On the complexity of approximation algorithms for American options

Denis Belomestny

Duisburg-Essen University

Paris, 16 May 2014

Bermudan Options

- Let $(X_t,\, 0 \leq t \leq T)$ be a Markov process valued in $\mathbb{R}^d$ and defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathrm{P})$
- Let $0 = t_0 < t_1 < t_2 < \ldots < t_J = T$ be a finite set of exercise opportunities

Definition

If exercised at time t_j, $j = 1, \ldots, J$, the option pays $g_j(X_{t_j})$, for some known functions $g_0, g_1, \ldots, g_J$ mapping $\mathbb{R}^d$ into $[0, \infty)$.

Notation

$$Z_j := X_{t_j}, \quad j = 1, \ldots, J.$$

Bermudan Options

- Let $\mathcal{T}_j$ be the set of stopping times taking values in $\{j, j+1, \ldots, J\}$

The equilibrium price $V_j^*(x)$ of the Bermudan option at time j in state x is its value under an optimal exercise policy:

$$V_j^*(z) = \sup_{\tau \in \mathcal{T}_j} \mathrm{E}[g_\tau(Z_\tau)|Z_j = z], \quad z \in \mathbb{R}^d.$$

Question

How to compute the fair price of the Bermudan options numerically?

Dynamic Programming Principle

- Continuation values

$$C_j^*(z) := \mathrm{E}[V_{j+1}^*(Z_j)|Z_j = z], \quad j = 0, \ldots, J-1,$$

satisfy the dynamic programming principle

$$\begin{aligned} C_J^*(z) &= 0, \\ C_j^*(z) &= \mathrm{E}[\max(g_{j+1}(Z_{j+1}), C_{j+1}^*(Z_{j+1}))|Z_j = z]. \end{aligned}$$

Observation

The use of the d.p.p. is relatively straightforward in low dimensions. However, many problems arising in practice are high dimensional !

Nested conditional expectations

- Problem: How to approximate the nested conditional expectations in the backward dynamic programming algorithm?
- Naive approach: Use nested simulations.

Infeasible: Computational cost explodes rapidly with the number of exercise dates.

Definition

Fast approximation methods (regression methods, stochastic mesh method and etc.) construct the estimates $C_1, \dots C_J$ recursively without use of nested simulations.

Comparison of fast approximation methods

Question

How to compare different fast approximation methods ?

Definition

A method has a complexity $\mathcal{C}(\varepsilon)$, if it requires $\mathcal{C}(\varepsilon)$ numerical operations to estimate V_0^* with the mean squared error ε.

Observation

Fast approximation methods may have rather large complexity.

Generic approximation algorithm

Given a sequence of estimates $C_{k,j}$, $j = 1, \ldots, J$, based on k training paths, we can define

$$V_0^{n,k} = \frac{1}{n} \sum_{r=1}^{n} g_{\tau_k^{(r)}}(Z_{\tau_k^{(r)}}^{(r)})$$

with

$$\tau_k^{(r)} = \inf\{0 \leq j \leq \mathcal{J} : g_j(Z_j^{(r)}) \geq C_{k,j}(Z_j^{(r)})\}.$$

1. Construction of the estimates $C_{k,j}$, $j = 1, \ldots, J$, on k training paths.
2. Construction of the low-biased estimate $V_0^{n,k}$ by evaluating functions $C_{k,j}$, $j = 1, \ldots, J$, on each of new n testing trajectories.

General classification

Assumption (AP)

For any $k \in \mathbb{N}$, the estimates $C_{k,0}(z), \ldots, C_{k,\mathcal{J}-1}(z)$ are defined on some filtered probability space $(\Omega^k, \mathcal{F}^k, \mathrm{P}^k)$ which is independent of $(\Omega, \mathcal{F}, \mathrm{P})$.

Assumption (AC)

For any $j = 1, \ldots, \mathcal{J}$, the cost of constructing $C_{k,j}$ on k training paths is of order $k \times k^{\varkappa_1}$ for some $\varkappa_1 > 0$ and the cost of evaluating $C_{k,j}(z)$ in a new point z is of order $k^{\varkappa_2}$ for some $\varkappa_2 > 0$.

General classification

Assumption (AQ)

There is a sequence of positive real numbers γ_k *with* $\gamma_k \to 0,\ k \to \infty$ *such that*

$$\mathrm{P}^k \left(\sup_{z \in \mathcal{S}} \left| C_{k,j}(z) - C_j^*(z) \right| > \eta \sqrt{\gamma_k} \right) < B_1 e^{-B_2 \eta}, \quad \eta > 0$$

for any compact set $\mathcal{S}$ *and some constants* $B_1 > 0$ *and* $B_2 > 0$.

General classification

Observation

Given (AC), the overall complexity of a fast approximation algorithm is proportional to

$$k^{1+\varkappa_1} + n \times k^{\varkappa_2},$$

where the first term represents the cost of constructing the estimates $C_{k,j}$, $j = 1, \ldots, J$, *on k training paths and the second one gives the cost of evaluating the estimated continuation values on n testing paths.*

Example: Global regression

- Simulate k trajectories of Z all starting from z at $j = 0$.
- Suppose that the estimates $C_{k,J-1}, \ldots, C_{k,j+1}$ are already constructed.
  - Let $\boldsymbol{\alpha}_j^k = (\alpha_{j,1}^k, \ldots, \alpha_{j,L}^k)$ be a solution of

$$\operatorname*{arginf}_{\boldsymbol{\alpha} \in \mathbb{R}^L} \sum_{i=1}^{k} \left[\zeta_{j+1,k}(Z_{j+1}^{(i)}) - \alpha_1 \psi_1(Z_j^{(i)}) - \ldots - \alpha_L \psi_L(Z_j^{(i)}) \right]^2$$

with $\zeta_{j+1,k}(z) = \max\left\{ g_{j+1}(z), C_{k,j+1}(z) \right\}.$

  - Define

$$C_{k,j}(z) = \alpha_{j,1}^k \psi_1(z) + \ldots + \alpha_{j,L}^k \psi_L(z), \quad z \in \mathbb{R}^d.$$

- It holds $\mathrm{comp}(\alpha_j^k) \sim k \cdot L^2 + \mathrm{comp}(\alpha_{j+1}^k)$, since $B\alpha_j^k = b$ with

$$B_{p,q} = \frac{1}{k}\sum_{i=1}^{k} \psi_p(Z_j^{(i)})\psi_q(Z_j^{(i)})$$

and

$$b_p = \frac{1}{k}\sum_{i=1}^{k} \psi_p(Z_j^{(i)})\zeta_{k,j+1}(Z_{j+1}^{(i)}),$$

$p, q \in \{1, \ldots, L\}$. Hence

$$\mathrm{comp}(\alpha_j^k) \sim (\mathcal{J} - j) \cdot k \cdot L^2.$$

- The estimates $C_{k,0}(z), \ldots, C_{k,\mathcal{J}-1}(z)$ satisfy (AQ) under some conditions. In order to guarantee (AQ) with $\gamma_k = k^{-\mu}$ for some $\mu > 0$, one has to take $L \asymp k^{\theta}$ for some $\theta > 0$, i.e., $\varkappa_1 = 2\theta$ and $\varkappa_2 = \theta$ in (AC).

Example: Stochastic Mesh Method

Broadie and Glasserman (2004):

$$C_{k,j}(z) = \frac{1}{k} \sum_{i=1}^{k} \zeta_{j+1,k}(Z_{j+1}^{(i)}) \cdot w_{ij}(z),$$

- Weights

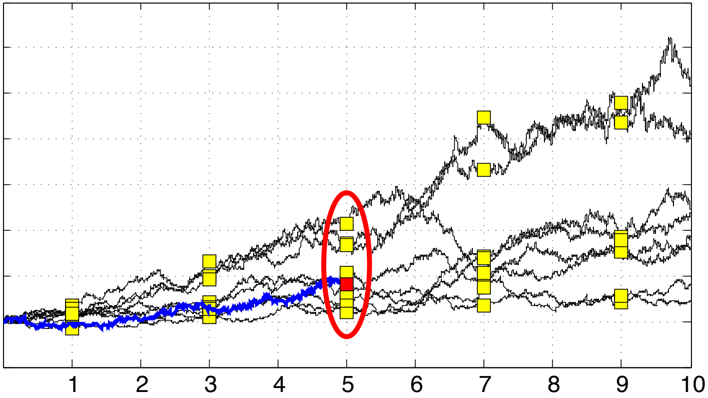
$$w_{ij}(z) = \frac{p_j(z, Z_{j+1}^{(i)})}{\frac{1}{k}\sum_{i=1}^{k} p_j(Z_j^{(i)}, Z_{j+1}^{(i)})},$$

- $p_j(x, y)$ is the transition density from $Z_j = x$ to $Z_{j+1} = y$.

Observation

The complexity of computing $C_{k,j}(z)$ for any fixed z is of order k, provided the transition density $p_j(x, y)$ is analytically known.

Example: Stochastic Mesh Method



$\varkappa_1 = 1$ and $\varkappa_2 = 1$ in (AC)

Lower bound for V_0^* via C_{kj}

- A suboptimal stopping rule:

$$\tau_k = \min\left\{0 \le j \le J : C_{k,j}(Z_j) \le g_j(Z_j)\right\}$$

  with $C_{k,J} \equiv 0$ by definition.
- Fix two natural numbers N and K, and define

$$V_0^{N,K} = \frac{1}{N}\sum_{r=1}^{N} g_{\tau_K^{(r)}}(Z^{(r)}_{\tau_K^{(r)}}).$$

Observation

$V_0^{N,K}$ is low biased, i.e., $\mathrm{E}[V_0^{N,K}] \le V_0^*$.

Idea

Compare different f. a. methods in terms of $\mathrm{E}[(V_0^{N,K} - V_0^*)^2]$.

Complexity of $V_0^{N,K}$

Exsercise boundary or margin assumption:

Assumption (AM)

There exist constants $A > 0$, $\delta_0 > 0$ *and* $\alpha > 0$ *such that*

$$\mathrm{P}\left(|C_j^*(Z_j) - g_j(Z_j)| \leq \delta\right) \leq A\delta^\alpha$$

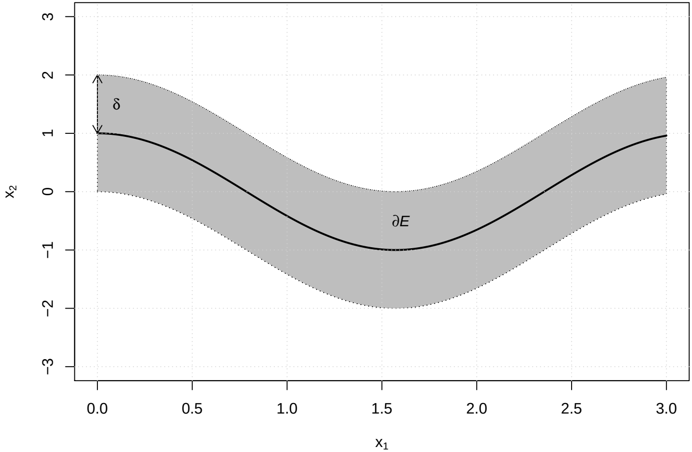
for all $j = 0, \ldots, J$, *and all* $\delta < \delta_0$.

Remark

- *Assumption (AM) characterises the behaviour of* Z *near the exercise boundary* ∂E *with*

$$E = \left\{(j, x) : g_j(x) \geq C_j^*(x)\right\}.$$

Boundary assumption



Complexity of $V_0^{N,K}$

Proposition

Let $\gamma_k = k^{-\mu}, \quad k \in \mathbb{N}$ *for some* $\mu > 0$. *Then for any* $\varepsilon > 0$ *the choice*

$$k^* = \varepsilon^{-\frac{2}{\mu(1+\alpha)}}, \quad n^* = \varepsilon^{-2}$$

leads to

$$\mathrm{E}\left[V_0^{n^*,k^*} - V_0^*\right]^2 \leq \varepsilon^2,$$

and the complexity of the estimate $V_0^{n^*,k^*}$ *is bounded from above by* $\mathcal{C}_{n^*,k^*}(\varepsilon)$ *with*

$$\mathcal{C}_{n^*,k^*}(\varepsilon) \lesssim \varepsilon^{-2\cdot\max\left(\frac{\varkappa_1+1}{\mu(1+\alpha)},\, 1+\frac{\varkappa_2}{\mu(1+\alpha)}\right)}, \quad \varepsilon \to 0.$$

Discussion

- The complexity of $V_0^{N,K}$ can be as large as ε^{-q} for any $q > 2$.
- The smaller is the margin parameter α, the larger is the complexity.

Remark

To compare: the complexity of computing $\mathrm{E}[g(X_T)]$ by Monte Carlo is of order ε^{-2}.

Question

Is it possible to reduce the complexity of $V_0^{N,K}$?

Multilevel idea

- Fix $L \in \mathbb{N}$ and $\mathbf{k} = (k_1, \ldots, k_L) \in \mathbb{N}^L$ with $k_1 < \ldots < k_L$.
- Let h_k be a sequence of positive numbers tending to 0.
- Write a telescopic sum

$$\mathrm{E}\left[g_{\tau_{k_L}}\left(Z_{\tau_{k_L}}\right)\right] = \mathrm{E}\left[g_{\tau_{k_0}}\left(Z_{\tau_{k_0}}\right)\right] + \sum_{l=1}^{L} \mathrm{E}\left[g_{\tau_{k_l}}\left(Z_{\tau_{k_l}}\right) - g_{\tau_{k_{l-1}}}\left(Z_{\tau_{k_{l-1}}}\right)\right]$$

with

$$\tau_k^{(r)} = \inf\left\{0 \leq j \leq J : g_j(Z_j^{(r)}) > C_{k,j}(Z_j^{(r)})\right\}.$$

Multilevel approach

Fix $\mathbf{n} = (n_1, \ldots, n_L) \in \mathbb{N}^L$ and simulate n_l trajectories of the process Z to approximate $\mathrm{E}\left[g_{\tau_{k_l}}\left(Z_{\tau_{k_l}}\right) - g_{\tau_{k_{l-1}}}\left(Z_{\tau_{k_{l-1}}}\right)\right]$.

Define

$$V_0^{\mathbf{n},\mathbf{k}} = \frac{1}{n_0}\sum_{r=1}^{n_0} g_{\tau_{k_0}^{(r)}}\left(Z^{(r)}_{\tau_{k_0}^{(r)}}\right) + \sum_{l=1}^{L}\frac{1}{n_l}\sum_{r=1}^{n_l}\left[g_{\tau_{k_l}^{(r)}}\left(Z^{(r)}_{\tau_{k_l}^{(r)}}\right) - g_{\tau_{k_{l-1}}^{(r)}}\left(Z^{(r)}_{\tau_{k_{l-1}}^{(r)}}\right)\right]$$

with

$$\tau_k^{(r)} = \inf\left\{0 \le j \le J : g_j(Z_j^{(r)}) > C_{k,j}(Z_j^{(r)})\right\}.$$

Multilevel approach: complexity

Proposition

Under the choice $k_l^* = k_0 \cdot \theta^l$, $l = 0, 1, \ldots, L$, *with* $\theta > 1$,

$$L = \left\lceil \frac{2}{\mu(1+\alpha)} \log_\theta \left(\varepsilon^{-1} \cdot k_0^{-\mu(1+\alpha)/2} \right) \right\rceil$$

and

$$n_l^* = \varepsilon^{-2} \left(\sum_{i=1}^{L} \sqrt{k_i^{(\varkappa_2 - \mu\alpha/2)}} \right) \cdot \sqrt{k_l^{(-\varkappa_2 - \mu\alpha/2)}}$$

the complexity of the estimate $V_0^{\mathbf{n},\mathbf{k}}$ *is bounded, up to a constant, from above by*

Multilevel approach: complexity

$$
\mathcal{C}_{\mathbf{n}^*,\mathbf{k}^*}(\varepsilon) \lesssim \begin{cases}
\varepsilon^{-2\cdot\max\left(\frac{\varkappa_1+1}{\mu(1+\alpha)},1\right)}, & 2\cdot\varkappa_2 < \mu\alpha \\
\varepsilon^{-2\cdot\frac{\varkappa_1+1}{\mu(1+\alpha)}}, & 2\cdot\varkappa_2 = \mu\alpha \text{ and } \frac{\varkappa_1+1}{\mu(1+\alpha)} > 1 \\
\varepsilon^{-2}\cdot(\log\varepsilon)^2, & 2\cdot\varkappa_2 = \mu\alpha \text{ and } \frac{\varkappa_1+1}{\mu(1+\alpha)} \leq 1 \\
\varepsilon^{-2\cdot\max\left(\frac{\varkappa_1+1}{\mu(1+\alpha)},1+\frac{\varkappa_2-\mu\alpha/2}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa_2 > \mu\alpha
\end{cases}
$$

Observation

The MLMC is superior to the standard MC as long as $\mu > 1/(1+\alpha)$.

Multilevel approach: complexity gain

For $\varkappa_1 = \varkappa_2 = \varkappa$,

$$\begin{cases} \varepsilon^{-2\cdot\min\left(\frac{\varkappa}{\mu(1+\alpha)},1-\frac{1}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa < \mu\alpha \\ \varepsilon^{-2\cdot\left(1-\frac{1}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa = \mu\alpha \text{ and } \frac{\varkappa+1}{\mu(1+\alpha)} > 1 \\ \varepsilon^{-2\cdot\frac{\varkappa}{\mu(1+\alpha)}}, & 2\cdot\varkappa = \mu\alpha \text{ and } \frac{\varkappa+1}{\mu(1+\alpha)} \leq 1 \\ \varepsilon^{-2\cdot\min\left(1-\frac{1}{\mu(1+\alpha)},\frac{\mu\alpha/2}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa > \mu\alpha \end{cases}$$

Conclusion

The largest complexity gain is of order ε^{-1}.

A two period Bermudan option:

- $C_0(z) = \mathrm{E}[g_1(Z_1)|Z_0 = z]$ is monotone increasing in z.
- $g_0(z)$ has a “digital” structure:

$$g_0(z) = \begin{cases} C_0(z_0) + \delta_0, & z < z_0, \\ C_0(z_0) - \delta_0, & z \geq z_0 \end{cases}$$

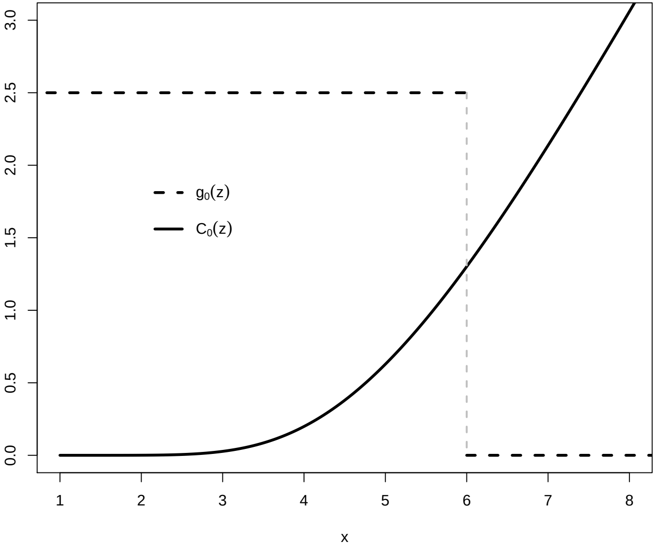
with some $z_0 \in \mathbb{R}$ and $\delta_0 < C_0(z_0)$.

Observation

It holds

$$\mathrm{P}(0 < |C_0(X_0) - g_0(X_0)| \leq \delta_0) = 0$$

and therefore $\alpha = \infty$.

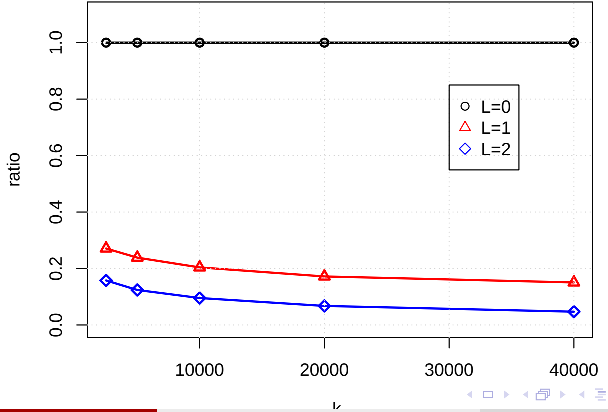

3.0
2.5
2.0
1.5
1.0
0.5
0.0
g0(z)
C0(z)
1
2
3
4
5
6
7
8
x

Numerical example

- $X = (X^1, \ldots, X^d)$ with
$$dX_t^i = (r - \delta)X_t^i dt + \sigma X_t^i dB_t^i,$$
- Payoff:
$$h(X_t) = e^{-rt}(\max(X_t^1, ..., X_t^d) - \kappa)^+.$$
- Benchmark parameters: $d = 2$, $r = 0.05$, $\delta = 0.1$, $t_j = jT/J$, $j = 0, ..., J$, with $T = 3$ and $J = 9$.
- We apply stochastic mesh method.

Numerical example

Variance reduction effect of the ML approach: the ratio of variances $(\sigma_L^*)^2/(\sigma_0^*)^2$ for $L = 0, 1, 2$.



Outlook

- The complexity of the fast approximation methods can be rather large ($> \varepsilon^{-2}$).
- The margin parameter α plays a crucial role in the complexity analysis.
- The complexity of the fast approximation methods can be significantly reduced using the proposed multilevel algorithm (sometimes up to the order ε^{-1}).

Open questions

- Are there other alternative approaches to reduce a complexity ?
- Is a further reduction of complexity possible ?
- What happens if $g_1, \ldots, g_L$ are not Lipschitz ?

Bibliography

Belomestny, D., Schoenmakers, J. (2012).
Multilevel dual approach for pricing American style derivatives, to appear in *Finance and Stochastics*.

Belomestny, D., Dickmann, F. and Nagapetyan, T. (2013).
Pricing American options via multi-level approximation methods. arXiv: 1303.1334.

Belomestny, D. (2011).
Pricing Bermudan options using regression: optimal rates of convergence for lower estimates. *Finance and Stochastics*, **15**(4), 655–683.