Pricing American options using martingale bases

Jérôme Lelong

Grenoble Alpes University

International Monte Carlo Conference, Paris – July 8, 2016

Outline

Framework (1)

- Consider a $d'-$dimensional financial market driven by a $d-$dimensional Brownian motion B, with $d' \leq d$.
- The discounted payoff process writes $\left(Z_t = e^{-\int_0^t r_s ds}\, \phi(S_t)\right)_{t \leq T}$.
  Assume $\mathbb{E}\left[\sup_t Z_t^2\right] < \infty$.
- Consider an American option. Its time–t discounted price is given by

$$U_t = \operatorname{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E}[Z_\tau | \mathcal{F}_t]$$

  where $\mathcal{T}_t$ is the set of all $\mathcal{F}-$ stopping times with values in $[t, T]$.

Dual price (1)

The *Snell envelope* process $(U_t)_{0\leq t\leq T}$ admits a Doob–Meyer decomposition

$$U_t = U_0 + M_t^\star - A_t^\star.$$

[Rogers, 2002]: $$U_0 = \inf_{M\in H_0^1} \mathbb{E}\left[\sup_{0\leq t\leq T} (Z_t - M_t)\right] = \mathbb{E}\left[\sup_{0\leq t\leq T} (Z_t - M_t^\star)\right]$$

- This problem admits more than a single solution.
- For any stopping time τ smaller than the largest optimal strategy,

$$U_0 = \inf_{M\in H_0^1} \mathbb{E}\left[\sup_{\tau\leq t\leq T} (Z_t - M_t)\right] = \mathbb{E}\left[\sup_{\tau\leq t\leq T} (Z_t - M_t^\star)\right].$$

Dual price (2)

- Some of the martingales M attaining the infimum are surely optimal

$$U_0 = \sup_{0 \le t \le T} (Z_t - M_t) \quad a.s.$$

- From [Schoenmakers et al., 2013], any martingale satisfying

$$\text{Var}\left(\sup_{0 \le t \le T} (Z_t - M_t) \right) = 0$$

  is surely optimal.
- From [Jamshidian, 2007], for any optimal stopping time τ and any surely optimal martingale M,

$$(M_{t\wedge\tau})_t = (M^\star_{t\wedge\tau})_t.$$

Dual price (3)

With our square integrability assumption, we can rewrite the minimization problem as

$$U_0 = \inf_{\substack{X \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}) \\ \text{s.t. } \mathbb{E}[X]=0}} \mathbb{E}\left[\sup_{0 \le t \le T} (Z_t - \mathbb{E}[X|\mathcal{F}_t])\right].$$

How to approximate $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ by a finite dimensional vector space in which conditional expectations are tractable in a closed form?

Wiener chaos expansion ($d = 1$)

Let H_i be the $i - th$ Hermite polynomial defined by

$$H_0(x) = 1; \qquad H_i(x) = (-1)^i \, e^{x^2/2} \frac{d^i}{dx^i}(e^{-x^2/2}), \text{ for } i \geq 1.$$

- $H_i' = H_{i-1}$ with the convention $H_{-1} = 0$.
- If $X, Y \sim \mathcal{N}(0, 1)$ and form a Gaussian vector,

$$\mathbb{E}[H_i(X)H_j(Y)] = i! \, (\mathbb{E}[XY])^i \; \mathbf{1}_{\{i=j\}}.$$

Truncated Wiener chaos expansion ($d = 1$)

Take a regular grid $0 = t_0 < t_1 < \cdots < t_n$ with step h.
Define the truncated Wiener chaos space of order p

$$\mathcal{H}_p = \text{span}\left\{ \prod_{i=1}^{n} H_{\alpha_i}(G_i) \ : \ \alpha \in \mathbb{N}^n, \ \|\alpha\|_1 = p \right\}$$

with $G_i = \frac{B_{t_i} - B_{t_{i-1}}}{\sqrt{h}}$.
For $F \in L^2(\Omega, \mathcal{F}_T)$, we introduce the truncated chaos expansion of order p

$$C_{p,n}(F) = \sum_{\alpha \in A_{p,n}} \lambda_\alpha \prod_{i \geq 1} H_{\alpha_i}(G_i)$$

where $A_{p,n} = \{\alpha \in \mathbb{N}^n \ : \ \|\alpha\|_1 \leq p\}$ with $\|\alpha\|_1 = \sum_{i \geq 0} \alpha_i$.
In the following we write,

$$C_{p,n}(F) = \sum_{\alpha \in A_{p,n}} \lambda_\alpha \widehat{H}_\alpha(G_1, \ldots, G_n)$$

Key property of the truncated Wiener chaos expansion

For $k \leq n$,

$$\mathbb{E}[C_{p,n}(F)|\mathcal{F}_{t_k}] = \sum_{\alpha \in A^k_{p,n}} \lambda_\alpha \, \widehat{H}_\alpha(G_1, \dots, G_n)$$

with $A^k_{p,n} = \{\alpha \in \mathbb{N}^n \ : \ \|\alpha\|_1 \leq p, \ \alpha_\ell = 0 \ \forall \ell > k\}$.

"Computing $\mathbb{E}[\cdot|\mathcal{F}_{t_k}]$" $\Leftrightarrow$ "Dropping all non $\mathcal{F}_{t_k}$ – measurable terms"

Extension to the multi–dimensional case

The truncated Wiener chaos of order $p \geq 0$ is given by

$$\left\{ \prod_{j=1}^{d} \widehat{H}_{\alpha^j}(G_1^j, \dots, G_n^j) \;:\; \alpha \in (\mathbb{N}^n)^d,\ \|\alpha\|_1 \leq p \right\}.$$

We introduce the truncated chaos expansion of order p of $F \in L^2(\Omega, \mathcal{F}_T)$

$$C_{p,n}(F) = \sum_{\alpha \in A_{p,n}^{\otimes d}} \lambda_\alpha \widehat{H}_\alpha^{\otimes d}(G_1, \dots, G_n) = C_{p,n}(\lambda)$$

where

$$A_{p,n}^{\otimes d} = \left\{ \alpha \in (\mathbb{N}^n)^d \;: \|\alpha\|_1 \leq p \right\},$$

$$\widehat{H}_\alpha^{\otimes d}(G_1, \dots, G_n) = \prod_{j=1}^{d} \widehat{H}_{\alpha_j}(G_1^j, \dots, G_n^j) \quad \forall \alpha \in (\mathbb{N}^n)^d.$$

Return to the American option price

We approximate the original problem

$$\inf_{\substack{X \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}) \\ \text{s.t. } \mathbb{E}[X]=0}} \mathbb{E}\left[\sup_{0 \le t \le T} (Z_t - \mathbb{E}[X|\mathcal{F}_t])\right]$$

by

$$\inf_{\substack{\lambda \in \mathbb{R}^{A_{p,n}^{\otimes d}} \\ \text{s.t. } \lambda_0 = 0}} V_{p,n}(\lambda) \tag{1}$$

with

$$V_{p,n}(\lambda) = \mathbb{E}\left[\max_{0 \le k \le n} (Z_{t_k} - \mathbb{E}[C_{p,n}(\lambda)|\mathcal{F}_{t_k}])\right].$$

Properties of the minimization problem (1)

Proposition 1

The minimization problem (1) has at least one solution.

- The function $V_{p,n}$ is clearly convex (maximum of affine functions).
- Not strongly convex but,

$$V_{p,n}(\lambda) \geq \frac{|\lambda|}{2} \inf_{\mu \in \mathbb{R}^{A_{p,n}^{\otimes d}}, |\mu|=1} \mathbb{E}\left[|C_{p,n}(\mu)|\right].$$

Properties of the minimization problem (2)

$$\mathcal{I}(\lambda, Z, G) = \{0 \le k \le n \ : \ \text{the pathwise maximum is attained at time } k\}.$$

Proposition 2

Let $p \ge 1$. Assume that

$$\forall 1 \le r \le k \le n,\ \forall F\ \mathcal{F}_{t_k} - measurable,\ F \in \mathcal{C}_{p-1,n},\ F \ne 0,$$
$$\exists\, 1 \le q \le d \text{ s.t. } \mathbb{P}\left(\forall t \in]t_{r-1}, t_r],\ D_t^q Z_{t_k} + F = 0 \;\middle|\; Z_{t_k} > 0\right) = 0.$$

Then, the function $V_{p,n}$ is differentiable at all points $\lambda \in \mathbb{R}^{A_{p,n}^{\otimes d}}$ with no zero component and its gradient $\nabla V_{p,n}$ is given by

$$\nabla V_{p,n}(\lambda) = \mathbb{E}\left[\mathbb{E}\left[\widehat{H}^{\otimes d}(G_1, \dots, G_n) \mid \mathcal{F}_{t_i}\right]_{|i=\mathcal{I}(\lambda,Z,G)}\right].$$

Properties of the minimization problem (3)

- Differentiability is ensured as soon as $\mathcal{I}(\lambda, Z, G)$ is a.s. reduced to a unique element: purpose of the blue condition.
- Alternative approach by [Belomestny, 2013]: use smoothing techniques instead (see [Nesterov, 2004]). General idea:

$$\text{Replace } \max_k a_k \text{ by } p^{-1} \log \left(\sum_k \exp(p\ a_k) \right).$$

- Let $\lambda^\sharp$ be a solution, $V_{p,n}(\lambda^\sharp_{p,n}) = \inf_\lambda V_{p,n}(\lambda)$. Then $\nabla V_{p,n}(\lambda^\sharp_{p,n}) = 0$.

Convergence to the true solution (1)

Proposition 3

The solution of the minimization problem (1), $V_{p,n}(\lambda^{\sharp}_{p,n})$, converges to the price of the American options when both p and n go to infinity and moreover

$$0 \leq V_{p,n}(\lambda^{\sharp}_{p,n}) - U_0 \leq 2 \left\| M^{\star}_T - C_{p,n}(M^{\star}_T) \right\|_2 .$$

- Consider a Bermudan option with exercising dates $t_0, \cdots, t_n$ and discounted payoff $(Z_{t_k})_k$ adapted to the discrete time filtration generated by the Brownian increments only. Then, $V_{p,n}(\lambda^{\sharp}_{p,n})$ converges to the price of the Bermudan option when p only goes to infinity.

Practically solving the optimization problem (1)

We approximate the solution of

$$V_{p,n}(\lambda^{\sharp}_{p,n}) = \inf_{\lambda \in A^{\otimes d}_{p,n}} V_{p,n}(\lambda) = \inf_{\lambda \in A^{\otimes d}_{p,n}} \mathbb{E}\left[\max_{0 \le k \le n} (Z_{t_k} - \mathbb{E}[C_{p,n}(\lambda)|\mathcal{F}_{t_k}])\right]$$

by introducing the well–known *Sample Average Approximation* (see [Rubinstein and Shapiro, 1993]) of $V_{p,n}$ defined by

$$V^m_{p,n}(\lambda) = \frac{1}{m} \sum_{i=1}^{m} \max_{0 \le k \le n} \left(Z^{(i)}_{t_k} - \mathbb{E}[C^{(i)}_{p,n}(\lambda)|\mathcal{F}_{t_k}]\right).$$

Note that the conditional expectation boils down to truncating the chaos expansion and hence is tractable in a closed form.

Practically solving the optimization problem (2)

For large enough m, $V^m_{p,N}$ is convex, a.s. differentiable and tends to infinity at infinity. Then, there exits $\lambda^m_{p,n}$ such that

$$V^m_{p,n}(\lambda^m_{p,n}) = \inf_{\lambda \in \mathbb{R}^{A^{\otimes d}_{p,n}}} V^m_{p,n}(\lambda).$$

Proposition 4

$V^m_{p,n}(\lambda^m_{p,n})$ converges a.s. to $V_{p,n}(\lambda^\sharp_{p,N})$ when $m \to \infty$.
The distance from $\lambda^m_{p,n}$ to the set of minimizers of $V_{p,n}$ converges to zero as m goes to infinity.

Practically solving the optimization problem (3)

Write $M_k(\lambda) = \mathbb{E}[C_{p,n}(\lambda)|\mathcal{F}_{t_k}]$ for $0 \le k \le n$.

Proposition 5

Assume $\lambda^{\sharp}_{p,n}$ *is unique. Then,*

$$\frac{1}{m}\sum_{i=1}^{m}\left(\max_{0\le k\le n} Z^{(i)}_{t_k} - M^{(i)}_k(\lambda^m_{p,n})\right)^2 - V^m_{p,n}(\lambda^m_{p,n})^2$$

is a convergent estimator of $\mathrm{Var}(\max_{k\le 0\le n} Z_{t_k} - M_k(\lambda^{\sharp}_{p,n}))$ *and moreover, if* $\lambda^m_{p,n}$ *is bounded,*

$$\lim_{m\to\infty} m \,\mathrm{Var}\left(V^m_{p,n}(\lambda^m_{p,n})\right) = \mathrm{Var}(\max_{k\le 0\le n} Z_{t_k} - M_k(\lambda^{\sharp}_{p,n})).$$

The algorithm: bespoke martingales

Define the first time the option goes in the money by

$$\tau_0 = \inf\{k \geq 0 \ : \ Z_{t_k} > 0\} \wedge n.$$

Consider martingales only starting once the option has been in the money

$$N_k(\lambda) = M_k(\lambda) - M_{k \wedge \tau_0}(\lambda).$$

In the dual price, "$\max_{0 \leq k \leq n}$" can be shrunk to "$\max_{\tau_0 \leq k \leq n}$".
Using Doob's stopping theorem, we have

$$\mathbb{E}\left[\max_{\tau_0 \leq k \leq n}(Z_{t_k} - M_k(\lambda))\right] = \mathbb{E}\left[\max_{\tau_0 \leq k \leq n}(Z_{t_k} - (M_k(\lambda) - M_{\tau_0}(\lambda)))\right]$$

The martingales $M(\lambda)$ or $N(\lambda)$ lead to the same minimum value.
The set of martingales N^λ is far more efficient from a practical point of view.

The algorithm: a gradient descent with line search

$x_0 \leftarrow 0, k \leftarrow 0, \gamma \leftarrow 1, d_0 \leftarrow 0, v_0 \leftarrow \infty$;
while *True* **do**
  Compute $v_{k+1/2} \leftarrow \tilde{V}^m_{p,n}(x_k - \gamma\alpha_k d_k)$;
  if $v_{k+1/2} < v_k$ **then**
    $x_{k+1} \leftarrow x_k - \gamma\alpha_k d_k$;
    $v_{k+1} \leftarrow v_{k+1/2}$;
    $d_{k+1} \leftarrow \nabla\tilde{V}^m_{p,n}(x_{k+1})$;
    if $\frac{|v_{k+1} - v_k|}{v_k} \le \varepsilon$ **then return**;
  else
    $\gamma \leftarrow \gamma/2$;
  end
end

The algorithm: a gradient descent with line search

$x_0 \leftarrow 0$, $k \leftarrow 0$, $\gamma \leftarrow 1$, $d_0 \leftarrow 0$, $v_0 \leftarrow \infty$;
while *True* **do**
  Compute $v_{k+1/2} \leftarrow \tilde{V}^m_{p,n}(x_k - \gamma \alpha_k d_k)$;
  if $v_{k+1/2} < v_k$ **then**
    $x_{k+1} \leftarrow x_k - \gamma \alpha_k d_k$;
    $v_{k+1} \leftarrow v_{k+1/2}$;
    $d_{k+1} \leftarrow \nabla \tilde{V}^m_{p,n}(x_{k+1})$;
    if $\frac{|v_{k+1} - v_k|}{v_k} \leq \varepsilon$ **then return**;
  else
    $\gamma \leftarrow \gamma/2$;
  end
end

Take $\alpha_\ell = \dfrac{\tilde{V}^m_{p,n}(x_\ell) - v^\sharp}{\left\| \nabla \tilde{V}^m_{p,n}(x_\ell) \right\|^2}$, see [Polyak, 1987], but with the European price instead of the American one for $v^\sharp$.

Some remarks on the algorithm

- Given the expression of $V^m_{p,n}$, both the value function and its gradient are computed at the same time without extra cost.

$$V_{p,n}(\lambda) = \mathbb{E}\left[\max_{\tau_0 \le k \le n} \left(Z_{t_k} - \mathbb{E}[\lambda \cdot H^{\otimes d}(G_1, \cdots, G_n) | \mathcal{F}_{t_k}]\right)\right],$$
$$= \mathbb{E}[Z_{t_{\mathcal{I}(\lambda,Z,G)}}] - \lambda \cdot \nabla \tilde{V}_{p,n}(\lambda).$$

- Checking the admissibility of a step γ costs as much as updating x_k.
- The algorithm is *almost* embarrassingly parallel:
  - Few iterations of the gradient descent are required (≈ 10).
  - Each iteration is fully parallel: each process treats its bunch of paths.
  - No demanding centralized computations
  - Very little communication: a few broadcasts only.

Parallel implementation

In parallel *Generate* $(G^{(1)}, Z^{(1)}), \dots, (G^{(m)}, Z^{(m)})$ m $x_0 \leftarrow 0 \in \mathbb{R}^{A_{p,n}^{\otimes d}}$;
while *True* **do**
  Broadcast x_ℓ, d_ℓ, γ, α_ℓ;
  In parallel *Compute* $\max_{\tau_0 \le k \le n}(Z_{t_k}^{(i)} - N_k^{(i)}(x_\ell - \gamma \alpha_\ell d_\ell))$;
  Make a reduction of *the above contributions to obtain* $\tilde{V}_{p,n}^m(x_{\ell+1/2})$ *and* $\nabla \tilde{V}_{p,n}^m(x_{\ell+1/2})$;
  $v_{\ell+1/2} \leftarrow \tilde{V}_{p,n}^m(x_\ell - \gamma \alpha_\ell d_\ell)$;
  if $v_{\ell+1/2} < v_\ell$ **then**
    $x_{\ell+1} \leftarrow x_\ell - \gamma \alpha_\ell d_\ell$;
    $v_{\ell+1} \leftarrow v_{\ell+1/2}$; $d_{\ell+1} \leftarrow \nabla \tilde{V}_{p,n}^m(x_{\ell+1})$;
    if $\frac{|v_{\ell+1} - v_\ell|}{v_\ell} \le \varepsilon$ **then return**;
  else
    $\gamma \leftarrow \gamma/2$;
  end
end

Basket option in the BS model

p	n	S_0	price	Stdev	time (sec.)	reference price
2	3	100	2.27	0.029	0.17	2.17
3	3	100	2.23	0.025	0.9	2.17
2	3	110	0.56	0.014	0.07	0.55
3	3	110	0.53	0.012	0.048	0.55
2	6	100	2.62	0.021	0.91	2.43
3	6	100	2.42	0.021	14	2.43
2	6	110	0.61	0.012	0.33	0.61
3	6	110	0.55	0.008	10	0.61

TAB.: Prices for the put basket option with parameters $T = 3$, $r = 0.05$, $K = 100$, $\rho = 0$, $\sigma^j = 0.2$, $\delta^j = 0$, $d = 5$, $\omega^j = 1/d$, $m = 20,000$.

Call option on the maximum of a basket

d	p	m	S_0	price	Stdev	time (sec.)	reference price
2	2	20,000	90	10.18	0.07	0.4	8.15
2	3	20,000	90	8.5	0.05	4.1	8.15
2	2	20,000	100	16.2	0.06	0.54	14.01
2	3	20,000	100	14.4	0.06	5.6	14.01
5	2	20,000	90	21.2	0.09	2	16.77
5	3	40,000	90	16.3	0.05	210	16.77
5	2	20,000	100	30.7	0.09	3.4	26.34
5	3	40,000	100	26.0	0.05	207	26.34

TAB.: Prices for the call option on the maximum of d assets with parameters $T = 3$, $r = 0.05$, $K = 100$, $\rho = 0$, $\sigma^j = 0.2$, $\delta^j = 0.1$, $n = 9$.

Scalability of the parallel algorithm

The tests were run on a BullX DLC supercomputer containing 3204 cores.

#processes	time (sec.)	efficiency
1	4365	1
2	2481	0.99
4	1362	0.90
16	282	0.84
32	272	0.75
64	87	0.78
128	52	0.73
256	34	0.69
512	10.7	0.59

TAB.: Scalability of the parallel algorithm on the 40−dimensional geometric put option described above with $T = 1$, $r = 0.0488$, $K = 100$, $\sigma^j = 0.3$, $\rho = 0.1$, $\delta^j = 0$, $n = 9$, $p = 2$, $m = 200,000$.

Conclusion

- Purely optimization approach. No need of an optimal strategy.
- The problem is in large dimension but convex.
- *Almost* embarrassingly parallel and scales very well.
- Can deal with path dependent options

- Belomestny, D. (2013).
  Solving optimal stopping problems via empirical dual optimization.
  Ann. Appl. Probab., 23(5):1988–2019.
- Jamshidian, F. (2007).
  The duality of optimal exercise and domineering claims: a Doob-Meyer decomposition approach to the Snell envelope.
  Stochastics, 79(1-2):27–60.
- Nesterov, Y. (2004).
  Smooth minimization of non-smooth functions.
  Mathematical Programming, 103(1):127–152.
- Polyak, B. T. (1987).
  Introduction to optimization.
  Optimization Software.
- Rogers, L. C. G. (2002).
  Monte Carlo valuation of American options.
  Math. Finance, 12(3):271–286.

- Rubinstein, R. Y. and Shapiro, A. (1993).
  Discrete event systems.
  Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester.
  Sensitivity analysis and stochastic optimization by the score function method.
- Schoenmakers, J., Zhang, J., and Huang, J. (2013).
  Optimal dual martingales, their analysis, and application to new algorithms for bermudan products.
  SIAM Journal on Financial Mathematics, 4(1):86–116.