A quantization algorithm for solving multi-dimensional discrete time optimal stopping problems(Ms#01-36 rev. 1)

Vlad BALLY [*] Gilles PAGÈS [†]

10.03.03

Abstract

A new grid method for computing the Snell envelope of a function of a $\mathbb{R}^d$-valued simulatable Markov chain $(X_k)_{0\le k\le n}$ is proposed. (This is a typical non linear problem that cannot be solved by the standard Monte Carlo method.) Every X_k is replaced by a "quantized approximation" $\widehat{X}_k$ taking its values in a grid Γ_k of size N_k. The n grids and their transition probability matrices make up a discrete tree on which a pseudo-Snell envelope is devised by mimicking the regular dynamic programming formula. We show, using Quantization Theory of random vectors, the existence of a set of optimal grids, given the total number N of elementary $\mathbb{R}^d$-valued quantizers. A recursive stochastic gradient algorithm, based on simulations of $(X_k)_{0\le k\le n}$, yields these optimal grids and their transition probability matrices. Some *a priori* error estimates based on the L^p-quantization errors $\|X_k-\widehat{X}_k\|_p$ are established. These results are applied to the computation of the Snell envelope of a diffusion approximated by its (Gaussian) Euler scheme. We apply these result to provide a discretization scheme for Reflected Backward Stochastic Differential Equations. Finally, a numerical illustration is carried out on a 2-dimensional American option pricing problem.

Key words: Numerical Probability, Markov chains, Snell envelope, Quantization of random variables, Reflected Backward Stochastic Differential Equation, American option pricing.

AMS classification: Primary 60G40, 91B28, Secondary 65C05, 65C20, 65C30, 65N50.

Introduction

Since the 40's, the theory of Markov Processes and Stochastic Calculus have provided a probabilistic interpretation for the solutions of linear Partial Differential Equations (PDE's) based on the Feynman-Kac formula. One of its most striking application is the emergence of the Monte Carlo method as an alternative to deterministic numerical algorithms for solving linear PDE's. It is widely known that the Monte Carlo method has two advantages: its rate of convergence does not depend upon the dimension d of the state space and is not affected by possible degeneracy of the second order terms of the

[*] Univ. du Maine, Labo. Stat. et Proc., B.P. 535, 72001 Le Mans Cedex et Projet MATHFI, INRIA-Rocquencourt (France). E-mail: bally@ccr.jussieu.fr

[†] Labo. de Probabilité et Modélisation, UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5. E-mail: gpa@ccr.jussieu.fr.

equation. As soon as $d \geq 4$, the probabilistic approach often remains the only numerical method available.

In the 90's the theory of Backward Stochastic Differential Equations ($BSDE$, see [49, 21, 22, 2],...) provided a probabilistic interpretation for nonlinear problems (semi-linear-PDE's, PDE's with obstacle,...). For example let us focus for a while on the problem of semi-linear PDE's with obstacle (in the weak sense):

$$\max((\partial_t + L)u + f(t,x,u), h(t,x) - u(t,x)) = 0, \quad 0 \leq t \leq T,\ u_T = h(T,.) \tag{1}$$

where $f : [0,T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, $h : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ are Lipschitz continuous and L is a second order differential operator defined on twice differentiable function on $\mathbb{R}^d$ by

$$Lu(x) := < b|\nabla u > (x) + \frac{1}{2}\mathrm{Trace}(\sigma^* \nabla^2 u\, \sigma)(x)$$

($b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathcal{M}(d \times q)$ are Lipschitz continuous functions, $|\,.\,|$ is for Euclidean norm on $\mathbb{R}^d$, $(\,.\,|\,.\,)$ for the inner product). The object involved in the probabilistic interpretation of (1) is the Reflected Backward Stochastic Differential Equation ($RBSDE$) associated with the diffusion process $(X_t)_{t\in[0,T]}$ solution of the Stochastic Differential Equation

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s \tag{2}$$

where $(B_t)_{t\in[0,T]}$ is a standard Brownian Motion on $\mathbb{R}^q$ (its completed filtration is denoted by $\underline{\mathcal{F}} := (\mathcal{F}_t)_{t\in[0,T]}$. The (solution) of the $RBSDE$ is defined as a triplet (Y, Z, K) of *square integrable*, $\underline{\mathcal{F}}$-progressively measurable processes satisfying

$$\begin{aligned} Y_t &= h(T, X_T) + \int_t^T f(s, X_s, Y_s)ds + K_T - K_t - \int_t^T Z_s dB_s, & (3)\\ Y_t &\geq h(t, X_t) \text{ and } \int_0^T (Y_t - h(t, X_t))\, dK_t = 0, & (4)\\ (K_t)_{t\in[0,T]} &\text{ has nondecreasing continuous paths and } K_0 := 0. \end{aligned}$$

The intuition on this definition is as follows: we wish to solve a $BSDE$ (*i.e.* (3) but we also ask Y to remain larger then the obstacle $h(t, X_t)$. Then we need a nondecreasing process K to "push" Y upwards. K is asked to be minimal: it pushes in the critical situation $Y_t = h(t, X_t)$ only. Z appears as the strategy to be used so that Y_t starts with Y_0 at time $t = 0$ and reaches $h(T, X_T)$ at time T *in a non anticipative way* although $h(T, X_T)$ depends on the whole information up to T.

Theorem *(El Karoui, Kapoudjan, Pardoux, Peng, Quenez, [22], 97) Assume that the following assumptions hold for some real constant $\gamma_0 > 0$:*

$$\begin{aligned} (Lip_{b,\sigma}) &\equiv \forall\, x, x' \in \mathbb{R}^d, \quad |\sigma(x) - \sigma(x')| \vee |b(x) - b(x')| \leq \gamma_0\, |x - x'|\,, & (5)\\ (Lip_f) &\equiv \forall\, t, t' \in \mathbb{R}_+,\ x, x' \in \mathbb{R}^d,\ y, y' \in \mathbb{R}, & \\ & \qquad |f(t,x,y) - f(t',x',y')| \leq \gamma_0(|t - t'| + |x - x'| + |y - y'|), & (6)\\ (Lip_h) &\equiv \forall\, t, t' \in \mathbb{R}_+,\ x, x' \in \mathbb{R}^d, \quad |h(t,x) - h(t',x')| \leq \gamma_0(|x - x'| + |t - t'|). & (7) \end{aligned}$$

Then the RBSDE (3) has a unique solution (Y, Z, K). Furthermore, the process Y admits the following representation

$$Y_t = u(t, X_t)$$

where u denotes the unique solution (in the viscosity sense) of (1).

Another approach is developed in [2]: the function u solves in a variational sense the PDE with obstacle h and u is the minimal solution for the corresponding variational inequality. Then, using the connections between variational inequalities and Optimal Stopping theory (see [9]) leads to represent the above process Y as a Snell envelope:

Proposition *If $(Y_t)_{t\in[0,T]}$ solves the RBSDE (3), then*

$$Y_t = \operatorname{esssup}_{\tau\in\mathcal{T}_t}\mathbb{E}\left(\int_t^\tau f(s,X_s,Y_s)ds + h(\tau,X_\tau)/\mathcal{F}_t\right) \tag{8}$$

where $\mathcal{T}_t$ is the set of $[t,T]$-valued $\underline{\mathcal{F}}$-stopping times.

When the function f *does not depend upon* Y_t, Equation (8) becomes an alternate definition of $(Y_t)_{t\in[0,T]}$ which then appears as the usual Snell envelope of a regular Optimal Stopping problem associated to the Brownian diffusion $(X_t)_{t\in[0,T]}$. Then f represents the *instantaneous gain* and h the *final gain*. Of course Optimal Stopping Theory for diffusions is a very classical topic in Probability Theory and its numerical aspects have been investigated since the very beginning of Numerical Probability, motivated by a wide range of applications in engineering. Let us mention the work by Kushner in [33] for elliptic diffusions and the book by Bensoussan & Lions [9]. However, Mathematical Finance made it still more strategic in the 80's: pricing an American option is in some way an almost generic Optimal Stopping problem (with $f\equiv 0$, $h\ge 0$ and X a nonnegative martingale).

In a few words, a (*vanilla*) American option is a contract that gives the right to receive once and only once $h(t,X_t)$ currency units, at a time t chosen between time 0 and the maturity $T>0$ of the contract. The possibly multidimensional (*nonnegative*) process $(X_t)_{t\in[0,T]}$ is called the *underlying asset* or *risky asset* process. If one assumes for the sake of simplicity that the interest rate is 0, classical arguments in the modeling of financial markets show that the price Y_t of such a contract is given at every time t by (8) (setting $f\equiv 0$), with respect to a so-called *risk-neutral* probability which makes the diffusion $(X_t)_{t\in[0,T]}$ a martingale. Among all possible models for the asset dynamics, the geometric Brownian Motion on $\mathbb{R}^q$:

$$dX_t = \sigma\,X_t\,dB_t, \qquad X_0 = x_0\in\mathbb{R}^d_+, \qquad \sigma\in\mathcal{M}(d,q),$$

is widely adopted. When $q=d$, it is known as the *Black & Scholes model.* Another question of interest is to know, at every time t, if there is some optimal stopping strategy to exercise this right in the future? An answer is provided by the (lowest) optimal stopping time given by $\tau_*^t := \inf\{s\ge t\,/\,Y_s = h(s,X_s)\}$ in the sense that τ_*^t satisfies $Y_{\tau_*^t} = \mathbb{E}(h_{\tau_*^t}/\mathcal{F}_t)$.

Historically, the underlying asset of the first massively traded American option contracts was 1-dimensional (a single stock). However, many American options, mostly traded "Over The Counter", have a much more complex structure depending on a whole basket of d underlying risky assets $X_t := (X_t^1,\dots,X_t^d)$. If one thinks *e.g.* to indices (Dow Jones, Dax, CAC40, etc), d is usually greater than 2, 3 or 4.

The usual numerical approach to solve (low dimensional) Optimal Stopping problems is essentially analytic: it consists in solving the variational inequality (1) using usual techniques of Numerical Analysis (finite differences, finite elements, finite volumes, ...). In spite of the loss of probabilistic interpretation, especially when implicit schemes are used, these methods are unrivalled in 1 or 2 dimension in terms of rate of converge. This holds similarly for $RBSDE$'s.

However, the own development of Mathematical Finance, mainly influenced by its probabilistic background, gave rise to algorithms directly derived from discretizations of the Snell envelope (8) (when $f \equiv 0$). The most famous method is undoubtedly the binomial tree made very popular in the financial world by its implementation and interpretation simplicity. Some very accurate rates of convergence are available for such models in the case of the vanilla American Put ($h(t,x) := \max(K-x,0)$) (see [37, 38, 42, 7]).

As the dimension increases, the implementation of analytic methods fail and the probabilistic interpretation becomes the key to any numerical approach. Although in this paper is discussed the more general case of the (h,f)-Snell envelopes to embody the numerical approximation of solutions of $RBSDE$, let us carry on with regular Optimal Stopping for a while. All probabilistic approaches roughly follow the same three steps:

– Time discretization: One approximates the $\underline{\mathcal{F}}$-adapted diffusion process $(X_t)_{t\in[0,T]}$ at times $t_k = \frac{kT}{n}$, $k = 0,\dots,n$, by an $(\widetilde{\mathcal{F}}_k)_{0\le k\le n}$-*Markov chain* $\widetilde{X} = (\widetilde{X}_k)_{0\le k\le n}$, where $\widetilde{\mathcal{F}}_k = \mathcal{F}_{t_k}$. The chain $\widetilde{X}$ is assumed *to be easily simulatable on a computer* (index k is then for absolute time t_k). Then one approximates the continuous time $\underline{\mathcal{F}}$-Snell envelope Y of X by the discrete time $\underline{\widetilde{\mathcal{F}}}$-Snell envelope of $\widetilde{X}$ defined by

$$\widetilde{U}_k \quad := \quad \text{esssup}\left\{\mathbb{E}\left(h(\theta\frac{T}{n}, \widetilde{X}_\theta/\widetilde{\mathcal{F}}_k\right),\ \theta \in \Theta_k\right\}$$

where Θ_k denotes the set of $\{k\dots n\}$-valued stopping times. The approximating process $\widetilde{X}$ will often be set as the Euler scheme $(\overline{X}_k)_{0\le k\le n}$ (with Gaussian increments) of the diffusion, but other choices are possible (Milshtein scheme, etc). In 1-dimension a binomial tree can be considered for a weak approximation. When samples $(X_0, X_{t_1},\dots,X_{t_n})$ of the diffusion are simulatable, *e.g.* because $X_t = \varphi(t,B_t)$ like in the Black & Scholes model, the best choice is of course to consider $\widetilde{X}_k = X_{t_k}$. Then, its $(\widetilde{\mathcal{F}}_k)_{0\le k\le n}$-Snell envelope will be simply denoted by $(U_k)_{0\le k\le n}$. In the accompanying paper [4] a detailed analysis of the resulting L^p-error is carried out: if $(Lip_{b,\sigma})$, (Lip_f), (Lip_h) hold and $\widetilde{X}_k = \overline{X}_k$ or X_{t_k}, then $\|Y_{kT/n} - \widetilde{U}_k\|_p = O(1/\sqrt{n})$; if, furthermore, h is *semi-convex* (see (16) below for a definition) then, $\|Y_{kT/n} - U_k\|_p = O(1/n)$.

– Dynamic Programming Principle: this discrete time Snell envelope associated to the obstacle $(h(t_k,\widetilde{X}_k))_{0\le k\le n}$ satisfies (see [48]) the following backward Dynamic Programming Principle

$$\widetilde{U}_n{:=}\,h(t_n,\widetilde{X}_n),\ \widetilde{U}_k{:=}\max\Big(h(t_k,\widetilde{X}_k), \mathbb{E}(\widetilde{U}_{k+1}/\widetilde{\mathcal{F}}_k)\Big) = \max\Big(h(t_k,\widetilde{X}_k), \mathbb{E}(\widetilde{U}_{k+1}/\widetilde{X}_k)\Big). \quad (9)$$

The main feature of this formula is that it involves at each step *the computation of conditional expectations*: this is the probabilistic counterpart for nonlinearity which makes the regular Monte Carlo method fails.

– Computation of conditional expectations: Numerical methods for the massive computation of conditional expectations can roughly be divided in three families: spatial discretization of $\widetilde{X}_k$, regression of (truncated) expansions on a basis of $L^2(\widetilde{X}_k)$ (see [45]), and representation formulae based on Malliavin calculus (like in [25]). The first two approaches are both finite-dimensional. An important drawback of the last two methods – especially in higher dimension – is that they directly depend on the obstacle process $(h(t_k,.))$. A spatial discretization method is "obstacle free" in the sense that it produces a discrete semi-group independently of any obstacle process and then works for any such obstacle process.

The *quantization tree method* that we propose and analyze in this paper belongs to the first family (spatial discretization). Its specificity is not to explode even in higher dimension. The starting idea is simple and shared by many grid methods (see [18, 15, 45]):

– First, at each time step k (that is t_k) one projects $\widetilde{X}_k$ onto a fixed grid $\Gamma_k := \{x_1^k, \ldots, x_{N_k}^k\}$ following *a closest neighbour rule i.e.* one sets

$$\widehat{X}_k := \sum_{1\le i\le N_k} x_i^k \mathbf{1}_{\{\widetilde{X}_k \in C_i^k\}},$$

where $(C_i^k)_{1\le i\le N_k}$ is a Borel partition of $\mathbb{R}^d$ such that $C_i^k \subset \{u \,/\, |u - x_i^k| = \min_{1\le \ell\le N_k} |u - x_\ell^k|\}$.

– As a second step, one "mimics" the above Dynamic Programming Principle (9) on the tree made up by the grids $\Gamma_0, \ldots, \Gamma_n$. The process $\widetilde{X}$ being simulatable it is possible to compute by simulation $\pi_{ij}^k := \mathbb{P}(\widetilde{X}_{k+1} \in C_j^{k+1} \,/\, \widetilde{X}_k \in C_i^k)$. Although $(\widehat{X}_k)_{0\le k\le n}$ is not a Markov chain, one may still define a pseudo-Snell envelope $(\widehat{U}_k)_{0\le k\le n}$ by setting

$$\widehat{U}_n := h(\widehat{X}_n) \quad \text{ and } \quad \widehat{U}_k := \max\left(h(\widehat{X}_k),\, \mathbb{E}(\widehat{U}_{k+1}/\widehat{X}_k)\right),\ 0 \le k \le n-1.$$

Since $\mathbb{E}(\widehat{U}_{k+1}/\widehat{X}_k = x_i^k) = \sum_{j=1}^{N_{k+1}} \pi_{ij}^k \widehat{u}_{k+1}(x_j^{k+1})$, a backward induction shows that $\widehat{U}_k := \widehat{u}_k(\widehat{X}_k)$ where $\widehat{u}_k$ satisfies the following backward dynamic programming formula

$$\begin{cases} \widehat{u}_n(x_i^n) &= h(x_i^n), \qquad 1 \le i \le N_n, \\ \widehat{u}_k(x_i^k) &= \max\left(h(x_i^k), \sum_{j=1}^{N_{k+1}} \widehat{u}_{k+1}(x_j^{k+1})\, \pi_{ij}^k\right),\ 1 \le i \le N_k,\ 0 \le k \le n-1, \end{cases} \tag{10}$$

that we will call *quantization tree algorithm.* One may reasonably expect the error $\widetilde{U}_k - \widehat{U}_k$ to be small. Indeed we are able to prove that, *for some specific choices of grids* Γ_k, and under some appropriate assumptions on the diffusion coefficients, for every $p \ge 1$,

$$\|\widehat{U}_k - \widetilde{U}_k\|_p \le C_p \sum_{k=1}^n \|\widehat{X}_k - \widetilde{X}_k\|_p \ = O\left(\frac{n^{1+\frac{1}{d}}}{N^{1/d}}\right).$$

Practical processing of the quantization tree algorithm (10) raises the following questions:

A. How to specify the grids $\Gamma_k := \{x_1^k, \ldots, x_{N_k}^k\}$? This means two things: first, how to choose in an optimal way the sizes N_k of the grids, given that $N_0 + N_1 + \cdots + N_n = N$, then how to choose the points x_i^k to keep the L^p-quantization errors $\|\widetilde{X}_k - \widehat{X}_k\|_p$ minimal?

B. How to compute the weights π_{ij}^k?

C. How to evaluate the error $\|\widetilde{U}_k - Y_{\frac{kT}{n}}\|_p$ (and $\|U_k - Y_{\frac{kT}{n}}\|_p$)?

D. What is the complexity of the quantization tree algorithm?

One crucial feature of the problem must be emphasized at this stage: whatever the selected methods to get optimal grids and weights can be, phases A, B and C are "one-shot": once the grids are settled, their weights and resulting L^p-quantization errors estimated, the computation of the pseudo-Snell envelope of $(h(t_k, \widehat{X}_k))_{0\le k\le n}$ using the quantization tree algorithm (10) is almost instantaneous on any computer. The numerical experiments carried out in Subsection 5.2 (pricing of American options) indicate that, in fact, the quantization tree grid optimization phase outlined below in Section 2.3 has a very reasonable cost (less than 15 *mn* on 1Ghz PC computer), given the fact that, once completed, one can price instantly any American option pay-off option in that model. Furthermore, in many applications, one may rely on the quantization of universal objects like the standard Brownian motion. In this latter case, the optimization of the quantization amounts by scaling to that of a Normal q-dim vector $\mathcal{N}(0; I_q)$ processed for once and stored on a CD-Rom.

Let us turn now to the optimization phases (A,B and C). Actually this is a very old story: since the early 50's people working in Signal Processing and Information Theory are concerned with the compression of the information contained in a continuous "signal" $(\widetilde{X}_k)$ using a finite number of "codebooks" (the points x_i^k's) in an optimal way (see subsection 2.1). Several deterministic algorithms have been designed for that purpose, essentially 1-dimensional signals. Among them, let us mention the Lloyd's Methods I (see [31]). Meanwhile, a sound mathematical theory of Quantization of probability distribution has been developed (see [30] for a recent monograph). In the 80's, with the emergence of Artificial Neural Networks, some new algorithmic aspects for quantization in higher dimension were investigated, mainly the *Competitive Learning vector Quantization* algorithm (and its variants) which appeared as a degenerate setting of the *Kohonen Self-Organizing Maps* (see [11, 23] and the references therein). This stochastic algorithmic approach will adapted below to Markov dynamics. It is based on massive simulations of independent copies of the random vector to be quantized. A first attempt to apply optimal multi-dimensional quantization to Numerical Probability seems to appear in [50].

As mentioned above, after Kushner's pioneering works, the pricing of American options raised up a new interest for numerical aspects of Optimal Stopping: let us mention among many others, the analysis of the convergence of the Snell envelope in an abstract approximation framework (see [40]) or the rate of convergence of the premium of a regular American put priced in a binomial tree toward its Black & Scholes counterpart (see [37] and the references therein).

In higher dimensions, several numerical methods have been designed and analyzed in the literature during the past ten years to solve Optimal Stopping problems as those naturally arising in Finance or, more generally, to process massive computations of conditional expectations. In the class of grid methods, one may cite the algorithm devised by Broadie and Glasserman (see [15]) for pricing multi-asset American options and the discretization scheme for for *BSDE*'s proposed by Chevance in [18] (the second is 1-dimensional but easy to extend to higher dimension). In both approaches the spatial discretization of discrete time approximation $(\widetilde{X}_k)_k$ consists of N independent copies of $(\widetilde{X}_k)_{0\leq k\leq n}$ which make up a grid of size N at every time step $k = 1, \ldots, n$. In [15], the transition between these grids are based on the likelihood ratios between $\widetilde{X}_k$ and $\widetilde{X}_{k+1}$. A convergence theorem is established without rate of convergence. In [18], $(\widetilde{X}_k)_{0\leq k\leq n}$ is in fact the Euler scheme of a diffusion and at every time step the grid is uniformly weighted by $1/N$. The transition weights are based on an empirical frequency approach based on a Monte Carlo simulation as well. Some *a priori* L^1-error bounds are proposed for functions h having finite variations. In [45], Longstaff and Schwartz develop a regression method based on truncated expansions in $L^2(\widetilde{X}_k)$. They introduce a dual dynamic programming principle for the lowest stopping time τ^* and then compute $\mathbb{E}(h(\tau^*, X_{\tau^*}))$ by a Monte Carlo simulation. The "Malliavin calculus" method was introduced in [25] by Fournié et al. and then developed in [26, 44]. These papers point out the importance of a localization procedure for variance reduction purpose. Optimal localization is investigated in [32] in 1-dimension and extended to d-dimension in [10].

Concerning a weak convergence approach to RSBDE discretization, let us mention [47], but also [13] or [14] (these last two are less directly focused on numerical aspects).

The paper is organized as follows. Section 1 is devoted to the computation of the Snell envelope of a discrete time $\mathbb{R}^d$-valued homogeneous Markov chain using a quantization tree. We start by an example, the discretization of a *RBSDE* in Subsection 1.1, to introduce the (h,f)-Snell envelope. In Subsection 1.2 we propose the (backward) quantization tree algorithm and derive some *a priori* L^p-error bounds using the L^p-quantization error. Section 2 is devoted to optimal quantization, from a theoretical point of view. Then, the extension of the *Competitive Learning Vector Quantization algorithm* to Markov chains is presented to process the numerical optimization of the grids and the computation of their transition weights. Section 3 briefly recalls some error bounds concerning the Monte Carlo estimation of the transition weights (for a fixed possibly not optimal quantization tree and in the linear case $f \equiv 0$). They are established in the accompanying paper [4]. In Section 4 a first comparison with the finite element method is carried out. In Section 5, the above results are applied to the discretization of *RBSDE*'s, with some *a priori* error bounds, when the diffusion is uniformly elliptic). We conclude by a numerical illustration: the pricing of American style exchange options.

NOTATIONS: • For any Lipschitz continuous function $\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}$ we denote by $[\varphi]_{Lip}$ its Lipschitz coefficient $[\varphi]_{Lip} := \sup_{x \neq y} \left|\frac{\varphi(x)-\varphi(y)}{x-y}\right|$.

• The set $\mathcal{M}(d \times q, \mathbb{R})$ of matrices with d rows, q columns and real-valued entries will be endowed with the norm $\|M\| := \sqrt{\mathrm{Tr}(MM^*)}$ where M^* denotes the transpose of M.

• For every finite set A, we denote by $|A|$ its cardinality.

• $\delta_{x,y}$ denotes the usual Kronecker symbol.

• For every $x \in \mathbb{R}$, $[x] := \max\{n \in \mathbb{Z} \,/\, n \leq x\}$ and $\lceil x \rceil := \min\{n \in \mathbb{Z} \,/\, n \geq x\}$.

1 Quantization of the Snell envelope of a Markov chain

Before dealing with the general case, let us look more precisely at the case of the *RBSDE* presented in the introduction.

1.1 Time discretization of a *RBSDE* by a (h,f)-Snell envelope

We follow the notations introduced in the introduction for *RBSDE*'s. It is natural to derive a discretization scheme for the solution of a *RBSDE* from the representation formula (8) following the approach described in the introduction. Let $t_k := \frac{kT}{n}$, $k = 0, \ldots, n$, denote the discretization epochs. One just needs to add a discretization term for the integral $\int_t^T f(s, X_s, Y_s)\, ds$.

– One first considers the homogeneous Markov chain $(X_{t_k})_{0\leq k\leq n}$. Its transition is $P_{\frac{T}{n}}(x,dy)$ where $P_t(x,dy)$ denotes the transition of the diffusion X. The discrete time $(\mathcal{F}_{t_k})_{0\leq k\leq n}$-$(h,f)$-*Snell envelope* $(U_k)_{0\leq k\leq n}$ of $(X_{t_k})_{0\leq k\leq n}$ is defined by

$$U_k := \operatorname{esssup}\left\{\mathbb{E}\left(h(\theta\frac{T}{n}, X_{\theta\frac{T}{n}}) + \frac{T}{n}\sum_{i=k+1}^{\theta} f(t_i, X_{t_i}, U_i) \,/\, \mathcal{F}_{t_k}\right),\ \theta \in \Theta_k\right\} \tag{11}$$

where Θ_k denotes the set of $\{k, \ldots, n\}$-valued $(\mathcal{F}_{t_k})_{0\leq k\leq n}$-stopping times (index k is for absolute discrete time in that formulation). When samples $(X_{t_1}, \ldots, X_{t_n})$ can easily be simulated, $(U_k)_{0\leq k\leq n}$ becomes the quantity of interest. When one deals with the Snell envelope of the Euler scheme, (U_n) remains a tool in the error analysis.

– The Gaussian Euler scheme with general step $\Delta > 0$ (here $\Delta = T/n$) is recursively defined by $\overline{X}^{\Delta}_0 := X_0$ and

$$\forall\, k\in \mathbb{N}, \qquad \overline{X}^{\Delta}_{k+1} := \overline{X}^{\Delta}_k + \Delta b(\overline{X}^{\Delta}_k) + \sigma(\overline{X}^{\Delta}_k)\sqrt{\Delta}\ \varepsilon_{k+1}. \tag{12}$$

were $\varepsilon_k := \frac{B_{k\Delta}-B_{(k-1)\Delta}}{\sqrt{\Delta}}$, $k\geq 1$, are i.i.d., $\mathcal{N}(0;I_q)$-distributed. The sequence $(\overline{X}^{\Delta}_k)_{0\leq k\leq n}$ is an homogeneous Markov chain with transition given on bounded Borel functions by

$$P^{\Delta}(f)(x) = \int_{\mathbb{R}^q} f\left(x+\Delta b(x)+\sqrt{\Delta}\,\sigma(x).u\right)e^{-\frac{|u|^2}{2}}\frac{du}{(2\pi)^{\frac{q}{2}}}. \tag{13}$$

The discrete time $(\mathcal{F}_{t_k})_{0\leq k\leq n}$-$(h,f)$-Snell envelope $(\overline{U}_k)_{0\leq k\leq n}$ of $(\overline{X}^{\Delta}_k)_{0\leq k\leq n}$ is

$$\overline{U}_k := \operatorname{esssup}\left\{\mathbb{E}\left(h(\theta\Delta,\overline{X}_\theta)+\frac{T}{n}\sum_{i=k+1}^{\theta} f(t_i,\overline{X}_i,\overline{U}_i)/\mathcal{F}_{t_k}\right),\ \theta\in\Theta_k\right\}. \tag{14}$$

One crucial fact for our purpose is that both transitions of interest $P_\Delta(x,dy)$ and $P^\Delta(x,dy)$ are *Lipschitz* in the sense of the following definition.

Definition 1 *A transition $(P(x,dy))_{x\in\mathbb{R}^d}$ is K-Lipschitz if,*

$$\forall\, g:\mathbb{R}^d\to\mathbb{R},\ \textit{Lipschitz continuous},\quad [Pg]_{Lip}\leq K[g]_{Lip}. \tag{15}$$

Proposition 1 *Assume that the drift $b:\mathbb{R}^d\to\mathbb{R}^d$ and the diffusion coefficient $\sigma:\mathbb{R}^d\to\mathcal{M}(d\times q)$ of the diffusion X are Lipschitz continuous. Set $\Delta := T/n$.*
(a) Euler scheme: *Then P^Δ is Lipschitz with ratio*

$$K^{Euler}_{\Delta} = \sqrt{1+\Delta\gamma_0(2+\gamma_0(1+\Delta))} = 1+\Delta\gamma_0\,(1+\gamma_0/2)+O(\Delta^2).$$

If furthermore, b and σ satisfy the so-called "asymptotic flatness" assumption

$$\exists\, a>0,\ \forall\, x,y\in\mathbb{R}^d,\qquad \frac{1}{2}\|\sigma(x)-\sigma(y)\|^2+(x-y|b(x)-b(y))\leq -a|x-y|^2$$

then $\quad K^{Euler}_{\Delta}\leq\sqrt{1-2a\Delta+\Delta^2\gamma_0^2} = 1-a\Delta+O(\Delta^2).$
(b) Diffusion: *The transition P_Δ is Lipschitz with ratio $K^{diff}_{\Delta} = \exp\left(\gamma_0(1+\gamma_0/2)\Delta\right)$.*

If furthermore, the asymptotic flatness assumption holds, then $K^{diff}_{\Delta} = \exp\left(-a\Delta\right)$.

Proof: (a) Let $g:\mathbb{R}^d\longrightarrow\mathbb{R}$ be a Lipschitz continuous function. Then, for every x, $y\in\mathbb{R}^d$,

$$\begin{aligned}
|P^\Delta(g)(x)-P^\Delta(g)(y)|^2 &\leq \mathbb{E}\left(\left|g\left(x+\Delta b(x)+\sqrt{\Delta}\,\sigma(x)\,\varepsilon_1\right)-g\left(y+\Delta b(y)+\sqrt{\Delta}\,\sigma(y)\,\varepsilon_1\right)\right|^2\right)\\
&\leq [g]^2_{Lip}\mathbb{E}\left(\left|x+\Delta b(x)+\sqrt{\Delta}\,\sigma(x)\,\varepsilon_1-(y+\Delta b(y)+\sqrt{\Delta}\,\sigma(y)\varepsilon_1)\right|^2\right)\\
&\leq [g]^2_{Lip}\Big(|x-y|^2+\Delta\|\sigma(x)-\sigma(y)\|^2+2\Delta(x-y|b(x)-b(y))\\
&\qquad +\Delta^2|b(x)-b(y)|^2\Big)\\
&\leq [g]^2_{Lip}\left(1+\Delta\gamma_0^2+2\Delta\gamma_0+\Delta^2\gamma_0^2\right)|x-y|^2.
\end{aligned}$$

The "asymptotically flat" case is established the same way round.

(b) Itô's formula implies (with obvious notations) that

$$|X^x_t-X^y_t|^2=|x-y|^2+\;2\int_0^t\Big((X^x_s-X^y_s|b(X^x_s)-b(X^y_s))+\frac{1}{2}\mathrm{Tr}((\sigma(X^x_s)-\sigma(X^y_s))(\sigma(X^x_s)-\sigma(X^y_s))^*\Big)ds$$

$$+\int_0^t(X^x_s-X^y_s|(\sigma(X^x_s)-\sigma(X^y_s))dB_s)\qquad\text{(true martingale)}$$

$$\mathbb{E}|X^x_t-X^y_t|^2\le\;|x-y|^2+\gamma_0(2+\gamma_0)\int_0^t\mathbb{E}|X^x_s-X^y_s|^2ds.$$

Gronwall's Lemma finally leads to $\mathbb{E}|X^x_t-X^y_t|^2\le|x-y|^2\exp\left(\gamma_0(2+\gamma_0)t\right)$.

In the"asymptotically flat" case, (16) implies that $t\mapsto\mathbb{E}|X^x_t-X^y_t|^2$ is differentiable and satisfies $\frac{d}{dt}\mathbb{E}|X^x_t-X^y_t|^2\le-a\,\mathbb{E}|X^x_t-X^y_t|^2$, whence the announced result. $\diamond$

Remarks: • The result of the above Proposition still holds if one considers an Euler scheme where the increments ε_k are simply square integrable, centered and normalized.

• The simplest "asymptotically flat" transitions are the Euler scheme of the Ornstein-Uhlenbeck process $dY_t:=-\frac{1}{2}Y_t\,dt+\sigma\,dB_t$ for which the property holds with $a=1/4$.

In the accompanying paper [4], an analysis of the L^p-error induced by considering the discrete time (h,f)-Snell envelopes $(U_k)_{0\le k\le n}$ and $(\overline{U}_k)_{0\le k\le n}$ instead of the process $(Y_t)_{t\in[0,T]}$ is carried out. The main result is summed up in the proposition below.

Proposition 2 *Assume that* $(Lip_{b,\sigma})$, (Lip_f), (Lip_h) *hold and that* $X_0=x\in\mathbb{R}^d$.
(a) Lipschitz continuous setting: *Let* $p\ge1$.

$$\forall\,k\in\{0,\ldots,n\},\qquad\left\|Y_{t_k}-\overline{U}_k\right\|_p+\|Y_{t_k}-U_k\|_p\le C_pe^{C_pT}(1+|x|)\frac{1}{\sqrt{n}}.$$

where C_p *is a positive real constant depending upon* p, b, σ, *and* f h *(by means of* γ_0*).*

(b) Semi-convex setting: *Assume furthermore that* f *is* $\mathcal{C}^{1,2,2}_b$ ([1]) *and that* h *is* semi-convex *in following sense*

$$\forall\,t\in[0,T],\;\forall\,x,\,y\in\mathbb{R}^d,\qquad h(t,y)-h(t,x)\ge(\delta_h(t,x)|y-x)-\rho|x-y|^2\tag{16}$$

where δ_h *is a bounded function on* $[0,T]\times\mathbb{R}^d$ *and* $\rho\ge0$. *Then, the* (h,f)*-Snell envelope of the discretized diffusion* $(X_{t_k})_{0\le k\le n}$ *satisfies for every* $p\ge1$,

$$\forall\,k\in\{0,\ldots,n\},\qquad\|Y_{t_k}-U_k\|_p\le C_pe^{C_pT}(1+|x|)\frac{1}{n}.$$

(the real constant C_p *is a priori different from that in the Lipschitz continuous case).*

Remarks: The semi-convexity assumption is a generalization of convexity which embodies smooth enough functions. This notion seems to have been introduced in [17] for pricing 1-dimensional American options. See also [38] for recent developments in Finance.

– If $h(t,.)$ is convex for every $t\in[0,T]$ with a bounded spatial derivative $\delta_h(t,.)$ (in the distribution sense), then h is semi-convex with $\rho=0$. Thus, it embodies *most American style pay-off functions used in Mathematical Finance* for options pricing like those involving the positive part of linear combination or extrema of the underlying traded asset).

– If $h(t,.)$ is $\mathcal{C}^1$ for every $t\in\mathbb{R}_+$ and $\frac{\partial h}{\partial x}(t,x)$ is ρ-Lipschitz in x, uniformly in t, then h is semi-convex (with $\delta_h(t,x):=s\frac{\partial h}{\partial x}(t,x)$).

[1] $\mathcal{C}^{1,2,2}_b$ is the set of $\mathcal{C}^{1,2,2}$ functions f whose existing partial derivatives are all bounded.

1.2 Quantization of the (h, f)-Snell envelope of a Lipschitz Markov chain

Let $(X_k)_{k\in\mathbb{N}}$ be an homogeneous $\mathbb{R}^d$-valued $(\mathcal{F}_k)_{k\in\mathbb{N}}$-Markov chain with transition $P(x, dy)$. Motivated by the former subsection, one is interested in computing the following (h, f)-Snell envelope $(U_k)_{0\leq k\leq n}$ related to a finite horizon n and to some functions $h := (h_k)_{0\leq k\leq n}$ and $f := (f_k)_{0\leq k\leq n}$ defined on $\{0, \ldots, n\} \times \mathbb{R}^d$ and $\{0, \ldots, n\} \times \mathbb{R}^d \times \mathbb{R}^d$ respectively.

$$U_k := \operatorname{esssup}\left\{\mathbb{E}\left(h_\theta(X_\theta)+\sum_{i=k+1}^{\theta}f_i(X_i, U_i)\,/\,\mathcal{F}_k\right), \theta\ \{k, \ldots, n\}\text{-valued } \mathcal{F}_l\text{-stopping time}\right\}. \quad (17)$$

In fact, the (h, f)-Snell envelope is simply connected with regular Snell envelope appearing in Optimal Stopping Theory: one checks that $V_k := U_k + \sum_{i=1}^{k} f_i(X_i, U_i)$ is the standard Snell envelope of the $\mathcal{F}_k$-adapted sequence $Z_k := h_k(X_k) + \sum_{i=1}^{k} f_i(X_i, U_i)$. Hence, following *e.g.* [48], V is the Snell envelope of Z *i.e.* it satisfies the following Backward Dynamic Programming Principle

$$V_n = Z_n \quad \text{and} \quad V_k = \max\left(Z_k, \mathbb{E}\left(V_{k+1}/\mathcal{F}_k\right)\right).$$

One derives that $(U_k)_{0\leq k\leq n}$ satisfies the following Dynamic Programming Principle

$$\left\{\begin{array}{lll} U_n & := & h_n(X_n), \\ U_k & := & \max\left(h_k(X_k), \mathbb{E}(U_{k+1} + f_{k+1}(X_{k+1}, U_{k+1})/\mathcal{F}_k)\right), \quad 0 \leq k \leq n-1. \end{array}\right. \quad (18)$$

So, it appears as the (h, f)-*Snell envelope* of X.

NOTATION: From now on, $\mathbb{E}(\,.\,/\mathcal{F}_k)$ will be simply denoted $\mathbb{E}_k(\,.\,)$.

At this stage, a straightforward induction using the Markov property shows that, for every $k\in \{0, \ldots, n\}$, $U_k = u_k(X_k)$ where u_k is recursively defined by

$$u_n := h_n, \qquad u_k := \max\left(h_k, P(u_{k+1} + f_{k+1}(\,.\,, u_{k+1})\,)\,\right), \; 0 \leq k \leq n-1. \quad (19)$$

Example: The time discretization of a *RBSDE* corresponds to functions

$$f_k(x, u) := \frac{T}{n} f\left(\frac{kT}{n}, x, u\right) \quad \text{and} \quad h_k(x) := h\left(\frac{kT}{n}, x\right), \; 0 \leq k \leq n. \quad (20)$$

1.2.1 The quantization tree algorithm: a pseudo-Snell envelope

The starting point of the method is to discretize at every step $k\in \{0, \ldots, n\}$ the random vector X_k using a $\sigma(X_k)$-measurable random vector $\widehat{X}_k$ that takes finitely many values. The random variable $\widehat{X}_k$ is called a *quantization* of X_k. One may always associate to $\widehat{X}_k$ a Borel function $q_k : \mathbb{R}^d \to \mathbb{R}^d$ such that $\widehat{X}_k = q_k(X_k)$. The function q_k is often called a *quantizer* (this terminology comes from Signal Processing and Information Theory, see section 2.1). It will be convenient to call *quantization grid* or simply *grid*, the finite subset $\Gamma_k := q_k(\mathbb{R}^d) = \widehat{X}_k(\Omega)$. The size of a grid Γ_k will be denoted N_k. The elements of a quantization grid are called *elementary quantizers*. We will denote by $N := N_0 + N_1 + \cdots + N_n$ the total number of elementary quantizers used to quantize all the X_k's, $0 \leq k \leq n$.

Then, we wish to approximate the Snell envelope $(U_k)_{0\leq k\leq n}$ by a sequence $(\widehat{U}_k)_{0\leq k\leq n}$ formally defined by a dynamic programming algorithm similar to (18) except that

– the random vector X_k is replaced by its quantization $\widehat{X}_k$ for every $k\in \{0, \ldots, n\}$,

– the conditional expectation $\mathbb{E}_k$ (*i.e.* the past of the filtration $\mathcal{F}$ up to time k) is replaced by the conditional expectation given $\widehat{X}_k$ *i.e.* $\mathbb{E}(\,.\,/\widehat{X}_k)$.

NOTATION: for the sake of simplicity, from now on, $\mathbb{E}(\,.\,/\widehat{X}_k)$ will denoted $\widehat{\mathbb{E}}_k(\,.\,)$.

Assume temporarily that for every $k \in \{0, 1, \dots, n\}$, we have access to an appropriate quantization $\widehat{X}_k = q_k(X_k)$ of X_k. The optimal choice of the grid Γ_k and the quantizer q_k that yield the best possible approximation will be investigated further on in section 2.1.

So, the *pseudo-Snell envelope* is defined by mimicking the original one (18) as follows:

$$\left\{\begin{array}{lcl} \widehat{U}_n & := & h_n(\widehat{X}_n), \\ \widehat{U}_k & := & \max\Big(h_k(\widehat{X}_k), \widehat{\mathbb{E}}_k(\widehat{U}_{k+1} + f_{k+1}(\widehat{X}_{k+1}, \widehat{U}_{k+1})\Big), \quad 0 \le k \le n-1. \end{array}\right. \tag{21}$$

The main reason for considering conditional expectation with respect to $\widehat{X}_k$ is that the sequence $(\widehat{X}_k)_{k\in\mathbb{N}}$ does not satisfy the Markov property. Then, the *quantization tree algorithm* simply consists in re-writing the pseudo-Snell envelope in distribution.

Proposition 3 *(Quantization tree algorithm) For every $k \in \{0, \dots, n\}$, let $\Gamma^k := \{x_1^k, \dots, x_{N_k}^k\}$ denote a quantization grid of the distribution $\mathcal{L}(X_k)$ and q_k its quantizer. For every $k \in \{0, \dots, n-1\}$, $i \in \{1, \dots, N_k\}$, $j \in \{1, \dots, N_{k+1}\}$*

$$\pi_{ij}^k := \mathbb{P}(\widehat{X}_{k+1} = x_j^{k+1}/\widehat{X}_k = x_i^k). \tag{22}$$

One defines the functions $\widehat{u}_k$ by the following backward induction

$$\begin{array}{rl} \widehat{u}_n(x_i^n) := & h_n(x_i^n), \qquad i \in \{0, \dots, N_n\}, \\ \widehat{u}_k(x_i^k) := & \max\left(h_k(x_i^k), \displaystyle\sum_{j=1}^{N_{k+1}} \pi_{ij}^k \left(\widehat{u}_{k+1}(x_j^{k+1}) + f_{k+1}(x_j^{k+1}, \widehat{u}_{k+1}(x_j^{k+1}))\right)\right), \\ & \qquad\qquad k \in \{0, \dots, n-1\},\ i \in \{1, \dots, N_k\}. \end{array} \tag{23}$$

Then, $\widehat{u}_k(\widehat{X}_k) = \widehat{U}_k$, $0 \le k \le n$, is the pseudo-Snell envelope defined by (21).

Note that if $\mathcal{L}(X_0) := \delta_{x_0}$, then $\widehat{u}_0(\widehat{X}_0) = \widehat{u}_0(x_0)$ is deterministic, otherwise

$$\mathbb{E}\,\widehat{u}_0(\widehat{X}_0) = \sum_{i=1}^{N_0} p_i^0\, \widehat{u}_0(x_i^0) \qquad \text{with} \qquad p_i^0 := \mathbb{P}(\widehat{X}_0 = x_i^0),\ 1 \le i \le N_0.$$

Implementing this procedure (23) on a computer raises two questions:

– How to estimate numerically the above coefficients π_{ij}^k?

– Is the complexity of the quantization tree algorithm acceptable?

FIRST APPROACH OF THE ESTIMATION PHASE: As far as practical implementation is concerned, the ability to compute the π_{ij}^k's (and the p_i^0's) at a reasonable cost is the key of the whole method. The most elementary solution is simply to process a wide range regular *Monte Carlo simulation of the Markov chain* $(X_k)_{0\le k\le n}$ (see Subsection 2.3). The estimation of the coefficients is based on the representation formula (22) of π_{ij}^k as expectations of simple functions of (X_k, X_{k+1}). Furthermore, the *a priori* error bounds for $\|U_k - \widehat{U}_k\|_p$ that will be derived in Theorem 1 below (see section 1.2.2) all rely on the

L^p-quantization errors $\|X_k - \widehat{X}_k\|_p$, $0 \le k \le n$, which can be simultaneously approximated. So, the parameters of interest can be evaluated provided that independent paths of the Markov chain $(X_k)_{0\le k\le n}$ can be simulated at a reasonable cost. This amounts *to the efficient simulation of some $P(x, dy)$-distributed random vectors for every $x \in \mathbb{R}^d$*.

We will see later on in Subsection 2.3.1 that this first approach can be improved by combining this Monte Carlo simulation with the grid optimization procedure.

Complexity of the algorithm: theory and practice A quick look at the structure of the quantization tree algorithm (23) shows that going from layer $k+1$ down to layer k needs $\kappa \times N_k N_{k+1}$ elementary computations (where $\kappa > 0$ denotes the average number of computations per link "$i \to j$"). Hence, the cost to complete the tree descent is

$$\begin{aligned} \text{Complexity} &= \kappa \times (N_0N_1 + \cdots + N_kN_{k+1} + \cdots + N_{n-1}N_n) \\ \text{so that} \qquad \kappa\,\frac{n}{(n+1)^2}N^2 &\le \text{Complexity} \le \kappa\,\frac{N^2}{4}. \end{aligned} \tag{24}$$

The lower bound holds for $N_k = \frac{N}{n+1}$, $0 \le k \le n$, the upper one for $N_0 = N_1 = \frac{N}{2}$, $N_k = 0$, $2 \le k \le n$, which is clearly unrealistic. The optimal dispatching (see *e.g.* the practical comments in Section 5.1) leads to a complexity close to the lower bound.

However this is a very pessimistic analysis of the complexity. In fact, in most examples – like the Euler scheme – the Markov transition $P(x, dy)$ is such that, at each step k, most coefficients of the quantized transition matrix $[\pi^k_{ij}]$ are so small that their estimates produced by the Monte Carlo simulation turn out to be 0! This is taken into account to speed up the computer procedure so that the practical complexity of the algorithm is $O(N)$. This can be compared to the complexity of a Cox-Ross-Rubinstein's one dimensional binomial tree with $\sqrt{2N}$ time steps which approximately contains N points.

1.2.2 Convergence and rate

The aim of this paragraph is to provide some *a priori* L^p-error bounds for $\|U_k - \widehat{U}_k\|_p$, $0 \le k \le n$, based on the L^p-quantization errors $\|X_k - \widehat{X}_k\|_p$, $0 \le k \le n$, (keep in mind that these quantities can simply be estimated during the Monte Carlo simulation of the chain).

The main necessary assumption on the Markov chain in this section is that its transition $P(x, dy)$ is *Lipschitz* (see Definition 1). This assumption is natural as emphasized by the above Proposition 1: the transitions of a diffusion and of its the Euler scheme are both Lipschitz if its coefficients are Lipschitz continuous. The first task is to evaluate the Lipschitz regularity of the functions u_k defined by (19) in that setting.

Proposition 4 *Assume that the functions h and f are Lipschitz continuous, uniformly with respect to k, that is, for every $k \in \{0, \dots, n\}$,*

$$\forall\, x,\, x' \in \mathbb{R}^d, \qquad |h_k(x) - h_k(x')| \le [h]_{Lip}|x - x'| \tag{25}$$

$$\forall\, x,\, x' \in \mathbb{R}^d, \forall\, u,\, u' \in \mathbb{R}, \quad |f_k(x, u) - f_k(x', u')| \le [f]_{Lip}(|x - x'| + |u - u'|) \tag{26}$$

If the transition P is K-Lipschitz, then the functions u_k defined by (19) are Lipschitz continuous. Furthermore, setting $L := K(1 + [f]_{Lip})$, one gets

- *if $L > 1$,* $\qquad [u_k]_{Lip} \le L^{n-k}\left([h]_{Lip} + \dfrac{K}{L-1}[f]_{Lip}\right),$

- *if* $L = 1$, $\quad [u_k]_{Lip} \;\leq\; [h]_{Lip} + (n-k)[f]_{Lip}$
- *if* $L < 1$, $\quad [u_k]_{Lip} \;\leq\; \max\Big([h]_{Lip}, \dfrac{K}{1-L}[f]_{Lip}\Big).$

Remark: If $f \equiv 0$ (*i.e.* regular Optimal Stopping), the above inequalities read as follows

$$[u_k]_{Lip} \leq (K \vee 1)^{n-k}[h]_{Lip}.$$

For practical applications, *e.g.* to the Euler scheme or to simulatable diffusions, $L \sim 1+c/n$ so the coefficient L^{n-k} does not explode as the number n of time steps goes to infinity.

Proof: As $u_n = h_n$, $[u_n]_{Lip} \leq [h]_{Lip}$. Then, using that

$$|\max(a,b) - \max(a',b')| \leq \max(|a-a'|, |b-b'|),$$

it easily follows from the dynamic programming equality (19) that

$$\begin{aligned}
[u_k]_{Lip} &\leq \max\Big([h]_{Lip}, [P(u_{k+1} + f_{k+1}(.,u_{k+1}))]_{Lip}\Big)\\
&\leq \max\Big([h]_{Lip}, K\left([u_{k+1}]_{Lip} + [f]_{Lip}(1 + [u_{k+1}]_{Lip})\right)\Big)\\
&\leq \max\Big([h]_{Lip}, L[u_{k+1}]_{Lip} + K[f]_{Lip}\Big).\\
\text{An induction shows that }\; [u_k]_{Lip} &\leq L^{-k}\max_{k\leq i\leq n}\Big(L^i[h]_{Lip} + (L^k + \cdots + L^{i-1})[f]_{Lip}\Big)\\
&\leq \max_{0\leq j\leq n-k}\left(L^j[h]_{Lip} + \frac{L^j-1}{L-1}K[f]_{Lip}\right)\ (\text{if } L \neq 1).
\end{aligned}$$

Inspecting the three announced cases completes the proof. $\diamond$

Now we pass to the main result of this section: some *a priori* estimates for $\|U_k - \widehat{U}_k\|_p$ as a function of the quantization error $\|X_k - \widehat{X}_k\|_p$.

Theorem 1 *Assume that the transition $(P(x,dy))_{x\in\mathbb{R}^d}$ is K-Lipschitz and that the functions h and f satisfy the Lipschitz assumptions (25) and (26). Then,*

$$\forall\, p \geq 1,\ \forall\, k \in \{0,\ldots,n\} \quad \|U_k - \widehat{U}_k\|_p \leq \frac{1}{(1+[f]_{Lip})^k}\sum_{i=k}^{n} d_i\,\|X_i - \widehat{X}_i\|_p \qquad \textit{with}$$

$$\begin{cases} d_i := \Big([h]_{Lip} + [f]_{Lip} + (2-\delta_{2,p})K\left(([u_{i+1}]_{Lip}+1)([f]_{Lip}+1)-1\right)\Big)(1+[f]_{Lip})^i,\ 0 \leq i \leq n-1,\\ d_n := ([h]_{Lip} + [f]_{Lip})(1+[f]_{Lip})^n \end{cases} \tag{27}$$

Proof: Set $\Phi_k := P(u_{k+1} + f_{k+1}(.,u_{k+1}))$, $k = 0,\ldots,n-1$, and $\Phi_n \equiv 0$ so that $\mathbb{E}(U_{k+1} + f_{k+1}(X_{k+1}, U_{k+1})/X_k) = \Phi_k(X_k)$. One defines similarly $\widehat{\Phi}_k$ by the equality $\widehat{\Phi}_k(\widehat{X}_k) := \widehat{\mathbb{E}}_k(\widehat{U}_{k+1} + f_{k+1}(\widehat{U}_{k+1}, \widehat{X}_{k+1}))$ and $\widehat{\Phi}_n \equiv 0$. Then

$$\begin{aligned}
|U_k - \widehat{U}_k| &\leq |h_k(X_k) - h_k(\widehat{X}_k)| \;+\; |\Phi_k(X_k) - \widehat{\Phi}_k(\widehat{X}_k)|\\
&\leq [h]_{Lip}|X_k - \widehat{X}_k| \;+\; |\Phi_k(X_k) - \widehat{\mathbb{E}}_k(\Phi_k(X_k))| + |\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k)|.
\end{aligned}$$

$$\begin{aligned}
\text{Now}\quad |\Phi_k(X_k) - \widehat{\mathbb{E}}_k\Phi_k(X_k)| &\leq |\Phi_k(X_k) - \Phi_k(\widehat{X}_k)| + \widehat{\mathbb{E}}_k|\Phi_k(X_k) - \widehat{\mathbb{E}}_k(\Phi_k(\widehat{X}_k))|\\
&\leq [\Phi_k]_{Lip}(|X_k - \widehat{X}_k| + \widehat{\mathbb{E}}_k|X_k - \widehat{X}_k|).
\end{aligned}$$

(Note that $\widehat{X}_k$ is $\mathcal{F}_k$-measurable.) Hence,

$$\|\Phi_k(X_k) - \widehat{\mathbb{E}}\Phi_k(X_k)\|_p \leq 2[\Phi_k]_{Lip}\|X_k - \widehat{X}_k\|_p$$

When $p = 2$, one may drop the factor 2 since the very definition of the conditional expectation as a projection in a Hilbert space implies that

$$\|\Phi_k(X_k) - \widehat{\mathbb{E}}\Phi_k(X_k)\|_2 \leq \|\Phi_k(X_k) - \Phi_k(\widehat{X}_k)\|_2 \leq [\Phi_k]_{Lip}\|X_k - \widehat{X}_k\|_2.$$

On the other hand, coming back to the definition of $\Phi_k(X_k)$ and $\widehat{\Phi}_k(\widehat{X}_k)$, one gets, using that $\widehat{\mathbb{E}}_k \circ \mathbb{E}_k = \widehat{\mathbb{E}}_k$ and that conditional expectation is a L^p-contraction,

$$\begin{aligned}
|\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k)| &\leq \widehat{\mathbb{E}}_k|U_{k+1} + f_{k+1}(X_{k+1}, U_{k+1}) - \widehat{U}_{k+1} - f_{k+1}(\widehat{X}_{k+1}, \widehat{U}_{k+1})| \\
\|\widehat{\mathbb{E}}_k(\Phi_k(X_k)) - \widehat{\Phi}_k(\widehat{X}_k)\|_p &\leq \|U_{k+1} + f_{k+1}(X_{k+1}, U_{k+1}) - \widehat{U}_{k+1} - f_{k+1}(\widehat{X}_{k+1}, \widehat{U}_{k+1})\|_p \\
&\leq (1 + [f]_{Lip})\|U_{k+1} - \widehat{U}_{k+1}\|_p + [f]_{Lip}\|X_{k+1} - \widehat{X}_{k+1}\|_p.
\end{aligned}$$

Finally, for every $k \in \{0, \ldots, n-1\}$,

$$\begin{aligned}
\|U_k - \widehat{U}_k\|_p &\leq [h]_{Lip}\|X_k - \widehat{X}_k\|_p + \|\Phi_k(X_k) - \widehat{\mathbb{E}}_k(\Phi_k(X_k))\|_p + \|\Phi_k(X_k) - \widehat{\Phi}_k(\widehat{X}_k)\|_p \\
&\leq (1 + [f]_{Lip})\|U_{k+1} - \widehat{U}_{k+1}\|_p + ([h]_{Lip} + (2 - \delta_{p,2})[\Phi_k]Lip)\|X_k - \widehat{X}_k\|_p \\
&\quad + [f]_{Lip}\|X_{k+1} - \widehat{X}_{k+1}\|_p.
\end{aligned}$$

Using that $\|U_n - \widehat{U}_n\|_p \leq [h]_{Lip}\|X_n - \widehat{X}_n\|_p$, standard computations yield

$$\begin{aligned}
\|U_k - \widehat{U}_k\|_p &\leq \sum_{i=k}^{n}\left([h]_{Lip} + (2 - \delta_{p,2})[\Phi_i]_{Lip} + \frac{[f]_{Lip}}{1+[f]_{Lip}}\right)(1+[f]_{Lip})^{i-k}\|X_i - \widehat{X}_i\|_p - [f]_{Lip}\|X_k - \widehat{X}_k\|_p \\
&\leq \frac{1}{(1 + [f]_{Lip})^k}\sum_{i=k}^{n}(1+[f]_{Lip})^i\left([h]_{Lip} + \frac{[f]_{Lip}}{1+[f]_{Lip}} + (2 - \delta_{p,2})[\Phi_i]_{Lip}\right)\|X_i - \widehat{X}_i\|_p.
\end{aligned}$$

Finally, the definition of Φ_i and the Lipschitz property of $P(x, dy)$ imply that

$[\Phi_i]_{Lip} = [P(u_{i+1} + f_{i+1}(\,.\,, u_{i+1}))]_{Lip} \leq K(1 + [f]_{Lip})[u_{i+1}]_{Lip} + K[f]_{Lip}$, $1 \leq i \leq n-1$. $\diamond$

1.2.3 Approximation of the (lowest) optimal stopping time

The second quantity of interest in Optimal Stopping Theory is the (set of) optimal stopping time(s). A stopping time τ_{opt} is optimal for the (h, f)-Snell envelope if

$$U_0 = \mathbb{E}_0\left(h_{\tau_{opt}}(X_{\tau_{opt}}) + \sum_{i=1}^{\tau_{opt}} f_i(X_i, U_i)\right).$$

One knows (see, *e.g.*, [48]) that the lowest optimal stopping time is given by

$$\tau_* := \min\{k \,/\, U_k = h_k(X_k)\}.$$

In case of non-uniqueness of the optimal stopping times, τ_* plays a special role because it turns out to be the easiest to approximate. Thus when dealing with quantization of Markov chains, it is natural to introduce its counterpart for the quantized process that is

$$\widehat{\tau}_* := \min\left\{k \,/\, \widehat{u}_k(\widehat{X}_k) = h_k(\widehat{X}_k)\right\}.$$

In fact, the estimation of the error $\|\tau_* - \widehat{\tau}_*\|_p$ seems out of reach, essentially because it is quite difficult to bound these stopping times from below. Nevertheless, we will be able to approximate τ_* (in probability). Let $\delta > 0$. Set

$$\left\{\begin{array}{lll}\tau_\delta & := & \min\left\{k \,/\, u_k(X_k) \le h_k(X_k) + \delta\right\}\\ \widehat{\tau}_\delta & := & \min\left\{k \,/\, \widehat{u}_k(\widehat{X}_k) \le h_k(\widehat{X}_k) + \delta\right\}.\end{array}\right.$$

Proposition 5 (a) *For every $\delta > 0$, $\tau_\delta \ge \tau_*$, $\widehat{\tau}_\delta \ge \widehat{\tau}_*$. Furthermore*

$$\tau_\delta \downarrow \tau_* \qquad \textit{and} \qquad \widehat{\tau}_\delta \downarrow \widehat{\tau}_* \qquad \textit{as} \qquad \delta \downarrow 0.$$

(These stopping times being integer-valued, τ_δ and $\widehat{\tau}_\delta$ are eventually equal to their limit.)

(b) *For every $\delta > 0$, $\mathbb{P}\left(\widehat{\tau}_\delta \notin [\tau_{\frac{3\delta}{2}}, \tau_{\frac{\delta}{2}}]\right) \le \dfrac{1}{\delta}\displaystyle\sum_{k=0}^{n}(kd_k + [h]_{Lip})\|\widehat{X}_k - X_k\|_1$.*

Proof: (a) is an obvious corollary of the definitions of τ_δ and $\widehat{\tau}_\delta$.

(b) Set $Z_k := u_k(X_k) - \widehat{u}_k(\widehat{X}_k) + h_k(\widehat{X}_k) - h_k(X_k)$. Then, one may write

$$\widehat{\tau}_\delta = \min\left\{k \,/\, u_k(X_k) \le h_k(X_k) + \delta + Z_k\right\}$$

so that, on the event $\{\max_{0\le k\le n}|Z_k| \le \delta/2\}$, $\tau_{\frac{3\delta}{2}} \le \widehat{\tau}_\delta \le \tau_{\frac{\delta}{2}}$. Subsequently

$$\mathbb{P}\left(\widehat{\tau}_\delta \notin [\tau_{\frac{3\delta}{2}}, \tau_{\frac{\delta}{2}}]\right) \le \mathbb{P}\left(\max_{0\le k\le n}|Z_k| > \delta/2\right) \le \frac{2}{\delta}\,\mathbb{E}\max_{0\le k\le n}|Z_k|.$$

Now, using the notations of Theorem 1, one has

$$\begin{array}{rcl} |Z_k| & \le & |\widehat{U}_k - U_k| + [h]_{Lip}|\widehat{X}_k - X_k| \\ \text{and, subsequently}\quad \mathbb{E}\displaystyle\max_{0\le k\le n}|Z_k| & \le & \displaystyle\sum_{k=0}^{n}\|\widehat{U}_k - U_k\|_1 + [h]_{Lip}\|\widehat{X}_k - X_k\|_1 \\ & \le & \displaystyle\sum_{k=0}^{n}(kd_k + [h]_{Lip})\|\widehat{X}_k - X_k\|_1. \quad \diamond \end{array}$$

2 Optimization of the quantization

After a short background on optimal quantization of a $\mathbb{R}^d$-valued random vector, this section is devoted to the optimal quantization method of a Markov chain. For a modern and rigorous overview of quantization of random vectors, see [30] and the references therein.

2.1 Optimal quantization of a random vector X

Let $X \in L^p_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$. Following the terminology introduced in 1.2.1, the L^p-quantization ($p \ge 1$) consists in studying the best possible L^p-approximation of X by a random vector $\widehat{X} := q(X)$ where $q : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel function (*quantizer*) taking at most N values called *elementary quantizers*. Set $\Gamma := q(\mathbb{R}^d) := \{x_1, \ldots, x_N\}$ (*quantization grid*), $x_1, \cdots, x_N \in \mathbb{R}^d$. Minimizing the L^p-quantization error $\|X - q(X)\|_p$ consists in two phases:

Phase 1. A grid $\Gamma \subset (\mathbb{R}^d)^N$, $|\Gamma| \le N$ being settled, find a/the Γ-valued quantizer q_Γ that minimizes $\|X - q_\Gamma(X)\|_p$ among all Γ-valued quantizers q (if some).

PHASE 2. Find a grid Γ with size $|\Gamma| \le N$ which achieves the infimum of $\|X - q_{\Gamma}(X)\|_{p}$ among all the grids having at most N points (if some).

The solution to Phase 1 is provided by any *Voronoi quantizer* of the grid Γ, also called *projection following the closest neighbour rule* and defined by

$$q_{\Gamma} := \sum_{x_i \in \Gamma} x_i \mathbf{1}_{C_i(\Gamma)}$$

where $(C_i(\Gamma))_{1 \le i \le N}$ is a Borel partition of $\mathbb{R}^d$ called *Voronoi tessellation* satisfying

$$C_i(\Gamma) \subset \{\xi \in \mathbb{R}^d \,/\, |x_i - \xi| = \min_{1 \le j \le N} |\xi - x_j|\}.$$

A given grid of size $N \ge 2$ clearly has infinitely many Voronoi tessellations, essentially due to median hyperplanes. However all the $C_i(\Gamma)$ have the same *convex* closure and boundary, included in at most $N-1$ hyperplanes. If the distribution of X weights no hyperplane, the Voronoi tessellation is $\mathbb{P}$-essentially unique.

The Voronoi Γ-quantization, *denoted* $\widehat{X}^{\Gamma} := q_{\Gamma}(X)$, induces a *$L^p$-quantization error* $\|X - \widehat{X}^{\Gamma}\|_{p}$ (in Information Theory $\|X - \widehat{X}^{\Gamma}\|_{p}^{p}$ is called *L^p-distortion*) which reads

$$\|X - \widehat{X}^{\Gamma}\|_{p}^{p} = \sum_{x_i \in \Gamma} \mathbb{E}\left(\mathbf{1}_{C_i(\Gamma)} |X - x_i|^p\right) = \mathbb{E}\left(\min_{1 \le i \le N} |X - x_i|^p\right) = \int_{\mathbb{R}^d} \min_{1 \le i \le N} |\xi - x_i|^p \, \mathbb{P}_{X}(d\xi). \quad (28)$$

Notice that the quantization error only depends on the distribution of X whereas the Voronoi quantizer q_{Γ} only depends on Γ (and the Euclidean norm). Equality (28) will be the key for the numerical optimization of the grid. Finally, one easily shows that

$$\|X - \widehat{X}^{\Gamma}\|_{p} = \min\left\{\|X - Y\|_{p},\, Y : \Omega \to \Gamma,\, |Y(\omega)| \le N\right\}. \quad (29)$$

To carry out Phase 2 (grid optimization), one derives from (28) that the quantization error $\|X - \widehat{X}^{\Gamma}\|_{p}$ behaves as *symmetric Lipschitz continuous function* of the components of the grid $\Gamma := \{x_1, \ldots, x_N\}$ (with the temporary convention that some elementary quantizers x_i may be "stuck" so that $|\Gamma| \le N$). One shows (see, *e.g.*, [1, 50, 30]) that

$$\Gamma \mapsto \|X - \widehat{X}^{\Gamma}\|_{p},\ |\Gamma| \le N, \text{ always reaches a minimum}$$

at some grid Γ^* which takes its values in the convex hull of the support of $\mathbb{P}_{X}$. One proceeds by induction: if $N = 1$, the existence of a minimum is obvious by convexity; then, one may assume w.l.o.g. that $|X(\Omega)| \ge N$. Then, $\{\Gamma \,/\, \|X - \widehat{X}^{\Gamma}\|_{p} \le m_{N-1} - \varepsilon,\, |\Gamma| \le N\}$ with $m_{N-1} := \min\{\|X - \widehat{X}^{\Gamma}\|_{p},\, |\Gamma| \le N-1\} = \|X - \widehat{X}^{\Gamma^{*,N-1}}\|_{p}$ is a nonempty compact set for small enough $\varepsilon > 0$ since it contains $\Gamma^{*,N-1} \cup \{\xi\}$ for some appropriate $\xi \in X(\Omega) \setminus \Gamma^{*,N-1}$. This grants the existence of an optimal grid $\Gamma^{*,N}$. Then, following (29), $\widehat{X}^{\Gamma^{*,N}}$ is the best L^p-approximation of X over the random vectors taking at most N values *i.e.*

$$\|X - \widehat{X}^{\Gamma^{*,N}}\|_{p} = \min\left\{\|X - Y\|_{p},\, Y : \Omega \to \mathbb{R}^d,\, |Y(\omega)| \le N\right\} \quad (30)$$

INSERT FIGURE 1 AROUND HERE

Fig.1: Optimal L^2-quantization of the Normal distribution with a 500-tuple and its Voronoi tessellation

We will need the following properties (see [50, 30] and references therein) in the sequel:

P1. If $\mathbb{P}_{_X}$ has an *infinite* support, any "*N-optimal*" grid Γ^* has N pairwise distinct components, $\mathbb{P}_{_X}(\cup_{i=1}^N \partial C_i(\Gamma^*)) = 0$ and $N \mapsto \min\limits_{|\Gamma|\le N} \|X - \widehat{X}^\Gamma\|_{_p}$ is decreasing.

P2. If the support of $\mathbb{P}_{_X}$ is *everywhere dense in its convex hull* $\mathcal{H}_X$, any N-optimal grid lies in $\mathcal{H}_X$ and any locally N-optimal grid lying in $\mathcal{H}_X$ has exactly N distinct components.

P3. The minimal L^p-quantization error goes to 0 as $N \to \infty$ *i.e.*, $\lim\limits_N \min\limits_{|\Gamma|\le N} \|X - \widehat{X}^\Gamma\|_{_p} = 0$.

As a matter of fact, set $\Gamma_{_N} := \{z_1, \dots, z_{_N}\}$ where $(z_k)_{k\in\mathbb{N}}$ is everywhere dense in $\mathbb{R}^d$. Then, $\min_{|\Gamma|\le N} \|X - \widehat{X}^\Gamma\|_{_p} \le \|X - \widehat{X}^{\Gamma_N}\|_{_p}$ goes to 0 by the Lebesgue Dominated Convergence Theorem.

At which rate does this convergence to zero hold turns out to be a much more challenging question. The answer, often called Zador Theorem, was completed by several authors (Zador, see [57], Bucklew & Wise, see [16] and finally Graf & Luschgy, see [30]).

Theorem 2 *(Asymptotics) If* $\mathbb{E}|X|^{p+\eta} < +\infty$ *for some* $\eta > 0$*, then*

$$\lim_N \left(N^{\frac{p}{d}} \min_{|\Gamma|\le N} \|X - \widehat{X}^\Gamma\|_{_p}^p \right) = J_{p,d} \left(\int |g|^{d/(d+p)}(u)\, du \right)^{1+p/d}, \tag{31}$$

where $\mathbb{P}_{_X}(du) = g(u).\lambda_d(du) + \nu$, $\nu \perp \lambda_d$ *(*λ_d *Lebesgue measure on* $\mathbb{R}^d$*). The constant* $J_{p,d}$ *corresponds to the case of the uniform distribution on* $[0,1]^d$.

Little is known about the true value of the constant $J_{p,d}$ except in 1-dimension where $J_{p,1} = \frac{1}{2^p(p+1)}$ and in 2-dimension (*e.g.* $J_{2,2} = \frac{5}{18\sqrt{3}}$, see [29, 30]). Nevertheless some bounds are available, based on the introduction of random quantization grids (see [57, 19]). Thus, one has $J_{p,d} \sim (\frac{d}{2\pi e})^{\frac{p}{2}}$ as d goes to infinity (see [30]).

Theorem 2 says that $\min_{|\Gamma|\le N} \|X - \widehat{X}^\Gamma\|_{_p} = O(N^{-\frac{1}{d}})$: this means that optimal quantization of a distribution $\mathbb{P}_{_X}$ produces (for every grid size N) some grids with the same rate as that obtained with uniform lattice grids (when $N = m^d$) for $U([0,1]^d)$-distributions (in fact, even then, uniform lattice grids are never optimal as soon as $d \ge 2$).

An example of application (numerical integration): on one hand for every $p \ge 1$

$$\|X - \widehat{X}^\Gamma\|_{_p}^p = \max\left\{ \int_{\mathbb{R}^d} |\varphi - \varphi \circ q_{_\Gamma}|^p d\mathbb{P}_{_X},\ \varphi \text{ Lipschitz continuous, } [\varphi]_{_{Lip}} \le 1 \right\}$$

(the equality stands for the function $\varphi : \xi \mapsto \min\limits_{x_i\in\Gamma} |\xi - x_i|$). This induces a propagation of the L^p-quantization error by Lipschitz functions (already used in the proof of Theorem 1). On the other hand, from a numerical viewpoint,

$$|\mathbb{E}\,\varphi(\widehat{X}^\Gamma) - \mathbb{E}\,\varphi(X)| \le [\varphi]_{_{Lip}} \|X - \widehat{X}^\Gamma\|_1 \le [\varphi]_{_{Lip}} \|X - \widehat{X}^\Gamma\|_{_p}. \tag{32}$$

with
$$\mathbb{E}\,\varphi(\widehat{X}^\Gamma) = \int_{\mathbb{R}^d} \varphi\, dq_{_\Gamma}(\mathbb{P}_{_X}) = \sum_{i=1}^N \mathbb{P}(X \in C_i(\Gamma))\, \varphi(x_i) \tag{33}$$

(The parameter of interest is mainly $p = 2$, for algorithmic reasons.) The numerical computation of $\mathbb{E}\,\varphi(\widehat{X}^\Gamma)$ for any (known) function φ relies on

the grid $\Gamma = \{x_1, \dots, x_{_N}\}$ and its "Voronoi $\mathbb{P}_{_X}$-weights" $(\mathbb{P}_{_X}(C_i(\Gamma)))_{1\le i\le N}$.

whereas the error evaluation relies on $\|X - \widehat{X}^\Gamma\|_{_p}$. See [50] and [24] for further numerical integration formulae (Hölder, $\mathcal{C}^2$, locally Lipschitz continuous functions).

2.2 Optimal quantization: how to get it?

As far as numerical applications of optimal quantization of a random vector X are concerned, it has been emphasized above that we need an algorithm which produces,

- an optimal (or at least a sub-optimal) grid $\Gamma^* := \{x_1^*, \ldots, x_N^*\}$,
- the $\mathbb{P}_X$-mass of the its Voronoi tessellation $(\mathbb{P}_X(C_i(\Gamma^*)))_{1\le i\le N}$,
- the resulting L^p-quantization error $\|X - \widehat{X}^{\Gamma^*}\|_p$.

2.2.1 One-dimensional quadratic setting ($d = 1$, $p = 2$):

In 1-dimension, an algorithm, known as the *Lloyd's Method I*, appears as a by-product of the uniqueness problem for optimal grids (in fact *stationary* grids, see (36) below): for every grid Γ of size N, one sets

$$T(\Gamma) = \left(\int_{C_i(\Gamma)} \xi\, \mathbb{P}_X(d\xi)/\mathbb{P}_X(C_i(\Gamma))\right)_{1\le i\le N}.$$

The grid $T(\Gamma)$ has size N and satisfies $\|X - \widehat{X}^{T(\Gamma)}\|_2 \le \|X - \widehat{X}^{\Gamma}\|_2$. If $\mathbb{P}_X$ has a log-*concave density function*, then T is contracting (see [31, 46, 55]). Its unique fixed point Γ^* is clearly an optimal grid and the resulting (deterministic) iterative algorithm: $\Gamma^{t+1} = T(\Gamma^t)$ converges exponentially fast toward Γ^*.

2.2.2 Multi-dimensional setting ($d \ge 2$):

The above Lloyd's Method I formally extends to higher dimension. Since there are always several sub-optimal quantizers, T can no longer be contracting, although it still converges to some stationary quantizer (see (36) below), usually not optimal, even locally. Its major drawback is that it involves at each step several integrals $\int_{C_i(\Gamma)} \varphi\, d\,\mathbb{P}_X$ (another is that it only works for $p = 2$). This suggests to randomize the Lloyd's Method I by computing these integrals by Monte Carlo simulations of $\mathbb{P}_X$-distributed random vectors.

There is another randomized procedure that can be called upon to find critical point of a function *when its gradient admits an integral representation* with respect to a probability distribution: the *stochastic gradient descent* (stochastic counterpart of the deterministic gradient descent). Let us recall briefly what this procedure is.

Stochastic gradient descent (a standard convergence result): Let V be a differentiable *potential* function $V : \mathbb{R}^M \to \mathbb{R}$ such that $\lim_{|y|\to+\infty} V(y) = +\infty$, $|\nabla V|^2 = O(V)$, ∇V is Lipschitz continuous, $\{\nabla V = 0\}$ is locally finite, and ∇V has an integral representation $\nabla V(y) = \mathbb{E}\, \nabla_y\, v(y, X)$, where X is a $\mathbb{R}^L$-valued random vector. Let $(\gamma_t)_{t\ge1}$ be a sequence of positive gain parameter satisfying $\sum_t \gamma_t = +\infty$ and $\sum_t \gamma_t^2 < +\infty$. Classical Stochastic Approximation Theory says that the recursive algorithm

$$Y^{t+1} = Y^t - \gamma_{t+1} \nabla_y\, v(Y^t, \xi^{t+1}), \quad \xi^t \ i.i.d., \ \xi^1 \overset{\mathcal{L}}{\sim} X, \tag{34}$$

a.s. converges toward some critical point $y^* \in \{\nabla V = 0\}$ of V (see [20, 34] among others, for various results in that direction). Under some additional assumptions, one shows that y^* is necessary a local minimum (see [54, 36, 12]).

Let us come back to our optimal quantization problem. We already noticed that Equation (28) defines a symmetric continuous function on $(\mathbb{R}^d)^N$, namely

$$\begin{aligned} D_N^{X,p}(x_1,\dots,x_N) &:= \|X-\widehat{X}^{\{x_1,\dots,x_N\}}\|_p^p = \mathbb{E}(d_N^{X,p}(x,X)) \quad (D \text{ for } D\text{istortion}),\\ \text{with } d_N^{X,p}(x_1,\dots,x_N,\xi) &:= \min_{1\le i\le N}|\xi-x_i|^p,\ (x_1,\dots,x_N)\in(\mathbb{R}^d)^N,\ \xi\in\mathbb{R}^d,\ \text{(local distortion)}. \end{aligned}$$

One shows (see [30] when $p=2$ or [50] otherwise) that for every $p>1$, $D_N^{X,p}$ is continuously differentiable at every N-tuple $x:=(x_1,\dots,x_N)\in(\mathbb{R}^d)^N$ such that $x_i\neq x_j,\ i\neq j$ and $\mathbb{P}_X(\cup_{i=1}^N\partial C_i(x))=0$. The gradient $\nabla D_N^{X,p}(x)$ is given by a formal differentiation, *i.e.*

$$\begin{aligned} \nabla D_N^{X,p}(x) &:= \mathbb{E}(\nabla_x d_N^{X,p}(x,X)) = \left(\int_{\mathbb{R}^d}\frac{\partial d_N^{X,p}(x,\xi)}{\partial x_i}\mathbb{P}_X(d\xi)\right)_{1\le i\le N} \qquad (35)\\ \text{with } \frac{\partial d_N^{X,p}}{\partial x_i}(x,\xi) &:= p\mathbf{1}_{C_i(x)}|x_i-\xi|^{p-1}\frac{x_i-\xi}{|x_i-\xi|},\quad 1\le i\le N,\quad (\text{set } \frac{\vec{0}}{\|\vec{0}\|}:=\vec{0}). \end{aligned}$$

Equality (35) still holds when $p=1$ if $\mathbb{P}_X$ is continuous. A grid $\{x_1,\dots,x_N\}$ (with N pairwise distinct components and $\mathbb{P}_X$-negligible Voronoi tessel boundaries) is

$$\textit{a stationary grid} \quad \text{if} \quad \nabla D_N^{X,p}(x_1,\dots,x_N)=0. \qquad (36)$$

Then, following **P1**, any optimal grid is stationary.

Setting $V:=D_N^{X,p}$ and plugging the above formula for $\nabla_x d_N^{X,p}$ in the abstract stochastic gradient procedure (34) yields (back with the grid notation $\Gamma^t=\{X_1^t,\dots,X_N^t\}$)

$$\Gamma^{t+1}=\Gamma^t-\frac{\gamma_{t+1}}{p}\nabla_x d_N^{X,p}(\Gamma^t,\xi^{t+1}),\quad \Gamma^0\subset\mathbb{R}^d,\ |\Gamma^0|=N,\quad \xi^t\ \textit{i.i.d.},\ \xi^1\sim\mathbb{P}_X \qquad (37)$$

where *the step sequence* $(\gamma_t)_{t\ge1}$ *is* $(0,1)$*-valued* and satisfies as usual $\sum_t\gamma_t=+\infty$ and $\sum_t\gamma_t^2<+\infty$. The fact that $\gamma_t\in(0,1)$ for every $t\in\mathbb{N}$ ensures that $|\Gamma^t|=N$.

Unfortunately, the assumptions that make the stochastic gradient descent *a.s.* converge are never satisfied by the L^p-distortion function $D_N^{X,p}$ which is not a true potential function for at least two reasons,

– the gradient $\nabla D_N^{X,p}$ does not exist at N-tuples of $(\mathbb{R}^d)^N$ having stuck components (although locally bounded). So, it cannot be Lipschitz or even Hölder continuous.

– the L^p-distortion $D_N^{X,p}(x)$ does not go to infinity as $|x|:=|x_1|+\cdots+|x_N|\to+\infty$ (but only if $\min_{1\le i\le N}|x_i|\to+\infty$).

However $D_N^{X,p}$ turns out to be a fairly good potential for practical implementation, especially in the quadratic case $p=2$ (see Fig.3 and the *CLVQ* algorithm below). One does not observe on simulations some components of Γ^t getting asymptotically stuck in (37) when $t\to+\infty$ in spite of the structure of the potential function. Although many stationary grids exist, it does converges toward a grid Γ^*, seemingly close to optimality.

This quadratic case corresponds to a very commonly implemented procedure in Automatic Classification called Competitive Learning Vector Quantization (CLVQ), also known as the Kohonen algorithm with 0 neighbour. Our specificity is that we need a great numerical accuracy far beyond qualitative features. This requires to run significantly more iterates of the procedure.

The *CLVQ* algorithm (quadratic case): The recursive procedure (37) can be developed as follows: set $\Gamma^t := \{X_1^t, \ldots, X_N^t\}$.

$$\text{Competitive phase:} \quad \text{select } i(t+1) \in \operatorname{argmin}_i |X_i^t - \xi^{t+1}| \text{ (closest neighbour)} \tag{38}$$

$$\text{Learning phase:} \quad \left\{ \begin{array}{lcl} X_{i(t+1)}^{t+1} & := & X_{i(t+1)}^t - \gamma_{t+1}(X_{i(t+1)}^t - \xi^{t+1}) \\ X_i^{t+1} & := & X_i^t, \quad i \neq i(t+1). \end{array} \right. \tag{39}$$

• *Practical aspects:* Concerning numerical implementations of the algorithm, notice that, at each step the grid Γ^{t+1} lives in the convex hull of Γ^t and ξ^{t+1} since the learning phase (39) is simply a homothety centered at ξ^{t+1} with ratio $1-\gamma_{t+1} > 0$ (see Fig. 2 below). This has a stabilizing effect on the procedure which explains why one verifies on simulations that the *CLVQ* algorithm does behave better than its non-quadratic counterparts. Finally, one often "refines" the *CLVQ* by processing a randomized Lloyd's I method (see [51]).

INSERT FIGURE 2 AROUND HERE

Fig.2: One step of the CLVQ algorithm (on the real line)

Figure 3 below shows some planar quantizations obtained using the *CLVQ* algorithm (see [51] for some simulations in the quadratic case).

INSERT FIGURE 3 AROUND HERE

Fig.3: $\mathbb{P}_X = U([0,1]^2)$, $\mathbb{P}_X := \mathcal{E}(1/2)^{\otimes 2}$, $\mathbb{P}_X := \mathcal{N}\left(0; \begin{bmatrix} 1 & \sqrt{2}/2 \\ \sqrt{2}/2 & 1 \end{bmatrix}\right)$.

The main drawback of the *CLVQ* algorithm is to be slow: roughly speaking, like usual recursive stochastic algorithms, its rate of convergence is ruled by a CLT at rate $\sqrt{\gamma_t}$ (cf.[20]). Moreover, at each iterate, the computation of the winning index i_0 in the cooperation phase is time consuming if N is large. Speeding up the algorithm needs to handle both aspects. First, one may use some deterministic sequences with low discrepancy instead of pseudo-random numbers to implement the algorithm (see, *e.g.* [43]) or call upon some averaging methods which reduce the variance in the CLT Theorem (see [53] and the references therein). To cut down the winning index search time, one may implement some fast (approximate) search procedures. For recent developments in that direction one may consult [28] (chapters 10.4, p.332 and 12.16, p.479).

The last practical question of interest is the choice of the starting grid Γ^0. This question is clearly connected with the existence of several local minima for the distortion. One way is to start from a random N-grid obtained by simulation. An alternative is to process a so-called *splitting method*: one adds progressively some (optimal) quantizers with smaller size to the current state grid: the heuristic idea is that if a N-grid Γ_N^* (almost) achieves the the minimal distortion, then the $N+\nu$-grid ($\nu \ll N$) $\Gamma_N^* \cup \Gamma_\nu^*$ is likely to be inside the attracting basin of the absolute minimum of $N+\nu$-distortion. These computational aspects are developed in [51] in the important framework of Gaussian distributions.

Finally, one can polish up the converging phase of the grid produced by the *CLVQ* stochastic gradient by processing a randomized Lloyd's Method 1 procedure.

• *Theoretical aspects:* some rigorous *a.s.* converging results have been established in [50], only for compactly supported distributions $\mathbb{P}_X$ on $\mathbb{R}^d$: when $d \geq 2$ this convergence only holds in the "Kushner & Clark" (or conditional) sense whereas standard *a.s.* convergence toward a true N-grid holds in 1-dimension (see Appendix 2 for details).

As a conclusion, the $CLVQ$ and its L^p-counterparts compress the information on $\mathbb{P}_X$ provided by the sequence $(\xi^t)_{t\in\mathbb{N}^*}$: it appears as a *compressed Monte Carlo Methods.*

Estimation of the companion parameters: The proposition below shows that the weight vector $(\mathbb{P}_X(C_i(\Gamma^*)))_{1\le i\le N}$ and the induced L^p-quantization error $\|X-\widehat{X}^{\Gamma^*}\|_p$ can be obtained *on line* for free as a by-product of the $CLVQ$ stochastic gradient descent as soon as Γ^t converges to Γ^* (this holds true for any $p\ge 1$).

Proposition 6 *Let $p\ge 1$. Assume that $X\in L^{p+\eta}$, $\eta>0$, and $\mathbb{P}_X$ weights no hyperplane. Set for every $t\ge 1$,*

- $p_i^t := \dfrac{|\{1\le s\le t\,/\,\xi^s\in C_i(\Gamma^{s-1})\}|}{t}$, *the empirical frequency of the ξ^s's falling in the i^{th} tessel $C_i(\Gamma^{s-1})$ of the (moving) Voronoi tessellation of Γ^{s-1} up to time t ($i=1,\ldots,N$).*
- $D^{\beta,t} := \dfrac{1}{t}\displaystyle\sum_{s=1}^{t}\min_{1\le i\le N}|\xi^s-X_i^{s-1}|^\beta$, $\beta\in(0,p+\eta)$, *the average of the lowest distance of ξ^s (to the power β) to the moving quantization grid $\Gamma^{s-1}=\{X_1^{s-1},\ldots,X_N^{s-1}\}$, $1\le s\le t$.*

Let Γ^* be a (stationary) grid (see (36)) and let $A_{\Gamma^*}:=\{\Gamma^t\to\Gamma^*\}$ be the set of convergence of Γ^t toward Γ^*. Then, on the event A_{Γ^*},

$$\begin{cases}\forall\, i\in\{1,\cdots,n\}, & p_i^t\xrightarrow{a.s.} p_i:=\mathbb{P}_X(C_i(\Gamma^*)) \quad as \quad t\to+\infty,\\ \forall\,\beta\in(0,p), & D^{\beta,t}\xrightarrow{a.s.} D_N^{X,\beta}(\Gamma^*) \quad as \quad t\to+\infty.\end{cases}$$

Remark: The above quantities p^t and $D^{\beta,t}$ obviously satisfy the recursive procedures

$$p_i^t=p_i^{t-1}-\frac{1}{t}\Big(p_i^{t-1}-\mathbf{1}_{\{\xi^t\in C_i(X^{t-1})\}}\Big)\,,\; D^{\beta,t}=D^{\beta,t-1}-\frac{1}{t}\Big(D^{\beta,t-1}-\min_{1\le i\le N}|X_i^{t-1}-\xi^t|^\beta\Big)\,. \quad (40)$$

In fact, the conclusion of Proposition 6 still holds true if $(\frac{1}{t})_{t\ge1}$ is replaced in (40) by any positive sequence $(\widetilde{\gamma}_t)_{t\ge1}$ satisfying $\sum_t\widetilde{\gamma}_t=+\infty$ and $\sum_t\widetilde{\gamma}_t^{(p+\varepsilon)/\beta}<+\infty$. Natural choices for $\widetilde{\gamma}_t$ are $\frac{1}{t}$ or the original step γ_t used in the $CLVQ$ algorithm (39), depending on the range of the simulation (see [51]).

Proof: For notational convenience, a generic N-tuple $x=(x_1,\ldots,x_N)\in(\mathbb{R}^d)^N$ will be denoted by its grid notation $\Gamma=\{x_1,\ldots,x_N\}$. Let $\Phi:(\mathbb{R}^d)^N\times\mathbb{R}^d\to\mathbb{R}$ be a Borel function satisfying

- $|\Phi(\Gamma,\xi)|\le M\,|\xi|^\beta$ for some real constants $M>0$ and $0\le\beta<p+\eta$,
- the function defined by $\varphi(\Gamma):=\displaystyle\int_{\mathbb{R}^d}\Phi(\Gamma,\xi)\,\mathbb{P}_X(d\xi)$ is bounded and continuous at Γ^*.

One checks that $(\Phi(\Gamma^t,\xi^{t+1})-\varphi(\Gamma^t))_{t\ge0}$ is a $L^{\frac{p+\eta}{\beta}}$-bounded sequence of martingale increments. Now the Chow Theorem (see *e.g.* [20], p.22) and $\sum_{t\ge1}t^{-\frac{p+\eta}{\beta}}<+\infty$, imply that the martingale $\sum_{1\le s\le t}\frac{\Phi(\Gamma^{s-1},\xi^s)-\varphi(\Gamma^{s-1})}{s}$ *a.s.* converges toward a finite random variable Z_∞. In turn, the Kronecker Lemma finally implies that $\frac{1}{t}\sum_{s=1}^t\Phi(\Gamma^{s-1},\xi^s)-\varphi(\Gamma^{s-1})\xrightarrow{a.s.}0$. Finally, the continuity of φ at Γ^* yields $\frac{1}{t}\sum_{s=1}^t\Phi(\Gamma^{s-1},\xi^s)\xrightarrow{a.s.}\varphi(\Gamma^*)$ on A_{Γ^*}.

One first applies this result to the indicator functions $\Phi_i(\Gamma,\xi):=\mathbf{1}_{C_i(\Gamma)}(\xi)$, $1\le i\le n$; the associated φ functions are continuous at Γ^* because $\mathbb{P}_X$ weights no hyperplane.

Then, set $\Phi(\Gamma,\xi) := \rho(\Gamma) \min_{1\le i\le N} \|\xi - x_i\|^\beta$ where ρ is a continuous $[0,1]$-valued function with compact support on $(\mathbb{R}^d)^N$ satisfying $\rho(\Gamma^*)=1$. $\varphi(\Gamma) := \rho(\Gamma) D_N^{X,\beta}(\Gamma)$ is continuous and, on A_{Γ^*}, $\Phi(\Gamma^s, \xi^{s+1}) = \min_{1\le i\le N} \|\xi^{s+1} - X_i^s\|^\beta$ for large enough s. $\diamondsuit$

2.3 Optimal quantization of a simulatable Markov chain

Let us pass now to the optimal quantization of a $\mathbb{R}^d$-valued Markov chain $(X_k)_{0\le k\le n}$ with transition $P(x,dy)$ and initial distribution $\mu_0 = \mathcal{L}(X_0)$. Assume that for every $x\in\mathbb{R}^d$, the distributions $P(x,dy)$ can be easily simulated on a computer as well as μ_0. A typical examples is the Gaussian Euler scheme of a diffusion (see subsection 1.1). Assume that,for every $k=0,\ldots,n$, $X_k\in L^{p+\eta}$ $(\eta>0)$, and that

$$\text{the distribution } of X_k \text{ weights no hyperplane in } \mathbb{R}^d. \tag{41}$$

2.3.1 The extended $CLVQ$ algorithms for Markov chain optimal quantization

The principle is to modify the Monte Carlo simulation shortly outlined in subsection 1.2.1 by processing a $CLVQ$ algorithm at each step k. One starts from a large scale Monte Carlo simulation of the Markov chain $(X_k)_{0\le k\le n}$ that we will denote by $\xi^0 := (\xi_0^0,\ldots,\xi_n^0)$, $\xi^1 := (\xi_0^1,\ldots,\xi_n^1),\ldots$, $\xi^t := (\xi_0^t,\ldots,\xi_n^t),\ldots$ Our aim is to produce for every $k\in\{0,\ldots,n\}$ some optimal grids $\Gamma_k := \{x_1^k,\ldots,x_{N_k}^k\}$ with size N_k, their transition kernels $[\pi_{ij}^k]$ and their L^p-quantization errors. Note that, if one sets

$$p_i^k := \mathbb{P}\left(X_k\in C_i(\Gamma_k)\right) \qquad \text{and} \qquad p_{ij}^k := \mathbb{P}\left(\{X_{k+1}\in C_j(\Gamma_{k+1})\}\cap\{X_k\in C_i(\Gamma_k)\}\right)$$

then $\pi_{ij}^k := \mathbb{P}\left(\widehat{X}_{k+1} = x_j^{k+1} / \widehat{X}_k = x_i^k\right) = \frac{p_{ij}^k}{p_i^k}$, $i=1,\ldots,N_k$, $j=1,\ldots,N_{k+1}$, $k=0,\ldots,n{-}1$.

$\boxed{p=2}$ In the quadratic case, the *extended CLVQ* algorithm reads as follows

1. Initialization phase:

• Initialize the $n+1$ starting grids $\Gamma_k^0 := \{x_1^{0,k},\ldots,x_{N_k}^{0,k}\}$, $k=0,\ldots,n$, of the $n+1$ $CLVQ$ algorithms that will quantize the distributions μ_k.

• Initialize the "marginal counter" vectors $\alpha_i^{k,0} := 0$, $i=1,\ldots,N_k$ for every $k=0,\ldots,n$.

• Initialize the "transition counters" $\beta_{ij}^{k,0} := 0$, $i=1,\ldots,N_k$, $j=1,\ldots,N_{k+1}$, $k=0,\ldots,n-1$.

2. Updating $t \rightsquigarrow t+1$: At step t, the $n+1$ grids Γ_k^t, $k=0,\ldots,n$, have been obtained. We use now the sample ξ^{t+1} to carry on the optimization process *i.e.* building up the grids Γ_k^{t+1}'s as follows. For every $k=0$ up to n:

• Simulation of ξ_k^{t+1} (using ξ_{k-1}^{t+1} if $k\ge 1$).

• Selection of the "winner" in the k^{th} $CLVQ$ algorithm *i.e.* the only index $i_k^{t+1}\in\{1,\ldots,N_k\}$ satisfying

$$\xi_k^{t+1}\in C_{i_k^{t+1}}(\Gamma_k^t).$$

• Updating of the k^{th} $CLVQ$ algorithm:

$$\forall\, i\in\{1,\ldots,N_k\},\quad \Gamma_{k,i}^{t+1} = \Gamma_{k,i}^t - \gamma_{t+1}\mathbf{1}_{\{i=i_k^{t+1}\}}(\Gamma_{k,i}^t - \xi_k^{t+1}).$$

- Updating of the k^{th} marginal counter vector $\alpha^{k,t} := (\alpha_i^{k,t})_{1\le i\le N_k}$:

$$\forall\, i\in\{1,\dots,N_k\},\quad \alpha_i^{k,t+1} := \alpha_i^{k,t} + \mathbf{1}_{\{i=i_k^{t+1}\}}.$$

- Updating of the (quadratic) distortion estimator $D^{k,t}$:

$$D^{k,t+1} := D^{k,t} - \frac{1}{t+1}(|\Gamma^t_{k,i_k^{t+1}} - \xi^{t+1}|^2 - D^{k,t}).$$

- Updating of the transition counters $\beta^{k,t} := (\beta_{ij}^{k,t})_{1\le i\le N_{k-1},1\le j\le N_k}$ $(k\ge 1)$

$$\forall\, i\in\{1,\dots,N_{k-1}\},\, \forall\, j\in\{1,\dots,N_k\},\ \beta_{ij}^{k-1,t+1} := \beta_{ij}^{k-1,t} + \mathbf{1}_{\{i=i_{k-1}^{t+1},\, j=i_k^{t+1}\}}.$$

- Updating the transition kernels $(\pi_{ij}^{k,t})_{1\le i\le N_{k-1},1\le j\le N_k}$ $(k\ge 1)$

$$\pi_{ij}^{k,t+1} := \frac{\beta_{ij}^{k,t+1}}{\alpha_i^{k,t+1}} \qquad \text{(possibly only once at the end of the simulation process!).}$$

Following Proposition 6 one has, on the event $\left\{\Gamma_k^t \to \Gamma_k^*\right\}$, $D^{k,t} \overset{t\to+\infty}{\to} D_{N_k}^{X_k,2}(\Gamma_k^*)$ and

$$\frac{\alpha^{k,t}}{t} \overset{t\to+\infty}{\longrightarrow} (p_i^{*,k})_{1\le i\le N_k} = (\mathbb{P}(X_k\in C_i(\Gamma_k^*)))_{1\le i\le N_k} \qquad \text{(thanks to (41)).}$$

The same martingale approach shows that, on the event $\{\Gamma_k^t \longrightarrow \Gamma_k^*\}\cap\left\{\Gamma_{k+1}^t \longrightarrow \Gamma_{k+1}^*\right\}$,

$$\frac{\beta^{k,t}}{t} \overset{t\to+\infty}{\longrightarrow} p_{ij}^{*,k} = (\mathbb{P}(X_k\in C_i(\Gamma_k^*),\ X_{k+1}\in C_j(\Gamma_{k+1}^*)))_{1\le i\le N_k,1\le j\le N_{k+1}}$$

so that, on the event $\left\{\Gamma_k^t \overset{t\to+\infty}{\longrightarrow} \Gamma_k^*,\ k=0,\dots,n\right\}$, for every $k\in\{1,\dots,n\}$,

$$\pi_{ij}^{k,t} \overset{t\to+\infty}{\longrightarrow} \pi_{ij}^{*,k},\quad 1\le i\le N_k,\ 1\le j\le N_{k+1}.$$

The main features of this algorithm are essentially those of the original $CLVQ$ algorithm. Moreover, note that the forward optimization of the grids and the weight computation are not recursive in k, so there is no deterioration of the optimization process as k increases.

$\boxed{p\neq 2}$ On designs similarly an *L^p-optimization procedure (extended L^p-CLVQ)* by using the general L^p-formula (37) in the grid updating phase.

2.3.2 *A priori* optimal dispatching of optimal quantizer sizes

The above grid optimization procedures for Markov chains (extended $CLVQ$ and its L^p-variants) produces optimal grids Γ_k for a given dispatching of their sizes N_k. This raises the following optimization problem:

How to dispatch *a priori* the sizes $N_0,\dots,N_n$ of the quantization grids, assumed to be L^p-optimal, if one wishes to use at most $N\ge N_0+\dots+N_n$ elementary quantizers?

Let $p\ge 1$. Formula (27) in Theorem 1 provides some positive real coefficients $d_0,\ d_1,\dots,d_n$ such that, for *any sequence of quantizations* $(\widehat{X}_k)_{0\le k\le n}$,

$$\|u_0(X_0)-\widehat{u}_0(\widehat{X}_0)\|_p \le \sum_{i=0}^n d_i\,\|X_i-\widehat{X}_i\|_p. \tag{42}$$

Then, the best we can do is to specify *the sizes N_k's that minimize the right hand of (42) when all the quantization vectors $\widehat{X}_k$'s are L^p-optimal i.e.*

$$\min_{(N_0+\cdots+N_n\le N)}\sum_{i=0}^n d_i\times\|X_i-\widehat{X}_i\|_p\quad\text{with } \|X_i-\widehat{X}_i\|_p=\min_{|\Gamma|\le N_i}\|X_i-\widehat{X}_i^\Gamma\|_p,\ 0\le i\le n.$$

This will produce as well an asymptotic bound for the resulting L^p-error $\|\widehat{u}_0(\widehat{X}_0)-u_0(X_0)\|_p$. The key is Theorem 2 which says that $N_k^{1/d}\min_{|\Gamma|\le N_k}\|X_k-\widehat{X}_k^\Gamma\|_p$ converges to some positive constant as $N_k\to+\infty$.

Proposition 7 *Assume that all the distributions $\mathcal{L}(X_k)$ have an absolutely continuous part φ_k, $0\le k\le n$. Let $N\in\mathbb{N}^*$. Set for every $i\in\{0,\ldots,n\}$,*

$$N_i:=\left[\underline{\rho}_iN\right]\quad\textit{and}\quad \underline{\rho}_i:=\frac{a_i}{\sum_{0\le j\le n}a_j}\quad\textit{with}\quad a_i:=\left(\|\varphi_i\|_{\frac{d}{d+p}}^{\frac{1}{p}}d_i\right)^{\frac{d}{d+1}},\ 0\le i\le n. \tag{43}$$

Assume that all the quantizations $\widehat{X}_k$ of the X_k's are L^p-optimal with size N_k. Then,

$$\overline{\lim_N}\,N^{\frac{1}{d}}\max_{0\le i\le n}\|u_i(X_i)-\widehat{u}_i^{(N_i)}(\widehat{X}_i)\|_p\le J_{p,d}^{\frac{1}{p}}\left(\sum_{i=0}^n a_i\right)^{1+\frac{1}{d}} \tag{44}$$

where $J_{p,d}$ is defined in Theorem 2.

The easy lemma below solves the "continuous bit allocation" problem.

Lemma 1 *Let $a_0,\ldots,a_n$ be some positive real numbers. Then the function $\rho:=(\rho_0,\ldots,\rho_n)\mapsto\sum_{i=0}^n a_i\rho_i^{-\frac{1}{d}}$ defined on the set $\{\rho_0+\ldots+\rho_n=1,\ \rho_i\ge0,\ 0\le i\le n\}$ reaches its minimum*

$$\textit{at}\qquad \underline{\rho}:=\left(\frac{a_i^{\frac{d}{d+1}}}{a_0^{\frac{d}{d+1}}+\cdots+a_n^{\frac{d}{d+1}}}\right)_{0\le i\le n}\qquad\textit{so that}\qquad \sum_{i=0}^n a_i\underline{\rho}_i^{-\frac{1}{d}}=(\sum_{i=0}^n a_i^{\frac{d}{d+1}})^{1+\frac{1}{d}}.$$

Note that this minimum value is nondecreasing as a function of the a_i's and that

$$(\sum_{j=0}^n a_j^{\frac{d}{d+1}})^{1+1/d}\le(n+1)^{\frac{1}{d}}\sum_{i=0}^n a_i.$$

Proof of Proposition 7: First, one rewrites (42) as follows

$$N^{\frac{1}{d}}\|u_0(X_0)-\widehat{u}_0^{(N_0)}(\widehat{X}_0)\|_p\le\sum_{i=0}^n d_i\left(\frac{N_i}{N}\right)^{-\frac{1}{d}}\left(N_i^{\frac{1}{d}}\|X_i-\widehat{X}_i\|_p\right)$$

keeping in mind that every $\widehat{X}_k$ is a L^p-optimal with size N_k. Then,

$$\begin{aligned}\overline{\lim_N}\,N^{\frac{1}{d}}\max_{0\le i\le n}\|u_i(X_i)-\widehat{u}_i^{(N_i)}(\widehat{X}_i)\|_p &\le \sum_{i=0}^n d_i\,\overline{\lim_N}\left\{\left(\frac{N}{N_i}\right)^{\frac{1}{d}}\left(N_i^{\frac{1}{d}}\|X_i-\widehat{X}_i\|_p\right)\right\}\\ &= \sum_{i=0}^n d_i\,\underline{\rho}_i^{-\frac{1}{d}}\,\overline{\lim_N}\left(N_i^{\frac{1}{d}}\|X_i-\widehat{X}_i\|_p\right).\end{aligned}$$

Now, as $N_i \to +\infty$ for every $i \in \{0, \ldots, n\}$, Theorem 2 (quantization error asymptotics) implies that $\overline{\lim_N} N_i^{\frac{1}{d}} \|X_i - \widehat{X}_i\|_p = (J_{p,d} \|\varphi_i\|_{\frac{d}{d+p}})^{\frac{1}{p}}$. One concludes by

$$\overline{\lim_N} N^{\frac{1}{d}} \max_{0 \le i \le n} \|u_i(X_i) - \widehat{u}_i^{(N_i)}(\widehat{X}_i)\|_p \le J_{p,d}^{\frac{1}{p}} \sum_{i=0}^{n} d_i \, \|\varphi_i\|_{\frac{d}{d+p}}^{\frac{1}{p}} \rho_i^{-\frac{1}{d}} \le J_{p,d}^{\frac{1}{p}} \left(\sum_{i=0}^{n} \|\varphi_i\|_{\frac{d}{d+p}}^{\frac{d}{p(d+1)}} d_i^{\frac{d}{d+1}} \right)^{1+1/d} . \ \diamond$$

Remark: The above asymptotic bound (44) is not fully satisfactory: if one wishes to optimize the number n of time steps as a function of the number N of elementary quantizers, one needs some *a priori* error bounds for a given couple (n, N). To this end we will need to control the distributions of the X_k's, namely to "dominate" them by a fixed distribution up to some affine time scaling. This can be done *e.g.* with some uniformly elliptic diffusions $(X_{t_k})_{0 \le k \le n}$, see subsection 5.1, or with their Doléans exponentials (see [5]).

3 Weight estimation in the quantization tree: the statistical error

In this short section are summed up some results developed in an accompanying paper [4]. First, although the question of the rate of convergence of the quantization grid optimization and the companion parameter estimation is natural, one must keep in mind that this is *a one shot phase of the process.*

We will not discuss rigorously the error induced on the coefficients π_{ij}^k's by the on-line estimation procedure (40): it would lead us too far, given the very partial results available on the rate of convergence of the $CLVQ$ algorithm. However, let us mention that, under some appropriate assumptions (see, *e.g.*, [20], p.52), one shows that a recursive stochastic algorithm like (34) with step γ_t satisfies a Central Limit Theorem (CLT) with rate $1/\sqrt{\gamma_t}$ as $t \to +\infty$. Thus, if $\gamma_t \sim \gamma/t$ (with γ large enough), a "standard" CLT holds like for the regular Monte Carlo method (see [51] for more details concerning the rate of convergence of the $CLVQ$ algorithm in 1-dimension).

A less ambitious but still challenging problem is the following: consider some fixed grids $\Gamma_0, \ldots, \Gamma_n$, their companion parameters and the related quantization tree algorithm (23). What is the error induced by the use of Monte Carlo estimated weights $\widetilde{\pi}_{ij}^k$ instead of the true weights π_{ij}^k. This third type of error – the *statistical error* – is extensively investigated in the accompanying paper [4] but let us sum up the situation (see also [3] for a TCL).

This error highly depends on the structure of the nonlinearity. In the linear case (no reflection: $h(t, .) := -\infty$ if $t \ne T$, and $f :\equiv 0$), the composition of the empirical frequency matrices $(\widetilde{\pi}_{ij}^k)_{0 \le k \le n}$ yields $\widetilde{\alpha}_i^n = \frac{1}{M} \sum_{\ell=1}^{M} \mathbf{1}_{\{\widehat{X}_n^\ell = x_i^n\}}$ so that $\widetilde{U}_n = \frac{1}{M} \sum_{\ell=1}^{M} h(T, \widehat{X}_n^\ell)$. This is but a standard Monte Carlo method for computing $\mathbb{E}\, h(T, X_T)$, ruled by the regular CLT: the statistical error is $O(\frac{1}{\sqrt{M}})$. In the nonlinear case, the empirical frequencies cannot be composed. In [4], we focus on the regular Snell envelope of $(h(X_k))_{0 \le k \le n}$ where h is a *bounded* Lipschitz continuous function ($f \equiv 0$) and X_k is either for the diffusion $X_{kT/n}$ (with Lipschitz continuous coefficients) or for its Euler scheme $\overline{X}_k$, $k = 0, \ldots, n$. Then, the statistical error depends on the regularity of the obstacle h as follows

$$\mathbb{E}|\widetilde{u}_0(x_0) - \widehat{u}_0(x_0)| \le C_{b,\sigma,h,T} \frac{\sqrt{nN} \sum_{k=1}^{n} \|X_k - \widehat{X}_k\|_2 + \rho_{n,N,M}}{\sqrt{M}}, \quad C_{b,\sigma,h,T} > 0, \tag{45}$$

with $\rho_{n,N,M} = \sqrt{n} + N^2/\sqrt{M}$ if h is semi-convex and X_k is the diffusion $X_{kT/n}$, and $\rho_{n,N,M} = n^{\frac{3}{4}} + N^2/\sqrt{nM}$ otherwise. Optimality of the quantizations $\widehat{X}_k$'s is not required.

4 Finite element method *vs* quantization: a first comparison

A quick comparison of the finite element method on one hand and of the quantization method on the other hand shows that there is a strong analogy between them: in both cases one computes the approximation of the solution at a finite number of points using a weighted sum, starting from a final condition still evaluated at a finite number of points. The aim of this short section is to get a slightly deeper insight of this analogy. In order to compare the two methods we will use the simplest possible example, the heat equation

$$(E) \quad \equiv \quad (\partial_t + \frac{1}{2}\Delta)u = 0, \quad u_T = f. \tag{46}$$

Let $<\varphi|\psi> := \int_{\mathbb{R}^d} \varphi(x)\psi(x)dx$ denotes the inner product of $L^2(\mathbb{R}^d, \lambda_d)$. The weak form of (E) is given by

$$<f|\varphi> - <u_t|\varphi> - \frac{1}{2}\int_t^T \sum_{i=1}^d < \frac{\partial u_s}{\partial x^i} | \frac{\partial \varphi}{\partial x^i} > ds = 0$$

for every $t \in [0,T]$ and every test function $\phi \in \mathcal{C}_c^2(\mathbb{R}^d, \lambda_d)$. The first step of both approaches is the time discretization. Let $n \in \mathbb{N}^*$ and $t_k = kh$, $h := T/n$ (we temporarily give up the notation Δ for the step because of the Laplacian). Then, we consider the discrete problem

$$<f|\phi> - <u_{t_{k_0}}|\phi> - \frac{h}{2}\sum_{k=k_0}^n \sum_{i=1}^d < \frac{\partial u_{t_k}}{\partial x^i} | \frac{\partial \phi}{\partial x^i} >= 0.$$

We use the dynamic programming principle in order to solve this problem by induction: we put $\overline{u}_n := f$ and we define $\overline{u}_k$ to be the solution of the elliptic problem

$$(\overline{E}_k) \quad \equiv \quad <\overline{u}_{k+1}|\phi> - <\overline{u}_k, \phi> - \frac{h}{2}\sum_{i=1}^d < \frac{\partial \overline{u}_k}{\partial x^i} | \frac{\partial \phi}{\partial x^i} >= 0.$$

Equation $(\overline{E}_k)$ can be seen either as a Dirichlet problem on the whole $\mathbb{R}^d$ with condition zero at infinity or as a restricted problem to a large enough ball. This is a technical point which requires some special treatment (an additional error appears if we restrict to a ball) but this is of no interest here.

At this stage, one calls upon the finite element method (or the quantization) in order to solve $(\overline{E}_k)$. In both cases, one builds up a grid $\Gamma := \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$ and looks for some weights π_{ij}^k in order to approximate $\overline{u}_k$ by

$$\widehat{u}_k(x_i) = \sum_{1 \le j \le N} \pi_{ij}^k \, \widehat{u}_{k+1}(x_j). \tag{47}$$

4.1 The Finite Element Method

In the finite element method, one tries to find the grid which best fits in the geometry of the problem. The same finality appears in the quantization method when we look for an optimal quantization. The simplest grid used in the finite element method is based on triangles. So each x_i is the vertex of several triangles and we may consider it as the centroid of the polygon completed by these triangles. We denote by $\mathcal{N}_i := \{x_{i_1}, \ldots, x_{i_k}\}$

the vertices of this polygon *i.e.* the points which are vertices of a triangle having x_i as a vertex. These are the neighbours of x_i.

Now, a grid Γ being defined, we focus on the construction of the weights π^k_{ij} and on their significance. We begin with the finite element method. One constructs the trials $T_i, 1 \leq i \leq N$, in the following way. One sets $T_i(x_i) := 1$ and $T_i(x_j) := 0$ for every $x_j \in \mathcal{N}_i$. Then one sets T_i to be linear on each triangle - so we get a pyramid centered at x_i whose basis is the polygon. The trial is null outside the polygon. The idea is to replace $(\overline{E}_k)$ by a finite dimensional problem to be solved in the space $\mathcal{S}_N$ spanned by $T_i, 1 \leq i \leq N$. Note that the trials are not orthogonal. Anyway each function $U \in \mathcal{S}_N$ may be written as $U(x) := \sum_{i=1}^N U_i\, T_i(x)$ (we use small letters for general functions in L^2 and capital letters for functions in $\mathcal{S}_N$). Note also that $U_i = U(x_i)$ and that

$$\frac{\partial U}{\partial x_i}(x) = \sum_{j=1}^N U_j \frac{\partial T_j}{\partial x_i}(x).$$

There are two ways of solving the finite dimensional problem, leading to the same result. The first one is the Galerkin method based on the weak form of the PDE and the second one is the Riesz-Reilich method based on the Dirichlet principle. We follow here the second idea. The solution of $(\overline{E}_k)$ is given by the function $u \in H^1_0(\mathbb{R}^d)$ which minimizes the energy

$$e(u) =: \int_{\mathbb{R}^d} \left(\frac{h}{2} \sum_{i=1}^d \left| \frac{\partial u}{\partial x_i}(x) \right|^2 - u(x)\overline{u}_{k+1}(x) \right) dx.$$

The discretization consists in solving the above minimum problem in $\mathcal{S}_N$. Since each function in this space may be identified with the vector $U := (U_1, \ldots, U_N)$, one may write the discrete problem as follows. Let

$$E(U) := \frac{h}{2} \sum_{i,j=1}^N U_i\, K_{ij}\, U_j - \sum_{i=1}^N U_i\, U_i^{k+1}$$

where $K_{ij} := \sum_{1 \leq r \leq N} < \frac{\partial T_i}{\partial x_p} | \frac{\partial T_j}{\partial x_r} >$ and $U_i^{k+1} := \widehat{u}_{k+1}(x_i)$, *i.e.* the coefficients of $\widehat{u}_{k+1}$. Note that $\overline{u}_{k+1}$ has been changed into $\widehat{u}_{k+1}$: the hat stresses that we are working with the approximation computed at the step $k+1$ of the dynamic principle algorithm. Now we have a finite dimensional problem whose solution is given by $U^k = \frac{2}{h} K^{-1} U^{k+1}$. This reads

$$\widehat{u}_k(x_i) = \sum_{1 \leq j \leq N} \pi^k_{ij} \widehat{u}_{k+1}(x_j) \qquad \text{with } \pi^k_{ij} := \frac{2}{h}(K)^{-1}_{ij} \qquad \text{(see Equation (47)).}$$

4.2 Quantization Method

We turn now to the quantization method where $\pi^k_{ij} = \mathbb{P}(X_{k+1} \in C_j(\Gamma_{k+1}) / X_k \in C_i(\Gamma_k))$. If one denotes by $P_h(x, dy) := \mathbb{P}(X_{k+1} \in dy \mid X_k = x)$, then

$$\pi^k_{ij} = \int_{C_i(\Gamma_k)} P_h(\mathbf{1}_{C_j(\Gamma_{k+1})})(\xi)\, d\,\mathbb{P} \circ X_k^{-1}(d\xi) \approx \frac{1}{h} P_h(\mathbf{1}_{C_j(\Gamma_{k+1})})(x_i).$$

So the part of the trial T_i in the finite element method is played here by the indicator function of the Voronoi tessel of x_i. Both $\frac{1}{h} T_i$ and $\frac{1}{h} \mathbf{1}_{C_i}$ are approximations of the Dirac mass at x_i. Finally we note that

$$\sum_{1 \leq j \leq N} \widehat{u}_{k+1}(x_j) \pi^k_{ij} \approx P_h \widehat{u}_{k+1}(x_i)$$

The Feynman-Kac formula shows that $\widehat{u}_k$ is the solution of $(\overline{E}_k)$ with final condition $\widehat{u}_{k+1}$.

4.3 Conclusion

In both methods the first step is a time discretization. Then, as a second step, both of them solve the same PDE problem $(\overline{E}_k)$ using a space approximation procedure. The finite element method relies on the variational principle whereas in the quantization method is based on the the probabilistic interpretation of the PDE. In the finite element method this leads to a linear system. Solving it amounts to inverting the sparse matrix K (only neighbours yield non-null entries). In the quantization method, we get directly the solution of the system: the weights are computed using the Monte Carlo method. Once again, the matrix K is sparse, numerically speaking since the Brownian motion does not go too far in one time step: many weights π^k_{ij} will be close to 0 except for the neighbours.

Finally, note that formula (47), whatever the way it comes up, reads as a finite difference scheme built on a grid which is no longer uniform with weights which are no longer $\frac{1}{h}$. It looks like a finite difference scheme *adapted to the geometry of the problem.*

Beyond these similarities, it appears that in the finite element method, *one projects the function* to be computed as an expectation whereas in the quantization method, *one projects the underlying process* X involved in the Feynman-Kac formula.

From a numerical point of view, there is a connection between the conditioning of the matrix K to be inverted in the finite element method and the asymptotic variance in the Central Limit Theorem that – heuristically – rules the rate of convergence of the $CLVQ$ algorithm: when K has a small eigenvalue, the variance of the $CLVQ$ is large *i.e.* it converges slowlier.

5 Applications to $RBSDE$'s and American option pricing

5.1 Back to $RBSDE$'s (and Optimal Stopping of Brownian diffusions)

In Subsection 1.1 we pointed out that (h,f)-Snell envelopes $(U_k)_{0\leq k\leq n}$ and $(\overline{U}_k)_{0\leq k\leq n}$ of the sampled diffusion $(X_{\frac{kT}{n}})_{0\leq k\leq n}$ or its Euler scheme $(\overline{X}_k)_{0\leq k\leq n}$ respectively provide natural discretization schemes for the$RBSDE$(according to the ability to simulate the diffusion). For a time discretization step $\frac{T}{n}$, these Snell envelopes are both related to the functions $f_k(x,u) := \frac{T}{n}f(t_k,x,u)$ and $h_k(x) := h(t_k,x)$, $k=0,\ldots,n$. They satisfy

$$U_n := h(T,X_T),\; U_k := \max\Big(h_k(X_{\frac{kT}{n}}), \mathbb{E}(U_{k+1}+f_k(X_{\frac{(k+1)T}{n}},U_{k+1})/\mathcal{F}_{\frac{kT}{n}})\Big), 0\leq k\leq n-1, \quad (48)$$

$$\overline{U}_n := h(T,\overline{X}_n),\; \overline{U}_k := \max\Big(h_k(\overline{X}_k), \mathbb{E}(\overline{U}_{k+1}+f_k(\overline{X}_{k+1},\overline{U}_{k+1})/\mathcal{F}_{\frac{kT}{n}})\Big), 0\leq k\leq n-1. \quad (49)$$

Throughout this section, we denote by $\overline{u}^{(n)}_k$ the function satisfying $\overline{U}_k := \overline{u}^{(n)}_k(\overline{X}_k)$. Lemma 2 below shows that the Lipschitz coefficients of functions $u^{(n)}_k$ and $\overline{u}^{(n)}_k$ do not explode as n goes to infinity. Its (easy) proof is left to the reader.

Lemma 2 *Assume that assumptions* $(Lip_{b,\sigma})$, (Lip_f), (Lip_h) *hold.*
(a) Diffusion: *the* (h,f)*-Snell envelope* $(U_k)_{0\leq k\leq n}$ *defined by (11) satisfies* $U_k = u^{(n)}_k(X_{t_k})$.

$$U_k = u^{(n)}_k(X_{t_k}) \quad \textit{with} \quad [u^{(n)}_k]_{Lip} \;\leq\; K_1\exp\left(K_0(T-t_k)\right)+\frac{\varepsilon_{n,k}}{n}$$
$$\textit{with} \quad K_1 := \gamma_0+\frac{1}{1+\gamma_0/2}, \qquad K_0 := \gamma_0(2+\gamma_0/2) \;\textit{ and }\; \lim_n \sup_{0\leq k\leq n}|\varepsilon_{n,k}| = 0.$$

If, furthermore, the transition is asymptotically flat with parameter $a > \gamma_0$, then

$$[u_k^{(n)}]_{Lip} \le K_2 + \frac{\varepsilon_{n,k}}{n}. \tag{50}$$

with $K_2 := \gamma_0 \max(1, 1/(a-\gamma_0))$ and $\lim_n \sup_{0\le k\le n} |\varepsilon_{n,k}| = 0$. When f does not depend on u (regular Optimal Stopping problems) then $K_2 := \gamma_0 \max(1, 1/a)$ and (50) holds as soon as the diffusion is asymptotically flat.

(b) Euler scheme: The (h, f)-Snell envelope $(\overline{U}_k)_{0\le k\le n}$ defined by (14) satisfies $\overline{U}_k = \overline{u}_k^{(n)}(\overline{X}_k)$ and the functions $u_k^{(n)}$ are Lipschitz continuous. The same bounds as those obtained for the Lipschitz coefficients of $u_k^{(n)}$ in item (a) hold.

Remarks: • In the asymptotically flat case, the minimal assumption on f is $a > [f]_{Lip}$.

• A less precise statement of Lemma 2 could be: there exist real constants $\tilde{K}_0$, $\tilde{K}_1 > 0$ only depending on b, σ, h and f such that,

$$\forall\, n \ge 1,\ \forall\, k \in \{0,\dots,n-1\},\quad [u_k^{(n)}]_{Lip} \le \tilde{K}_1 e^{\tilde{K}_0(T-t_k)}.$$

Furthermore, if the diffusion is "asymptotically flat" enough, one may set $\tilde{K}_0 := 0$.

Lemma 2 yields the following bounds for the coefficients $d_i^{(n)}$ defined by Equation (27).

Proposition 8 *Let $p \ge 1$. For every fixed n and every $k \in \{0,\dots,n\}$,*

$$\begin{aligned} \forall\, i \in \{k,\dots,n\},\quad \frac{d_i^{(n)}}{(1+\frac{T}{n}\gamma_0)^k} &= \left(\gamma_0 + (2-\delta_{2,p})K_1 e^{K_0(T-t_i)}\right) e^{\gamma_0(t_i-t_k)} + \frac{\varepsilon_{i,k,n}}{n} \\ \textit{with}\quad \lim_n \max_{0\le k\le i\le n} |\varepsilon_{i,k,n}| &= 0 \quad \textit{so that}\quad d^\infty := \sup_{n\ge 1} \max_{0\le i\le n} d_i^{(n)} < +\infty. \end{aligned}$$

We use the notations about *RBSDE* introduced in the introduction. The aim of this paragraph is to derive some *a priori* error bounds for $\|Y_{t_k} - \widehat{u}_k(\widehat{X}_k)\|_p$ as a function of the total number N of elementary quantizers and of the number n of time steps, all real constants depending on b, σ, h and T. To this end, we will need some precise estimates for the probability densities of the diffusion process $(X_t)_{t\in[0,T]}$ in the uniformly elliptic case. Assume that diffusion parameters b and σ satisfy the following assumptions

$$\begin{aligned} &(i)\quad \sigma\sigma^* \ge \varepsilon_0 I_d,\ \varepsilon_0 > 0 \quad \text{(uniform ellipticity)} \\ &(ii)\quad (b,\, \sigma \in \mathcal{C}_b^\infty(\mathbb{R}^d)) \ \text{ or } \ (b,\, Db,\, \sigma, D\sigma, D^2\sigma \ \text{ Lipschitz continuous and bounded}). \end{aligned} \tag{51}$$

Then, there exist two real constants α, $\beta > 0$ such that the diffusion process $(X_t^x)_{t\in[0,T]}$ starting at x has a probability density function $p_t(x, y)$ at every time $t \in [0, T]$ satisfying

$$p_t(x,y) \le \frac{\alpha}{(2\pi t)^{\frac{d}{2}}} \exp\left(-\frac{|y-x|^2}{2\beta t}\right). \tag{52}$$

The *bounded Lipschitz setting* is due to Friedman (Theorem 4.5 p.141 and 5.4 p.148-149 in [27]), the *smooth sub-linear setting* follows from [35] by Kusuoka & Stroock. When the diffusion is only hypoelliptic and satisfies a non degeneracy Hormander type assumption, it is also established in [35] that (52) holds with an exponent $d' \ge d/2$. This could lead to different dispatching rules (provided that d' is known). An alternative approach could be to rely on similar results established in [8] for an "excited" version of the Euler scheme.

Theorem 3 *Assume $(Lip_{b,\sigma})$, (Lip_f), (Lip_h) hold, that the diffusion $(X_t)_{t\in[0,T]}$ satisfies (51) and $X_0 = x$. For $N \geq n+1$, assign $N_k := \left\lceil \frac{t_k^{\frac{d}{2(d+1)}} N}{t_1^{\frac{d}{2(d+1)}}+\cdots+t_n^{\frac{d}{2(d+1)}}} \right\rceil \geq 1$ elementary quantizers to the optimal quantization grid Γ_k of $\overline{X}_k$ or X_{t_k}, $1 \leq k \leq n$ and set $N_0 = 1$. (Note that in fact $N \leq N_0 + \cdots + N_n \leq N + n + 1$).*

(a) Diffusion: *Let $\widehat{X}_k$ denote the optimal L^p-quantization of the diffusion X_{t_k}. Then,*

$$\forall\, p \geq 1, \qquad \max_{0\leq k\leq n} \|Y_{t_k} - u_k(\widehat{X}_k)\|_p \leq C_p e^{C_pT}\left(\frac{n^{1+\frac{1}{d}}}{N^{\frac{1}{d}}} + \frac{1+|x|}{n^{\theta}}\right) \tag{53}$$

where $\theta = 1$ if h is semi-convex and f is $\mathcal{C}_b^{1,2,2}$ and $\theta = 1/2$ otherwise.

(b) Euler scheme: *Let $\widehat{X}_k$ denote the optimal L^p-quantization of the Euler scheme $\overline{X}_k$. Then, the above error bound holds with $\theta = 1/2$.*

Practical comments about the dispatching: • One checks using $t_k = \frac{kT}{n}$ that

$$N_k \sim \frac{3d+2}{2(d+1)}\left(\frac{k}{n}\right)^{\frac{d}{2(d+1)}}\frac{N}{n} \quad \text{as} \quad n \to +\infty,\ N \to +\infty \quad \text{with} \quad n = o(N).$$

Note that the optimized ratio N_k/N of the elementary quantifiers assigned to time t_k marginally depends upon the dimension d since

$$\frac{N_k}{N} \approx \frac{3}{2n}\sqrt{\frac{k}{n}} \quad \text{when } d \text{ becomes "large" (say } d \geq 5). \tag{54}$$

• Then, the (theoretical) complexity of the algorithm is approximately $\frac{9}{8}\frac{N^2}{n}\kappa$ which is close to the lowest possible one (see (24)).

• Note that, for example in the Lipschitz continuous setting, if $N := \left\lceil n^{1+\frac{3d}{2}} \right\rceil$

$$\max_{0\leq k\leq n} \|Y_{t_k} - \widehat{u}_k(\widehat{X}_k)\|_p \leq \frac{C_p\, e^{C_pT}(1+|x|)}{\sqrt{n}}.$$

Proof of Theorem 3: (a) One derives from Proposition 2 and Theorem 1 applied to the diffusion $(X_{t_k})_{0\leq k\leq n}$ that, for every $p \geq 1$,

$$\begin{aligned} &\max_{0\leq k\leq n} \|Y_{t_k} - U_k\|_p &\leq& \ \frac{C_p(x)}{n^{\theta}} \quad \text{with } C_p(x) := C_pe^{C_pT}(1+|x|),\ C_p > 0,\\ \text{and} \quad &\max_{0\leq k\leq n} \|U_k - \widehat{U}_k\|_p &\leq& \ \sum_{k=0}^{n} d_k^{(n)}\|X_{t_k} - \widehat{X}_k\|_p. \end{aligned}$$

Proposition 8 shows that, for every $n \geq 1$ and every $k \in \{0,\ldots,n\}$,

$$\begin{aligned} d_k^{(n)} &= \gamma_0\left(1 + C_{b,\sigma,p}e^{(\gamma_0+C_{b,\sigma})(T-t_k)}\right)e^{\gamma_0 t_k} + \frac{\varepsilon_{k,n}}{n} \leq C_{\gamma_0,T,p} + \frac{|\varepsilon_{k,n}|}{n}\\ \text{with} \quad C_{\gamma_0,T,p} &:= \gamma_0 e^{\gamma_0 T}\left(1 + \frac{(2-\delta_{2,p})(1+C_{\gamma_0})}{C_{\gamma_0}}e^{(\gamma_0+C_{\gamma_0})T}\right),\ C_{\gamma_0} := \gamma_0(1+\gamma_0/2) \end{aligned}$$

and $\lim_n \max_{0\le k\le n} |\varepsilon_{k,n}| = 0$. First set $\widehat{X}_0 := X_0 = x$. Setting $\varepsilon_n := \max_{0\le k\le n} |\varepsilon_{k,n}|$, one has

$$N^{\frac{1}{d}} \max_{0\le k\le n} \|Y_{t_k} - \widehat{u}_k(\widehat{X}_k)\|_p \le \left(C_{\gamma_0,T,p} + \frac{\varepsilon_n}{n}\right) \sum_{k=0}^{n} \left(\frac{N}{N_k}\right)^{\frac{1}{d}} \times N_k^{\frac{1}{d}} \min_{|\Gamma|\le N_k} \|X_{t_k} - \widehat{X_{t_k}}^{\Gamma}\|_p + \frac{C_p(x)}{n^\theta}.$$

Now, the Gaussian domination Inequality (52) implies that, for every $k \in \{1,\dots,n\}$

$$\forall\, x,\, y \in \mathbb{R}^d, \qquad p_{t_k}(x,y) \le \alpha\beta^{\frac{d}{2}} \pi_{x+\sqrt{\beta_{t_k}}\, Z}(y) \quad \text{where} \quad Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0; I_d)$$

and $\pi_Y(x)$ is for the probability density function of a random vector Y. Hence, for every $k \in \{1,\dots,n\}$, $N_k \ge 1$, and every grid $\Gamma := \{v_1,\dots,v_{N_k}\} \subset \mathbb{R}^d$ of size N_k

$$\begin{aligned} \|X_{t_k} - \widehat{X}^{\Gamma}_{t_k}\|_p &\le \alpha\beta^{\frac{d}{2}} \| \min_{1\le \ell\le N_k} |v_\ell - x - \sqrt{\beta t_k}\, Z|\, \|_p, \qquad (55)\\ &= \alpha\beta^{\frac{d}{2}} \sqrt{\beta\, t_k}\, \|Z - \widehat{Z}^{(\Gamma - x)/\sqrt{\beta t_k}}\|_p. \end{aligned}$$

$$\text{Hence} \qquad \min_{\Gamma,\, |\Gamma|\le N_k} \|X_{t_k} - \widehat{X}^{\Gamma}_{t_k}\|_p \le \alpha\beta^{\frac{d}{2}} \sqrt{\beta\, t_k} \min_{|\Gamma|\le N_k} \|Z - \widehat{Z}^{\Gamma}\|_p. \qquad (56)$$

Applying Theorem 2 to Z (*i.e.* to the Normal distribution $\mathcal{N}(0; I_d)$ on $\mathbb{R}^d$) yields

$$\min_{|\Gamma|\le N_k} \|Z - \widehat{Z}^{\Gamma}\|_p \le (1+\eta_N)^{\frac{1}{p}} \widetilde{J}_{p,d}\, \|\pi_Z\|^{\frac{1}{p}}_{\frac{d}{d+p}} = (1+\eta_N)^{\frac{1}{p}} \widetilde{J}_{p,d}\, (1+p/d)^{\frac{d+p}{2p}} \sqrt{2\pi}. \qquad (57)$$

where $\widetilde{J}_{p,d} := J^{\frac{1}{p}}_{p,d}$ and $\lim_N \eta_N = 0$. Combining (56) and (57) yield

$$\begin{aligned} N_k^{\frac{1}{d}} \min_{|\Gamma|\le N_k} \|X_{t_k} - \widehat{X}^{\Gamma}_{t_k}\|_p &\le \alpha\beta^{\frac{d+1}{2}} \widetilde{J}_{p,d} (1+\eta_{N_k})^{\frac{1}{p}} (1+p)^{\frac{d+p}{2p}} \sqrt{2\pi t_k}\,, \quad k \in \{0,\dots,n\}.\\ N^{\frac{1}{d}} \max_{0\le k\le n} \|Y_{t_k} - \widehat{u}_k(\widehat{X}_k)\|_p &\le \left(C_{\gamma_0,T,p} + \frac{\varepsilon_n}{n}\right)(1 + \max_{1\le k\le n} \eta_{N_k}) \sqrt{2\pi}\, \widetilde{J}_{p,d} (1+\frac{p}{d})^{\frac{d+p}{2p}} \beta^{\frac{d+1}{2}} \sum_{k=1}^{n} \left(\frac{N}{N_k}\right)^{\frac{1}{d}} \sqrt{t_k} + \frac{C_p(x) N^{\frac{1}{d}}}{n^\theta}\\ &\le C_{n,\gamma_0,T,p,d} \sum_{1\le k\le n} \rho_k^{-\frac{1}{d}} \sqrt{t_k} + \frac{C_p(x) N^{\frac{1}{d}}}{n^\theta} \end{aligned}$$

where $\rho_k \propto t_k^{\frac{d}{2(d+1)}}$, $1 \le k \le n$, $\rho_1 + \cdots + \rho_n = 1$ and $N_k := \lceil \rho_k N \rceil \ge 1$ and $C_{n,\gamma_0,T,p,d}$ is bounded as $n \to \infty$ by a real constant $\overline{C}_{\gamma_0,T,p,d}$. Following Lemma 1,

$$N^{\frac{1}{d}} \max_{0\le k\le n} \|Y_{t_k} - \widehat{u}_k(\widehat{X}_k)\|_p \le \overline{C}_{\gamma_0,T,p,d} \left(\sum_{k=1}^{n} t_k^{\frac{d}{2(d+1)}}\right)^{1+\frac{1}{d}} + \frac{C_p(x) N^{\frac{1}{d}}}{n^\theta}.$$

Now, Jensen inequality implies that $\sum_{1\le k\le n} t_k^{\frac{d}{2(d+1)}} \le T^{\frac{d}{2(d+1)}}\, n$. Setting $\overline{C}_p := T^{\frac{d}{2(d+1)}} \overline{C}_{\gamma_0,T,p,d}$, yields the announced bound (53) by setting C_p at the appropriate value.

(*b*) The main modification lies in the above Inequality (55). With the same notations

$$\begin{aligned} \|\overline{X}_k - \widehat{\overline{X}_k}^{\Gamma}\|_p &= \| \min_{1\le \ell\le N_k} |v_\ell - \overline{X}_k| \|_p\\ &\le \| \min_{1\le \ell\le N_k} |v_\ell - X_{t_k}|^p \|_p + \|X_{t_k} - \overline{X}_k\|_p\\ &\le \|X_{t_k} - \widehat{X}^{\Gamma}_{t_k}\|_p + C_p e^{C_p T}(1+|x|)\frac{1}{\sqrt{n}}, \end{aligned}$$

using classical L^p-error bounds for the Euler scheme. The rest of the proof is the same. $\diamondsuit$

One easily derives an optimized choice for the number n of time steps.

Corollary 1 (a) Lipschitz setting (quantization of the Euler scheme or of the diffusion): *The optimal number n of time steps and the resulting error bound satisfy*

$$n \approx \left(\frac{2d}{d+1}C_p(x)\right)^{\frac{2}{3d+2}} N^{\frac{2}{3d+2}} \qquad \textit{and} \qquad |u_0(x) - \widehat{u}_0(x)| = O\left(N^{-\frac{1}{3d+2}}\right) = O\left(\frac{1}{\sqrt{n}}\right)$$

where $C_p(x) \le C_p e^{C_p T}(1+|x|)$.

(b) Semi-convex setting (quantization of the diffusion): *The optimal number n of time steps and the resulting error bound are*

$$n \approx \left(\frac{d}{d+1}C_p(x)\right)^{\frac{d}{2d+1}} N^{\frac{1}{2d+1}} \qquad \textit{and} \qquad |u_0(x) - \widehat{u}_0(x)| = O\left(N^{-\frac{1}{2d+1}}\right) = O\left(\frac{1}{n}\right).$$

5.2 Numerical pricing of American exchange options by quantization

5.2.1 The test model

One considers two risky assets, a stock S^1 with a geometric dividend rate λ and a stock S^2 without dividend. The interest rate r is deterministic and constant. Assume that (S^1, S^2) follows a Black & Scholes dynamics, so that, under the *risk-neutral* probability $\mathbb{P}$, one has

$$\begin{aligned} dS^1_t &= S^1_t((r-\lambda)\,dt + \sigma_1 dB^1_t),\ S^1_0 := s^1_0 > 0, \\ dS^2_t &= S^2_t(r\,dt + \sigma_2 dB^2_t), \qquad S^2_0 := s^2_0 > 0 \end{aligned}$$

where (B^1, B^2) is a 2-dimensional Brownian motions with covariance $< B^1, B^2 >_t = \rho\, t$, $\rho \in [-1, 1]$. One checks that the discounted traded assets

$$\widetilde{S}^1_t := e^{-rt}(e^{\lambda t} S^1_t) \qquad \text{and} \qquad \widetilde{S}^2_t := e^{-rt} S^2_t, \qquad t \in [0,T],$$

make up a 2-dimensional $\mathbb{P}$-martingale with respect to the filtration $\underline{\mathcal{F}}^B$ of the 2-dimensional Brownian motion $B := (B^1, B^2)$. The diffusion $S := (S^1, S^2)$ is obviously not *uniformly* elliptic but $(\ln S^1, \ln S^2)$ clearly is.

An American exchange option with exchange rate μ is the right to exchange once and only once at any time $t \in [0,T]$, μ units of asset S^2 for one unit of asset S^1. Or, put some way round, the right to buy one unit of asset S^1 at the price of μ units of asset S^2.

The discounted premium of such an option is defined as the $(h, 0)$-Snell envelope of

$$h_t := e^{-rt}(S^1_t - \mu\, S^2_t)_+ = \max(e^{-\lambda t}\widetilde{S}^1_t - \mu\, \widetilde{S}^2_t, 0).$$

where x_+ denotes the positive part of the real number x. Since h_t does not depend upon the interest rate r, we may *assume without loss of generality that* $r := 0$. So, if $\mathcal{E}x_t$ denotes the price of this exchange American option at time t,

$$\mathcal{E}x_t := \text{esssup}_{t \le \tau \le T}\mathbb{E}(h_\tau/\mathcal{F}_t) := \mathcal{E}(t, S^1_t, S^2_t, \rho, \sigma_1, \sigma_2, \lambda).$$

One noticeable feature of this derivative is that the premium of its European counterpart, *i.e.* the right to exchange *at time* T, μS^2 and S^1, admits a closed form given by

$$\begin{aligned} \mathit{Ex}_t &:= E(T-t, S^1_t, \mu\, S^2_t, \widetilde{\sigma}, \lambda), \quad \widetilde{\sigma} := \sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2} \\ E(t,x,y,\sigma,\lambda) &:= x\, e^{-\lambda t}\mathrm{erf}(d_1(t,x,y,\sigma,\lambda)) - y\,\mathrm{erf}(d_1(t,x,y,\sigma,\lambda)-\sigma\sqrt{t}) \\ d_1(t,x,y,\sigma,\lambda) &:= \frac{\ln\left(\frac{x}{y}\right)+\left(\frac{\sigma^2}{2}-\lambda\right)t}{\sigma\sqrt{t}} \quad \text{and} \quad \mathrm{erf}(x) := \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\frac{du}{\sqrt{2\pi}}. \end{aligned}$$

American exchange options own two characteristics: the (discounted) contingent claim h_t only depends on the state variable $S_t := (S^1_t, S^2_t)$ at time t and the European option related to h_T has a closed form. For such options, one uses the premium of the European option as a control variate to reduce the computations to the "residual" American part of the option. Namely, keeping in mind that $r = 0$, set

$$\forall\, t \in [0,T],\ k_t := h_t - \mathit{Ex}(T-t, S^1_t, \mu\, S^2_t, \widetilde{\sigma}, \lambda).$$

Since $\left(E(T-t, S^1_t, \mu\, S^2_t, \widetilde{\sigma}, \lambda)\right)_{t\in[0,T]}$ is a $\mathbb{P}$-martingale, it follows that

$$\mathcal{E}x_t = \mathrm{esssup}_{t\le\tau\le T} k_\tau + \mathit{Ex}_t.$$

So, pricing the American exchange option amounts to pricing the American option having k_t as discounted contingent claim. The interesting fact for numerical purpose is that $k_T \equiv 0$ and that k_t is always smaller than h_t. Moreover, standard computations show that one may write $k_t := k(\ln \widetilde{S}^1_t, \ln \widetilde{S}^2_t)$ where k is a Lipschitz continuous function.

5.2.2 Practical implementation and results

In a Black & Scholes model, the exact simulation of $(S_{t_k})_{0\le k\le n}$ at times $0 =: t_0 < t_1 < t_2 < \cdots < t_k < \cdots < t_n := T$ is possible. In fact, for numerical purpose, it is more convenient to consider the couple $(\ln \widetilde{S}^1_t, \ln \widetilde{S}^2_t)$ as the underlying variables of the pricing problem. One simulates recursively this process at times $t_k := \frac{k}{n}T$, $0 \le k \le n$, by setting $\widetilde{S}^1_0 := s^1_0 > 0$, $\widetilde{S}^2_0 := s^2_0 > 0$ and, for every $k \in \{0, n\ldots, n-1\}$,

$$\begin{aligned} \ln(\widetilde{S}^1_{t_{k+1}}) &:= \ln(\widetilde{S}^1_{t_k}) + (-\frac{\sigma_1^2}{2}\Delta + \sigma_1\sqrt{\Delta}\,\varepsilon_{2k}) \\ \ln(\widetilde{S}^2_{t_{k+1}}) &:= \ln(\widetilde{S}^2_{t_k}) + (-\frac{\sigma_2^2}{2}\Delta + \sigma_2\sqrt{\Delta}(\rho\,\varepsilon_{2k} + \sqrt{1-\rho^2}\,\varepsilon_{2k+1})) \end{aligned}$$

where ε_k is an i.i.d. sequence of Normal random variables and $\Delta := \frac{T}{n}$.

The parameters of the options has been set as follows $T := 1$ (1 year, $\sigma_1 = \sigma_2 := 20\%$, $\lambda := 5\%$, $S^1_0 = 40$, $S^2_0 := 36$ or 44, $\mu := 1$ and $\rho \in \{-0.8; 0; 0.8\}$ (the price does not depend upon the interest rate r).

The quantization has been processed with $N := 5\,722$ $\mathbb{R}^2$-valued elementary quantizers dispatched on 25 layers using the above optimal dispatching rule (43). The estimation of the weights (and of the quadratic quantization error) have been carried out using $M := 10^6$ trials in the $CLVQ$ algorithm. The reference solution labeled VZ is computed by a $2D$-finite difference algorithm devised by S. Villeneuve and A. Zanette in [56].

The "quantization" price of European style options have been computed using the number N_{25} of elementary quantizers on the last layer (25^{th}), namely $N_{25} := 299$. The numerical experiments have been carried out with $\mu = 1$. Of course, once the quantization is performed, the pricing of any American style options for any (reasonable) value of μ simply needs to re-run the (pseudo-)dynamic programming formula whose C.P.U. cost is negligible. One could parametrize the starting values s_0^1 and s_0^2 the same way round.

$S_0^2 := 36, \quad \rho$	$-0,8$	0	$0,8$	$S_0^2 := 44, \quad \rho$	$-0,8$	0	$0,8$
Euro. B & S	$6,6547$	$5,2674$	$3,0674$	Euro. B &S	$3,6390$	$2,2289$	$0,3217$
Euro. Quantiz	$6,6297$	$5,2558$	$3,0639$	Euro. Quantiz	$3,6133$	$2,2117$	$0,3151$
Am. VZ.	$6,9754$	$5,6468$	$4,0000$	Am. VZ	$3,7692$	$2,3364$	$0,3595$
Am. Quantiz	$6,9812$	$5,6520$	$4,000$	Am. Quantiz	$3,7726$	$2,3398$	$0,3610$

Fig. 4 below displays the global results obtained for 25 maturities from (approximately) 2 weeks up to 1 year.

One important and promising fact is that quite similar results have been obtained by directly quantizing the Standard Brownian Motion itself instead of the geometric brownian motions of the above Black & Scholes model. The first noticeable fact is that the function h_k are no longer Lipschitz continuous: it pleads in favour of the robustness of the method. This robustness of the method could play a role in the future development of this approach to finance since the quantization of a S.B.M. is parameter free (except for the dimension!). Furthermore, the quantization of every layer can be achieved, up to an appropriate dilatation, by using some pre-computed tables of optimal quantization of the standard d-dimensional Gaussian vector([2]). At this stage, if using these pre-computed grids, it remains to estimate the transition weights π_{ij}^k using a standard Monte Carlo simulation. Exploratory experiments showed that 10 000 trials are enough to obtain as accurate results as above for American exchange options (the total C.P.U. time cost for this weight estimation is then $25\,s$). The quantization tree descent itself is instantaneous.

These results augur well of the future comparisons with former pricing methods for multidimensional American options (see [45, 15]). The simulations have been performed by J. Printems (Univ. Paris 12). Some extensive investigations will be carried out in [5].

5.2.3 Provisional remarks

There is a possible alternative to optimal quantization: one may build the grids of the quantization tree by settling the first $\overline{N} = N/n$ random paths of the Markov chain. The resulting theoretical quantization errors $a.s.$ still follow a $O(\overline{N}^{-1/d})$-rate, with a worse sharp rate (see [19]). The companion parameter estimation is carried out by a standard Monte Carlo simulation.

Some developments (quantized hedging) are proposed in [5], some first order schemes based on correctors obtained using Malliavin calculus are proposed in [6].

Annex: Partial $a.s.$-convergence results for the $CLVQ$ algorithm

As the distortion $D_N^{X,2}$ does not enjoy the standard properties of a potential for a stochastic gradient, we are lead to make the following restrictive assumption:

$$(\mathcal{C}) \quad \equiv \quad \text{supp}(\mathbb{P}_X) \text{ is a compact set.} \tag{58}$$

hence, the convex hull of supp($\mathbb{P}_X$) is a convex compact set.

[2] Generating such optimal quantization tables is carried out by a splitting method. At each step, 10^6 $CLVQ$ trials are processed. The C.P.U. time for producing an optimal $N+1$-grid from an optimal N-grid grows from $5\,s$ (if $N = 25$) up to $60\,s$ ($N = 300$).

$A.s.$-convergence in the 1-dimensional case In that very special setting, standard Stochastic Approximation Theory applies and we obtain a satisfactory *a.s.* convergence result. The convex hull of supp($\mathbb{P}_X$) is an interval $[a,b]$. One first notices that the set of N-grids is one-to-one with the simplex

$$\Sigma_N^{a,b,+} := \{\Gamma := (x_1,\cdots,x_N) \in (a,b)^N \,/\, a < x_1 < \cdots < x_N < b\}$$

which is invariant for the algorithm provided that the starting value $\Gamma^0 \in \Sigma_N^{a,b,+}$. Then,

$$\forall\, \Gamma := (x_1,\cdots,x_N) \in \Sigma_N^{a,b,+}, \qquad \nabla D_N^{X,2}(\Gamma) := 2\left(\int_{\widetilde{x}_i}^{\widetilde{x}_{i+1}} (x_i - \xi)\,\mathbb{P}_X(d\xi)\right)_{1\le i\le N} \tag{59}$$

where $\widetilde{x}_1 := a$, $\widetilde{x}_i := \frac{x_i + x_{i-1}}{2}$, $\widetilde{x}_{N+1} := b$. Consequently, the distribution $\mathbb{P}_X$ being continuous (*i.e.* no single ξ is weighted), $\nabla D_N^{X,2}$ has a *continuous extension* on the compact set $\overline{\Sigma}_n^{a,b,+}$.

Theorem 4 *([50]) (a) If $\Gamma^0 \in \Sigma_N^{a,b,+}$, then the algorithm (37) lives in $\Sigma_N^{a,b,+}$. Furthermore if $\mathbb{P}_X$ is continuous, if $\{\nabla D_N^{X,2} = 0\} \cap \overline{\Sigma}_N^{a,b,+}$ is finite and if the step γ_t satisfies the usual decreasing step assumption $\sum_{t\ge1}\gamma_t = +\infty$ and $\sum_{t\ge1}\gamma_t^2 < +\infty$, then*

$$\Gamma^t \xrightarrow{a.s.} \Gamma^* \in \{\nabla D_N^{X,2} = 0\} \qquad \textit{(see [50]).}$$

(b) Moreover, if $\mathbb{P}_X$ has a log-concave density *then $\{\nabla D_N^{X,2} = 0\} \cap \overline{\Sigma}_N^{a,b,+} = \mathrm{argmin}_{\overline{\Sigma}_N^{a,b,+}} D_N^{X,2} = \{\Gamma^*\}$ (see [30, 55, 31, 41]), with $\Gamma^* = \{a + \frac{2k-1}{2N}(b-a),\ 1 \le k \le N\}$ if $X \sim U([0,1])$.*

The multi-dimensional case ($d \ge 2$) In this multi-dimensional setting, only partial results are proved, even for bounded distributions μ. Let us mention once again that the singularity of $\nabla D_N^{X,2}$ makes the standard theory inefficient. The result below follows from a specific proof. The assumption on the stimuli distribution is now the following

$$(\mathcal{D}_X) \quad \equiv \quad \mathbb{P}_X \text{ has a bounded density } f \text{ with a compact convex support.} \tag{60}$$

Roughly speaking, Theorem 5 below says that, *a.s.*, either the N components of Γ^t remain parted and converge to some stationary quantizer of $D_N^{X,2}$ or they get stuck into M aggregates which will converge, up to some subsequence, toward a stationary quantizer of $D_M^{X,2}$. This is of very little help for simulations since it does not say precisely how often the elementary quantizers remain parted. In the worst case – all the components get stuck into a single aggregate at the expectation of $\mathbb{P}_X$ – the resulting quantization would be simply useless. These partial theoretical results seem very pessimistic in view of the practical performance of the CLVQ: no aggregation phenomenon is usually observed when the algorithm is appropriately initialized (random or splitting method).

Theorem 5 *(See* [50]*) assume that the above assumption $(\mathcal{D}_X)$ holds and that the step sequence satisfies $\sum_{t\ge1}\gamma_t = +\infty$ and $\sum_{t\ge1}\gamma_t^2 < +\infty$. Then*

(a) $\mathbb{P}$-a.s., either the elements of Γ^t remain asymptotically parted or at least two elements of Γ^t get asymptotically stuck as $t \to +\infty$ i.e.

$$\underline{\lim}_t \mathrm{dist}(\Gamma^t, {}^c\mathcal{S}_N) > 0 \quad \textit{or} \quad \lim_t \mathrm{dist}(\Gamma^t, {}^c\mathcal{S}_N) = 0$$

where $\mathcal{S}_N$ denotes the set of N-tuples with pairwise distinct components.

(b) On the event $\{\underline{\lim}_t \mathrm{dist}(\Gamma^t, {}^c\mathcal{S}_N) > 0\}$ (asymptotically parted components), there exists a "level" $\delta^* > 0$ and a connected component Γ^* of $\{\nabla D_N^{X,2} = 0\} \cap \{D_N^{X,2} = \delta^*\}$ s.t.

$$\Gamma^t \xrightarrow{a.s.} \Gamma^* \quad \textit{as} \quad t \to +\infty.$$

(c) On the event $\big\{\underline{\lim}_t \mathrm{dist}(\Gamma^t, {}^c\mathcal{S}_N) = 0\big\}$, the components get definitely stuck that is

$$\mathrm{dist}(\Gamma^t, {}^c\mathcal{S}_N) \overset{t\to+\infty}{\longrightarrow} 0 \quad \textit{as} \quad t \to +\infty.$$

There is a partition $I_1 \cup \cdots \cup I_M$ of $\{1,\cdots,N\}$ along which the components Γ^t make M aggregates as $t \to +\infty$. At least one of the limiting values of Γ^t is a zero of $\nabla D_M^{X,2}$.

References

[1] Abaya E., Wise G. (1992) On the existence of optimal quantizers, *IEEE on Information Theory*, **38**, n^{o}2, 937-946.

[2] Bally V., Caballero M.E., Fernandez B., El Karoui N. (2002) Reflected BSDE's, PDE's and Variational inequalities, technical report n^{0} 4455, Projet MATFI, INRIA (France).

[3] Bally V. (2002) The Central Limit Theorem for a non linear algorithm based on quantization, preprint INRIA n^{0} 4629, to appear in *Proceedings of the Royal Society.*

[4] Bally V., Pagès G. (2001) Error analysis of the quantization algorithm for obstacle problems, forthcoming in *Stoch. Proc. and their Appl.*, technical report $n°$642, Labo. Proba. & Modèles Aléatoires, Univ. Paris 6 (France).

[5] Bally V., Pagès G., Printems J. (2002) A quantization tree method for pricing and hedging multi-dimensional American options, technical report $n°$753, Labo. Proba. & Modèles Aléatoires, Univ. Paris 6 (France).

[6] Bally V., Pagès G., Printems J. (2003) First order schemes in the numerical quantization method, *Mathematical Finance* (selected papers of the colloquium *Applications of Malliavin calculus to Finance, 2001)*, **13**, 1-16.

[7] Bally V., Saussereau B. (2002) Approximation of the Snell envelope and computation of American option prices in dimension one, *ESSAIM P&S*, **16**, 1-21.

[8] Bally V., Talay, D. (1996) The law of the Euler scheme for stochastic differential equations (II): Convergence rate of the density, *Monte Carlo Methods and Applications*, **2**, 93-128.

[9] Bensoussan A., Lions J.L. (1982) *Applications of the Variational Inequalities in Stochastic Control*, North Holland, or (1978) *Applications des inéquations variationnelles en contrôle stochastique*, Dunod, Paris.

[10] Bouchard-Denize B., Touzi N. (2002) Discrete time approximation and Monte-Carlo simulation of backward stochastic differential equations, technical report $n°$766, Labo. Proba. & Modèles aléatoires, Univ. Paris 6 (France).

[11] Bouton C., Pagès G. (1997) About the multidimensional Competitive Learning Vector Quantization algorithm with constant gain, *The Annals of Applied Probability*, **7**, n^{o}3, 679-710.

[12] Brandière O., Duflo M. (1996) Les algorithmes stochastiques contournent-ils les pièges?, *Ann. de l'Inst. H. Poincaré*, Proba. & Stat. **32**, n^{o}3, 395-427.

[13] Briand P., Delyon B., Mémin J. (2002) On the robustness of backward stochastic differential equations. *Stochastic Process. Appl.*, **97**, no. 2, 229-253.

[14] Briand P., Delyon B., Mémin J. (2001) Donsker-type theorem for BSDEs. *Electron. Comm. Probab.*, **6**, 1-14 (electronic).

[15] Broadie M., Glasserman P. (1997) Pricing American-Style Securities Using Simulation, *Journal of Economic Dynamics and Control*, **21**, n^{0}8-9, 1323-1352.

[16] Bucklew J., Wise G. (1982) Multidimensional Asymptotic Quantization Theory with r^{th} Power distortion Measures, *IEEE on Inform. Th., Special issue on Quantization*, **28**, n^{o}2, 239-247.

[17] Caverhill A.P., Webber N. (1990) American options: theory and numerical analysis, in *Options: recent advances in theory and practice*, Manchester University Press.

[18] Chevance D. (1997) Numerical methods for backward stochastic differential equations, *Numerical Methods in Finance*, L. Rogers and D. Talay eds., Publications of the Newton Institute series, Cambridge University Press.

[19] Cohort P. (2003) Limit Theorems for the Random Normalized Distortion, forthcoming in *Ann. of Appl. Proba.*

[20] Duflo, M. (1997), *Random Iterative Models*, Coll. Applications of Mathematics, **34**, Springer-Verlag, Berlin, 1997.

[21] El Karoui N., Peng S., Quenez M.C. (1997) Backward stochastic differential equations in finance, *Math. Finance*, **7**, No.1, 1-71.

[22] El Karoui N., Kapoudjan C., Pardoux É., Peng S., Quenez M.C. (1997) Reflected solutions of Backward Stochastic Differential Equations and related obstacle problems for PDE's, *The Ann. of Proba.*, **25**, No 2, 702-737.

[23] Fort J.C., Pagès G. (1995) On the *a.s.* convergence of the Kohonen algorithm with a general neighborhood function, *The Ann. of Applied Proba.*, **5**, n$^{\text{O}}$4, 1177-1216.

[24] Fort J.C., Pagès G. (2002) Asymptotics of optimal quantizers for some scalar distributions, *Journal of Comput. and Applied Mathematics*, **146**, 253-275.

[25] Fournié É., Lasry J.M., Lebouchoux J., Lions P.L., Touzi N. (2001) Aplications of Malliavin calculus to Monte Carlo methods in Finance II, *Finance & Stochastics*, **3**, 391-412.

[26] Fournié É., Lasry J.M., Lebouchoux J., Lions P.L. (2001) Aplications of Malliavin calculus to Monte Carlo methods in Finance II, *Finance & Stochastics*, **5**, 201-236.

[27] Friedman A. (1975) *Stochastic Differential Equations and Applications*, Academic Press, **1**.

[28] Gersho A., Gray R. (1992) *Vector Quantization and Signal Compression*, MA Kluwer, Boston, 732 p.

[29] Gersho A., Gray R. (eds.) (1982) *IEEE on Inform. Th., Special issue on Quantization*, **28**.

[30] Graf S., Luschgy H. (2000) *Foundations of quantization for random vectors*, Lecture Notes in Mathematics n$^{\text{O}}$1730, Springer, 230p.

[31] Kieffer J. (1982) Exponential rate of Convergence for the Lloyd's Method I, *IEEE on Information Theory, Special issue on Quantization*, **28**, n$^{\text{O}}$2, 205-210.

[32] Kohatsu-Higa A., Petterson R. (2001) Variance reduction methods for simulation of densities on Wiener space, *SIAM Journal of Numerical Analysis*, to appear.

[33] Kushner H.J. (1977) *Probability Methods for Approximation in Stochastic Control and for Elliptic Equations*, Academic Press, New-York, 243p.

[34] Kushner H.J., Yin G.G. (1997) *Stochastic Approximation Algorithms and Applications*, Springer, New York.

[35] Kusuoka S., Stroock D. (1985) Applications of the Malliavin calculus, part II, *J. Fac. Sci. Univ. Tokyo*, **32**, 1-76.

[36] Lazarev V.A. (1992) Convergence of stochastic approximation procedures in the case of a regression equation with several roots. Translated from *Problemy Pederachi Informatsii*, vol. **28**, $n^0$1.

[37] Lamberton D. (1998) Error estimates for the binomial approximation of the American put option, *Annals of Applied Probability*, **8**, n^0 1, 206-233.

[38] Lamberton D. (2002) Brownian optimal stopping and random walks, *Applied Mathematics and Optimization*, **45**, 283-324.

[39] Lamberton D., Lapeyre B. (1996) Introduction to Stochastic Calculus applied to Finance, Chapman & Hall, 185p.

[40] Lamberton D., Pagès G. (1990) Sur l'approximation des réduites, *Ann. Inst. Poincaré*, **26**, $n^0$2, 331-355.

[41] Lamberton D., Pagès G. (1996) On the critical points of the 1-dimensional Competitive Learning Vector Quantization Algorithm, *Proceedings of the ESANN'96*, Bruges (Belgium).

[42] Lamberton D., Rogers L.C. (2000) Optimal stopping and embedding, *Journal of Applied Probability*, **37**, 1143-1148.

[43] Lapeyre B., Pagès G., Sab K. (1990) Sequences with low discrepancy. Generalisation and application to Robbins-Monro Algorithm, *Statistics*, **21**, n^{o}2, 251-272.

[44] Lions P.L., Régnier H. (2002) Calcul des prix et des sensibilités d'une option américaine par une méthode de Monte Carlo, working paper.

[45] Longstaff F.A., Schwartz E.S. (2001) Valuing American options by simulation: a simple least-squares approach, *Review of Financial Studies*, **14**, 113-148.

[46] Lloyd S.P. (1982) Least squares quantization in PCM, *IEEE on Information Theory, Special issue on Quantization*, **28**, $n^0 2$, 129-137 (adapted from a Bell Memo. (1957)).

[47] Ma J., Protter P., San Martín J., Torres S. (2002) Numerical method for backward stochastic differential equations, *Ann. Appl. Proba.*, **12**, no. 1, 302-316.

[48] Neveu J. (1971) *Martingales à temps discret*, Masson, Paris, 215p.

[49] Pardoux É., Peng S. (1992) Backward SDE's and quasi-linear PDE's, in *Stochastic Partial Differential Equations and their Applications*, B.L. Rozovskii and R.B. Sowers editors, Lectures Notes in Control and Inform. Sci. 176, Springer.

[50] Pagès G. (1997) A space vector quantization method for numerical integration, *Journal of Computational and Applied Mathematics*, **89**, 1-38.

[51] Pagès G., Printems J. (2003) Optimal quantization for numerics: the Gaussian case, technical report, Labo. Proba. & Modèles aléatoires, Univ. Paris 6 (France).

[52] Shiriaev A.N. (1984) *Probability, Graduate Text in Mathematics*, Springer, New-York, 577 p.

[53] Pelletier M. (2000) Asymptotic almost sure efficiency of averaged stochastic algorithms, *SIAM J. Control Optim.*, **39**, n^0 1, 49-72.

[54] Pemantle R. (1990) Nonconvergence to unstable points in urn models and stochastic approximations, *Annals of Probability*, **18**, n^0 2, 698-712.

[55] Trushkin A. (1982) Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions, *IEEE on Information Theory, Special issue on Quantization*, **28**, $n^0 2$, 187-198.

[56] Villeneuve S., Zanette A. (2002) Parabolic A.D.I. methods for pricing american option on two stocks, forthcoming in *Mathematics of Operation Research*, **27**, $n^0 1$, 121-149.

[57] Zador P. (1982) Asymptotic quantization error of continuous signals and the quantization dimension, *IEEE on Information Theory, Special issue on Quantization*, **28**, $n^0 2$, 139-148.

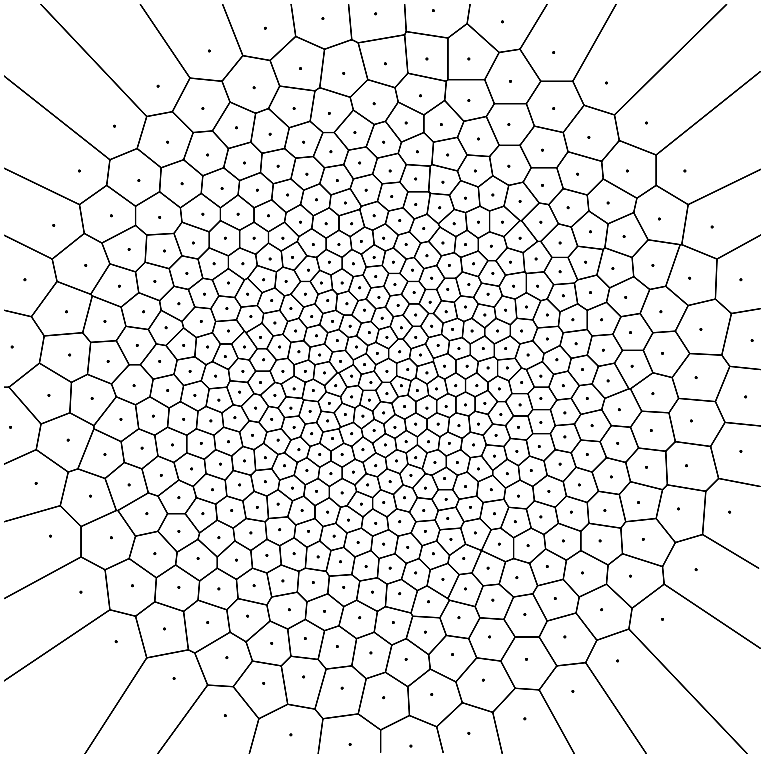

Fig.1: Optimal L^2-quantization of the Normal distribution $\mathcal{N}(0; I_2)$ with a 500-tuple and its Voronoi tessellation

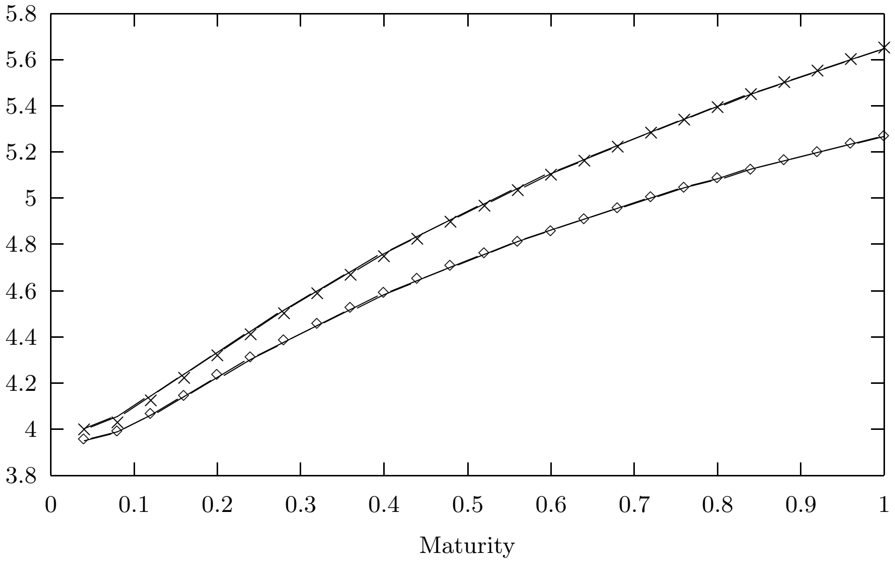

Fig.4: AM and EURO style option prices as a function of the maturity ($S_0^1 := 40$, $S_0^2 := 36$, $\rho := 0$)
− depicts the reference prices (V&Z for AM style and B&S for EURO style options),
× depicts the quantization price for AM style options,
◇ depicts the quantization price for EURO style options.