

Sensitivity Analysis for Monte Carlo Simulation of Option Pricing

MICHAEL C. FU
College of Business and Management,
University of Maryland, College Park, MD 20742, USA
e-mail: mfu@umd5.umd.edu

JIAN-QIANG HU*
Department of Manufacturing Engineering,
Boston University, 44 Cummington Street, Boston, MA 02215, USA
e-mail: hu@enga.bu.edu

November 1993; revised September 1994, June 2005

ABSTRACT

Monte Carlo simulation is one alternative for analyzing options markets when the assumptions of simpler analytical models are violated. We introduce techniques for the sensitivity analysis of option pricing which can be efficiently carried out in the simulation. In particular, using these techniques, a single run of the simulation would often provide not only an estimate of the option value but also estimates of the sensitivities of the option value to various parameters of the model. Both European and American options are considered, starting with simple analytically tractable models to present the idea and proceeding to more complicated examples. We then propose an approach for the pricing of options with early exercise features by incorporating the gradient estimates in an iterative stochastic approximation algorithm. The procedure is illustrated in a simple example estimating the option value of an American call. Numerical results indicate that the additional computational effort required over that required to estimate a European option is relatively small.

Keywords: sensitivity analysis, option pricing, Monte Carlo simulation, stochastic approximation algorithm.

*J.Q. Hu was supported in part by the National Science Foundation under grants EID-9212122 and DDM-9212368.

Option pricing is an important area of research in the finance community. Due to the complexity of the underlying dynamics, analytical models for option pricing entail many restrictive assumptions, so for real-world applications approximate numerical methods are employed. One such method is Monte Carlo simulation. Boyle (1977) was among the first to propose using Monte Carlo simulation to study option pricing. Since then, many researchers, e.g., Figlewski (1992), Hull and White (1987), Johnson and Shanno (1987), and Scott (1987), have employed Monte Carlo simulation for analyzing options markets. The advantage of the approach is its generality in being able to model “imperfect” market conditions not easily captured in analytically tractable models; as Boyle (1977) has stated, “the Monte Carlo method should prove most valuable in situations where it is difficult if not impossible to proceed using a more accurate approach.” Its disadvantages are its computational inefficiency when compared to most other numerical methods and its tendency to result in an un insightful black box treatment of the model.

Given that Monte Carlo simulation is to be employed, there are ways to make the technique both more efficient and more insightful. For example, Boyle (1977) used control variates and antithetic variates to reduce variance in his studies. However, he makes the statement, “To obtain option values corresponding to different current stock prices a set of simulation trials has to be carried out for each starting stock price.” Presumably, this statement would also apply equally to other parameters of the system, as well. One step in reducing this inefficiency would be the availability of sensitivity estimates that did not resort to the “brute force” method of numerous additional simulation trials. The purpose of this paper is to introduce gradient estimation techniques from the stochastic discrete-event simulation community to option valuation models. (To avoid confusion with its usage in the lexicon of finance, we will avoid the use of the word “derivative” and use either of the two terms *gradient* or *sensitivity*.) These techniques provide the simulation user with an efficient way to obtain estimates for sensitivities of the option value with respect to various parameters of the model, both input parameters such as the current stock price, the striking price, the time to expiration, the volatility, and the interest rate, and “decision” parameters such as the early exercise threshold levels for the American option.

We briefly review the mechanics of an option. The chief components are the striking price and the expiration date (or time to expiration). In a *call* option, the holder of the option has the right to purchase the underlying asset (e.g., shares of a stock) at the striking price. In the case of a European

option, this right can only be exercised on the expiration date. In the case of an American option, this right can be exercised at any time up to and including the expiration date. In either case, the option does not have to be exercised at all on the expiration date if the stock price is below the striking price (“out of the money” in finance terminology). A *put* option, on the other hand, is identical except that it is the right to *sell* at the striking price. Calls and puts can themselves be either bought or sold, which in combination with opportunities to buy and sell the asset itself, creates numerous hedging strategy possibilities. The dynamics of the underlying asset are generally described by a stochastic differential equation, usually containing diffusion processes and jump processes. Due to the complexity of these dynamics, exact “easy” expressions for the option price are not obtainable except for the simplest models, e.g., for the justly celebrated Black-Scholes model. This difficulty necessitates the use of approximate numerical methods, with Monte Carlo simulation being one of the most general and easily applied methods.

The chief benefits of the gradient estimation techniques that we introduce in this paper and which we hope to illustrate through our examples are the following:

- implementation of these techniques requires very little additional overhead in the simulation;
- the estimation is computationally efficient compared to the multiple runs that would be needed to construct finite difference estimates for each parameter of interest;
- the estimators can handle both jump processes and continuous processes;
- the estimators have lower variance (generally) than naive finite difference estimates;
- the simultaneous availability of sensitivity estimates opens up the possibility for on-line adjustment of certain decision parameters (e.g., exercise thresholds).

The last item leads to the possibility of using the estimates in the difficult problem of pricing options with early exercise features via Monte Carlo simulation. In particular, by viewing the pricing of an American call as an optimization problem, the gradient with respect to early exercise threshold levels can be incorporated into a stochastic approximation algorithm to estimate the option value, thus refuting the widely held believe that

“Monte Carlo simulation can only be used for European-style options” (Hull 1993, p.363).

Numerical results that we report for a few examples indicate that the algorithm converges very quickly.

Stochastic derivative estimation in Monte Carlo simulation has been an active research area in the study of discrete-event systems such as queueing systems. The two most widely used techniques are perturbation analysis (PA) and the likelihood ratio (LR) method (also known as the score function method). Monographs for the former are Ho and Cao (1991) and Glasserman (1991), and a recent monograph for the latter is Rubinstein and Shapiro (1993). In this work, we will concentrate on PA, which is a sample path method for estimating the gradient. Roughly speaking, if J_T is a sample performance over some time horizon $[0, T]$, $E[J_T]$ its expectation, and θ a parameter of interest, then our interest is in estimating the sensitivity $\partial E[J_T]/\partial\theta$. Infinitesimal perturbation analysis (IPA) simply takes the estimate $\partial J_T/\partial\theta$, so in order for it to be unbiased, the following must be satisfied:

$$E \left[\frac{\partial J_T}{\partial\theta} \right] = \frac{\partial E[J_T]}{\partial\theta}. \quad (1)$$

This is usually established by finding conditions on the system under which the dominated convergence theorem can be applied to exchange the limit (differentiation) and integral (expectation). In applications to queueing, this usually means a.s. (almost sure) continuity of J_T with respect to θ . What makes this a non-trivial problem in general is the dependence of J_T on θ being non-explicit, i.e., there may be a dynamic, recursive relationship. For (smoothly) continuously changing stock prices (e.g., governed by a diffusion process), however, the problem is usually trivial, as we shall see. However, the presence of jumps, e.g., in the form of ex-dividends or more generally a jump process, makes the problem more interesting, and sometimes the conditions for (1) are violated.

When the conditions for unbiasedness of the IPA estimator are violated, an alternative means to IPA employing conditional Monte Carlo estimator can be used. This approach has come to be known as smoothed perturbation analysis (SPA). By conditioning on some appropriate set of random variables represented by say Z , we hope to be able to satisfy the following less stringent condition:

$$E \left[\frac{\partial E[J_T|Z]}{\partial\theta} \right] = \frac{\partial E[J_T]}{\partial\theta}. \quad (2)$$

The term “smoothed” comes about from the usual smoothing property of a conditional expectation (Gong and Ho 1987), which would require milder conditions under which the exchange in (2) holds over those required for the exchange in (1).

We now briefly compare and contrast our work to other related work. Early work in the area of sensitivity analysis for option pricing includes Broadie and Glasserman (1996), which uses the likelihood ratio method alluded to earlier and perturbation analysis, but the use of perturbation analysis is restricted to infinitesimal perturbation analysis, and they consider European and Asian options, but without early exercise features, as in American options. With respect to using Monte Carlo simulation to perform pricing of options with early exercise features, early work includes Tilley (1993) and Grant, Vora, and Weeks (1997). The former analyzes only American options, whereas the latter considers more general path-dependent options such as Asian options. However, their approaches differ totally from ours, in that they do not employ gradient estimation techniques with stochastic approximation; rather, Monte Carlo simulation is used to perform the backwards induction of the dynamic programming formulation to determine the optimal exercise threshold levels and then the option value. In light of Hull's earlier quote, we strongly believe that all new alternatives which open up the possibility of using Monte Carlo simulation for the pricing of early exercise options merit further consideration and research.

The rest of the paper is organized as follows. In section 1, we apply PA to a simple European call option, including the celebrated Black-Scholes model as a special case. In section 2, we consider a more complicated example of an American call option on an underlying asset with dividends. In section 3, we show how the PA estimates can be incorporated in a stochastic approximation algorithm to iteratively determine the optimal early exercise threshold levels, and hence be used to estimate the value of options with early exercise features. Empirical performance of this algorithm is provided via numerical results on American options, considering both analytically tractable Black-Scholes-type models (diffusion model giving lognormally-distributed stock prices; these constitute the most celebrated models for options pricing in the finance literature), and intractable, non-Markovian, non-continuous (i.e., jump) models.

1 A Simple European Call

We begin by introducing the following notation to be used throughout the rest of the paper:

S_t = the asset price at time t ,

S_0 = the initial stock asset,

r = the annualized riskless interest rate (compounded continuously),

μ, σ = parameters of the distribution of the underlying stock,

K = the striking price of the option contract,

T = the lifetime (expiration date) of the option contract,

J_T = the net present value return of the option on its expiration.

We are without loss of generality designating the present time as time 0. Usually, σ will represent the volatility of the underlying stock, and μ the drift or some other mean-related parameter. We assume that S_t and J_T are random variables, whereas the rest are constants.

In this section, we consider a simple European call where stock prices change continuously, and the exchange in (1) can be established in a straightforward manner. The option value for a European call is given by

$$J_T = (S_T - K)^+ e^{-rT} \triangleq \max(S_T - K, 0) e^{-rT}. \quad (3)$$

We are interested in $E[J_T]$, the expected return of the option at its expiration date, and its sensitivity $\partial E[J_T]/\partial\theta$, where θ is the parameter of interest (e.g., $S_0, r, \sigma, \mu, K, T$). We wish to derive an estimator for $\partial E[J_T]/\partial\theta$. Since r and T are assumed constant, we can write $E[J_T] = e^{-rT} E[(S_T - K)^+]$, so

$$\frac{\partial E[J_T]}{\partial\theta} = e^{-rT} \left[\frac{\partial E[(S_T - K)^+]}{\partial\theta} - E[(S_T - K)^+] \frac{\partial(rT)}{\partial\theta} \right],$$

and hence the problem reduces to finding an unbiased estimator for $\partial E[(S_T - K)^+]/\partial\theta$.

The IPA derivative estimator from differentiating (3) is

$$\begin{aligned} \frac{\partial J_T}{\partial\theta} &= e^{-rT} \begin{cases} \frac{\partial(S_T - K)}{\partial\theta} - (S_T - K) \frac{\partial(rT)}{\partial\theta} & \text{for } S_T > K \\ 0 & \text{otherwise} \end{cases} \\ &= e^{-rT} \left[\frac{\partial(S_T - K)}{\partial\theta} - (S_T - K) \frac{\partial(rT)}{\partial\theta} \right] \mathbf{1}\{S_T > K\}, \end{aligned} \quad (4)$$

where $\mathbf{1}\{\cdot\}$ is the set indicator function.

Theorem 1. If S_t is a.s. continuous and a.s. piecewise differentiable with respect to $\theta \in \Theta$ for some open set Θ , and

$$E \left[\sup_{\theta \in \Theta} S_T \right] < \infty, \quad E \left[\sup_{\theta \in \Theta} \left| \frac{\partial S_T}{\partial \theta} \right| \right] < \infty,$$

(the second “sup” taken over only those points in Θ where the derivative exists), then the IPA estimator given by (4) is an unbiased estimator for $\partial E[J_T]/\partial \theta$.

Remark. Thus, for European options with smoothly changing stock prices, IPA works.

Proof. The proof essentially follows Glasserman (1991). If S_t is a.s. continuous with respect to θ , then $(S_T - K)^+$ and hence J_T are a.s. continuous and a.s. piecewise differentiable with respect to θ (its only point of nondifferentiability at $S_T = K$). Therefore, based on a generalized mean value theorem (p.15 in Glasserman 1991), there exists a $\Delta > 0$ such that

$$\left| \frac{J_T(\theta + \epsilon) - J_T(\theta)}{\epsilon} \right| \leq \sup_{0 \leq \epsilon \leq \Delta} \left| \frac{\partial J_T(\theta + \epsilon)}{\partial \theta} \right|.$$

Furthermore, we have by the second set of conditions $E \left[\sup_{0 \leq \epsilon \leq \Delta} dJ_T(\theta + \epsilon)/\partial \theta \right] \leq \infty$, so using the dominated convergence theorem, we have

$$\begin{aligned} \frac{\partial E[J_T]}{\partial \theta} &= \lim_{\epsilon \rightarrow 0} \frac{E[J_T(\theta + \epsilon)] - E[J_T(\theta)]}{\epsilon} \\ &= E \left[\lim_{\epsilon \rightarrow 0} \frac{J_T(\theta + \epsilon) - J_T(\theta)}{\epsilon} \right] \quad \text{applying the dominated convergence theorem} \\ &= E \left[\frac{\partial J_T}{\partial \theta} \right]. \end{aligned}$$

Hence, the estimator given by (4) is an unbiased estimator for $\partial E[J_T]/\partial \theta$. □

The stochastic derivatives for each of the five different parameters are given in Table 1. From a simulation trial used to estimate of S_T , the table can be used to get estimates of $\partial E[J_T]/\partial \theta$ for each parameter of interest, along with the usual estimate of $E[J_T]$ itself via $e^{-rT}(S_T - K)^+$. Notice that no distributional assumptions need be made a priori.

Assuming the stock price is independent of the striking price, the sensitivity estimate with respect to the striking price K is already determined, and thus an unbiased estimate for $\partial E[J_T]/\partial K$ is given by $\partial J_T/\partial K = -e^{-rT} \mathbf{1}\{S_T > K\}$. This result follows from (4) and Theorem 1, but can also be established easily in a direct manner, since

$$E[J_T] = e^{-rT} E[(S_T - K)^+] = e^{-rT} \int_K^\infty (x - K) dF_{S_T}(x),$$

Table 1: IPA Derivative Estimators for the European Option

θ	$\frac{\partial J_T}{\partial \theta}$ for $S_T > K$ (0 otherwise)
K	$-e^{-rT}$
S_0	$e^{-rT} \frac{\partial S_T}{\partial S_0}$
r	$e^{-rT} [T(K - S_T) + \frac{\partial S_T}{\partial r}]$
σ	$e^{-rT} \frac{\partial S_T}{\partial \sigma}$
T	$e^{-rT} [r(K - S_T) + \frac{\partial S_T}{\partial T}]$

$$\text{so } \partial E[J_T]/\partial K = e^{-rT} \int_K^\infty (-1) dF_{S_T}(x) = E[-e^{-rT} \mathbf{1}\{S_T > K\}].$$

Some other special cases which may be of practical interest are when the parameters S_0 and/or σ are scale or location parameters. If they are *scale* parameters of the stock price, i.e., if they double then the price of the stock would double, then

$$\frac{\partial S_T}{\partial \theta} = \frac{S_T}{\theta}.$$

If they are *location* parameters of the stock price, i.e., if they are increased (or decreased) some amount then the price of the stock would increase (or decrease) by the same amount, then

$$\frac{\partial S_T}{\partial \theta} = 1.$$

Special Case: Black-Scholes

We illustrate the derivatives for the special case of Black-Scholes (see, e.g., Hull 1993). The stock price S_t is assumed to follow the dynamics given by the stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dZ_t,$$

where dZ_t is the standard Wiener process whose increments are uncorrelated, and μ and σ^2 are the annualized drift and variance rate of the underlying stock, respectively. Risk-neutral valuation (e.g., Cox and Ross 1976; see also Harrison and Pliska 1981) justifies $\mu = r$, and the solution of this differential equation is given by Ito's equation, which yields a lognormally distributed random variable

$$S_t = S_0 \exp[(r - \sigma^2/2)t + \sigma\sqrt{t}Z], \tag{5}$$

where $Z \sim N(0, 1)$, a normal random variable with mean 0 and standard deviation 1. Note that the starting price S_0 is a scale parameter for the stock price S_t .

Table 2: IPA Derivative Estimator for Black-Scholes

θ	$\frac{\partial J_T}{\partial \theta}$ for $S_T > K$ (0 otherwise)
K	$-e^{-rT}$
S_0	$e^{-rT} \frac{S_T}{S_0}$
r	$e^{-rT} KT$
σ	$e^{-rT} S_T(-\sigma T + \sqrt{T}Z)$
T	$e^{-rT} \left[S_T \left(-\frac{\sigma^2}{2} + \frac{\sigma Z}{2\sqrt{T}} \right) + Kr \right]$

Straightforward differentiation of (5) for (4) gives the estimates in Table 2. Since S_t is continuous with respect to S_0, r, σ , and t , Theorem 1 applies, and the estimators are unbiased. However, since Black-Scholes is analytically tractable, unbiasedness can be established directly by calculating the left-hand and right-hand sides of the required condition (1), instead of invoking the dominated convergence theorem to exchange the two operators. First denote

$$d_1 = \frac{1}{\sigma\sqrt{T}} \left(\ln \frac{S_0}{K} + (r + \sigma^2/2)T \right), \quad d_2 = d_1 - \sigma\sqrt{T}.$$

Then we have (e.g., Stoll and Whaley 1993, p.319)

$$\begin{aligned} E[J_T(S_0, K; r, \sigma)] &= S_0 N(d_1) - K e^{-rT} N(d_2), \\ \int_{-d_2}^{\infty} dN(z) &= N(d_2), \\ \int_{-d_2}^{\infty} S_T dN(z) &= S_0 e^{rT} N(-d_2), \\ \int_{-d_2}^{\infty} z dN(z) &= N'(d_1), \\ \int_{-d_2}^{\infty} z S_T dN(z) &= \sigma\sqrt{T} S_0 e^{rT} N(-d_2) + S_0 e^{rT} N'(-d_2), \end{aligned}$$

where $N(\cdot)$ denotes the standard normal c.d.f. Also note that $S_T > K$ is equivalent to $Z > -d_2$, therefore,

$$\begin{aligned} E \left[\frac{\partial J_T}{\partial K} \mathbf{1}\{S_T > K\} \right] &= -e^{-rT} N(d_2), \\ E \left[\frac{\partial J_T}{\partial S_0} \mathbf{1}\{S_T > K\} \right] &= N(-d_2), \\ E \left[\frac{\partial J_T}{\partial r} \mathbf{1}\{S_T > K\} \right] &= K T e^{-rT} N(d_2), \\ E \left[\frac{\partial J_T}{\partial \sigma} \mathbf{1}\{S_T > K\} \right] &= S_0 \sqrt{T} N'(-d_2), \\ E \left[\frac{\partial J_T}{\partial T} \mathbf{1}\{S_T > K\} \right] &= \frac{\sigma S_0}{2\sqrt{T}} N'(-d_2) + K r e^{-rT} N(d_2), \end{aligned}$$

which are consistent with the analytical derivatives obtained by differentiating the expression for $E[J_T]$ above.

2 An American Call on a Dividend-Paying Stock

In this section, we consider a more complicated option model: an American call option on a stock which distributes a cash dividend of amount D_j at time $t_j = \sum_{i=1}^j \tau_i$ ($\tau_j > 0$), $j = 1, \dots, \eta(T)$, where $\eta(T)$ is the number of dividends distributed during the lifetime of the call contract, τ_1 denotes the time until the first ex-dividend point, $\tau_i, i = 2, \dots, \eta(T) - 1$ denote the time between subsequent ex-dividends, and $\tau_{\eta(T)}$ denotes the time from the last ex-dividend point to the expiration date. Following standard models (e.g., Stoll and Whaley 1993), we assume that after the ex-dividend, the stock price drops by the amount of the dividend, i.e., $S_{t_j^+} = S_{t_j^-} - D_j$. (Any other drop can be handled without loss of generality.) For notational convenience, we also denote $\tau_{\eta(T)+1} = T - \sum_{i=1}^{\eta(T)} \tau_i, t_0 = 0, t_{\eta(T)+1} = T$. We will assume that the dividend amounts $\{D_j\}$ are known (deterministic). Although an American call option can be exercised at any time before the expiration date T , under the assumption of no transaction costs, it is well-known that the option should only be exercised – if at all – right before an ex-dividend date or at the expiration date (e.g., Stoll and Whaley 1993). Thus, we assume that the following threshold exercise policy is adopted: there is a stock price $s_j (\geq K)$ associated with t_j such that the option is exercised if $S_{t_j^-} > s_j$.

The sample performance can be written as

$$J_T = e^{-rT} \left(\sum_{i=1}^{\eta(T)} \left[\prod_{j=1}^{i-1} \mathbf{1}\{S_{t_j^-} \leq s_j\} \right] \mathbf{1}\{S_{t_i^-} > s_i\} (S_{t_i^-} - K) e^{r(T-t_i)} + \prod_{j=1}^{\eta(T)} \mathbf{1}\{S_{t_j^-} \leq s_j\} (S_T - K)^+ \right), \quad (6)$$

and we are interested in estimating $\partial E[J_T]/\partial\theta$. It should be clear that S_t is no longer a.s. continuous with respect to its parameters, since jumps occur at each ex-dividend point. Hence, the conditions of Theorem 1 are not met, and (4) will be biased for this model. We note again that since r and T are constant,

$$\frac{\partial E[J_T]}{\partial\theta} = e^{-rT} \left[\frac{\partial E[\hat{J}_T]}{\partial\theta} - E[\hat{J}_T] \frac{\partial(rT)}{\partial\theta} \right], \quad J_T = e^{-rT} \hat{J}_T,$$

where

$$\hat{J}_T = \sum_{i=1}^{\eta(T)} \left[\prod_{j=1}^{i-1} \mathbf{1}\{S_{t_j^-} \leq s_j\} \right] \mathbf{1}\{S_{t_i^-} > s_i\} (S_{t_i^-} - K) e^{r(T-t_i)} + \prod_{j=1}^{\eta(T)} \mathbf{1}\{S_{t_j^-} \leq s_j\} (S_T - K)^+.$$

Thus, the more complicated function \hat{J}_T replaces $(S_T - K)^+$ from the model of the previous section. The European call option can be thought of as the special case of $s_j = \infty$ for all j .

We will assume that aside from the discrete drops at ex-dividend points, the stock price changes continuously, i.e., according to

$$h(Z; S, t, \mu, \sigma),$$

which gives the stock price after duration t from the present, net of the present value of escrowed dividends, given current stock price S and random vector Z , which is independent of the parameters. Defining \tilde{S}_t as the corresponding process, we have

$$\tilde{S}_t = h(Z; \tilde{S}_0, t, \mu, \sigma).$$

For example, for the Black-Scholes log-normal distribution, we would have

$$h(Z; S, t, r, \sigma) = S e^{(r-\sigma^2/2)t + \sigma\sqrt{t}Z},$$

where Z is a standard $N(0, 1)$ normal variable. The relationship between \tilde{S}_t and S_t is given by

$$S_t = \tilde{S}_t + \sum_{i=j+1}^{\eta(T)} D_i e^{-r(t_i-t)}, \quad \text{for } t_j \leq t < t_{j+1}, \quad j = 0, 1, \dots, \eta(T).$$

In particular, just prior to ex-dividend points where early exercise decisions are made, we have

$$S_{t_j^-} = \tilde{S}_{t_j^-} + \sum_{i=j}^{\eta(T)} D_i \exp\left(-r \sum_{k=j+1}^i \tau_k\right), \quad j = 1, \dots, \eta(T).$$

(The equation also holds at the non-ex-dividend terminal point $j = \eta(T) + 1$.)

For illustrative purposes, we first consider $\eta(T) = 1$, i.e., there is a single ex-dividend payable during the lifetime of the contract $[0, T]$. We will drop the subscript on the dividend, i.e., $D_1 \triangleq D$, and $S_{t_1^+} = S_{t_1^-} - D$. Then, we have

$$\hat{J}_T = \mathbf{1}\{S_{t_1^-} > s\}(S_{t_1^-} - K)e^{r(T-t_1)} + \mathbf{1}\{S_{t_1^-} \leq s\}(S_T - K)^+, \quad (7)$$

where $S_{t_1^-} = h(Z_1; \tilde{S}_0, \tau_1, \mu, \sigma) + D$, $S_T = h(Z_2; S_{t_1^-} - D, \tau_2, \mu, \sigma)$ and Z_1 and Z_2 are two random variables with distribution functions $F_1(\cdot)$ and $F_2(\cdot)$ and density functions $f_1(\cdot)$ and $f_2(\cdot)$. To simplify notation, we will usually omit the explicit display of the dependence on μ and σ .

The PA estimator for $\partial E[J_T]/\partial\theta$, the derivation of which is included in the appendix, is given by

$$\left(\frac{\partial \hat{J}_T}{\partial \theta}\right)_{PA} = e^{-rT} \left[\left(\frac{\partial \hat{J}_T}{\partial \theta}\right)_{PA} - \hat{J}_T \frac{\partial}{\partial \theta}(rT) \right], \quad (8)$$

where

$$\begin{aligned} \left(\frac{\partial \hat{J}_T}{\partial \theta}\right)_{PA} &= \frac{\partial h^{-1}(y^*)}{\partial \theta} f_1(h^{-1}(y^*)) \left(E[J_T | S_{t_1^-} = s^-] - E[J_T | S_{t_1^-} = s^+] \right) \\ &\quad + \mathbf{1}\{S_{t_1^-} > s\} \frac{\partial}{\partial \theta} \left[(S_{t_1^-} - K)e^{r(T-t_1)} \right] + \mathbf{1}\{S_{t_1^-} \leq s\} \frac{\partial}{\partial \theta} (S_T - K)^+, \end{aligned} \quad (9)$$

where we have defined

$$E[J_T | S_{t_1^-} = s^-] = E[(S_T - K)^+ | S_{t_1^-} = s^-] = E[(h(Z_2; s - D, \tau_2) - K)^+], \quad (10)$$

$$E[J_T | S_{t_1^-} = s^+] = (s - K)e^{r(T-t_1)}, \quad (11)$$

$$y^* = (s - D; \tilde{S}_0, \tau_1). \quad (12)$$

Note that the first term ($J_T | S_{t_1^-} = s^-$) will also have to be estimated separately. Intuitively, the estimator resembles the form of the general estimator derived in Fu and Hu (1992) with $\frac{\partial h^{-1}(y^*)}{\partial \theta} f_1(h^{-1}(y^*))$ in (9) representing $\lim_{\Delta s \rightarrow 0} P(S_{t_1^-} \leq s + \Delta s, S_{t_1^-} > s) / \Delta s$.

In the appendix, we provide a proof of the following theorem on the unbiasedness of the estimators:

Theorem 2. Under the following assumptions, $E[J_T]$ is differentiable and the PA estimator given by (8)-(12) is unbiased for $\partial E[J_T] / \partial \theta$, $\theta \in \Theta$ for some open set Θ , i.e., $E[(\partial J_T / \partial \theta)_{PA}] = \partial E[J_T] / \partial \theta$.

A1. $h(\cdot)$ is continuously differentiable and is strictly increasing;

A2. On its support, $f_1(\cdot)$ is continuous and bounded, i.e., there exists $k > 0$ such that $f_1(\cdot) < k$;

A3.

$$E \left[\sup_{\theta \in \Theta} S_{t_1^-} \right] < \infty, E \left[\sup_{\theta \in \Theta} \left| \frac{\partial S_{t_1^-}}{\partial \theta} \right| \right] < \infty, E \left[\sup_{\theta \in \Theta} S_T \right] < \infty, E \left[\sup_{\theta \in \Theta} \left| \frac{\partial S_T}{\partial \theta} \right| \right] < \infty.$$

and

$$E \left[\sup_{\theta \in \Theta} S_T \middle| Z_2 \right] < \infty, E \left[\sup_{\theta \in \Theta} \left| \frac{\partial S_T}{\partial \theta} \right| \middle| Z_2 \right] < \infty \text{ a.s.}$$

Proof. The proof is provided in the appendix. The key is the usual one of showing an exchange of differentiation (limit) and expectation (integration) operators in the derivation of the estimators.

We now illustrate the estimators with two examples.

Example 1: $h(Z; S, t, r, \sigma) = S e^{(r - \sigma^2/2)t + \sigma \sqrt{t} Z}$ and $f_1(x) = f_2(x) = e^{-x^2/2} / \sqrt{2\pi}$, i.e., we assume the stock price follows the Black-Scholes log-normal distribution. The inverse is given by $h^{-1}(y; S, t, r, \sigma) = (\ln(y/S) - (r - \sigma^2/2)t) / (\sigma \sqrt{t})$, so we have

$$\begin{aligned} h^{-1}(y^*) &= \frac{1}{\sigma \sqrt{\tau_1}} \left(\ln \frac{s - D}{\tilde{S}_0} - (r - \sigma^2/2)\tau_1 \right); \\ \tilde{S}_0 &= S_0 - D e^{-r\tau_1}; \\ \tilde{S}_{t_1^-} &= \tilde{S}_0 e^{(r - \sigma^2/2)\tau_1 + \sigma \sqrt{\tau_1} Z_1}; \\ S_{t_1^-} &= \tilde{S}_{t_1^-} + D = \tilde{S}_0 e^{(r - \sigma^2/2)\tau_1 + \sigma \sqrt{\tau_1} Z_1} + D; \\ \tilde{S}_T &= S_T = (S_{t_1^-} - D) e^{(r - \sigma^2/2)\tau_2 + \sigma \sqrt{\tau_2} Z_2} = \tilde{S}_0 e^{(r - \sigma^2/2)T + \sigma \sqrt{\tau_1} Z_1 + \sigma \sqrt{\tau_2} Z_2}. \end{aligned}$$

Case 1. $\theta = K$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial K} &= 0; \\ \frac{\partial}{\partial K} [(S_{t_1^-} - K)e^{r\tau_2}] &= -e^{r\tau_2}; \\ \frac{\partial}{\partial K} (S_T - K)^+ &= -\mathbf{1}\{S_T > K\}.\end{aligned}$$

Case 2. $\theta = S_0$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial S_0} &= -\frac{1}{\tilde{S}_0 \sigma \sqrt{\tau_1}}; \\ \frac{\partial}{\partial S_0} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} \tilde{S}_{t_1^-} / \tilde{S}_0; \\ \frac{\partial}{\partial S_0} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} S_T / \tilde{S}_0.\end{aligned}$$

Case 3. $\theta = s$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial s} &= \frac{1}{(s - D)\sigma \sqrt{\tau_1}}; \\ \frac{\partial}{\partial s} [(S_{t_1^-} - K)e^{r\tau_2}] &= 0; \\ \frac{\partial}{\partial s} (S_T - K)^+ &= 0.\end{aligned}$$

Case 4. $\theta = r$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial r} &= -\frac{\sqrt{\tau_1}}{\sigma} (De^{-r\tau_1} / \tilde{S}_0 + 1); \\ \frac{\partial}{\partial r} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} [\tau_2(S_{t_1^-} - K) + \tau_1 \tilde{S}_{t_1^-} (1 + De^{-r\tau_1} / \tilde{S}_0)]; \\ \frac{\partial}{\partial r} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} S_T (T + \tau_1 De^{-r\tau_1} / \tilde{S}_0).\end{aligned}$$

Case 5. $\theta = \sigma$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial \sigma} &= -\frac{1}{\sigma^2 \sqrt{\tau_1}} \ln \frac{s - D}{\tilde{S}_0} + \frac{\sqrt{\tau_1}}{2} + \frac{r\sqrt{\tau_1}}{\sigma^2}; \\ \frac{\partial}{\partial \sigma} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} \tilde{S}_{t_1^-} (-\sigma\tau_1 + \sqrt{\tau_1} Z_1); \\ \frac{\partial}{\partial \sigma} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} (-\sigma T + \sqrt{\tau_1} Z_1 + \sqrt{\tau_2} Z_2) S_T.\end{aligned}$$

Case 6. $\theta = \tau_1$ (τ_2 fixed).

$$\frac{\partial h^{-1}(y^*)}{\partial \tau_1} = -\frac{1}{2\sigma\sqrt{\tau_1}} \left(\frac{1}{\tau_1} \ln \frac{s - D}{\tilde{S}_0} + 2rDe^{-r\tau_1} / \tilde{S}_0 + 2(r - \sigma^2/2) \right);$$

$$\begin{aligned}\frac{\partial}{\partial \tau_1} \left[(S_{t_1^-} - K)e^{r\tau_2} \right] &= e^{r\tau_2} \tilde{S}_{t_1^-} \left(r - \frac{\sigma^2}{2} + \frac{\sigma Z_1}{2\sqrt{\tau_1}} \right); \\ \frac{\partial}{\partial \tau_1} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} S_T \left(r - \frac{\sigma^2}{2} + \frac{\sigma Z_1}{2\sqrt{\tau_1}} + rDe^{-r\tau_1}/\tilde{S}_0 \right).\end{aligned}$$

Case 7. $\theta = \tau_2$ (τ_1 fixed).

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial \tau_2} &= 0; \\ \frac{\partial}{\partial \tau_2} \left[(S_{t_1^-} - K)e^{r\tau_2} \right] &= e^{r\tau_2} r (S_{t_1^-} - K); \\ \frac{\partial}{\partial \tau_2} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} S_T \left(r - \frac{\sigma^2}{2} + \frac{\sigma Z_2}{2\sqrt{\tau_2}} \right).\end{aligned}$$

Case 8. $\theta = D$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial D} &= \frac{1}{\sigma\sqrt{\tau_1}} \left(\frac{-1}{s-D} + e^{-r\tau_1}/\tilde{S}_0 \right); \\ \frac{\partial}{\partial D} \left[(S_{t_1^-} - K)e^{r\tau_2} \right] &= e^{r\tau_2} (-e^{-r\tau_1} \tilde{S}_{t_1^-}/\tilde{S}_0 + 1); \\ \frac{\partial}{\partial D} (S_T - K)^+ &= -\mathbf{1}\{S_T > K\} e^{-r\tau_1} S_T/\tilde{S}_0.\end{aligned}$$

Example 2: Every m days, $\tilde{S}_t = \tilde{S}_0 e^{rt} X$, where $X \sim U(1 - \sigma\sqrt{t}, 1 + \sigma\sqrt{t})$, where $\sigma\sqrt{m} < 1$, i.e., $h = h_i$ for $t \in [im, (i+1)m]$, where $h_i(Z; S, t, r, \sigma) = Se^{rt}(1 + \sigma\sqrt{t}Z)$, $Z \sim U(-1, 1)$, and $f_1(x) = f_2(x) = 0.5$ for $x \in (-1, 1)$. Note that this distribution is not memoryless (the process is non-Markovian), and that σ does not correspond precisely to the annualized variance rate, though it is a measure of dispersion. We can write h as

$$h(S, t, r, \sigma) = Se^{rt} \left[\prod_{j=1}^i (1 + \sigma\sqrt{m}Z_j) \right] (1 + \sigma\sqrt{t - im}Z_{i+1}) \text{ for } t \in [im, (i+1)m].$$

For simplicity, we will assume $\tau_1 = m$ and $\tau_2 < m$ (there is no difficulty in considering more general cases, aside from algebra), so we have

$$\begin{aligned}h^{-1}(y^*) &= \frac{1}{\sigma\sqrt{\tau_1}} \left(\frac{s-D}{\tilde{S}_0 e^{r\tau_1}} - 1 \right); \\ \tilde{S}_0 &= S_0 - De^{-r\tau_1}; \\ \tilde{S}_{t_1^-} &= \tilde{S}_0 e^{r\tau_1} (1 + \sigma\sqrt{\tau_1}Z_1); \\ S_{t_1^-} &= \tilde{S}_{t_1^-} + D = \tilde{S}_0 e^{r\tau_1} (1 + \sigma\sqrt{\tau_1}Z_1) + D; \\ \tilde{S}_T &= S_T = (S_{t_1^-} - D)e^{r\tau_2} (1 + \sigma\sqrt{T - \tau_1}Z_2) = \tilde{S}_0 e^{rT} (1 + \sigma\sqrt{\tau_1}Z_1)(1 + \sigma\sqrt{\tau_2}Z_2).\end{aligned}$$

Case 1. $\theta = K$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial K} &= 0; \\ \frac{\partial}{\partial K} [(S_{t_1^-} - K)e^{r\tau_2}] &= -e^{r\tau_2}; \\ \frac{\partial}{\partial K} (S_T - K)^+ &= -\mathbf{1}\{S_T > K\}.\end{aligned}$$

Case 2. $\theta = S_0$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial S_0} &= -\frac{s - D}{\tilde{S}_0^2 \sigma \sqrt{\tau_1} e^{r\tau_1}}; \\ \frac{\partial}{\partial S_0} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} \tilde{S}_{t_1^-} / \tilde{S}_0; \\ \frac{\partial}{\partial S_0} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} S_T / \tilde{S}_0.\end{aligned}$$

Case 3. $\theta = s$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial s} &= \frac{1}{\sigma \sqrt{\tau_1} \tilde{S}_0 e^{r\tau_1}}; \\ \frac{\partial}{\partial s} [(S_{t_1^-} - K)e^{r\tau_2}] &= 0; \\ \frac{\partial}{\partial s} (S_T - K)^+ &= 0.\end{aligned}$$

Case 4. $\theta = r$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial r} &= -\frac{(s - D)\sqrt{\tau_1}}{\sigma \tilde{S}_0 e^{r\tau_1}} (De^{-r\tau_1} / \tilde{S}_0 + 1); \\ \frac{\partial}{\partial r} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} [\tau_2(S_{t_1^-} - K) + \tau_1 \tilde{S}_{t_1^-} (1 + De^{-r\tau_1} / \tilde{S}_0)]; \\ \frac{\partial}{\partial r} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} S_T (T + \tau_1 De^{-r\tau_1} / \tilde{S}_0).\end{aligned}$$

Case 5. $\theta = \sigma$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial \sigma} &= -\frac{1}{\sigma^2 \sqrt{\tau_1}} \left(\frac{s - D}{\tilde{S}_0 e^{r\tau_1}} - 1 \right); \\ \frac{\partial}{\partial \sigma} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} (\tilde{S}_{t_1^-} - \tilde{S}_0 e^{r\tau_1}) / \sigma; \\ \frac{\partial}{\partial \sigma} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} \tilde{S}_0 e^{rT} (2\sigma \sqrt{\tau_1} Z_1 \sqrt{\tau_2} Z_2 + \sqrt{\tau_1} Z_1 + \sqrt{\tau_2} Z_2).\end{aligned}$$

Case 6. $\theta = \tau_1$ (τ_2 fixed).

$$\frac{\partial h^{-1}(y^*)}{\partial \tau_1} = -\frac{s - D}{2\sigma \sqrt{\tau_1} \tilde{S}_0 e^{r\tau_1}} \left(\left(1 - \frac{\tilde{S}_0 e^{r\tau_1}}{s - D} \right) / \tau_1 + 2r \left(1 + De^{-r\tau_1} / \tilde{S}_0 \right) \right);$$

$$\begin{aligned}\frac{\partial}{\partial \tau_1} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} \left(r\tilde{S}_{t_1^-} + \tilde{S}_0 e^{r\tau_1} \frac{\sigma}{2\sqrt{\tau_1}} Z_1 + rDe^{-r\tau_1} \tilde{S}_{t_1^-} / \tilde{S}_0 \right); \\ \frac{\partial}{\partial \tau_1} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} \left(rS_T + \tilde{S}_0 e^{rT} \frac{\sigma}{2\sqrt{\tau_1}} Z_1 (1 + \sigma\sqrt{\tau_2} Z_2) + rDe^{-r\tau_1} S_T / \tilde{S}_0 \right).\end{aligned}$$

Case 7. $\theta = \tau_2$ (τ_1 fixed).

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial \tau_2} &= 0; \\ \frac{\partial}{\partial \tau_2} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} r(S_{t_1^-} - K); \\ \frac{\partial}{\partial \tau_2} (S_T - K)^+ &= \mathbf{1}\{S_T > K\} \left(rS_T + \tilde{S}_0 e^{rT} (1 + \sigma\sqrt{\tau_1} Z_1) \frac{\sigma}{2\sqrt{\tau_2}} Z_2 \right).\end{aligned}$$

Case 8. $\theta = D$.

$$\begin{aligned}\frac{\partial h^{-1}(y^*)}{\partial D} &= \frac{s - D}{\sigma\sqrt{\tau_1}\tilde{S}_0 e^{r\tau_1}} \left(\frac{-1}{s - D} + e^{-r\tau_1} / \tilde{S}_0 \right); \\ \frac{\partial}{\partial D} [(S_{t_1^-} - K)e^{r\tau_2}] &= e^{r\tau_2} (-e^{-r\tau_1} \tilde{S}_{t_1^-} / \tilde{S}_0 + 1); \\ \frac{\partial}{\partial D} (S_T - K)^+ &= -\mathbf{1}\{S_T > K\} e^{-r\tau_1} S_T / \tilde{S}_0.\end{aligned}$$

Analysis for general $\eta(T) \geq 1$ carried out in Fu et al. (2000) leads to the PA estimator

$$\begin{aligned}\sum_{i=1}^{\eta(T)} \mathbf{1} \left\{ \bigcap_{j=1}^{i-1} S_{t_j^-} \leq s_j \right\} \frac{\partial h^{-1}(y_i^*)}{\partial \theta} f_i(h^{-1}(y_i^*)) &\left\{ E \left[L \left[\bigcap_{j=1}^{i-1} S_{t_j^-} \leq s_j, S_{t_i^-} = s_i^- \right] - (s_i - K)e^{-rt_i} \right\} \\ &+ \sum_{i=1}^{\eta(T)} \mathbf{1} \left\{ \bigcap_{j=1}^{i-1} S_{t_j^-} \leq s_j, S_{t_i^-} > s_i \right\} \frac{\partial}{\partial \theta} [(S_{t_i^-} - K)e^{-rt_i}] \\ &+ \mathbf{1}\{S_{t_1^-} \leq s_1, \dots, S_{t_{\eta(T)}^-} \leq s_{\eta(T)}\} \frac{\partial}{\partial \theta} [(S_T - K)^+ e^{-rT}],\end{aligned}$$

where
$$y_i^* = \left(s_i - \sum_{j=i}^{\eta(T)} D_j \exp \left(-r \sum_{k=i+1}^j \tau_k \right); \tilde{S}_{t_{i-1}}, \tau_i, \mu, \sigma \right), \quad i = 1, \dots, \eta(T).$$

Numerical Results

Numerical results for a large number of parameter settings — 486 cases for Example 1 and 162 cases for Example 2 — are reported here (from Fu and Hu 1993). We compare the PA estimators with the performance of finite difference estimates. The “precision” of the estimate will be represented by the standard error. For the finite difference estimate, there is obviously a choice that needs to be made with respect to the size of the difference. Larger differences lead to lower variance but higher bias, since you would be less likely to be estimating the gradient unless the performance curve is very linear. Exact results via the Geske-Roll-Whaley model (cf. Stoll and Whaley 1993 or Hull 1993) are also provided for Example 1, the lognormal case.

For Example 1, three different settings for each of K , S_0 , σ , D , and the (τ_1, τ_2) pair were chosen, and two settings for r , giving a total of $(3^5)(2) = 486$ cases:

$$\begin{aligned} K &= 40, 50, 60; \\ S_0 &= 40, 50, 60; \\ \sigma &= 0.1, 0.3, 0.5; \\ (\tau_1, \tau_2) &= (60, 30), (60, 90), (5, 30); \\ D &= 0.5, 1.0, 2.0; \\ r &= 0.05, 0.10. \end{aligned}$$

Example 2 used the identical settings, with the exception of using only the first choice for the (τ_1, τ_2) pair, giving a reduced total number of 162 cases. The value of s used was calculated via the Geske-Roll-Whaley model, so it will be optimal for Example 1, but not for Example 2; this however, just means that in Example 1, the estimate for $E[J_T]$ will estimate the option value, whereas in Example 2, the estimate will merely estimate the option payoff at the same early exercise threshold. In the next section, we will discuss a method that can be used to determine the optimal s values via simulation for any given distribution of stock prices.

For the finite difference estimation, we used a difference of 0.1 for the derivatives with respect to K and S_0 , and 0.001 for the derivatives with respect to r , σ , τ_1 , τ_2 , and D . Large differences were also tried in order to reduce the variance, but the bias was increase dramatically, and so these simulations are not reported here. The results are given in the form of a mean \pm standard error based on 36,100 simulation replications, with “PA” indicating the perturbation analysis estimates and “FD” indicating the finite difference estimates. Tables 3 through 56 give the results for Example 1 (lognormal), where exact results are also included, and Tables 57 through 74 give the results for

Example 1 (uniform). The minimal variation due to the one extra “off-line” simulation needed to estimate $E[J_T | S_{t_1}^- = s^-]$ is not included in the sample variance calculations. The precision can be further reduced by employing other variance reduction techniques such as control variates and antithetic variates, as pointed out by Boyle (1977). However, these methods could also reduce variance for PA, as well as the FD estimate, so the relative accuracy of the two should be unchanged.

For the most part, the perturbation analysis estimates did better than the finite difference estimates, and sometimes quite a bit better. In addition, in terms of raw computation time, the reduction was on the order of 5, versus a “best” reduction of 8, since 8 derivatives were estimated (the one with respect to s was omitted in this part, but are used in the next subsection). The results indicate that the sensitivity estimate with respect to the volatility is usually the noisiest, for both PA and FD.

3 Application to Option Pricing of American Calls

3.1 Formulation as An Optimization Problem

Now, for an American call, we incorporate the gradient estimate with respect to the early exercise threshold parameter into a stochastic approximation algorithm (see, e.g., Fu 1994) to determine the optimal setting of this parameter, where the “best guess” of the optimal setting is updated iteratively based on the gradient estimate. Let $g(\theta) = \nabla_{\theta} E[J_T(\theta)]$, where ∇_{θ} denotes the gradient with respect to θ , which in general could be a vector of parameters. In our application, J_T is the return of the option as a function of θ , which is the scalar early exercise threshold parameter s for the single ex-dividend case, and a vector when there are multiple ex-dividend points. The option value can then be defined as the point at which $E[J_T(\theta)]$ is maximized with respect to θ . The basic underlying assumption of the stochastic approximation algorithm is that the original problem can be solved by finding the zero of the gradient, i.e., by finding θ_* , the optimal exercise threshold level, such that $g(\theta_*) = 0$. Of course, in practice, this may lead only to local optimality. Formulation of American call options in this manner was also pointed out by Welch and Chen (1991) for the application of analytical techniques.

Since the problem is a maximization problem, the stochastic approximation algorithm takes the following form:

$$\theta^{(n+1)} = \Pi_{\Theta} \left(\theta^{(n)} + a_n \hat{g}_n \right), \quad (13)$$

where $\theta^{(n)}$ is the parameter setting at the beginning of iteration n , \hat{g}_n is an estimate of $g(\theta^{(n)})$ from iteration n , a_n is a (positive) sequence of step sizes, and Π_{Θ} is a projection onto some set Θ , e.g., the positive real numbers \mathfrak{R}^+ . For our example, we have the unbiased estimate defined by (8)-(12):

$$\hat{g}_n = e^{-rT} \frac{\partial h^{-1}(y^*)}{\partial \theta} f(h^{-1}(y^*)) \left(E[J_T | S_{t_1^-} = s^-] - E[J_T | S_{t_1^-} = s^+] \right), \quad (14)$$

where the last two terms in (9) are zero, since the underlying stock process is independent of the exercise threshold (Case 3 in either of the two examples).

To guarantee a.s. convergence of the algorithm to the optimum, certain conditions on the nature of the noise in the gradient estimate, the sequence of step sizes, and the uniqueness of the optimum are required. For example, the results in Kushner and Clark (1978) yield the following convergence result:

Theorem 3. Assume that $g(\theta)$ is continuous w.r.t. θ , and that

$$\sum a_n = \infty, \quad \sum a_n^2 < \infty, \quad E[\hat{g}] = dE[g(\theta)]/ds, \quad E[\hat{g}^2] \text{ bounded, on } \theta \in \Theta.$$

Then, if $g(\theta)$ has a unique zero $\theta_* \in \Theta$ s.t. $g(\theta) > 0 \forall \theta < \theta_*$, then $\theta^{(n)} \rightarrow \theta_*$ w.p.1 for the projection algorithm.

The conditions on the step sizes are satisfied, for example, by the harmonic series $a_n = a/n$ (for some constant a). In the harmonic series sequence of step sizes, a decrease is taken at every iteration. In practice, the harmonic series often leads to rather slow convergence. Modifications such as decreasing the step size only if the gradient direction has changed from the previous iteration work better in practice; this is the so-called accelerated harmonic series. We also note that the best achievable asymptotic convergence rate for the algorithm – obtained when an unbiased estimate is used – is $n^{-1/2}$ (cf. Kushner and Clark 1978), the same (slow) rate as when Monte Carlo simulation is used for estimation only.

3.2 Numerical Examples

The main considerations in applying the algorithm in practice are the following:

- Getting a gradient estimate \hat{g} .
- Choosing a step size a_n .
- Choosing an observation length for each iteration.
- Choosing a starting point for the algorithm.
- Choosing a stopping rule for the algorithm.

For our example, we took $\Theta = [K, 5K]$, with the value $\theta = 5K$ essentially meaning the American option is equivalent to the European option for most realistic values of the parameters. For the step sizes, we used the so-called accelerated harmonic series described before, with $a = 100$. We took $\theta_0 = K$, i.e., the initial exercise threshold level was simply set to the striking prices. We considered fixed observation lengths of both 10 and 100. Increasing the observation length gradually is another option that we did not implement. Because we just wanted to get an idea as to the behavior and performance of the algorithm, we did not implement a stopping rule, but instead investigated the improvement of the option value at various points in the iteration.

We ran three different cases corresponding to varying the dividend amounts $D = 0.5, 1.0, 2.0$, and keeping the other parameters at the following values:

$$K = S_0 = 50, \quad r = 0.10, \quad \sigma = 0.30, \quad (\tau_1, \tau_2) = (60, 30).$$

We used the lognormal distribution of Example 1 so we could track the performance of our algorithm by calculating the objective function (expected option payoff, e.g., via Stoll and Whaley 1993) as the values of s was changed according to the gradient algorithm. We would expect the curve for $D = 0.5$ to be relatively flat with respect to the exercise threshold, since a lower cash dividend makes the behavior closer to a European-type option, whereas $D = 2.0$ should show decidedly more peakedness. In Figure 1, this behavior is borne out, where the expected option payoff as a function of the early exercise threshold is plotted for the three dividend values. Of course, this information was not used in the algorithm, since it would not be available for actual cases where simulation was needed. We ran the algorithm for 9 different replications to get the average performance. A detailed graphical representation of the improvement in estimating the option call value is shown in Figures 2 through 7, which depicts the expected option payoff (on the vertical axis) at the value of the exercise threshold after a given number of simulations (on the horizontal axis) was used. Thus, the point 10 on the horizontal axis (in 100s of simulations) corresponds to 100 iterations for the case where the observation length is 10, and only 10 iterations for the case where the observation length is 100. The actual option value is represented by the horizontal line, and the mean of the objective function over the 9 replications is shown by the thicker plot, with the thinner plots above and below bracketing the mean by the standard error. There are six figures, since there are two observation lengths used for each of the three cases.

The results show that the algorithm converges quite quickly. In all the figures except Figure 3, convergence to within a penny of the actual option value is achieved within 1000 simulations, and in Figure 3, it is achieved within 7000 simulations. Compared with the numerical results reported in Fu and Hu (1993), this is much less effort than is needed to simply estimate an option payoff to within a penny. In other words, the additional effort needed to estimate an American option using Monte Carlo simulation over what is needed to estimate a European option in this case is negligible. Finally, we reiterate that Example 1 was used just so the performance of the algorithm could be easily tracked. As far as the algorithm itself is concerned, the intractable Example 2 could just have easily been used.

References

- [1] Boyle, P.P., "Options: A Monte Carlo Approach," *Journal of Financial Economics*, **4**, 323-338, 1977.
- [2] Broadie, M. and Glasserman, P., "Estimating Security Price Derivatives Using Simulation," *Management Science*, **42**, 269-285, 1996.
- [3] Cox, J.C. and Ross, S.A., "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics*, **3**, 145-166, 1976.
- [4] Figlewski, S., "Options Arbitrage in Imperfect Markets," *J. Finance*, **XLIV**, 1289-1311, 1989.
- [5] Fu, M. C., "Optimization via Simulation: A Review," *Annals of Operations Research*, Vol. 53, pp.199-248, 1994.
- [6] Fu, M.C. and Hu, J.Q., "Extensions and Generalizations of Smoothed Perturbation Analysis in a Generalized Semi-Markov Process Framework," *IEEE Transactions Automatic Control*, **37**, pp.1483-1500, 1992.
- [7] Fu, M.C. and Hu, J.Q., "Sensitivity Analysis for Monte Carlo Simulation of Option Pricing," Working Paper, University of Maryland, College of Business and Management, August 1993.
- [8] Fu, M.C., Wu, R., Gürkan, G., and Demir, A.Y., "A Note on Perturbation Analysis Estimators for American-Style Options," *Prob. in the Eng. Inf. Sciences*, **14**, 385-392, 2000.
- [9] Glasserman, P., *Gradient Estimation Via Perturbation Analysis*, Kluwer Academic, 1991.
- [10] Gong, W.B. and Ho, Y.C., "Smoothed Perturbation Analysis of Discrete-Event Dynamic Systems," *IEEE Transactions on Automatic Control*, **32**, pp.858-867, 1987.
- [11] Grant, D., Vora, G., and Weeks, D., "Path-Dependent Options: Extending the Monte Carlo Simulation Approach," *Management Science*, **43**, 1589-1602, 1997.
- [12] Harrison, J.M. and Pliska, S., "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and their Applications*, **11**, pp.215-260, 1981.
- [13] Ho, Y.C. and Cao, X.R., *Discrete Event Dynamic Systems and Perturbation Analysis*, Kluwer Academic, 1991.
- [14] Hull, J.C., *Options, Futures, and Other Derivative Securities*, 2nd edition, Prentice Hall, 1993.
- [15] Hull, J.C. and White, A., "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, **42**, 281-300, 1987.
- [16] Johnson, H. and Shanno, D., "Option Pricing When the Variance Is Changing," *Journal of Financial and Quantitative Analysis*, **22**, 143-151, 1987.
- [17] Kushner, H.J. and Clark, D.C., *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, New York, 1978.
- [18] Rubinstein, R.Y. and Shapiro, A., *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, John Wiley & Sons, 1993.
- [19] Scott, L.O., "Option Pricing When the Variance Changes Randomly: Theory, Estimation, and An Application," *Journal of Financial and Quantitative Analysis*, **22**, 419-438, 1987.
- [20] Stoll, H.R. and Whaley, R.E., *Futures and Options*, South-Western, 1993.
- [21] Tilley, J., "Valuing American Options in a Path Simulation Model," Morgan Stanley working paper; also, *Transactions of the Society of Actuaries*, **45**, pp.83-104, 1993.
- [22] Welch, R.L. and Chen, D.M., "Static Optimization of American Contingent Claims," *Advances in Futures and Options Research*, **5**, 175-184, 1991.

Appendix: Proof of Theorem 2

In this section, we provide the proof of Theorem 2 by way of deriving the PA estimator for $\partial E[J_T]/\partial\theta$, given by Equation (8). First we will do the derivation assuming certain exchanges of operators are valid, which we will prove by using the conditions of the theorem statement. Recall that to simplify notation, we will usually omit the explicit display of the dependence on μ and σ .

Taking the expectation of the first term of \hat{J}_T , given by Equation (7), we have

$$\begin{aligned}
& E \left[\mathbf{1}\{S_{t_1^-} > s\} (S_{t_1^-} - K) e^{r\tau_2} \right] \\
&= E \left[\mathbf{1}\{h(Z_1; \tilde{S}_0, \tau_1) + D > s\} (h(Z_1; \tilde{S}_0, \tau_1) + D - K) e^{r\tau_2} \right] \\
&= E \left[\mathbf{1}\{Z_1 > h^{-1}(s - D; \tilde{S}_0, \tau_1)\} (h(Z_1; \tilde{S}_0, \tau_1) + D - K) e^{r\tau_2} \right] \\
&= \int_{h^{-1}(y^*)}^{\infty} (h(x; \tilde{S}_0, \tau_1) + D - K) e^{r\tau_2} f_1(x) dx,
\end{aligned}$$

where we have defined $y^* = (s - D; \tilde{S}_0, \tau_1)$. Differentiating, we have

$$\begin{aligned}
& \frac{\partial}{\partial\theta} E \left[\mathbf{1}\{S_{t_1^-} > s\} (S_{t_1^-} - K) e^{r\tau_2} \right] \\
&= -\frac{\partial h^{-1}(y^*)}{\partial\theta} (h(h^{-1}(y^*); \tilde{S}_0, \tau_1) + D - K) e^{r\tau_2} f_1(h^{-1}(y^*)) \\
&\quad + \int_{h^{-1}(y^*)}^{\infty} \frac{\partial}{\partial\theta} \left((h(x; \tilde{S}_0, \tau_1) + D - K) e^{r\tau_2} \right) f_1(x) dx \tag{15}
\end{aligned}$$

$$= -\frac{\partial h^{-1}(y^*)}{\partial\theta} (s - K) e^{r\tau_2} f_1(h^{-1}(y^*)) + E \left[\mathbf{1}\{S_{t_1^-} > s\} \frac{\partial}{\partial\theta} \left((S_{t_1^-} - K) e^{r\tau_2} \right) \right]. \tag{16}$$

On the other hand,

$$\begin{aligned}
& E \left[\mathbf{1}\{S_{t_1^-} \leq s\} (S_T - K)^+ \right] \\
&= E \left[\mathbf{1}\{S_{t_1^-} \leq s\} (h(Z_2; S_{t_1^-} - D, \tau_2) - K)^+ \right] \\
&= E \left[\int_{-\infty}^{h^{-1}(y^*)} (h(Z_2; h(x; \tilde{S}_0, \tau_1), \tau_2) - K)^+ f_1(x) dx \right],
\end{aligned}$$

so assuming an interchange of differentiation and expectation, we have

$$\begin{aligned}
& \frac{\partial}{\partial\theta} E \left[\mathbf{1}\{S_{t_1^-} \leq s\} (S_T - K)^+ \right] \\
&= E \left[\frac{\partial}{\partial\theta} \left(\int_{-\infty}^{h^{-1}(y^*)} (h(Z_2; h(x; \tilde{S}_0, \tau_1), \tau_2) - K)^+ f_1(x) dx \right) \right] \tag{17} \\
&= E \left[\frac{\partial h^{-1}(y^*)}{\partial\theta} (h(Z_2; h(h^{-1}(y^*); \tilde{S}_0, \tau_1), \tau_2) - K)^+ f_1(h^{-1}(y^*)) \right]
\end{aligned}$$

$$+ \int_{-\infty}^{h^{-1}(y^*)} \frac{\partial}{\partial \theta} \left((h(Z_2; h(x; \tilde{S}_0, \tau_1), \tau_2) - K)^+ f_1(x) \right) dx \quad (18)$$

$$= E \left[\frac{\partial h^{-1}(y^*)}{\partial \theta} (h(Z_2; s - D, \tau_2) - K)^+ f_1(h^{-1}(y^*)) + \mathbf{1}\{S_{t_1^-} \leq s\} \frac{\partial}{\partial \theta} (S_T - K)^+ \right]. \quad (19)$$

Combining (16) and (19), we have

$$\begin{aligned} \frac{\partial E[\hat{J}_T]}{\partial \theta} &= \frac{\partial h^{-1}(y^*)}{\partial \theta} f_1(h^{-1}(y^*)) E \left[(h(Z_2; s - D, \tau_2) - K)^+ - (s - K) e^{r\tau_2} \right] \\ &\quad + E \left[\mathbf{1}\{S_{t_1^-} > s\} \frac{\partial}{\partial \theta} \left((S_{t_1^-} - K) e^{r\tau_2} \right) \right] + E \left[\mathbf{1}\{S_{t_1^-} \leq s\} \frac{\partial}{\partial \theta} (S_T - K)^+ \right], \end{aligned}$$

resulting in the PA estimator given by Equations (8)-(12).

Now we proceed with the actual proof. As we said earlier, the key is to show that we can exchange the differentiation (limit) and expectation (integration) operators in deriving (15), (17) and (18). Let us first consider (15). Based on (A1) we know that $h^{-1}(y^*)$ exists, $(h(x; \tilde{S}_0, \tau_1) - K) e^{r\tau_2} f_1(x)$ is continuous with respect to x and $(h(x; \tilde{S}_0, \tau_1) - K) e^{r\tau_2} f_1(x)$ and $h^{-1}(y^*)$ are continuously differentiable with respect to θ . On the other hand we have based on (A3) that

$$\begin{aligned} &\int_{h^{-1}(y^*)}^{\infty} \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} \left(h(x; \tilde{S}_0, \tau_1) + D - K \right) e^{r\tau_2} \right| f_1(x) dx \\ &= E \left[\mathbf{1}\{S_{t_1^-} > s\} \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} \left((S_{t_1^-} - K) e^{r\tau_2} \right) \right| \right] \\ &\leq E \left[\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} \left((S_{t_1^-} - K) e^{r\tau_2} \right) \right| \right] < \infty. \end{aligned}$$

Hence using the chain rule (here as Leibniz's rule for differentiating an integral), and the dominated convergence theorem, we obtain

$$\begin{aligned} &\frac{\partial}{\partial \theta} \int_{h^{-1}(y^*)}^{\infty} (h(x; \tilde{S}_0, \tau_1) + D - K) e^{r\tau_2} f_1(x) dx \\ &= - \frac{\partial h^{-1}(y^*)}{\partial \theta} (h(h^{-1}(y^*); \tilde{S}_0, \tau_1) + D - K) e^{r\tau_2} f_1(h^{-1}(y^*)) \\ &\quad + \int_{h^{-1}(y^*)}^{\infty} \frac{\partial}{\partial \theta} \left((h(x; \tilde{S}_0, \tau_1) + D - K) e^{r\tau_2} \right) f_1(x) dx \\ &= - \frac{\partial h^{-1}(y^*)}{\partial \theta} (s - K) e^{r\tau_2} f_1(h^{-1}(y^*)) + E \left[\mathbf{1}\{S_{t_1^-} > s\} \frac{\partial}{\partial \theta} \left((S_{t_1^-} - K) e^{r\tau_2} \right) \right]. \end{aligned}$$

This verifies (15). We now turn to verification of (18). We note that

$$(h(Z_2; h(x; \tilde{S}_0, \tau_1) - D, \tau_2) - K)^+ f_1(x)$$

is a.s. continuous with respect to x , a.s. continuous and piecewise differentiable over Θ , and a.s. differentiable at every $\theta \in \Theta$; furthermore a.s.

$$\begin{aligned}
& \int_{-\infty}^{h^{-1}(y^*)} \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} \left((h(Z_2; h(x; \tilde{S}_0) - D, \tau_2) - K)^+ \right) \right| f_1(x) dx \\
&= E \left[\sup_{\theta \in \Theta} \left| \mathbf{1}\{S_{t_1^-} \leq s\} \frac{\partial}{\partial \theta} (S_T - K)^+ \right| \middle| Z_2 \right] \\
&\leq E \left[\sup_{\theta \in \Theta} \left| \frac{\partial S_T}{\partial \theta} \right| \middle| Z_2 \right] + \sup_{\theta \in \Theta} \left| \frac{\partial K}{\partial \theta} \right| \\
&< \infty
\end{aligned}$$

the last inequality following from (A3). Repeating the argument we used for verifying (15), we can easily show that (18) is correct.

Finally, we verify (17). Since

$$\int_{-\infty}^{h^{-1}(y^*)} (h(Z_2; h(x; \tilde{S}_0, \tau_1) - D, \tau_2) - K)^+ f_1(x) dx$$

is a.s. continuous and piecewise differentiable with respect to θ , based on the dominated convergence theorem it suffices for us to show that

$$E \left[\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} \left(\int_{-\infty}^{h^{-1}(y^*)} (h(Z_2; h(x; \tilde{S}_0, \tau_1) - D, \tau_2) - K)^+ f_1(x) dx \right) \right| \right] < \infty.$$

On the other hand, based on (18) we know that

$$\begin{aligned}
& \frac{\partial}{\partial \theta} \left(\int_{-\infty}^{h^{-1}(y^*)} (h(Z_2; h(x; \tilde{S}_0, \tau_1) - D, \tau_2) - K)^+ f_1(x) dx \right) \\
&= (h(Z_2; s - D, \tau_2) - K)^+ f_1(h^{-1}(y^*)) + \mathbf{1}\{S_{t_1^-} \leq s\} \frac{\partial}{\partial \theta} (S_T - K)^+ \quad \text{a.s.},
\end{aligned}$$

hence we only need to show

$$E \left[\sup_{\theta \in \Theta} \left| (h(Z_2; s - D, \tau_2) - K)^+ f_1(h^{-1}(y^*)) + \mathbf{1}\{S_{t_1^-} \leq s\} \frac{\partial}{\partial \theta} (S_T - K)^+ \right| \right] < \infty,$$

which immediately follows from (A2) and (A3), completing our proof. \square