Revisit of stochastic mesh method for pricing American options

Guangwu Liu, L. Jeff Hong

Abstract

From an importance sampling viewpoint, Broadie and Glasserman [M. Broadie, P. Glasserman, A stochastic mesh method for pricing high-dimensional American options, Journal of Computational Finance 7 (4) (2004) 35–72] proposed a stochastic mesh method to price American options. In this paper, we revisit the method from a conditioning viewpoint, and derive some new weights.

1. Introduction

In this paper, we treat American options as derivative securities which can be exercised at a finite number of dates prior to the maturity. They are sometimes called Bermudan options. To price an American option using Monte Carlo simulation, one may formulate it as a dynamic programming problem, and then estimate the holding value of the option backwards at each exercise date recursively. To approximate the holding value, Longstaff and Schwartz [1] and Tsitsiklis and Van Roy [2] use a regression approach by employing a sequence of basis functions, and Broadie and Glasserman [3] design a stochastic mesh method. In this paper, we focus on the stochastic mesh method. Basically the method estimates the holding value at each exercise date by a weighted average of the option values at the next exercise date on all sample paths. Along this line of research, Avramidis and Hyden [4] consider the efficiency improvement of the method, and Avramidis and Matzinger [5] show the convergence of the stochastic mesh estimators.

A key feature of the stochastic mesh method is how to derive the weight functions. Broadie and Glasserman [3] take an importance sampling viewpoint and derive weights of each exercise date based on the information of the next exercise date. Therefore, we call them forward-looking weights. In this paper we revisit this problem, and consider it from a conditioning viewpoint. From this viewpoint, we may derive the same forward-looking weights of [3]. Furthermore, we may also derive new weights that use not only the information of the next exercise date but also the information of the last exercise date. Therefore, we call them binocular weights. To illustrate how the idea works, we compare these two weights for the Black-Scholes model. The numerical results show that the forward-looking weights have smaller variances, but the binocular weights have smaller biases. We also demonstrate how to apply the forward-looking and binocular weights to the variance-gamma model. We show that both weights cannot be implemented efficiently. However, the conditioning viewpoint allows us to exploit more information in the simulation and develop other weights that can be implemented efficiently. A simple numerical study shows that the new weights work well for the variance-gamma model.

The rest of the paper is organized as follows. In Section 2 we review some preliminary knowledge on pricing American options and the stochastic mesh method. Then in Section 3 we analyze the problem from a conditioning viewpoint, and derive the forward-looking and binocular weights. In Section 4 we consider two examples to explain how to apply the new weights, and illustrate their performances through simple numerical examples.

2. Preliminaries

Let $S_t$ denote the price at time $t$ of the underlying asset whose price dynamics follows a Markov process on $\mathbb{R}^2$. Suppose that $0 = t_0 < t_1 < \ldots < t_m = T$ are the exercise dates. Without loss of generality, we assume that $t_i+1 - t_i = \Delta t$ for all $i = 0, 1, \ldots, m-1$. We write $S_t$ for $S_{t_i}$ for simplicity of notation. Moreover, suppose that $n$ independent sample paths of $(S_0, S_1, \ldots, S_m)$ are generated through Monte Carlo simulation. We denoted the $j$th sample path by $(S_0^j, S_1^j, \ldots, S_m^j)$.

Let $L(x)$ denote the discounted payoff of the option at $t_i$ if it is exercised, and $H_i(x)$ denote the holding value of the option at $t_i$, when $S_i = x$. Let $V(x)$ denote the value of the option at $t_i$ when $S_i = x$. Then a backwards recursion algorithm for
pricing the American option can be expressed as $V_m = L_m(x)$ and $V_i(x) = \max\{L_i(x), H_i(x)\}, i = 0, 1, \ldots, m - 1$. The holding value $H_i(x)$ satisfies $H_i(x) = e^{-r\Delta t} E \{ V_{i+1}(S_{i+1}) | S_i = x \}$, where the expectation is taken with respect to the risk-neutral measure and $r$ is the risk-free interest rate. Then the price of the American option at time 0 is $V_0(S_0)$.

The major difficulty of pricing the American option is how to estimate the holding value $H_i(x)$ for any state $x$. Broadie and Glasserman [3] propose a stochastic mesh method to do it. The key feature of the method is that for any $x$, it evaluates $H_i(x)$ by exploiting all the nodes at time $t_i+1$, i.e., $S_i^1, \ldots, S_i^{L_i}$. Essentially they choose an appropriate weight function $w(i, x, S_i^i)$ such that $H_i(x)$ can be estimated by

$$H_i(x) = e^{-r\Delta t} \frac{1}{n} \sum_{j=1}^{n} V_{i+1}(S_{i+1}^j) \cdot w(i, x, S_i^i),$$

where $V_{i+1}(x) = \max\{L_{i+1}(x), H_{i+1}(x)\}$. The key issue of the stochastic mesh method is how to choose an appropriate weight function $w(i, x, S_i^i, S_{i+1}^i)$. Broadie and Glasserman [3] analyze this problem from an importance sampling viewpoint. A weight function they suggest is

$$w(i, x, S_i^i) = \frac{f_i(x, S_i^i)}{\frac{1}{n} \sum_{j=1}^{n} f_i(S_i^j, S_{i+1}^j)}, \quad (1)$$

where $f_i(x, y)$ is the transition density from $S_i = x$ to $S_{i+1} = y$.

3. Estimating the holding value $H_i(x)$

For convenience of exposition, throughout the paper we work with one-dimensional case, i.e., $d = 1$, though the analysis can be easily extended to high-dimensional case. We also assume for convenience that densities involved in the analysis are continuous. In this section, we analyze the conditional expectation form of $H_i(x)$, and apply the conditioning techniques to obtain a different insight of the stochastic mesh method. Note that

$$H_i(x) = e^{-r\Delta t} \int_0^\infty V_{i+1}(y) f_i(x, y) \, dy = \lim_{\epsilon \to 0^+} \frac{E \{ e^{-r\Delta t} V_{i+1}(S_{i+1}) : 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}} \}}{E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}} \}}, \quad (2)$$

where $f(-, -)$ denotes the joint density of $(S_i, S_{i+1})$, the first equality serves as a definition of the conditional expectation (see, e.g., Example 1.4 in Page 221 of [6]), and the second equality results from simply applying the mean value theorem if $\int_0^\infty V_{i+1}(y) f_i(x, y) \, dy$ is continuous at $x$.

Essentially Eq. (2) expresses the conditional expectation $H_i(x)$ as a limit of expectations. However, it is of little practical value because of the limit operator. Intuitively, with more information, e.g., the transition densities, one may remove the limit operator and express $H_i(x)$ in terms of expectations. In the following two subsections, we apply the conditioning techniques to incorporate more information so that we can express $H_i(x)$ in terms of expectations.

3.1. Forward-looking weights

Based on Eq. (2) and by conditioning on $S_{i+1}$, we have

$$H_i(x) = \lim_{\epsilon \to 0} \frac{E \{ e^{-r\Delta t} V_{i+1}(S_{i+1}) : 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}} \}}{E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}} \}} = \lim_{\epsilon \to 0} \frac{E \{ e^{-r\Delta t} V_{i+1}(S_{i+1}) : 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i+1} \}}{E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i+1} \}}.$$

With some regularity conditions, we can take the limit into the expectations. Then we have $H_i(x) = E \{ e^{-r\Delta t} V_{i+1}(S_{i+1}) w(i, x, S_{i+1}) \}$, where

$$w(i, x, S_{i+1}) = \lim_{\epsilon \to 0} \frac{E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i+1} \}}{E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}} \}} \lim_{\epsilon \to 0} \frac{\frac{1}{2\epsilon} E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i+1} \}}{\frac{1}{2\epsilon} E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}} \}}.$$

Denote by $f_i(\cdot)$ the marginal density of $S_i$. Note that, by the mean value theorem, $\lim_{\epsilon \to 0} \frac{1}{2\epsilon} E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}} \} = f_i(x)$. Furthermore, by Bayes’ rule and the mean value theorem,

$$\lim_{\epsilon \to 0} \frac{1}{2\epsilon} E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i+1} \} = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} f_i(u) f_i(u, S_{i+1}) \, du = \int_{x-\epsilon}^{x+\epsilon} f_i(u)(x, S_{i+1}) \, du,$$

Then

$$w(i, x, S_{i+1}) = f_i(x, S_{i+1}) f_i(S_{i+1}).$$

The above weight $w(i, x, S_{i+1})$ involves two densities, $f_i(x, \cdot)$ and $f_i(\cdot, S_{i+1})$. In practice $f_i(x, \cdot)$ is often known, since it is the actual transition density that is used to generate the sample paths of the underlying asset price. However, the explicit expression of $f_i(\cdot, S_{i+1})$ is typically unknown. By conditioning on $S_{i+1}$, we have $f_i(\cdot, S_{i+1}) = E \{ f_i(\cdot, \cdot) \}$. Then $f_i(1)(S_{i+1})$ can be estimated by $\frac{1}{2}\sum_{j=1}^{L_{i+1}} f_i(S_{i+1}^j)$. Therefore, we obtain exactly the weight function defined in Eq. (1). We refer to this weight function as a forward-looking weight function, since it can condition on $S_{i+1}$, the price of the underlying at the next exercise date.

3.2. Binocular weights

Motivated by the bridge sampling techniques used in Monte Carlo simulation (see, for instance, [7]), we condition on $S_{i-1}$ and $S_{i+1}$ to derive new weight functions for estimating the holding value $H_i(x)$.

Let $f_{j_{i-1}, j_{i+1}}(-)$ denote the conditional density of $S_i$ given $S_{i-1} = v_1$ and $S_{i+1} = v_2$. Based on Eq. (2) and by conditioning on $S_{i-1}$ and $S_{i+1}$, we have

$$H_i(x) = \lim_{\epsilon \to 0} \frac{E \{ e^{-r\Delta t} V_{i+1}(S_{i+1}) E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i-1} \} \}}{E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}} \}} = E \{ e^{-r\Delta t} V_{i+1}(S_{i+1}) \cdot w(i, x, S_{i-1}, S_{i+1}) \},$$

where, under some regularity conditions, interchanging the limit and expectation yields

$$w(i, x, S_{i-1}, S_{i+1}) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i-1} \} = \int_{0}^{\epsilon} f_{j_{i-1}, j_{i+1}}(x, S_{i-1}, S_{i+1}) \, dx.$$

Specifically, suppose that for any $u_1, u_2 > 0$, $f_{j_{i-1}, j_{i+1}}(-)$ is continuous and bounded in a neighborhood of $x$ with $g(u_1, u_2)$ being an upper bound, and there exists a random variable $K$ with finite mean such that $V_{i+1}(S_{i+1}) \cdot g(S_{i-1}, S_{i+1}) \leq K$ with probability 1 (w.p.1). Then $V_{i+1}(S_{i+1}) \geq E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i-1}, S_{i+1} \} \geq V_{i+1}(S_{i+1}) \cdot g(S_{i-1}, S_{i+1}) \leq K$ w.p.1 for small $\epsilon$, and hence by the dominated convergence theorem,

$$\lim_{\epsilon \to 0} \frac{1}{2\epsilon} E \{ e^{-r\Delta t} V_{i+1}(S_{i+1}) E \{ 1_{\{x-\epsilon \leq S_i \leq x+\epsilon\}}, S_{i-1} \} \} = E \{ e^{-r\Delta t} V_{i+1}(S_{i+1}) \cdot f_{j_{i-1}, j_{i+1}}(x, S_{i-1}, S_{i+1}) \}.$$
Numerical results of forward-looking and binocular weights for the Black–Scholes model.

<table>
<thead>
<tr>
<th>n</th>
<th>Call on one asset</th>
<th>Max-option on five assets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>Forward</td>
<td>Bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stdev</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rmse(%)</td>
</tr>
<tr>
<td>Binocular</td>
<td>Bias</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>Stdev</td>
<td>0.525</td>
</tr>
<tr>
<td></td>
<td>Rmse(%)</td>
<td>6.8</td>
</tr>
</tbody>
</table>

since \( \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ \left( 1_{x \leq t < x+1} \right) \right] \rho_{n} \rho_{n+1} = f_{i-1,i+1}(x, \rho_{n}, \rho_{n+1}) \) w.p.1. Then Eq. (3) holds. These regularity conditions hold quite generally, e.g., for the Black–Scholes model.

For many models, the expression of \( f_{i-1,i+1}(x, \rho_{1}, \rho_{n}) \) is known or can be approximated based on the bridge sampling techniques. For instance, if \( S_{t} \) follows a geometric Brownian motion, then \( f_{i-1,i+1}(x, \rho_{1}, \rho_{n}) \) can be computed explicitly. Since the marginal density \( f_{1}(x) \) is typically unknown, we may use \( f_{i-1,i+1}(x, \rho_{1}, \rho_{n}) \) to estimate it. Note that \( f_{i}(x) = \mathbb{E}[f_{i-1,i+1}(x, \rho_{1}, \rho_{n})] \), then \( f_{i}(x) \) can be unbiasedly estimated by \( \frac{1}{n} \sum_{k=i}^{n} f_{i-1,i+1}(x, \rho_{1}, \rho_{n}) \).

Then we obtain a new weight function

\[
\tilde{w}(i, x, \rho_{1}, \rho_{n}) = \frac{f_{i-1,i+1}(x, \rho_{1}, \rho_{n})}{\frac{1}{n} \sum_{k=i}^{n} f_{i-1,i+1}(x, \rho_{1}, \rho_{n})}.
\]

Since the new weight function use the information on both sides of the current exercise date, we refer to it as a binocular weight.

4. Examples

4.1. Black–Scholes model

Suppose that the price of the underlying asset follows a geometric Brownian motion (GBM) under the risk-neutral measure, i.e.,

\[
dS_{t} = (r - \delta) S_{t} dt + \sigma S_{t} dB_{t},
\]

where \( r \) is the risk-free interest rate, \( \delta \) the dividend rate, \( \sigma \) the volatility, and \( B_{t} \) a standard Brownian motion.

We first derive the forward-looking weight. Note that \( S_{t} = S_{0} \exp[(r - \delta - \sigma^{2}/2)t + \sigma B_{t}] \), where \( \sigma \) is typically unknown, we may use \( S_{t} \) to estimate it. Then Eq. (4) holds.

To illustrate the performance of the binocular weight and compare it to the forward-looking weight of [3], we consider two options under the Black–Scholes model, a call underlying one asset and a max-option underlying five assets. These two examples are cited from Glasserman [8].

Table 1

The numerical results are summarized in Table 1, where we show the bias, standard deviation (stdev), and relative root mean square error (rmse, defined as the percentage of root mean square error relative to true price) of the estimators. From the table we can see that the binocular weight has smaller bias while the forward-looking weight has smaller variance. Intuitively, large variance of the binocular weights results from that weights are more dispersed when conditioning on both \( S_{t} \) and \( S_{t+1} \), and hence have large variances. Numerical results also suggest that binocular weight leads to smaller bias, which may be helpful in finding tighter bounds of the price. The bias issue deserves further investigation in the future.

4.2. Variance-Gamma model

We consider the variance-gamma (VG) model. Pricing American options under VG model has been studied by Hirsa and Madan [9]. Our aim is not to propose a competing method, but to explore more insights of the stochastic mesh method. We show that both the forward-looking and binocular weights are practically infeasible. However, the proposed conditioning techniques can still be applied to derive other weights that can be implemented efficiently.

Following the notation of Avramidis and L’Ecuyer [7], we let \( B(t) = B(t; \theta, \sigma) \) be a Brownian motion with drift parameter \( \theta \) and variance parameter \( \sigma \), and \( G(t) = G(t; \mu, \nu) \) be a gamma process independent of \( B(t) \), with drift \( \mu \geq 0 \) and volatility \( \nu > 0 \). A VG process with parameters \( (\theta, \sigma, \nu) \) is defined by

\[
X(t) = B(t; \mu, \nu) + \sum_{k \geq 1} \left( \frac{\nu}{\sigma^2} \right) \Gamma\left( \frac{1}{2}, \frac{\nu}{2\sigma^2} \right),
\]

where \( \mu \) is the risk-free interest rate, \( \delta \) is the dividend rate, and \( \omega = \log(1 - \theta \nu - \sigma^2\nu/2)/\nu \). To ensure that \( E(S_{t}) < \infty \) for all \( t > 0 \), we require that \( \theta + \sigma^2/2 > 1 \).

To analyze this model, we first review two schemes of simulating the VG process. The first one is simulating it as a gamma time-changed Brownian motion, while the second one simulating it via a Brownian bridge. For details of these schemes, one is referred to Fu [10] and Avramidis and L’Ecuyer [7]. We will use the two schemes to develop the forward-looking and binocular weights respectively.

We first look at the scheme of simulating VG process as a gamma time-changed Brownian motion. We let \( X_{t} \) and \( G_{t} \) denote \( X(t) \) and \( G(t) \), respectively, and independently generate \( \Delta G_{i} := G_{i+1} - G_{i} \) from a gamma distribution \( \Gamma(\Delta t/v, \nu) \) at \( Z_{i} \) from a standard normal distribution. Then we have \( X_{i+1} = X_{i} + \theta \Delta G_{i} + \sigma \sqrt{G_{i} Z_{i}} \).

Therefore the binocular weight of Eq. (4) can be applied.

To illustrate the performance of the binocular weight and compare it to the forward-looking weight of [3], we consider two options under the Black–Scholes model, a call underlying one asset and a max-option underlying five assets. These two examples are cited from Glasserman [8].

\[
\tilde{f}(x, u) = \int_{0}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp\left( \frac{(\log(u/x) - (\omega + r - \delta) t - \theta y)^2}{2\sigma^2} \right) dy.
\]

where \( \gamma \) is the \( \Gamma(\Delta t/v, \nu) \) density. Plugging \( \tilde{f}(x, u) \) in Eq. (1) we obtain the forward-looking weights.
The VG process can also be simulated via a Brownian bridge (see, e.g., Avramidis and L’Ecuyer [7]). In particular, given $X_{i-1}$, $X_{i+1}$, $G_{i-1}$ and $G_{i+1}$, $X_i$ can be simulated by a two-step algorithm. Particularly, $X_i$ can be simulated by
\[ X_i = YX_{i+1} + (1 - Y)X_{i-1} + \sqrt{Y(1 - Y)}(G_{i+1} - G_{i-1}) \cdot Z, \] (7)
where $Y$ is generated from a beta distribution $\beta(\Delta t/\nu, \Delta t/\nu)$, and $Z$ is a standard normal random variable.

Note that, by Eq. (5), conditioning on $S_{i-1}$ and $S_{i+1}$ is equivalent to conditioning on $X_{i-1}$ and $X_{i+1}$. However, the density of $X_i$ (thus the density of $S_i$) is difficult to derive given $X_{i-1}$ and $X_{i+1}$. Therefore, we suggest to condition on $(X_i, G_{i-1})$ and $(X_{i+1}, G_{i+1})$, which include all the information in the steps $i-1$ and $i+1$. Then by simple algebra, the conditional density of $S_i$ given $(X_{i-1}, G_{i-1}) = (v_1, u_1)$ and $(X_{i+1}, G_{i+1}) = (v_2, u_2)$ can be written as
\[ f_{i|v_1,v_2,u_1,u_2}(x, u_1, u_2, v_2) = \int_0^1 \phi \left( \frac{\log(S_1/v_1) - \log(v_2/u_2) - (1 - \nu)u_1 + \nu x}{\sqrt{\nu} \sqrt{1 - \nu}(u_2 - u_1)} \right) g(y) dy. \] (8)
where $g(\cdot)$ is the $\beta(\Delta t/\nu, \Delta t/\nu)$ density. Plugging $f_{i|v_1,v_2,u_1,u_2}$ in Eq. (4) we obtain the binocular weights.

Although both the forward-looking and binocular weights are applicable to this example, they are practically infeasible. As shown in Eqs. (6) and (8), both $f_i(x, u)$ and $f_{i|v_1,v_2,u_1,u_2}$ need numerically integrating, which requires a prohibitive amount of effort, and makes it practically infeasible.

Fortunately, we may use the conditioning technique to obtain other weights that are easier to implement. This approach provides us some flexibility in choosing the conditioning quantities. By conditioning on some appropriate quantities we may obtain weights that are practically applicable. We illustrate how to do so.

Rather than conditioning on $X_{i-1}, X_{i+1}, G_{i-1}$ and $G_{i+1}$, we additionally condition on $G_i$. Then similar to the previous analysis, a new weight function can be expressed as:
\[ w(i, x, X_{i-1}, X_{i+1}, G_{i-1}, G_{i+1}, G_i) = \frac{f_i^G(x, X_{i-1}, X_{i+1}, G_{i-1}, G_{i+1}, G_i)}{\sum_{i=1}^N f_i^G(x, X_{i-1}, X_{i+1}, G_{i-1}, G_{i+1}, G_i)}, \] (9)
where $f_i^G(x, v_1, v_2, u_1, u_2, u)$ is the conditional density of $S_i$ given $X_{i-1} = v_1, X_{i+1} = v_2, G_{i-1} = u_1, G_{i+1} = u_2$ and $G_i = u$. By simple algebra, we can easily find that
\[ f_i^G(x, v_1, v_2, u_1, u_2, u) = \frac{1}{\sqrt{2\pi}} \phi \left( \frac{x - \log(S_0/v_1)}{\sqrt{\nu}} \right) \frac{1}{\sqrt{2\pi}} \phi \left( \frac{u - r - \delta i - p}{\sqrt{\nu}} \right), \]
where $p = [(u - u_1)v_2 + (u_2 - u)v_1]/(u_2 - u_1)$ and $q = \sqrt{(u - u_1)(u_2 - u)}/(u_2 - u_1)$.

Then by Eq. (9), we may compute the weight and use it to price American options. To illustrate the performance of the new weight function, we consider a call option under the VG model. Similar to the settings in [9], we let $T = 0.5616$, $r = 5.41\%$, $\delta = 1.2\%$, $\sigma = 20.72\%$, $\nu = 0.5022$, $\theta = -0.2290$, $S_0 = 1369.4$, $K = 1200$ and $m = 10$. The numerical results are summarized in Table 2. From the table we see that the estimator works well. However, bias of the estimator is relatively high, which results from the backward induction procedure, as well as the introduction of a sample mean as the denominator of the weights. The analysis of the bias is an issue that is worthy of further investigation in the future.

Acknowledgements

The authors thank the associate editor and referees for their valuable comments. This research was partially supported by Hong Kong Research Grants Council grant CERG 613907. An earlier version of this paper has been accepted by Proceedings of the 2008 Winter Simulation Conference.

References