
Vibrato Monte Carlo sensitivities

M.B. Giles

Oxford-Man Institute of Quantitative Finance
Oxford University Mathematical Institute
24–29 St. Giles, Oxford, U.K. OX1 3LB
<http://people.maths.ox.ac.uk/~gilesm/>

Summary. We show how the benefits of the pathwise sensitivity approach to computing Monte Carlo Greeks can be extended to discontinuous payoff functions through a combination of the pathwise approach and the Likelihood Ratio Method. With a variance reduction modification, this results in an estimator which for timestep h has a variance which is $O(h^{-1/2})$ for discontinuous payoffs and $O(1)$ for continuous payoffs. Numerical results confirm the variance is much lower than the $O(h^{-1})$ variance of the Likelihood Ratio Method, and the approach is also compatible with the use of adjoints to obtain multiple first order sensitivities at a fixed cost.

1 Introduction

Monte Carlo simulation is the most popular approach in computational finance for determining the prices of financial options. This is partly due to its computational efficiency for high-dimensional problems involving multiple assets, interest rates or exchange rates, and partly due to its relative simplicity and the ease with which it can be parallelised across large compute clusters. However, the accurate calculation of prices is only one objective of Monte Carlo simulation. Even more important in some ways is the calculation of the sensitivities of the prices to various input parameters. These sensitivities, known collectively as the “Greeks”, are important for risk analysis and mitigation through hedging.

The pathwise sensitivity approach (also known as Infinitesimal Perturbation Analysis) is one of the standard techniques for computing these sensitivities [12]. Giles and Glasserman have recently introduced a particularly efficient implementation of this approach using adjoint techniques [11] which are related to the use of reverse mode automatic differentiation [13, 9]. This makes it possible to calculate an unlimited number of first order sensitivities at a total cost which is comparable to the cost of the original pricing calculation.

However, the pathwise approach is not applicable when the financial payoff function is discontinuous, and even when the payoff is continuous and piece-

wise differentiable, the use of scripting languages in real-world implementations means it can be very difficult in practice to evaluate the derivative of very complex financial products. One solution to these problems is to use the Likelihood Ratio Method (LRM) but its weaknesses are that the variance of the resulting estimator is usually $O(h^{-1})$, where h is the timestep for the path discretisation, and it can not be combined efficiently with the adjoint approach.

Building on the ideas of l'Ecuyer on hybrid pathwise/LRM sensitivity calculations [15, 16], this paper presents a new idea which combines the adjoint pathwise approach for the stochastic path evolution with LRM for the payoff evaluation. Through the use of a control variate for variance reduction, the variance of the resulting estimator is $O(h^{-1/2})$ when the payoff is discontinuous, and $O(1)$ when it is continuous. Numerical examples show it is very much more efficient than the standard LRM approach.

2 Pathwise and LRM sensitivities

Consider the approximate solution of the general SDE driven by Brownian motion,

$$dS_t = a(S, t) dt + b(S, t) dW_t, \quad (1)$$

using the Euler discretisation,

$$\widehat{S}_{n+1} = \widehat{S}_n + a(\widehat{S}_n, t_n) h + b(\widehat{S}_n, t_n) \Delta W_n. \quad (2)$$

The Brownian increments ΔW_n can be defined to be a linear transformation of a vector of independent unit Normal random variables Z .

The goal now is to efficiently estimate the expected value of some financial payoff function $f(S)$, and numerous first order sensitivities of this value with respect to different input parameters such as the volatility or one component the initial conditions $S(0)$. In the simplest cases, $f(S)$ is a function of the value of the underlying solution $S(T)$ at the final time T , but in more general cases it might depend on the values at intermediate times as well.

The pathwise sensitivity approach can be viewed as expressing the expectation as an integral with respect to Z :

$$\widehat{V} \equiv \mathbb{E} [f(\widehat{S})] = \int f(\widehat{S}(Z, \theta)) p_Z(Z) dZ. \quad (3)$$

Here θ represents a generic input parameter, and the probability density function for Z is

$$p_Z(Z) = (2\pi)^{-d/2} \exp(-|Z|^2/2),$$

where d is the dimension of the vector Z .

If the drift, volatility and payoff functions are all differentiable, Eq. (3) may be differentiated to give

$$\frac{\partial \widehat{V}}{\partial \theta} = \int \frac{\partial f}{\partial \widehat{S}} \frac{\partial \widehat{S}}{\partial \theta} p_Z(Z) dZ, \quad (4)$$

with $\frac{\partial \widehat{S}}{\partial \theta}$ being obtained by differentiating Eq. (2) to obtain

$$\frac{\partial \widehat{S}_{n+1}}{\partial \theta} = \frac{\partial \widehat{S}_n}{\partial \theta} + \left(\frac{\partial a_n}{\partial \widehat{S}_n} \frac{\partial \widehat{S}_n}{\partial \theta} + \frac{\partial a_n}{\partial \theta} \right) h + \left(\frac{\partial b_n}{\partial \widehat{S}_n} \frac{\partial \widehat{S}_n}{\partial \theta} + \frac{\partial b_n}{\partial \theta} \right) \Delta W_n. \quad (5)$$

By considering the limit of a sequence of regularised functions, it can be proved that Eq.(4) remains valid when the payoff function is continuous and piecewise differentiable, and the numerical estimate obtained by averaging over M independent path simulations

$$M^{-1} \sum_{m=1}^M \frac{\partial f}{\partial \widehat{S}}(\widehat{S}^{(m)}) \frac{\partial \widehat{S}^{(m)}}{\partial \theta}$$

is an unbiased estimate for $\partial \widehat{V} / \partial \theta$ with a variance which is $O(M^{-1})$, independent of h , if $f(S)$ is Lipschitz and the drift and volatility functions satisfy the standard conditions [14].

Performing a change of variables, the expectation can also be expressed as

$$V \equiv \mathbb{E} [f(\widehat{S})] = \int f(\widehat{S}) p_S(\widehat{S}, \theta) d\widehat{S}, \quad (6)$$

where $p_S(\widehat{S}, \theta)$ is the probability density function for \widehat{S} which will depend on all of the inputs parameters.. If this is known, Eq.(6) can be differentiated to give

$$\frac{\partial \widehat{V}}{\partial \theta} = \int f \frac{\partial p_S}{\partial \theta} d\widehat{S} = \int f \frac{\partial (\log p_S)}{\partial \theta} p_S d\widehat{S} = \mathbb{E} \left[f \frac{\partial (\log p_S)}{\partial \theta} \right].$$

which can be estimated using the unbiased estimator

$$M^{-1} \sum_{m=1}^M f(\widehat{S}^{(m)}) \frac{\partial \log p_S(\widehat{S}^{(m)})}{\partial \theta}$$

This is the Likelihood Ratio Method. Its great advantage is that it does not require the differentiation of $f(\widehat{S})$. This makes it applicable to cases in which the payoff is discontinuous, and it also simplifies the practical implementation because banks often have complicated flexible procedures through which traders specify payoffs. However, it does have a number of limitations.

One limitation is that it requires that $\partial S / \partial W$ is an invertible matrix. This condition, known as absolute continuity, is not satisfied in a few important applications such as the LIBOR market model [12]. Other drawbacks of LRM are that in most cases it gives an estimator with a variance which is $O(M^{-1}h^{-1})$, becoming infinite as $h \rightarrow 0$ [12], and there is no way to efficiently incorporate adjoint techniques and hence the computational cost is proportional to the number of first order sensitivities which are needed.

3 Vibrato Monte Carlo

We now introduce a hybrid combination of pathwise and LRM sensitivity calculation, applying the pathwise approach to the differentiable path simulation, and using LRM for the discontinuous payoff evaluation. The idea of combining pathwise and LRM approaches is not new. L'Ecuyer [15, 16] presented a general framework in which the two approaches are just special cases of a more general estimator, and Chen and Glasserman [3] have recently shown that the use of Malliavin calculus [6, 7] can also be viewed as a hybrid pathwise/LRM combination.

The novelty in the present paper lies in the precise form of the hybrid combination, and the variance reduction which is achieved, making it a very practical method for finance applications with a discontinuous payoff function.

3.1 Conditional expectation

The Oxford English Dictionary describes “vibrato” as “a rapid slight variation in pitch in singing or playing some musical instruments”. The analogy to Monte Carlo methods is the following: whereas a path simulation in a standard Monte Carlo calculation produces a precise value for the output values from the underlying stochastic process, in the vibrato Monte Carlo approach the output values have a narrow probability distribution.

This is a generalisation of the technique of conditional expectation discussed by Glasserman in section 7.2.3 of his book [12] as a solution to the problem of discontinuous payoffs. In his example, a path simulation for a scalar SDE is performed in the usual way for the first $N-1$ timesteps, at each timestep taking a value for the Wiener increment ΔW^n which is a sample from the appropriate Gaussian distribution, and then using (2) to update the solution. On the final timestep, one instead considers the full distribution of possible values for ΔW^N . This gives a Gaussian distribution for \widehat{S}_N at time T , with probability density function

$$p_S(\widehat{S}_N) = \frac{1}{\sqrt{2\pi} \sigma_W} \exp\left(-\frac{(\widehat{S}_N - \mu_W)^2}{2\sigma_W^2}\right) \quad (7)$$

where

$$\mu_W = \widehat{S}_{N-1} + a(\widehat{S}_{N-1}, T-h) h, \quad \sigma_W = b(\widehat{S}_{N-1}, T-h) \sqrt{h},$$

with $a(S, t)$ and $b(S, t)$ being the drift and volatility of the SDE described in (1). Hence, the conditional expectation for the value of a digital payoff with strike K ,

$$f(S(T)) = H(S(T) - K) \equiv \begin{cases} 1, & S(T) > K \\ 0, & S(T) \leq K \end{cases}$$

is

$$\mathbb{E}[f(\widehat{S}_N) | \widehat{S}_{N-1}] = \int_{-\infty}^{\infty} H(\widehat{S}_N - K) p_S(\widehat{S}_N) d\widehat{S}_N = \Phi\left(\frac{\mu_W - K}{\sigma_W}\right)$$

where $\Phi(\cdot)$ is the cumulative Normal distribution function. A Monte Carlo estimator for the option value is therefore

$$M^{-1} \sum_m \mathbb{E}[f(\widehat{S}_N) | \widehat{S}_{N-1}^{(m)}] \equiv M^{-1} \sum_m \Phi\left(\frac{\mu_W^{(m)} - K}{\sigma_W^{(m)}}\right)$$

and because the conditional expectation $\mathbb{E}[f(\widehat{S}_N) | \widehat{S}_{N-1}]$ is a differentiable function of the input parameters the pathwise sensitivity approach can now be applied.

There are two difficulties in using this form of conditional expectation in real-world financial applications. This first is that the integral arising from the conditional expectation will often become a multi-dimensional integral without an obvious closed-form value (e.g. consider a digital option based on the median of a basket of 20 stocks), and the second is that it requires a change to the often complex software framework used to specify payoffs.

The solution is to use a Monte Carlo estimate of the conditional expectation, and use LRM to obtain its sensitivity. Thus, the technique combines pathwise sensitivity for the path calculation with LRM sensitivity for the payoff evaluation.

3.2 Vibrato Monte Carlo

The idea is very simple; adopting the conditional expectation approach, each path simulation for a particular discrete set of Wiener increments W computes a Gaussian conditional probability distribution $p_S(\widehat{S}_N | W)$ for the state of the path approximation at time T . For a scalar SDE, if μ_W and σ_W are the mean and standard deviation for given W , then

$$\widehat{S}_N(W, Z) = \mu_W + \sigma_W Z,$$

where Z is a unit Normal random variable. The expected payoff can then be expressed as

$$V = \mathbb{E}_W \left[\mathbb{E}_Z [f(\widehat{S}_N) | W] \right] = \int \left\{ \int f(\widehat{S}_N) p_S(\widehat{S}_N | W) d\widehat{S}_N \right\} p_W(W) dW.$$

The outer expectation/integral is an average over the different values for the discrete Wiener increments, while the inner conditional expectation/integral is averaging over Z .

To compute the sensitivity to the input parameter θ , the first step is to apply the pathwise sensitivity approach for fixed W to obtain $\partial\mu_W/\partial\theta$, $\partial\sigma_W/\partial\theta$. We then apply LRM to the inner conditional expectation to get

$$\frac{\partial \widehat{V}}{\partial \theta} = \mathbb{E}_W \left[\frac{\partial}{\partial \theta} \mathbb{E}_Z \left[f(\widehat{S}_N) \mid W \right] \right] = \mathbb{E}_W \left[\mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \theta} \mid W \right] \right],$$

where p_S is defined in (7) and

$$\frac{\partial(\log p_S)}{\partial \theta} = \frac{\partial(\log p_S)}{\partial \mu_W} \frac{\partial \mu_W}{\partial \theta} + \frac{\partial(\log p_S)}{\partial \sigma_W} \frac{\partial \sigma_W}{\partial \theta}.$$

The Monte Carlo estimators for \widehat{V} and $\partial \widehat{V} / \partial \theta$ have the form

$$M^{-1} \sum_{m=1}^M \widehat{Y}^{(m)}, \quad M^{-1} \sum_{m=1}^M \widehat{Y}_\theta^{(m)},$$

where $\widehat{Y}^{(m)}$ is an unbiased estimator for $\mathbb{E}_Z \left[f(\widehat{S}_N) \mid W \right]$ and $\widehat{Y}_\theta^{(m)}$ is an unbiased estimator for $\mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \theta} \mid W \right]$.

Although the discussion so far has had an option based on the value of a single output at the terminal time T , it will be shown later that the idea extends very naturally to multiple outputs at the final time, producing a multivariate Gaussian distribution, and to payoffs which are also dependent on values at intermediate times.

3.3 Efficient estimators

It is important that $\widehat{Y}^{(m)}$ and $\widehat{Y}_\theta^{(m)}$ have low variance to minimise the number of path simulations which must be performed to achieve a given accuracy. For a particular path, using the identity

$$\mathbb{E}_Z[h(Z)] = \mathbb{E}_Z[h(-Z)]$$

for all functions $h(Z)$ due to the symmetry of the probability distribution for Z , it follows that

$$\begin{aligned} \mathbb{E}_Z \left[f(\widehat{S}_N) \right] &= f(\mu_W) + \mathbb{E}_Z \left[f(\mu_W + \sigma_W Z) - f(\mu_W) \right] \\ &= f(\mu_W) + \mathbb{E}_Z \left[\frac{1}{2} \left(f(\mu_W + \sigma_W Z) - 2f(\mu_W) + f(\mu_W - \sigma_W Z) \right) \right]. \end{aligned}$$

If $f(S)$ is differentiable, the second term is the expectation of a quantity which has a very small mean and variance, and so a single Z sample will be sufficient for the estimator.

For a scalar SDE

$$\log p_S = -\log \sigma_W - \frac{(\widehat{S}_N - \mu_W)^2}{2\sigma_W^2} - \frac{1}{2} \log(2\pi)$$

and for a given W ,

$$\begin{aligned} \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \theta} \right] &= \frac{\partial \mu_W}{\partial \theta} \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \mu_W} \right] \\ &\quad + \frac{\partial \sigma_W}{\partial \theta} \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \sigma_W} \right]. \end{aligned}$$

Looking at the first of the two expectations, then

$$\begin{aligned} \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \mu_W} \right] &= \mathbb{E}_Z \left[\frac{\widehat{S}_N - \mu_W}{\sigma_W^2} f(\widehat{S}_N) \right] \\ &= \mathbb{E}_Z \left[\frac{Z}{\sigma_W} f(\mu_W + \sigma_W Z) \right] \\ &= \mathbb{E}_Z \left[\frac{Z}{2\sigma_W} (f(\mu_W + \sigma_W Z) - f(\mu_W - \sigma_W Z)) \right]. \end{aligned}$$

If $f(S)$ is Lipschitz, this is the expectation of a quantity which is $O(1)$ in magnitude. One sample may be sufficient, but if the computational cost of evaluating the payoff is small compared to the path calculation, it is better to use several samples; we analyse this later. If $f(S)$ is discontinuous, then for paths near the discontinuity the expectation is of a quantity which is $O(\sigma_W^{-1}) = O(h^{-1/2})$ and multiple samples should be used to estimate the expected value.

Similarly, using the additional result that $\mathbb{E}_Z[Z^2 - 1] = 0$,

$$\begin{aligned} \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \sigma_W} \right] &= \mathbb{E}_Z \left[\left(-\frac{1}{\sigma_W} + \frac{(\widehat{S}_N - \mu_W)^2}{\sigma_W^3} \right) f(\widehat{S}_N) \right] \\ &= \mathbb{E}_Z \left[\frac{Z^2 - 1}{\sigma_W} f(\mu_W + \sigma_W Z) \right] \\ &= \mathbb{E}_Z \left[\frac{Z^2 - 1}{\sigma_W} (f(\mu_W + \sigma_W Z) - f(\mu_W)) \right] \\ &= \mathbb{E}_Z \left[\frac{Z^2 - 1}{2\sigma_W} (f(\mu_W + \sigma_W Z) - 2f(\mu_W) + f(\mu_W - \sigma_W Z)) \right]. \end{aligned}$$

The expression within this expectation is in general no larger than for the previous expectation, and so the same set of samples will suffice.

3.4 Multivariate generalisation

These estimators can be generalised to the case of multiple assets with a multivariate Gaussian distribution conditional on the discrete set of Wiener increments. If μ_W is now the column vector of means, and Σ_W is the covariance matrix, then \widehat{S}_N can be written as

$$\widehat{S}_N(W, Z) = \mu_W + CZ,$$

where Z is a vector of uncorrelated unit Normal variables and C is any matrix such that $\Sigma_W = CC^T$. Provided Σ_W is non-singular, the joint probability density function for S is

$$\log p_S = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (\widehat{S}_N - \mu_W)^T \Sigma_W^{-1} (\widehat{S}_N - \mu_W) - \frac{1}{2} d \log(2\pi),$$

where d is the dimension of Z . Differentiating this (see [5, 17]) gives

$$\frac{\partial \log p_S}{\partial \mu_W} = \Sigma_W^{-1} (\widehat{S}_N - \mu_W) = C^{-T} Z,$$

and

$$\frac{\partial \log p_S}{\partial \Sigma_W} = -\frac{1}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} (\widehat{S}_N - \mu_W) (\widehat{S}_N - \mu_W)^T \Sigma^{-1} = \frac{1}{2} C^{-T} (ZZ^T - I) C^{-1},$$

and then for a given W

$$\begin{aligned} \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \theta} \right] &= \left(\frac{\partial \mu_W}{\partial \theta} \right)^T \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \mu_W} \right] \\ &\quad + \text{Trace} \left(\frac{\partial \Sigma_W}{\partial \theta} \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \Sigma_W} \right] \right), \end{aligned}$$

where the trace of a matrix is the sum of its diagonal elements. To obtain efficient estimators, we again use the symmetry property to get

$$\mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \mu_W} \right] = \mathbb{E}_Z \left[\frac{1}{2} \left(f(\mu_W + CZ) - f(\mu_W - CZ) \right) C^{-T} Z \right],$$

and we use $\mathbb{E}_Z[ZZ^T - I] = 0$ to give

$$\begin{aligned} \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \Sigma_W} \right] \\ = \mathbb{E}_Z \left[\frac{1}{4} \left(f(\mu_W + CZ) - 2f(\mu_W) + f(\mu_W - CZ) \right) C^{-T} (ZZ^T - I) C^{-1} \right]. \end{aligned}$$

If the payoff also depends on values at intermediate times τ_j , not just at maturity, these can be handled by omitting the simulation time t_n closest to each measurement time τ_j , using for that time interval $[t_{n_1}, t_{n+1}]$ a timestep which is twice as big as usual. Using Brownian interpolation conditional on the values $\widehat{S}_{n\pm 1}$ and with constant drift and volatility based on \widehat{S}_{n-1} results in a Normal distribution for $\widehat{S}(\tau_j)$ (see for example [12]). Collectively, the values $\widehat{S}(\tau_j)$ form a set with a multivariate Normal distribution, with the values at different times being independently distributed. One can then apply the theory above to obtain the sensitivities.

The Likelihood Ratio Method is not applicable when the covariance matrix Σ is singular. This situation occurs, for example, in the LIBOR market model driven by a single Brownian motion [1]. A solution is to introduce an additional diffusion in the final timestep. If the extra diffusion is of the same magnitude, this will introduce an $O(h)$ bias in the expected payoff and its sensitivity, but this bias is of the same order of magnitude as the weak convergence error associated with the Euler approximation.

3.5 Optimal number of samples

The variance analysis in the Appendix proves that if W and Z are independent random variables, then for any function $f(W, Z)$ the estimator

$$\widehat{Y}_{M,N} = M^{-1} \sum_{m=1}^M \left(N^{-1} \sum_{n=1}^N f(W^{(m)}, Z^{(m,n)}) \right)$$

with independent samples $W^{(m)}$ and $Z^{(m,n)}$ is an unbiased estimator for

$$\mathbb{E}_{W,Z} [f(W, Z)] \equiv \mathbb{E}_W \left[\mathbb{E}_Z [f(W, Z)] \right],$$

and its variance is

$$\mathbb{V}[\widehat{Y}_{M,N}] = M^{-1} \mathbb{V}_W \left[\mathbb{E}_Z [f(W, Z)] \right] + (MN)^{-1} \mathbb{E}_W \left[\mathbb{V}_Z [f(W, Z)] \right].$$

Applying this general result to our vibrato estimators with N samples for Z for each simulation path, the variance is of the form

$$v_1 M^{-1} + v_2 (MN)^{-1},$$

and the cost of computing $\widehat{Y}_{M,N}$ is proportional to

$$c_1 M + c_2 MN,$$

where c_1 corresponds to the path calculation and c_2 corresponds to the payoff evaluation. For a fixed computational cost, the variance can be minimised by minimising the product

$$(v_1 + v_2 N^{-1}) (c_1 + c_2 N) = v_1 c_2 N + v_1 c_1 + v_2 c_2 + v_2 c_1 N^{-1},$$

which gives the optimum value for N as being

$$N_{opt} = \sqrt{\frac{v_2 c_1}{v_1 c_2}}.$$

In a standard path calculation $c_1 = O(h^{-1})$ since the cost is proportional to the number of timesteps, and c_2 is $O(1)$, independent of h . If the payoff is

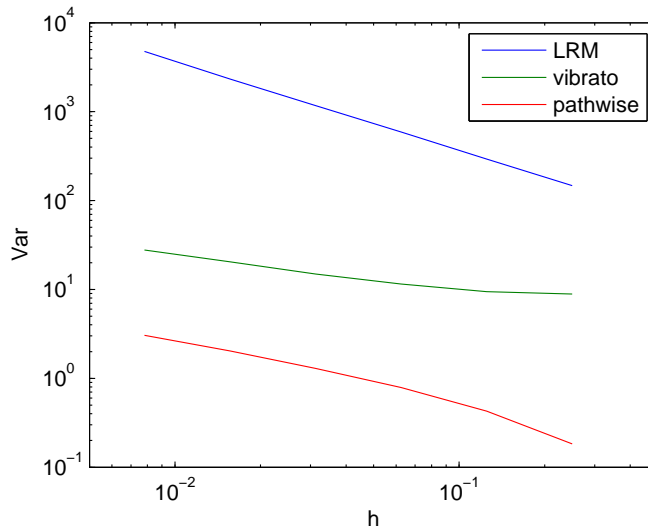


Fig. 1. Comparison of Vega variance for LRM, pathwise and vibrato estimators.

Lipschitz, then v_1 and v_2 are both $O(1)$ and so $N_{opt} = O(h^{-1/2})$. On the other hand, if the payoff is discontinuous then there will be a $O(h^{1/2})$ fraction of paths which are near the discontinuity and for which $\mathbb{E}_Z[f(W, Z)] = O(h^{-1/2})$ and $\mathbb{V}_Z[f(W, Z)] = O(h^{-1})$. This leads to v_1 and v_2 both being $O(h^{-1/2})$ and so again $N_{opt} = O(h^{-1/2})$.

In both cases, as $h \rightarrow 0$, the variance is asymptotically equal to $v_1 M^{-1}$ and the cost is asymptotically equal to $c_1 M$. Thus the use of the vibrato technique does not, to leading order, increase the variance or the computational cost compared to the use of exact conditional expectation in the few cases for which this exists in a simple closed form.

3.6 Numerical results

We consider a 2-dimensional Geometric Brownian Motion,

$$\begin{aligned} dS_t^{(1)} &= r S_t^{(1)} dt + \sigma^{(1)} S_t^{(1)} dW_t^{(1)} \\ dS_t^{(2)} &= r S_t^{(2)} dt + \sigma^{(2)} S_t^{(2)} dW_t^{(2)} \end{aligned}$$

with parameters $r=0.05$, $\sigma^{(1)}=0.2$, $\sigma^{(2)}=0.3$ and correlation $\rho=0.5$ between the driving Brownian motions. The payoff function is chosen to be a digital call paying a discounted value of $\exp(-rT)$ if and only if the value of $S_1(T)$ exceeds the strike K . Parameter values $T=1, K=100$ are used. This very simple example is chosen so that in Figure 1 we can compare the variance for the vibrato calculation to the variance of both the LRM method and also the pathwise method in combination with the analytic conditional expectation.

The figure shows the increase in the variance of the estimator for one of the Vegas, $\partial V/\partial\sigma^{(1)}$, as the timestep h is reduced. We see the rapid increase in the variance of the LRM method which is $O(h^{-1})$ asymptotically, and the much slower $O(h^{-1/2})$ growth in the variance of the two sets of results based on the pathwise approach. The difference between the pathwise and vibrato results is due to the number of Z samples used in the vibrato method. Only one sample was used for the results presented here; increasing this number will lead to the vibrato variance converging to the variance of the pathwise method using the analytic conditional expectation. It is very striking how much larger the LRM variance is. With 2 timesteps it is already 10 times larger than the vibrato method with a single Z sample, while for 128 timesteps it is 200 times larger.

4 Adjoint pathwise sensitivity implementation

The vibrato approach is completely compatible with an adjoint calculation of the path sensitivity. Following the terminology used by the Automatic Differentiation community [13, 4, 2], forward mode sensitivity calculation starts with a perturbation to an input and derives the corresponding perturbation to all subsequent variables within a computer program, with an binary operation (e.g. addition or multiplication) of the form

$$c = f(a, b)$$

giving rise to the linear perturbation equation

$$\dot{c} = \frac{\partial c}{\partial a} \dot{a} + \frac{\partial c}{\partial b} \dot{b}$$

where \dot{c} denotes the derivative of c with respect to the perturbed input parameter.

By contrast, the reverse (or adjoint) mode starts with the fact that the final output of interest has unit sensitivity with respect to itself, and then works backward through the sequence of computer instructions, to determine the sensitivity of the final output to changes in the input parameters of each instruction. Assuming that a and b are only used for the computation of c in the above example (i.e. they are not used as inputs for any other calculation) the corresponding adjoint equations are

$$\begin{aligned} \bar{a} &= \frac{\partial c}{\partial a} \bar{c} \\ \bar{b} &= \frac{\partial c}{\partial b} \bar{c} \end{aligned}$$

where \bar{a} represents the sensitivity of the final output to changes in a .

The key point of the adjoint approach is that it can compute the sensitivity of a single output quantity (such as a payoff function) to an unlimited number of input parameters (such as initial price, interest rate, volatility, etc.) at a total cost which is little more than the original calculation. For more on this approach and its use in computational finance, see [11, 9].

Using the vibrato approach, in the scalar case one would first compute the path, and then estimate the quantities

$$\mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \mu_W} \right], \quad \mathbb{E}_Z \left[f(\widehat{S}_N) \frac{\partial(\log p_S)}{\partial \sigma_W} \right]$$

using one or more Z samples. These values correspond to $\overline{\mu_W}$ and $\overline{\sigma_W}$, the sensitivity of the expected payoff for that path to changes in μ_W and σ_W . This is the starting information required for the reverse pass of the adjoint path calculation.

In the multi-dimensional case, the adjoint initialisation is

$$\overline{\mu_W} = \mathbb{E}_Z \left[\frac{1}{2} \left(f(\mu_W + CZ) - f(\mu_W - CZ) \right) C^{-T} Z \right],$$

and

$$\overline{\Sigma_W} = \mathbb{E}_Z \left[\frac{1}{4} \left(f(\mu_W + CZ) - 2f(\mu_W) + f(\mu_W - CZ) \right) C^{-T} (ZZ^T - I) C^{-1} \right].$$

5 Conclusions and future work

In this paper we have introduced a new idea of vibrato Monte Carlo calculations. This is a generalisation of the use of conditional expectation for payoff smoothing, and leads to a hybrid method for calculating sensitivities, applying pathwise sensitivity analysis to the path simulation, and the Likelihood Ratio Method to the payoff evaluation. This offers the computational efficiency of the pathwise method, particularly when combined with an adjoint implementation, together with the greater generality and ease-of-implementation of LRM.

Although the paper discusses only first order sensitivities, the approach extends naturally to higher order derivatives. A similar variance reduction construction for second order derivatives leads to an estimator with a variance which is $O(h^{-1/2})$ for payoffs which are continuous but have a discontinuous derivative, and $O(h^{-3/2})$ for payoffs which are discontinuous.

Another direction for future research is the use of the vibrato idea for multilevel Monte Carlo analysis [10]. Analytic conditional expectation is currently used to treat discontinuous payoffs to obtain improved convergence rates with the Milstein scheme [8]. The vibrato approach will allow this to be generalised to multivariate cases.

6 Acknowledgments

This research has been supported by the Oxford-Man Institute of Quantitative Finance and by a fellowship from the UK Engineering and Physical Sciences Research Council.

References

1. A. Brace, D. Gatarek, and M. Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7:127–155, 1997.
2. M. Bücker, G. Corliss, P. Hovland, U. Naumann, and B. Norris, editors. *Automatic Differentiation: Applications, Theory and Implementations*. Springer-Verlag, 2006.
3. N. Chen and P. Glasserman. Malliavin Greeks without Malliavin calculus. *Stochastic Processes and their Applications*, 117:1689–1723, 2007.
4. G. Corliss, C. Faure, A. Griewank, L. Hascoët, and U. Naumann, editors. *Automatic Differentiation: From Simulation to Optimization*. Springer-Verlag, 2001.
5. P.S. Dwyer. Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association*, 62(318):607–625, 1967.
6. E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3:391–412, 1999.
7. E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance, II. *Finance and Stochastics*, 5:201–236, 2001.
8. M.B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer-Verlag, 2007.
9. M.B. Giles. Monte Carlo evaluation of sensitivities in computational finance. Technical Report NA07/12, Oxford University Computing Laboratory, 2007.
10. M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):981–986, 2008.
11. M.B. Giles and P. Glasserman. Smoking adjoints: fast Monte Carlo Greeks. *RISK*, January 2006.
12. P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.
13. A. Griewank. *Evaluating derivatives : principles and techniques of algorithmic differentiation*. SIAM, 2000.
14. P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, 1992.
15. P. L’Ecuyer. A unified view of the IPA, SF and LR gradient estimation techniques. *Management Science*, 36(11):1364–1383, 1990.
16. P. L’Ecuyer. On the interchange of derivative and expectation for likelihood ratio derivative estimators. *Management Science*, 41(4):738–748, 1995.
17. J.R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 1988.

A Variance analysis

Theorem 1. *If W and Z are independent random variables, then*

$$\widehat{Y}_{M,N} = M^{-1} \sum_{m=1}^M \left(N^{-1} \sum_{n=1}^N f(W^{(m)}, Z^{(m,n)}) \right)$$

with independent samples $W^{(m)}$ and $Z^{(m,n)}$ is an unbiased estimator for

$$\mathbb{E}_{W,Z} [f(W, Z)] \equiv \mathbb{E}_W \left[\mathbb{E}_Z [f(W, Z)] \right],$$

and its variance is

$$\mathbb{V}[\widehat{Y}_{M,N}] = M^{-1} \mathbb{V}_W \left[\mathbb{E}_Z [f(W, Z)] \right] + (MN)^{-1} \mathbb{E}_W \left[\mathbb{V}_Z [f(W, Z)] \right].$$

Proof. For a given W ,

$$\widehat{Y}_N(W) \equiv N^{-1} \sum_{n=1}^N f(W, Z^{(n)})$$

has expectation $\mathbb{E}_Z [f(W, Z)]$ and variance $N^{-1} \mathbb{V}_Z [f(W, Z)]$, and therefore

$$\mathbb{E}_Z \left[\left(\widehat{Y}_N(W) \right)^2 \right] = \left(\mathbb{E}_Z [f(W, Z)] \right)^2 + N^{-1} \mathbb{V}_Z [f(W, Z)].$$

Taking the expectation over W gives the result that $\widehat{Y}_{1,N}$ has expectation

$$\mathbb{E}[\widehat{Y}_{1,N}] = \mathbb{E}_W \left[\mathbb{E}_Z [f(W, Z)] \right]$$

and its square has expectation

$$\mathbb{E} \left[\widehat{Y}_{1,N}^2 \right] = \mathbb{E}_W \left[\left(\mathbb{E}_Z [f(W, Z)] \right)^2 \right] + N^{-1} \mathbb{E}_W \left[\mathbb{V}_Z [f(W, Z)] \right],$$

from which it follows that its variance is

$$\begin{aligned} & \mathbb{V}[\widehat{Y}_{1,N}] \\ &= \mathbb{E} \left[\widehat{Y}_{1,N}^2 \right] - \left(\mathbb{E}[\widehat{Y}_{1,N}] \right)^2 \\ &= \mathbb{E}_W \left[\left(\mathbb{E}_Z [f(W, Z)] \right)^2 \right] - \left(\mathbb{E}_W \left[\mathbb{E}_Z [f(W, Z)] \right] \right)^2 + N^{-1} \mathbb{E}_W \left[\mathbb{V}_Z [f(W, Z)] \right] \\ &= \mathbb{V}_W \left[\mathbb{E}_Z [f(W, Z)] \right] + N^{-1} \mathbb{E}_W \left[\mathbb{V}_Z [f(W, Z)] \right]. \end{aligned}$$

The final result follows from the fact that $\mathbb{E}[\widehat{Y}_{M,N}] = \mathbb{E}[\widehat{Y}_{1,N}]$ and $\mathbb{V}[\widehat{Y}_{M,N}] = M^{-1} \mathbb{V}[\widehat{Y}_{1,N}]$.