On Hölder fields clustering

Benoît CADRE*, Quentin PARIS

IRMAR, ENS Cachan Bretagne, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman, 35170 Bruz, France
cadre, paris@bretagne.ens-cachan.fr

Abstract

Based on n randomly drawn vectors in a Hilbert space, we study the k-means clustering scheme. Here, clustering is performed by computing the Voronoi partition associated with centers that minimize an empirical criterion, called distorsion. The performance of the method is evaluated by comparing the theoretical distorsion of empirical optimal centers to the theoretical optimal distorsion. Our first result states that, provided the underlying distribution satisfies an exponential moment condition, an upper bound for the above performance criterion is $O(1/\sqrt{n})$. Then, motivated by a broad range of applications, we focus on the case where the data are real-valued random fields. Assuming that they share a Hölder property in quadratic mean, we construct a numerically simple k-means algorithm based on a discretized version of the data. With a judicious choice of the discretization, we prove that the performance of this algorithm matches the performance of the classical algorithm.

Index Terms — Clustering, k-means, Unsupervised learning, Random fields, Hilbert space, Empirical risk minimization.

1 Introduction

Clustering methods aim at partitioning a complex data set into a series of piecewise groups, or *clusters*, each of which may then be regarded as a separate class of data, thus reducing overall data complexity. This unsupervised learning problem is one of the most widely used tool in exploratory data analysis since in various areas of science, e.g. social science, biology, oceanography, meteorology, finance or computer science, practitioners try to get a first intuition about their data by

*corresponding author

identifying meaningful groups of observations. General references on the subject are to be found in Duda et al (2000), Gersho and Gray (1992), Graf and Luschgy (2000), Gray and Neuhoff (1998) and Linder (2001) among others.

We focus on the so-called *k-means* clustering scheme. It has been extensively studied in the case where the data belongs to a finite-dimensional space, see for instance Pollard (1981,1982a, 1982b), Abaya and Wise (1984), Chou (1994), Linder et al (1994), Bartlett et al (1998), Linder (2000, 2001), Antos (2005) and Antos et al (2005). In this paper, we study the infinite-dimensional setting, in the continuity of the papers by Biau et al (2008) and Maurer and Pontil (2010) for instance. This setting is motivated by a broad range of applications, e.g. turbulence, image analysis, finance, speech recognition, Kinematics of chemical reactions, etc. Here, the data to be clustered are modelled by random fields, e.g. random variables taking values in $\mathbb{L}^2([0,1]^s)$. To mention a few of them, fractional Brownian fields (Lindstrøm, 1993, Mandelbrot, 1997, Mandelbrot and van Ness, 1968) have been proved to be the key tools in the modelling of long-term dependency, for instance in the analysis of river level height (Kärner, 2001) or turbulence (Frisch, 1995). Other examples are given by the class of Brownian diffusion processes or Lévy fields, that are central objects in financial mathematics (Cont and Tankov, 2003, Lamberton and Lapeyre, 1996), and Markov processes that modelize the kinematics of chemical reactions (van Kampen, 2007) and evolutionary dynamics (Kimmel and Axelrod, 2002).

The following observations will guide the study of this paper : as seen above, many relevant random variables in stochastic modelling are unbounded, and they share a Hölder type property in quadratic mean. This leads to first study in Section 2 the performances of the k-means scheme in the case of unbounded Hilbert-valued random variables (Theorem 2.1). In fact, the k-means algorithm requires to minimize a function over the infinite-dimensional sample space (see Section 2). To circumvent this drawback, we consider in Section 3 the case of $\mathbb{L}^2([0,1]^s)$-valued random variables (random fields) and we construct a numerically simple algorithm involving discretized versions of the fields, i.e. fields defined over a finite grid. Assuming that the random fields share a Hölder property, we prove that the performance of this algorithm matches the performance of the classical k-means scheme (Theorem 3.1).

2 Clustering in Hilbert space

2.1 *k*-means algorithm

We now recall the general clustering context, in which the observation space $(\mathscr{H}, \|.\|)$ is a Hilbert space. In this setting, the data to be clustered is a sequence of independent $\mathscr{H}$-valued random observations $X_1, \cdots, X_n$ with the same distribution as a generic square integrable random variable X with distribution μ. We focus on the k-means algorithm, which prescribes a criterion for partitioning the sample space into k clusters, by minimizing the *empirical distorsion*

$$W_k(\mathbf{c}, \mu_n) = \frac{1}{n} \sum_{i=1}^{n} \min_{j=1,\cdots,k} \|X_i - c_j\|^2,$$

over all centers $\mathbf{c} = (c_1, \cdots, c_k) \in \mathscr{H}^k$. Here, μ_n is the empirical measure defined by

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \in A\}},$$

for all Borel set $A \subset \mathscr{H}$. Associated with the centers c_j's are the convex polyhedrons S_j of all points in $\mathscr{H}$ closer to c_j than to any other center. Then, $\{S_1, \cdots, S_k\}$ forms a partition of $\mathscr{H}$, called the Voronoi partition, and the S_j's are the clusters of interest. From a theoretical point of view, the performance of a clustering scheme given by the centers $\mathbf{c} = (c_1, \cdots, c_k) \in \mathscr{H}^k$ is evaluated by the *theoretical distorsion*

$$W_k(\mathbf{c}, \mu) = \mathbb{E} \min_{j=1,\cdots,k} \|X - c_j\|^2.$$

Clustering methods aim at approximating the *clustering risk*, defined by

$$W_k^\star(\mu) = \inf_{\mathbf{c} \in \mathscr{H}^k} W_k(\mathbf{c}, \mu).$$

More precisely, the performance of a clustering scheme based on the empirical centers $\mathbf{c}_n = (c_{n1}, \cdots, c_{nk})$ is evaluated by

$$W_k(\mathbf{c}_n, \mu) - W_k^\star(\mu).$$

Since the early work from Hartigan (1975, 1978), many authors have contributed to the study of k-means clustering based on a minimizer $\mathbf{c}_n$ of the empirical distorsion, namely

$$W_k(\mathbf{c}_n, \mu_n) = \inf_{\mathbf{c} \in \mathscr{H}^k} W_k(\mathbf{c}, \mu_n). \tag{2.1}$$

Proof of the existence of $\mathbf{c}_n$ is to be found for instance in Graf and Luschgy (2000, Theorem 4.12). It is proved in Corollary 2.1 in Biau et al (2008) that, provided $\|X\|$ is bounded by M, then for any $\delta \in]0,1[$

$$W_k(\mathbf{c}_n,\mu) - W_k^\star(\mu) \leq \frac{12kM^2 + 4M\sqrt{-2\ln\delta}}{\sqrt{n}} \tag{2.2}$$

with probability at least $1-\delta$. The proof of (2.2) consists in two steps: first establish the result in mean, and then conclude with the McDiarmid Inequality (e.g. see Devroye et al, 1996). In the case of non-bounded random variables, however, such a proof can not hold because McDiarmid's Inequality is based on a boundedness property of the random variables. Our first task is to extend this results to the case of unbounded random variables.

2.2 Clustering unbounded random variables

In the whole paper, we denote by $\mathbf{c}_n = (c_{n1}, \cdots, c_{nk}) \in \mathscr{H}^k$ a vector that minimizes the empirical clustering risk in $\mathscr{H}^k$:

$$W_k(\mathbf{c}_n,\mu_n) = \inf_{\mathbf{c}\in\mathscr{H}^k} W_k(\mathbf{c},\mu_n). \tag{2.3}$$

Then, $\{S_{n1}, \cdots, S_{nk}\}$ stands for the Voronoi partition of $\mathscr{H}$ which is associated with the centers $c_{n1} \in S_{n1}, \cdots, c_{nk} \in S_{nk}$ (for a definition and properties of the Voronoi partition, we refer the reader to Chapter 1 in Graf and Luschgy, 2000). We know from Lemma 1 in Linder (2001) that for all $j = 1, \cdots, k$, the center c_{nj} has the following expression:

$$c_{nj} = \frac{\sum_{i=1}^n X_i \mathbf{1}_{\{X_i \in S_{nj}\}}}{\sum_{i=1}^n \mathbf{1}_{\{X_i \in S_{nj}\}}}. \tag{2.4}$$

The aim of the section is to study the performance of the empirical centers $\mathbf{c}_n = (c_{n1}, \cdots, c_{nk}) \in \mathscr{H}^k$.

We shall assume that, for some $\tau > 0$,

$$\mathbb{E}\, e^{\tau\|X\|} < \infty, \tag{2.5}$$

and we denote by $R(\mu)$ the quantity:

$$R(\mu) = \frac{1}{\tau}\left(1 + \omega(\mu) + \ln \mathbb{E}\, e^{\tau\|X\|}\right),$$

where $\omega(\mu) = \ln^-[W^\star_{k-1}(\mu) - W^\star_k(\mu)]$, if $\ln^- x = \max(0, -\ln x)$ stands for the negative part of $\ln x$. Recall that $W^\star_{k-1}(\mu) > W^\star_k(\mu)$ when the support of μ contains at least k points (e.g., see Theorem 4.12 in Graf and Luschgy, 2000).

Theorem 2.1. *Assume that* (2.5) *holds and the support of* μ *contains at least k points. There exists a universal constant* $C > 0$ *such that for all* $\delta \in]0,1[$*, one has*

$$W_k(\mathbf{c}_n, \mu) - W^\star_k(\mu) \leq CkR(\mu)^2 \sqrt{\frac{\ln(1/\delta)}{n}},$$

with probability $(1-\delta) - Ke^{-rn^{1/5}}$*, where* $r, K > 0$ *only depend on* $\mathbb{E}e^{\tau\|X\|}$ *and k.*

Due to the generality of the situation under study, the obtained value for the numerical constant C is large. However, the interest of Theorem 2.1 is to point out the contribution of each parameter, especially n, δ and μ.

Examples:

1. *Bounded random variable*. Though our study is not fully adapted to the bounded case, it is of importance to compare the previous result to its equivalent as given in Biau et al (2008). In the case where μ has a bounded support, i.e. $\|X\| \leq M$ for some $M > 0$, then τ can be chosen arbitrarily large, say $\tau = \infty$. Theorem 2.1 reveals that there exists a universal constant $C > 0$ such that for all $\delta \in]0,1[$,

$$W_k(\mathbf{c}_n, \mu) - W^\star_k(\mu) \leq CkM^2 \sqrt{\frac{\ln(1/\delta)}{n}},$$

with probability $(1-\delta) - Ke^{-rn^{1/5}}$, where $r, K > 0$ only depend on M and k. In this result, the contribution of each parameter n, k, δ and M is very close to that of Corollary 2.1 in Biau et al (2008).

2. *Diffusion process*. We let $\mathscr{H} = \mathbb{L}_2[0,1]$ be the set of square integrable and real-valued functions on $[0,1]$, and we assume that $X = (X(t))_{t\in[0,1]}$ is a solution to the stochastic differential equation:

$$\mathrm{d}X(t) = b(t, X(t))\mathrm{d}W(t) + \sigma(t, X(t))\mathrm{d}t, \tag{2.6}$$

where $W = (W(t))_{t\in[0,1]}$ is a standard one-dimensional Brownian motion, and b, σ are real-valued Borel functions defined on $[0,1] \times \mathbb{R}$. (For an overview on stochastic differential equations, we refer the reader to the book by Revuz and Yor, 1999). In a recent paper, Huang (2009) finds a wide class of diffusion processes such as (2.6) so that the exponential condition (2.5) holds. For simplicity, we shall assume for this example the stronger conditions that functions b and σ are bounded. In this case, we can prove (see the Appendix) that, for some judicious choice of $\tau > 0$,

$$R(\mu) \leq C'(\sup|b| + \sup|\sigma|)(1 + \omega(\mu)), \tag{2.7}$$

for some numerical constant $C' > 0$. Then by Theorem 2.1, there exists a universal constant $C > 0$ such that for all $\delta \in]0,1[$,

$$W_k(\mathbf{c}_n, \mu) - W_k^\star(\mu) \leq Ck(\sup|b| + \sup|\sigma|)^2(1 + \omega(\mu))^2\sqrt{\frac{\ln(1/\delta)}{n}},$$

with probability $(1-\delta) - Ke^{-rn^{1/5}}$, where $r, K > 0$ only depend on $\mathbb{E}e^{\tau\|X\|}$ and k.

3 Hölder fields clustering

3.1 Numerical step in the clustering of random fields

The k-means algorithm assigns the data $x \in \mathscr{H}$ to the j-th cluster if for all $\ell = 1, \cdots, k$,

$$\|x - c_{nj}\| \leq \|x - c_{n\ell}\|,$$

where $\mathbf{c}_n = (c_{n1}, \cdots, c_{nk})$ is defined by (2.3); that is, if c_{nj} is the nearest-neighbor of x among $\{c_{n1}, \cdots, c_{nk}\}$. However, when the dimension of $\mathscr{H}$ is infinite, the step (2.3) is numerically unrealistic since it involves a minimization in the space $\mathscr{H}^k$. To circumvent this drawback, Biau et al (2008) proposed to make use of the so-called *random projections method*, leading to a minimization step in a finite dimensional space. Roughly speaking, they proved that a projected version of $\mathbf{c}_n$ in a $(n\ln n)$-dimensional space has similar performance than $\mathbf{c}_n$ (Corollary 3.1 in Biau et al, 2008).

If the previous approach has the advantage to hold whatever is the situation, it leads to many complications in term of the computational complexity (e.g., computation of the random projections, orthonormal representation of the data), and it does not fully exploit the particular form of the data under study. Indeed, as mentioned in the introduction, classical stochastic modelling often leads to consider diffusion processes, fractional Brownian fields, Lévy fields, etc. The common point here, is that most of these random fields satisfy a Hölder property. The aim of the next subsection is to exploit this fact in order to derive a simple and tractable method for clustering, in which the numerical step (2.3) is computationaly feasible.

3.2 Discretized fields

For the rest of the section, we assume that $\mathscr{H} = \mathbb{L}_2([0,1]^s)$ is the set of square integrable and real-valued functions defined on $[0,1]^s$. (Here, the choice of $[0,1]^s$ instead of a general space is for convenience of the reader and simplicity of the statements; it turns out that the case of a compact s-dimensional space could be considered as well.) In this setting, the random variable $X = (X(t))_{t\in[0,1]^s}$ is a random field taking values in $\mathscr{H}$.

Our goal is to evaluate the performance of the natural method that simply consists of creating the empirical quantizer, based on the partial information carried by the data that are evaluated over a common finite grid $\mathscr{S} = \{t_1,\cdots,t_d\}$ of $[0,1]^s$. With this respect, the natural questions that arise are: Where must the points of $\mathscr{S}$ be located and what must the size of $\mathscr{S}$ be, so that the performance of an optimal quantizer associated with the discretized data is comparable to that of $\mathbf{c}_n$?

We fix a partition $\pi = \{V_1,\cdots,V_d\}$ of $[0,1]^s$ such that $t_i \in V_i$ and $\text{vol}(V_i) = v$ for all $i = 1,\cdots,d$. Note that the cells V_i have the same volume. In the sequel, the number d is referred to as the *discretization level*. To any function $x = (x(t))_{t\in[0,1]^s} \in \mathbb{L}_2([0,1]^s)$, we associate the *discretized function* $x^\pi = (x^\pi(t))_{t\in[0,1]^s}$ defined as follows:

$$x^\pi(t) = x(t_i) \quad \text{if} \quad t \in V_i,$$

for all $i = 1,\cdots,d$. In the sequel, $\mathbb{L}_2^\pi([0,1]^s)$ is the Hilbert space defined by

$$\mathbb{L}_2^\pi([0,1]^s) = \{x^\pi, x \in \mathbb{L}_2([0,1]^s)\}.$$

Let now μ_n^π be the empirical measure associated with the transformed data $X_1^\pi,\cdots,X_n^\pi$, i.e. for any borel set $A \subset \mathbb{L}_2^\pi([0,1]^s)$:

$$\mu_n^\pi(A) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i^\pi \in A\}}.$$

We define the discretized empirical centers $\mathbf{c}_n^\pi = (c_{n1}^\pi,\cdots,c_{nk}^\pi)$ by

$$W_k(\mathbf{c}_n^\pi,\mu_n^\pi) = \inf_{\mathbf{c}\in(\mathbb{L}_2^\pi([0,1]^s))^k} W_k(\mathbf{c},\mu_n^\pi). \tag{3.1}$$

Observe that since the cells V_j have the same volume v, then for any $\mathbf{c} = (c_1, \cdots, c_k) \in (\mathbb{L}_2^\pi([0,1]^s))^k$:

$$W_k(\mathbf{c}, \mu_n^\pi) = \frac{v}{n} \sum_{i=1}^{n} \min_{j=1,\cdots,k} \sum_{\ell=1}^{d} \big(X_i(t_\ell) - c_j(t_\ell)\big)^2.$$

Then, in comparison with the minimization step (2.3), step (3.1) only requires to minimize the function

$$(\mathbb{R}^d)^k \ni (c_1, \cdots, c_k) \mapsto \sum_{i=1}^{n} \min_{j=1,\cdots,k} |\hat{X}_i - c_j|^2,$$

hence a minimization over a $d \times k$-dimensional space. In the previous formula, $|.|$ stands for the euclidean norm on $\mathbb{R}^d$ and $\hat{X}_i$ is the random vector with coordinates $(X_i(t_1), \cdots, X_i(t_d))$. The aim now is to find the value of d, and the location of the t_i's so that the center $\mathbf{c}_n^\pi$ has similar performance than the center $\mathbf{c}_n$.

3.3 Result

In this subsection, we shall make use of a mean Hölder condition on the random field X. As mentioned in the introduction, this is a mild assumption which is satisfied by most of the relevant random fields that arise in stochastic modelling, such as fractional Brownian fields or diffusion processes for instance. We assume that X satisfies the inequality:

$$\mathbb{E}|X(s) - X(t)|^2 \leq L|s-t|^h, \ \forall s,t \in [0,1]^s, \tag{3.2}$$

for some $h > 0$ and $L > 0$. Here, $|.|$ stands for the euclidean norm in $[0,1]^s$. Observe that assumption (3.2) does not mean that X has Hölder sample-paths, as illustrated by the case of the Poisson process (for which $h = 1$).

For simplicity, we assume that $X(0) = 0$ a.s. and a stronger property than (2.5), namely that for some $\tau > 0$,

$$\mathbb{E}\, e^{\tau\|X\|_\infty} < \infty, \tag{3.3}$$

if $\|.\|_\infty$ stands for the supremum norm of $X = (X(t))_{t\in[0,1]^s}$. Furthemore, we let

$$R_\infty(\mu) = \frac{1}{\tau}\left(1 + \omega(\mu) + \ln \mathbb{E} e^{\tau\|X\|_\infty}\right).$$

The next result shows how to choose the partition π and the discretization level d so that the empirical center $\mathbf{c}_n^\pi$ has similar performance than $\mathbf{c}_n$.

Theorem 3.1. *Assume that* (3.2) *and* (3.3) *hold, and the support of* μ *contains at least k points. If* $d = [n^{s/h}]$ *and each* V_i *is an hypercube with edge length* $1/d^{1/s}$, *then there exists a universal constant* $C > 0$ *such that for all* $\delta \in]0,1[$, *one has*

$$W_k(\mathbf{c}_n^\pi, \mu) - W_k^\star(\mu) \leq \frac{C}{\sqrt{n}} \left(Ls^{h/2} + kR_\infty(\mu)^2 \sqrt{\ln(1/\delta)} \right),$$

with probability $(1-\delta) - Ke^{-rn^{1/5}}$, *for some* $r, K > 0$ *that only depend on* $\mathbb{E}e^{\tau \|X\|_\infty}$ *and k.*

Except for some specific cases, e.g. the diffusion process (2.6) in which $h = 1$, the exact value of h is usually unknown. A lot of attention has been paid to the estimation of h in the litterature, for example, in the case of Gaussian or Lévy fields. With this respect, we refer the reader to the recent papers by Breton et al (2009), Coeurjolly (2008), Lacaux and Loubès (2007) and the references therein.

With our method based on a Hölder property of the random fields, the step (3.1) leads to a minimization in a $k \times [n^{s/h}]$-dimensional space. Hence, except for the case $s > h$ of long-range dependence, we improve the approach by Biau et al (2008). Nevertheless, we recall that our approach is simpler from a computational point of view.

Examples:

1. *Diffusion process.* Suppose that X is the process defined by (2.6), with bounded coefficients terms b and σ. Burkholder Inequality and the exponential Inequality for martingales (see Revuz and Yor, 1999) ensures that assumptions (3.2) and (3.3) are satisfied for any $\tau > 0$, with $h = 1$ and $L = 4(\sup b^2 + \sup \sigma^2)$. Hence by Theorems 3.1 and 2.1 the discretized centers $\mathbf{c}_n^\pi$ provide an upper bound which matches the upper bound for $\mathbf{c}_n$, at least when the discretization level d is n.

2. *Fractional Brownian field.* Assume that X is the fractional Brownian field of index $H \in (0,2)$ (for an overview on fractional Brownian motion/field, we refer the reader to Pipiras and Taqqu, 2003 and Lindstrøm, 1993). In this case, both conditions (3.3) and (3.2) hold, with $L = 1$ and $h = H$. Hence, we deduce from Theorems 3.1 and 2.1 that the empirical quantizers $\mathbf{c}_n^\pi$ and $\mathbf{c}_n$ have similar behavior, provided the discretization level d is chosen so that $d = n^{s/H}$. In the case $s = 1$ with a positive correlation (i.e. $H > 1$), the process has an aggregation behavior, and the numerical step (3.1) is reduced to a minimization in a $n^{1/H} \times k$-dimensional space. From a computational point of view, we considerably improve

the complexity of the minimization procedure, in comparison with the diffusion process above for instance.

4 Proofs

4.1 Proof of Theorem 2.1

From now on, we assume that the exponential moment condition (2.5) holds. We let for any measure ν on $\mathscr{H}$ and $\mathbf{c} = (c_1, \cdots, c_k) \in \mathscr{H}^k$:

$$\bar{W}_k(\mathbf{c}, \nu) = \int_{\mathscr{H}} \min_{j=1,\cdots,k} \left[-2 < x, c_j > + \|c_j\|^2\right] \nu(\mathrm{d}x),$$

where $< .,. >$ stands for the scalar product in $\mathscr{H}$. Furthemore, we denote by B_ρ the closed centered ball in $\mathscr{H}$ with radius ρ. In the proofs, the letter C stands for a positive real number whose value may change from line to line.

Lemma 4.1. *Let $\rho, t > 0$ and denote $\kappa = (\ln \mathbb{E} e^{\tau\|X\|})/\tau$. Then,*

$$\mathbb{P}\left(\sup_{\mathbf{c}\in B_\rho^k} |\bar{W}_k(\mathbf{c}, \mu_n) - \bar{W}_k(\mathbf{c}, \mu)| \geq \frac{16k}{\sqrt{n}}\rho\left(\rho + \sqrt{\mathbb{E}\|X\|^2}\right) + 4t\right)$$
$$\leq 4\exp\left(-\frac{nt^2}{16\rho^2(\mathbb{E}\|X\|^2 + \rho^2)}\right) + 12\exp\left(-\frac{nt}{C\rho(\rho+\kappa)\ln n}\right).$$

PROOF. We set

$$b = \frac{2k}{\sqrt{n}}\rho\left(\rho + \sqrt{\mathbb{E}\|X\|^2}\right).$$

The classical symmetrisation argument (see Devroye et al, 1996, pp. 193-195) reveals that

$$\mathbb{P}\Big(\sup_{\mathbf{c}\in B_\rho^k} |\bar{W}_k(\mathbf{c}, \mu_n) - \bar{W}_k(\mathbf{c}, \mu)| \geq 8b + 4t\Big) \leq 4\mathbb{P}\Big(Z \geq 2b + t\Big), \qquad (4.1)$$

where we put:

$$Z = \sup_{\mathbf{c}\in B_\rho^k} \frac{1}{n}\left|\sum_{i=1}^n \sigma_i \min_{j=1,\cdots,k} \ell_{c_j}(X_i)\right|,$$

with $\ell_c(x) = -2 < x, c > + \|c\|^2$ for $x, c \in \mathscr{H}$, and $\sigma_1, \cdots, \sigma_n$ are i.i.d. Rademacher random variables, independent from the data $X_1, \cdots, X_n$. Proof of Theorem 2.1 in

Biau et al (2008) reveals that $\mathbb{E}Z \leq b$. According to Theorem 4 in Adamczak (2008), we then have:

$$\begin{aligned}
\mathbb{P}(Z \geq 2b+t) &\leq \mathbb{P}(Z \geq 2\mathbb{E}Z+t) \\
&\leq \exp\left(-\frac{nt^2}{4\sigma^2}\right) \\
&\quad +3\exp\left(-\frac{nt}{C\|\max_{1\leq i\leq n}\sup_{\mathbf{c}\in B_\rho^k}\left|\ell_{\mathbf{c}}(X_i)\right|\|_\psi}\right), \qquad (4.2)
\end{aligned}$$

where $\sigma^2 = \sup_{\mathbf{c}\in B_\rho^k}\mathbb{E}\ell_{\mathbf{c}}(X)^2$ and $\|.\|_\psi$ stands for the Orlicz norm associated with $\psi(x) = e^x - 1$, $x \in \mathbb{R}$. Recall that for a real random variable Y, the Orlicz norm is defined by:

$$\|Y\|_\psi = \inf\left\{\lambda > 0 : \mathbb{E}\psi\left(\frac{|Y|}{\lambda}\right) \leq 1\right\}.$$

We now proceed to bound the terms in (4.2). By definition of $\ell_{\mathbf{c}}$, we have:

$$\sigma^2 \leq 4\rho^2(\mathbb{E}\|X\|^2 + \rho^2). \qquad (4.3)$$

Moreover, according to the Pisier Inequality (cf Pisier, 1983),

$$\begin{aligned}
\|\max_{1\leq i\leq n}\sup_{\mathbf{c}\in B_\rho^k}\left|\ell_{\mathbf{c}}(X_i)\right|\|_\psi &\leq C\|\sup_{\mathbf{c}\in B_\rho^k}\left|\ell_{\mathbf{c}}(X)\right|\|_\psi \ln n \\
&\leq 2C\rho(\rho + \|X\|_\psi)\ln n. \qquad (4.4)
\end{aligned}$$

Since $\|X\|_\psi \leq \kappa$, we finally obtain the announced result, combining inequalities (4.1), (4.2), (4.3) and (4.4) $\square$

In the sequel, R is a positive real number that satisfies both conditions:

$$4\mathbb{E}\|X\|^2\mathbf{1}_{B_R^c}(X) < \frac{1}{2}[W_{k-1}^\star(\mu) - W_k^\star(\mu)], \text{ and} \qquad (4.5)$$

$$(R/2-r)^2\mu(B_r) > \frac{1}{2}[W_{k-1}^\star(\mu) + W_k^\star(\mu)] \text{ for some } r \in]0, R/2[. \qquad (4.6)$$

Lemma 4.2. *There exists $K, r > 0$ that only depend on $\mathbb{E}e^{\tau\|X\|}$ and k such that*

$$\mathbb{P}\left(\mathbf{c}_n \notin B_{3R}^k\right) \leq Ke^{-rn^{1/5}}.$$

PROOF. We fix $\varepsilon > 0$ such that

$$4\mathbb{E}\|X\|^2 \mathbf{1}_{B_R^c}(X) < \varepsilon < \frac{1}{2}[W_{k-1}^\star(\mu) - W_k^\star(\mu)].$$

Since R satisfies (4.5) and (4.6), one can prove that (e.g. see the proof of Theorem 4.12 in Graf and Luschgy, 2000):

$$\mathbf{c} \notin B_{3R}^k \Rightarrow W_k(\mathbf{c}, \mu) \geq W_k^\star(\mu) + \varepsilon.$$

Hence,

$$\mathbb{P}\left(\mathbf{c}_n \notin B_{3R}^k\right) \leq \mathbb{P}\left(W_k(\mathbf{c}_n, \mu) - W_k^\star(\mu) \geq \varepsilon\right)$$

But, since for all $\mathbf{c} = (c_1, \cdots, c_k) \in \mathscr{H}^k$,

$$W_k(\mathbf{c}, \mu) = \mathbb{E}\|X\|^2 + \mathbb{E} \min_{j=1,\cdots,k}\left[-2 < X, c_j > + \|c_j\|^2\right], \tag{4.7}$$

we also have

$$\mathbb{P}\left(\mathbf{c}_n \notin B_{3R}^k\right) \leq \mathbb{P}\left(\bar{W}_k(\mathbf{c}_n, \mu) - \bar{W}_k^\star(\mu) \geq \varepsilon\right),$$

where $\bar{W}_k^\star(\mu) = \inf_{\mathscr{H}^k} \bar{W}_k^\star(., \mu)$. According to Theorem 4.12 in Graf and Luschgy (2000), there exists $\eta_0 > 0$ such that $\bar{W}_k^\star(\mu) = \inf_{B_{\eta_0}^k} \bar{W}_k(., \mu)$. Let $\eta \geq \eta_0$. Observing that

$$\bar{W}_k(\mathbf{c}_n, \mu_n) = \inf_{B_\eta^k} \bar{W}_k(., \mu_n),$$

when $\mathbf{c}_n \in B_\eta^k$, we deduce that

$$\begin{aligned} |\bar{W}_k(\mathbf{c}_n, \mu) - \bar{W}_k^\star(\mu)| &\leq |\bar{W}_k(\mathbf{c}_n, \mu) - \bar{W}_k(\mathbf{c}_n, \mu_n)| + \left|\bar{W}_k(\mathbf{c}_n, \mu_n) - \inf_{\mathbf{c} \in B_\eta^k} \bar{W}_k(\mathbf{c}, \mu)\right| \\ &\leq 2 \sup_{\mathbf{c} \in B_\eta^k} |\bar{W}_k(\mathbf{c}, \mu_n) - \bar{W}_k(\mathbf{c}, \mu)| \end{aligned} \tag{4.8}$$

provided $\mathbf{c}_n \in B_\eta^k$. Consequently,

$$\begin{aligned} \mathbb{P}(\mathbf{c}_n \notin B_{3R}^k) &\leq \mathbb{P}\left(\bar{W}_k(\mathbf{c}_n, \mu) - \bar{W}_k^\star(\mu) \geq \varepsilon, \mathbf{c}_n \in B_\eta^k\right) + \mathbb{P}\left(\mathbf{c}_n \notin B_\eta^k\right) \\ &\leq \mathbb{P}\left(\sup_{\mathbf{c} \in B_\eta^k} |\bar{W}_k(\mathbf{c}, \mu_n) - \bar{W}_k(\mathbf{c}, \mu)| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\mathbf{c}_n \notin B_\eta^k\right). \end{aligned} \tag{4.9}$$

We apply Lemma 4.1 with t such that

$$\frac{16k}{\sqrt{n}}\eta\left(\eta+\sqrt{\mathbb{E}\|X\|^2}\right)+4t=\frac{\varepsilon}{2},$$

and $\rho=\eta$ such that t defined above is a positive real number. For the choice of $\eta=n^{1/5}+\ln n/\tau$, we can then find $s>0$ such that

$$\mathbb{P}\Big(\sup_{\mathbf{c}\in B_\eta^k}|\bar{W}_k(\mathbf{c},\mu_n)-\bar{W}_k(\mathbf{c},\mu)|\geq\frac{\varepsilon}{2}\Big)\leq 16\exp(-sn^{1/5}).$$

Moreover, according to (2.4) we get $\|c_{nj}\|\leq\max_{i=1,\cdots,n}\|X_i\|$ for all $j=1,\cdots,k$. Consequently,

$$\begin{aligned}\mathbb{P}\left(\mathbf{c}_n\notin B_\eta^k\right) &\leq k\max_{j=1,\cdots,k}\mathbb{P}\left(c_{nj}\notin B_\eta\right)\\ &\leq k\mathbb{P}\Big(\max_{i=1,\cdots,n}\|X_i\|\geq\eta\Big)\\ &\leq kn\mathbb{P}\left(\|X\|\geq\eta\right)\\ &\leq kne^{-\tau\eta}\mathbb{E}e^{\tau\|X\|}.\end{aligned}$$

We deduce from above and (4.9) that

$$\mathbb{P}\left(\mathbf{c}_n\notin B_{3R}^k\right)\leq\left(16+k\mathbb{E}e^{\tau\|X\|}\right)e^{-rn^{1/5}},$$

for some constant $r>0$, hence the lemma. □

We are now in position to prove Theorem 2.1.

PROOF OF THEOREM 2.1. Let $b>0$ be such that

$$b\geq\frac{4k}{\sqrt{n}}R\left(R+\sqrt{\mathbb{E}\|X\|^2}\right). \tag{4.10}$$

For $t>0$, we derive from (4.7) and (4.8) the chain of inequalities:

$$\begin{aligned}\mathbb{P}\left(W_k(\mathbf{c}_n,\mu)-W_k^\star(\mu)\geq 8b+8t\right) &= \mathbb{P}\left(\bar{W}_k(\mathbf{c}_n,\mu)-\bar{W}_k^\star(\mu)\geq 8b+8t\right)\\ &\leq \mathbb{P}\left(\bar{W}_k(\mathbf{c}_n,\mu)-\bar{W}_k^\star(\mu)\geq 8b+8t,\mathbf{c}_n\in B_{3R}^k\right)\\ &\quad+\mathbb{P}\left(\mathbf{c}_n\notin B_{3R}^k\right)\\ &\leq \mathbb{P}\Big(\sup_{\mathbf{c}\in B_{3R}^k}|\bar{W}_k(\mathbf{c},\mu)-\bar{W}_k(\mathbf{c},\mu_n)|\geq 4b+4t\Big)\\ &\quad+\mathbb{P}\left(\mathbf{c}_n\notin B_{3R}^k\right).\end{aligned}$$

By Lemmas 4.1 and 4.2, we deduce that

$$\begin{aligned}\mathbb{P}\left(W_k(\mathbf{c}_n,\mu)-W_k^\star(\mu)\geq 8b+8t\right) &\leq 4\exp\left(-\frac{nt^2}{4R^2(\mathbb{E}\|X\|^2+R^2)}\right)\\ &\quad +12\exp\left(-\frac{nt}{CR(R+\kappa)\ln n}\right)+Ke^{-rn^{1/5}},\end{aligned}$$

for some $r,K>0$ that only depend on k and $\mathbb{E}e^{\tau\|X\|}$, and where $\kappa=(\ln\mathbb{E}e^{\tau\|X\|})/\tau$. Then, any choice of t so that

$$t\geq\sqrt{\frac{\ln(8/\delta)}{n}4R^2(\mathbb{E}\|X\|^2+R^2)}+\frac{\ln(24/\delta)}{n}CR(R+\kappa)\ln n \tag{4.11}$$

gives

$$\mathbb{P}\left(W_k(\mathbf{c}_n,\mu)-W_k^\star(\mu)\geq 8b+8t\right)\leq\delta+Ke^{-rn^{1/5}}.$$

It is proved in the Appendix that there exists a numerical constant $A>0$ such that if

$$R\geq A\,\frac{1+\omega(\mu)+\ln\mathbb{E}e^{\tau\|X\|}}{\tau} \tag{4.12}$$

then R satisfies both conditions (4.5) and (4.6). Using the upper bound

$$\mathbb{E}\|X\|^2\leq C\left(\frac{1+\ln\mathbb{E}e^{\tau\|X\|}}{\tau}\right)^2 \tag{4.13}$$

proved in the Appendix, we deduce that a choice of t like

$$t=C\left(\frac{1+\omega(\mu)+\ln\mathbb{E}e^{\tau\|X\|}}{\tau}\right)^2\sqrt{\frac{\ln(1/\delta)}{n}},$$

satisfies (4.11). We finally use a similar argument for the choice of b in (4.10), hence the theorem. □

4.2 Proof of Theorem 3.1

In the sequel, μ^π stands for the law of X^π.

Lemma 4.3. *(i) For all* $\mathbf{c}\in\mathscr{H}^k$, $|W_k(\mathbf{c},\mu^\pi)-W_k(\mathbf{c},\mu)|\leq 4Ls^{h/2}d^{-h/(2s)}$.

(ii) We have $|W_k^\star(\mu^\pi)-W_k^\star(\mu)|\leq 4Ls^{h/2}d^{-h/(2s)}$.

PROOF. Proofs of *(i)* and *(ii)* are similar. We only prove *(ii)*. According to Lemma 3 in Linder (2001) and (3.2),

$$\begin{aligned}
\left|W_k^\star(\mu^\pi)^{1/2} - W_k^\star(\mu)^{1/2}\right|^2 &\leq \mathbb{E}\|X^\pi - X\|^2 \\
&= \sum_{p=1}^{d} \int_{V_p} \mathbb{E}\left|X(t_p) - X(t)\right|^2 \mathrm{d}t \\
&\leq L \sum_{p=1}^{d} \int_{V_p} |t_p - t|^h \mathrm{d}t \\
&\leq \frac{Ls^{h/2}}{d^{h/s}},
\end{aligned}$$

because $\mathrm{vol}(V_p) = 1/d$ and $\mathrm{diam}(V_p) = \sqrt{s}/d^{1/s}$ for all $p = 1, \cdots, d$. Consequently,

$$\begin{aligned}
|W_k^\star(\mu^\pi) - W_k^\star(\mu)| &\leq 2\max\left(W_k^\star(\mu^\pi)^{1/2}, W_k^\star(\mu)^{1/2}\right)\frac{\sqrt{Ls^{h/2}}}{d^{h/(2s)}} \\
&\leq 2\left(W_k^\star(\mu)^{1/2} + \frac{\sqrt{Ls^{h/2}}}{d^{h/(2s)}}\right)\frac{\sqrt{Ls^{h/2}}}{d^{h/(2s)}} \\
&\leq 2\left(\sqrt{L}s^{h/4} + \frac{\sqrt{Ls^{h/2}}}{d^{h/(2s)}}\right)\frac{\sqrt{Ls^{h/2}}}{d^{h/(2s)}} \\
&\leq \frac{4Ls^{h/2}}{d^{h/(2s)}},
\end{aligned}$$

since $W_k^\star(\mu) \leq \mathbb{E}\|X\|^2 \leq Ls^{h/2}$, hence *(ii)*. □

We are now in position to prove Theorem 3.1.

PROOF OF THEOREM 3.1. We shall apply Theorem 2.1 to the measure μ^π. Then, observe that the constants $K, r > 0$ of the remainder term $Ke^{-rn^{1/5}}$ may be chosen so that they only depend on k and $\mathbb{E}e^{\tau\|X\|_\infty}$, because

$$\mathbb{E}e^{\tau\|X^\pi\|} \leq \mathbb{E}e^{\tau\|X^\pi\|_\infty} \leq \mathbb{E}e^{\tau\|X\|_\infty}.$$

Since $\mathbb{E}\|X\|^2 \leq Ls^{h/2}$, we deduce from Lemma 4.3 and Theorem 2.1 applied with the measure μ^π, that for some numerical constant $C>0$:

$$\begin{aligned} W_k(\mathbf{c}_n^\pi,\mu)-W_k^\star(\mu) &= [W_k(\mathbf{c}_n^\pi,\mu)-W_k(\mathbf{c}_n^\pi,\mu^\pi)]+[W_k(\mathbf{c}_n^\pi,\mu^\pi)-W_k^\star(\mu^\pi)] \\ &\quad +[W_k^\star(\mu^\pi)-W_k^\star(\mu)] \\ &\leq C\left(\frac{Ls^{h/2}}{d^{h/(2s)}}+kR_\infty(\mu)^2\sqrt{\frac{\ln(1/\delta)}{n}}\right) \end{aligned}$$

with probability $(1-\delta)-Ke^{-rn^{1/5}}$, hence the result. □

5 Appendix

PROOF OF (2.7). We first proceed to bound $\mathbb{E}\exp(\tau\|X\|)\leq\mathbb{E}\exp(\tau\|X\|_\infty)$, where $\|.\|_\infty$ stands for the supremum norm. Denote by $(Z(t))_{t\in[0,1]}$ the continuous-time martingale defined for all $t\in[0,1]$ by

$$Z(t)=\int_0^t b(s,X(s))\mathrm{d}W(s).$$

Since σ is bounded, say $\bar\sigma=\sup|\sigma|$, we have for any $\tau>0$:

$$\mathbb{E}\,e^{\tau\|X\|_\infty}\leq e^{\tau\bar\sigma}\,\mathbb{E}\,e^{\tau\|Z\|_\infty}. \tag{5.1}$$

Hence one only needs to bound the rightmost term. Observe that the quadratic variation $\langle Z\rangle$ of Z satisfies

$$\langle Z\rangle_t=\int_0^t b^2(s,X(s))\mathrm{d}s\leq\bar b^2 t,$$

for all $t\in[0,1]$, where $\bar b$ stands for the supremum of $|b|$. Therefore, by Doob's exponential Inequality for continuous martingales (see Revuz-Yor, 1999):

$$\mathbb{E}\,e^{\tau\|Z\|_\infty}=\int_0^\infty\mathbb{P}\left(\|Z\|_\infty\geq\frac{v}{\tau}\right)e^v\mathrm{d}v\leq 2\int_0^\infty e^{-v^2/(2\tau^2\bar b^2)}e^v\mathrm{d}v\leq\sqrt{2\pi}\,\tau\bar b\,e^{\tau^2\bar b^2/2}.$$

According to (5.1), we then have for all $\tau>0$:

$$\mathbb{E}\,e^{\tau\|X\|_\infty}\leq\sqrt{2\pi}\,\tau\,\bar b\,e^{\tau\bar\sigma+\tau^2\bar b^2/2}.$$

Then, it is an easy exercise to deduce that

$$\inf_{\tau>0}R(\mu)\leq\inf_{\tau\geq1/\bar b}\frac{1}{\tau}\left(1+\omega(\mu)+\ln\mathbb{E}\,e^{\tau\|X\|}\right)\leq C(\sup|b|+\sup|\sigma|)(1+\omega(\mu)),$$

for some numerical constant $C > 0$, hence the result. □

PROOF OF (4.13). Since the function $x \mapsto (\ln x)^2$ defined on $]\exp(1), \infty[$ is concave, one has according to the Jensen Inequality:

$$\begin{aligned}
\mathbb{E}\|X\|^2 &\leq \left(\frac{2}{\tau}\right)^2 + \frac{1}{\tau^2}\mathbb{E}\left(\ln e^{\tau\|X\|}\right)^2 \mathbf{1}_{\{\|X\|\geq 1/\tau\}} \\
&\leq \left(\frac{2}{\tau}\right)^2 + \frac{1}{\tau^2}\left(\ln \mathbb{E}\, e^{\tau\|X\|}\right)^2 \\
&\leq 4\left(\frac{1+\ln \mathbb{E}\, e^{\tau\|X\|}}{\tau}\right)^2,
\end{aligned}$$

hence the result. □

PROOF OF (4.12). The task is to give an upper bound for a positive number R that satisfies both conditions (4.5) and (4.6). On one hand, it is an easy exercise to prove that (4.5) is satisfied for all R greater than

$$A_1 \frac{\ln \mathbb{E} e^{\tau\|X\|} + \omega(\mu)}{\tau},$$

for some numerical constant $A_1 > 0$. On the other hand, using the inequality (4.13), we deduce that (4.6) is verified for all R greater than

$$A_2 \frac{1+\ln \mathbb{E} e^{\tau\|X\|}}{\tau},$$

for some numerical constant $A_2 > 0$. Hence a combination of the above inequalities gives the result. □

References

Abaya, E.A. and Wise, G.L. (1984). Convergence of vector quantizers with applications to optimal quantization, *SIAM J. Appl. Math.*, pp. 183–189

Adamczak, R. (2008). A tail inequality for suprema of unbounded emirical processes with applications to Markov chains, *Electron. J. Probab.*, pp. 1000–1034.

Antos, A. (2005). Improved minimax bounds on the test and training distortion of empirically designed vector quantizers, *IEEE Trans. Inform. Theory*, pp. 4022–4032.

Antos, A., Györfi, L. and György, A. (2005). Improved convergence rates in empirical vector quantizer design, *IEEE Trans. Inform. Theory*, pp. 4013–4022.

Bartlett, P.L. (2003). Prediction algorithms: complexity, concentration and convexity, in *Proceedings of the 13th IFAC Symposium on System Identification*, pp. 1507–1517.

Bartlett, P.L., Linder, T. and Lugosi, G. (1998). The minimax distorsion redundancy in empirical quantizer design, *IEEE Trans. Inform. Theory*, pp. 1802–1813.

Biau, G., Devroye, L. and Lugosi, G. (2008). On the performance of clustering in Hilbert spaces, *IEEE Trans. Inform. Theory*, pp. 781–790.

Bosq, D. (2000). *Linear processes in function spaces. Theory and applications*, Lecture Notes in Statistics 149, Springer-Verlag, New-York.

Breton, J.C., Nourdin, I. and Peccati, G. (2009). Exact confidence intervals for the Hurst parameter of a fractional Brownian motion, *Electron. J. Stat.*, pp. 416-425.

Chou, P.A. (1994). The distorsion of vector quantizers trained on n vectors decreases to the optimum at $O_P(1/n)$, *IEEE Trans. Inform. Theory*, pp. 457–457.

Coeurjolly, J.-F. (2008). Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles, *Ann. Statist.*, pp. 1404-1434.

Cont, R. and Tankov, P. (2003). *Financial Modelling with Jump Processes*, Second Edition, Chapmann and Hall, CRC Press, London.

Devroye, L., Györfi, L. and Lugosi, G. (1996). *A probabilistic theory of pattern recognition*, Springer-Verlag, New-York.

Duda, R.O., Hart, P.E. and Stork, D.G. (2000). *Pattern Classification*, Second Edition, Wiley-Interscience, New York.

Frisch, U. (1995). *Turbulences*, Cambridge University Press, Cambridge.

Gersho, A. and Gray, R.M. (1992). *Vector Quantization and Signal Compression*, Kluwer Academic Press, Boston.

Graf, S. and Luschgy, H. (2000). *Foundations of quantization for probability distributions*, Lectures Notes in Mathematics 1730, Springer-Verlag, New-York.

Gray, R.M. and Neuhoff, D.L. (1998). Quantization. *IEEE Trans. Inform. Theory*, pp. 2325–2384.

Hartigan, J.A. (1975). *Clustering Algorithms*, Wiley, New-York.

Hartigan, J.A. (1978). Asymptotic distributions for clustering criteria, *Ann. Statist.*, pp. 117–131.

Huang, W. (2009). Exponential integrability of Itô's processes, *J. Math. Anal. Appl.*, pp. 427-433.

Kärner, O. (2001). Comments on Hurst exponent, *Geophysical research Letters*, pp. 3825-3826.

Kimmel, M. and Axelrod, D.E. (2002). *Branching Processes in Biology*, Springer, Berlin.

Lacaux, C. and Loubès, J.-M. (2007). Hurst exponent estimation of fractional Lévy motions, *Alea*, pp. 143-164.

Lamberton, D. and Lapeyre, B. (1996). *Introduction to Stochastic Calculus applied to Finance*, Chapman and Hall, CRC Press, London.

Linder, T. (2000). On the training distortion of vector quantizers, *IEEE Trans. Inform. Theory*, pp. 1617–1623.

Linder, T. (2001). *Learning-theoretic methods in vector quantization*, Lecture Notes for the Advanced School on the Principle of Nonparamteric Learning, Udine, Italy, July 9-13.

Linder, T., Lugosi, G. and Zeger, K. (1994). Rates of convergence in the source coding theorem, in empirical quantizer design, and in universal lossy source coding, *IEEE Trans. Inform. Theory*, pp. 1728–1740.

Lindstrøm, T. (1993). Fractional Brownian fields as integrals of white noise, *Bull. London Math. Soc.*, pp. 83-88.

Mandelbrot, B. (1997). *Fractals and scaling in finance*. Selected Works of Benoit B. Mandelbrot. Discontinuity, concentration, risk. Selecta Vol. E, with a forward by R. E. Gomory. Springer-Verlag, New York.

Mandelbrot, B. and van Ness, J. (1968). Fractional Brownian motion, fractional noises and applications. *SIAM Rev.*, pp. 422–437.

Maurer, A. and Pontil, M. (2010). K-dimensional coding schemes in Hilbert spaces, *IEEE Trans. Inform. Theory*, to appear.

Pipiras, V. and Taqqu, M.S. (2003). Fractional calculus and its connection to fractional Brownian motion, in: *Long Range Dependence*, pp. 166-201, Birkhäuser, Basel.

Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Banach spaces, harmonic analysis, and probability theory*, Lecure Notes in Math. **995**, pp. 123–154, Springer, Berlin.

Pollard, D. (1981). Strong consistency of k-means clustering, *Ann. Statist.*, pp. 135–140.

Pollard, D. (1982). A central limit theorem for k-means clustering, *Ann. Probab.*, pp. 199–205.

Pollard, D. (1982). Quantization and the method of k-means, *IEEE Trans. Inform. Theory*, pp. 1728–1740.

Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, Springer-Verlag, New-York.

van Kampen, N.G. (2007). *Stochastic Processes in Physics and Chemistry, 3rd Edition*, Elsevier, New-York.