Optimizing the exercise boundary for the holder of an American option over a parametric family

ICASQF 2016

Jose Vidal Alcala-Burgos

CONACYT and CIMAT-Merida

June 17 2016

American options

The holder has the right to exercise the option at any time $t \in [0, T]$ and get a payoff $V(X_t, t)$, where the coordinates of $X_t \in \mathbb{R}^d$ are the prices of some underlying assets at time t. The diffusive process X_t solves

$$dX = a(X)dt + \sigma(X)dW \tag{1}$$

with generator $\mathcal{L} = \nabla \cdot a + \frac{1}{2}\mathrm{tr}(\sigma\sigma^T D^2)$ where D^2 is the Hessian. The optimization problem is to find

$$\sup_{\tau \in \mathcal{S}_{0,T}} E[V(X_\tau, \tau)]\,, \tag{2}$$

where $\mathcal{S}_{0,T}$ is the set of all stopping times bounded by the final time T.

Free Boundary

Let $w(x,t)$ be the solution of the optimization problem starting with prices x at the time t. Clearly, $w \geq V$, and if we denote by Ω the points (x,t) for which $w > V$, then w solves the parabolic PDE

$$w_t + \mathcal{L}w = 0\,, \qquad \text{when } (x,t) \in \Omega \tag{3}$$

with boundary conditions

$$\begin{aligned} w(x,t) &= V(x,t)\,, \qquad \text{for } (x,t) \in \partial\Omega \text{ or } t = T\,, \\ \nabla w(x,t) &= \nabla V(x,t)\,, \qquad \text{for } (x,t) \in \partial\Omega\,. \end{aligned} \tag{4}$$

The boundary conditions (4) determine $\partial\Omega$.

Numerical methods

1. **Finite differences**
   - Works up to $d \simeq 4$.
   - Computational work has exponential growth with d.
   - Computational memory has exponential growth with d.
2. **Longstaff and Schwartz.**
   - Works up to $d \simeq 10$.
   - It is a Monte Carlo method, but the confidence intervals have exponential growth with d.

Our objective is to design a Monte Carlo method such that the confidence intervals have polynomial growth with d.

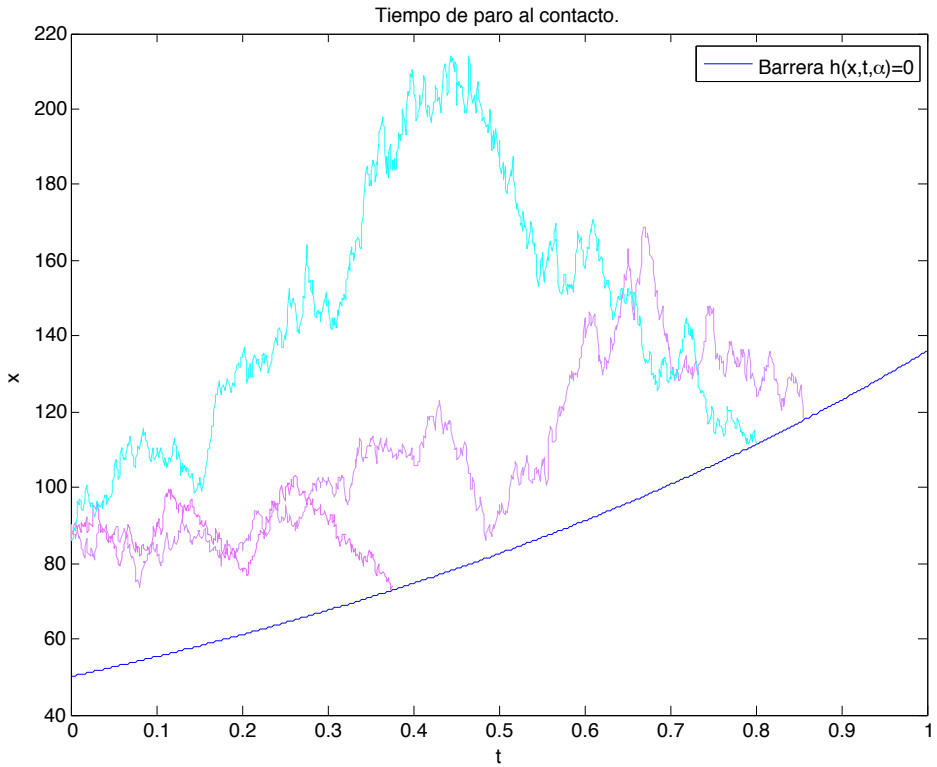
Parametrized free boundary

We optimize only over stopping times given by a level curve.

$$\tau_\theta = \inf \{ t \geq 0 | h(X_t, t, \theta) \leq 0 \} \ . \tag{5}$$

Our objective is to maximize the function

$$f(\theta) \ = \ E\left[V(X_{\tau_\theta}, \tau_\theta) \right] \ .$$



Stochastic optimization

We have random variables V_θ, indexed by the parameter θ. The aim is to maximize

$$f(\theta) = E[V_\theta]\,. \tag{6}$$

Difficulty:

- $f(\theta)$ is evaluated by the Monte Carlo method, which means that evaluations are expensive and have a significant statistical error.

$$\widehat{f}(\theta) = \frac{1}{L}\sum_{k=1}^{L} V_k \approx f(\theta) + \frac{\sigma_\theta}{\sqrt{L}} Z, \qquad Z \sim N(0,1)\,. \tag{7}$$

Sensitivity Analysis

Solution: Find a random variable W_θ that satisfies

$$\nabla f(\theta) \approx E[W_\theta] \,. \tag{8}$$

Desiderata:

- Low *bias*,

$$bias = E[W_\theta] - \nabla f(\theta) \,.$$

- Evaluation of W_θ should be as hard as evaluating V_θ.
- The standard deviations of W_θ and V_θ should be of the same order.

In general, there is no rule for the construction of W_θ. The construction in each application is the sensitivity analysis.

Robbins-Monro optimization

A general iteration takes the form

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n - \frac{1}{n^{\gamma}} K_n \, W_n \, , \tag{9}$$

- *Stochastic Gradient Descent:* $K_n = \lambda I$, $\gamma = 1.0$
- *Optimal:* $K_n \approx H^{-1}$ where H is the Hessian of f evaluated at α_*.
- Big difference when H is an ill conditioned matrix
- Since W_{θ_n} has mean $\nabla f(\theta_n)$, the slope of a least square fit of the points (θ_n, W_{θ_n}) will have a slope approximately equal to the derivative of the map $\theta \to \nabla f(\theta)$; this is, the Hessian of f.

Linear regression approximation of H^{-1}

Let X_n and Y_n be the corresponding $n \times (p+1)$ and $n \times p$ matrices with entries

$$x_k = (\widehat{\theta}_k, 1), \qquad y_k = W_k \, ,$$

consider the linear regression $Y = XB$ and denote the first p rows of a matrix M by $\overline{M}$. We calculate the natural estimators

$$\begin{aligned} B_n &= (X_n^T X_n)^{-1} X_n^T Y, \qquad H_n = \overline{B_n} \\ G_n &= \overline{B_n}^{-1}, \qquad K_n = \frac{G_n + G_n^T}{2} \, , \end{aligned} \tag{10}$$

and use K_n as our H^{-1} estimator and $\gamma = 0.6$.

- The estimator is $\overline{\theta_n} = \frac{1}{n}(\widehat{\theta}_1 + \ldots + \widehat{\theta}_n)$, Polyak averaging.
- Similar methods (with $\gamma = 1$ and no Polyak averaging) were analyzed by Lai and Robbins in 1981, without numerical simulations.
- The resulting optimization is Affine Invariant:

$$B\,\theta_O(f \circ B) = \theta_O(f), \qquad B : \mathcal{S} \longrightarrow \mathcal{S} \quad \text{linear} \, . \tag{11}$$

PDE representation (Feynman-Kac).

- The function $u(x,t,\theta) = E_{x,t}[V(X_{\tau_\theta}, \tau_\theta)]$ satisfies the PDE

$$u_t + \mathcal{L}u = 0\,, \qquad \text{cuando } h(x,t,\theta) > 0 \tag{12}$$

with boundary conditions

$$u(x,t,\theta) = V(x,t)\,, \qquad \text{cuando } h(x,t,\theta) = 0 \text{ o } t = T\,. \tag{13}$$

- The function $\dot{u}(x,t,\theta) = \nabla_\theta u(x,t,\theta)$ satisfies the same PDE but with boundary conditions

$$\dot{u}(x,t,\theta) = \left(\frac{\partial u}{\partial n} - \frac{\partial V}{\partial n}\right) \frac{\dot{h}}{\|\nabla h\|} \text{ cuando } h(x,t,\theta) = 0 \tag{14}$$

$$\dot{u}(x,T,\theta) = 0\,, \tag{15}$$

The vector $n = \frac{\nabla h}{\|\nabla h\|}$ is the inner normal of the hypersurface $h(x,t,\theta) = 0$ in R^d.

Probabilistic representation of $\dot{u}$

The Feynman-Kac formula gives

$$\nabla f(\theta) = \dot{u}(X_0, 0, \theta) = E_{X_0,0}\left[\left(\frac{\partial u}{\partial n} - \frac{\partial V}{\partial n}\right)\frac{\dot{h}}{\|\nabla h\|}\right]. \tag{16}$$

All functions are evaluated at $(X_{\tau_\theta}, \tau_\theta, \theta)$. It is tempting to define

$$W_\theta = \left(\frac{\partial u}{\partial n} - \frac{\partial V}{\partial n}\right)\frac{\dot{h}}{\|\nabla h\|}$$

There is an obvious obstacle: We need to evaluate $\frac{\partial u}{\partial n}$ at the boundary.

Our solution

We introduce a simulation algorithm for diffusion processes (time increment Δt) which is very efficient close to the boundary, and therefore, useful when estimating $\frac{\partial u}{\partial n}$. The (X_τ, τ) simulation will be denoted by $(\widehat{X}_{\widehat{\tau}}, \widehat{\tau})$. Main properties:

- $E[V(\widehat{X}_{\widehat{\tau}}, \widehat{\tau})] = E_{x,t}[V(X_\tau, \tau)] + O(\varepsilon\, \Delta t^\gamma)$
- $\mathrm{Var}\ V(\widehat{X}_{\widehat{\tau}}, \widehat{\tau}) = O(\varepsilon)$
- $E[\widehat{\tau}] = O(\varepsilon)$

where ε is the distance to the boundary $h = 0$.

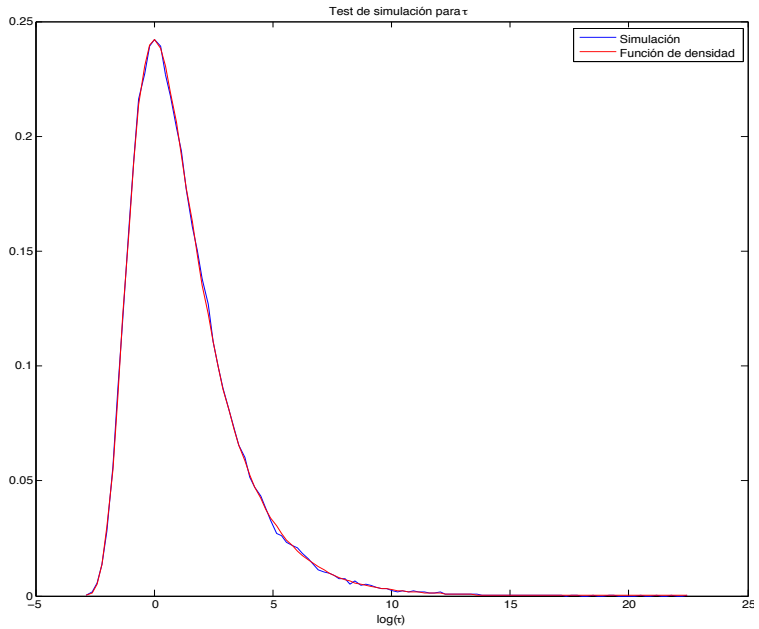
We have $\gamma = 1$ when $d = 1$ and $\gamma = 1/2$ in high dimensions ($d > 1$). The bias, variance and simulation time are $O(\varepsilon)$.

Stopping time simulation.

Assume $\overline{X}_t \in \mathbb{R}$ is a Brownian motion and the linearized boundary is $\overline{h}(x,t) = c - x$, this is, the line $x = c > 0$. The density function for $\overline{\tau}$ is given by

$$P(\overline{\tau} \in t + dt) = \frac{1}{\sqrt{2\,\pi t^3}}\, c\, e^{-c^2/2t}\, dt + o(dt)\,. \tag{17}$$

Changing variables with $z = c/\sqrt{t}$ we obtain $\overline{\tau} = c^2/Z^2$ with $Z \sim N(0,1)$.



Short Paths

Since paths started close to the boundary do not last long, we can generate many paths and take an average:

$$\widehat{u} = \varepsilon \sum_{i=1}^{1/\varepsilon} V(\widehat{X}_{\widehat{\tau}_i}, \widehat{\tau}_i) \,, \tag{18}$$

and we obtain

- $E[\widehat{u}] = E_{x,t}[V(X_\tau, \tau)] + O(\varepsilon\, \Delta t^\gamma)$
- $\operatorname{Var} \widehat{u} = O(\varepsilon^2)$
- $E[Work(\widehat{u})] = \frac{1}{\epsilon} \frac{E[\widehat{\tau}]}{\Delta t} = O(\frac{1}{\Delta t})$

The time it takes to generate $\frac{1}{\varepsilon}$ paths started at distance ε from the boundary is the same time it takes to generate one path far from the boundary. We call this technique the Splash Method.

Normal Derivative

The estimator for the normal derivative at a point x at the boundary $h = 0$ is of course

$$D = \lim_{\varepsilon \to 0} \frac{\widehat{u}(x + \varepsilon n) - V(x, 0)}{\varepsilon} .$$

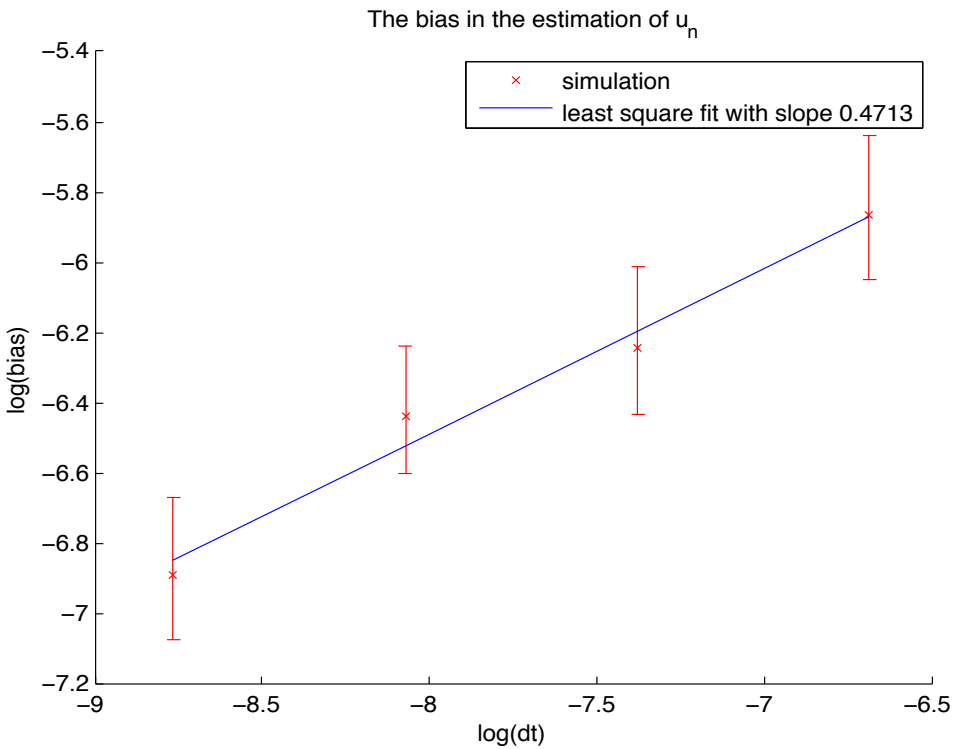
The properties of the estimator $\widehat{u}$ give

- $E[D] = \partial_n u + O(\Delta t^{\gamma})$
- $\operatorname{Var} D = O(1)$

The only task left is to evaluate the limit as $\epsilon \to 0$ in the definition of D.

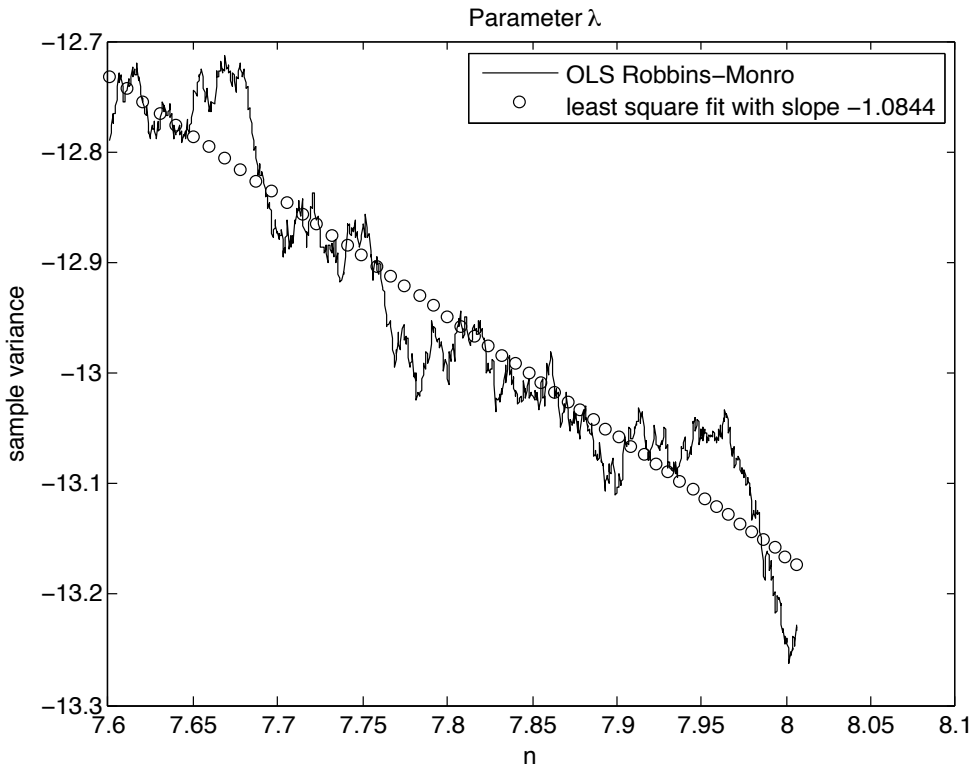
Numerical simulation of D, $d = 3$

- $d\ln(X) = \sigma\, dW$
- $u(x,t) = \exp\left(-\frac{1}{2}(T-t)\xi^T(\sigma\,\sigma^T)\,\xi\right)\cos(\xi^T\,\ln(x))\,.$
- $h(x,t) = (1+\frac{t}{2})^2 - \|x-c\|^2\,.$



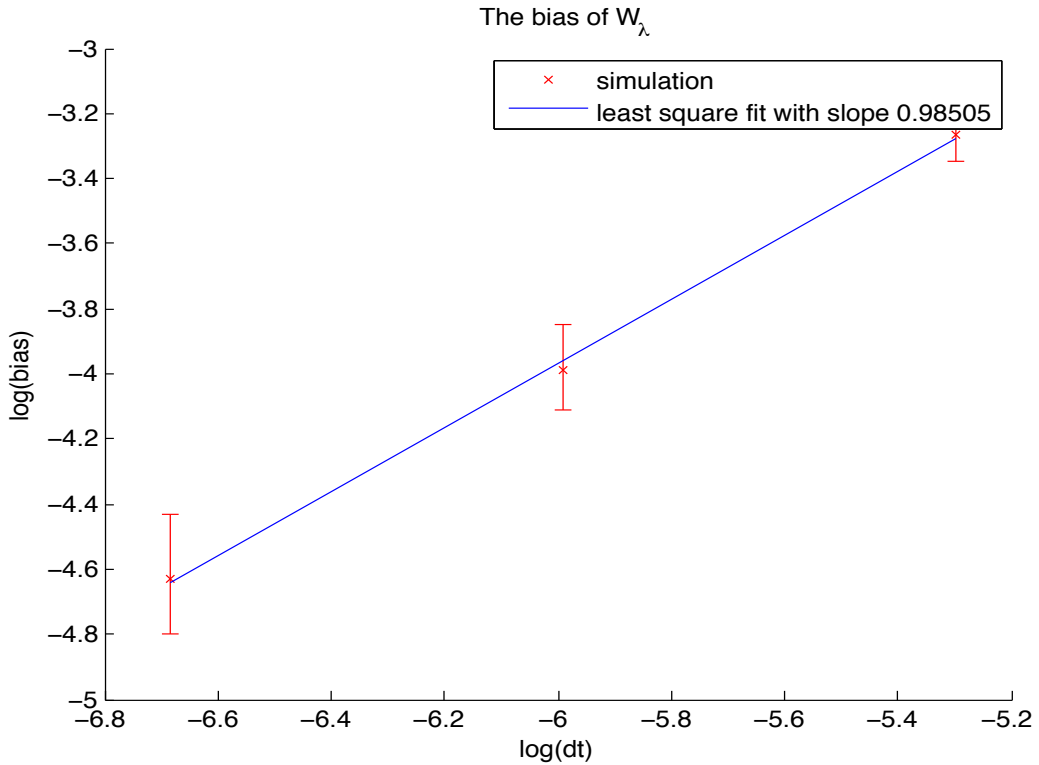
Numerical simulations. $d = 1$ and two parameters.

- $dX = r\,X\,dt + \sigma\,X dW$ Risk neutral process
- $h(x, t, S, \lambda) = x - S\,e^{\lambda\,t}$ Exponential boundary
- $V(x, t) = e^{-r\,t}\,(K - x)_+$ American put option
- We can calculate $u(x, t, S, \lambda)$ analytically.
- Convergence speed is measured by $\mathrm{Var}\,(S_n)$ y $\mathrm{Var}\,(\lambda_n)$.



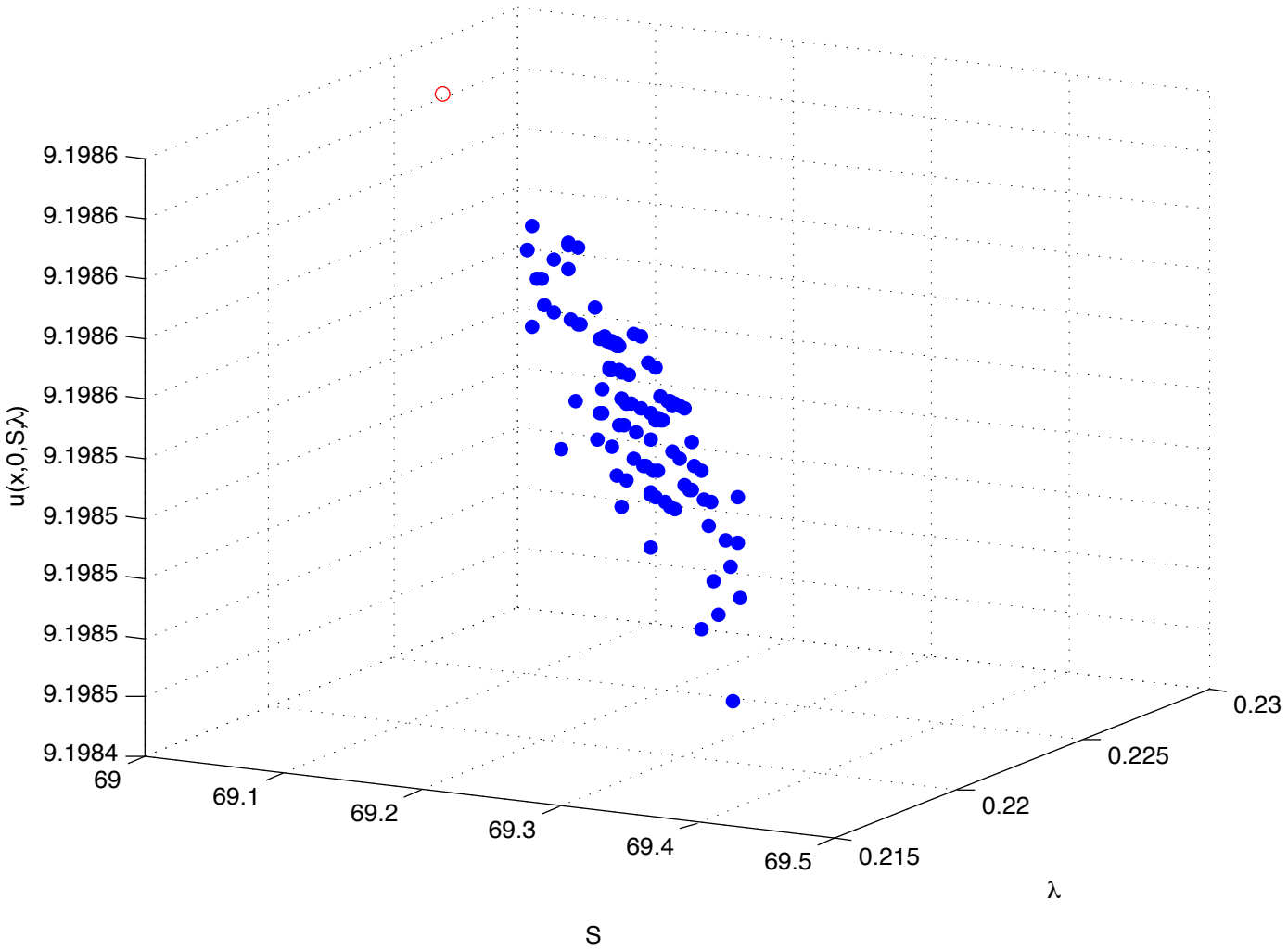
W_θ bias

- bias $= \nabla f(\theta) - E[W_\theta]$
- The bias is $O(\Delta t)$ for $d = 1$ and $O(\sqrt{\Delta t})$ for $d > 1$.



100 simulations of 3000 steps

The price of the put option with the simulation parameters is $9.22 .



References

- José Vidal Alcalá Burgos, Optimizing the exercise boundary for the holder of an American option over a parametric family, Ph.D. Thesis, ProQuest (2012).

  `http://gradworks.umi.com/35/24/3524127.html`
- Code for the *affine invariant optimization* is available at https://github.com/vidalalcala/sopt-ols

Efficient online update

There is an online update algorithm for K_n, the estimator of H^{-1}.

$$
\begin{aligned}
s_{n+1} &= \frac{1}{1 + x_{n+1} P_n x_{n+1}^T} \\
u_{n+1} &= s_{n+1} \overline{P_n x_{n+1}^T} \\
v_{n+1} &= y_{n+1} - x_{n+1} \overline{B_n} \\
t_{n+1} &= \frac{1}{1 + v_{n+1} G_n u_{n+1}} \\
G_{n+1} &= G_n - t_{n+1} G_n u_{n+1} v_{n+1} G_n \\
\overline{B_{n+1}} &= \overline{B_n} + s_{n+1} P_n x_{n+1}^T v_{n+1} \\
P_{n+1} &= P_n - s_{n+1} P_n x_{n+1}^T x_n P_n^T
\end{aligned}
\tag{19}
$$

- $P_n = (X_n^T X_n)^{-1}$ is the precision matrix.
- $O(p^2)$ paralellizable operations.

Remove bias , $\varepsilon \to 0$.

$$E_\varepsilon[V(\widehat{X}_{\widehat{\tau}}, \widehat{\tau})] - V(x,0)] = E_\varepsilon[V(\widehat{X}_{\widehat{\tau}}, \widehat{\tau}) - V(x,0) \,|\, \widehat{\tau} \geq \Delta t] P_\varepsilon(\widehat{\tau} \geq \Delta t) + E_\varepsilon[(V(\widehat{X}_{\widehat{\tau}}, \widehat{\tau}) - V(x,0))\, 1_{\widehat{\tau} < \Delta t}]$$

- $E_\varepsilon[V(\widehat{X}_{\widehat{\tau}}, \widehat{\tau}) - V(x,0) \,|\, \widehat{\tau} \geq \Delta t] = O(1)\,, \quad S(y) = \frac{y}{\Delta t} e^{-y^2/2\Delta t}$
- $P_\varepsilon(\widehat{\tau} \geq \Delta t) = O(\varepsilon)$
- $E_\varepsilon[V(\widehat{X}_{\widehat{\tau}}, \widehat{\tau}) - V(x,0)\, 1_{\widehat{\tau} < \Delta t}] = O(\varepsilon)$

Table: Stochastic optimization parameters

Δt	0.01	Time step .
T	1.0	Final time.
σ	0.3	SDE coefficient
r	0.07	SDE coefficient
X_0	100	Initial position for the SDE
K	100	Strike price in the payoff $V(x,t) = e^{-r\,t}\,(K-x)_+$
S_{min}	60	Lower bound for the parameter S
S_{max}	90	Upper bound for the parameter S
λ_{min}	-0.01	Lower bound for the parameter λ
λ_{max}	0.7	Upper bound for the parameter λ
L	50	The number of paths started at position X_0 at time 0
S_{path}	50	The number of samples of $\overline{D}$ per path hitting the boundary

Table: Initial condition for the parameters (S, λ)

S_0	80.7642	Initial condition for the parameters.
S_1	70.8288	Initial condition for the parameters.
S_2	79.0376	Initial condition for the parameters.
S_3	76.8769	Initial condition for the parameters.
λ_0	0.556173	Initial condition for the parameters.
λ_1	0.279206	Initial condition for the parameters.
λ_2	0.0235858	Initial condition for the parameters.
λ_3	0.286322	Initial condition for the parameters.

Multidimensional simulation, $d > 1$.

Assume $x^p = 0$ y $\overline{a} = 0$. Linear change of coordinates:

- $\overline{X}_s = AZ_s$
- $Z_s - Z_0$ is a Brownian motion in $\mathbb{R}^d$
- $\overline{h}(\overline{X}_s, s) = 0$ if and only if $e_1^T Z_s = \beta s$.

The stopping time only depends on the new first coordinate and the stopping time is independent of the remaining coordinates. Requirements:

- $AA^T = \overline{\sigma}\,\overline{\sigma}^T$, orthogonal columns with the inner product $(\overline{\sigma}\,\overline{\sigma}^T)^{-1}$
- $Ae_1 = \frac{1}{\sqrt{\nabla h^T \overline{\sigma}\,\overline{\sigma}^T \nabla h}} \overline{\sigma}\,\overline{\sigma}^T \nabla h$, Conormal vector.
- $\beta = -\frac{h_t}{\sqrt{\nabla h^T \overline{\sigma}\,\overline{\sigma}^T \nabla h}}$.

The stopping time for $e_1^T Z_s$ is denoted by $\overline{\tau}$. Define

- $e_i^T Z_{\overline{\tau}} = e_i^T Z_0 + \sqrt{\overline{\tau}} N_i, \;\; i \neq 1$
- $(\overline{X}_{\overline{\tau}}, \overline{\tau}) = (AZ_{\overline{\tau}}, \overline{\tau})$

with N_i standard normal variables.