

Longstaff Schwartz algorithm and Neural Network regression

B. Lapeyre

Ecole des Ponts ParisTech

Conférence in honor of Nicole El Karoui

INTRODUCTION

- ▶ Ongoing work with Jérôme Lelong, ENSIMAG
- ▶ Computing an American option involving a large number of assets remains a numerically difficult task
- ▶ A hope : Neural Network (NN) can (may) help to reduce the computational burden.
- ▶ Some previous works using NN for optimal stopping (not LS algorithm though)
 - ▶ Michael Kohler, Adam Krzyżak, and Nebojsa Todorovic. Pricing of high-dimensional American options by neural networks. *Math. Finance*, 20(3):383–410, 2010
 - ▶ Sebastian Becker, Patrick Cheridito, and Arnulf Jentzen. Deep optimal stopping. *arXiv preprint arXiv:1804.05394*, 2018
- ▶ What can be proved for a LS algorithm using NN approximation?

COMPUTING BERMUDEAN OPTIONS PRICES

- ▶ A discrete time (discounted) payoff process $(Z_{T_k})_{0 \leq k \leq N}$ adapted to $(\mathcal{F}_{T_k})_{0 \leq k \leq N}$. $\max_{0 \leq k \leq N} |Z_{T_k}| \in L^2$.
- ▶ The (discounted) value of the Bermudan option at times $(T_k)_{0 \leq k \leq N}$ is given by

$$U_0 = \sup_{\tau \text{ s. t.}} \mathbb{E}(Z_\tau).$$

- ▶ Standard dynamic programming algorithm (\rightarrow “Tsitsiklis-Van Roy” type algorithms).

$$(1) \quad \begin{cases} U_{T_N} = Z_{T_N} \\ U_{T_k} = \max(Z_{T_k}, \mathbb{E}[U_{T_{k+1}} | \mathcal{F}_{T_k}]) \end{cases}$$

- ▶ Using optimal stopping times τ_k at time k rewrites as (\rightarrow “Longstaff-Schwartz” algorithm)

$$(2) \quad \begin{cases} \tau_N = T_N, & \text{then for } 0 < k < N \\ \tau_k = T_k \mathbf{1}_{\{Z_{T_k} \geq \mathbb{E}[Z_{\tau_{k+1}} | \mathcal{F}_{T_k}]\}} + \tau_{k+1} \mathbf{1}_{\{Z_{T_k} < \mathbb{E}[Z_{\tau_{k+1}} | \mathcal{F}_{T_k}]\}}, \\ U_0 = \max(Z_0, \mathbb{E}[Z_{\tau_1}]). \end{cases}$$

THE MARKOVIAN CONTEXT

- ▶ Markovian context : $(X_t)_{0 \leq t \leq T}$ is a Markov process and $Z_{T_k} = \phi_k(X_{T_k})$.

$$\mathbb{E}[Z_{\tau_{k+1}} | \mathcal{F}_{T_k}] = \mathbb{E}[Z_{\tau_{k+1}} | X_{T_k}] = \psi_k(X_{T_k})$$

- ▶ Because of L^2 assumption, ψ_k can be computed by a minimization problem

$$\inf_{\psi \in L^2(\mathcal{L}(X_{T_k}))} \mathbb{E} \left[|Z_{\tau_{k+1}} - \psi(X_{T_k})|^2 \right]$$

- ▶ LS = using equation (2), if you have an approximation of τ_{k+1} using simulation, you can obtain by minimization an approximation of $\mathbb{E}[Z_{\tau_{k+1}} | X_{T_k}]$ and so obtain an approximation of τ_k .

DIFFERENT NUMERICAL STRATEGIES

- ▶ Longstaff/Schwartz type algorithms rely on direct approximation of *stopping times* and use of *the same simulated paths* for all time steps (obvious and large computational gains).
- ▶ The standard numerical (LS) approach : approximate the space L^2 by a finite dimensional vector space (polynomials, ...)
- ▶ We investigate the use of Neural Networks to approximate ψ_k .
- ▶ Kohler et al. [2010]: neural networks but in a different context (approximation of the value function Tsitsiklis and Roy [2001], equation (1)) and re-simulation of the paths at each time steps.

LS: TRUNCATION STEP

- ▶ $(g_k, k \geq 1)$ is an $L^2(\mathcal{L}(X))$ basis and $\Phi_p(X, \theta) = \sum_{k=1}^p \theta_k g_k(X)$.
- ▶ Backward approximation of iteration policy using (2),

$$\begin{cases} \hat{\tau}_N^p = T_N \\ \hat{\tau}_n^p = T_n \mathbf{1}_{\{Z_{T_n} \geq \Phi_p(X_{T_n}; \hat{\theta}_n^p)\}} + \hat{\tau}_{n+1}^p \mathbf{1}_{\{Z_{T_n} < \Phi_p(X_{T_n}; \hat{\theta}_n^p)\}} \end{cases}$$

- ▶ with conditional expectation computed using a Monte Carlo minimization problem : $\hat{\theta}_n^p$ is a minimizer of

$$\inf_{\theta} \mathbb{E} \left(\left| \Phi_p(X_{T_n}; \theta) - Z_{\hat{\tau}_{n+1}^p} \right|^2 \right).$$

- ▶ Price approximation : $U_0^p = \max \left(Z_0, \mathbb{E} \left(Z_{\hat{\tau}_1^p} \right) \right)$.

THE LS ALGORITHM

- ▶ $(g_k, k \geq 1)$ is an $L^2(\mathcal{L}(X))$ basis and $\Phi_p(X, \theta) = \sum_{k=1}^p \theta_k g_k(X)$.
- ▶ Paths $X_{T_0}^{(m)}, X_{T_1}^{(m)}, \dots, X_{T_N}^{(m)}$ and payoff paths $Z_{T_0}^{(m)}, Z_{T_1}^{(m)}, \dots, Z_{T_N}^{(m)}$, $m = 1, \dots, M$.
- ▶ Backward approximation of iteration policy,

$$\begin{cases} \hat{\tau}_N^{p,(m)} = T_N \\ \hat{\tau}_n^{p,(m)} = T_n \mathbf{1}_{\{Z_{T_n}^{(m)} \geq \Phi_p(X_{T_n}^{(m)}; \hat{\theta}_n^{p,M})\}} + \hat{\tau}_{n+1}^{p,(m)} \mathbf{1}_{\{Z_{T_n}^{(m)} < \Phi_p(X_{T_n}^{(m)}; \hat{\theta}_n^{p,M})\}} \end{cases}$$

- ▶ with conditional expectation computed using a Monte Carlo minimization problem : $\hat{\theta}_n^{p,M}$ is a minimizer of

$$\inf_{\theta} \frac{1}{M} \sum_{m=1}^M \left| \Phi_p(X_{T_n}^{(m)}; \theta) - Z_{\hat{\tau}_{n+1}^{p,(m)}}^{(m)} \right|^2.$$

- ▶ Price approximation : $U_0^{p,M} = \max \left(Z_0, \frac{1}{M} \sum_{m=1}^M Z_{\hat{\tau}_1^{p,(m)}}^{(m)} \right)$.

REFERENCE PAPERS

- ▶ Description of the algorithm :

F.A. Longstaff and R.S. Schwartz. [Valuing American options by simulation : A simple least-square approach.](#)
Review of Financial Studies, 14:113–147, 2001.

- ▶ Rigorous approach :

E. Clément, D. Lamberton, and P. Protter. [An analysis of a least squares regression method for american option pricing.](#)
Finance and Stochastics, 6(4):449–471, 2002.

- U_0^p converge to U_0 , $p \rightarrow +\infty$
- $U_0^{p,M}$ converge to U_0^p , $M \rightarrow +\infty$ a.s.
- “almost” a central limit theorem

THE MODIFIED ALGORITHM

- ▶ In LS algorithm replace approximation on a Hilbert basis $\Phi_p(.; \theta)$ by a Neural Network.
- ▶ The optimization problem is non linear, non convex, ...
- ▶ Aim : extend the proof of (a.s.) convergence results

A QUICK VIEW OF NEURAL NETWORKS

- ▶ In short a NN : $x \rightarrow \Phi_p(x, \theta) \in \mathbb{R}$, $\theta \in \mathbb{R}^d$, d large
- ▶ $\Phi_p = A_L \circ \sigma \circ A_{L-1} \circ \cdots \circ \sigma \circ A_1$, $L \geq 2$, L fixed
- ▶ $A_l(x_l) = w_l x_l + \beta_l$ (affine functions)
- ▶ p “maximum number of neurons per layer” (i.e. sizes of the w_l matrix)
- ▶ σ a fixed non linear (called *activation function*) applied component wise
- ▶ $\theta :=$ parameter (w, β) of all the different layers
- ▶ Restriction to the compact set $\Theta_p = \{\theta : |\theta| \leq \gamma_p\}$ and assume $\lim_{p \rightarrow \infty} \gamma_p = \infty$.
- ▶ $NN_p = \{\Phi_p(\cdot, \theta) : \theta \in \Theta_p\}$
- ▶ $NN_\infty = \cup_{p \in \mathbb{N}} NN_p$

HYPOTHESIS H

- ▶ there exist $\delta > 0$ and $q \geq 1$

$$\begin{cases} |\Phi_p(x, \theta_1) - \Phi(x, \theta_2)| \leq \kappa |\theta_1 - \theta_2|^\delta (1 + |x|^q) \\ |\Phi_p(x, \theta_1)| \leq \kappa_q (1 + |x|^q) \end{cases}$$

- ▶ $\mathbb{E}[|X_t|^{2q}] < \infty$ for all $0 \leq t \leq T$.
- ▶ For all $p, n < N$, $\mathbb{P}(Z_{T_n} = \Phi_p(X_{T_n}; \theta_n^p)) = 0$.
- ▶ the family of optimization problems admit a *unique* minimizer θ_n^p

$$\inf_{\theta \in \Theta_p} \mathbb{E} \left(\left| \Phi_p(X_{T_n}; \theta) - Z_{\widehat{\tau}_{n+1}^p} \right|^2 \right).$$

THE RESULT

Theorem

Under hypothesis H

- ▶ *Convergence of the approximation in p*

$$\lim_{p \rightarrow \infty} \mathbb{E}[Z_{\tau_n^p} | \mathcal{F}_{T_n}] = \mathbb{E}[Z_{\tau_n} | \mathcal{F}_{T_n}] \text{ in } L^2(\Omega) \quad (\text{i.e. } U_0^p \rightarrow U_0).$$

- ▶ *SLLN : for every $k = 1, \dots, N$,*

$$\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M Z_{\widehat{\tau}_k^{p,(m)}}^{(m)} = \mathbb{E} \left[Z_{\tau_k^p} \right] \quad a.s. \quad (\text{i.e. } U_0^{p,M} \rightarrow U_0^p)$$

CONVERGENCE OF THE APPROXIMATION IN p

A simple consequence of the Hornik theorem (Hornik [1991]).

- ▶ Also known as “Universal Approximation Theorem”.

Theorem (Hornik)

Assume that the function σ is non constant and bounded. Let μ denote a probability measure on \mathbb{R}^r , then NN_∞ is dense in $L^2(\mathbb{R}^r, \mu)$.

- ▶ Corollary : If for every p , $\alpha_p \in \Theta_p$ is a minimizer of

$$\inf_{\theta \in \Theta_p} \mathbb{E}[|\Phi_p(X; \theta) - Y|^2],$$

$(\Phi_p(X; \alpha_p))_p$ converges to $\mathbb{E}[Y|X]$ in $L^2(\Omega)$ when $p \rightarrow \infty$.

- ▶ proof of the convergence of the “non-linear approximation” $\Phi_p(X; \theta)$.

CONVERGENCE OF MONTE-CARLO APPROXIMATION

- ▶ p fixed, $N \rightarrow +\infty$
- ▶ Minimisation problem is now non linear, need more abstract arguments to prove convergence
- ▶ Two ingredients (“old” results)
- ▶ First result : approximation of minimization problems

Lemma (Rubinstein and Shapiro [1993])

- ▶ $(f_n)_n$ defined on a compact set $K \subset \mathbb{R}^d$. $v_n = \inf_{x \in K} f_n(x)$
- ▶ x_n a sequence of minimizers $f_n(x_n) = \inf_{x \in K} f_n(x)$.

If $(f_n)_n$ converges uniformly on K to a continuous function f . $v^* = \inf_{x \in K} f(x)$, then $v_n \rightarrow v^*$.

Moreover if x^* is the unique minimizer of f , then $x_n \rightarrow x^*$.

CONVERGENCE OF MONTE-CARLO APPROXIMATION

- ▶ Second result : SLLN in Banach spaces
 (Ledoux and Talagrand [1991], date back to Mourier [1953]).

Lemma

Let $(\xi_i)_{i \geq 1}$ i.i.d. \mathbb{R}^m -valued, $h : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}$. If

- ▶ a.s., $\theta \in \mathbb{R}^d \mapsto h(\theta, \xi_1)$ is continuous,
- ▶ $\forall K > 0$, $\mathbb{E} \left[\sup_{|\theta| \leq K} |h(\theta, \xi_1)| \right] < +\infty$.

Then

$$\lim_{n \rightarrow \infty} \sup_{|\theta| \leq K} \left| \frac{1}{n} \sum_{i=1}^n h(\theta, \xi_i) - \mathbb{E}[h(\theta, \xi_1)] \right| = 0 \quad \text{a.s.}$$

CONVERGENCE OF MONTE-CARLO APPROXIMATION

- ▶ Direct application of this 2 results give convergence of

$$\widehat{\theta}_N^{p,M} \rightarrow \theta_N^p$$

- ▶ Using backward iteration, following Clément et al. [2002], give, for $n \leq N$

$\widehat{\theta}_n^{p,M}$ converges to θ_n^p a.s. as $M \rightarrow \infty$.

- ▶ Convergence of $\widehat{\theta}_n^{p,M}$ needed to use backward recurrence
- ▶ ...

NUMERICAL ASPECTS NOT COVERED HERE

- ▶ Numerical experiments yet to be done (work in progress)
- ▶ Numerical resolution of the optimization problem ...
- ▶ Stochastic gradient descent is used to obtain a (local?) minimum
- ▶ In principle less efficient than direct computation of the minimum (in the linear case)
- ▶ ...

BIBLIOGRAPHIE I

- Sebastian Becker, Patrick Cheridito, and Arnulf Jentzen. Deep optimal stopping. *arXiv preprint arXiv:1804.05394*, 2018.
- E. Clément, D. Lamberton, and P. Protter. An analysis of a least squares regression method for american option pricing. *Finance and Stochastics*, 6(4):449–471, 2002.
- Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251 – 257, 1991.
- Côme Huré, Huyêñ Pham, Achref Bachouch, and Nicolas Langrené. Deep neural networks algorithms for stochastic control problems on finite horizon, part i: convergence analysis. *arXiv preprint arXiv:1812.04300*, 2018.
- Michael Kohler, Adam Krzyżak, and Nebojsa Todorovic. Pricing of high-dimensional American options by neural networks. *Math. Finance*, 20(3):383–410, 2010.
- Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Springer-Verlag, Berlin, 1991.
- F.A. Longstaff and R.S. Schwartz. Valuing American options by simulation : A simple least-square approach. *Review of Financial Studies*, 14:113–147, 2001.
- Edith Mourier. Eléments aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré*, 13:161–244, 1953.

BIBLIOGRAPHIE II

Reuven Y. Rubinstein and Alexander Shapiro. *Discrete event systems*. John Wiley & Sons Ltd., 1993. Sensitivity analysis and stochastic optimization by the score function method.

J.N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American-style options. *IEEE Trans. Neural Netw.*, 12(4):694–703, 2001.