Pricing Bermudan options using random forests

Jérôme Lelong (joint work with Zineb El Filali Ech-chafiq, Pierre Henry Labordère)

Univ. Grenoble Alpes, LJK

Premia 24 delivery meeting
April 14, 2022

Computing Bermudan 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 time-T_k discounted value of the Bermudan option is given by

$$U_{T_k} = \text{esssup}_{\tau\in\mathcal{T}_{T_k}} \mathbb{E}[Z_\tau|\mathcal{F}_{T_k}]$$

where $\mathcal{T}_t$ is the set of all $\mathcal{F}-$ stopping times with values in $\{T_k, T_{k+1}, ..., T\}$.

- From the Snell enveloppe theory, we derive the standard dynamic programming algorithm ($\rightarrow$ "Tsistsiklis-Van Roy" type algorithms).

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

The policy iteration approach...

Let τ_k be the smallest optimal stopping time after T_k.

$$(2) \qquad \begin{cases} \tau_N = T_N \\ \tau_k = T_k \mathbf{1}_{\{Z_{T_k} \geq \mathbb{E}[Z_{\tau_{k+1}} | \mathcal{F}_{T_k}]\}} + \tau_{k+1} \mathbf{1}_{\{Z_{T_k} < \mathbb{E}[Z_{\tau_{k+1}} | \mathcal{F}_{T_k}]\}}. \end{cases}$$

This is a dynamic programming principle on the policy not on the value function $\rightarrow$ "Longstaff-Schwartz" algorithm.
This approach has the practitioners' favour for its robustness.

Difficulty: how to compute the conditional expectations?

. . . in a Markovian context

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

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

  where ψ_k is a measurable function.

- Thanks to the L^2 assumption, ψ_k can be computed by a least-square problem

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

  How to approximate $L^2(\mathcal{L}(X_{T_k}))$?

Many different numerical strategies

- The standard numerical (LS) approach: approximate the space L^2 by a finite dimensional vector space (polynomials, local polynomials, ...)
- Some previous works using deep learning for optimal stopping
  - Michael Kohler, Adam Krzyżak, and Nebojsa Todorovic. Pricing of high-dimensional american options by neural networks. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 20(3):383–410, 2010
  - S. Becker, P. Cheridito, and A. Jentzen. Deep optimal stopping. *Journal of Machine Learning Research*, 20(74):1–25, 2019a
  - Sebastian Becker, Patrick Cheridito, Arnulf Jentzen, and Timo Welti. Solving high-dimensional optimal stopping problems using deep learning, 2019b
  - Bernard Lapeyre and Jérôme Lelong. Neural network regression for Bermudan option pricing. *Monte Carlo Methods Appl.*, 27(3):227–247, 2021

From local approximations to regression trees

- Local approximations look appealing but they do not scale in high dimension.
- In high dimension, the main difficulty is to determine the underlying partition. $\rightarrow$ it cannot be a regular grid.
- Actually, no need of a fine approximation away from the exercising boundary.
- Regression trees provide an algorithmic way to find a "good" grid.
- Theoretical difficulties:
  - Not a linear approximation
  - The approximation does not solve a minimization problem

Regression trees

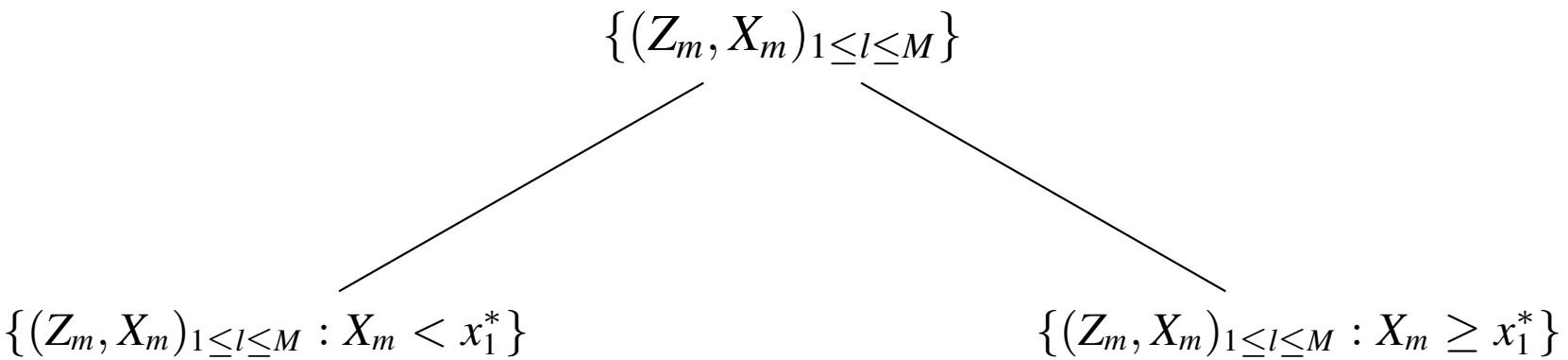
Consider some samples $(Z_m, X_m)_{1\leq m\leq M}$ with values in $\mathbb{R}\times[0,1]$. We want to approximate $\mathbb{E}[Z|X]$. The iterative procedure writes as

$$\inf_{z,z',x^*}\sum_{m=1}^{M}\left(Z_m-\left(z\mathbf{1}_{\{X_m>x^*\}}+z'\mathbf{1}_{\{X_m\leq x^*\}}\right)\right)^2.$$

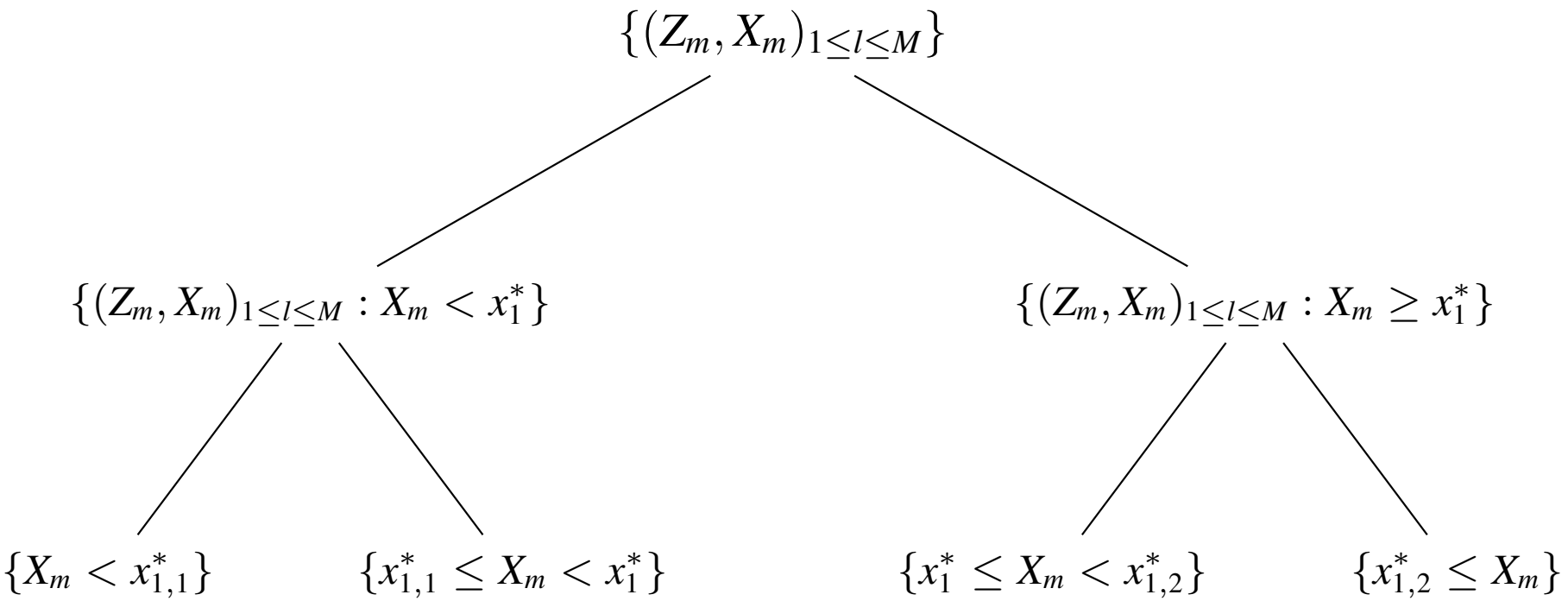
For a fixed x^*, z and z' are known explicitly

$$z=\frac{\sum_{i=1}^{M}Z_i\mathbf{1}_{\{X_i>x^*\}}}{\sum_{i=1}^{M}\mathbf{1}_{\{X_i>x^*\}}};\quad z'=\frac{\sum_{i=1}^{M}Z_i\mathbf{1}_{\{X_i\leq x^*\}}}{\sum_{i=1}^{M}\mathbf{1}_{\{X_i\leq x^*\}}}.$$

Regression trees



Regression trees



Regression trees

- It is sufficient to determine x^* at each level, known as the split point.
- To ensure convergence of the tree, we modify the procedure
  - With a small probability q, choose x^* as the midpoint;
  - With probability $1-q$, choose the optimal $x*$.
- When M goes to infinity,

$$z = \mathbb{E}[Z|\mathbf{1}_{\{X>x^*\}}]; \quad z' = \mathbb{E}[Z|\mathbf{1}_{\{X\leq x^*\}}].$$

- When X is multidimensional, we first choose (at random) the direction of the new split and then apply the one dimensional procedure.

Regression trees

A regression tree $\mathcal{T}_p(X)$ can be written as follows

$$\mathcal{T}_p(X) = \sum_{i=1}^{2^p} \alpha_p^i \mathbf{1}_{\left\{X \in \prod_{j=1}^d [a_p^i(j) - a_p^{i-1}(j))\right\}}$$

with $\left([a_p^{i-1}, a_p^i)\right)_{1 \leq i \leq 2^p}$ forming a partition of $[0,1]^d$ and

$$\alpha_p^i = \mathbb{E}\left[Z | X \in [a_p^{i-1}, a_p^i)\right].$$

$\mathcal{T}_p^M$ denotes the tree built on the sample of size M.

Convergence of the regression tree approximation

Proposition

$$\lim_{p\to\infty} \mathbb{E}\left[\left|\mathcal{T}_p(X) - \mathbb{E}[Y|X]\right|^2\right] = 0.$$

Proposition

For every p, $\lim_{M\to\infty} \mathcal{T}_p^M(X_1) \to \mathcal{T}_p(X_1)$ *a.s.*

In practice, use random forests

Consider a fraction of the initial data chosen at random and build a regression tree $\mathcal{T}_{p,k}^M$. Repeat this procedure, to obtain a collection of B trees: $(\mathcal{T}_{p,k}^M)_{1\leq k\leq B}$. The random forest approximation is defined by

$$\frac{1}{B}\sum_{k=1}^{B}\mathcal{T}_{p,k}^M(X).$$

It is an ensemble method so it looses explainability but provides a more accurate approximation.

The LS algorithm

Longstaff-Schwartz type algorithms rely on direct approximation of *stopping times* and use of *the same sample paths* for all time steps (large computational savings).

- Backward approximation of iteration policy using (2),

$$\begin{cases} \widehat{\tau}_N^{p,} = N \\ \widehat{\tau}_k^p = T_k \, \mathbf{1}_{\left\{Z_{T_k} \geq \mathcal{T}_p(X_{T_k})\right\}} + \widehat{\tau}_{k+1}^p \, \mathbf{1}_{\left\{Z_k < \mathcal{T}_p(X_{T_k})\right\}} \end{cases}$$

- where $\mathcal{T}_p(X_{T_k})$ is a regression tree approximation of $\mathbb{E}[Z_{\widehat{\tau}_{k+1}^p} | X_{T_k}]$.
- Price approximation: $U_0^p = \max\left(Z_0, \mathbb{E}\left(Z_{\widehat{\tau}_1^p}\right)\right)$.

The LS algorithm

- Paths $X_{T_0}^{(m)}, X_{T_1}^{(m)}, \ldots, X_{T_N}^{(m)}$ and payoff paths $Z_{T_0}^{(m)}, Z_{T_1}^{(m)}, \ldots, Z_{T_N}^{(m)}$, $m = 1, \ldots, M$.
- Backward approximation of iteration policy,

$$\begin{cases} \widehat{\tau}_N^{p,(m)} = T_N \\ \widehat{\tau}_k^{p,(m)} = T_n \mathbf{1}_{\left\{Z_{T_k}^{(m)} \geq \mathcal{T}_p^M(X_{T_k}^{(m)})\right\}} + \widehat{\tau}_{k+1}^{p,(m)} \mathbf{1}_{\left\{Z_{T_n}^{(m)} < \mathcal{T}_p^M(X_{T_k}^{(m)})\right\}} \end{cases}$$

- whee $\mathcal{T}_p^M(X_{T_k}^{(m)})$ is a regression tree approximation of "$\mathbb{E}[Z_{\widehat{\tau}_{k+1}^{p,M}} | X_{T_k}^{(m)}]$".
- Price approximation: $U_0^{p,M} = \max\left(Z_0, \frac{1}{M}\sum_{m=1}^M Z_{\widehat{\tau}_1^{p,(m)}}^{(m)}\right)$.

Some related works

- 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: Emmanuelle Clément, Damien Lamberton, and

  Philip 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, $p \to +\infty$
  - $U_0^{p,M}$ converge to U_0^p, $M \to +\infty$ a.s.
  - A central limit theorem but the limiting variance is unknown.

The modified algorithm

- In LS algorithm replace the approximation on a Hilbert basis $\Phi_p(.;\theta)$ by a random forest. This is not a vector space approximation (non linear).
- A random tree does not even write as the solution of a minimization problem.
- Our aim: extending the proof of the (a.s.) convergence results and performing some numerical tests. The linear structure played a key role in Clément et al. [2002]'s analysis.

The result

Theorem 1

Under good assumptions

- *The RF approximation converges when the depth p of the trees goes to infinity.*

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

- *SLLN: For every p and $k = 1, \dots, N$,*

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

Put option in the Heston model

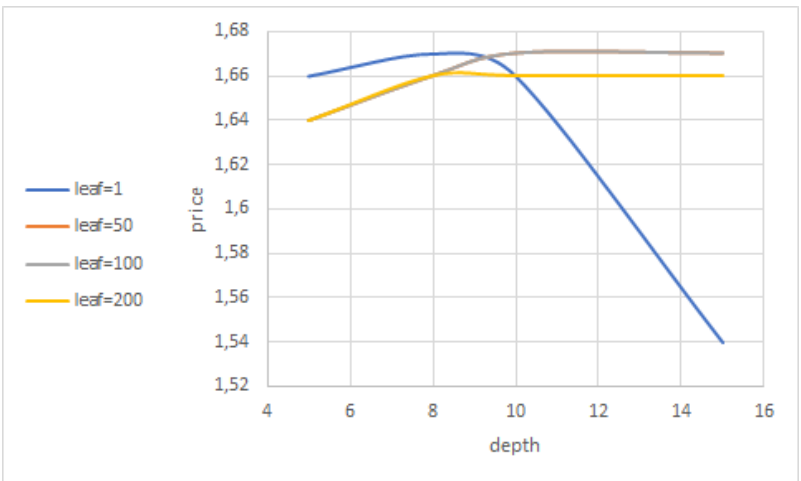


Figure: Regression trees

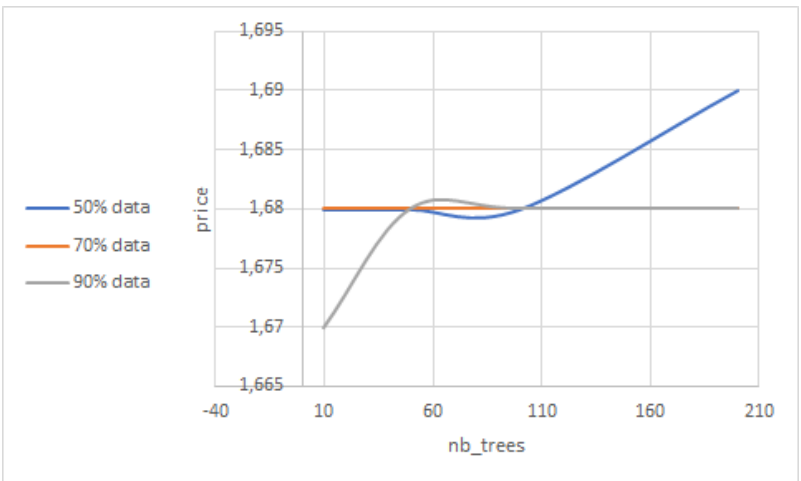


Figure: Random forests

A call on maximum in the Black Scholes model

depth	samples per leaf	price
50	50	66,89
50	100	66.88
100	50	67.13
100	100	67.31
200	50	67.16
200	100	67.28

Table: A call option on the maximum of 50 assets with regression trees

nb trees	sample fraction	price
10	50%	68,32
10	70%	68,32
10	90%	68,29

Table: A call option on the maximum of 50 assets with random forests

Conclusion

- Learn the continuation value using a RF instead of a polynomial regression.
- RF do not help much for low dimensional problems but do scale better.
- Reasonable local approximation in high dimension.
- Use random forests with many “small” trees.

Bibliography (1)

S. Becker, P. Cheridito, and A. Jentzen. Deep optimal stopping. *Journal of Machine Learning Research*, 20(74):1–25, 2019a.

Sebastian Becker, Patrick Cheridito, Arnulf Jentzen, and Timo Welti. Solving high-dimensional optimal stopping problems using deep learning, 2019b.

Emmanuelle Clément, Damien Lamberton, and Philip 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, Huyên 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. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 20(3):383–410, 2010.

Bernard Lapeyre and Jérôme Lelong. Neural network regression for Bermudan option pricing. *Monte Carlo Methods Appl.*, 27(3):227–247, 2021.

Bibliography (2)

Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. ISBN 3-540-52013-9. Isoperimetry and processes.

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.

Reuven Y. Rubinstein and Alexander Shapiro. *Discrete event systems*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, 1993. ISBN 0-471-93419-4. 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.