Finance Stochast. 6, 449–471 (2002)

Finance and Stochastics

An analysis of a least squares regression method for American option pricing

Emmanuelle Clément[1], **Damien Lamberton**[1], **Philip Protter**[2]

[1] Équipe d'Analyse et de mathématiques appliquées, Université de Marne-la-Vallée, 5 Bld Descartes, Champs-sur-marne, 77454 Marne-la-Vallée Cedex 2, France
[2] Operations Research and Industrial Engineering Department, Cornell University, Ithaca, NY 14853-3801, USA (e-mail: protter@orie.cornell.edu)

Abstract. Recently, various authors proposed Monte-Carlo methods for the computation of American option prices, based on least squares regression. The purpose of this paper is to analyze an algorithm due to Longstaff and Schwartz. This algorithm involves two types of approximation. Approximation one: replace the conditional expectations in the dynamic programming principle by projections on a finite set of functions. Approximation two: use Monte-Carlo simulations and least squares regression to compute the value function of approximation one. Under fairly general conditions, we prove the almost sure convergence of the complete algorithm. We also determine the rate of convergence of approximation two and prove that its normalized error is asymptotically Gaussian.

Key words: American options, optimal stopping, Monte-Carlo methods, least squares regression

JEL Classification: G10, G12, G13

Mathematics Subject Classification (1991): 90A09, 93E20, 60G40

1 Introduction

The computation of American option prices is a challenging problem, especially when several underlying assets are involved. The mathematical problem to solve is

---

The research on this paper has been stimulated by the "projet mathfi" seminar on the topic "Monte-Carlo methods for American options", during the fall 2000. The talks given by P. J.F. Delmas, B. Jourdain, E. Temam were especially helpful. (Supported in part by NSF Grant #DMS-9971720 and NSA Grant #MDA904-00-1-0035)

Manuscript received: April 2001; final version received: January 2002

an optimal stopping problem. In classical diffusion models, this problem is associated with a variational inequality, for which, in higher dimensions, classical PDE methods are ineffective.

Recently, various authors introduced numerical methods based on Monte-Carlo techniques (see, among others, [1–5, 9, 12]). The starting point of these methods is to replace the time interval of exercise dates by a finite subset. This amounts to approximating the American option by a so called *Bermudan* option. A control of the error caused by this restriction to discrete stopping times is generally easy to obtain (see, for instance, [8], Remark 1.4). Throughout the paper, we concentrate on the discrete time problem.

The solution of the discrete optimal stopping problem reduces to an effective implementation of the dynamic programming principle. The conditional expectations involved in the iterations of dynamic programming cause the main difficulty for the development of Monte-Carlo techniques. One way of treating this problem is to use least squares regression on a finite set of functions as a proxy for conditional expectation. This idea (which already appeared in [5]) is one of the main ingredients of two recent papers by Longstaff and Schwartz [9], and by Tsitsiklis and Van Roy [12].

The purpose of the present paper is to analyze the least squares regression method proposed by Longstaff and Schwartz [9], which seems to have become popular among practitioners. In fact, we will consider a variant of their approach (see Remark 2.1). In order to present our results more precisely, we will distinguish two types of approximation in their algorithm. Approximation one: replace conditional expectations in the dynamic programming principle by projections on a finite set of functions taken from a suitable basis. Approximation two: use Monte-Carlo simulations and least squares regression to compute the value function of the first approximation. Approximation two will be referred to as *the Monte-Carlo procedure*. In practice, one chooses the number of basis functions and runs the Monte-Carlo procedure.

We will prove that the value function of approximation one approaches (with probability one) the value function of the initial optimal stopping problem as the number m of functions goes to infinity. We then prove that for a fixed finite set of functions, we have almost sure convergence of the Monte-Carlo procedure to the value function of the first approximation. We also establish a type of central limit theorem for the rate of convergence of the Monte-Carlo procedure, thus providing the asymptotic normalized error. We note that partial convergence results are stated in [9], together with excellent empirical results, but with no study of the rate of convergence. On the other hand, convergence (but not the rate nor the error distribution) is provided in [12] for a somewhat different algorithm. We also refer to [12] for a discussion of accumulation of errors as the number of possible exercise dates rows. We believe that our methods could be applied to analyze the rate of convergence of the Tsitsiklis-Van Roy approach, but we will concentrate on the Longstaff-Schwartz method.

Mathematically, the most technical part of our work concerns the Central Limit Theorem for the Monte-Carlo procedure. One might think that the methods developed for the analysis of asymptotic errors in statistical estimation based on stochastic optimization (see, for instance, [6, 7, 11]) are applicable to our problem. However, the algorithm does not seem to fit in this setting for two reasons: the lack of regularity of the value function as a function of the parameters and the recursive nature of dynamic programming.

The paper is organized as follows. In Sect. 2, a precise description of the least squares regression method is given and the notation is established. In Sect. 3, we prove the convergence of the algorithm. In Sect. 4, we study the rate of convergence of the Monte-Carlo procedure.

2 The algorithm and notations

2.1 Description of the algorithm

As mentioned in the introduction, the first step in all probabilistic approximation methods is to replace the original optimal stopping problem in continuous time by an optimal stopping problem in discrete time. Therefore, we will present the algorithm in the context of discrete optimal stopping.

We will consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, equipped with a discrete time filtration $(\mathcal{F}_j)_{j=0,\ldots,L}$. Here, the positive integer L denotes the (discrete) time horizon. Given an adapted payoff process $(Z_j)_{j=0,\ldots,L}$, where Z_0, Z_1,..., Z_L are square integrable random variables, we are interested in computing

$$\sup_{\tau\in\mathcal{T}_{0,L}} \mathbb{E} Z_\tau,$$

where $\mathcal{T}_{j,L}$ denotes the set of all stopping times with values in $\{j,\ldots,L\}$.

Following classical optimal stopping theory (for which we refer to [10], Chap. 6), we introduce the Snell envelope $(U_j)_{j=0,\ldots,L}$ of the payoff process $(Z_j)_{j=0,\ldots,L}$, defined by

$$U_j = \text{ess-}\sup_{\tau\in\mathcal{T}_{j,L}} \mathbb{E}\left(Z_\tau \mid \mathcal{F}_j\right), \quad j = 0,\ldots,L.$$

The dynamic programming principle can be written as follows:

$$\begin{cases} U_L = Z_L \\ U_j = \max\left(Z_j, \mathbb{E}\left(U_{j+1} \mid \mathcal{F}_j\right)\right), \quad 0 \le j \le L-1. \end{cases}$$

We also have $U_j = \mathbb{E}\left(Z_{\tau_j} \mid \mathcal{F}_j\right)$, with

$$\tau_j = \min\{k \ge j \mid U_k = Z_k\}.$$

In particular $\mathbb{E}U_0 = \sup_{\tau\in\mathcal{T}_{0,L}} \mathbb{E}Z_\tau = \mathbb{E}Z_{\tau_0}$.

The dynamic programming principle can be rewritten in terms of the optimal stopping times τ_j, as follows:

$$\begin{cases} \tau_L = L \\ \tau_j = j\mathbf{1}_{\{Z_j \ge \mathbb{E}(Z_{\tau_{j+1}}\mid\mathcal{F}_j)\}} + \tau_{j+1}\mathbf{1}_{\{Z_j < \mathbb{E}(Z_{\tau_{j+1}}\mid\mathcal{F}_j)\}}, \quad 0 \le j \le L-1, \end{cases}$$

This formulation in terms of stopping rules (rather than in terms of value functions) plays an essential role in the least squares regression method of Longstaff and Schwartz.

The method also requires that the underlying model be a Markov chain. Therefore, we will assume that there is an $(\mathcal{F}_j)$-Markov chain $(X_j)_{j=0,\dots,L}$ with state space $(E, \mathcal{E})$ such that, for $j = 0, \dots, L$,

$$Z_j = f(j, X_j),$$

for some Borel function $f(j, \cdot)$. We then have $U_j = V(j, X_j)$ for some function $V(j, \cdot)$ and $\mathbb{E}\left(Z_{\tau_{j+1}} \mid \mathcal{F}_j\right) = \mathbb{E}\left(Z_{\tau_{j+1}} \mid X_j\right)$. We will also assume that the initial state $X_0 = x$ is deterministic, so that U_0 is also deterministic.

The first approximation consists of approximating the conditional expectation with respect to X_j by the orthogonal projection on the space generated by a finite number of functions of X_j. Let us consider a sequence $(e_k(x))_{k \geq 1}$ of measurable real valued functions defined on E and satisfying the following conditions:

$\mathbf{A_1}$: For $j = 1$ to $L-1$, the sequence $(e_k(X_j))_{k \geq 1}$ is total in $L^2(\sigma(X_j))$.

$\mathbf{A_2}$: For $j = 1$ to $L-1$ and $m \geq 1$, if $\sum_{k=1}^{m} \lambda_k e_k(X_j) = 0$ a.s. then $\lambda_k = 0$ for $k = 1$ to m.

For $j = 1$ to $L-1$, we denote by P_j^m the orthogonal projection from $L^2(\Omega)$ onto the vector space generated by $\{e_1(X_j), \dots, e_m(X_j)\}$ and we introduce the stopping times $\tau_j^{[m]}$:

$$\begin{cases} \tau_L^{[m]} = L \\ \tau_j^{[m]} = j \mathbf{1}_{\left\{Z_j \geq P_j^m\left(Z_{\tau_{j+1}^{[m]}}\right)\right\}} + \tau_{j+1}^{[m]} \mathbf{1}_{\left\{Z_j < P_j^m\left(Z_{\tau_{j+1}^{[m]}}\right)\right\}}, \quad 1 \leq j \leq L-1, \end{cases}$$

From these stopping times, we obtain an approximation of the value function:

$$U_0^m = \max\left(Z_0, \mathbb{E} Z_{\tau_1^{[m]}}\right). \tag{2.1}$$

Recall that $Z_0 = f(0, x)$ is deterministic. The second approximation is then to evaluate numerically $\mathbb{E} Z_{\tau_1^{[m]}}$ by a Monte-Carlo procedure. We assume that we can simulate N independent paths $(X_j^{(1)}), \dots, (X_j^{(n)}), \dots (X_j^{(N)})$ of the Markov chain (X_j) and we denote by $Z_j^{(n)}$ ($Z_j^{(n)} = f(j, X_j^{(n)})$) the associated payoff for $j = 0$ to L and $n = 1$ to N. For each path n, we then estimate recursively the stopping times $(\tau_j^{[m]})$ by:

$$\begin{cases} \tau_L^{n,m,N} = L \\ \tau_j^{n,m,N} = j \mathbf{1}_{\left\{Z_j^{(n)} \geq \alpha_j^{(m,N)} \cdot e^m(X_j^{(n)})\right\}} & 1 \leq j \leq L-1, \\ \qquad + \tau_{j+1}^{n,m,N} \mathbf{1}_{\left\{Z_j^{(n)} < \alpha_j^{(m,N)} \cdot e^m(X_j^{(n)})\right\}}, \end{cases}$$

Here, $x \cdot y$ denotes the usual inner product in $I\!R^m$, e^m is the vector valued function $(e_1, \ldots, e_m)$ and $\alpha_j^{(m,N)}$ is the least square estimator:

$$\alpha_j^{(m,N)} = \arg \min_{a \in I\!R^m} \sum_{n=1}^{N} \left(Z^{(n)}_{\tau_{j+1}^{n,m,N}} - a \cdot e^m(X_j^{(n)}) \right)^2 ,$$

Remark that for $j = 1$ to $L-1$, $\alpha_j^{(m,N)} \in I\!R^m$. Finally, from the variables $\tau_j^{n,m,N}$, we derive the following approximation for U_0^m:

$$U_0^{m,N} = \max \left(Z_0, \frac{1}{N} \sum_{n=1}^{N} Z^{(n)}_{\tau_1^{n,m,N}} \right). \tag{2.2}$$

In the next section, we prove that, for any fixed m, $U_0^{m,N}$ converges almost surely to U_0^m as N goes to infinity, and that U_0^m converges to U_0 as m goes to infinity. Before stating these results, we devote a short section to notation. We also mention that the above algorithm is not exactly the Longstaff-Schwartz algorithm as their regression involves only in-the-money paths (see Remark 2.1).

2.2 Notation

Throughout the paper we denote by $|x|$ the Euclidean norm of a vector x in $I\!R^d$.

For $m \geq 1$ we denote by $e^m(x)$ the vector $(e_1(x), \ldots, e_m(x))$ and for $j = 1$ to $L - 1$ we define α_j^m so that

$$P_j^m(Z_{\tau_{j+1}^{[m]}}) = \alpha_j^m \cdot e^m(X_j) \tag{2.3}$$

We remark that, under A_2, the m dimensional parameter α_j^m has the explicit expression:

$$\alpha_j^m = (A_j^m)^{-1} I\!E(Z_{\tau_{j+1}^{[m]}} e^m(X_j)), \tag{2.4}$$

for $j = 1$ to $L - 1$, where A_j^m is an $m \times m$ matrix, with coefficients given by

$$(A_j^m)_{1 \leq k,l \leq m} = I\!E(e_k(X_j) e_l(X_j)). \tag{2.5}$$

Similarly, the estimators $\alpha_j^{(m,N)}$ are equal to

$$\alpha_j^{(m,N)} = (A_j^{(m,N)})^{-1} \frac{1}{N} \sum_{n=1}^{N} Z^{(n)}_{\tau_{j+1}^{n,m,N}} e^m(X_j^{(n)}), \tag{2.6}$$

for $j = 1$ to $L - 1$, where $A_j^{(m,N)}$ is an $m \times m$ matrix, with coefficients given by

$$(A_j^{(m,N)})_{1 \leq k,l \leq m} = \frac{1}{N} \sum_{n=1}^{N} e_k(X_j^{(n)}) e_l(X_j^{(n)}). \tag{2.7}$$

Note that $\lim_{N\to\infty} A_j^{(m,N)} = A_j^m$ almost surely. Therefore, under A_2, the matrix $A_j^{(m,N)}$ is invertible for N large enough. We also define $\alpha^m = (\alpha_1^m, \dots, \alpha_{L-1}^m)$ and $\alpha^{(m,N)} = (\alpha_1^{(m,N)}, \dots, \alpha_{L-1}^{(m,N)})$.

Given a parameter $a^m = (a_1^m, \dots, a_{L-1}^m)$ in $I\!R^m \times \dots \times I\!R^m$ and deterministic vectors $z = (z_1, \dots, z_L) \in I\!R^L$ and $x = (x_1, \dots, x_L) \in E^L$, we define a vector field $F = (F_1, \dots, F_L)$ by:

$$\begin{aligned} F_L(a^m, z, x) &= z_L \\ F_j(a^m, z, x) &= z_j \mathbf{1}_{\{z_j \ge a_j^m \cdot e^m(x_j)\}} + F_{j+1}(a^m, z, x) \mathbf{1}_{\{z_j < a_j^m \cdot e^m(x_j)\}}, \\ &\quad \text{for } j = 1, \dots, L-1. \end{aligned}$$

We have

$$F_j(a^m, z, x) = z_j \mathbf{1}_{B_j^c} + \sum_{i=j+1}^{L-1} z_i \mathbf{1}_{B_j \dots B_{i-1} B_i^c} + z_L \mathbf{1}_{B_j \dots B_{L-1}},$$

with

$$B_j = \{z_j < a_j^m \cdot e^m(x_j)\}.$$

We remark that $F_j(a^m, Z, X)$ does not depend on $(a_1^m, \dots, a_{j-1}^m)$ and that we have

$$\begin{aligned} F_j(\alpha^m, Z, X) &= Z_{\tau_j^{[m]}} \\ F_j(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}) &= Z^{(n)}_{\tau_j^{n,m,N}}. \end{aligned}$$

For $j = 2$ to L, we denote by G_j the vector valued function

$$G_j(a^m, z, x) = F_j(a^m, z, x) e^m(x_{j-1}),$$

and we define the functions ϕ_j and ψ_j by

$$\phi_j(a^m) = I\!E F_j(a^m, Z, X) \tag{2.8}$$

$$\psi_j(a^m) = I\!E G_j(a^m, Z, X). \tag{2.9}$$

Observe that with this notation, we have

$$\alpha_j^m = (A_j^m)^{-1} \psi_{j+1}(\alpha^m), \tag{2.10}$$

and similarly, for $j = 1$ to $L - 1$,

$$\alpha_j^{(m,N)} = (A_j^{(m,N)})^{-1} \frac{1}{N} \sum_{n=1}^{N} G_{j+1}(\alpha^{(m,N)}, Z^{(n)}, X^{(n)}). \tag{2.11}$$

Remark 2.1 In [9], the regression involves only in the money paths, i.e. samples for which $Z_j^{(n)} > 0$. This seems to be more efficient numerically. In order to stick to this type of regression, the above description of the algorithm should be modified as follows. Use

$$\hat{\tau}_j^{[m]} = j\mathbf{1}_{\{Z_j \geq \hat{\alpha}_j^m \cdot e(X_j)\} \cap \{Z_j > 0\}} + \hat{\tau}_{j+1}^{[m]} \mathbf{1}_{\{Z_j < \hat{\alpha}_j^m \cdot e(X_j)\} \cup \{Z_j = 0\}},$$

instead of $\tau_j^{[m]}$ for $j \leq L-1$, with

$$\hat{\alpha}_j^m = \arg\min_{a \in \mathbb{R}^m} \mathbb{E}\mathbf{1}_{\{Z_j > 0\}} \left(Z_{\hat{\tau}_{j+1}^{[m]}} - a \cdot e(X_j) \right)^2.$$

We analogously define $\hat{\tau}_j^{n,m,N}$, $\hat{\alpha}_j^{(m,N)}$, $\hat{F}_j$ and $\hat{G}_j$. The convergence results still hold for this version of the algorithm (with similar proofs), provided assumptions A_1 and A_2 are replaced by

$\hat{\mathbf{A}}_\mathbf{1}$: For $j = 1$ to $L-1$, the sequence $(e_k(X_j))_{k \geq 1}$ is total in $L^2(\sigma(X_j), \mathbf{1}_{\{Z_j > 0\}} d\mathbb{P})$.

$\hat{\mathbf{A}}_\mathbf{2}$: For $1 \leq j \leq L-1$ and $m \geq 1$, if $\mathbf{1}_{\{Z_j > 0\}} \sum_{k=1}^m \lambda_k e_k(X_j) = 0$ a.s. then $\lambda_k = 0$ for $1 \leq k \leq m$.

3 Convergence

The convergence of U_0^m to U_0 is a direct consequence of the following result.

Theorem 3.1 *Assume that A_1 is satisfied, then, for $j = 1$ to L, we have*

$$\lim_{m \to +\infty} \mathbb{E}(Z_{\tau_j^{[m]}} | \mathcal{F}_j) = \mathbb{E}(Z_{\tau_j} | \mathcal{F}_j),$$

in L^2.

Proof We proceed by induction on j. The result is true for $j = L$. Let us prove that if it holds for $j+1$, it is true for j $(j \leq L-1)$. Since

$$Z_{\tau_j^{[m]}} = Z_j \mathbf{1}_{\{Z_j \geq \alpha_j^m \cdot e^m(X_j)\}} + Z_{\tau_{j+1}^{[m]}} \mathbf{1}_{\{Z_j < \alpha_j^m \cdot e^m(X_j)\}},$$

for $j \leq L-1$, we have

$$\begin{aligned}
&\mathbb{E}(Z_{\tau_j^{[m]}} - Z_{\tau_j} | \mathcal{F}_j) \\
&= (Z_j - \mathbb{E}(Z_{\tau_{j+1}} | \mathcal{F}_j)) \left(\mathbf{1}_{\{Z_j \geq \alpha_j^m \cdot e^m(X_j)\}} - \mathbf{1}_{\{Z_j \geq \mathbb{E}(Z_{\tau_{j+1}} | \mathcal{F}_j)\}} \right) \\
&+ \mathbb{E}(Z_{\tau_{j+1}^{[m]}} - Z_{\tau_{j+1}} | \mathcal{F}_j) \mathbf{1}_{\{Z_j < \alpha_j^m \cdot e^m(X_j)\}}.
\end{aligned}$$

By assumption, the second term of the right side of the equality converges to 0 and we just have to prove that B_j^m defined by

$$B_j^m = (Z_j - \mathbb{E}(Z_{\tau_{j+1}} | \mathcal{F}_j)) \left(\mathbf{1}_{\{Z_j \geq \alpha_j^m \cdot e^m(X_j)\}} - \mathbf{1}_{\{Z_j \geq \mathbb{E}(Z_{\tau_{j+1}} | \mathcal{F}_j)\}} \right),$$

converges to 0 in L^2. Observe that

$$\begin{aligned}
|B_j^m| &= |Z_j - \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)||\mathbf{1}_{\{\mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j) > Z_j \geq \alpha_j^m \cdot e^m(X_j)\}} \\
&\quad - \mathbf{1}_{\{\alpha_j^m \cdot e^m(X_j) > Z_j \geq \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)\}}| \\
&\leq |Z_j - \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)|\mathbf{1}_{\{|Z_j - \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)| \leq |\alpha_j^m \cdot e^m(X_j) - \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)|\}} \\
&\leq |\alpha_j^m \cdot e^m(X_j) - \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)| \\
&\leq |\alpha_j^m \cdot e^m(X_j) - P_j^m(\mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j))| + |P_j^m(\mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)) \\
&\quad - \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)|.
\end{aligned}$$

But

$$\alpha_j^m \cdot e^m(X_j) = P_j^m(Z_{\tau_{j+1}^{[m]}}) = P_j^m(\mathbb{E}(Z_{\tau_{j+1}^{[m]}}|\mathcal{F}_j)),$$

since P_j^m is the orthogonal projection on a subspace of the space of $\mathcal{F}_j$-measurable random variables. Consequently

$$\begin{aligned}
||B_j^m||_2 &\leq ||\mathbb{E}(Z_{\tau_{j+1}^{[m]}}|\mathcal{F}_j) - \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)||_2 \\
&\quad + ||P_j^m(\mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)) - \mathbb{E}(Z_{\tau_{j+1}}|\mathcal{F}_j)||_2.
\end{aligned}$$

The first term of the right side of this inequality tends to 0 by the induction hypothesis and the second one by A_1. □

In what follows, we fix the value m and we study the properties of $U_0^{m,N}$ as N the number of Monte-Carlo simulations, goes to infinity. For notational simplicity, we drop the superscript m throughout the rest of the paper.

Theorem 3.2 *Assume that for $j = 1$ to $L - 1$, $\mathbb{P}(\alpha_j \cdot e(X_j) = Z_j) = 0$. Then $U_0^{m,N}$ converges almost surely to U_0^m as N goes to infinity. We also have almost sure convergence of $\frac{1}{N}\sum_{n=1}^{N} Z^{(n)}_{\tau_j^{n,m,N}}$ towards $\mathbb{E}Z_{\tau_j^{[m]}}$ as N goes to infinity, for $j = 1, \ldots, L$.*

Note that with the notation of the preceding section, we have to prove that

$$\lim_N \frac{1}{N}\sum_{n=1}^{N} F_j(\alpha^{(N)}, Z^{(n)}, X^{(n)}) = \phi_j(\alpha), \quad 1 \leq j \leq L. \tag{3.1}$$

The proof is based on the following lemmas.

Lemma 3.1 *For $j = 1$ to $L - 1$, we have:*

$$|F_j(a, Z, X) - F_j(b, Z, X)| \leq \left(\sum_{i=j}^{L} |Z_i|\right) \left(\sum_{i=j}^{L-1} \mathbf{1}_{\{|Z_i - b_i \cdot e(X_i)| \leq |a_i - b_i||e(X_i)|\}}\right).$$

Proof Let $B_j = \{Z_j < a_j \cdot e(X_j)\}$ and $\tilde{B}_j = \{Z_j < b_j \cdot e(X_j)\}$. We have:

$$\begin{aligned} F_j(a,Z,X) - F_j(b,Z,X) = Z_j(\mathbf{1}_{B_j} - \mathbf{1}_{\tilde{B}_j}) \\ + \sum_{i=j+1}^{L-1} Z_i(\mathbf{1}_{B_j \ldots B_{i-1}B_i^c} - \mathbf{1}_{\tilde{B}_j \ldots \tilde{B}_{i-1}\tilde{B}_i^c}) \\ + Z_L(\mathbf{1}_{B_j^c \ldots B_{L-1}^c} - \mathbf{1}_{\tilde{B}_j^c \ldots \tilde{B}_{L-1}^c}). \end{aligned}$$

But

$$\begin{aligned} |\mathbf{1}_{B_j} - \mathbf{1}_{\tilde{B}_j}| &= \mathbf{1}_{\{a_j \cdot e(X_j) \le Z_j < b_j \cdot e(X_j)\}} + \mathbf{1}_{\{b_j \cdot e(X_j) \le Z_j < a_j \cdot e(X_j)\}} \\ &\le \mathbf{1}_{\{|Z_j - b_j \cdot e(X_j)| \le |a_j - b_j||e(X_j)|\}} \end{aligned}$$

Moreover

$$\begin{aligned} |\mathbf{1}_{B_j \ldots B_{i-1}B_i^c} - \mathbf{1}_{\tilde{B}_j \ldots \tilde{B}_{i-1}\tilde{B}_i^c}| &\le \sum_{k=j}^{i-1} |\mathbf{1}_{B_k} - \mathbf{1}_{\tilde{B}_k}| + |\mathbf{1}_{B_i^c} - \mathbf{1}_{\tilde{B}_i^c}| \\ &= \sum_{k=j}^{i} |\mathbf{1}_{B_k} - \mathbf{1}_{\tilde{B}_k}|, \end{aligned}$$

this gives

$$|F_j(a,Z,X) - F_j(b,Z,X)| \le \sum_{i=j}^{L} |Z_i| \sum_{i=j}^{L-1} |\mathbf{1}_{B_i} - \mathbf{1}_{\tilde{B}_i}|.$$

Combining these inequalities, we obtain the result of Lemma 3.1. □

Lemma 3.2 *Assume that for $j = 1$ to $L-1$, $I\!P(\alpha_j \cdot e(X_j) = Z_j) = 0$ then $\alpha_j^{(N)}$ converges almost surely to α_j.*

Proof We proceed by induction on j. For $j = L-1$, the result is a direct consequence of the law of large numbers. Now, assume that the result is true for $i = j$ to $L-1$. We want to prove that it is true for $j-1$. We have

$$\alpha_{j-1}^{(N)} = (A_{j-1}^{(N)})^{-1} \frac{1}{N} \sum_{n=1}^{N} G_j(\alpha^{(N)}, Z^{(n)}, X^{(n)}).$$

By the law of large numbers, we know that $A_{j-1}^{(N)}$ converges almost surely to A_{j-1} and it remains to prove that $\frac{1}{N}\sum_{n=1}^{N} G_j(\alpha^{(N)}, Z^{(n)}, X^{(n)})$ converges to $\psi_j(\alpha)$. From the law of large numbers, we have the convergence of $\frac{1}{N}\sum_{n=1}^{N} G_j(\alpha, Z^{(n)}, X^{(n)})$ to $\psi_j(\alpha)$ and it suffices to prove that :

$$\lim_{N \to +\infty} \frac{1}{N} \sum_{n=1}^{N} \Big(G_j(\alpha^{(N)}, Z^{(n)}, X^{(n)}) - G_j(\alpha, Z^{(n)}, X^{(n)}) \Big) = 0.$$

We note $G_N = \frac{1}{N}\sum_{n=1}^{N}\Big(G_j(\alpha^{(N)}, Z^{(n)}, X^{(n)}) - G_j(\alpha, Z^{(n)}, X^{(n)})\Big)$. We have :

$$\begin{aligned}|G_N| &\leq \frac{1}{N}\sum_{n=1}^{N}|e(X_{j-1}^{(n)})||F_j(\alpha^{(N)}, Z^{(n)}, X^{(n)}) - F_j(\alpha, Z^{(n)}, X^{(n)})| \\ &\leq \frac{1}{N}\sum_{n=1}^{N}|e(X_{j-1}^{(n)})|\sum_{i=j}^{L}|Z_i^{(n)}|\sum_{i=j}^{L-1}\mathbf{1}_{\{|Z_i^{(n)}-\alpha_i\cdot e(X_i^{(n)})|\leq|\alpha_i^{(N)}-\alpha_i||e(X_i^{(n)})|\}}.\end{aligned}$$

Since, for $i = j$ to $L-1$, $\alpha_i^{(N)}$ converges almost surely to $\alpha_i^{(N)}$, we have for each $\epsilon > 0$:

$$\begin{aligned}&\limsup|G_N| \leq \limsup\frac{1}{N}\sum_{n=1}^{N}|e(X_{j-1}^{(n)})|\sum_{i=j}^{L}|Z_i^{(n)}|\sum_{i=j}^{L-1} \\ &\mathbf{1}_{\{|Z_i^{(n)}-\alpha_i\cdot e(X_i^{(n)})|\leq\epsilon|e(X_i^{(n)})|\}} \\ &= \mathbb{E}|e(X_{j-1})|\sum_{i=j}^{L}|Z_i|\sum_{i=j}^{L-1}\mathbf{1}_{\{|Z_i-\alpha_i\cdot e(X_i)|\leq\epsilon|e(X_i)|\}},\end{aligned}$$

where the last equality follows from the law of large numbers. Letting ϵ go to 0, we obtain the convergence to 0, since for $j = 1$ to $L-1$, $\mathbb{P}(\alpha_j\cdot e(X_j) = Z_j) = 0$. □

The Proof of Theorem 3.2 is similar to the proof of Lemma 3.2. Therefore, we omit it.

4 Rate of convergence of the Monte-Carlo procedure

4.1 Tightness

In this section we are interested in the rate of convergence of $\frac{1}{N}\sum_{n=1}^{N}Z_{\tau_j^{n,N}}^{(n)}$, for $j = 1$ to L. Recall that m is fixed and that Z_j ($1 \leq j \leq L$) and $e(X_j)$ ($1 \leq j \leq L-1$) are square integrable variables.

We assume that:

$\mathbf{H_1}$: $\forall j = 1,\ldots,L-1$, $\limsup_{\epsilon\to 0}\dfrac{\mathbb{E}\bar{Y}\mathbf{1}_{\{|Z_j-\alpha_j\cdot e(X_j)|\leq\epsilon|e(X_j)|\}}}{\epsilon} < +\infty$, where

$$\bar{Y} = \left(1+\sum_{i=1}^{L}|Z_i|+\sum_{i=1}^{L-1}|e(X_i)|\right)\left(1+\sum_{i=1}^{L-1}|e(X_i)|\right). \tag{4.1}$$

Note that H_1 implies that $\mathbb{P}(Z_j = \alpha_j\cdot e(X_j)) = 0$ and, consequently, under H_1 we know from Sect. 3 that $\frac{1}{N}\sum_{n=1}^{N}F_j(\alpha^{(N)}, Z^{(n)}, X^{(n)})$ converges almost surely to $\phi_j(\alpha)$. Remark too that H_1 is satisfied if the random variable $(Z_j - \alpha_j\cdot e(X_j))$ has a bounded density near zero and the variables Z_j and $e(X_j)$ are bounded.

Theorem 4.1 *Under H_1, the sequences*

$$\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_j(\alpha^{(N)},Z^{(n)},X^{(n)})-\phi_j(\alpha))\right)_{N\geq 1}$$

($j=1,\ldots,L$) and $(\sqrt{N}(\alpha_j^{(N)}-\alpha_j))_{N\geq 1}$ ($j=1,\ldots,L-1$) are tight.

The Proof of Theorem 4.1 is based on the following Lemma.

Lemma 4.1 *Let $(U^{(n)},V^{(n)},W^{(n)})$ be a sequence of identically distributed random variables with values in $[0,+\infty)^3$ such that*

$$\limsup_{\epsilon\to 0}\frac{\mathbb{E}\left(W^{(1)}\mathbf{1}_{\{U^{(1)}\leq\epsilon V^{(1)}\}}\right)}{\epsilon}<+\infty,$$

and (θ_N) a sequence of positive random variables such that $(\sqrt{N}\theta_N)$ is tight, then the sequence $\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}W^{(n)}\mathbf{1}_{\{U^{(n)}\leq\theta_N V^{(n)}\}}\right)_{N\geq 1}$ is tight.

Proof Let $\sigma_N(\theta)=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}W^{(n)}\mathbf{1}_{\{U^{(n)}\leq\theta V^{(n)}\}}$. Observe that σ_N is a non decreasing function of θ. Let $A>0$, we have

$$\begin{aligned}
\mathbb{P}(\sigma_N(\theta_N)\geq A)&\leq\mathbb{P}(\sigma_N(\theta_N)\geq A,\sqrt{N}\theta_N\leq B)+\mathbb{P}(\sqrt{N}\theta_N>B)\\
&\leq\mathbb{P}(\sigma_N(\tfrac{B}{\sqrt{N}})\geq A)+\mathbb{P}(\sqrt{N}\theta_N>B)\\
&\leq\tfrac{1}{A}\mathbb{E}\sigma_N(\tfrac{B}{\sqrt{N}})+\mathbb{P}(\sqrt{N}\theta_N>B)\\
&=\tfrac{\sqrt{N}}{A}\mathbb{E}(W^{(1)}\mathbf{1}_{\{U^{(1)}\leq\frac{B}{\sqrt{N}}V^{(1)}\}})+\mathbb{P}(\sqrt{N}\theta_N>B).
\end{aligned}$$

From the assumption on $(U^{(n)},V^{(n)},W^{(n)})$ and the tightness of $(\sqrt{N}\theta_N)$, we deduce easily the tightness of $\sigma_N(\theta_N)$. □

Proof of Theorem 4.1 We know from the classical Central Limit Theorem that the sequence $(1/\sqrt{N})\sum_{n=1}^{N}(F_j(\alpha,Z^{(n)},X^{(n)})-\phi_j(\alpha))$ is tight and it remains to prove the tightness of $\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_j(\alpha^{(N)},Z^{(n)},X^{(n)})-F_j(\alpha,Z^{(n)},X^{(n)}))$, for $j=1$ to L. Similarly, to prove the tightness of $(\sqrt{N}(\alpha_j^{(N)}-\alpha_j))_{N\geq 1}$, for $j=1$ to $L-1$, we just have to prove the tightness of $\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(G_j(\alpha^{(N)},Z^{(n)},X^{(n)})-G_j(\alpha,Z^{(n)},X^{(n)}))$ (see Sect. 2 for the notation). We proceed by induction on j. The tightness of $\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_L(\alpha^{(N)},Z^{(n)},X^{(n)})-F_L(\alpha,Z^{(n)},X^{(n)}))$ is obvious and that of $(\sqrt{N}(\alpha_{L-1}^{(N)} - \alpha_{L-1}))$ follows from the Central Limit Theorem for the sequence $\left(Z_L^{(n)} e(X_{L-1}^{(n)})\right)$ and the almost sure convergence of the sequence $(A_{L-1}^{(N)})_{N\in\mathbb{N}}$.

Assume that $\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_i(\alpha^{(N)}, Z^{(n)}, X^{(n)}) - F_i(\alpha, Z^{(n)}, X^{(n)}))$ and $(\sqrt{N}(\alpha_{i-1}^{(N)} - \alpha_{i-1}))$ are tight for $i = j$ to L. We set

$$\overline{F}_N = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_{j-1}(\alpha^{(N)}, Z^{(n)}, X^{(n)}) - F_{j-1}(\alpha, Z^{(n)}, X^{(n)})).$$

Now from Lemma 3.1, we have :

$$|\overline{F}_N| \leq \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\bar{Y}^{(n)}\sum_{i=j-1}^{L-1}\mathbf{1}_{\{|Z_i^{(n)} - \alpha_i \cdot e(X_i^{(n)})| \leq |\alpha_i - \alpha_i^{(N)}| \cdot |e(X_i^{(n)}|\}},$$

with

$$\bar{Y}^{(n)} = \left(1 + \sum_{i=1}^{L}|Z_i^{(n)}| + \sum_{i=1}^{L-1}|e(X_i^{(n)})|\right)\left(1 + \sum_{i=1}^{L-1}|e(X_i^{(n)})|\right). \quad (4.2)$$

From Lemma 4.1 and by the induction hypothesis, we deduce that $\overline{F}_N$ is tight. In the same way, we prove that $(\sqrt{N}(\alpha_{j-2}^{(N)} - \alpha_{j-2}))$ is tight.

4.2 A central limit theorem

We prove in this section that under some stronger assumptions than in Sect. 4.1, the vector $\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(Z_{\tau_j^{n,N}}^{(n)} - \mathbb{E}Z_{\tau_j^{[m]}})\right)_{j=1,\ldots,L}$ converges weakly to a Gaussian vector. With the preceding notation, we have

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(Z_{\tau_j^{n,N}}^{(n)} - \mathbb{E}Z_{\tau_j^{[m]}}) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_j(\alpha^{(N)}, Z^{(n)}, X^{(n)}) - \phi_j(\alpha)). \quad (4.3)$$

In the following, we will denote by Y the pair (Z, X) and by $Y^{(n)}$ the pair $(Z^{(n)}, X^{(n)})$. We will also use $\bar{Y}$ and $\bar{Y}^{(n)}$ as defined in (4.1) and (4.2).

We will need the following hypothesis: $\mathbf{H_1^*}$: For $j = 1$ to $L - 1$, there exists a neighborhood V_j of α_j, $\eta_j > 0$ and $k_j > 0$ such that for $a_j \in V_j$ and for $\epsilon \in [0, \eta_j]$,

$$\mathbb{E}(\bar{Y}\mathbf{1}_{\{|Z_j - a_j \cdot e(X_j)| \leq \epsilon |e(X_j)|\}}) \leq \epsilon k_j.$$

$\mathbf{H_2}$: For $j = 1$ to L, Z_j and $e(X_j)$ are in L^p for all $1 \leq p < +\infty$. $\mathbf{H_3}$: For $j = 1$ to $L - 1$, ϕ_j and ψ_j are $\mathcal{C}^1$ in a neighborhood of α.

Observe that H_1^* is stronger than H_1.

Theorem 4.2 *Under H_1^*, H_2, H_3, the vector*

$$\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(Z^{(n)}_{\tau_j^{n,N}} - \mathbb{E}Z_{\tau_j^{[m]}}), \sqrt{N}(\alpha_j^{(N)} - \alpha_j)\right)_{j=1,\dots,L-1}$$

converges in law to a Gaussian vector as N goes to infinity.

For the proof of Theorem 4.2, we will use the following decomposition :

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_j(\alpha^{(N)}, Y^{(n)}) - \phi_j(\alpha)) =$$
$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\Big(F_j(\alpha^{(N)}, Y^{(n)}) - F_j(\alpha, Y^{(n)}) - (\phi_j(\alpha^{(N)}) - \phi_j(\alpha))\Big)$$
$$+\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_j(\alpha, Y^{(n)}) - \phi_j(\alpha)) + \sqrt{N}(\phi_j(\alpha^{(N)}) - \phi_j(\alpha)).$$

From the classical Central Limit Theorem, we know that $\left(\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(F_j(\alpha, Y^{(n)}) - \phi_j(\alpha))\right)_j$ converges in law to a Gaussian vector. Moreover, we have

$$\sqrt{N}(\alpha_j^{(N)} - \alpha_j) = (A_j^{(N)})^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\Big(G_{j+1}(\alpha^{(N)}, Y^{(n)}) - \psi_{j+1}(\alpha)\Big)$$
$$-(A_j)^{-1}\sqrt{N}(A_j^{(N)} - A_j)(A_j^{(N)})^{-1}\psi_{j+1}(\alpha),$$

where $A_j^{(N)}$ converges almost surely to A_j and $\sqrt{N}(A_j^{(N)} - A_j)$ converges in law. We have the similar decomposition

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(G_{j+1}(\alpha^{(N)}, Y^{(n)}) - \psi_{j+1}(\alpha)) =$$
$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\Big(G_{j+1}(\alpha^{(N)}, Y^{(n)}) - G_{j+1}(\alpha, Y^{(n)}) - (\psi_{j+1}(\alpha^{(N)}) - \psi_{j+1}(\alpha))\Big)$$
$$+\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(G_{j+1}(\alpha, Y^{(n)}) - \psi_{j+1}(\alpha)) + \sqrt{N}(\psi_{j+1}(\alpha^{(N)}) - \psi_{j+1}(\alpha)). \quad (4.4)$$

Using these decompositions and Theorem 4.3 below, together with the differentiability of the functions ϕ and ψ, Theorem 4.2 can be proved by induction on j.

Theorem 4.3 *Under H_1^*, H_2, H_3, the variables*

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\Big(F_j(\alpha^{(N)}, Y^{(n)}) - F_j(\alpha, Y^{(n)}) - (\phi_j(\alpha^{(N)}) - \phi_j(\alpha))\Big)$$

and

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\Big(G_{j+1}(\alpha^{(N)},Y^{(n)})-G_{j+1}(\alpha,Y^{(n)})-(\psi_{j+1}(\alpha^{(N)})-\psi_{j+1}(\alpha))\Big)$$

converge to 0 *in* L^2*, for* $j=1$ *to* $L-1$.

Remark 4.1 If we try to compute the covariance matrix of the limiting distribution, we see from Theorem 4.3 and the above decompositions that it depends on the derivatives of the functions ϕ_j and ψ_j at α. This means that the estimation of this covariance matrix may prove difficult. Indeed, the derivation of an estimator for the derivative of a function is typically harder than for the function itself.

Remark 4.2 Theorem 4.3 can be viewed as a way of centering the F_j-increments (resp. the G_{j+1}-increments) between α and $\alpha^{(N)}$ by the ϕ_j-increments (resp. the ψ_{j+1}-increments). One way to get some intuition of the proof of Theorem 4.3 is to observe that if the sequence $(\alpha^{(N)})_{N\in\mathbb{N}}$ were *independent* of the $Y^{(n)}$'s, the convergence in L^2 would reduce to the convergence of $(\alpha^{(N)})_{N\in\mathbb{N}}$ to α. Indeed, we would have to consider expectations of the type $\frac{1}{N}\mathbb{E}\left(\sum_{n=1}^{N}\xi_n\right)^2$ with variables ξ_n which are centered and iid, conditionally on $\alpha^{(N)}$. The main difficulty in the proof of Theorem 4.3 comes from the fact that $\alpha^{(N)}$ is *not* independent of the $Y^{(n)}$'s. On the other hand, we do have identically distributed random variables and we will exploit symmetry arguments and the independence of $\alpha^{(N-1)}$ and $Y^{(N)}$.

For the proof of Theorem 4.3, we need to control the increments of F_j and G_{j+1}. Lemma 4.2 relates these increments to indicator functions. Lemma 4.3 will enable us to localize $\alpha^{(N)}$ near α. We will then develop recursive techniques adapted to dynamic programming (see Lemma 4.3 and Lemma 4.5).

In the following, we denote by $I(Y_i,a_i,\epsilon)$ the function

$$I(Y_i,a_i,\epsilon)=\mathbf{1}_{\{|Z_i-a_i\cdot e(X_i)|\le\epsilon|e(X_i)|\}}.$$

Note that

$$I(Y_i,a_i,\epsilon)\le I(Y_i,b_i,\epsilon+|b_i-a_i|). \tag{4.5}$$

The following Lemma is essentially a reformulation of Lemma 3.1.

Lemma 4.2 *For* $j=1$ *to* $L-1$*, and* a,b *in* $(\mathbb{R}^m)^{L-1}$*, we have*

$$\begin{aligned}|F_j(a,Y)-F_j(b,Y)| &\le \bar{Y}\sum_{i=j}^{L-1}I(Y_i,a_i,|a_i-b_i|)\\ |G_j(a,Y)-G_j(b,Y)| &\le \bar{Y}\sum_{i=j}^{L-1}I(Y_i,a_i,|a_i-b_i|)\end{aligned}$$

Lemma 4.3 *Assume H_1^* and H_2, then for $j = 1$ to $L-1$, there exists $C_j > 0$ such that for all $\delta > 0$,*

$$\mathbb{P}(|\alpha_j^{(N)} - \alpha_j| \geq \delta) \leq \frac{C_j}{\delta^4 N^2}.$$

Proof Let us recall that if $(U_n)_n$ is a sequence of i.i.d. variables such that $EU_1^4 < +\infty$, we have

$$\forall \delta > 0, \quad \mathbb{P}\left(|\frac{1}{N}\sum_{n=1}^{N} U_n - \mathbb{E}U_1| \geq \delta\right) \leq \frac{C}{\delta^4 N^2} \tag{4.6}$$

Observe that

$$\alpha_j^{(N)} - \alpha_j = (A_j^{(N)})^{-1}\frac{1}{N}\sum_{n=1}^{N}(G_{j+1}(\alpha^{(N)}, Y^{(n)}) - \psi_{j+1}(\alpha))$$
$$-(A_j)^{-1}(A_j^{(N)} - A_j)(A_j^{(N)})^{-1}\psi_{j+1}(\alpha),$$

We set $\Omega_j^\epsilon = \{||A_j^{(N)} - A_j|| \leq \epsilon\}$ and we choose ϵ such that $||(A_j^{(N)})^{-1}|| \leq 2||(A_j)^{-1}||$ on Ω_j^ϵ. From (4.6), we know that $\mathbb{P}((\Omega_j^\epsilon)^c) \leq \frac{C_j}{\epsilon^4 N^2}$, for $j = 1$ to $L-1$ and that, on Ω_j^ϵ,

$$|\alpha_j^{(N)} - \alpha_j| \leq K' \left|\frac{1}{N}\sum_{n=1}^{N}(G_{j+1}(\alpha^{(N)}, Y^{(n)}) - \psi_{j+1}(\alpha))\right| + K\epsilon.$$

Now, since $G_L(\alpha^{(N)}, Y^{(n)}) = Z_L^{(n)} e(X_{L-1}^{(n)})$, we deduce from (4.6) applied with this choice of U_n that

$$\mathbb{P}(|\alpha_{L-1}^{(N)} - \alpha_{L-1}| \geq \delta) \leq \frac{C_{L-1}}{(\delta - K\epsilon)^4 N^2} + \frac{C_{L-1}}{\epsilon^4 N^2}.$$

Choosing $\epsilon = \rho\delta$ with ρ small enough, we obtain:

$$\mathbb{P}(|\alpha_{L-1}^{(N)} - \alpha_{L-1}| \geq \delta) \leq \frac{C_{L-1}}{\delta^4 N^2}.$$

Assume now that the result of Lemma 4.3 is true for $j+1, \ldots, L-1$. We will prove that $\mathbb{P}(|\alpha_j^{(N)} - \alpha_j| \geq \delta) \leq \frac{C_j}{\delta^4 N^2}$.

We have

$$\frac{1}{N}\sum_{n=1}^{N}(G_{j+1}(\alpha^{(N)}, Y^{(n)}) - \psi_{j+1}(\alpha)) = \frac{1}{N}\sum_{n=1}^{N}(G_{j+1}(\alpha^{(N)}, Y^{(n)})$$
$$-G_{j+1}(\alpha, Y^{(n)})) + \frac{1}{N}\sum_{n=1}^{N}(G_{j+1}(\alpha, Y^{(n)}) - \psi_{j+1}(\alpha)).$$

From Lemma 4.2, we obtain on Ω_j^ϵ,

$$|\alpha_j - \alpha_j^{(N)}| \leq K\epsilon + \frac{1}{N}\sum_{n=1}^{N} \bar{Y}^{(n)} \sum_{i=j+1}^{L-1} I(Y_i^{(n)}, \alpha_i, |\alpha_i - \alpha_i^{(N)}|)$$
$$+\frac{K'}{N}\left|\sum_{n=1}^{N}(G_{j+1}(\alpha, Y^{(n)}) - \psi_{j+1}(\alpha))\right|.$$

The last term can be treated using 4.6. Therefore, it suffices to prove that

$$\forall \delta > 0, \quad I\!P(S_N \geq \delta) \leq \frac{C}{\delta^4 N^2},$$

where $S_N = \frac{1}{N}\sum_{n=1}^{N} \bar{Y}^{(n)} \sum_{i=j+1}^{L-1} I(Y_i^{(n)}, \alpha_i, |\alpha_i - \alpha_i^{(N)}|)$. But

$$I\!P(S_N \geq \delta) \leq I\!P\left(\frac{1}{N}\sum_{n=1}^{N} \bar{Y}^{(n)} \sum_{i=j+1}^{L-1} I(Y_i^{(n)}, \alpha_i, \epsilon) \geq \delta\right)$$
$$+\sum_{i=j+1}^{L-1} I\!P\left(|\alpha_i^{(N)} - \alpha_i| \geq \epsilon\right).$$

By assumption, for $i = j+1$ to $L-1$, we have $I\!P(|\alpha_i^{(N)} - \alpha_i| \geq \epsilon) \leq \frac{C_i}{\epsilon^4 N^2}$. Moreover we know from H_1^* that $\sum_{i=j+1}^{L-1} I\!E\bar{Y}^{(n)} I(Y_i^{(n)}, \alpha_i, \epsilon) \leq \epsilon K$, with $K = \sum_{i=j+1}^{L-1} k_j$, so we have $\delta - \sum_{i=j+1}^{L-1} I\!E\bar{Y}^{(n)} I(Y_i^{(n)}, \alpha_i, \epsilon) \geq \delta - \epsilon K$ and, using (4.6) again, we see that

$$I\!P\left(\frac{1}{N}\sum_{n=1}^{N} \bar{Y}^{(n)} \sum_{i=j+1}^{L-1} I(Y_i^{(n)}, \alpha_i, \epsilon) \geq \delta\right) \leq \frac{C}{(\delta - \epsilon K)^4 N^2}\ .$$

Choosing $\epsilon = \rho\delta$, with ρ small enough, we obtain the result of Lemma 4.3. □

Before stating other technical results in preparation for the proof of Theorem 4.3, we introduce the following notations. Given $k \in \{1, 2, \ldots, L\}$, λ and μ in $I\!R^+$, we define a sequence of random vectors $\left(\mathcal{U}_i^{(N-k)}(\lambda, \mu), i = 1, \ldots, L-1\right)$, by the recursive relations

$$\mathcal{U}_{L-1}^{(N-k)}(\lambda, \mu) = \frac{\lambda}{N}$$
$$\mathcal{U}_i^{(N-k)}(\lambda, \mu) = \frac{\lambda}{N} + \frac{\mu}{N}\sum_{n=1}^{N-k} \bar{Y}^{(n)} \sum_{j=i+1}^{L-1} I\left(Y_j^{(n)}, \alpha_j^{(N-k)}, \mathcal{U}_j^{(N-k)}(\lambda, \mu)\right),$$
$$(1 \leq i \leq L-2).$$

With this definition, $\mathcal{U}_i^{(N-k)}(\lambda,\mu)$ is obviously $\sigma(Y^{(1)},\ldots,Y^{(N-k)})$-measurable. We also observe that it is a *symmetric* function of $(Y^{(1)},\ldots,Y^{(N-k)})$ (because $\alpha^{(N-k)}$ depends symmetrically on $Y^{(1)},\ldots,Y^{(N-k)}$). The next lemma establishes a useful relation between $\mathcal{U}^{(N-k)}$ and $\mathcal{U}^{(N-k-1)}$.

Lemma 4.4 *Assume H_2. There exist positive constants C,u,v such that for each $N\in\mathbb{N}$, one can find an event Ω_N with $\mathbb{P}(\Omega_N^c)\le C/N^2$ and, on the set Ω_N we have, for $k\in\{1,\ldots,L\}$ and $i\in\{1,\ldots L-1\}$,*

$$\begin{aligned}&\mathcal{U}_i^{(N-k)}(\lambda,\mu)+|\alpha_i^{(N-k)}-\alpha_i^{(N-k-1)}|\\&\le\mathcal{U}_i^{(N-k-1)}\left(\lambda+(L\mu+u)\bar{Y}^{(N-k)},v+\mu\right).\end{aligned}$$

Proof We have

$$\alpha_i^{(N-k)}=(A_i^{(N-k)})^{-1}\frac{1}{N-k}\sum_{n=1}^{N-k}G_{i+1}(\alpha^{(N-k)},Y^{(n)}).$$

Since $A_i^{(N)}$ is the mean of N iid random variables with moments of all orders and mean A_i, we can find Ω_N, with $\mathbb{P}(\Omega_N^c)=O(1/N^2)$, on which $||(A_i^{(N-k)})^{-1}||\le 2||A_i^{-1}||$, for $k=1,\ldots,L$, $i=1,\ldots,L-1$. On this set, we have, for some positive constant C,

$$\begin{aligned}|\alpha_i^{(N-k)}-\alpha_i^{(N-k-1)}|\le C\frac{\bar{Y}^{(N-k)}}{N-k}\left(1+\frac{1}{N-k-1}\sum_{n=1}^{N-k-1}\bar{Y}^{(n)}\right)+\\\frac{C}{N-k}\sum_{n=1}^{N-k-1}\left(G_{i+1}(\alpha^{(N-k)},Y^{(n)})-G_{i+1}(\alpha^{(N-k-1)},Y^{(n)})\right).\end{aligned}\tag{4.7}$$

Here we have used the inequality $||(N-k)A^{(N-k)}-(N-k-1)A^{(N-k-1)}||\le\bar{Y}^{(N-k)}$. We may choose Ω_N in such a way that $\frac{1}{N-k-1}\sum_{n=1}^{N-k-1}\bar{Y}^{(n)}$ remains bounded on Ω_N. Note that, for $i=L-1$, the last sum in (4.7) vanishes. Using Lemma 4.2 for $i\le L-2$, we have, on Ω_N,

$$\begin{aligned}&|\alpha_i^{(N-k)}-\alpha_i^{(N-k-1)}|\le\frac{u\bar{Y}^{(N-k)}}{N}\\&+\frac{v}{N}\sum_{n=1}^{N-k-1}\bar{Y}^{(n)}\sum_{j=i+1}^{L-1}I\left(Y_j^{(n)},\alpha_j^{(N-k-1)},|\alpha_j^{(N-k)}-\alpha_j^{(N-k-1)}|\right)\end{aligned}\tag{4.8}$$

for some constants u and v.

To complete the proof of the lemma, we observe that

$$\begin{aligned}&I\left(Y_j^{(n)},\alpha_j^{(N-k)},\mathcal{U}_j^{(N-k)}(\lambda,\mu)\right)\le\\&I\left(Y_j^{(n)},\alpha_j^{(N-k-1)},\mathcal{U}_j^{(N-k)}(\lambda,\mu)\right.\\&\left.+|\alpha_j^{(N-k)}-\alpha_j^{(N-k-1)}|\right).\end{aligned}$$

Now, for $i \leq L-2$, by going back to the recursive definition of $\mathcal{U}_i^{(N-k)}(\lambda, \mu)$ and separating the $N-k$-th term of the sum, we obtain

$$\mathcal{U}_i^{(N-k)}(\lambda,\mu) \leq \frac{\lambda + L\mu\bar{Y}^{(N-k)}}{N} + \frac{\mu}{N}\sum_{n=1}^{N-k-1}\bar{Y}^{(n)}\sum_{j=i+1}^{L-1} I\left(Y_j^{(n)}, \alpha_j^{(N-k-1)}, \mathcal{U}_j^{(N-k)}(\lambda,\mu) + |\alpha_j^{(N-k)} - \alpha_j^{(N-k-1)}|\right) \tag{4.9}$$

Now let $V_i = \mathcal{U}_i^{(N-k)}(\lambda,\mu) + |\alpha_i^{(N-k)} - \alpha_i^{(N-k-1)}|$. By combining (4.8) and (4.9) we get

$$V_i \leq \frac{\lambda + (L\mu + u)\bar{Y}^{(N-k)}}{N} + \frac{\mu + v}{N}\sum_{n=1}^{N-k-1}\bar{Y}^{(n)}\sum_{j=i+1}^{L-1} I\left(Y_j^{(n)}, \alpha_j^{(N-k-1)}, V_j\right)$$

□

Lemma 4.5 *Assume H_1^* and H_2. For all $\varepsilon \in (0,1]$ and for all $\mu \geq 0$, there exists a constant $C_{\varepsilon,\mu}$ such that*

$$\forall\lambda \geq 0, \quad \forall i \in \{1,\ldots,L-1\}, \quad \mathbb{E}\mathcal{U}_i^{(N-2)}(\lambda,\mu) \leq \frac{C_{\varepsilon,\mu}(1+\lambda)}{N^{1-\varepsilon}}.$$

Proof We will prove by induction on k $(= L, L-1, \ldots, 2)$ that

$$\sup_{k-1\leq i\leq L-1} \mathbb{E}\mathcal{U}_i^{(N-k)}(\lambda,\mu) \leq \frac{C_{\varepsilon,\mu}(1+\lambda)}{N^{1-\varepsilon}} \tag{4.10}$$

We obviously have (4.10) for $k = L$, since $\mathcal{U}_{L-1}^{(N-L)}(\lambda,\mu) = \lambda/N$.

We now assume that (4.10) holds for $k+1$ and will prove it for k. For $i \geq k-1$, we have, using the symmetry of $\mathcal{U}_j^{(N-k)}(\lambda,\mu)$ with respect to $Y^{(1)}, \ldots, Y^{(N-k)}$,

$$\mathbb{E}\mathcal{U}_i^{(N-k)}(\lambda,\mu) = \frac{\lambda}{N} + \frac{\mu(N-k)}{N}\mathbb{E}\left(\bar{Y}^{(N-k)}\sum_{j=i+1}^{L-1} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k)}, \mathcal{U}_j^{(N-k)}(\lambda,\mu)\right)\right)$$

For $j = i+1, \ldots, L-1$, we have, using (4.5), Lemma 4.4, and the notation

$$\mathcal{V}_j^{(N-k-1)}(\lambda,\mu) = \mathcal{U}_j^{(N-k-1)}\left(\lambda + (L\mu + u)\bar{Y}^{(N-k)}, v + \mu\right),$$

$$\mathbb{E}\left(\bar{Y}^{(N-k)} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k)}, \mathcal{U}_j^{(N-k)}(\lambda,\mu)\right)\right) \leq$$
$$\mathbb{E}\bar{Y}^{(N-k)} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k-1)}, \mathcal{U}_j^{(N-k)}(\lambda,\mu) + |\alpha_j^{(N-k)} - \alpha_j^{(N-k-1)}|\right)$$
$$\leq \mathbb{E}\bar{Y}^{(N-k)} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k-1)}, \mathcal{V}_j^{(N-k-1)}(\lambda,\mu)\right) + \mathbb{E}\bar{Y}^{(N-k)} \mathbf{1}_{\Omega_N^c}$$

Note that $\mathbb{E}\bar{Y}^{(N-k)}\mathbf{1}_{\Omega_N^c} \leq ||\bar{Y}||_{L^2}\sqrt{\mathbb{P}(\Omega_N^c)} = O(1/N)$.

At this point we would like to use H_1^* and the induction hypothesis. However, we have to be careful because $\mathcal{V}_j^{(N-k-1)}(\lambda,\mu)$ depends on $Y^{(N-k)}$. For $j \in \{i+1,\dots,L-1\}$, we write

$$\mathbb{E}\bar{Y}^{(N-k)} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k-1)}, \mathcal{V}_j^{(N-k-1)}(\lambda,\mu)\right) = \sum_{l=1}^{\infty} A_l,$$

with

$$A_l = \mathbb{E}\bar{Y}^{(N-k)} \mathbf{1}_{\{l-1\leq \bar{Y}^{(N-k)}<l\}} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k-1)}, \mathcal{V}_j^{(N-k-1)}(\lambda,\mu)\right)$$
$$\leq \left(l\mathbb{P}(\bar{Y}^{(N-k)} \geq l-1)\right)^{1/p}$$
$$\left(\mathbb{E}\bar{Y}^{(N-k)} \mathbf{1}_{\{\bar{Y}^{(N-k)}<l\}} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k-1)}, \mathcal{V}_j^{(N-k-1)}(\lambda,\mu)\right)\right)^{1-\frac{1}{p}},$$

for all $p \in (1,+\infty)$ (Hölder). Now,

$$\mathbb{E}\bar{Y}^{(N-k)} \mathbf{1}_{\{\bar{Y}^{(N-k)}<l\}} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k-1)}, \mathcal{V}_j^{(N-k-1)}(\lambda,\mu)\right) \leq$$
$$\mathbb{E}\bar{Y}^{(N-k)} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k-1)}, \mathcal{U}_j^{(N-k-1)}\left(\lambda + (L\mu+u)l, v+\mu\right)\right),$$

and we may condition with respect to $\sigma(Y^1,\dots,Y^{(N-k-1)})$ and use Lemma 4.3 and H_1^* to obtain

$$\mathbb{E}\bar{Y}^{(N-k)} \mathbf{1}_{\{\bar{Y}^{(N-k)}<l\}} I\left(Y_j^{(N-k)}, \alpha_j^{(N-k-1)}, \mathcal{V}_j^{(N-k-1)}(\lambda,\mu)\right) \leq$$
$$C\mathbb{E}\mathcal{U}_j^{(N-k-1)}\left(\lambda + (L\mu+u)l, v+\mu\right) + \frac{C}{N^2}.$$

We can now apply the induction hypothesis, and we easily deduce (4.10) for k, using the fact that $\mathbb{P}(\bar{Y}^{(N-k)} \geq l-1) = o(1/l^m)$ for all $m \in \mathbb{N}$. □

Proof of Theorem 4.3 We prove that

$$\lim_N \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left(F_j(\alpha^{(N)}, Y^{(n)}) - F_j(\alpha, Y^{(n)}) - (\phi_j(\alpha^{(N)}) - \phi_j(\alpha))\right) = 0$$

in L^2. The proof is similar for the second term of the Theorem. We introduce the notation $\Delta_j(a,b,Y) = F_j(a,Y) - F_j(b,Y) - (\phi_j(a) - \phi_j(b))$. We have to prove that

$$\lim_N \frac{1}{N}\mathbb{E}\left(\sum_{n=1}^{N} \Delta_j(\alpha^{(N)}, \alpha, Y^{(n)})\right)^2 = 0.$$

Remark that for $n = 1$ to N, the pairs $(\alpha^{(N)}, Y^{(n)})$ and $(\alpha^{(N)}, Y^{(1)})$ have the same law, and for $n \neq m$, $(\alpha^{(N)}, Y^{(n)}, Y^{(m)})$ and $(\alpha^{(N)}, Y^{(N)}, Y^{(N-1)})$ have the same distribution. So we obtain

$$\frac{1}{N}\mathbb{E}(\sum_{n=1}^{N} \Delta_j(\alpha^{(N)}, \alpha, Y^{(n)}))^2 = \mathbb{E}\Delta_j^2(\alpha^{(N)}, \alpha, Y^{(1)})$$
$$+(N-1)\mathbb{E}\Delta_j(\alpha^{(N)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha, Y^{(N)}).$$

But $|\Delta_j(\alpha^{(N)}, \alpha, Y^{(1)})| \leq 2\left(\bar{Y}^{(1)} + \mathbb{E}\bar{Y}^{(1)}\right)$. Since the sequence $(\alpha^{(N)})$ goes to α almost surely and $\mathbb{P}(Z_j = \alpha_j \cdot e(X_j)) = 0$ for $j = 1$ to $L-1$ by assumption, we deduce that $\Delta_j(\alpha^{(N)}, \alpha, Y^{(1)})$ goes to 0 almost surely. Consequently, we obtain that $\mathbb{E}\Delta_j^2(\alpha^{(N)}, \alpha, Y^{(1)})$ tends to 0. It remains to prove that

$$\lim_{N} N\mathbb{E}\Delta_j(\alpha^{(N)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha, Y^{(N)}) = 0. \tag{4.11}$$

We observe that

$$\mathbb{E}\left(\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N)})|Y^{(1)}, \ldots, Y^{(N-1)}\right) = 0,$$

since $\mathbb{E}\left(F_j(\alpha^{(N-2)}, Y^{(N)})|Y^{(1)}, \ldots, Y^{(N-1)}\right) = \phi_j(\alpha^{(N-2)})$ almost surely. This gives

$$\mathbb{E}\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N)}) = 0,$$

and we just have to prove that

$$\lim_{N} N\mathbb{E}\Big(\Delta_j(\alpha^{(N)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha, Y^{(N)})$$
$$-\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N)})\Big) = 0.$$

We have the equality

$$\Delta_j(\alpha^{(N)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha, Y^{(N)})$$
$$-\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N)})$$
$$= \Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha, Y^{(N)})$$
$$+\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N)}),$$

We want to prove that

$$\lim_{N\to\infty} N\mathbb{E}\Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha, Y^{(N)}) = 0 \tag{4.12}$$

and

$$\lim_{N\to\infty} N\mathbb{E}\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N)}) = 0. \tag{4.13}$$

Both equalities can be proved in a similar manner. We give the details for (4.13).

First, note that given any $\eta > 0$, we have, using H_2 and Hölder's inequality,

$$\mathbb{E}\left(\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N)})\mathbf{1}_{\{|\alpha^{(N)}-\alpha| \geq N^{-\eta}\}}\right) \leq C_p\left(\mathbb{P}(|\alpha^{(N)} - \alpha| \geq N^{-\eta})\right)^{1/p},$$

for all $p > 1$. We know from Lemma 4.3 that $\mathbb{P}(|\alpha^{(N)} - \alpha| \geq N^{-\eta}) \leq C/N^{2-4\eta}$. Therefore, if $\eta < 1/4$,

$$\lim_{N\to\infty} N\mathbb{E}\left(\Delta_j(\alpha^{(N-2)}, \alpha, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N)}) \mathbf{1}_{\{|\alpha^{(N)}-\alpha| \geq N^{-\eta}\}}\right) = 0.$$

On the other hand, we have, using Lemma 4.2,

$$\left|\Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N-1)})\right| \leq \bar{Y}^{(N-1)} \sum_{i=j}^{L-1} I\left(Y_i^{(N-1)}, \alpha_i^{(N-2)}, \left|\alpha_i^{(N)} - \alpha_i^{(N-2)}\right|\right) + \left|\phi_j(\alpha^{(N)}) - \phi_j(\alpha^{(N-2)}\right|,$$

and

$$\left|\Delta_j(\alpha^{(N)}, \alpha, Y^{(N)})\right| \leq \bar{Y}^{(N)} \sum_{i=j}^{L-1} I\left(Y_i^{(N)}, \alpha_i, \left|\alpha_i - \alpha_i^{(N)}\right|\right) + \left|\phi_j(\alpha) - \phi_j(\alpha^{(N)}\right|.$$

By the same reasoning as in the proof of Lemma 4.4, we have positive constants C, s and t such that, for each $N \in \mathbb{N}$, one can find a set Ω_N with $\mathbb{P}(\Omega_N^c) \leq C/N^2$, on which

$$\left|\alpha_i^{(N)} - \alpha_i^{(N-2)}\right| \leq \mathcal{U}_i^{(N-2)}\left(s(\bar{Y}^{(N-1)} + \bar{Y}^{(N)}), t\right), \quad i = j, \ldots, L-1.$$

Using Lemma 4.3 and H_3, we may also assume that, on Ω_N,

$$\left|\phi_j(\alpha^{(N)}) - \phi_j(\alpha^{(N-2)})\right| \leq K \sum_{i=j}^{L-1} \left|\alpha_i^{(N)} - \alpha_i^{(N-2)}\right|,$$

and

$$\left|\phi_j(\alpha^{(N)}) - \phi_j(\alpha)\right| \leq K \sum_{i=j}^{L-1} \left|\alpha_i^{(N)} - \alpha_i\right|$$

for some positive constant K.

We now have, for $\eta < 1/4$,

$$\mathbb{E}\Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N-1)})\Delta_j(\alpha^{(N)}, \alpha, Y^{(N)}) =$$
$$\sum_{l=j}^{L-1} \mathbb{E}\Delta_j(\alpha^{(N)}, \alpha^{(N-2)}, Y^{(N-1)}) \left(\bar{Y}^{(N)} I\left(Y_l^{(N)}, \alpha_l, N^{-\eta}\right)\right.$$
$$\left.+KN^{-\eta}\right) + o(1/N).$$

In order to prove (4.12), it now suffices to show that, for $j \le i, l \le L-1$,

$$\lim_{N\to\infty} N\mathbb{E}\left(\bar{Y}^{(N-1)} I\left(Y_i^{(N-1)}, \alpha_i^{(N-2)}, V_i^{(N-2)}\right)\right.$$
$$\left.+V_j^{(N-2)}\right)\left(\bar{Y}^{(N)} I\left(Y_l^{(N)}, \alpha_l, N^{-\eta}\right) + N^{-\eta}\right) = 0,$$

with the notation

$$V_i^{(N-2)} = \mathcal{U}_i^{(N-2)}\left(s(\bar{Y}^{(N-1)} + \bar{Y}^{(N)}), t\right).$$

We will only prove that

$$\lim_{N\to\infty} N\mathbb{E}\left(\bar{Y}^{(N-1)} I\left(Y_i^{(N-1)}, \alpha_i^{(N-2)}, V_i^{(N-2)}\right)\right.$$
$$\left.\bar{Y}^{(N)} I\left(Y_l^{(N)}, \alpha_l, N^{-\eta}\right)\right) = 0,$$

since the other terms are easier to control. For $m \in \mathbb{N}$, let

$$A_m = \{m - 1 \le \bar{Y}^{(N-1)} + \bar{Y}^{(N)} < m\}.$$

We have

$$\mathbb{E}\left(\bar{Y}^{(N-1)} I\left(Y_i^{(N-1)}, \alpha_i^{(N-2)}, V_i^{(N-2)}\right)\bar{Y}^{(N)} I\left(Y_l^{(N)}, \alpha_l, N^{-\eta}\right)\mathbf{1}_{A_m}\right) \le$$
$$\mathbb{E}\left(\bar{Y}^{(N-1)} I\left(Y_i^{(N-1)}, \alpha_i^{(N-2)}, \mathcal{U}_i^{(N-2)}(sm, t)\right)\bar{Y}^{(N)} I\left(Y_l^{(N)}, \alpha_l, N^{-\eta}\right)\right) =$$
$$\mathbb{E}\bar{Y}^{(N-1)} I\left(Y_i^{(N-1)}, \alpha_i^{(N-2)}, \mathcal{U}_i^{(N-2)}(sm, t)\right)\mathbb{E}\bar{Y}^{(N)} I\left(Y_l^{(N)}, \alpha_l, N^{-\eta}\right).$$

Here we have used the fact that $Y^{(N)}$ is independent of $(Y^{(1)}, \dots, Y^{(N-1)})$. We now condition with respect to $\sigma(Y^{(1)}, \dots, Y^{(N-2)})$ in the first expectation and we use H_1^* and Lemma 4.5 to obtain

$$\mathbb{E}\left(\bar{Y}^{(N-1)} I\left(Y_i^{(N-1)}, \alpha_i^{(N-2)}, V_i^{(N-2)}\right)\bar{Y}^{(N)} I\left(Y_l^{(N)}, \alpha_l, N^{-\eta}\right)\mathbf{1}_{A_m}\right)$$
$$\le C_\varepsilon \frac{1+m}{N^{1-\varepsilon+\eta}}.$$

Here ε is an arbitrary positive number and, by taking $\varepsilon < \eta < 1/4$, summing up over m and using Hölder's inequality we easily complete the proof.

References

1. Bally, V., Pagès, G.: A quantization algorithm for solving multidimensional optimal stopping problems, Preprint, December 2000
2. Broadie, M., Glasserman, P.: Pricing American-style securities using simulation. J. Econ. Dyn. Control **21**, 1323–1352 (1997)
3. Broadie, M., Glasserman, P., Jain, G.: Enhanced Monte-Carlo estimates for American option prices. J. Derivatives **5**, 25–44 (1997)
4. Broadie, M., Glasserman, P.: A stochastic mesh method for pricing high-dimensional American options. Preprint, November 1997
5. Carrière, J.: Valuation of Early-Exercice Price of options Using Simulations and Nonparametric Regression. Insurance: Math. Econ. **19**, 19–30 (1996)
6. Dupacova, J., Wets, R.: Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. Ann. Stat. **16** (4), 1517–1549 (1988)
7. King A.J., Rockafellar, T.R.: Asymptotic theory for solutions in statistical estimation and stochastic programming. Math. Oper. Res. **18** (1), 148–162 (1993)
8. Lamberton, D., Brownian optimal stopping and random walks. Appl. Math. Optimization (forthcoming)
9. Longstaff, F.A., Schwartz, E.S.: Valuing American options by simulation: a simple least-squares approach. Rev. Financial Stud. **14**, 113–148 (2001)
10. Neveu J.: Discrete-parameter Martingales. Amsterdam: North Holland 1975
11. Shapiro A.: Asymptotic properties of statistical estimators in stochastic programming. Ann. Stat. **17** (2), 841–858 (1989)
12. Tsitsiklis, J.N., Van Roy, B.: Regression methods for pricing complex American-style options. Preprint, August 2000. IEEE Transactions on Neural Networks (forthcoming)