Actually, we do have a theoretical upper-bound for the rate of convergence to 0 of $\min_{C^n} E_n^\alpha$ where E_n^α denotes the magnitude with respect to the (Lebesgue) measure $d\omega^1 \ldots d\omega^d$. We guess it is optimal without knowing how to prove it in dimensions $d \ge 2$. In fact, we can compute the true value of $\min_{[0,1]^n} E_n^\alpha$ when $d = 1$ (on the unit interval), and then we use the following result.

Proposition 2 *Let μ be a probability measure on $[0, 1]^d$ and μ_k its kth marginal on $[0, 1]$ (i.e. such that $\int_{[0,1]^d} f(u^k) \mu(du^1 \ldots du^d) = \int_{[0,1]} f(u^k) \mu_k(du^k)$). Then*

• *Majorization : for every decomposition $n = n_1 \ldots n_d$ and every adapted "grid"*
$x := x^1 \otimes \ldots \otimes x^d := \left((x_{i_1}^1, \ldots, x_{i_d}^d) \right)_{1 \le i_k \le n_k, 1 \le k \le d} \in ([0, 1]^d)^n$

$$E_n^{2,\mu}(x) = \sum_{k=1}^{d} E_{n_k}^{2,\mu_k}\left((x_i^k)_{1 \le i \le n_k}\right). \tag{3}$$

- *Minorization : For every $\alpha \in [0,2]$ and $x \in ([0,1]^d)^n$, il vient*

$$E_n^{\alpha,\mu}(x) \ge d^{\frac{\alpha}{2}-1} \sum_{k=1}^{d} E_n^{\alpha,\mu_k}\left((x_i^k)_{1 \le i \le n}\right). \tag{4}$$

It follows that, in dimension d, if $\alpha \in [0,2]$, for every $n \in \mathbb{N}^*$,

$$\frac{d^{\frac{\alpha}{2}-1}}{(\alpha+1)2^\alpha}\frac{1}{n^\alpha} \le \min_{x \in ([0,1]^d)^n} E_n^\alpha(x) \le \left(\frac{d}{12}\right)^{\frac{\alpha}{2}}\frac{1}{[n]^{\frac{\alpha}{d}}} = O\left(\frac{1}{n^{\frac{\alpha}{d}}}\right). \tag{5}$$

The inequality on the left holds as an equality whenever $d = 1$.

4 How to reach the minimal tesselation : the stochastic gradient

$E_n^{\alpha,\mu}$ being differentiable and its gradient $\nabla E_n^{\alpha,\mu}$ having an integral representation, this strongly suggests to implement a stochastic gradient to get x^*. On a numerical point of view notice that this requires to know how to simulate on a computer the distribution μ. The algorithm displays as

$$X^{t+1} = X^t - \varepsilon_{t+1}H_\alpha(X^t, \omega^{t+1}), \quad X^0 = x \in \overset{\bullet}{C}{}^n$$

where $H_\alpha(x,\omega) = \left(\dfrac{x_i - \omega}{|x_i - \omega|^{2-\alpha}}1_{C_i(x)}(\omega)\right)_{1 \le i \le n}$ and $(\varepsilon_t)_{t \le 1}$ is a sequence of

non negatives steps satisfying the usual "decreasing step" assumption :

$$\sum_t \varepsilon_t = +\infty \quad \text{and} \quad \sum_t \varepsilon_t^2 < +\infty.$$

When $\alpha = 2$ and $\varepsilon_t \in [0,1]$, is but the Kohonen algorithm with 0 neighbour.

Applying the O.D.E. method (see Kushner & Clark, cf.[8]), one obtains the "conditional" *a.s.* convergence of the C^n-valued vectors X^t to x^* as $t \to +\infty$. Furthermore, the related weights $(\mu(C_i(x^*)))_{1 \le i \le n}$ and $E_n^{\alpha,\mu}(x^*)$ can also be reached by the mean of this stochastic algorithm which is crucial to be able to process actual numerical integration with the n-uplet x^*.

Theorem 3 *Let $x^* \in argminE_n^{\alpha,\mu}$ et $\alpha \in [1,2]$.*
(i) $X^t \xrightarrow{a.s.} x^*$ on $\{X^t \in K$ infinitly often$\}$ where K is a compact set contained in the attracting area Γ_{x^*} of x^* for the autonomous differential system $\dot{x} = -\nabla E_n^{\alpha,\mu}(x)$.

226

(ii) Let, for every $i \in \{1,\dots,n\}$, $\mu_i^t := \dfrac{\operatorname{card}\{1 \le s \le t / \omega^s \in C_i(X^{s-1})\}}{t}$ the frequency of the "stimuli" ω^s falling in the ith "facet" of the moving tesselation related to X^{s-1}. Then

$$
\begin{cases}
\mu_i^t \xrightarrow[t]{\textbf{P}} \mu(C_i(x^*)) & \text{on } A_\infty := \{X^t \longrightarrow x^* \text{ when } t \to +\infty\}\, 1 \le i \le n, \\
\dfrac{1}{t}\sum_{s=1}^{t}\min_i |\omega^s - X_i^{s-1}|^\alpha \xrightarrow{p.s.} E_n^{\alpha,\mu}(x^*) & \text{on } A_\infty.
\end{cases}
$$

Some additional remarks on the (non-)uniqueness of x^*, the existence of non global local minima and the problems it may cause can be found in [10].

Some numerical tests were processed in a 2-dimensional setting with $\alpha = 2$, $C = [0,1]^2$ and the usual (Lebesgue) measure $d\omega^1 d\omega^2$. Our purpose was to compute $x^{(n),*}$ for $n \in \{1,\dots,100\}$ and the related minimum $E_n^2 = \min_{C^n} E_n^2$ by the above stochastic gradient descent. Recall that, in this setting, this is the Kohonen algorithm with 0 neighbour! As $\alpha = d = 2$, we know that $\min_{C^n} E_n^2 = O(\frac{1}{n})$, so we computed $n \times \min_{C^n} E_n^2$ as well to get the optimal constant. For every n, we first sampled in $[0,1]^2$ 10^6 uniformly distributed random 2-dim vectors, the step ε_t being (essentially) constant (its starting value ε_1 being chosen as a decreasing function of the number n of points) x. It was the "searching phase. Then, we processed a converging phase by sampling 2.10^6 additional independant uniform vectors with a slowly decreasing step $\varepsilon_t = \varepsilon_{10^6} \frac{250}{250+t-10^6}$.

We verified on these simulations that the optimal "uniform" quantization of the unit square was performed as n increases by the regular hexagonal pavement, except for the (predictable) edge effects near the boundary of the square. As far as $\min_{([0,1]^2)^n} E_n^2$ is concerned we obtained the following results :

n	1	2	3	4	5	6	7	8	9
$\min E_n^2$	0,1667	0,1042	0,0662	0,0417	0,0353	0,0278	0,0253	0,0214	0,0185
$n \min E_n^2$	0,1667	0,2084	0,1986	0,1666	0,1765	0,1667	0,1771	0,1712	0,1667

10	11	12	13	14	15	16	...	25	...
0,0170	0,0156	0,0142	0,01317	0,01207	0,01125	0,01034	...	0,006626	...
0,1700	0,1716	0,1704	0,1712	0,1690	0,16875	0,16632	...	0,165650	...

29	...	36	...	49	...	64	...	81	...	100
0,005758	...	0,0045995	...	0,003365	...	0,002571	...	0,002026	...	0,001641
0,166982	...	0,165582	...	0,164885	...	0,1645444	...	0,164106	...	0,163900

Actually $n \min E_n^2$ is very slowly converging to $\frac{5}{18\sqrt{3}} \approx 0,1604\dots$ as n goes to infinity. $\frac{5}{18\sqrt{3}}$ is the value of $\int_H |x - \omega|^2 d\omega^1 d\omega^2$, where H dnotes the regular unit hexagon with center x.

As a first conclusion, let say that, on a strictly theoretical point of view, if $\mu(d\omega) = \lambda(d\omega)$, the Voronoï method is asymptotically faster than, for example, the Monte Carlo method as far as

$$\frac{\alpha}{d} \geq \frac{1}{2} \quad \text{i.e.} \quad d \leq 2\alpha, \ \alpha \in [0, 2] \quad \text{(i.e. } d \leq 4 \text{ at the best).}$$

However, these bounds must be considered very carefully: the above numerical tests pointed out that the efficiency of the method remains great in much higher dimensional settings especially for small value of n ($n \leq 100$). One must keep in mind that all the bounds that hold in Monte Carlo or quasi-Monte Carlo method are asymptotic and, numerically speaking, usually do not hold before $n \geq 1\,000$ ... as we could check when processing the stochastic gradient...

References

[1] H. Amann, *Ordinary Differential Equations*, de Gruyter (Studies in Mathematics, 13), Berlin, New-York, 1990, 458 p.

[2] C. Bouton, G. Pagès, Self-organization and convergence of the one-dimensional Kohonen algorithm with non uniformly distributed stimuli, pré-publ. n°102 du Labo. de Probabilités de l'Univ. Paris 6 (France), Avril 1992. To appear in *Stochastic Processes and their Applications*.

[3] C. Bouton, G. Pagès, About the multidimensional Kohonen algorithm with 0 neighbour, en préparation.

[4] J.P.-Borel, G. Pagès, I.J. Xiao, Suites à discrépance faible et intégration numérique, *Probabilités Numériques*, N. Bouleau, D. Talay éds., Collection didactique, INRIA, 1992, 204 p.

[5] M. Cottrell, J.C. Fort, Etude d'un algorithme d'auto-organisation, *Annales de l'Institut Henri Poincaré*, 23, 1986, n°1, p.1-20.

[6] T. Kohonen, Self-organizing maps: optimization approaches, *Artificial Neural Networks*, T. Kohonen et al. éditeurs, Elsevier Science Publishers B.V., 1991, p.981-990.

[7] L. Kuipers, H. Niederreiter, *Uniform distribution of Sequences*, New-York, Wiley, 1974, 390 p.

[8] H.J. Kushner, D.S. Clark, Stochastic Approximation for constrained and unconstrained systems, *Applied Math. Science Series*, 26, Springer, 1978.

[9] H. Niederreiter, Quasi-Monte Carlo methods and pseudo-random numbers (survey), *Bull. of the American Math. Soc.*, 84, 1978, p. 957-1041.

[10] G. Pagès, Mosaïques de Voronoï, algorithmes de quantification de l'espace et intégration numérique, pré-publication n°138 du labo. de Proba. de l'Univ. Paris 6 (Décembre 92).

[11] A.H. Stroud, *Approximate Calculation of multiple integrals*, Prentice Hall, Englewood Cliffs (E.U.), 1971, 431 p.