Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

REVISIT OF STOCHASTIC MESH METHOD FOR PRICING AMERICAN OPTIONS

Guangwu Liu
L. Jeff Hong

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, China

ABSTRACT

We revisit the stochastic mesh method for pricing American options, from a conditioning viewpoint, rather than the importance sampling viewpoint of Broadie and Glasserman (1997). Starting from this new viewpoint, we derive the weights proposed by Broadie and Glasserman (1997) and show that their weights at each exercise date use only the information of the next exercise date (therefore, we call them forward-looking weights). We also derive new weights that exploit not only the information of the next exercise date but also the information of the last exercise date (therefore, we call them binocular weights). We show how to apply the binocular weights to the Black-Scholes model, more general diffusion models, and the variance-gamma model. We demonstrate the performance of the binocular weights and compare to the performance of the forward-looking weights through numerical experiments.

1 INTRODUCTION

The pricing of American options is one of the challenging problems in financial engineering. By the term American options, we refer to derivative securities which can be early-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 approximate the value of the American option backwards recursively.

To approximate the value of the option at each exercise date, Tsitsiklis and Van Roy (1999) and Longstaff and Schwartz (2001) use a regression approach by employing a sequence of basis functions, and Broadie and Glasserman (1997) design a stochastic mesh method. In this paper, we focus on the stochastic mesh method. Basically the stochastic mesh method approximates the option value by using weight functions which explore the information contained in the simulation, e.g., the density information. Along this line of research, Avramidis and Hyden (1999) consider the the efficiency improvement of the method, and Avramidis and Matzinger (2004) show the convergence of the stochastic mesh estimators. Other subsequent work includes Broadie, Glasserman, and Ha (2000) and Broadie, Glasserman, and Jain (1997).

A key feature of the stochastic mesh method is how to derive the weight functions. Broadie and Glasserman (1997) 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 can derive the same weights of Broadie and Glasserman (1997). Furthermore, we can 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 to apply the binocular weights, we study how to apply them to the Black-Scholes model and more general diffusion models. We compare these two weights for the Black-Scholes model through some simple and preliminary numerical experiments. 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 binocular weights to the variance-gamma model. Note that the forward-looking weights are difficult to implement for this model since they require a large amount of computational effort. A simple numerical study shows that the binocular 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 several examples to illustrate how to apply the forward-looking and binocular weights, followed by numerical study in Section 5. We conclude the paper in Section 6.

2 PRELIMINARIES

Let S_t denote the price at time t of the underlying asset whose price dynamics follows a Markov process on $\mathbb{R}^d$. Suppose that $0 = t_0 < t_1 < \ldots < t_m = T$ are exercise opportunities (also called exercise dates), i.e., the American option can be exercised at t_i for any $i \in \{0, 1, \ldots, m\}$. Without loss of generality, we assume that $t_{i+1} - t_i = \tau$ for all $i = 0, 1, \ldots, m-1$. We write S_i 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, denoted by $\{S_0, S_1^j, \ldots, S_m^j\}$ the j-th sample path.

Let $L_i(x)$ denote the payoff function of the American option from exercise at date t_i when $S_i = x$, and $V_i(x)$ denote the value of the option at date t_i when $S_i = x$. Then a backwards recursion algorithm for pricing the American option can be expressed as

$$\begin{aligned} V_m(x) &= L_m(x) \\ V_i(x) &= \max\left(L_i(x), e^{-r\tau}\mathrm{E}\left[V_{i+1}(S_{i+1}) | S_i = x\right]\right), \\ &\quad i = 0, 1, \ldots, m-1, \end{aligned}$$

where the expectation is taken under the risk-neutral measure, r is the risk-free interest rate. For simplicity we only consider the interest rate as a constant, while it can be extended to more complicated models of interest rates. Then the price of the American option at time 0 is $V_0(S_0)$.

For $i = 0, 1, \ldots, m$, let $H_i(x)$ be the holding value of the option at date t_i when $S_i = x$, i.e.,

$$H_i(x) = e^{-r\tau}\mathrm{E}\left[V_{i+1}(S_{i+1}) | S_i = x\right].$$

Then the major difficulty of pricing the American option reduces to how to estimate the holding value $H_i(x)$ for any state x.

Broadie and Glasserman (1997) propose a stochastic mesh method to price the American option. 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., $S_{i+1}^1, \ldots, S_{i+1}^n$. Essentially they choose an appropriate weight function $w(i, x, S_i, S_{i+1})$ such that $H_i(x)$ can be estimated by

$$\bar{H}_i(x) = e^{-r\tau}\frac{1}{n}\sum_{j=1}^n \bar{V}_{i+1}(S_{i+1}^j) \cdot w(i, x, S_i^j, S_{i+1}^j),$$

where $\bar{V}_{i+1}(x) = \max\{L_{i+1}(x), \bar{H}_{i+1}(x)\}$. The key issue of the stochastic mesh method is how to choose an appropriate weight function $w(i, x, S_i, S_{i+1})$. Broadie and Glasserman (1997) analyze this problem from an importance sampling viewpoint. One of the weight functions they suggest is

$$\overline{w}_2(i, x, S_{i+1}) = \frac{f_i(x, S_{i+1})}{\frac{1}{n}\sum_{j=1}^n f_i(S_i^j, S_{i+1})},$$

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

3 ESTIMATING THE HOLDING VALUE $H_i(x)$

Note that

$$\begin{aligned} H_i(x) &= e^{-r\tau}\mathrm{E}\left[V_{i+1}(S_{i+1}) | S_i = x\right] \\ &= \lim_{\varepsilon \to 0^+} \frac{\mathrm{E}\left[e^{-r\tau}V_{i+1}(S_{i+1}) \cdot 1_{\{x-\varepsilon \le S_i \le x+\varepsilon\}}\right]}{\mathrm{E}\left[1_{\{x-\varepsilon \le S_i \le x+\varepsilon\}}\right]} \qquad (1) \\ &\approx \mathrm{E}\left[e^{-r\tau}V_{i+1}(S_{i+1}) \cdot \frac{1_{\{x-\varepsilon \le S_i \le x+\varepsilon\}}}{\mathrm{E}\left[1_{\{x-\varepsilon \le S_i \le x+\varepsilon\}}\right]}\right], \end{aligned}$$

when ε is small. Based on this expression, an estimator of $H_i(x)$ can be

$$\bar{H}_i^\varepsilon(x) = e^{-r\tau}\frac{1}{n}\sum_{j=1}^n \bar{V}_{i+1}^\varepsilon\left(S_{i+1}^j\right) \cdot w^\varepsilon(x, S_i^j),$$

where

$$w^\varepsilon(x, S_i^j) = \frac{1_{\{x-\varepsilon_n \le S_i \le x+\varepsilon_n\}}}{\frac{1}{n}\sum_{k=1}^n 1_{\{x-\varepsilon_n \le S_k \le x+\varepsilon_n\}}},$$

or more generally,

$$w^\varepsilon(x, S_i^j) = K\left(\frac{S_i^j - x}{\varepsilon_n}\right) \Bigg/ \left[\frac{1}{n}\sum_{k=1}^n K\left(\frac{S_i^k - x}{\varepsilon_n}\right)\right]$$

by the kernel method (Bosq 1998) where K is a kernel density function, e.g., the standard normal density function. To ensure the convergence of the kernel estimators, by Bosq (1998), we need to select ε_n such that $\varepsilon_n \to 0$ and $n\varepsilon_n \to \infty$ as $n \to \infty$.

In the above kernel estimator, $\bar{V}_{i+1}^\varepsilon(x) = \max\left(L_{i+1}(x), \bar{H}_{i+1}^\varepsilon(x)\right)$ and $\bar{V}_m^\varepsilon(x) = L_m(x)$. An advantage of this estimator is that it does not require any density information, but only the sample paths S_i^j, $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. However, its performance is typically poor since it essentially exploits only the information in the S_i^js which are close to x. Generally, kernel estimators have a rate of convergence of $(n\varepsilon_n)^{-1/2}$ which is slower than the typical $n^{-1/2}$.

Basically, the weight function $w(x, S_i^j)$ is crucial to the performance of the estimator. Intuitively, with further information, e.g., the densities, one may obtain better weights, and hence better estimators for $H_i(x)$, which have faster rate of convergence. For instance, the rate of convergence of the estimators of Broadie and Glasserman (1997) is $n^{-1/2}$.

In the following two subsections, we apply conditioning approach to Equation (1) to incorporate more information in the weight functions, and derive estimators that have better rate of convergence.

3.1 Forward-Looking Weights

Note that by conditioning on S_{i+1}, we have

$$\begin{aligned} H_i(x) &= \lim_{\varepsilon\to 0} \frac{\mathrm{E}\left[e^{-r\tau}V_{i+1}(S_{i+1})\cdot 1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]}{\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]} \\ &= \lim_{\varepsilon\to 0} \frac{\mathrm{E}\left(\mathrm{E}\left[e^{-r\tau}V_{i+1}(S_{i+1})\cdot 1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\middle| S_{i+1}\right]\right)}{\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]} \\ &= \lim_{\varepsilon\to 0} \frac{\mathrm{E}\left(e^{-r\tau}V_{i+1}(S_{i+1})\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\middle| S_{i+1}\right]\right)}{\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]}. \end{aligned}$$

With some regularity conditions, we can take the limit into the expectations. Then we have

$$H_i(x) = \mathrm{E}\left[e^{-r\tau}V_{i+1}(S_{i+1})w(i,x,S_{i+1})\right],$$

where

$$w(i,x,S_{i+1}) = \lim_{\varepsilon\to 0} \frac{\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\middle| S_{i+1}\right]}{\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]}.$$

Suppose that the transition density of S_{i+1} given $S_i = x$ and the marginal density of S_i are available, denoted by $f_i(x,\cdot)$, and $f(i,\cdot)$ respectively. When they are smooth, we have

$$\begin{aligned} & w(i,x,S_{i+1}) \\ &= \frac{\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\middle| S_{i+1}\right]}{\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]} \\ &= \frac{\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_{x-\varepsilon}^{x+\varepsilon} f(i,u)f_i(u,S_{i+1})\,du}{f(i+1,S_{i+1})} \\ &\quad \times \frac{1}{\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_{x-\varepsilon}^{x+\varepsilon} f(i,u)\,du} \\ &= \frac{f(i,x)f_i(x,S_{i+1})}{f(i+1,S_{i+1})}\cdot\frac{1}{f(i,x)} = \frac{f_i(x,S_{i+1})}{f(i+1,S_{i+1})}, \end{aligned}$$

where the third equality follows from the mean value theorem.

The above weight $w(i,x,S_{i+1})$ involves two density functions, $f_i(x,\cdot)$ and $f(i+1,\cdot)$. In practice $f_i(x,\cdot)$ is usually known or can be calculated, since it is actually the transition density which is used to generate the sample pathes of the underlying asset price. However, the explicit expression of $f(i+1,\cdot)$ is often unknown or can not be easily calculated, except for some simple models, e.g., S_t following geometric Brownian motion. When the explicit expression of $f(i+1,\cdot)$ is unknown, we may estimate it by using the transition densities. Since

$$f(i+1,v) = \mathrm{E}\left[f_i(S_i,v)\right],$$

then $f(i+1,S_{i+1})$ can be approximated by a sample mean, i.e.,

$$\widehat{f}(i+1,S_{i+1}) = \frac{1}{n}\sum_{k=1}^{n} f_i(S_i^k,S_{i+1}).$$

Therefore, we obtain two weights, denoted by $\overline{w}_1$ and $\overline{w}_2$ respectively,

$$\overline{w}_1(i,x,S_{i+1}) = \frac{f_i(x,S_{i+1})}{f(i+1,S_{i+1})}, \qquad (2)$$

$$\overline{w}_2(i,x,S_{i+1}) = \frac{f_i(x,S_{i+1})}{\frac{1}{n}\sum_{j=1}^{n} f_i(S_i^j,S_{i+1})}. \qquad (3)$$

We refer to these two weights as forward-looking weights, since they are obtained by conditioning on S_{i+1}, the sample paths in the next exercise date.

In fact, the forward-looking weights derived here are special cases of the weights in Broadie and Glasserman (1997). As shown in Broadie and Glasserman (1997), the weight $\overline{w}_1(i,x,S_{i+1})$ may lead to estimator whose variance grows exponentially with the number of exercise opportunities, while $\overline{w}_2(i,x,S_{i+1})$ can avoid this problem.

Generally speaking, the weights in Broadie and Glasserman (1997) exploit the information of the next exercise date, and they are obtained from an importance sampling viewpoint, rather than the conditioning viewpoint in our analysis. In their work the weights can be generally expressed as $w(i,x,S_{i+1}) = f_i(x,S_{i+1})/g_{i+1}(S_{i+1})$, where $g_{i+1}(\cdot)$ is the density of S_{i+1} from which the mesh points S_{i+1}^j's are actually generated. Emphasis should be given to that $f_i(x,S_{i+1})$ is the transition density under risk-neutral measure while the marginal density $g_{i+1}(\cdot)$ may not be under risk-neutral measure. Since the choice of $g_{i+1}(\cdot)$ is crucial to the performances of the estimators, Broadie and Glasserman (1997) suggest a good choice of $g_{i+1}(\cdot)$, $g_{i+1}(u) = \frac{1}{n}\sum_{k=1}^{n} f_i(S_i^k,u)$, which is called average density function. Then the weight function becomes $\overline{w}_2(i,x,S_{i+1})$. Intuitively, the average density function is equivalent to generating n independent paths of S_t and then "forgetting" the path to which each S_i^j belongs (see, e.g., Broadie and Glasserman 1997 or Avramidis and Hyden 1999). For more details of the weights in Broadie and Glasserman (1997), one is referred to Glasserman (2004) for a comprehensive overview.

3.2 Binocular Weights

Notice that the forward-looking weights are obtained by conditioning on the information of the next exercise date. Now we take one step further: what if we conditions on not only the information of the next exercise date, but also the information of the last exercise date? Motivated by the usage of Brownian bridge sampling in Monte Carlo methods, hopefully we may obtain new weights. We describes how we can do so. Since these new weights use the information on both sides of the current exercise date, we refer to them as binocular weights.

By conditioning on S_{i-1} and S_{i+1}, we have,

$$\begin{aligned} H_i(x) & \\ = & \lim_{\varepsilon\to 0} \frac{\mathrm{E}\left(e^{-r\tau}\mathrm{E}\left[V_{i+1}(S_{i+1})\cdot 1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\middle| S_{i-1},S_{i+1}\right]\right)}{\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]} \\ = & \lim_{\varepsilon\to 0} \frac{\mathrm{E}\left(e^{-r\tau}V_{i+1}(S_{i+1})\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\middle| S_{i-1},S_{i+1}\right]\right)}{\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]}. \end{aligned}$$

With some regularity conditions we can take the limit inside the expectation. Then

$$H_i(x) \quad = \quad \mathrm{E}\left[e^{-r\tau}V_{i+1}(S_{i+1})\cdot w(i,x,S_{i-1},S_{i+1})\right],$$

where

$$w(i,x,S_{i-1},S_{i+1}) = \frac{\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\middle| S_{i-1},S_{i+1}\right]}{\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]}.$$

Let $f_{i|i-1,i+1}(\cdot,v_1,v_2)$ denote the conditional density of S_i given $S_{i-1}=v_1$ and $S_{i+1}=v_2$, and we assume that it is a smooth function. Then by the mean value theorem,

$$w(i,x,S_{i-1},S_{i+1}) = \frac{f_{i|i-1,i+1}(x,S_{i-1},S_{i+1})}{f(i,x)}.$$

For many models, the expression of $f_{i|i-1,i+1}(\cdot,v_1,v_2)$ is known or can be approximated based on the bridge sampling techniques. For instance, if S_t follows a geometric Brownian motion, then $f_{i|i-1,i+1}(\cdot,v_1,v_2)$ can be obtained using the result for Brownian bridge. It can also be calculated or approximated in other models, e.g., the variance-gamma model.

Since the marginal density $f(i,x)$ is typically unknown except for some very simple models of S_t, we may use $f_{i|i-1,i+1}(\cdot,v_1,v_2)$ to estimate it. Note that

$$f(i,x) = \mathrm{E}\left[f_{i|i-1,i+1}(\,\cdot\,,S_{i-1},S_{i+1})\right],$$

then $f(i,x)$ can be unbiasedly estimated by

$$\hat{f}(i,x) = \frac{1}{n}\sum_{k=1}^{n} f_{i|i-1,i+1}(x,S_{i-1}^k,S_{i+1}^k).$$

Then we obtain two new weights:

$$\widetilde{w}_1(i,x,S_{i-1},S_{i+1}) \quad = \quad \frac{f_{i|i-1,i+1}(x,S_{i-1},S_{i+1})}{f(i,x)}, \qquad (4)$$

$$\widetilde{w}_2(i,x,S_{i-1},S_{i+1}) \quad = \quad \frac{f_{i|i-1,i+1}(x,S_{i-1},S_{i+1})}{\frac{1}{n}\sum_{k=1}^{n} f_{i|i-1,i+1}(x,S_{i-1}^k,S_{i+1}^k)}. \qquad (5)$$

Note the denominator of $\widetilde{w}_1(i,x,S_{i-1},S_{i+1})$ is exactly the marginal density $f(i,x)$, while in $\widetilde{w}_2(i,x,S_{i-1},S_{i+1})$ it is replaced by an average. As we have discussed for the forward-looking weights $\overline{w}_1(i,x,S_{i+1})$ and $\overline{w}_2(i,x,S_{i+1})$, using the marginal density may lead to estimator whose variance grows exponentially with the number of exercise opportunities, while the use of an average can avoid this problem. We conjecture that the binocular weights have the similar properties, and we indeed observe this phenomenon in numerical experiments. Therefore, we recommend to use $\widetilde{w}_2(i,x,S_{i-1},S_{i+1})$ when both can be implemented.

4 EXAMPLES

In this section we use several examples to illustrate how the forward-looking and binocular weights can be applied. We first consider the Black-Schole model where the underlying asset follows a geometric Brownian motion, and then general diffusion models. At the end we consider the variance-gamma model, which is a Lévy process.

For the Black-Scholes model, both forward-looking and binocular weights can be derived, while for general diffusion models, forward-looking weights can be derived but binocular weights need to be approximated. Forward-looking weights can also be derived for the variance-gamma model, but it is not practical to implement them since they are expressed in terms of expectations and hence require to be evaluated using extra simulations which may be computationally intensive. However, binocular weights with explicit forms can be derived for the variance-gamma model, which can be implemented practically.

4.1 Black-Scholes Model

Suppose that the price of the underlying asset follows a Geometric Brownian motion under the risk neutral measure, i.e.,

$$\frac{dS_t}{S_t} = (r-\delta)dt + \sigma dB_t,$$

where r is the risk-free interest rate, δ is the dividend rate, σ is the volatility, and B_t is a standard Brownian motion.

Then we have $S_i = S_0 e^{(r-\delta-\frac{1}{2}\sigma^2)t_i+\sigma B_{t_i}}$, and by elementary calculation,

$$f_i(u,v) = \frac{1}{v\sigma\sqrt{\tau}}\phi\left(\frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{v}{u}\right)-(\mu-\sigma^2/2)\tau\right]\right),$$
$$f(i,x) = \frac{1}{x\sigma\sqrt{t_i}}\phi\left(\frac{1}{\sigma\sqrt{t_i}}\left[\log\left(\frac{x}{S_0}\right)-\left(\mu-\sigma^2/2\right)t_i\right]\right),$$

where $\phi(\cdot)$ denotes the standard normal density.

Plug the transition density $f_i(u,v)$ and the marginal density $f(i,x)$ in Equations (2) and (3), then the forward-looking weights can be obtained.

To consider the binocular weights, we need to first obtain the conditional density of S_i given S_{i-1} and S_{i+1}. To do so, we use the result of a Brownian bridge. Conditioning on B_{i-1} and B_{i+1}, B_i is a Brownian bridge where B_i represents B_{t_i} for simplicity of notation. Particularly, $B_i \sim 1/2\,[B_{i-1}+B_{i+1}]+\sqrt{\tau/2}\cdot Z$ (see, e.g., Avramids and L'Ecuyer 2006), where Z follows a standard normal distribution, and the operator "$\sim$" stands for equivalence in distribution. Then conditioning on S_{i-1} and S_{i+1}, we can easily obtain that

$$S_i \quad \sim \quad \sqrt{S_{i-1}\cdot S_{i+1}}\cdot e^{\sigma\sqrt{\tau/2}\cdot Z}.$$

Then by some simple algebra, we have

$$f_{i|i-1,i+1}(x,v_1,v_2) = \frac{1}{x\sigma\sqrt{\tau/2}}\phi\left(\frac{1}{\sigma\sqrt{\tau/2}}\log\frac{x}{\sqrt{v_1\cdot v_2}}\right).$$

Therefore, the binocular weights of Equations (4) and (5) can also be applied.

4.2 General Diffusion Models

Suppose that the price of the underlying asset follows the diffusion process:

$$dS_t = \mu(t,S_t)dt + \sigma(t,S_t)dB_t.$$

We use Euler scheme to discretize S_t (Glasserman 2004). Under the scheme,

$$S_{i+1} = S_i + \mu(t_i,S_i)\tau + \sigma(t_i,S_i)\sqrt{\tau}Z_{i+1}, \quad i = 0,1,\cdots,m-1,$$

where $\{Z_1,Z_2,\cdots,Z_m\}$ are independent standard normal random variables. To simplify the notation, we let $\mu_i(S_i)$ and $\sigma_i(S_i)$ denote $\mu(t_i,S_i)$ and $\sigma(t_i,S_i)$ respectively.

For general diffusion models where the drift μ and volatility σ depend on t and S_t, it is easy to derive the forward-looking weights since the transition density $f_i(x,\cdot)$ can be calculated based on the discretization scheme, while it is not easy to derive the binocular weights since the conditional density $f_{i|i-1,i+2}(\cdot,v_1,v_2)$ is not easy to obtain. However, based on the discretization scheme, we are able to derive explicit expressions for approximations of the binocular weights. We describe how this can be done. Recall that conditioning on B_{i-1} and B_{i+1}, $B_i \sim 1/2\,(B_{i-1}+B_{i+1})+\sqrt{\tau/2}\cdot Z$, where Z is a standard normal random variable. Then

$$B_i - B_{i-1} \quad \sim \quad \frac{1}{2}(B_{i+1}-B_{i-1})+\sqrt{\tau/2}\cdot Z. \tag{6}$$

By Euler scheme, when the step size τ is small, approximately we have

$$S_i = \quad S_{i-1}+\mu_{i-1}(S_{i-1})\tau+\sigma_{i-1}(S_{i-1})\,[B_i-B_{i-1}], \tag{7}$$
$$S_{i+1} \approx \quad S_{i-1}+\mu_{i-1}(S_{i-1})2\tau+\sigma_{i-1}(S_{i-1})\,[B_{i+1}-B_{i-1}]. \tag{8}$$

Then combining Equations (6), (7) and (8) together we have, conditioning on S_{i-1} and S_{i+1},

$$S_i \sim \frac{1}{2}[S_{i-1}+S_{i+1}]+\sigma_{i-1}(S_{i-1})\sqrt{\tau/2}\cdot Z.$$

Then by some simple algebra, we have

$$f_{i|i-1,i+1}(x,v_1,v_2) = \frac{1}{\sigma_{i-1}(v_1)\sqrt{\tau/2}}\phi\left(\frac{x-\frac{1}{2}(v_1+v_2)}{\sigma_{i-1}(v_1)\sqrt{\tau/2}}\right).$$

Moreover, the transition density $f_i(x,\cdot)$ can be easily obtained:

$$f_i(x,u) = \frac{1}{\sigma(t_i,x)\sqrt{\tau}}\phi\left(\frac{u-x-\mu(t_i,x)\tau}{\sigma(t_i,x)\sqrt{\tau}}\right).$$

Then plugging $f_i(x,u)$ and $f_{i|i-1,i+1}(x,v_1,v_2)$ in Equations (3) and (5) respectively, we obtain the forward-looking weight and the binocular weight.

4.3 Variance-Gamma Model

The forward-looking and binocular weights work not only for the diffusion processes, but also for some other models. In this example, we consider the variance gamma model.

Following the notation in Avramidis and L'Ecuyer (2006). Let $B(t;\theta,\sigma)$ be a Brownian motion with drift parameter θ and variance parameter σ. Let $G(t;\mu,\nu)$ be a gamma process independent of $B(t;\theta,\sigma)$, with drift $\mu>0$ and volatility $\nu>0$. Then $G(0;\mu,\nu)=0$, the process G has independent increments, and the increments follow a Gamma distribution, i.e., $G(t+\delta;\mu,\nu)-G(t;\mu,\nu)\sim\Gamma(\delta\mu^2/\nu,\nu/\mu)$ for $t\geq 0$ and $\delta>0$.

A variance gamma process with parameters (θ, σ, ν) is defined by

$$X = \{X(t) = X(t;\theta,\sigma,\nu) = B(G(t;1,\nu),\theta,\sigma), t \geq 0\},$$

which is obtained by subjecting the Brownian motion to a random time change following a gamma process with parameter $\mu = 1$.

Then the risk neutralized asset price process S_t is

$$S_t = S_0 \exp\{(\omega + r - \delta)t + X(t)\},$$

where r is the risk-free interest rate, δ is the dividend rate, and the constant $\omega = \log(1 - \theta\nu - \sigma^2\nu/2)/\nu$ is chosen so that the discounted value of a portfolio invested in the asset is a martingale. In particular, $\mathrm{E}(S_t) = S_0 \exp[(r-\delta)t]$. Then we require

$$(\theta + \sigma^2/2)\nu < 1,$$

which ensures that $\mathrm{E}(S_t) < \infty$ for all $t > 0$. We assume that this requirement is satisfied in this example.

To analyze this model, we first review two schemes of simulating the variance gamma process. The first one is simulating it as 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 (2007) and Avramidis and L'Ecuyer (2006). With the first scheme, we will derive the transition density $f_i(x,\cdot)$, while with the second scheme we obtain the conditional density of S_i given S_{i-1}, S_{i+1}. Then the weights obtained by plugging these densities in Equations (3) and (5) can be applied.

We first look at the scheme of simulating variance gamma process as Gamma time-changed Brownian motion. For simplicity of notation, we let X_i and G_i denote $X(t_i)$ and $G(t_i)$ respectively from now on. We independently generate $\Delta G_i := G_{i+1} - G_i$ according to a Gamma distribution and Z_i from a standard normal distribution, which are independent of the past r.v.s. Particularly, let $\Gamma(a,b)$ denote the Gamma distribution with shape parameter a and scale parameter b, and (a,b) the normal distribution with mean a and variance b, then $\Delta G_i \sim \Gamma(\tau/\nu, \nu)$, and $Z_i \sim (0,1)$. Then we have

$$X_{i+1} = X_i + \theta\ \Delta G_i + \sigma\sqrt{\Delta G_i}\ Z_i.$$

By simple algebra we obtain the transition density of S_{i+1} given $S_i = x$:

$$\begin{aligned} & f_i(x,u) \\ &= \int_0^\infty \frac{1}{u\sigma\sqrt{y}}\phi\left(\frac{\log(u/x) - (\omega + r - \delta)\tau - \theta y}{\sigma\sqrt{y}}\right)\gamma(y)\,dy \\ &= \mathrm{E}^W\left[\frac{1}{u\sigma\sqrt{W}}\phi\left(\frac{\log(u/x) - (\omega + r - \delta)\tau - \theta W}{\sigma\sqrt{W}}\right)\right], \end{aligned}$$

where $\gamma(\cdot)$ is the density of the random variable which follows a $\Gamma(\tau/\nu, \nu)$ distribution, and the expectation is taken over W. Then we can use $f_i(x,u)$ to obtain forward-looking weights.

The variance-gamma process can also be simulated via Brownian bridge. 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. Let $\beta(a,b)$ denote the beta distribution with parameters a and b, then in the first step, we generate $Y \sim \beta(\tau/\nu, \tau/\nu)$, and let

$$G_i = G_{i-1} + (G_{i+1} - G_{i-1})Y. \tag{9}$$

Then in the second step, we generate $Z \sim \big(0, (G_{i+1} - G_i)\,\sigma^2 Y\big)$, and let

$$X_i = YX_{i+1} + (1-Y)X_{i-1} + Z. \tag{10}$$

Then we have

$$\begin{aligned} X_i &= YX_{i+1} + (1-Y)X_{i-1} + \sigma\sqrt{(G_{i+1} - G_i)Y}\cdot Z_1 \\ &= YX_{i+1} + (1-Y)X_{i-1} + Y_1 \cdot Z_1, \end{aligned}$$

where Z_1 is a standard normal random variable independent of Y, and $Y_1 = \sigma\sqrt{Y(1-Y)\,(G_{i+1} - G_{i-1})}$.

With the above bridge sampling scheme, we derive the conditional density of S_i given S_{i-1}, S_{i+1}, G_{i-1} and G_{i+1} by simple algebra. Specifically,

$$\begin{aligned} & f_{i|i-1,i+1}(x, S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}) \\ &= \int_0^1 \frac{1}{x\sigma\sqrt{y(1-y)[G_{i+1} - G_{i-1}]}}\tilde{\phi}(y)g(y)\,dy \\ &= \mathrm{E}^Y\left[\frac{1}{x\sigma\sqrt{Y(1-Y)[G_{i+1} - G_{i-1}]}}\tilde{\phi}(Y)\right], \end{aligned}$$

where

$$\tilde{\phi}(y) = \phi\left(\frac{\log\left(\frac{x}{S_{i+1}^{y}\cdot S_{i-1}^{1-y}}\right) + (2y-1)(w + r - \delta)\tau}{\sigma\sqrt{y(1-y)[G_{i+1} - G_{i-1}]}}\right),$$

$g(y)$ is the density of the random variable Y which follows a $\beta(\tau/\nu, \tau/\nu)$ distribution, and the expectation is taken for Y. Then the corresponding binocular weight can be obtained.

So far we have derived the forward-looking and binocular weights following exactly the analysis in the previous sections. For these weights, though the transition density $f_i(x, S_{i+1})$ and the conditional density $f_{i|i-1,i+1}(x, S_{i-1}, S_{i+1}, G(t_{i-1}), G(t_{i+1}))$ can be estimated by running extra Monte Carlo simulations, it may not be easy to implement in practice because of the huge computational effort required. Fortunately, we may obtain other weights which are much easier to implement, in the light of the conditioning viewpoint of the weights. This viewpoint 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 we can do so.

Rather than condition only on $(S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1})$, we additionally condition on G_i. Then similar to the previous analysis, a binocular weight can be expressed as:

$$\begin{aligned} & w(i, x, S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}, G_i) \\ = & \frac{\lim_{\varepsilon\to 0} \mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}} \middle| S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}, G_i\right]}{\lim_{\varepsilon\to 0} \mathrm{E}\left[1_{\{x-\varepsilon\le S_i\le x+\varepsilon\}}\right]} \\ = & \frac{f_c(i, x, S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}, G_i)}{\mathrm{E}\left[f_c(i, x, S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}, G_i)\right]}, \end{aligned} \tag{11}$$

where $f_c(i, x, S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}, G_i)$ is the conditional density of S_i given $(S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}, G_i)$.

Using Equations (9) and (10), by elementary algebra we can obtain

$$f_c(i, x, S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}, G_i) = \frac{1}{x\sigma\sqrt{l_i}}\phi(U_i),$$

where $\Delta G_i = G_{i+1} - G_i$,

$$p_i = \frac{\Delta G_{i-1}}{\Delta G_{i-1} + \Delta G_i}, \quad l_i = \frac{\Delta G_{i-1} \cdot \Delta G_i}{\Delta G_{i-1} + \Delta G_i},$$

and

$$U_i = \frac{\log\left[x/\left(S_{i+1}^{p_i} \cdot S_{i-1}^{1-p_i}\right)\right] + (\omega + r - \delta)(2p_i - 1)\tau}{\sigma\sqrt{l_i}}.$$

5 NUMERICAL STUDY

In the previous section we have shown how to applied the forward-looking and binocular weights for several examples. To illustrate the performances of these weights, we conduct numerical experiments for the Black-Scholes model and the variance-gamma model.

We consider an American call option underlying an asset following the Black-Scholes model, i.e., the underlying asset price follows a geometric Brownian motion. The option expires in three years and can be exercised at any of 10 equally spaced exercise opportunities. The payoff upon exercise at t_i is $(S_i - K)^+$, with $K = 100$ and $S_0 = 100$, volatility $\sigma = 0.2$, interest rate $r = 5\%$, and dividend yield $\delta = 10\%$. We have known that the price of this American call option is 7.98, obtained from a binomial lattice (see age 469 of Glasserman 2004). We use it as a benchmark value to test the performances of different weights. We conduct

Table 1: Results of forward-looking and binocular weights for the Black-Scholes model

	n	500	1000	1500	2000
forward	mean	8.281	8.131	8.075	8.048
	Var	0.186	0.090	0.051	0.037
	bias	0.301	0.151	0.095	0.048
binocular	mean	8.128	8.051	8.026	8.007
	Var	0.276	0.128	0.085	0.063
	bias	0.148	0.071	0.046	0.027

1000 replications to estimate the error of the estimators. We observed that the estimators using weights $\overline{w}_1$ and $\widetilde{w}_1$ which involve marginal densities, have large errors (the standard deviations can be in a order of 10^4 while the true value is 7.98). These large errors are due to some extreme large observations occasionally. This phenomenon coincides with the proof in Broadie and Glasserman (1997) that use of marginal densities in weights may lead to estimators whose variances grow exponentially with the number of exercise opportunities. Then we mainly compare the estimators using the weights $\overline{w}_2$ and $\widetilde{w}_2$. The comparison results are presented in Table 1, where we show the mean, variance and bias of the estimators correspond to forward-looking weight and binocular weight respectively. From the table we can see that binocular weight has smaller bias while the the forward-looking weight has smaller variance.

We also consider an American put option under the variance-gamma model, to illustrate the performance of the stochastic mesh method using the weight $w(i, x, S_{i-1}, S_{i+1}, G_{i-1}, G_{i+1}, G_i)$. In the experiments, we let $T = 0.5616$, $r = 5.41\%$, $\delta = 1.2\%$, $\sigma = 20.72\%$, $\nu = 0.5022$, $\theta = -0.2290$, $S_0 = 1369.4$ and $K = 1200$. These settings are cited from Hirsa and Madan (2003). We let $m = 10$ and the value of the American option at current time is approximately 35.56. By using the binocular weights as in Equation (11), the numerical results of the stochastic mesh estimator are summarized in Table 2. From the table we can see that and standard deviation (stdev) of the estimator decreases as the sample size increases.

6 CONCLUSIONS

In this paper we revisit the stochastic mesh method from a conditioning viewpoint. Based on this new viewpoint, binocular weights of the stochastic mesh method are derived, which exploit the information on both sides of the current exercise date. Though binocular weights may not be superior to the existing forward-looking weights, they can be applied

Table 2: Results of the binocular weights for the variance-gamma model

n	500	1000	1500	2000
mean	40.40	39.53	39.28	38.97
stdev	4.43	3.28	2.63	2.21

to some models, e.g., the variance-gamma model, where the forward-looking weights may not be applied efficiently.

For future research, it would be interesting to compare the forward-looking weights and the binocular weights for high dimensional problems, to examine whether the extra information has benefits in the estimation.

ACKNOWLEDGMENTS

This research was partially supported by Hong Kong Research Grants Council grant CERG 613907.

REFERENCES

Avramidis, A. N., and P. Hyden. 1999. Efficiency improvements for pricing American options with a stochastic mesh. *Proceedings of the 1999 Winter Simulation Conference*, 344-350.

Avramidis, A. N., and P. L'Ecuyer. 2006. Efficient Monte Carlo and quasi-Monte Carlo option pricing under the variance Gamma model. *Management Science*, **52**:1930-1944.

Avramidis, A. N., and H. Matzinger. 2004. Convergence of the stochastic mesh estimator for pricing American options. *Journal of Computational Finance*, **7**(4).

Bosq, D. 1998. *Nonparametric Statistics for Stochastic Processes.* Second Edition. Springer, New York.

Broadie, M., and P. Glasserman. 1997. A stochastic mesh method for pricing high-dimensional American options. PaineWebber Working Papers in Money, Economics and Finance #PW9804, Columbia Business School, New York.

Broadie, M., P. Glasserman, and Z. Ha. 2000. Pricing American options by simulation using a stochastic mesh with optimized weights. In *Probabilistic Constrained Optimization: Methodology and Applications*, ed S. Uryasev. Kluwer Academic Publishers, Norwell, Mass.

Broadie, M., P. Glasserman, and G. Jain. 1997. Enhanced Monte Carlo estimators for American option prices. *Journal of Derivatives*, **5**(1):25-44.

Fu, M. 2007. Variance-Gamma and Monte Carlo. In *Advances in Mathematical Finance*, ed Fu et al. Birkhauser, Boston.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer, New York.

Hirsa, A., and D. B. Madan. 2003. Pricing American options under variance gamma. *Journal of Computational Finance*, **7**(2):63-80.

Longstaff, F. A., and E. S. Schwartz. 2001. Valuing American options by simulation: a simple least-square approach. *Review of Financial Studies*, **14**:113-147.

Tsitsiklis, J., and B. Van Roy. 1999. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Transactions on Automatic Control*, **44**:1840-1851.

AUTHOR BIOGRAPHIES

GUANGWU LIU is a Ph.D. student in the Department of Industrial Engineering and Logistics Management at The Hong Kong University of Science and Technology. His research interests include simulation methodology and financial engineering. His e-mail address is `<liugw@ust.hk>`.

L. JEFF HONG is an assistant professor in industrial engineering and logistics management at The Hong Kong University of Science and Technology. His research interests include Monte-Carlo method, sensitivity analysis, financial engineering and simulation optimization. He is currently an associate editor of *Operations Research*, *Naval Research Logistics* and *ACM Transactions on Modeling and Computer Simulation*. His e-mail address is `<hongL@ust.hk>`.