ESAIM: PROCEEDINGS AND SURVEYS, October 2015, Vol. 51, p. 272-292
A. Garivier et al, Editors

RECENT ADVANCES IN VARIOUS FIELDS OF NUMERICAL PROBABILITY *, **

C.E. Bréhier[1, 2], P.E. Chaudru de Raynal[3], V. Lemaire[4], F. Panloup[5] and C. Rey[6]

Abstract. The goal of this paper is to present a series of recent contributions on some various problems of numerical probability. Beginning with the Richardson-Romberg Multilevel Monte-Carlo method which, among other fields of applications, is a very efficient method for the approximation of diffusion processes, we focus on some adaptive multilevel splitting algorithms for rare event simulation. Then, the third part is devoted to the simulation of McKean-Vlasov forward and decoupled forward-backward stochastic differential equations by some cubature algorithms. Finally, we tackle the problem of the weak error estimation in total variation norm for a general Markov semi-group.

1. Introduction

During the last decades, probabilistic numerical methods received a growing attention. This can be explained by several reasons: the need for giving some answers, even approximate ones, is increasing in many domains such as (for instance) Biology, Physics, Finance, industrial ingeneering or Big Data related problems. This first argument must be associated with two other ones: the exponential growth of the capacity of computers combined with the *a priori* simplicity of crude Monte-Carlo methods (which is certainly not the only probabilistic numerical method but probably the most well-known).

In this paper, we would like to present four recent contributions on the topic. Before going further, let us note that although the works presented below are related to various numerical problems, they have in common to combine algorithmic efficiency with theoretical sharpness. In this sense, they are in the continuity of the numerous works on the topic of several French researchers since the 1980's.

* *C.-E. B. would like to thank his collaborators M. Gazeau, L. Goudenège, T. Lelièvre, M. Rousset and L. Tudela. He would also like to thank N. Chenavier for a careful reading of a draft version of his contribution.*

** *Fabien Panloup would like to thank the other authors for the contributions presented in this paper.*

[1] Université Paris-Est, CERMICS (ENPC), 6-8-10 Avenue Blaise Pascal, Cité Descartes , F-77455 Marne-la-Vallée, France; e-mail: brehierc@cermics.enpc.fr

[2] INRIA Paris-Rocquencourt, Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay, France; e-mail: charles-edouard.brehier@inria.fr

[3] IUT de Chambery, LAMA, UMR 5127,Université de Savoie, Campus Scientifique,73376 Le Bourget-du-Lac Cedex, France, E-mail: pe.deraynal@univ-savoie.fr

[4] Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC, Case 188, 4 pl. Jussieu, F-75252 Paris Cedex 5, France, E-mail: vincent.lemaire@upmc.fr

[5] Institut de Mathématiques de Toulouse, Université Paul Sabatier&INSA Toulouse, 135, av. de Rangueil, F-31077 Toulouse Cedex 4, France, E-mail: fabien.panloup@math.univ-toulouse.fr

[6] CERMICS-École des Ponts Paris Tech, 6 et 8 avenue Blaise Pascal, Cité Descartes - Champs sur Marne, 77455 Marne la Vallée Cedex 2, France, E-mail: reyc@cermics.enpc.fr

Article published online by EDP Sciences and available at http://www.esaim-proc.org or http://dx.doi.org/10.1051/proc/201551015

In Section 2, Vincent Lemaire introduces a new estimator called Multilevel Richardson-Romberg (ML2R) which combines the higher order bias cancellation of the Multistep Richardson-Romberg method introduced in [61] and the variance control resulting from the stratification in the Multilevel Monte Carlo (MLMC) method introduced in [29] (see also [37, 43]). In particular, it is proved that that one can build to build a combination which outperforms MLMC. This note is based on a paper of the author with G. Pagès (see [55]).

In Section 3, Charles-Edouard Bréhier focuses on recent algorithms for the estimation of rare events. After a brief review of reduction of variance strategies, especially the Multilevel Splitting approach (see [42]), the author introduces different adaptive versions of the latter and states a series of its properties: the goal is to build a theoretically unbiased estimator in an idealized setting and then to explain and to test numerically how the algorithm should be modified to preserve the unbiasedness property in a more general setting. This note is mainly based on the following joint works: [16] and [15].

In Section 4, Paul-Éric Chaudru de Raynal shortly describes a cubature algorithm for the computation of the expectation of some functionals of the solution of a decoupled Mc Kean-Vlasov Forward Backward Stochastic Differential Equation (MKV-FBSDE) type. This work is mainly based on a collaboration with C.A. Garcìa Trillos (see [25]).

Finally, in Section 5, Clément Rey considers the problem of the estimation of the weak error related to the discretization schemes of a given Markov process. More precisely, for a general $\mathbb{R}^d$-valued Markov process, the author gives some estimates of the total variation distance between the marginals of the process and its discretization scheme under some general conditions (including the non-degeneracy of the semi-group). As an example, an application to the estimation of the weak error related to the Ninomiya-Victoir scheme (for diffusions) is given. This part is based on a joint work with V. Bally.

2. Multilevel Richardson-Romberg extrapolation

We propose and analyze a new estimator called Multilevel Richardson-Romberg (ML2R) which combines the higher order bias cancellation of the Multistep Richardson-Romberg method introduced in [61] and the variance control resulting from the stratification in the Multilevel Monte Carlo (MLMC) method introduced in [29] (see also [37, 43]). The Multilevel Monte Carlo method has been extensively applied to various fields of numerical probability recently. We refer to the webpage http://people.maths.ox.ac.uk/gilesm/mlmc_community.html for a comprehensive list of references in this field.

2.1. Framework

We aim at computing a very accurate approximation of $I_0 = \mathbf{E}[Y_0]$ by a Monte Carlo type estimator where the (non-degenerate) random variable $Y_0 \in \mathbf{L}^2(\mathbf{P})$ cannot be simulated (exactly) at a reasonable cost. Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space and $(Y_h)_{h\in\mathcal{H}}$ be a family of real-valued random variables in $\mathbf{L}^2(\mathbf{P})$ associated to Y_0 where $\mathcal{H} \subset (0, \mathbf{h}]$ is an admissible subset of parameters having 0 as a limiting value and such that $\frac{\mathcal{H}}{n} \subset \mathcal{H}$ for every integer $n \geqslant 1$. Typically the random variable Y_h results from a time discretization scheme of parameter h or from an inner approximation in a nested Monte Carlo so that we will speak of h as the *bias parameter* in what follows. We assume that the family $(Y_h)_{h\in\mathcal{H}}$ satisfies the following two conditions which formalize the approximates Y_0

$$\exists\, \alpha > 0, c_k, \bar{R} \geqslant 1, \quad \mathbf{E}[Y_h] - \mathbf{E}[Y_0] = \sum_{k=1}^{\bar{R}} c_k h^{\alpha k} + o(h^{\alpha(\bar{R}+1)}), \tag{Bias error}$$

$$\exists\, \beta > 0, V_1 \geqslant 0, \quad \|Y_h - Y_0\|_2^2 = \mathbf{E}\left[|Y_h - Y_0|^2\right] \leqslant V_1 h^{\beta}. \tag{Strong error}$$

Note that the parameters α, β, V_1, c_k and $\bar{R}$ are structural parameters which depend on the family $(Y_h)_{h\in\mathcal{H}}$. In the sequel, we set $\theta = \sqrt{V_1/\operatorname{var}(Y_0)}$.

Thus the following two problems of numerical probability can be formalized within this framework.

Brownian diffusion approximation: Let $X = (X_t)_{t\in[0,T]}$ be a Brownian diffusion with drift function b and diffusion coefficient σ, and its continuous Euler scheme $\bar{X}^h = (\bar{X}^h_t)_{t\in[0,T]}$ with bias parameter $h = \frac{T}{n}$, defined by

$$\bar{X}^h_t = X_0 + \int_0^t b\big(\bar{X}^h_{\underline{s}}\big)\mathrm{d}s + \int_0^t \sigma\big(\bar{X}^h_{\underline{s}}\big)\mathrm{d}W_s, \quad \text{where} \quad \underline{s} = kh \text{ on } [kh,(k+1)h)$$

In the case $Y_0 = f(X_T)$ where f is a real-valued β–Hölder continuous function, $\beta \in (0,1]$, we consider the approximation $Y_h = f(\bar{X}^h_T)$. It is classical background that (Strong error) is satisfied with β. Concerning (Bias error), the weak error expansion is valid with $\alpha = 1$ and we refer to [67] if f is a smooth function and to [8, 35] under ellipticity or hypo-ellipticity conditions. As for weak error, the functional case $Y_0 = F\big((X_t)_{t\in[0,T]}\big)$ is much more challenging. We can refer to [2] for results in this direction and references on this topic.

Nested Monte Carlo: The purpose is to compute by simulation quantities of the form

$$\mathbf{E}\left[f\big(\mathbf{E}\left[X \mid Y\right]\big)\right] \quad \text{with} \quad X = F(Z,Y)$$

where Y and Z are two independent random variables taking values in $\mathbf{R}^{q_Y}$ and $\mathbf{R}^{q_Z}$ respectively, and $F : \mathbf{R}^{q_Z} \times \mathbf{R}^{q_Y} \to \mathbf{R}$ is a Borel function. To comply with our framework, we set

$$\mathcal{H} = \{1/K,\, K \geqslant 1\}, \quad Y_0 = f\big(\mathbf{E}\left[X \mid Y\right]\big), \quad Y_{\frac{1}{K}} = f\left(\frac{1}{K}\sum_{k=1}^{K} F(Z_k, Y)\right)$$

where $(Z_k)_{k\geqslant 1}$ is an *i.i.d.* sequence of random variables with the same distribution as Z and independent of Y. If f is smooth enough we prove that the assumptions (Bias error) and (Strong error) are satisfied with $\alpha = \beta = 1$ (see [55] for details). For more references on the nested Monte Carlo in the field of financial risk and actuary we refer to [12, 28, 40].

2.2. Crude Monte Carlo estimator

A crude Monte Carlo estimator to approximate $I_0 = \mathbf{E}\left[Y_0\right]$ can be summarized as follows. The starting point is of course to fix a parameter $h \in \mathcal{H}$ (this bias parameter must be chosen carefully). So, let $(Y_h^{(k)})_{k\geqslant 1}$ be a sequence of independent copies of Y_h and the estimator $I_h^N = \frac{1}{N}\sum_{k=1}^N Y_h^{(k)}$. By the strong law of large numbers and the central limit theorem we have a well-known control of the renormalized *statistical error* $\sqrt{N}(I_h^N - \mathbf{E}\left[Y_h\right])$ which behaves as a centered Gaussian variable with variance $\mathrm{var}(Y_h)$ as N goes to infinity. On the other hand, there is a *bias error* due to the approximation of I_0 by $I_h = \mathbf{E}\left[Y_h\right]$.

A natural choice for measuring the random error $I_h^N - I_0$ is to consider the mean squared error $\mathbf{E}\big[(I_h^N - I_0)^2\big] = \big\|I_h^N - I_0\big\|_2^2$. Our aim is to minimize the cost of the simulation for a given target error, say $\varepsilon > 0$. The numerical cost of this biased Monte Carlo procedure is given by $\mathrm{Cost}(I_h^N) = Nh^{-1}$. We look for the optimal parameters $h^*(\varepsilon)$ and $N^*(\varepsilon)$ solution to the minimization problem

$$\big(h^*(\varepsilon), N^*(\varepsilon)\big) = \underset{\|I_h^N - I_0\|_2 \leqslant \varepsilon}{\operatorname{argmin}} \mathrm{Cost}(I_h^N). \tag{1}$$

Proposition 2.1. *For a prescribed L^2-error $\varepsilon > 0$, the optimal parameters $h^*(\varepsilon)$ and $N^*(\varepsilon)$ solution to (1) are given by*

$$h^*(\varepsilon) = \frac{(1+2\alpha)^{-\frac{1}{2\alpha}}}{\big|c_1\big|^{\frac{1}{\alpha}}}\varepsilon^{\frac{1}{\alpha}}, \quad N^*(\varepsilon) = \left(1 + \frac{1}{2\alpha}\right)\frac{\mathrm{var}(Y_0)\left(1 + \theta(h^*(\varepsilon))^{\frac{\beta}{2}}\right)^2}{\varepsilon^2}, \tag{2}$$

with $\theta = \sqrt{V_1/\operatorname{var}(Y_0)}$. *Furthermore, we have*

$$\limsup_{\varepsilon\to 0} \varepsilon^{2+\frac{1}{\alpha}} \operatorname{Cost}\left(I^{N^*(\varepsilon)}_{h^*(\varepsilon)}\right) \leqslant \left|c_1\right|^{\frac{1}{\alpha}} \left(1+\frac{1}{2\alpha}\right)(1+2\alpha)^{\frac{1}{2\alpha}} \operatorname{var}(Y_0).$$

This well-known result points out that the cost of the crude Monte Carlo estimator implemented with optimal parameters (2) grows as $\varepsilon^{-(2+\frac{1}{\alpha})}$ when ε goes to zero. In his seminal paper [29] Giles proves that the cost of a Multilevel Monte Carlo estimator can grow as $\varepsilon^{-2}\log^2(1/\varepsilon)$ in the most common case $\beta = 1$ and as ε^{-2} (which is optimal for a Monte Carlo method) in the case $\beta > 1$.

2.3. **Multilevel estimators**

We now present a specific implementation of the general Multilevel Richardson-Romberg (ML2R) estimator introduced in [55]. With this particular design the ML2R estimator is a weighted version of the classical Multilevel Monte Carlo (MLMC) estimator. More preciselly, for a depth level parameter $R \in \{2,\dots,\bar{R}\}$ and a well chosen sequence of real weights $(\mathbf{W}_j)_{2\leqslant j\leqslant R}$ we consider the weighted Multilevel estimator based on the following telescopic summation

$$\mathbf{E}\left[Y_{h_R}\right] = \mathbf{E}\left[Y_h\right] + \sum_{j=2}^{R} \mathbf{W}_j\, \mathbf{E}\left[Y_{h_j} - Y_{h_{j-1}}\right], \tag{3}$$

where $h_j = M^{-(j-1)}h$, $j = 1,\dots,R$ is a geometrically decreasing sequence of bias parameters depending on $M \geqslant 2$ a fixed parameter. We denote by $n_j = M^{(j-1)}$ the j-th *refiner coefficient* of the initial bias parameter $h \in \mathcal{H}$. Weights $(\mathbf{W}_j)_j$ are discussed after a brief reminder on construction of the Multilevel estimator.

At each level $j \in \{1,\dots,R\}$ the computation of $\mathbf{E}\left[Y_{h_j} - Y_{h_{j-1}}\right]$ is performed by a standard Monte Carlo procedure. The key point of a Multilevel estimator is that for each level we consider a number $N_j = \lceil N q_j \rceil$ of scenarios where $q = (q_1,\dots,q_R) \in \mathcal{S}_+(R) = \left\{q \in (0,1)^R, \sum_{j=1}^R q_j = 1\right\}$ in which Y_{h_j} and $Y_{h_{j-1}}$ are perfectly correlated. More precisely we consider R copies of the biased family denoted $Y^{(j)} = (Y^{(j)}_h)_{h\in\mathcal{H}}$, $j \in \{1,\dots,R\}$ attached to *independant* random copies $Y^{(j)}_0$ of Y_0. The specific version of ML2R presented here then writes

$$I^N_{h,R,M,q} = \frac{1}{N_1}\sum_{k=1}^{N_1} Y^{(1),k}_h + \sum_{j=2}^{L} \frac{\mathbf{W}_j}{N_j} \sum_{k=1}^{N_j} \left(Y^{(j),k}_{h_j} - Y^{(j),k}_{h_{j-1}}\right) \tag{4}$$

where for every j, $(Y^{(j),k})_{k\geqslant 1}$ is a sequence of independent copies of $Y^{(j)}$. This estimator depends now on the parameters h, R, M and ratios $(q_j)_{1\leqslant j\leqslant R}$ of the total number of scenarios N. These parameters must be carrefuly tuned to obtain an estimator with an optimal complexity for a given L^2–error. The MLMC estimator introduced by Giles corresponds to the case $\mathbf{W}_j = 1$.

We now briefly present the construction of the weights $(\mathbf{W}_j)_{1\leqslant j\leqslant R}$ which are the highlight of our estimator. Note that by Abel's transform, the telescopic summation (3) rewrites

$$\mathbf{E}\left[Y_{h_R}\right] = \sum_{j=1}^{R} \mathbf{w}_j\, \mathbf{E}\left[Y_{h_j}\right] \quad \text{with} \quad \mathbf{w}_j = \mathbf{W}_j - \mathbf{W}_{j+1},\ j = 1,\dots,R$$

where $\mathbf{W}_{R+1} = 0$. The derivative weights $(\mathbf{w}_j)_{1\leqslant j\leqslant R}$ are defined as in [61] in order to kill all terms in front of the coefficients c_r in (Bias error). This yields

$$\sum_{j=1}^{R} \mathbf{w}_j\, \mathbf{E}\left[Y_{h_j}\right] = \mathbf{E}\left[Y_0\right] + \mathcal{O}(h^{\alpha R}).$$

Such weights $(\mathbf{w}_j)_{1\leqslant j\leqslant R}$ are obtained as the unique solution to a Vandermonde system and we get the explicit expression (which depends uniquely on α, R and M) by standard Cramer formulas

$$\mathbf{w}_i = \frac{(-1)^{R-i} M^{-\frac{\alpha}{2}(R-i)(R-i+1)}}{\prod_{j=1}^{i-1}(1-M^{-j\alpha})\prod_{j=1}^{R-i}(1-M^{-j\alpha})}.$$

From now on, we are looking for the optimal parameters h^*, R^*, $q^* = (q_j^*)_{j=1,\dots,R^*}$, M^* minimizing the numerical cost for a given L^2–error $\varepsilon > 0$ (for clarity, we do not write the dependance in ε). We decompose this global optimization problem in different steps and we first obtain the optimal allocation strategy $(q_1^*,\dots,q_R^*)$ as a function of other fixed parameters. Note that this is a robust allocation which only depends on assumption (Strong error)

$$\begin{cases} q_1^* = \mu^*(1+\theta h^{\frac{\beta}{2}}), \\ q_j^* = \mu^* \theta h^{\frac{\beta}{2}} \left(|\mathbf{W}_j(R,M)| \frac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1}+n_j}} \right), \quad j = 2,\dots,R, \end{cases} \tag{5}$$

where μ^* is the normalizing constant such that $\sum_{j=1}^R q_j = 1$. Secondly, we obtain h^* as a function of R and finally we optimize in R. We obtain that $R^*(\varepsilon)$ must grow like $\sqrt{\log(1/\varepsilon)}$ as ε goes to zero. This result is to compare to the MLMC estimator where $R^*(\varepsilon)$ grows like $\log(1/\varepsilon)$. Parameter $N^*(\varepsilon)$ easily follows

$$N^*(\varepsilon) = \left(1+\frac{1}{2\alpha R}\right) \frac{\mathrm{var}(Y_0)\left(1+\theta h^{\frac{\beta}{2}} \sum_{j=1}^R |\mathbf{W}_j(R,M)| \left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2}{\varepsilon^2 \sum_{j=1}^R q_j (n_{j-1}+n_j)} \tag{6}$$

Theorem 2.2. *Assume that* $\lim_{R\to+\infty} |c_R|^{\frac{1}{R}} = \widetilde{c} \in (0,+\infty)$. *For a prescribed* L^2*–error* $\varepsilon > 0$, *the ML2R estimator* $I^N_{h,R,M,q}$ *implemented with a fixed* $M \geqslant 2$ *and the following parameters*

$$R^*(\varepsilon) = \left\lfloor \frac{1}{2} + \frac{\log(\mathbf{h}\,\widetilde{c}^{\frac{1}{\alpha}})}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(\mathbf{h}\,\widetilde{c}^{\frac{1}{\alpha}})}{\log(M)}\right)^2 + 2\frac{\log(1/\varepsilon)}{\alpha\log(M)}} \right\rfloor$$
$$h^*(\varepsilon) = (1+2\alpha R)^{\frac{1}{2\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} M^{-\frac{R-1}{2}}$$

q^* *given in* (5) *and* N^* *given in* (6), *satisfies*

$$\limsup_{\varepsilon\to 0} v(\beta,\varepsilon)\,\mathrm{Cost}\left(I^N_{h,R,M,q}\right) \leqslant K(\alpha,\beta,M)$$

$$\text{with } v(\beta,\varepsilon) = \begin{cases} \varepsilon^2 \left(\log(1/\varepsilon)\right)^{-1} & \text{if } \beta = 1, \\ \varepsilon^2 & \text{if } \beta > 1, \\ \varepsilon^2 e^{-\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(1/\varepsilon)\log(M)}} & \text{if } \beta < 1. \end{cases}$$

Note that in the case $\beta = 1$, ML2R is asymptotically more efficient than MLMC by a factor $\log(1/\varepsilon)$ which goes to $+\infty$ as ε goes to 0. This factor $\log(1/\varepsilon)$ is significant and provides a substantial gain in numerical simulations especially when we look for an accurate result (ε small). Our estimator is half way between MLMC and an unbiased Monte Carlo simulation (of cost ε^{-2}).

In the case $\beta < 1$, ML2R is asymptotically much more efficient than MLMC estimator by a factor $\varepsilon^{-\frac{1-\beta}{\alpha}} e^{-\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(M)\log(1/\varepsilon)}}$ which goes to $+\infty$ as ε goes to 0 in a very steep way. It seems clear that it is for this setting that our estimator is the most powerful with respect to regular MLMC. In the case $\beta > 1$, the two

Multilevel estimators achieve the same asymptotic rate ε^{-2} corresponding to a virtual unbiased Monte Carlo method based on the direct simulation of Y_0.

2.4. **Numerical examples**

Barrier option (without Brownian bridge method)

We consider an up-and-out call option in a Black-Scholes model discretized by a naïve Euler scheme to illustrate the case $\beta = 0.5$ and $\alpha = 0.5$. This path-dependent option with strike $K = 100$ and barrier $L = 120 > K$ is defined by its path-dependent payoff

$$\varphi(x) = e^{-rT}(x(T) - K)_+ \mathbf{1}_{\left\{\max_{t\in[0,T]} x(t) \leqslant L\right\}}, \quad x \in \mathcal{C}([0,T], \mathbf{R}).$$

The parameters of the Black-Scholes model are $s_0 = 100$, $r = 0$, $\sigma = 0.15$ and $T = 1$. For this set of parameters the reference price is $I_0 = 1.855225$.

Nested Monte Carlo for compound option pricing

A compound option is simply an option on an option. The exercise payoff of a compound option involves the value of another option. A compound option then has two expiration dates $T_1 < T_2$ and two strike prices K_1 and K_2. We consider here the example of a European style *Put* on a *Call* where the underlying risky asset S is still given by a Black-Scholes process with parameters (r, σ). On the first expiration date T_1, the holder has the right to sell a new *Call* option using the strike price K_1. The new *Call* has expiration date T_2 and strike price K_2. The payoff of such a *Put*-on-*Call* option writes

$$(K_1 - \mathbf{E}\left[(S_{T_2} - K_2)_+ \mid S_{T_1}\right])_+ .$$

To comply with our framework, we set $\mathcal{H} = \{1/K,\, K \geqslant 1\}$

$$Y_0 = f\big(\mathbf{E}\left[S_{T_2} \mid S_{T_1}\right]\big), \quad Y_{\frac{1}{K}} = f\left(\frac{1}{K}\sum_{k=1}^{K} F(Z^k, S_{T_1})\right)$$

where $(Z^k)_{k\geqslant 1}$ is an *i.i.d.* $\mathcal{N}(0;1)$–distributed sequence, $f(x) = (K_1 - x)_+$ and F such that

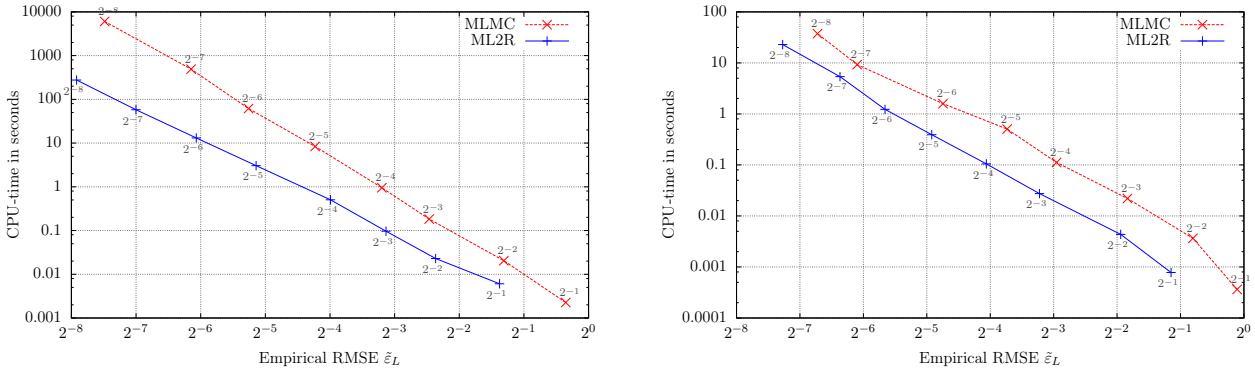
$$S_{T_2} = F(G, S_{T_1}) = S_{T_1} e^{(r - \frac{\sigma^2}{2})(T_2 - T_1) + \sigma\sqrt{T_2 - T_1} Z}.$$

The parameters used for the underlying process $(S_t)_{t\in[0,T_2]}$ are $S_0 = 100$, $r = 0.03$ and $\sigma = 0.3$. The parameters of the *Put*-on-*Call* payoff are $T_1 = 1/12$, $T_2 = 1/2$ and $K_1 = 6.5$, $K_2 = 100$. The reference price is $I_0 = 1.36857$.

Numerical results and comments

We compare the two estimators MLMC and ML2R on these two examples. We first do a crude evaluation of the structural parameters $\mathrm{var}(Y_0)$ and V_1, and consider that $c_1 = \tilde{c} = 1$. Then, for a fixed value of $\varepsilon > 0$, the optimal parameters R, h, the optimal allocation strategy $(q_j)_{1\leqslant j\leqslant R}$ and the optimal sample size N are expressed as a function of $M \geqslant 2$ (the unique free parameter). Finally we determine $M \in \{2, \ldots, 10\}$ which minimises the resulting cost of the estimator.

Using this fully automatic procedure we compare the empirical L^2–error or empirical root mean squared error (RMSE) of both estimators for $\varepsilon = 2^{-k}$, $k = 1, \ldots, 8$. The empirical RMSE is computed using 200 replications of the estimator with the indicated reference price. In figure 1 (a) and (b) are depicted the CPU–time of one estimator with respect to the empirical RMSE. The speed-up of ML2R over MLMC is significant on these two highly biased examples. For $\varepsilon = 0.1$, the ML2R run about 33 times faster than MLMC in the Barrier option pricing example and about 10 times faster in the nested Monte Carlo example. For more details on the numerical aspects we refer to [55].



(A) Barrier option discretized by an Euler scheme

(B) Compound option pricing by a nested Monte Carlo

FIGURE 1. Log-log plot of the CPU–time in seconds as a function of empirical L^2–error (RMSE).
Each point corresponds to a prescribed value of the expected L^2–error $\varepsilon = 2^{-k}$ for $k = 1, \ldots, 8$.

3. Analysis of the Adaptive Multilevel Splitting algorithms for rare event simulation

3.1. Introduction

The estimation of rare events is a crucial problem in many applications and has been the subject of many studies in the last decades, since the pioneering works on Monte-Carlo methods in the 1950s; see for instance Chapter 6 in [31], [38], [63]. For simplicity, we mainly focus on the problem of the estimation of the probability of a rare event, although it is not the only interesting question: for instance, in the metastability context, one also wishes to understand the behavior of the so-called reactive trajectories. We start with a description of the crude Monte-Carlo method and of its limitations. We then briefly review some reduction of variance strategies, especially the Multilevel Splitting approach introduced in [42] and revisited recently, starting from [30].

We then focus on adaptive versions of the latter, and describe its properties in two different situations. Firstly, in an idealized setting, we state the algorithm and we emphasize on the unbiasedness of the estimator of the probability. Secondly, we explain how the algorithm should be modified to handle a situation which involves a Markov chain dynamics: we numerically demonstrate that unbiasedness is preserved provided that the adaptive approach is appropriately implemented.

3.2. Crude Monte-Carlo method

Assume that X is a non-negative, real random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and that given $a > 0$, we want to estimate the probability $p = \mathbb{P}(X > a)$. When $a \to +\infty$, the probability p converges to 0.

We can use the Monte-Carlo procedure: if $(X_m)_{m \in \mathbb{N}^*}$ denotes a sequence of independent and identically distributed random variables, with the same law as X, for any integer $M \in \mathbb{N}^*$, the probability p can be approximated by the empirical average

$$\overline{p}_M^{\mathrm{MC}} = \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{X_m > a}.$$

It is clear that $\mathbb{E}[\overline{p}_M^{\mathrm{MC}}] = \mathbb{E}[\mathbb{1}_{X>a}] = p$ (thus $\overline{p}_M^{\mathrm{MC}}$ is an unbiased estimator of p, for any fixed value of M) and that $\mathrm{Var}(\overline{p}_M^{\mathrm{MC}}) = \frac{p(1-p)}{M}$. Moreover, when $M \to +\infty$, the Central Limit Theorem implies that $\sqrt{M}\left(\overline{p}_M^{\mathrm{MC}} - p\right)$ converges in law to a Gaussian distribution $\mathcal{N}(0, p(1-p))$, with variance $p(1-p)$; asymptotic (symmetric) confidence intervals for the estimation of p are thus of the form $[\overline{p}_M^{\mathrm{MC}} - \frac{q_\alpha \sigma_M}{M}, \overline{p}_M^{\mathrm{MC}} + \frac{q_\alpha \sigma_M}{M}]$, where q_α denotes

the quantile of level α of the standard Gaussian distribution: $1-\alpha = \mathbb{P}(|Z| \leq q_\alpha)$ if $Z \sim \mathcal{N}(0,1)$, and $\sigma_M^2 = \frac{1}{M}\sum_{m=1}^M \left(\mathbb{1}_{X_m>a} - \overline{p}_M^{\mathrm{MC}}\right)^2 \to p(1-p)$ is the empirical variance.

The length of the confidence interval goes to 0 when $p \to 0$; however, meaningful interpretations can only be given using the relative error $\epsilon_M^{\mathrm{MC}} = \frac{q_\alpha \sigma_M}{\overline{p}_M^{\mathrm{MC}}\sqrt{M}} \sim \frac{q_\alpha\sqrt{p(1-p)}}{p\sqrt{M}}$. For a given tolerance $\epsilon > 0$, the control $\epsilon_M^{\mathrm{MC}} \leq \epsilon$ requires a number of realizations M of the order $1/p \to +\infty$, which starts to be prohitive on standard computers when (approximately) $p \leq 10^{-9}$, especially when the simulation of a single X_m has a non-negligible cost.

3.3. **Reduction of variance strategies**

Two main strategies were initiated in the 1950s: importance sampling and importance splitting.

We can apply the basic importance sampling strategy for the approximation of $p = \mathbb{E}[\mathbb{1}_{X>a}]$. The (zero-variance) optimal change of measure $\tilde{\mathbb{P}}$ is given by the conditional distribution $\mathbb{P}(\cdot|X>a)$ given the (rare) event $\{X>a\}$. As usual in this context, sampling according to this distribution requires the knowledge of the unknown quantity p - which appears here as a normalizing constant. See [31], Sections 5 and 6.1.

For specific models, some importance sampling strategies have been proposed, based for instance on large deviations theory. However the introduction of an importance function always modifies the model, which is inappropriate for complex systems - think of an industrial code, that you might not be allowed to view and modify. Instead, we prefer to see the model as a black box and to propose non-intrusive strategies.

3.4. **Multilevel Splitting**

The principle is as follows. We split the (small) probability $p = \mathbb{P}(X>a)$ into a (telescoping) product of conditional probabilities which are easier to estimate. The latter are defined using an increasing sequence of levels $a_0 < \ldots < a_n$, with $a_0 = 0$ and $a_n = a$: we have $p = \mathbb{P}(X>a_0)\prod_{j=1}^n \mathbb{P}(X>a_j|X>a_{j-1})$. Here the sequence of levels must be fixed *a priori*, but its choice is crucial for the improvement of the estimation.

We define $p_0 = \mathbb{P}(X>a_0)$ and $p_j = \mathbb{P}(X>a_j|X>a_{j-1})$ for $1 \leq j \leq n$, and we estimate independently each p_j thanks to the Crude Monte-Carlo procedure introduced above. Each computation involves an ensemble of size $M \in \mathbb{N}^*$ (independent of j for simplicity), and the associated estimator of p_j is denoted by $\overline{p}_M^{(j)}$. Notice that in practice a way to sample according to conditional distributions of X knowing $X > a_{j-1}$ is required.

Since the random variables $(\overline{p}_M^{(j)})_{0\leq j\leq n}$ are independent by construction, the quantity $\overline{p}^{\mathrm{MS}} = \prod_{j=0}^n \overline{p}_M^{(j)}$ is an unbiased estimator of $\prod_{j=0}^n p_j = p$. One easily compute that when $M \to +\infty$ the variance satisfies $M\mathrm{Var}\left(\overline{p}^{\mathrm{MS}}\right) \to (n+1)p^2\left(-1 + \frac{1}{n+1}\sum_{j=1}^{n+1}\frac{1}{p_j}\right)$. Under the constraint $\prod_{j=0}^n p_j = p$, the asymptotic variance is then minimized for a fixed n when $p_j = p^{1/(n+1)}$ for any $0 \leq j \leq n$; then the optimum for n fixed brings $M\mathrm{Var}\left(\overline{p}^{\mathrm{MS}}\right) \to (n+1)p^2\left(p^{-1/(n+1)} - 1\right) \geq -p^2\log(p)$, the infimum being obtained when $n \to +\infty$.

We thus need to choose n as large as possible, and then the levels a_j such that $p_j = p^{1/(n+1)}$.

In practice, we cannot *a priori* choose the levels in a nearly optimal way, such that the conditional probabilities are all equal, without any further knowledge of the distribution of X. One way to improve the Multilevel Splitting algorithm is to compute random levels adaptively.

3.5. **Adaptive Multilevel Splitting, ideal case**

Many Adaptive versions of the Multilevel Splitting approach have been proposed in the recent years, and are found with different names in the literature; see for instance [3], [20], [23], [22], [24] , [21], [34], [65], [64].

Let $X_1, \ldots, X_n$ be i.i.d; we assume that the common distribution has a continuous cumulative distribution function (see Assumption 1). If $m^n = \min(X_i)_{1\leq i\leq n}$, we have the equality $\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i>m^n} = 1 - \frac{1}{n}$; although m^n is a random quantity, it can be interpreted as a level a_1, which gives a probability $p_0 = 1 - \frac{1}{n}$. The basic version (with $k=1$ below) of the AMS algorithm uses this idea recursively, sampling according to conditional distributions at each iteration, until the current level is above a.

More generally, $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i > m_k^n} = 1 - \frac{k}{n}$ where m_k^n denotes the k-th order statistics of $X_1, \dots, X_n$. The generalization of the algorithm for $k > 1$ is motivated by several arguments: it could give a higher reduction of variance than for $k = 1$; and in fact it is necessary to understand it to implement correctly the AMS algorithm in the case of discrete time Markov processes, see Section 3.6.

3.5.1. *The ideal case*

We now review the results from [16] and [15]. The algorithm 1 defined below is well-posed and yields a family of unbiased estimators under the following condition.

Assumption 1. *X is a real-valued positive random variable which admits a continuous cumulative distribution function (c.d.f.) $F : t \mapsto \mathbb{P}(X \leqslant t)$.*

For any $x \in \mathbb{R}^+$, we denote by $\mathcal{L}(X|X > x)$ the distribution of X conditional on $\{X > x\}$. In the ideal case, we make the additional assumption that we know in practice how to sample from $\mathcal{L}(X|X > x)$.

3.5.2. *The AMS algorithm*

The definition of the Adaptive Multilevel Splitting Method (see Algorithm 1) depends on several parameters given by the user:

- $n \in \mathbb{N} \setminus \{0, 1\}$ the number of replicas, which remains constant at each iteration;
- $k \in \{1, \dots, n-1\}$ the number of resampled particles at each iteration of the algorithm.

For any $k \in \{1, \dots, n\}$, we write $Y_{(k)}$ for the k-th order statistics independent and identically distributed real valued random variables $Y_1, \dots, Y_n$, with continuous c.d.f: almost surely $Y_{(1)} < \dots < Y_{(k)} < \dots < Y_{(n)}$.

Algorithm 1 (Adaptive Multilevel Splitting)**.**
Initialization: *Define $Z^0 = 0$. Sample n i.i.d. realizations $X_1^0, \dots, X_n^0$, with the law $\mathcal{L}(X)$.*
Define $Z^1 = X_{(k)}^0$, the k-th order statistics of the sample $X^0 = (X_1^0, \dots, X_n^0)$, and σ^1 the (a.s.) unique associated permutation: $X_{\sigma^1(1)}^0 < \dots < X_{\sigma^1(n)}^0$.
Set $j = 1$.
Iterations (on $j \geq 1$): *While $Z^j < a$:*
Conditional on Z^j, sample k independent random variables $(\chi_1^j, \dots, \chi_k^j)$ distributed with $\mathcal{L}(X|X > Z^j)$.
Resample particles as follows: if $(\sigma^j)^{-1}(i) \leq k$ (i.e. $X_i^{j-1} \leq Z^j$), then $X_i^j = \chi_{(\sigma^j)^{-1}(i)}^j$; else $X_i^j = X_i^{j-1}$.
Define $Z^{j+1} = X_{(k)}^j$, the k-th order statistics of $X^j = (X_1^j, \dots, X_n^j)$, and σ^{j+1} the associated permutation.
Finally increment $j \leftarrow j + 1$.
End of the algorithm: *Define $J^{n,k} = j - 1$ as the (random) number of iterations; thus $Z^{J^{n,k}} < a \leq Z^{J^{n,k}+1}$.*

The estimator of the probability p is defined by

$$\hat{p}^{n,k} = \frac{1}{n}\mathrm{Card}\left\{i;\, X_i^{J^{n,k}} \geqslant a\right\}\left(1 - \frac{k}{n}\right)^{J^{n,k}}. \tag{7}$$

3.5.3. *Theoretical results*

The main result from [15] is that the AMS algorithm 1 yields unbiased estimators given by (7) for the probability p, for any values of the parameters n and k: in particular no asymptotic regime is required.

Theorem 3.1. *For any $n \in \mathbb{N}^*$, $k \in \{1, \dots, n-1\}$, and $a > 0$, such that $p = \mathbb{P}(X > a) > 0$, $\hat{p}^{n,k}$ is an unbiased estimator of the probability p:*

$$\mathbb{E}[\hat{p}^{n,k}] = p. \tag{8}$$

To analyze how efficient the algorithm is depending on n and k, one can look at the variance; more precisely, in the regime when a and k are fixed and $n \to +\infty$, a Central Limit Theorem holds, see [16]:

Theorem 3.2. *We have the following convergence in law, when $n \to +\infty$, and for fixed k and a (such that $p = \mathbb{P}(X > a) > 0$):*

$$\sqrt{n}\big(\hat{p}^{n,k} - p\big) \to \mathcal{N}\big(0, -p^2 \log(p)\big). \tag{9}$$

We observe that the asymptotic variance $-p^2 \log(p)$ is the same as for the Multilevel Splitting algorithm with fixed, optimally chosen levels; moreover it does not depend on k. Notice that this result justifies a reduction of the variance with respect to the Crude Monte-Carlo procedure: the relative error is now of size $\frac{q_\alpha p\sqrt{-\log(p)}}{pn}$, and the required number of interacting replicas n is of size $\sqrt{-\log(p)} = \mathrm{o}\left(\frac{1}{p}\right)$.

For this adaptive algorithm, a better notion of cost involves the mean number of iterations - which is of order n/k: we prove in [15] that this cost is minimal when $k = 1$. Therefore our analysis reveals that taking $k > 1$ in the ideal case produces no improvement compared with $k = 1$; however it is an important first step for the definition of appropriate algorithms in more complicated situations, for which the unbiasedness is preserved, see Section 3.6.

Notice that when $k = 1$ the proof of Theorem 3.1 was already given in [34], where it was established that $J^{n,1}$ follows a Poisson distribution with parameter $-n \log(p)$; the proof of Theorem 3.2 in this case also follows by simple computations and the use of the delta-method.

The fundamental argument in [15] and [16] is the embedding of the estimation problem of p into the problem of estimating the conditional probability $p(x) = \mathbb{P}(X > a | X > x)$ for any $x \in [0, a]$; for that purpose, we introduce the estimator $\hat{p}^{n,k}(x)$ obtained thanks to Algorithm 1 where in the initialization step the random variables are generated according to $\mathcal{L}(X|X > x)$. Then Theorem 3.1 is a consequence of two facts:

- $p^{n,k} : x \in [0, a] \mapsto \mathbb{E}[\hat{p}^{n,k}(x)]$ is a solution of the functional equation in q (for some function $\theta_p^{n,k}$)

$$q(x) = \int_x^a \left(1 - \frac{k}{n}\right) q(y) f_{n,k}(y; x) dy + \theta_p^{n,k}(x), \tag{10}$$

  where $f_{n,k}(\cdot; x)$ is the density of the k-th order statistics in a sample of size n with law $\mathcal{L}(X|X > x)$;
- $x \in [0, a] \mapsto p(x) = \mathbb{P}(X > a | X > x)$ is the unique solution of (10).

Equation (10) has a very natural interpretation: after the first iteration, if the current level Z is lower than a, we can consider that the algorithm restarts at Z; the factor $\left(1 - \frac{k}{n}\right)$ accounts for the first iteration.

To obtain the Central Limit Theorem 3.2 and to derive cost estimates, a more refined analysis is necessary, carried out in the case when the random variable X is exponentially distributed. This case is generic since thanks to Assumption 1, we have $U = F(X)$ is uniformly distributed in $(0, 1)$, and $-\log(U)$ is exponentially distributed. No close expressions for solutions of appropriate functional equations is available for finite n, and we rely on an asymptotic analysis of some linear ODEs.

3.6. **Beyond the ideal case**

The content of this section is based on [14].

3.6.1. *A Markov chain*

Now we would like to explain how to implement a modified AMS scheme based on Algorithm 1, in the case where the random variable X depends on the evolution of a Markov chain $\overline{X} = (X_\ell)_{\ell \in \mathbb{N}}$, given by

$$X_{\ell+1} = X_\ell - \Delta t + \sqrt{2\beta^{-1}\Delta t} G_\ell \quad X_0 = x_0, \tag{11}$$

where the random variables $(G_\ell)_{\ell \in \mathbb{N}}$ are independent standard Gaussian random variables. In the numerical simulations below we choose $\Delta t = 0.1$. The parameter $\beta > 0$ can be interpreted as an inverse temperature.

Notice that this dynamics corresponds to a discretization of the following SDE $dX_t = -dt + \sqrt{2\beta^{-1}} dB_t$ using the Euler-Maruyama method, with a time-step size $\Delta t > 0$.

Given $b_1 < b_2$, such that $x_0 \in (b_1, b_2)$, we want to estimate the probability $p = p(x_0, \beta) = \mathbb{P}(\tau_{b_2} < \tau_{b_1})$, where $\tau_{b_2} = \inf\{\ell \geqslant 1; X_\ell > b_2\}$ and $\tau_{b_1} = \inf\{\ell \geqslant 1; X_\ell < b_1\}$. From now on $x_0 = 1$, $b_1 = 0.1$ and $b_2 = 1.9$.

It is clear that for a fixed initial condition x_0, $p(x_0, \beta) \to 0$ when $\beta \to +\infty$. Notice that $p = \mathbb{P}(X \geq b_2)$, where $X = \sup_{0 \leq \ell \leq \min(\tau_{b_1}, \tau_{b_2})} X_\ell$. Therefore X takes values in $[x_0, b_2]$, and we have $\mathbb{P}(X = x_0) > 0$ and $\mathbb{P}(X = b_2) = p > 0$; the latter identities imply that the c.d.f. of X is not continuous.

Another essential difference with the above ideal framework is that we are not able to sample according to conditional distributions; instead we split the trajectories defining the replicas. The resampling step in Algorithm 1 is then modified as follows: to sample each of the new trajectories, we pick at random one of the non-resampled replicas, that we call $\overline{Y} = (Y_\ell)_{\ell \in \mathbb{N}}$. Then we sample $\overline{\mathcal{X}} = (\mathcal{X}_\ell)_{\ell \in \mathbb{N}}$ by a splitting of $\overline{Y}$ at a (stopping) time T: we define $\mathcal{X}_\ell = Y_\ell$ if $\ell \leq T$, and then sample a new trajectory starting at Y_T using the Markov dynamics (11). If the current level is Z, we propose two choices for T, which lead to different algorithms due to possible equalities of random variables X_i^j:

$$T^- = \inf\{\ell \geqslant 1; Y_\ell \geqslant Z\} \quad \text{or} \quad T^+ = \inf\{\ell \geqslant 1; Y_\ell > Z\}.$$

Using $T = T^-$ (defined with a large inequality), we define an Algorithm, which is referred to as AMS($-$). In order to use $T = T^+$, we observe that due to possible equalities, we have to pick at random a trajectory $\overline{Y}$ such that $T^+ < +\infty$; if it happens that no replica satisfies this condition, the algorithm stops and we set the estimator to 0. The associated algorithm is referred to as AMS($+$).

It turns out that both modified algorithms AMS($-$) and AMS($+$) lead to a biased estimation of the probability p. A reason for this is that we might have $Z^j = Z^{j+1}$ at some step j, because of possible equalities of levels. To correct this issue, a random number of replicas $K^j \geq k$ has to be killed and resampled at each iteration j: if we define $K^j = \sup\left\{i \in \{1, \dots, n\}; X_{(i)}^{j-1} \leq Z^j\right\}$, we now kill replicas such that $(\sigma^j)^{-1}(i) \leq K^j$, and sample K^j new trajectories using the splitting strategy with the time T^+. Notice that we can have $K^j = n$: in this case no replica can be split, the algorithm stops and the estimator is set to 0.

We thus obtain an *unbiased* AMS algorithm. If in the definition (7) of the estimator $\hat{p}^{n,k}$ we replace $\left(1 - \frac{k}{n}\right)^{J^{n,k}}$ with the product $\prod_{j=1}^{J^{n,k}} \left(1 - \frac{K^j}{n}\right)$, we still get an unbiased estimator of p.

3.6.2. *Numerical tests*

We start with numerical results for the Generalized AMS algorithm, in Table 3.1 for $\beta = 8$ and $\beta = 24$.

Results are obtained with $n \in \{10, 50, 100\}$ and $k = 1$; in the case $n = 50$, we also take $k = 10$. The estimator $\overline{p}_N$ is defined as a Monte-Carlo average over an ensemble of size $N = 6.10^6$: namely

$$\overline{p}_N = \frac{1}{N} \sum_{m=1}^{N} \hat{p}_m^{n,k}$$

where $(\hat{p}_m^{n,k})_{1 \leq m \leq N}$ are computed using independent realizations, with the parameters n and k. We also include the relative error $\epsilon_N = 100 \frac{1.96 \sigma_N}{\sqrt{N} \overline{p}_N}$ (in percent) and the length of the confidence intervals: $\delta_N = 2 \frac{1.96 \sigma_N}{\sqrt{N}}$, where the empirical variance is $\sigma_N^2 = \frac{1}{N} \sum_{m=1}^{N} \left(\hat{p}_m^{n,k} - \overline{p}_N\right)^2$.

The results when $\beta = 8$ can be compared with the probability estimated with a direct Monte-Carlo estimation, with a sample of size $N' = 6.10^8$: we estimate $\overline{p}_{N'}^{\text{CMC}} = 3.595\ 10^{-4}$, with a relative error 0.2%; the corresponding confidence interval is $[3.580\ 10^{-4}, 3.610\ 10^{-4}]$ and contains all the values of Table 3.1.

We also give a comparison with the algorithms AMS($-$) and AMS($+$) in Table 3.2, where $k = 1$.

Notice that the choice of a relatively large value for the latter is motivated by the following consideration: if one of the algorithms introduces a bias in the estimation of the probability, it should increase when the time-step Δt increase - since in the limit $\Delta t \to 0$ we recover the idealized case, which exhibits no bias. As a consequence, it is easier to exhibit a bias (if it exists) than for a smaller value of Δt. Therefore, Table 3.1 can be interpreted as numerical validation of the theoretical unbiasedness result for the unbiased AMS algorithm, while Table 3.2

n	10	50	50	100	10	50	50	100
k	1	1	10	1	1	1	10	1
β	8	8	8	8	24	24	24	24
$\overline{p}_N$	$3.597\ 10^{-4}$	$3.596\ 10^{-4}$	$3.596\ 10^{-4}$	$3.597\ 10^{-4}$	$1.20\ 10^{-10}$	$1.22\ 10^{-10}$	$1.21\ 10^{-10}$	$1.21\ 10^{-10}$
δ_N	$0.012\ 10^{-4}$	$0.004\ 10^{-4}$	$0.004\ 10^{-4}$	$0.003\ 10^{-4}$	$0.29\ 10^{-10}$	$0.03\ 10^{-10}$	$0.04\ 10^{-10}$	$0.01\ 10^{-10}$
ϵ_N	0.09 %	0.03 %	0.03 %	0.02 %	6 %	0.6 %	0.8 %	0.2 %

TABLE 3.1. Results for the unbiased AMS algorithm

version	AMS(+)	AMS(+)	AMS(−)	AMS(+)	AMS(+)	AMS(−)
n	10	100	100	10	100	100
β	8	8	8	24	24	24
$\overline{p}_N$	$2.958\ 10^{-4}$	$3.257\ 10^{-4}$	$1.741\ 10^{-4}$	$5.079\ 10^{-11}$	$6.049\ 10^{-11}$	$1.397\ 10^{-12}$
δ_N	$0.012\ 10^{-4}$	$0.003\ 10^{-4}$	$0.034\ 10^{-4}$	$1.523\ 10^{-11}$	$0.066\ 10^{-11}$	$0.142\ 10^{-12}$

TABLE 3.2. Results for the modified algorithms AMS(−) and AMS(+)

exhibits the bias of the other modifications AMS(−) and AMS(+) - leading to a very poor underestimated value, with possibly several orders of magnitude of difference.

3.7. **Conclusion and perspectives**

We have reviewed here some recent results on the Adaptive Multilevel Splitting Algorithm for the estimation of a rare event probability. We have focused in Section 3.5 on the so-called ideal case and on the consistency property: we obtain a family of unbiased estimators.

However, this ideal case is quite restrictive both theoretically - the cumulative distribution function must be continuous - and from a practical point of view - we need to be able to sample according to conditional distributions. Most of the interesting cases are excluded, such as the problem involving a discrete time Markov process described above, in dimension possibly greater than 1.

The unbiased AMS algorithm introduced in Section 3.6 allows us to preserve the unbiasedness; moreover the numerical simulations indicate that the modifications made in the algorithm are crucial. Further theoretical and numerical investigations are contained in [14], especially in higher dimensional situations. The algorithm requires to use an importance function ξ taking values in $\mathbb{R}$: the unbiasedness holds independently of the choice of ξ, but it is an open important question to study its impact on the efficiency.

Finally, unbiasedness of the unbiased AMS algorithm is not restricted to the estimation of the probability: it also provides information about the so-called reactive trajectories, see for instance [13] for preliminary simulations for the stochastic Allen-Cahn equation.

4. A CUBATURE BASED ALGORITHM TO SOLVE MCKEAN-VLASOV FORWARD AND DECOUPLED FORWARD-BACKWARD STOCHASTIC DIFFERENTIAL EQUATION

We shortly describe how the cubature algorithm of Lyons and Victoir [56] can be adapted in order to compute expectation of some functional of the solution of a decoupled Mc Kean-Vlasov Forward Backward Stochastic Differential Equation (MKV-FBSDE) type. This work has been done with C.A. Garcìa Trillos and we refer to [25] for more details.

4.1. **The system**

We call a McKean-Vlasov Forward Backward Stochastic Differential equation (MKV-FBSDE in short) the following FBSDE system:

$$\left\{\begin{array}{l} dX_t = \sum_{i=0}^{d} V_i(t, X_t, \mathbb{E}\varphi_i(X_t)) \circ dB_t^i \\ dY_t = -f(s, X_t, Y_t, Z_t, \mathbb{E}\varphi_f(X_t, Y_t))dt + Z_t dB_t^{1:d} \ - \\ X_0 = x, \quad Y_T = \phi(X_T) \end{array}\right. \tag{12}$$

for any t in $[0,T]$, $T>0$ be given. Here $B_t^{1:d} := (B_t^1, \cdots, B_t^d)^*,\ t \geq 0$ (where "$*$" stands for the transpose) is a d-dimensional Brownian motion defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ and $B_t^0 = t$. The notation "$\circ$" stands for the Stratonovitch integral. The $\{V_i,\ i=0,\cdots,d\}$ is a family of smooth[1] vector fields from $[0,T]\times\mathbb{R}^d\times\mathbb{R}$ to $\mathbb{R}$ as well as the functions $\varphi_i,\ i=0,\cdots,d$ from $\mathbb{R}^d$ to $\mathbb{R}$, φ_f from $\mathbb{R}^d\times\mathbb{R}$ to $\mathbb{R}$ and the mapping f from $[0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}$ to $\mathbb{R}$. The function ϕ is also supposed to be a smooth function from $\mathbb{R}^d$ to $\mathbb{R}$.

These processes may be regarded as a limit approximation for interacting systems with large number of particles. They appeared initially in statistical mechanics, but are now used in many fields because of the big range of applications requiring large populations interactions. For example, they are used in finance, as factor stochastic volatility models [10] or uncertain volatility models [36]; in economics, in the theory of "mean field games" recently developed by J.M. Lasry and P.L. Lions in a series of papers [50–53] (see the work of Carmona and Delarue [18] for the probabilistic analysis) and also in physics, neuroscience, dynamic population models etc... Here, this system can be also viewed as the probabilistic representation of a control problem of a marked player in a mean field environment [25].

The note of Sznitman [66] gives a beautiful overview on the topic of systems with a large number of particles. A proof on the existence and uniqueness of such a process can be found in [17] (see also [19] for more general setting).

4.2. **A cubature based algorithm for MKV-FBSDE processes**

We give the main idea to construct a cubature based approximation scheme for computing quantities such as $\mathbb{E}[\phi(X_T^x)]$ or $Y_t^{t,x},\ t\in[0,T]$ defined in (12).

4.2.1. *Preliminaries: Cubature on Wiener space.*

Cubature on Wiener space has been introduced by Lyons and Victoir [56]. Let us shortly described how this method works.

For a given positive integer m and a positive real T, we call m-cubature on Wiener space (the set of continuous functions from $[0,T]$ to $\mathbb{R}^d$) a discrete probability measure $\mathbb{Q}_T$ with finite support on the set of continuous function from $[0,T]$ to $\mathbb{R}^d$ with bounded variations such that the expectations of the Stratonovitch iterated integrals up to order m are the same under the Wiener and the cubature measure.

Then, instead of solving an SDE (in Stratonovitch form)

$$dX_t = \sum_{i=0}^{d} V_i(X_t) \circ dB_t^i,$$

on $[0,T]$, we solve a system of weighted ordinary differential equations having the same form as this SDE but where the Brownian motion is replaced by a continuous with bounded variation path $\omega_j : [0,T] \to \mathbb{R}^d$ associated

[1] say bounded and infinitely differentiable with bounded derivatives of any order.

to a weight λ_j, $j \in \{1, \dots, n\}$ where $n = \text{card}\{\text{supp}(\mathbb{Q}_T)\}$:

$$d\hat{X}_t^j = V_0(\hat{X}_t^j)dt + \sum_{i=1}^{d} V_i(\hat{X}_t^j)dw_j^i(t), \quad j = 1, \dots, n,$$

on $[0, T]$, with the convention $w_j^0(t) = t$.

It is well seen that such method allows to approach the expectation of smooth functional of the SDE's solution by polynomials : by using the definition of the cubature, together with a Taylor expansion we obtain that, for all smooth function F:

$$(\mathbb{E}_{\mathbb{P}} - \mathbb{E}_{\mathbb{Q}_T})[F(X_T)] \leq CT^{(m+1)/2} \max_{j \leq m+2} \left\| \frac{\partial^j}{\partial x^j} F \right\|_{\infty}, \tag{13}$$

by using classical estimates on the remainder of the Taylor stochastic expansion, see [44] for example.

Obviously, if the constant C above, or if the time length interval T are not small enough, the error could be large. In order to take benefit from the method, we have to use the Markovian structure of the process : we do not apply the cubature on all the time interval, but on each small step of a subdivision $T_0 = 0 < T_1 < \cdots < T_N = T$ of that interval.

This approach leads to the construction of a tree, each node of the tree having a number n (depending on m) of child. We call this object a cubature tree.

4.2.2. *Cubature for MKV-FBSDE*

The main issue in the case of a MKV-FBSDE is the McKean-Vlasov dependence that appears in the coefficients. This dependence breaks the Markov property (in the usual sense) of the process so that it is not possible to apply, a priori, many classical analysis tools. In order to handle this problem, the idea consists in taking benefit on the following crucial observation: *given the law of the solution of the system,* (12) *is a classical time inhomogeneous FBSDE* (the law just acts as a time dependent parameter).

Let $(\nu_t)_{0 \leq t \leq T}$ be a family of probability measures on $\mathbb{R}^d$, and let us fix the law in the McKean-Vlasov terms of (12) to be $(\nu_t)_{0 \leq t \leq T}$. For this modified system, we may apply a classical cubature FBSDE scheme (the time dependence of the coefficients being handled as an additional dimension). The trick consists in taking advantage of the decoupled setting: we first build a cubature tree (depending of the order of the cubature) and then go back along the nodes of the tree by computing the backward process as conditional expectation of each node. We refer to [26] or [27] for a detailed description of such algorithm.

Obviously, at each step of the scheme, we pay the price of using an arbitrary probability measure as parameter for the coefficients instead of the law of the process. Therefore this law has to be chosen carefully in order to keep a good control on the error and achieve convergence.

A first remark allows to reduce the problem: as in the classical FBSDE case, we can show that there exists a deterministic decoupling field $u : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ such that, for all t in $[0, T]$,

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

Hence, the family of probability measures $(\nu_t)_{0 \leq t \leq T}$ has to be chosen as a proxy of the law of the forward component only. A "good choice" is then to take at each step of the cubature tree the discrete law $\hat{\mu}$ given by the Dirac measures at the solutions of the ODEs along the cubature paths and the corresponding aggregated weights.

4.2.3. *A simple example*

Before giving our results, we explain the main idea of it by means of an example on dimension one, using a cubature of order $m = 3$, for a discretization with 2 steps. In this case, we may use a cubature formula with

$n = 2$ paths given by $\{+t, -t\}$, and associated weights: $\{\lambda_1 = 1/2, \lambda_2 = 1/2\}$. The algorithm is illustrated in Figure 2. We initialize the tree at a given point x, and the law at T_0 as δ_x. Then, we find two descendants given as the solution of an ODE that uses the position X, the two cubature paths, and an approximated law using the information at time 0. Each descendant will have a weight equal to the product of the weight of its parent times the weight given to the corresponding cubature path. Once all nodes at time T_1 are calculated, we obtain the discrete measure $\hat{\mu}_{T_1}$, the law approximation at time T_1. The process is then repeated for each node at time T_1 to reach the final time $T = T_2$.

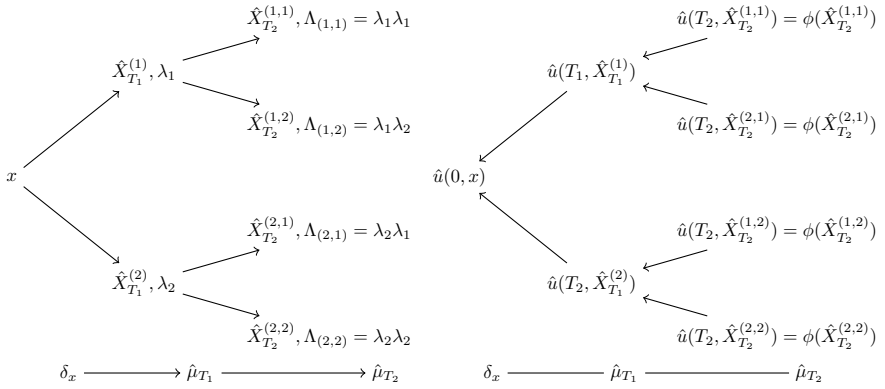


FIGURE 2. Left: Cubature tree. Right: Backward scheme.

Figure 2 right illustrates the idea behind the backward approximation: the approximated function $\hat{u}$ is defined first at the leaves of the constructed tree, and then back-propagates using the approximated law to obtain $\hat{u}$ at previous times. The back-propagation is made by conditional expectation: average with respect to the weight of each cubature path.

4.2.4. *Results*

In [25] it is shown that this algorithm with this choice of approximation law leads to a first order approximation scheme of Y (more precisely, the infinity norm of the approximation error decreases linearly with the number of discretization steps). It is also possible to obtain higher orders of approximation by correcting some terms in the algorithm and to relax the regularity of ϕ in (12) to Lipschitz, provided the vector fields are uniformly elliptic. In all these cases, the algorithm can be parametrized in order to obtain an approximation of the law of the forward component of any order.

5. Approximation of Markov semigroup in total variation distance

5.1. **Introduction**

Framework. We consider the d dimensional Markov chain

$$X^n_{k+1} = \psi_{k+1}(X^n_k, \frac{Z_{k+1}}{\sqrt{n}}), \qquad k = 0, \cdots, n-1 \tag{14}$$

where $(\psi_k)_{k\in\{1,\cdot,n\}}$ is a sequence of smooth functions from: $\mathbb{R}^d \times \mathbb{R}^N$ to $\mathbb{R}^d$ such that $\psi(x, 0) = x$ and $Z_k \in \mathbb{R}^N$, $k \in \{1, \cdots, n\}$, is a sequence of independent random variables. We denote $\mu^n_{k+1}(x, dy) = P(X^n_{k+1} \in dy \mid X^n_k = x)$ the transition kernels of this chain. We fix $\delta = 1/n$ and consider the time grid $t_k = k\delta$. Moreover we introduce a Markov process $(X_t)_{t\geqslant 0}$ and define $\mu_{k+1}(x, dy) = P(X_{t_{k+1}} \in dy \mid X_{t_k} = x)$. Finally we denote by $(P^n_k)_{k\in\{0,\cdot,n\}}$ (respectively $(P_{t_k})_{k\in\{0,\cdot,n\}}$) the discrete semigroup induced by μ^n (resp. μ) such that for $f : \mathbb{R}^d \mapsto \mathbb{R}$, $P^n_0 f = f$ and $P^n_{k+1} f = \mu^n_{k+1} P^n_k f$ (resp. $P_0 f = f$ and $P_{t_{k+1}} f = \mu_{k+1} P_{t_k} f$).

Preliminary result. A first standard result is the following. Assume that there exists $h > 0$, $p \in \mathbb{N}$ such that for every $f \in \mathcal{C}^p(\mathbb{R}^d)$, $k \in \{1, \cdots, n\}$,

$$\left| \int f(y)\mu_k^n(x, dy) - \int f(y)\mu_k(x, dy) \right| \leqslant C\|f\|_{p,\infty}\delta^{1+h} \tag{15}$$

where $\|f\|_{p,\infty}$ designates the supremum norm of f and of its derivatives up to order p. Then, for every $T > 0$,

$$\max_{t_k \leq T} |\mathbb{E}[f(X_k^n)] - \mathbb{E}[f(X_{t_k})]| \leqslant C\|f\|_{p,\infty}\delta^h. \tag{16}$$

It means that $(X_k)_{k\in\{0,\cdot,n\}}$ is an approximation scheme of order h for the Markov process $(X_t)_{t\geqslant 0}$. In the case of the Euler scheme for diffusion processes, this result, with $h = 1$, has first been proved in the seminal papers of Milstein [57] and of Talay and Tubaro [68] (see also [44]). A very rich literature on this topic followed. Similar results were obtained in various situations: diffusion processes with jumps (see [62], [39]) or diffusion processes with boundary conditions (see [33], [32], [11]) for instance. See [41] for an overview of this subject. More recently, approximation schemes of higher order ($h = 2$), based on a cubature method, have been introduced and studied by Kusuoka [48], Lyons [56], Victoir, Ninomiya [58], Alfonsi [1], Kohatsu-Higa and Tankov [45]. Our abstract result intends to cover all these situations.

5.2. **Estimates of the error in the total variation distance**

Our first result concerns convergence in total variation distance: we want to obtain (16) with $\|f\|_{p,\infty}$ replaced by $\|f\|_\infty$. In the case of the Euler scheme for diffusion processes, a first result of this type has been obtained by Bally and Talay [8], [9] using the Malliavin calculus (see also Guyon [35]). Afterwords Konakov, Menozzi and Molchanov [46], [47] obtained similar results using a parametrix method. Recently Kusuoka [49] obtained estimates of the error in total variation distance for the Victoir Ninomiya scheme. Our second result establishes the rate of convergence of the density function and its derivatives.

*Abstract regularization properties.*We first remark that the crucial property which allows to replace $\|f\|_{p,\infty}$ by $\|f\|_\infty$ in (16) is the regularization property of the semigroup. Let us give a precise statement. Let $\eta > 0$, $p \in \mathbb{N}$ be fixed and f a $\mathbb{R}$-valued measurable function on $\mathbb{R}^d$. We say that a semigroup $(P_k)_{k\in\{0,\cdot,n\}}$ satisfies $R_{p,\eta}$ if

$$R_{p,\eta} \qquad \|P_k f\|_{p,\infty} \leqslant \frac{C}{t_k^\eta}\|f\|_\infty . \tag{17}$$

We consider also a dual regularization property: let P^* be the dual semigroup of P (that is $\langle P_k^* g, f\rangle = \langle g, P_k f\rangle$ with the scalar product in $L^2(\mathbb{R}^d)$). The semigroup P satisfies $R_{p,\eta}^*$ if

$$R_{p,\eta}^* \qquad \|P_k^* f\|_{p,1} \leqslant \frac{C}{t_k^\eta}\|f\|_1 , \tag{18}$$

where $\|f\|_{p,1}$ designates the L^1 norm of f and of its derivatives up to order p. Finally, we give a stronger assumption $\overline{R}_{p,\eta}$ which implies both $R_{p,\eta}$ and $R_{p,\eta}^*$. We say that $\overline{R}_{p,\eta}$ is satisfied if for all multi-indexes α and β with $|\alpha| + |\beta| = p$, we have

$$\overline{R}_{p,\eta} \qquad \left\|\partial^\alpha P_k \partial^\beta f\right\|_\infty \leqslant \frac{C}{t_k^{\eta(|\alpha|+|\beta|)}}\|f\|_\infty . \tag{19}$$

In addition to (15), we will also suppose that the following dual estimate of the error in short time holds between the measure μ and μ^n:

$$|\langle g, (\mu_k^n - \mu_k)f\rangle| \leqslant C\,\|g\|_{p,1}\,\|f\|_\infty\,\delta^{1+h}. \tag{20}$$

Having this properties at hand we can state the following result.

Theorem. *We fix T, $h>0$, $p\in\mathbb{N}$ and we assume that the short time estimates (15) and (20) hold (with this p and h). Moreover, we assume that (17) holds for $(P_{t_k})_{k\in\{0,\cdot,n\}}$ and that (18) holds for $(P_k^n)_{k\in\{0,\cdot,n\}}$. Then, there exists some constants $C>0$ such that*

$$\forall 0<S\leqslant T,\quad \sup_{S\leqslant t_k\leqslant T}\|P_{t_k}f-P_k^nf\|_\infty\leqslant\frac{C}{S^{\eta p}}\delta^h\|f\|_\infty. \tag{21}$$

Approximation Markov chain. In this section we focus on the approximation scheme and the dual regularization property. We are not able to prove $R^*_{p,\eta}$ for $(P_k^n)_{k\in\{0,\cdot,n\}}$ but we can overcome this issue. Indeed, we introduce the following modification (regularization) of P^n : we define

$$\forall k\in\{0,\cdots,n\},\quad \overline{X}_k^n=\delta^{1+h}G+X_k^n, \tag{22}$$

with G a standard normal random variable independent from X^n and we define also

$$\overline{P}_k^nf(x)=\mathbb{E}[f(\overline{X}_k^n)|\overline{X}_0^n=x]. \tag{23}$$

We will be able to prove that $\overline{P}^n$ satisfies $\overline{R}_{p,\eta}$ and moreover $\overline{P}_k^nf$ is close to P_k^nf in total variation distance. Combining these two properties we can deduce that (29) holds even if P^n does not satisfy $R^*_{p,\eta}$. Then, the following step consists in giving sufficient conditions on the approximation scheme to obtain such a $\overline{P}^n$. We will use some integration by parts formulas of Malliavin type based on the noise $Z_k\in\mathbb{R}^N$. In order to do it, we assume that the law of each Z_k is lower bounded by the Lebesgue measure: there exists some $z_k\in\mathbb{R}^N$ and $r_*,\varepsilon_*>0$ such that for every measurable set $A\subset B_{r_*}(z_k)$ one has

$$P(Z_k\in A)\geqslant\varepsilon_*\lambda(A) \tag{24}$$

where λ is the Lebesgue measure. If this property holds then one may use a "splitting method" in order to represent Z_k as

$$\frac{Z_k}{\sqrt{n}}=\chi_kU_k+(1-\chi_k)V_k$$

where χ_k,U_k,V_k are independent random variables, χ_k is a Bernoulli random variable and $\sqrt{n}U_k\sim\varphi_{r_*}(u)du$ with $\varphi_{r_*}\in\mathcal{C}^\infty(\mathbb{R}^N)$. This decomposition allows us to use the abstract Malliavin calculus based on U_k, developed in [4] and [5], in order to obtain integration by parts formulas and then to prove $\overline{R}_{p,\eta}$. The crucial point is that the density φ_{r_*} of $\sqrt{n}U_k$ is smooth and we control its logarithmic derivatives. This allows to derive integration by parts formulas and to obtain convenient estimates for the weights which appear in these formulas. It is worth mentioning that a variant of the Malliavin calculus based on a similar splitting method has already been used by Nourdin and Poly [59] (see also [60] and [54]). They use the so-called Γ-calculus (see *e.g.* [7] for background). Finally we will make the following assumptions

$$\forall p\in\mathbb{N},\quad 1\wedge\max_{k\in\{0,\cdot,n\}}\mathbb{E}[|Z_k|^p]<\infty, \tag{25}$$

$$\forall r\in\mathbb{N}^*,\quad \sup_{k\in\{0,\cdots,n\}}\sum_{1\leqslant|\beta|\leqslant r}\sum_{0\leqslant|\alpha|\leqslant r-|\beta|}\sup_{x\in\mathbb{R}^d}\sup_{z\in\mathbb{R}^N}|\partial_x^\alpha\partial_z^\beta\psi_k|<\infty, \tag{26}$$

$$\exists\lambda_*>0,\ \forall k\in\{1,\cdots,n\},\quad \inf_{x\in\mathbb{R}^d}\inf_{|\zeta|=1}\sum_{i=1}^N\langle\partial_{z_i}\psi_k(x,0),\zeta\rangle^2\geqslant\lambda_*. \tag{27}$$

Distance estimations. Now we can state our main result.

Theorem. *Consider a Markov semigroup P_{t_k} and its approximation scheme P_k^n, $k \in \{0, \cdots, n\}$ defined above. We fix T, $h > 0$, $p \in \mathbb{N}$ and we assume that the short time estimates (15) and (20) hold (with this p and h). Moreover, we assume (24), (25), (26) and (27) and that (17) holds for the semigroup P_{t_k}. Then, we have the following results.*

A. *The regularization $\overline{P}^n$ of P^n defined by (23) satisfies*

$$\sup_{k\in\{0,\cdot,n\}} \|P_k^n f - \overline{P}_k^n f\|_\infty \leqslant C\delta^h \|f\|_\infty. \tag{28}$$

B. *For every $S > 0$ we have*

$$\sup_{S\leqslant t_k\leqslant T} \|P_{t_k} f - P_k^n f\|_\infty \leqslant \frac{C}{S^{\eta p}} \delta^h \|f\|_\infty . \tag{29}$$

C. *We denote $\overline{p}_k^n$ the density funtion of $\overline{P}_k^n$. For every $t > 0$, $P_t(x, dy) = p_t(x, y)dy$ with $(x, y) \to p_t(x, y)$ belonging to $\mathcal{C}^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ and for every $R, S, \varepsilon > 0$ and every multi-indexes α, β we have*

$$\sup_{S\leqslant t_k\leqslant T} \sup_{|x|+|y|\leqslant R} \left|\partial_x^\alpha \partial_y^\beta p_{t_k}(x, y) - \partial_x^\alpha \partial_y^\beta \overline{p}_k^n(x, y)\right| \leqslant C\delta^{h(1-\varepsilon)} \tag{30}$$

with a constant C which depends on R, S, ε and on $|\alpha| + |\beta|$ (and may blow up as $\varepsilon \downarrow 0$).

In contrast with (16) and (29) which have a rather elementary proof once we have the appropriate regularization properties, the estimate (30) is based on a non trivial interpolation result from [6]. Notice however that those estimates are sub-optimal (because of $\varepsilon > 0$).

5.3. The Ninomiya Victoir scheme

We illustrate this theorem when X^n is the Ninomiya Victoir scheme for a diffusion process. This is a variant of the result already obtained by Kusuoka [49] in the case where Z_k has a Gaussian distribution (and so the standard Malliavin calculus is available). Since in our paper Z_k has an arbitrary distribution, our result may be seen as an invariance principle as well. We consider the d dimensional diffusion process

$$dX_t = \sum_{i=1}^N V_i(X_t) \circ dW_t^i + V_0(X_t)dt \tag{31}$$

with $V_0, V_i \in \mathcal{C}_b^\infty(\mathbb{R}^d; \mathbb{R}^d)$, $i = 1, \cdots, N$ and $W = (W^1, \cdots, W^N)$ a Brownian motion and $\circ dW_t^i$ denotes the Stratonovich integral with respect to W^i. The infinitesimal operator of this Markov process is

$$A = V_0 + \frac{1}{2}\sum_{k=1}^N V_k^2 \tag{32}$$

with the notation $Vf(x) = \langle V(x), \nabla f(x)\rangle$. Let us define $\exp(V)(x) := \Phi_V(x, 1)$ where Φ_V solves the deterministic equation $\Phi_V(x, t) = x + \int_0^t V(\Phi_V(x, s))ds$. We present now the Ninomiya Victoir scheme. We consider a sequence ρ_k, $k \in \mathbb{N}$ of independent Bernoulli random variables and we define $\psi_k : \mathbb{R}^d \times \mathbb{R}^{N+1} \to \mathbb{R}^d$ in the following way

$$\psi_k(x, w) = \exp(w^0 V_0) \circ \exp(w^1 V_1) \circ \cdot \circ \exp(w^N V_N) \circ \exp(w^0 V_0)(x), \quad \text{if } \rho_k = 1,$$
$$\psi_k(x, w) = \exp(w^0 V_0) \circ \exp(w^N V_N) \circ \cdot \circ \exp(w^1 V_1) \circ \exp(w^0 V_0)(x), \quad \text{if } \rho_k = -1.$$

Here $w = (w^0, w^1, \cdots, w^N)$. with $w_k^0 = 1/2n$, $w_k^i = Z_k^i/\sqrt{n}$, for $i = 1, \cdots, N$. Moreover $Z_k^i,\ i = 1, \cdots, d,\ k \in \mathbb{N}$ are independent random variables which verify (24) and moreover satisfy the following moment conditions:

$$E(Z_k^i) = E((Z_k^i)^3) = 0, \qquad E((Z_k^i)^2) = 1, \qquad E((Z_k^i)^4) = 6. \tag{33}$$

In the original paper of Ninomiya Victoir, the random variables Z_k^i are standard normal distributed, and then verify (24). The new point here is that we do not require that Z_k follows this particular law anymore but only the weaker assumptions (24) and (33). We also denote $t_k = Tk/n$. One step of our scheme is given by

$$X_{k+1}^n = \psi_{k+1}(X_k^n, w_{k+1}). \tag{34}$$

Under an ellipticity condition we are able to give an estimate of the total variation distance between a diffusion process of the form (31) and its Ninomiya Victoir scheme.

Theorem. *Suppose that $V_i \in \mathcal{C}_b^\infty(\mathbb{R}^d; \mathbb{R}^d), i = 0, ..., N$ and moreover*

$$\inf_{|\xi|=1} \sum_{i=1}^N \langle V_i(x), \xi \rangle^2 \geqslant \lambda_* > 0 \qquad \forall x \in R^d. \tag{35}$$

Then for every $0 < S \leqslant T$ and every bounded and measurable function $f : \mathbb{R}^d \to \mathbb{R}$

$$\sup_{S \leqslant t_k \leqslant T} |E(f(X_{t_k})) - E(f(X_k^n))| \leqslant \frac{C}{S^{1/2}} \delta^2 \|f\|_\infty . \tag{36}$$

References

[1] Aurélien Alfonsi. High order discretization schemes for the CIR process: application to affine term structure and Heston models. *Math. Comp.*, 79(269):209–237, 2010.

[2] A. Alfonsi, B. Jourdain, and A. Kohatsu-Higa. Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme. *The Annals of Applied Probability*, (To appear in), 2013.

[3] S.K. Au and J.L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Journal of Probabilistic Engineering Mechanics*, 16:263–277, 2001.

[4] V. Bally and E. Clément. Integration by parts formula and applications to equations with jumps. *Probab. Theory Related Fields*, 151(3-4):613–657, 2011.

[5] V. Bally and L. Caramellino. On the distance between probability density functions. November 2013.

[6] V. Bally and L. Caramellino. Convergence and regularity of probability laws by using an interpolation method. *arXiv preprint arXiv:1409.3118*, 2014.

[7] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[8] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104(1):43–60, 1996.

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

[10] L. Bergomi. Smile dynamics III. SSRN Scholarly Paper ID 1493308, Social Science Research Network, Rochester, NY, October 2009.

[11] M. Bossy, E. Gobet, and D. Talay. A symmetrized Euler scheme for an efficient approximation of reflected diffusions. *J. Appl. Probab.*, 41(3):877–889, 2004.

[12] M. Broadie, Y. Du, and C.C. Moallemi. Efficient risk estimation via nested sequential simulation. *Management Science*, 57(6):1172–1194, 2011.

[13] C.E. Bréhier, M. Gazeau, L. Goudenège and M. Rousset. Analysis and simulations of rare events for SPDEs. *ESAIM Proceedings and Reviews*, 48:364–384, 2015.

[14] C.E. Bréhier, M. Gazeau, L. Goudenège, T. Lelièvre and M. Rousset. Unbiasedness of some generalized Adaptive Multilevel Splitting algorithms. *arXiv preprint arXiv:1505.02674*, 2015.

[15] C.E. Bréhier, T. Lelièvre and M. Rousset. Analysis of adaptive multilevel splitting algorithms in an idealized case. *ESAIM P&S*, in press.

[16] C.E. Bréhier, L. Goudenège and L. Tudela. Central limit theorem for adaptive multilevel splitting estimators in an idealized setting. *arXiv preprint arXiv:1501.01399*, 2015.
[17] R. Buckdahn, J. Li, and S. Peng. Mean-field backward stochastic differential equations and related partial differential equations. *Stochastic Processes and their Applications*, 119(10):3133–3154, 2009.
[18] R. Carmona and F. Delarue. Probabilistic analysis of mean-field games. *SIAM J. Control Optim.*, 51(4): 2705–2734, 2013.
[19] R. Carmona and F. Delarue. Mean field forward-backward stochastic differential equations. *Electron. Commun. Probab.*, 18(68), 15 p., 2013.
[20] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Stat. Comput.*, 22(3):795–808, 2012.
[21] F. Cérou, P. Del Moral, A. Guyader and F. Malrieu. Fluctuations of adaptive multilevel splitting. *arXiv preprint arXiv:1408.6366*, 2014.
[22] F. Cérou and A. Guyader. Adaptive particle techniques and rare event estimation. *Conference Oxford sur les méthodes de Monte Carlo séquentielles, ESAIM Proc.*, 19.
[23] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.*, 25(2):417–443, 2007.
[24] F. Cérou, A. Guyader, T. Lelièvre and D. Pommier. A multiple replica approach to simulate reactive trajectories. *J. Chem. Phys.*, 134:054108.
[25] P. E. Chaudru de Raynal and C. A. Garcia Trillos. A cubature based algorithm to solve decoupled McKean-Vlasov forward backward stochastic differential equations. arXiv e-print 1307.6427, July 2013.
[26] D. Crisan and K. Manolarakis. Solving backward stochastic differential equations using the cubature method. application to nonlinear pricing. In *Progress in analysis and its applications*, page 389–397. World Sci. Publ., Hackensack, NJ, 2010.
[27] Dan Crisan and Konstantinos Manolarakis. Second order discretization of backward SDEs. *Ann. Appl. Probab.*, 24(2) 652–678, 2014.
[28] L. Devineau and S. Loisel. Construction d'un algorithme d'accélération de la méthode des "simulations dans les simulations" pour le calcul du capital économique Solvabilité II. *Bulletin Français d'Actuariat*, 10(17):188–221, 2009.
[29] M.B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.
[30] P. Glasserman, P. Heidelberger, P. Shahabuddin and T. Zajic. Multilevel splitting for estimating rare event probabilities. *Operations Research*, 47(4):585–600, 1999.
[31] P. W. Glynn. S. Asmussen. *Stochastic simulation: Algorithms and analysis*. Springer, 2007.
[32] E. Gobet and S. Menozzi. Stopped diffusion processes: boundary corrections and overshoot. *Stochastic Process. Appl.*, 120(2):130–162, 2010.
[33] E. Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic Process. Appl.*, 87(2):167–197, 2000.
[34] A. Guyader, N. Hengartner and E. Matzner-Løber. Simulation and estimation of extreme quantiles and extreme probabilities. *Appl. Math. Optim.*, 64(2):171–196, 2011.
[35] J. Guyon. Euler scheme and tempered distributions. *Stochastic Processes and their Applications*, pages 877–904, 2006.
[36] J. Guyon and P.H. Labordere. The smile calibration problem solved. SSRN Scholarly Paper ID 1885032, Social Science Research Network, Rochester, NY, July 2011.
[37] S. Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.
[38] J.M. Hammersley and D.C. Handscomb. *Monte Carlo methods*. Methuen & Co. Ltd., 1965.
[39] J. Jacod, T.G. Kurtz, S. Méléard, and P. Protter. The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(3):523–558, 2005.
[40] L. Jeff Hong and S. Juneja. Estimating the mean of a non-linear function of conditional expectation. In M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference*, pages 1223–1236. IEEE, 2009.
[41] B. Jourdain and A. Kohatsu-Higa. *A Review of Recent Results on Approximation of Solutions of Stochastic Differential Equations*, volume 65 of *Progress in Probability*. Springer Basel, 2011.
[42] H. Kahn and T.E. Harris. Estimation of particle transmission by random sampling. *National Bureau of Standards, Appl. Math. Series*, 12:27–30, 1951.
[43] A. Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *Ann. Appl. Probab.*, 15(4):2681–2705, 2005.
[44] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
[45] A. Kohatsu-Higa and P. Tankov. Jump-adapted discretization schemes for Lévy-driven SDEs. *Stochastic Process. Appl.*, 120(11):2258–2285, 2010.
[46] V. Konakov and S. Menozzi. Weak error for stable driven stochastic differential equations: expansion of the densities. *J. Theoret. Probab.*, 24(2):454–478, 2011.

[47] V. Konakov, S. Menozzi, and S. Molchanov. Explicit parametrix and local limit theorems for some degenerate diffusion processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(4):908–923, 2010.
[48] S. Kusuoka. Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. In *Advances in mathematical economics. Vol. 6*, volume 6 of *Adv. Math. Econ.*, pages 69–83. Springer, Tokyo, 2004.
[49] S. Kusuoka. Gaussian K-scheme: justification for KLNV method. In *Advances in mathematical economics. Vol. 17*, volume 17 of *Adv. Math. Econ.*, pages 71–120. Springer, Tokyo, 2013.
[50] J. M. Lasry and P.L. Lions. Jeux à champ moyen. i. le cas stationnaire. *Comptes Rendus Mathématique. Académie des Sciences. Paris*, 343(9):619–625, 2006.
[51] J. M. Lasry and P.L. Lions. Jeux à champ moyen. II. horizon fini et contrï¿½le optimal. *Comptes Rendus Mathématique. Académie des Sciences. Paris*, 343(10):679–684, 2006.
[52] J. M. Lasry and P.L. Lions. Large investor trading impacts on volatility. *Annales de l'Institut Henri Poincaré. Analyse Non Linéaire*, 24(2):311–323, 2007.
[53] J.M. Lasry and P.L. Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.
[54] M. Ledoux, I. Nourdin, and G. Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *ArXiv e-prints*, March 2014.
[55] V. Lemaire and G. Pagès. Multilevel richardson-romberg extrapolation. *arXiv preprint arXiv:1401.1177*, 2014.
[56] T. Lyons and N. Victoir. Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 460(2041):169–198, 2004. Stochastic analysis with applications to mathematical finance.
[57] G.N. Milstein. Weak approximation of solutions of systems of stochastic differential equations. In *Numerical Integration of Stochastic Differential Equations*, volume 313 of *Mathematics and Its Applications*, pages 101–134. Springer Netherlands, 1995.
[58] S. Ninomiya and N. Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance*, 15(1-2):107–121, 2008.
[59] I. Nourdin and G. Poly. An invariance principle under the total variation distance. 15 pages, October 2013.
[60] I. Nourdin, G. Peccati, and Y. Swan. Entropy and the fourth moment phenomenon. *J. Funct. Anal.*, 266(5):3170–3207, 2014.
[61] G. Pagès. Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity. *Monte Carlo Methods Appl.*, 13(1):37–70, 2007.
[62] P. Protter and D. Talay. The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, 25(1):393–423, 1997.
[63] G. Rubino and B. Tuffin. *Rare Event Simulation using Monte Carlo Methods*. Wiley, 2009.
[64] E. Simonnet. Combinatorial analysis of the adaptive last particle method. *Statistics and Computing*, pages 1–20, 2014.
[65] J. Skilling. Nested sampling for general bayesian computation. *Bayesian Analysis*, 1(4):833–859, 2006.
[66] A.S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX-1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.
[67] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.
[68] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.