Computing Greeks using multilevel path simulation

Sylvestre Burgos, Michael B. Giles

Abstract We investigate the extension of the multilevel Monte Carlo method [4, 5] to the calculation of Greeks. The pathwise sensitivity analysis [8] differentiates the path evolution and effectively reduces the smoothness of the payoff. This leads to new challenges: the use of naive algorithms is often impossible because of the inapplicability of pathwise sensitivities to discontinuous payoffs.
These challenges can be addressed in three different ways: payoff smoothing using conditional expectations of the payoff before maturity [8]; an approximation of the above technique using path splitting for the final timestep [1]; the use of a hybrid combination of pathwise sensitivity and the Likelihood Ratio Method [6]. We discuss the strengths and weaknesses of these alternatives in different multilevel Monte Carlo settings.

1 Introduction

In mathematical finance, Monte Carlo methods are used to compute the price of an option by estimating the expected value $\mathbb{E}(P)$. P is the payoff function that depends on an underlying asset's scalar price $S(t)$ which satisfies an evolution SDE of the form

$$\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W_t, \quad 0 \le t \le T, \quad S(0) \text{ given.} \tag{1}$$

This is just one use of Monte Carlo in finance. In practice the prices are often quoted and used to calibrate our market models; the option's sensitivities to market param-

Sylvestre Burgos
Oxford-Man Institute of Quantitative Finance, University of Oxford,
e-mail: sylvestre.burgos@maths.ox.ac.uk

Michael B. Giles
Oxford-Man Institute of Quantitative Finance, University of Oxford,
e-mail: mike.giles@maths.ox.ac.uk

eters, the so-called Greeks, reflect the exposure to different sources of risk. Computing these is essential to hedge portfolios and is therefore even more important than pricing the option itself. This is why our research focuses on getting fast and accurate estimates of Greeks through Monte Carlo simulations.

1.1 Multilevel Monte Carlo

Let us first recall important results from [4] and [5]. We consider a standard Monte Carlo method using a discretisation with first order weak convergence (e.g. the Milstein scheme). Achieving a root-mean square error of $O(\varepsilon)$ requires a variance of order $O(\varepsilon^2)$, hence $O(\varepsilon^{-2})$ independent paths. It also requires a discretisation bias of order $O(\varepsilon)$, thus $O(\varepsilon^{-1})$ timesteps, giving a total computational cost $O(\varepsilon^{-3})$.

Giles' multilevel Monte Carlo technique reduces this cost to $O(\varepsilon^{-2})$ under certain conditions. The idea is to write the expected payoff with a fine discretisation using 2^L uniform timesteps as a telescopic sum. Let $\widehat{P}_\ell$ be the simulated payoff with a discretisation using 2^ℓ uniform timesteps,

$$\mathbb{E}(\widehat{P}_L) = \mathbb{E}(\widehat{P}_0) + \sum_{\ell=1}^{L} \mathbb{E}(\widehat{P}_\ell - \widehat{P}_{\ell-1}) \tag{2}$$

We then use Monte Carlo estimators using N_ℓ independent samples

$$\mathbb{E}(\widehat{P}_\ell - \widehat{P}_{\ell-1}) \approx \widehat{Y}_\ell = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(\widehat{P}_\ell^{(i)} - \widehat{P}_{\ell-1}^{(i)} \right) \tag{3}$$

The small corrective term $\widehat{P}_\ell^{(i)} - \widehat{P}_{\ell-1}^{(i)}$ comes from the difference between a fine and a coarse discretisation of the same driving Brownian motion. Its magnitude depends on the strong convergence properties of the scheme used. Let V_ℓ be the variance of a single sample $\widehat{P}_\ell^{(i)} - \widehat{P}_{\ell-1}^{(i)}$. The next theorem shows that what determines the efficiency of the multilevel approach is the convergence rate of V_ℓ as $\ell \to \infty$.

To ensure a better efficiency we may modify (3) and use different estimators of $\widehat{P}$ on the fine and coarse levels of $\widehat{Y}_\ell$,

$$\mathbb{E}(\widehat{P}_L) = \mathbb{E}(\widehat{P}_0) + \sum_{\ell=1}^{L} \mathbb{E}\left(\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c \right) \tag{4}$$

$\widehat{P}_\ell^f$, $\widehat{P}_{\ell-1}^c$ are the estimators using respectively 2^ℓ and $2^{\ell-1}$ steps in the computation of $\widehat{Y}_\ell$. The telescoping sum property is maintained provided that

$$\mathbb{E}\left(\widehat{P}_\ell^f \right) = \mathbb{E}\left(\widehat{P}_\ell^c \right). \tag{5}$$

Theorem 1. *Let P be a function of a solution to (1) for a given Brownian path $W(t)$; let $\widehat{P}_\ell$ be the corresponding approximation using the discretisation at level ℓ, i.e. with 2^ℓ steps of width $h_\ell = 2^{-\ell}T$.*

If there exist independent estimators $\widehat{Y}_\ell$ of computational complexity C_ℓ based on N_ℓ samples and there are positive constants $\alpha \geq \frac{1}{2}, \beta, c_1, c_2, c_3$ such that

1. $\mathbb{E}(\widehat{Y}_\ell) = \begin{cases} \mathbb{E}(\widehat{P}_0) & \text{if } l = 0 \\ \mathbb{E}(\widehat{P}_\ell - \widehat{P}_{\ell-1}) & \text{if } \ell > 0 \end{cases}$
2. $|\mathbb{E}(\widehat{P}_\ell - P)| \leq c_1 h_\ell^\alpha$
3. $\mathbb{V}(\widehat{Y}_\ell) \leq c_2 h_\ell^\beta N_\ell^{-1}$
4. $C_\ell \leq c_3 N_\ell h_\ell^{-1}$

Then there is a constant c_4 such that for any $\varepsilon < e^{-1}$, there are values for L and N_ℓ resulting in a multilevel estimator $\widehat{Y} = \sum_{\ell=0}^{L} \widehat{Y}_\ell$ with a mean-square-error $MSE = \mathbb{E}((\widehat{Y} - \mathbb{E}(P))^2) < \varepsilon^2$ with a complexity C bounded by

$$C \leq \begin{cases} c_4 \varepsilon^{-2} & \text{if } \beta > 1 \\ c_4 \varepsilon^{-2} (\log \varepsilon)^2 & \text{if } \beta = 1 \\ c_4 \varepsilon^{-2-(1-\beta)/\alpha} & \text{if } 0 < \beta < 1 \end{cases} \tag{6}$$

Proof. See [5].

We usually know α thanks to the literature on weak convergence. Results in [9] give $\alpha = 1$ for the Milstein scheme, even in the case of discontinuous payoffs. β is related to strong convergence and is in practice what determines the efficiency of the multilevel approach. Its value depends on the payoff and may not be known *a priori*.

1.2 Monte Carlo Greeks

Let us briefly recall two classic methods used to compute Greeks in a Monte Carlo setting: the pathwise sensitivities and the Likelihood Ratio Method. More details can be found in [2], [3] and [8].

Pathwise sensitivities

Let $\widehat{S} = (\widehat{S}_k)_{k \in [0,N]}$ be the simulated values of the asset at the discretisation times and $\widehat{W} = (\widehat{W}_k)_{k \in [1,N]}$ be the corresponding set of independent Brownian increments. The value of the option V is estimated by $\widehat{V}$ defined as

$$V = \mathbb{E}[P(S)] \approx \widehat{V} = \mathbb{E}\left[P(\widehat{S})\right] = \int P(\widehat{S}) p(\theta, \widehat{S}) \, \mathrm{d}\widehat{S}$$

Assuming that the payoff $P(\widehat{S})$ is Lipschitz, we can use the chain rule and write

$$\frac{\partial \widehat{V}}{\partial \theta} = \frac{\partial}{\partial \theta} \int P\left(\widehat{S}(\theta, \widehat{W})\right) p(\widehat{W}) \mathrm{d}\widehat{W} = \int \frac{\partial P(\widehat{S})}{\partial \widehat{S}} \frac{\partial \widehat{S}(\theta, \widehat{W})}{\partial \theta} p(\widehat{W}) \mathrm{d}\widehat{W}$$

where $\mathrm{d}\widehat{W} = \prod_{k=1}^{N} \mathrm{d}\widehat{W}_k$ and $p(\widehat{W}) = \prod_{k=1}^{N} p(\widehat{W}_k)$ is the joint probability density function of the normally distributed independent increments $(\widehat{W}_k)_{k \in [1,N]}$.

We obtain $\dfrac{\partial \widehat{S}}{\partial \theta}$ by differentiating the discretisation of (1) with respect to θ and iterating the resulting formula.The limitation of this technique is that it requires the payoff to be Lipschitz and piecewise differentiable.

Likelihood Ratio Method

The Likelihood Ratio Method starts from

$$\widehat{V} = \mathbb{E}\left[P(\widehat{S})\right] = \int P(\widehat{S}) p(\theta, \widehat{S})\, \mathrm{d}\widehat{S} \tag{7}$$

The dependence on θ comes through the probability density function $p(\theta, \widehat{S})$; assuming some conditions discussed in [3] and in section 7 of [8], we can write

$$\frac{\partial \widehat{V}}{\partial \theta} = \int P(\widehat{S}) \frac{\partial p(\widehat{S})}{\partial \theta}\, \mathrm{d}\widehat{S} = \int P(\widehat{S}) \frac{\partial \log p(\widehat{S})}{\partial \theta} p(\widehat{S})\, \mathrm{d}\widehat{S} = \mathbb{E}\left[P(\widehat{S}) \frac{\partial \log p(\widehat{S})}{\partial \theta}\right] \tag{8}$$

$$\text{with} \qquad \mathrm{d}\widehat{S} = \prod_{k=1}^{N} \mathrm{d}\widehat{S}_k \qquad \text{and} \qquad p(\widehat{S}) = \prod_{k=1}^{N} p\left(\widehat{S}_k \middle| \widehat{S}_{k-1}\right)$$

The main limitation of the method is that the estimator's variance is $O(N)$, increasing without limit as we refine the discretisation.

1.3 Multilevel Monte Carlo Greeks

By combining the elements of sections 1.1 and 1.2 together, we write

$$\frac{\partial V}{\partial \theta} = \frac{\partial \mathbb{E}(P)}{\partial \theta} \approx \frac{\partial \mathbb{E}(\widehat{P}_L)}{\partial \theta} = \frac{\partial \mathbb{E}(\widehat{P}_0)}{\partial \theta} + \sum_{\ell=1}^{L} \frac{\partial \mathbb{E}(\widehat{P}_\ell - \widehat{P}_{\ell-1})}{\partial \theta} \tag{9}$$

As in (3), we define the multilevel estimators

$$\widehat{Y}_0 = N_0^{-1} \sum_{i=1}^{M} \frac{\partial \widehat{P}_0^{(i)}}{\partial \theta} \qquad \text{and} \qquad \widehat{Y}_\ell = N_\ell^{-1} \sum_{i=1}^{N_\ell} \left(\frac{\partial \widehat{P}_\ell^{(i)}}{\partial \theta} - \frac{\partial \widehat{P}_{\ell-1}^{(i)}}{\partial \theta} \right) \tag{10}$$

where $\dfrac{\partial \widehat{P}_0}{\partial \theta}, \dfrac{\partial \widehat{P}_{\ell-1}}{\partial \theta}, \dfrac{\partial \widehat{P}_\ell}{\partial \theta}$ are computed with the techniques presented in section 1.2.

2 European call

We consider the Black-Scholes model: the asset's evolution is modelled by a geometric Brownian motion $\mathrm{d}S(t) = r\,S(t)\mathrm{d}t + \sigma\,S(t)\mathrm{d}W_t$. We use the Milstein scheme for its good strong convergence properties. For timesteps of width h,

$$\widehat{S}_{n+1} = \widehat{S}_n \cdot \left(1 + r\,h + \sigma\,\Delta W_n + \frac{\sigma^2}{2}\left(\Delta W_n^2 - h\right)\right) := \widehat{S}_n \cdot D_n \tag{11}$$

The payoff of the European call is $P = (S_T - K)^+ = \max(0, S_T - K)$. We illustrate the techniques by computing Delta and Vega, the sensitivities to the asset's initial value S_0 and to its volatility σ. We take a time to maturity $T = 1$.

2.1 Pathwise sensitivities

Since the payoff is Lipschitz, we can use pathwise sensitivities. The differentiation of equation (11) gives

$$\frac{\partial \widehat{S}_0}{\partial S_0} = 1, \qquad \frac{\partial \widehat{S}_{n+1}}{\partial S_0} = \frac{\partial \widehat{S}_n}{\partial S_0} \cdot D_n$$

$$\frac{\partial \widehat{S}_0}{\partial \sigma} = 0, \qquad \frac{\partial \widehat{S}_{n+1}}{\partial \sigma} = \frac{\partial \widehat{S}_n}{\partial \sigma} \cdot D_n + \widehat{S}_n \left(\Delta W_n + \sigma(\Delta W_n^2 - h)\right)$$

To compute $\widehat{Y}_\ell$ we use a fine and a coarse discretisation with $N_f = 2^\ell$ and $N_c = 2^{\ell-1}$ uniform timesteps respectively.

$$\widehat{Y}_\ell = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left[\left(\frac{\partial P}{\partial S_{N_f}} \frac{\partial \widehat{S}_{N_f}^{(i)}}{\partial \theta} \right)^{(\ell)} - \left(\frac{\partial P}{\partial S_{N_c}} \frac{\partial \widehat{S}_{N_c}^{(i)}}{\partial \theta} \right)^{(\ell-1)} \right] \tag{12}$$

We use the same driving Brownian motion for the fine and coarse discretisations: we first generate the fine Brownian increments $\widehat{W} = (\Delta W_0, \Delta W_2, \ldots, \Delta W_{N_f-1})$ and then use $\widehat{W}^c = (\Delta W_0 + \Delta W_1, \ldots, \Delta W_{N_f-2} + \Delta W_{N_f-1})$ as the coarse level's increments.

To assess the order of convergence of $\mathbb{V}(\widehat{Y}_L)$, we take a sufficient number of samples so that the Monte Carlo error of our simulations will not influence the results. We plot $\log(\mathbb{V}(\widehat{Y}_\ell))$ as a function of $\log(h_\ell)$ and use a linear regression to measure the slope for the different estimators. The theoretical results on convergence are asymptotic ones, therefore the coarsest levels are not relevant: hence we perform the linear regression on levels $\ell \in [3,8]$. This gives a numerical estimate of the parameter β in Theorem 1. Combining this with the theorem, we get an estimated complexity of the multilevel algorithm. This gives the following results :

Estimator	β	MLMC Complexity
Value	≈ 2.0	$O(\varepsilon^{-2})$
Delta	≈ 0.8	$O(\varepsilon^{-2.2})$
Vega	≈ 1.0	$O(\varepsilon^{-2}\log\varepsilon^2)$

Giles has shown in [4] that $\beta = 2$ for the value's estimator. For Greeks, the convergence is degraded by the discontinuity of $\frac{\partial P}{\partial S} = \mathbf{1}_{S>K}$: a fraction $O(h_\ell)$ of the paths has a final value $\widehat{S}$ which is $O(h_\ell)$ from the discontinuity K. For these paths, there is a $O(1)$ probability that $\widehat{S}_{N_f}^{(\ell)}$ and $\widehat{S}_{N_c}^{(\ell-1)}$ are on different sides of the strike K, implying $\left(\frac{\partial P}{\partial S_{N_f}}\frac{\partial \widehat{S}_{N_f}}{\partial \theta}\right)^{(\ell)} - \left(\frac{\partial P}{\partial S_{N_c}}\frac{\partial \widehat{S}_{N_c}}{\partial \theta}\right)^{(\ell-1)}$ is $O(1)$. Thus $\mathbb{V}(\widehat{Y}_\ell) = O(h_\ell)$, and $\beta = 1$ for the Greeks.

2.2 Pathwise sensitivities and Conditional Expectations

We have seen that the payoff's lack of smoothness prevents the variance of Greeks' estimators $\widehat{Y}_\ell$ from decaying quickly and limits the potential benefits of the multilevel approach. To improve the convergence speed, we can use conditional expectations as explained in section 7.2 of [8]. Instead of simulating the whole path, we stop at the penultimate step and then for every fixed set $\widehat{W} = (\Delta W_k)_{k\in[0,N-2]}$, we consider the full distribution of $\left(\widehat{S}_N|\widehat{W}\right)$. With $a_n = a\left(\widehat{S}_{N-1}(\widehat{W}),(N-1)h\right)$ and $b_n = b\left(\widehat{S}_{N-1}(\widehat{W}),(N-1)h\right)$, we can write

$$\widehat{S}_N(\widehat{W},\Delta W_{N-1}) = \widehat{S}_{N-1}(\widehat{W}) + a_n(\widehat{W})h + b_n(\widehat{W})\,\Delta W_{N-1} \tag{13}$$

We hence get a normal distribution for $\left(\widehat{S}_N|\widehat{W}\right)$.

$$p(\widehat{S}_N|\widehat{W}) = \frac{1}{\sigma_{\widehat{W}}\sqrt{2\pi}}\exp\left(-\frac{\left(\widehat{S}_N-\mu_{\widehat{W}}\right)^2}{2\sigma_{\widehat{W}}^2}\right) \tag{14}$$

with

$$\mu_{\widehat{W}} = \widehat{S}_{N-1} + a\left(\widehat{S}_{N-1},(N-1)h\right)h$$
$$\sigma_{\widehat{W}} = b\left(\widehat{S}_{N-1},(N-1)h\right)\sqrt{h}$$

If the payoff function is sufficiently simple, we can evaluate analytically $\mathbb{E}\left[P\left(\widehat{S}_N\right)|\widehat{W}\right]$. Using the tower property, we get

$$\widehat{V} = \mathbb{E}\left[P(\widehat{S}_N)\right] = \mathbb{E}_{\widehat{W}}\left[\mathbb{E}_{\Delta W_N}\left[P(\widehat{S}_N)\middle|\widehat{W}\right]\right] \approx \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\left[P\left(\widehat{S}_N^{(m)}\right)\middle|\widehat{W}^{(m)}\right] \quad (15)$$

In the particular case of geometric Brownian motion and a European call option, we get (16) where ϕ is the normal probability density function, Φ the normal cumulative distribution function, $\alpha = (1+rh)\widehat{S}_{N-1}(\widehat{W})$ and $\beta = \sigma\sqrt{h}\widehat{S}_{N-1}(\widehat{W})$.

$$\mathbb{E}(P(\widehat{S}_N)|\widehat{W}) = \beta\,\phi\left(\frac{\alpha-K}{\beta}\right) + (\alpha-K)\,\Phi\left(\frac{\alpha-K}{\beta}\right) \quad (16)$$

This expected payoff is infinitely differentiable with respect to the input parameters. We can apply the pathwise sensitivities technique to this smooth function at time $(N-1)\,h$. The multilevel estimator for the Greek is then

$$\widehat{Y}_\ell = \frac{1}{N_\ell}\sum_{1}^{N_\ell}\left[\left(\frac{\partial\widehat{P}_f^{(i)}}{\partial\theta}\right)^{(\ell)} - \left(\frac{\partial\widehat{P}_c^{(i)}}{\partial\theta}\right)^{(\ell-1)}\right] \quad (17)$$

At the fine level we use (16) with $h = h_f$ and $\widehat{W}_f = (\Delta W_0, \Delta W_2, \ldots, \Delta W_{N_f-2})$ to get $\mathbb{E}(P(\widehat{S}_{N_f})|\widehat{W}_f)$. We then use

$$\left(\frac{\partial\widehat{P}_f}{\partial\theta}\right)^{(\ell)} = \frac{\partial\widehat{S}_{N_f-1}}{\partial\theta}\frac{\partial\mathbb{E}(P(\widehat{S}_{N_f})|\widehat{W}_f)}{\partial S_{N_f-1}} + \frac{\partial\mathbb{E}(P(\widehat{S}_{N_f})|\widehat{W}_f)}{\partial\theta} \quad (18)$$

At the coarse level, directly using $\mathbb{E}(P(\widehat{S}_{N_c})|\widehat{W}_c)$ leads to an unsatisfactorily low convergence rate of $\mathbb{V}(\widehat{Y}_\ell)$. As explained in (4) we use a modified estimator. The idea is to include the final fine Brownian increment in the computation of the expectation over the last coarse timestep. This guarantees that the two paths will be close to one another and helps achieve better variance convergence rates.

$\widehat{S}$ still follows a simple Brownian motion with constant drift and volatility on all coarse steps. With $\widehat{W}_c = (\Delta W_0 + \Delta W_1, \ldots, \Delta W_{N_f-4} + \Delta W_{N_f-3})$ and given that the Brownian increment on the first half of the final step is ΔW_{N_f-2} , we get

$$p(\widehat{S}_{N_c}|\widehat{W}_c, \Delta W_{N_f-2}) = \frac{1}{\sigma_{\widehat{W}_c}\sqrt{2\pi}}\exp\left(-\frac{\left(\widehat{S}_{N_c}-\mu_{\widehat{W}_c}\right)^2}{2\sigma_{\widehat{W}_c}^2}\right) \quad (19)$$

with

$$\mu_{\widehat{W}_c} = \widehat{S}_{N_c-1}(\widehat{W}_c) + a\left(\widehat{S}_{N_c-1}, (N_c-1)h_c\right)h_c + b\left(\widehat{S}_{N_c-1}, (N_c-1)h_c\right)\Delta W_{N_f-2}$$

$$\sigma_{\widehat{W}_c} = b\left(\widehat{S}_{N_c-1}, (N_c-1)h_c\right)\sqrt{h_c/2}$$

From this distribution we derive $\mathbb{E}\left[P(\widehat{S}_{N_c}) \,\middle|\, \widehat{W}_c, \Delta W_{N_f-2}\right]$, which for the particular application being considered, leads to the same payoff formula as before with $\alpha_c = (1 + r h_c + \sigma \Delta W_{N_f-2})\widehat{S}_{N_c-1}(\widehat{W}_c)$ and $\beta_c = \sigma\sqrt{h_c}\,\widehat{S}_{N_c-1}(\widehat{W}_c)$. Using it as the coarse level's payoff does not introduce any bias. Using the tower property we check that it satisfies condition (5),

$$\mathbb{E}_{\Delta W_{N_f-1}}\left[\mathbb{E}\left[P(\widehat{S}_{N_c})\,\middle|\,\widehat{W}_c, \Delta W_{N_f-2}\right]\,\middle|\,\widehat{W}_c\right] = \mathbb{E}\left[P(\widehat{S}_{N_c})\,\middle|\,\widehat{W}_c\right]$$

Our numerical experiments show the benefits of the conditional expectation technique on the European call:

Estimator	β	MLMC Complexity
Value	≈ 2.0	$O(\varepsilon^{-2})$
Delta	≈ 1.5	$O(\varepsilon^{-2})$
Vega	≈ 2.0	$O(\varepsilon^{-2})$

A fraction $O(\sqrt{h_\ell})$ of the paths arrive in the area around the strike where the conditional expectation $\dfrac{\partial \mathbb{E}(P(\widehat{S}_N)|\widehat{W})}{\partial \widehat{S}_{N_f-1}}$ is neither close to 0 nor 1. In this area, its slope is $O(h_\ell^{-1/2})$. The coarse and fine paths differ by $O(h_\ell)$, we thus have $O(\sqrt{h_\ell})$ difference between the coarse and fine Greeks' estimates. Reasoning as in [4] we get $\mathbb{V}_{\widehat{W}}(\mathbb{E}_{\Delta W_{N-1}}(...|\widehat{W})) = O(h_\ell^{3/2})$ for the Greeks' estimators. This is the convergence rate observed for Delta; the higher convergence rate of Vega is not explained yet by this rough analysis and will be investigated in our future research.

The main limitation of this approach is that in many situations it leads to complicated integral computations. Path splitting, to be discussed next, may represent a useful numerical approximation to this technique.

2.3 Split pathwise sensitivities

This technique is based on the previous one. The idea is to avoid the tricky computation of $\mathbb{E}\left[P(\widehat{S}_{N_f})|\widehat{W}_f\right]$ and $\mathbb{E}\left[P(\widehat{S}_{N_c})|\widehat{W}_c, \Delta W_{N_f-2}\right]$. Instead, as detailed in section 5.5 of [1], we get numerical estimates of these values by "splitting" every path simulation on the final timestep.

At the fine level: for every simulated path $\widehat{W}_f = (\Delta W_0, \Delta W_2, \ldots, \Delta W_{N_f-2})$, we simulate a set of d final increments $(\Delta W^{(i)}_{N_f-1})_{i\in[1,d]}$ which we average to get

$$\mathbb{E}\left[P(\widehat{S}_{N_f})\,\middle|\,\widehat{W}_f\right] \approx \frac{1}{d}\sum_{i=1}^{d} P(\widehat{S}_{N_f}(\widehat{W}_f, \Delta W^{(i)}_{N_f-1})) \tag{20}$$

At the coarse level we use $\widehat{W}_c = (\Delta W_0 + \Delta W_1, \ldots, \Delta W_{N_f-4} + \Delta W_{N_f-3})$. As before (still assuming a constant drift and volatility on the final coarse step), we improve the convergence rate of $\mathbb{V}(\widehat{Y}_\ell)$ by reusing ΔW_{N_f-2} in our estimation of $\mathbb{E}\left[P(\widehat{S}_{N_c}) \,\middle|\, \widehat{W}_c\right]$. We can do so by constructing the final coarse increments as $(\Delta W^{(i)}_{N_c-1})_{i\in[1,d]} = (\Delta W_{N_f-2} + (\Delta W^{(i)}_{N_f-1}))_{i\in[1,d]}$ and using these to estimate

$$\mathbb{E}(P(\widehat{S}_{N_c})|\widehat{W}_c) = \mathbb{E}\left[P(\widehat{S}_{N_c}) \,\middle|\, \widehat{W}_c, \Delta W_{N_f-2}\right] \approx \frac{1}{d}\sum_{i=1}^{d} P\left(\widehat{S}_{N_c}(\widehat{W}_c, \Delta W^{(i)}_{N_c-1})\right)$$

To get the Greeks, we simply compute the corresponding pathwise sensitivities.

We now examine the influence of d the number of splittings on the estimated complexity.

Estimator	d	β	MLMC Complexity
Value	10	≈ 2.0	$O(\varepsilon^{-2})$
	500	≈ 2.0	$O(\varepsilon^{-2})$
Delta	10	≈ 1.0	$O(\varepsilon^{-2}(\log\varepsilon)^2)$
	500	≈ 1.5	$O(\varepsilon^{-2})$
Vega	10	≈ 1.6	$O(\varepsilon^{-2})$
	500	≈ 2.0	$O(\varepsilon^{-2})$

As expected this method yields higher values of β than simple pathwise sensitivities: the convergence rates increase and tend to the rates offered by conditional expectations as d increases and the approximation gets more precise.

Taking a constant number of splittings d for all levels is actually not optimal; for Greeks we can write the variance of the estimator as

$$\begin{aligned}\mathbb{V}(\widehat{Y}_\ell) = &\frac{1}{N_\ell}\mathbb{V}_{\widehat{W}_f}\left[\mathbb{E}\left[\left(\frac{\partial \widehat{P}_f}{\partial\theta}\right)^{(\ell)} - \left(\frac{\partial \widehat{P}_c}{\partial\theta}\right)^{(\ell-1)} \middle| \widehat{W}_f\right]\right] \\ &+ \frac{1}{N_\ell d}\mathbb{E}_{\widehat{W}_f}\left[\mathbb{V}\left[\left(\frac{\partial \widehat{P}_f}{\partial\theta}\right)^{(\ell)} - \left(\frac{\partial \widehat{P}_c}{\partial\theta}\right)^{(\ell-1)} \middle| \widehat{W}_f\right]\right]\end{aligned} \tag{21}$$

As explained in section 2.2 we have $\mathbb{V}_{\widehat{W}_f}(\mathbb{E}(...|\widehat{W}_f)) = O(h_\ell^{3/2})$ for the Greeks. We also have $\mathbb{E}_{\widehat{W}_f}(\mathbb{V}(...|\widehat{W}_f)) = O(h_\ell)$ for similar reasons. We optimise the variance at a fixed computational cost by choosing d such that the two terms of the sum are of similar order. Taking $d = O(h_\ell^{-1/2})$ is therefore optimal.

2.4 Vibrato Monte Carlo

Since the previous method uses pathwise sensitivity analysis, it is not applicable when payoffs are discontinuous. To address this limitation, we use the Vibrato Monte Carlo method introduced by Giles [6]. This hybrid method combines pathwise sensitivities and the Likelihood Ratio Method.

We consider again equation (15). We now use the Likelihood Ratio Method on the last timestep and with the notations of section 2.2 we get

$$\frac{\partial \widehat{V}}{\partial \theta} = \mathbb{E}_{\widehat{W}}\left[\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial(\log p(\widehat{S}_N|\widehat{W}))}{\partial \theta}\,\middle|\,\widehat{W}\right]\right] \tag{22}$$

We can write $p(\widehat{S}_N|\widehat{W}))$ as $p(\mu_{\widehat{W}},\sigma_{\widehat{W}})$. This leads to the estimator

$$\begin{aligned}\frac{\partial \widehat{V}}{\partial \theta} \approx \frac{1}{N_\ell}\sum_{m=1}^{N_\ell} &\left(\frac{\partial \mu_{\widehat{W}^{(m)}}}{\partial \theta}\mathbb{E}_{\Delta W_{N-1}}\left[P(\,\widehat{S}_N)\frac{\partial(\log p)}{\partial \mu_{\widehat{W}}}\,\middle|\,\widehat{W}^{(m)}\right]\right.\\ &\left.+\frac{\partial \sigma_{\widehat{W}^{(m)}}}{\partial \theta}\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial(\log p)}{\partial \sigma_{\widehat{W}}}\,\middle|\,\widehat{W}^{(m)}\right]\right)\end{aligned} \tag{23}$$

We compute $\dfrac{\partial \mu_{\widehat{W}^{(m)}}}{\partial \theta}$ and $\dfrac{\partial \sigma_{\widehat{W}^{(m)}}}{\partial \theta}$ with pathwise sensitivities.
With $\widehat{S}_N^{(m,i)} = \widehat{S}_N(\widehat{W}^{(m)},\Delta W_{N-1}^{(i)})$, we substitute the following estimators into (23)

$$\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial(\log p)}{\partial \mu_{\widehat{W}}}\,\middle|\,\widehat{W}^{(m)}\right] \approx \frac{1}{d}\sum_{i=1}^{d}\left(P\left(\widehat{S}_N^{(m,i)}\right)\frac{\widehat{S}_N^{(m,i)}-\mu_{\widehat{W}^{(m)}}}{\sigma^2_{\widehat{W}^{(m)}}}\right)$$

$$\mathbb{E}_{\Delta W_{N-1}}\left[P\left(\widehat{S}_N\right)\frac{\partial(\log p)}{\partial \sigma_{\widehat{W}}}\,\middle|\,\widehat{W}^{(m)}\right] \approx \frac{1}{d}\sum_{i=1}^{d}P\left(\widehat{S}_N^{(m,i)}\right)\left(\frac{\left(\widehat{S}_N^{(m,i)}-\mu_{\widehat{W}^{(m)}}\right)^2}{\sigma^3_{\widehat{W}^{(m)}}}-\frac{1}{\sigma_{\widehat{W}^{(m)}}}\right)$$

In a multilevel setting: at the fine level we can use (23) directly. At the coarse level, for the same reasons as in section 2.3, we reuse the fine brownian increments to get efficient estimators. We take

$$\begin{aligned}&\widehat{W}_c = (\Delta W_0+\Delta W_1,\ldots,\Delta W_{N_f-4}+\Delta W_{N_f-3})\\ &(\Delta W_{N_c-1}^{(i)})_{i\in[1,d]} = (\Delta W_{N_f-2}+(\Delta W_{N_f-1}^{(i)}))_{i\in[1,d]}\end{aligned} \tag{24}$$

We use the tower property to verify that condition (5) is verified on the last coarse step. With the notations of equation (19) we derive the following estimators

$$
\begin{aligned}
&\mathbb{E}_{\Delta W_{N_c-1}}\left[P\left(\widehat{S}_{N_c}\right)\frac{\partial(\log p_c)}{\partial\mu_{\widehat{W}_c}}\,\middle|\,\widehat{W}_c^{(m)}\right]\\
&\qquad=\mathbb{E}\left[\mathbb{E}\left[P\left(\widehat{S}_{N_c}\right)\frac{\partial(\log p_c)}{\partial\mu_{\widehat{W}_c}}\,\middle|\,\widehat{W}_c^{(m)},\Delta W_{N_f-2}\right]\middle|\,\widehat{W}_c^{(m)}\right]\\
&\qquad\approx\frac{1}{d}\sum_{i=1}^{d}\left(P\left(\widehat{S}_{N_c}^{(m,i)}\right)\frac{\widehat{S}_{N_c}^{(m,i)}-\mu_{\widehat{W}_c^{(m)}}}{\sigma^2_{\widehat{W}_c^{(m)}}}\right)\\
&\mathbb{E}_{\Delta W_{N_c-1}}\left[P\left(\widehat{S}_{N_c}\right)\frac{\partial(\log p)}{\partial\sigma_{\widehat{W}}}\,\middle|\,\widehat{W}_c^{(m)}\right]\\
&\qquad=\mathbb{E}\left[\mathbb{E}\left[P\left(\widehat{S}_{N_c}\right)\frac{\partial(\log p)}{\partial\sigma_{\widehat{W}}}\,\middle|\,\widehat{W}_c^{(m)},\Delta W_{N_f-2}\right]\middle|\,\widehat{W}_c^{(m)}\right]\\
&\qquad\approx\frac{1}{d}\sum_{i=1}^{d}P\left(\widehat{S}_{N_c}^{(m,i)}\right)\left(-\frac{1}{\sigma_{\widehat{W}_c^{(m)}}}+\frac{\left(\widehat{S}_{N_c}^{(m,i)}-\mu_{\widehat{W}_c^{(m)}}\right)^2}{\sigma^3_{\widehat{W}_c^{(m)}}}\right)
\end{aligned}
\tag{25}
$$

Our numerical experiments show the following convergence rates for $d = 10$:

Estimator	β	MLMC Complexity
Value	≈ 2.0	$O(\varepsilon^{-2})$
Delta	≈ 1.5	$O(\varepsilon^{-2})$
Vega	≈ 2.0	$O(\varepsilon^{-2})$

As in section 2.3, this is an approximation of the conditional expectation technique, and so the same convergence rates was expected.

3 European digital call

The European digital call's payoff is $P = \mathbf{1}_{S_T>K}$. The discontinuity of the payoff makes the computation of Greeks more challenging. We cannot apply pathwise sensitivities, and so we use conditional expectations or Vibrato Monte Carlo.

With the same notation as in section 2.2 we compute the conditional expectations of the digital call's payoff.

$$
\mathbb{E}(P(\widehat{S}_{N_f})|\widehat{W}) = \Phi\left(\frac{\alpha-K}{\beta}\right) \qquad \mathbb{E}(P(\widehat{S}_{N_c})|\widehat{W}_c,\Delta W_{N_f-2}) = \Phi\left(\frac{\alpha_c-K}{\beta_c}\right)
$$

The simulations give

Estimator	β	MLMC Complexity
Value	≈ 1.4	$O(\varepsilon^{-2})$
Delta	≈ 0.5	$O(\varepsilon^{-2.5})$
Vega	≈ 0.6	$O(\varepsilon^{-2.4})$

The Vibrato technique can be applied in the same way as with the European call. We get

Estimator	β	MLMC Complexity
Value	≈ 1.3	$O(\varepsilon^{-2})$
Delta	≈ 0.3	$O(\varepsilon^{-2.7})$
Vega	≈ 0.5	$O(\varepsilon^{-2.5})$

The analysis presented in section 2.2 explains why we expected $\beta = 3/2$ for the value's estimator. A fraction $O(\sqrt{h})$ of all paths arrive in the area around the payoff where $(\partial \mathbb{E}(P(\widehat{S}_N)|\widehat{W})/\partial \widehat{S}_{N-1})$ is not close to 0 ; there its derivative is $O(h_\ell^{-1})$ and we have $|\widehat{S}_{N_f} - \widehat{S}_{N_c}| = O(h_\ell)$. For these paths, we thus have $O(1)$ difference between the fine and coarse Greeks' estimates. This explains the experimental $\beta \approx 1/2$.

4 European lookback call

The lookback call's value depends on the values that the asset takes before expiry. Its payoff is $P(T) = (S_T - \min_{t \in [0,T]} (S_t))$.

As explained in [4], the natural discretisation $\widehat{P} = (\widehat{S}_N - \min_n \widehat{S}_n)$ is not satisfactory. To regain good convergence rates, we approximate the behaviour within each fine timestep $[t_n, t_{n+1}]$ of width h_f as a simple Brownian motion with constant drift a_n^f and volatility b_n^f conditional on the simulated values $\widehat{S}_n^f$ and $\widehat{S}_{n+1}^f$. As shown in [8] we can then simulate the local minimum

$$\widehat{S}_{n,min}^f = \frac{1}{2}\left(\widehat{S}_n^f + \widehat{S}_{n+1}^f - \sqrt{\left(\widehat{S}_{n+1}^f - \widehat{S}_n^f\right)^2 - 2(b_n^f)^2 h_f \log U_n}\right) \tag{26}$$

with U_n a uniform random variable on $[0,1]$. We define the fine level's payoff this way choosing $b_n^f = b(\widehat{S}_n^f, t_n)$ and considering the minimum over all timesteps to get the global minimum of the path.

At the coarse level we still consider a simple Brownian motion on each timestep of width $h_c = 2h_f$. To get high strong convergence rates, we reuse the fine increments by defining a midpoint value for each step

$$\widehat{S}_{n+1/2}^c = \frac{1}{2}\left(\widehat{S}_n^c + \widehat{S}_{n+1}^c - b_n^c(\Delta W_{n+1/2} - \Delta W_n)\right), \tag{27}$$

where $(\Delta W_{n+1/2} - \Delta W_n)$ is the difference of the corresponding fine Brownian increments on $[t_{n+1/2}, t_{n+1}]$ and $[t_n, t_{n+1/2}]$. Conditional on this value, we then define the minimum over the whole step as the minimum of the minimum over each half step, that is

$$\widehat{S}^c_{n,min} = \min\left[\frac{1}{2}\left(\widehat{S}^c_n + \widehat{S}^c_{n+1/2} - \sqrt{\left(\widehat{S}^c_{n+1/2} - \widehat{S}^c_n\right)^2 - (b^c_n)^2 h_c \log U_{1,n}}\right),\right.$$
$$\left.\frac{1}{2}\left(\widehat{S}^c_{n+1/2} + \widehat{S}^c_{n+1} - \sqrt{\left(\widehat{S}^c_{n+1} - \widehat{S}^c_{n+1/2}\right)^2 - (b^c_n)^2 h_c \log U_{2,n}}\right)\right] \quad (28)$$

where $U_{1,n}$ and $U_{2,n}$ are the values we sampled to compute the minima of the corresponding timesteps at the fine level. Once again we use the tower property to check that condition (5) is verified and that this coarse-level estimator is adequate.

Using the treatment described above, we can then apply straighforward pathwise sensitivities to compute the multilevel estimator. This gives the following results:

Estimator	β	MLMC Complexity
Value	≈ 1.9	$O(\varepsilon^{-2})$
Delta	≈ 1.9	$O(\varepsilon^{-2})$
Vega	≈ 1.3	$O(\varepsilon^{-2})$

For the value's estimator, Giles, Debrabant and Rössler [7] have proved that $\mathbb{V}(\widehat{Y}_l) = O(h_\ell^{2-\delta})$ for all $\delta > 0$, thus we expected $\beta \approx 2$. In the Black & Scholes model, we can prove that Delta $= (V/S_0)$. We therefore expected $\beta \approx 2$ for Delta too. The strong convergence speed of Vega's estimator cannot be derived that easily and will be analysed in our future research.

Unlike the regular call option, the payoff of the lookback call is perfectly smooth and so therefore there is no benefit from using conditional expectations and associated methods.

5 European barrier call

Barrier options are contracts which are activated or deactivated when the underlying asset S reaches a certain barrier value B. We consider here the down-and-out call for which the payoff can be written as

$$P = (S_T - K)^+ \mathbf{1}_{\min\limits_{t\in[0,T]}(S_t) > K} \quad (29)$$

Both the naive estimators and the approach used with the lookback call are unsatisfactory here: the discontinuity induced by the barrier results in a higher variance than before. Therefore we use the approach developed in [4] where we compute the probability p_n that the minimum of the interpolant crosses the barrier within each timestep. This gives the conditional expectation of the payoff conditional on the Brownian increments of the fine path:

$$\widehat{P}^f = (\widehat{S}^f_{N_f} - K)^+ \prod_{n=0}^{N_f-1} \left(1 - \widehat{p}^f_n\right) \quad (30)$$

with

$$\widehat{p}^f_n = \exp\left(\frac{-2(\widehat{S}^f_n - B)^+(\widehat{S}^f_{n+1} - B)^+}{(b^f_n)^2\, h_f}\right)$$

At the coarse level we define the payoff similarly: we first simulate a midpoint value $\widehat{S}^c_{n+1/2}$ as before and then define $\widehat{p}^c_n$ the probability of not hitting B in $[t_n, t_{n+1}]$, that is the probability of not hitting B in $[t_n, t_{n+1/2}]$ and $[t_{n+1/2}, t_{n+1}]$. Thus

$$\widehat{P}^c = (\widehat{S}^c_{N_c} - K)^+ \prod_{n=0}^{N_c-1} (1 - \widehat{p}^c_n) = (\widehat{S}^c_{N_c} - K)^+ \prod_{n=0}^{N_c-1} ((1 - \widehat{p}_{n,1})(1 - \widehat{p}_{n,2})) \tag{31}$$

with

$$\widehat{p}_{n,1} = \exp\left(\frac{-2(\widehat{S}^c_n - B)^+(\widehat{S}^c_{n+1/2} - B)^+}{(b^c_n)^2\, h_f}\right)$$
$$\widehat{p}_{n,2} = \exp\left(\frac{-2(\widehat{S}^c_{n+1/2} - B)^+(\widehat{S}^c_{n+1} - B)^+}{(b^c_n)^2\, h_f}\right)$$

5.1 Pathwise sensitivities

The multilevel estimators $\widehat{Y}_\ell = \left(\widehat{P}^f\right)^{(\ell)} - \left(\widehat{P}^c\right)^{(\ell-1)}$ are Lipschitz with respect to all $(\widehat{S}^f_n)_{n=1\ldots N_f}$ and $(\widehat{S}^c_n)_{n=1\ldots N_c}$, so we can use pathwise sensitivities to compute the Greeks. Our numerical simulations give

Estimator	β	MLMC Complexity
Value	≈ 1.6	$O(\varepsilon^{-2})$
Delta	≈ 0.6	$O(\varepsilon^{-2.4})$
Vega	≈ 0.6	$O(\varepsilon^{-2.4})$

Giles proved $\beta = \frac{3}{2} - \delta$ ($\delta > 0$) for the value's estimator. We are currently working on a numerical analysis supporting the observed convergence rates for the Greeks.

5.2 Conditional Expectations

The low convergence rates observed in the previous section come from both the discontinuity at the barrier and from the lack of smoothness of the call around K. To address the latter, we can use the techniques described in section 1. Since path splitting and Vibrato Monte Carlo offer rates that are at best equal to those of conditional expectations, we have therefore implemented conditional expectations and obtained the following results:

Estimator	β	MLMC Complexity
Value	≈ 1.7	$O(\varepsilon^{-2})$
Delta	≈ 0.7	$O(\varepsilon^{-2.3})$
Vega	≈ 0.7	$O(\varepsilon^{-2.3})$

We see that the maximum benefits of these techniques are only marginal. The barrier appears to be responsible for most of the variance of the multilevel estimators.

Conclusion and future work

In this paper we have shown for a range of cases how multilevel techniques can be used to reduce the computational complexity of Monte Carlo Greeks.

Smoothing a Lipschitz payoff with conditional expectations reduces the complexity to $O(\varepsilon^{-2})$. From this technique we derive the Path splitting and Vibrato methods: they offer the same efficiency and avoid intricate integral computations. Payoff smoothing and Vibrato also enable us to extend the computation of Greeks to discontinuous payoffs where the pathwise sensitivity approach is not applicable. Numerical evidence shows that with well-constructed estimators these techniques provide computational savings even with exotic payoffs.

So far we have mostly relied on numerical estimates of β to estimate the complexity of the algorithms. Our current analysis is somewhat crude ; this is why our current research now focuses on a rigorous numerical analysis of the algorithms' complexity.

References

[1] S. Asmussen and P. Glynn. *Stochastic Simulation*. Springer, New York, 2007.
[2] M. Broadie and P. Glasserman. "Estimating security price derivatives using simulation". *Management Science*, 42, 2 (1996), 269-285.
[3] P. L'Ecuyer. "A unified view of the IPA, SF and LR gradient estimation techniques". *Management Science*, 36, 11 (1990), 1364-1383.
[4] 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*, 343-358. Springer-Verlag, 2007.
[5] M.B. Giles. "Multilevel Monte Carlo path simulation". *Operations Research*, 56, 3 (2008), 607-617.
[6] M.B. Giles. "Vibrato Monte Carlo sensitivities". In P. L'Ecuyer and A. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2008*, 369-382. Springer, 2009.
[7] M.B. Giles, K. Debrabant and A. Rössler. "Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation". In preparation.
[8] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.
[9] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.