Multilevel dimension reduction Monte-Carlo simulation for high-dimensional stochastic models in finance

Duy-Minh Dang\(^1\), Qifan Xu\(^1\), and Shangzhe Wu\(^1\)

School of Mathematics and Physics, The University of Queensland, St Lucia, Brisbane, 4072, Australia
duyminh.dang@uq.edu.au, qifan.xu@uq.net.au, shangzhe.wu@uq.net.au

Abstract
One-way coupling often occurs in multi-dimensional stochastic models in finance. In this paper, we develop a highly efficient Monte Carlo (MC) method for pricing European options under a \(N\)-dimensional one-way coupled model, where \(N\) is arbitrary. The method is based on a combination of (i) the powerful dimension and variance reduction technique, referred to as drMC, developed in Dang et. al (2014), that exploits this structure, and (ii) the highly effective multilevel MC (mlMC) approach developed by Giles (2008). By first applying Step (i), the dimension of the problem is reduced from \(N\) to 1, and as a result, Step (ii) is essentially an application of mlMC on a 1-dimensional problem. Numerical results show that, through a careful construction of the ml-dr estimator, improved efficiency expected from the Milstein timestepping with first order strong convergence can be achieved. Moreover, our numerical results show that the proposed ml-drMC method is significantly more efficient than the mlMC methods currently available for multi-dimensional stochastic problems.

Keywords: Monte Carlo, multilevel, conditional Monte Carlo, dimension reduction, variance reduction

1 Introduction

The rapid growth of financial markets over the past few decades has spawned many intellectually challenging problems. These problems have evolved from the simple one-factor Black-Scholes (BS) model to highly-complex multi-factor models that capture crucial features of important practical applications. For instance, pricing models for option contracts have been extended from the BS model to include stochastic variance, e.g. via the very popular square-root Cox-Ingersoll-Ross (CIR) model [1], and/or a stochastic interest rate, e.g. via a one/two-factor Hull-White model [2, 4, 3]. In addition, the financial markets have become more diverse, with trading not only of stocks, but also of many types of financial derivatives. For example, cross-currency interest rate derivatives have become increasingly important and are traded in large quantities in Over-the-Counter markets [10, 8].

Since in practice, closed-form solutions rarely exist for most pricing problems, numerical methods must be used. For high-dimensional (stochastic) models, the Monte-Carlo (MC) approach is very popular, due to the fact that the complexity of MC methods increases linearly...
with respect to the number of dimensions. Using an ordinary MC approach, referred to as ordMC, with a (time) discretization method having first-order weak convergence, such as the simple Euler-Maruyama, the computational cost to achieve a root-mean-square error (RMSE) of $\epsilon$ is $O(\epsilon^{-3})$.

The multi-level MC (mlMC) approach, developed in [5], is based on the multi-grid idea for iterative solutions of PDEs, but applied to MC path calculations. More specifically, the mlMC approach combines simulations with different numbers of timestep sizes to achieve the same level of accuracy obtained by ordMC at the finest timestep size, but at a much lower computational cost. It is well-known that the efficiency of the mlMC method primarily depends on the strong convergence of the (time) discretization method (e.g. see [6, 7], among many others). With the first-order strong convergence Milstein discretization, to achieve a RMSE of $\epsilon$, the computational cost is reduced to $O(\epsilon^{-2})$, even for discontinuous and path-dependent payoffs. This offers a significant saving compared to the simple Euler-Maruyama discretization which has only half-order strong convergence, and hence, $O(\epsilon^{-2}(\log(1/\epsilon))^{2})$ computational cost. As a result, there has been much interest recently in the computational finance community in using mlMC with the Milstein discretization. However, a disadvantage of the Milstein scheme is that, for multi-dimensional models, except in some special cases, it usually requires simulation of iterated Itô integrals, also known as Lévy areas, and this is usually very slow. In [7], it is shown that, through the construction of a suitable antithetic mlMC estimator, it is possible to avoid simulating Lévy areas, but still achieve the overall complexity $O(\epsilon^{-2})$ for a RMSE of $\epsilon$. To the best of our knowledge, this is the only mlMC method currently available in the literature that deals with multi-dimensional models. Note that this method still requires multi-dimensional MC simulations.

Along a different line of MC research, in [3], we develop a highly-efficient and easy-to-implement dimension reduction approach for MC methods, referred to as drMC, for plain-vanilla European options written on multi-dimensional stochastic models with one-way coupling. The underlying idea of the drMC approach is to combine (i) the conditional MC technique, and (ii) the derivation of closed-form solution of the conditional partial differential equations (PDEs) via the Fourier transform. The drMC approach reduces the dimension of an $N$-dimensional ($N$-dim) problem from $N$ to 1, provided the model is suitably coupled. As a result, the variance associated with the stochastic factors in the model, except that of the factor we condition on, is completely removed from the option price. This often results in a significant variance reduction.

In this paper, we present a highly efficient ml-drMC pricing method for plain-vanilla European options based on multi-dimensional stochastic models with one-way coupling. The ml-drMC method essentially consists of two stages. In the first stage, by applying the drMC method, we can reduce the dimension of an $N$-dim problem from $N$ to 1. In the second stage, we apply the multilevel technique with the Milstein discretization to the stochastic factor we condition on in the first stage. Under the drMC framework, however, the option price can be expressed as an expectation of a non-conventional path-dependent payoff, a fact that poses challenges in constructing a multilevel estimator that preserves the improved efficiency expected from the Milstein discretization in the second stage. Nonetheless, we show that such a multilevel estimator can be devised. As an illustration to the underlying idea, as well as to demonstrate the efficiency and robustness of the ml-drMC method for some high-dimensional problems, we present the ml-drMC method in a very general cross-currency context with stochastic variance, multi-factor domestic and foreign short rates, and full correlations between factors.

We highlight two important potential advantages of the ml-drMC over the antithetic mlMC method of [7] in this context. Due to the model dimension reduction from $N$ to 1 of the first stage, the second stage is essentially an application of the multilevel technique to a 1-dim problem, and hence, naturally avoids simulating Lévy areas. In addition, since the first stage also often results in a significant variance reduction, it is expected that the ml-drMC is
significantly more efficient than the antithetic mlMC approach of [7]. As a result, the ml-drMC method is (potentially) a significantly more efficient and better alternative to the antithetic mlMC method of [7], provided that the models are suitably coupled. However, no convergence proofs have yet been constructed for ml-drMC, and so the paper relies on the numerical results to demonstrate the effectiveness of the method.

2 Cross-currency model

We consider the following cross-currency model under domestic risk-neutral measure $Q$.

$$dS(t) = S(t) \left( (r_d(t) - r_f(t)) \, dt + \sqrt{\nu(t)} \, dW_s(t) \right),$$  

$$(1a)$$

$$r_d(t) = \sum_{i=1}^{m} X_i(t) + \gamma_d(t), \quad dX_i(t) = -\kappa_d(t) \, X_i(t) \, dt + \sigma_d(t) \, dW_{d,i}(t), \quad X_i(0) = 0,$$  

$$(1b)$$

$$r_f(t) = \sum_{i=1}^{l} Y_i(t) + \gamma_f(t), \quad dY_i(t) = -\kappa_f(t) \, Y_i(t) \, dt + \sigma_f(t) \, dW_{f,i}(t) - \rho_{S,f} \sigma_f(t) \sqrt{\nu(t)} \, dt, \quad Y_i(0) = 0,$$  

$$(1c)$$

$$dr(t) = \kappa_{\nu} (\bar{\nu} - \nu(t)) \, dt + \sigma_{\nu} \sqrt{\nu(t)} \, dW_{\nu}(t).$$  

$$(1d)$$

Here, $S(t)$, $r_d(t)$, $r_f(t)$ and $\nu(t)$, respectively, represent the spot foreign exchange (FX) rate, the domestic short rate, the foreign short rate, and the variance of the spot FX rate. The spot FX rate is defined as the number of units of domestic currency per one unit of foreign currency. The functions $\kappa_d(t)$, $\sigma_d(t)$, $\kappa_f(t)$, and $\sigma_f(t)$, are deterministic functions of $t$, with $\kappa_d(t)$, and $\kappa_f(t)$ being the mean-reversion rates. The functions $\gamma_d(t)$ and $\gamma_f(t)$ are also deterministic, and they, respectively, capture the current domestic and foreign term structures. They are defined as

$$\gamma_i(t) = r_i(0) e^{-\kappa_i t} + \kappa_i \int_0^t e^{-\kappa_i (t-s)} \theta_i(s) \, ds, \quad i \in \{d, f\},$$  

(2)

where $\theta_i$ are deterministic, and represent the interest rates’ mean levels. The constants $\kappa_{\nu}$ and $\bar{\nu}$ are the mean-reversion rate and the long-term mean of the variance, respectively, and $\sigma_{\nu}$ is the constant instantaneous volatility of the variance process. Also, $W_s(t)$, $W_{d,i}(t)$, $W_{f,i}(t)$, and $W_{\nu}(t)$ are standard correlated Brownian motions (BMs) under measure $Q$ with correlation matrix $Q = [\rho]_{ij}$, i.e. $Q$ is symmetric positive semi-definite. Note that the “quanto” drift adjustments, $-\rho_{S,f} \sigma_f(t) \sqrt{\nu(t)}$, for $dY_i(t)$, $i = 1, \ldots, l$, come from changing from the foreign risk-neutral measure to the domestic risk-neutral one. We emphasize the generality of the cross-currency model (1a)-(1d), as well as its strong suitability for modeling FX products, especially long-dated (maturities of 30 years or more) hybrid FX products, such as Power-Reverse Dual-Currency (PRDC) swaps [2, 4].

3 Dimension reduction MC (drMC)

In this section, we briefly review the drMC approach. The reader is referred to [3] for a detailed proof of the approach and relevant discussions. One of the key steps in developing the drMC approach is to decide which factor we should condition on. Given the way that the model (1) is coupled, we can condition on either $r_d$ or $\nu$.\footnote{Note that, due to the quanto term, $\nu$ is in fact coupled to $r_f$, and hence we should not condition on $r_f$.} However, we choose to condition on $\nu$, since

1585
the Gaussian dynamics of \( r_d \) and \( r_f \) allow for exact computation of the expectation of the terms \( \exp\left(\int_0^t r_d(s)\,ds\right) \) and \( \exp\left(\int_0^t r_f(s)\,ds\right) \), as explained below, whereas the CIR dynamics of \( \nu \) do not. We decompose the BM processes into a linear combination of independent BM processes \( \tilde{W}_i(t), i = 1, \ldots, m + l + 2, \)

\[
(W_s(t), W_{d_1}(t), \ldots, W_{d_m}(t), W_{f_1}(t), \ldots, W_{f_l}(t), W_{\nu}(t))^\top = A \left( \tilde{W}_1(t), \tilde{W}_2(t), \ldots, \tilde{W}_{m+1}(t), \tilde{W}_{m+2}(t), \ldots, \tilde{W}_{m+l+1}(t), \tilde{W}_{m+l+2}(t) \right)^\top, \tag{3}
\]

where \( A \equiv \left[ a_{ij} \right]_{(m+l+2) \times (m+l+2)} \in \mathbb{R}^{(m+l+2) \times (m+l+2)} \), obtained using a Cholesky factorization, is an upper triangular matrix with \( a_{i,i}(m+l+2) = 1 \), and thus \( \tilde{W}_{m+l+2}(t) = W_{\nu}(t) \).

Let \( z(t) = \log(S(t)), 0 \leq t \leq T \). We denote by \( V(z(t), t, r_d(t), r_f(t), \nu(t)) \equiv V(z(t), t, \cdot) \) the price at time \( t \) of a plain-vanilla European option under the model (1), with payoff \( \Phi(S(T)) \). We also denote by \( \hat{V}(\xi, t, \cdot) \) and \( \hat{\Phi}(\cdot) \) the Fourier transforms of \( V(z(t), t, \cdot) \) and \( \Phi(z) \), respectively. For the rest of this paper, let the expectation and the variance operators be denoted by \( \mathbb{E}(\cdot) \) and \( \mathbb{V}(\cdot) \), respectively. Also, the subscript \( \{v_t\} \) is used to denote that the associated quantity is conditional on the \( \nu(t) \) path only, i.e. the quantity does not depend on the BM paths associated with \( S, r_d, \) or \( r_f \).

The first step in the drMC approach is to derive the the conditional PDE in terms of \( z \), which can be expressed in the following form:

\[
\partial_t u + \frac{a_{11}^2}{2} \nu(t) \partial_{zz} u + \left( \mu(t) - \frac{a_{11}^2}{2} \nu(t) \right) \partial_z u - r_d(t) u = 0. \tag{4}
\]

Here, the “drift” \( \mu(t) \) is a function of \( r_d(t), r_f(t), \nu(t) \) and the BMs \( \tilde{W}_k(t), k = 2, \ldots, m + l + 2 \). Then, by applying the Fourier transform on both sides of (4), we obtain an ordinary differential equation in \( \hat{u} \), the Fourier transform of \( u \), which can then be solved in closed-form. This allows us to write \( \hat{V}(\xi, 0, \cdot) \) as

\[
\hat{V}(\xi, 0, \cdot) = \mathbb{E} \left[ \hat{\Phi}(\xi) \exp \left( \sum_{k=2}^{m+l+1} \psi_k \left( \xi; \{\tilde{W}_k(t)\}, \{\tilde{W}_{m+l+2}(t)\} \right) \right) \right]. \tag{5}
\]

Here, the quantities \( \psi_k \left( \xi; \{\tilde{W}_k(t)\}, \{\tilde{W}_{m+l+2}(t)\} \right) \equiv \psi_k(\xi; \cdot), k = 1, \ldots, m + l + 1 \), depend only on the independent BMs \( \tilde{W}_k(t) \) and \( \tilde{W}_{m+l+2}(t) \), and the quantity \( \psi_{m+l+2} \left( \xi; \{\tilde{W}_{m+l+2}(t)\} \right) \equiv \psi_{m+l+2}(\xi; \cdot) \) only depends on the BM \( \tilde{W}_{m+l+2}(t) \). This indicates that, if we condition on \( \tilde{W}_{m+l+2}(t) \), then \( \psi_k(\xi; \cdot), k = 1, \ldots, m + l + 1 \), are independent, and that \( \psi_{m+l+2}(\xi; \cdot) \) is a constant.

Using the above fact and iterated conditioning, we have

\[
\hat{V}(\xi, 0, \cdot) = \mathbb{E} \left[ \hat{\Phi}(\xi) \exp \left( \psi_{m+l+2}(\xi; \cdot) \right) \prod_{k=2}^{m+l+1} \mathbb{E} \left[ \exp \left( \psi_k(\xi; \cdot) \right) \mid \mathcal{F}_0, \{\tilde{W}_{m+l+2}(t)\} \right] \right]. \tag{6}
\]

Conditional on \( \tilde{W}_{m+l+2}(t) \), each of \( \psi_k(\xi; \cdot), k = 2, \ldots, m + l + 1 \) is, in fact, Gaussian, due to the Gaussian dynamics of \( r_d \) and \( r_f \). Thus, using the fact that, for a random variable \( X \sim \text{Normal}(0, \sigma^2) \) and \( y \in \mathbb{R}, \mathbb{E}[e^{yX}] = e^{\frac{1}{2}y^2\sigma^2} \), we can write (6) in the following form

\[
\hat{V}(\xi, 0, \cdot) = \mathbb{E} \left[ \hat{\Phi}(\xi) \exp \left( -G_{\{\nu\}} \xi^2 + iF_{\{\nu\}} \xi + H_{\{\nu\}} \right) \right]. \tag{7}
\]
Finally, reduced to the multilevel dimension and variance reduction Monte-Carlo, the linearity of expectation and Fourier transform, we obtain
\[ F_{\{\nu_i\}} = \frac{1}{2} \int_0^T \nu(t) \, dt + \int_0^T (\gamma_d(t) - \gamma_f(t)) \, dt - \sum_{k=2}^{l+m+1} \int_0^T \left( \sum_{j=1}^m a_{(j+1),k} \beta_d(t) \left( \sum_{j=1}^m a_{(j+1),k} \beta_d(t) - \sum_{j=1}^l a_{(j+m+1),k} \beta_f(t) \right) \right) \, dt + \sum_{j=1}^m a_{(j+1),(m+l+2)} \int_0^T \beta_d(t) \, dW_\nu(t) - \sum_{j=1}^l \int_0^T \beta_f(t) \, dW_\nu(t) \]
\[ + a_{1,(m+l+2)} \int_0^T \sqrt{\nu(t)} \, dW_\nu(t) + \sum_{j=1}^l \rho_{S,f} \int_0^T \beta_f(t) \sqrt{\nu(t)} \, dt \]
\[ - \sum_{k=2}^{l+m+1} \sum_{j=1}^m a_{1,k} a_{(j+1),k} \beta_d(t) \sqrt{\nu(t)} \, dt , \]
\[ G_{\{\nu_i\}} = \frac{a^2_{\{1\}}}{2} \int_0^T \nu(t) \, dt + \frac{1}{2} \sum_{k=2}^{m+l+1} \int_0^T \left( \sum_{j=1}^m a_{(j+1),k} \beta_d(t) - \sum_{j=1}^l a_{(j+m+1),k} \beta_f(t) + a_{1,k} \sqrt{\nu(t)} \right)^2 \, dt , \]
\[ H_{\{\nu_i\}} = -\sum_{j=1}^m a_{(j+1),(m+l+2)} \int_0^T \beta_d(t) \, dW_\nu(t) - \int_0^T \gamma_d(t) \, dt + \frac{1}{2} \sum_{k=2}^{l+m+1} \int_0^T \left( \sum_{j=1}^m a_{(j+1),k} \beta_d(t) \right)^2 \, dt . \]

In (9), \( \beta_d(t), i = 1, \ldots, m, \) and \( \beta_f(t), i = 1, \ldots, l, \) are defined as
\[ \beta_d(t) = \sigma_d(t) \int_t^T e^{-\int_t^{t'} \kappa_d(t'') \, dt''} \, dt' , \quad \beta_f(t) = \sigma_f(t) \int_t^T e^{-\int_t^{t'} \kappa_f(t'') \, dt''} \, dt' . \]

Finally, applying the inverse Fourier transform to both sides of (7), together with the convolution theorem, along with the linearity of expectation and Fourier transform, we obtain
\[ V(z(0), 0, r_d(0), r_f(0), \nu(0)) = \mathbb{E} \left[ \int_{-\infty}^{+\infty} \Phi(z(0) - \bar{z}) \sqrt{4\pi G_{\{\nu_i\}}} \exp \left( H_{\{\nu_i\}} - \frac{(\bar{z} + F_{\{\nu_i\}})^2}{4G_{\{\nu_i\}}} \right) \, d\bar{z} \right] . \]

Equation (11) is the main result of [3].

As an illustrative example, in this paper, we consider a standard put option with the payoff function defined by \( \Phi(z) = (K - e^z)^+ \), where \( K \) is the strike price. In this case, (11) is further reduced to
\[ V(z(0), 0, r_d(0), r_f(0), \nu(0)) = \mathbb{E} \left[ e^{H_{\{\nu_i\}}} \left( K \left( 1 - \text{erf} \left( \frac{F_{\{\nu_i\}} + \log\left( \frac{S(0)}{K} \right)}{2\sqrt{G_{\{\nu_i\}}} } \right) \right) \right) - S(0) e^{F_{\{\nu_i\}} + G_{\{\nu_i\}}} \left( 1 - \text{erf} \left( \frac{F_{\{\nu_i\}} + 2G_{\{\nu_i\}} + \log\left( \frac{S(0)}{K} \right)}{2\sqrt{G_{\{\nu_i\}}} } \right) \right) \right] , \]
\[ (12) \]
where \( \text{erf}(\cdot) \) is the error function. For use later in the paper, we denote by \( P \) the quantity inside the expectation formula (12). That is

\[
P \equiv P_{\nu_{n}} = \frac{e^{H_{\nu_{n}}}}{2} \left( K \left( 1 - \text{erf} \left( \frac{F_{\nu_{n}} + \log\left( \frac{S_{0}}{K} \right)}{2 \sqrt{G_{\nu_{n}}}} \right) \right) 
- S_{0} e^{F_{\nu_{n}} + G_{\nu_{n}}} \left( 1 - \text{erf} \left( \frac{F_{\nu_{n}} + 2G_{\nu_{n}} + \log\left( \frac{S_{0}}{K} \right)}{2 \sqrt{G_{\nu_{n}}}} \right) \right) \right)
\]

(13)

It is important to note that, due to the dependence of \( F_{\nu_{n}}, G_{\nu_{n}}, \) and \( H_{\nu_{n}} \) on the variance paths, \( P \) is indeed a (non-conventional) path-dependent payoff. As noted in [6], construction of the multilevel estimator for path-dependent options requires special attentions in order to achieve improved efficiency in the case of the Milstein discretization. We discuss this in the next section.

4 Multilevel drMC (ml-drMC)

In the ml-drMC approach, we apply the multilevel technique to the variance factor \( \nu(t) \), which is driven by the BM \( \tilde{W}_{m} \). For simplicity, for the rest of the paper, let \( W(t) \equiv \tilde{W}_{m+2} \). For the simulation of \( \nu(t) \), we use a drift-implicit Milstein scheme that preserves the positivity of the original CIR model (1d), and has a good strong convergence property, recently established in [9]. More specifically, given a timestep size \( h = T/N \), the Milstein discretization of (1d) is

\[
\begin{align*}
\hat{\nu}_{n+1} &= \hat{\nu}_{n} + \kappa (\bar{\nu} - \hat{\nu}_{n+1}) h + \sigma \sqrt{\hat{\nu}_{n}^{+}} \Delta W_{n} + 0.25 \sigma^{2} ( (\Delta W_{n})^{2} - h ) \\
\Rightarrow \hat{\nu}_{n+1} &= \frac{\hat{\nu}_{n} + \kappa \nu h + \sigma \sqrt{\hat{\nu}_{n}^{+}} \Delta W_{n} + 0.25 \sigma^{2} ( (\Delta W_{n})^{2} - h )}{1 + h \kappa},
\end{align*}
\]

(14)

where \( \hat{\nu}_{n} \) denotes the discrete approximation at \( t_{n} = nh, n = 0, \ldots, N-1 \), \( \Delta W_{n} = W(t_{n+1}) - W(t_{n}) = \text{Normal}(0, h) \), and \( \hat{\nu}_{n}^{+} = \max(\hat{\nu}_{n}, 0) \).

Consider multiple sets of simulations of \( \nu(t) \) with different timesteps sizes \( h = h_{\ell} = \frac{T}{\tau \ell}, \ell = 0, \ldots, L \), and \( \tau \) the refinement factor, so level \( \ell \) has \( \tau \) times more timesteps than level \( \ell - 1 \). In all of our experiments, we use the refinement factor \( \tau = 2 \). For a given BM path \( W(t) \), we denote by \( \hat{P}_{\ell}, \ell = 0, \ldots, L \), an approximation to the payoff \( P \), defined in (13), using the discretization scheme (14) with timestep size \( h_{\ell} \). Note the key identity underlying the mlMC method

\[
E(\hat{P}_{L}) = E(\hat{P}_{0}) + \sum_{\ell=1}^{L} E[\hat{P}_{\ell} - \hat{P}_{\ell-1}].
\]

(15)

We denote by \( \hat{Y}_{0} \) an estimator for \( E(\hat{P}_{0}) \), and by \( \hat{Y}_{\ell}, \ell = 1, \ldots, L \), an estimator for \( E[\hat{P}_{\ell} - \hat{P}_{\ell-1}] \) using \( M_{\ell} \) simulation paths. In the simplest scheme, the estimator \( \hat{Y}_{\ell} \) is a mean of \( M_{\ell} \) paths, i.e.

\[
\hat{Y}_{\ell} = \frac{1}{M_{\ell}} \sum_{m=1}^{M_{\ell}} (\hat{P}_{\ell}^{(m)} - \hat{P}_{\ell-1}^{(m)}).
\]

(16)

A key point in the mlMC approach is that the quantity \( \hat{P}_{\ell}^{(m)} - \hat{P}_{\ell-1}^{(m)} \) comes from two discrete approximations with different timestep sizes, but are based on the same BM path. We denote by \( \hat{Y} \) the combined estimator, defined as \( \hat{Y} = \sum_{\ell=0}^{L} \hat{Y}_{\ell} \). The idea of mlMC is to independently
estimate each $\hat{Y}_\ell$, $\ell = 1, \ldots, L$, in such a way that, for a given computational cost, the variance of the combined estimator, namely $\mathbb{V}(Y)$, is minimized. As showed in [5], this can be achieved by choosing $M_\ell$ proportional to $\sqrt{\nu(h)}$. Thus, the convergence of the sample variance $V_\ell$ as $\ell \to \infty$ is very important to the efficiency of the methods, since it determines an optimal choice of $M_\ell$, the number of sample paths used in each level. In the case of Asian options, which have path-dependent payoffs, it has been shown in [6] that $V_\ell = \mathcal{O}(h^2)$ when the Milstein discretization method is used, but special treatments to the estimator must be applied.

For the rest of the paper, the super-scripts “f” and “c” are used to denote the dependence of the quantities on fine and coarse levels, respectively. (This is not to be confused with the sub-script “f” used to indicate association with the “foreign” interest rate factor.) We now discuss how to construct the estimator $\hat{Y}_\ell$ for $\mathbb{E}[\hat{P}_\ell - \hat{P}_{\ell-1}]$ defined in (16). The estimator $\hat{Y}_\ell$ is an average of $M_\ell$ values from $M_\ell$ independent path simulations. For each BM $W(t)$ input, the value is of the form $\hat{P}_\ell^f - \hat{P}_{\ell-1}^c$, where $\hat{P}_\ell^f$ is a fine-path estimate using timestep size $h_\ell = T/2^\ell$, and $\hat{P}_{\ell-1}^c$ is the corresponding coarse-path estimate using timestep size $h_\ell = T/2^{\ell-1}$. As showed in [6], it is required that $\mathbb{E}[\hat{P}_{\ell-1}^c] = \mathbb{E}[\hat{P}_{\ell-1}^f]$, so that the key identity (15) is satisfied. In the case of European options, we can simply define $\hat{P}_{\ell-1}^c$ and $\hat{P}_{\ell-1}^f$ to be the same. In our case, however, due to the path-dependency of the payoff $P$, the definition of $\hat{P}_{\ell-1}^c$ needs information from the discrete approximation of $\hat{P}_\ell^f$, which is not available when $\hat{P}_{\ell-1}^c$ is computed. This is done in order to reduce the variance of the estimator, but the relation $\mathbb{E}[\hat{P}_{\ell-1}^c] = \mathbb{E}[\hat{P}_{\ell-1}^f]$ must be satisfied. Similar observations are made in [6] in the context of Asian options.

Now recall the coefficients $F^{(v)}_{\{v\}}$, $G^{(v)}_{\{v\}}$, and $H^{(v)}_{\{v\}}$ of the payoff $P$, which are given in (9). To compute these coefficients, we need to evaluate the following stochastic integrals: $\int_0^T \nu(t)dt$, $\int_0^T \sqrt{\nu(t)}dW(t)$, $\int_0^T \beta_i(t)\sqrt{\nu(t)}dt$, and $\int_0^T \beta_i(t)dW(t)$, $i = 1, \ldots, m$; $\int_0^T \beta_f(t)\sqrt{\nu(t)}dt$, and $\int_0^T \beta_f(t)dW(t)$, $i = 1, \ldots, l$. First, we consider approximating $\int_0^T \nu(t)dt$ on a fine-path using $h_\ell = T/N_\ell$, where $N_\ell = 2^\ell$. The treatment that we use here is similar to the treatment proposed in [6] in the context of Asian options, which is based on Brownian interpolation. More specifically, we use the trapezoidal integration scheme

$$\int_0^T \nu(t)dt = \sum_{n=0}^{N_\ell-1} \frac{h_\ell}{2} \left( \nu_n^f + \nu_{n+1}^f \right),$$

(17)

which can be viewed as averages of a piecewise linear interpolation. (The above trapezoidal integration scheme is shown to result in $V_\ell = \mathcal{O}(h^2)$ in the context of arithmetic average Asian options [6]).

Regarding $\int_0^T \sqrt{\nu(t)}dW(t)$, we proceed as follows. By first integrating (1d) from $t_n$ to $t_n + h$ for $\nu(t)$, and then rearranging, we obtain

$$\int_{t_n}^{t_n+h} \sqrt{\nu(t)}dW(t) = \nu(t_n) - \nu(t_n) - \kappa \nu h + \kappa \int_{t_n}^{t_n+h} \nu(t)dt.$$

(18)

Thus, (17) and (18) gives rise to the follow approximation scheme for $\int_0^T \sqrt{\nu(t)}dW(t)$

$$\int_0^T \sqrt{\nu(t)}dW(t) \approx \frac{\nu_{N_\ell}^f - \nu(0) - \kappa h + \kappa \sum_{n=0}^{N_\ell-1} \frac{h_\ell}{2} \left( \nu_n^f + \nu_{n+1}^f \right)}{\sigma}.$$  

(19)

For the rest of the stochastic integrals, namely $\int_0^T \beta_f(t)\sqrt{\nu(t)}dt$, $\int_0^T \beta_f(t)dW(t)$, $i = 1, \ldots, m$, $\int_0^T \beta_i(t)\sqrt{\nu(t)}dt$, $\int_0^T \beta_i(t)dW(t)$, $i = 1, \ldots, l$, we use a
trapezoidal integration rule, similar to (17). The respective quantities for the coarse-path are constructed similarly. Once these stochastic integral for the fine- and coarse-paths, are computed, it is straightforward to compute the respective coefficients \( F_i \), \( G_i \), and \( H_i \), where \( i = \{ f, c \} \). Then, \( \hat{P}^i_f \) and \( \hat{P}^i_c \) can be computed.

5 Numerical results

For the numerical experiments, we consider a 6-D cross-currency model, i.e. model (1) with \( m = l = 2 \). For simplicity, for the domestic and foreign two-factor HW models, we assume \( \theta_d \) and \( \theta_f \) in (2)) are constant. Thus all the deterministic integrals in \( G_{\{v_i\}} \), \( F_{\{v_i\}} \) and \( H_{\{v_i\}} \) can be computed analytically. The quantities \( G_{\{v_i\}} \), \( F_{\{v_i\}} \) and \( H_{\{v_i\}} \) defined in (9) can further be reduced. For brevity, we omit these reduced formulas, which can be found in [3].

For comparison purposes, we also implement an antithetic mlMC method combined with a Milstein discretization scheme for the above 6D model, as developed in [7]. We refer to this method as anti-mlMC. For this method, due to the non-linearity of the diffusion coefficient in the price process \( S(t) \), we work with \( \log(S(t)) \) instead [7]. Following [7], given a timestep size \( h = T/N \), the Milstein scheme for the 6-D cross-currency model under consideration with the Lévy area terms set to zero is given by:

\[
\lambda(S_{n+1}) = \lambda(S_n) + (\hat{r}_{d,n} - \hat{r}_{f,n} - 0.5\hat{\nu}_n)n + \sqrt{\hat{\nu}_n} \Delta W_{n,n} + 0.5\hat{\nu}_n ((\Delta W_{S,n})^2 - h) + 0.25\sigma_n((\Delta W_{S,n})\Delta W_{\nu,n} - \rho_{S,\nu} h),
\]

\[
\hat{r}_{d,n+1} = \sum_{i=1}^{2} X_{i,n} + \gamma_{d,n}, \quad \hat{X}_{i,n+1} = \hat{X}_{i,n} - \kappa_d \hat{X}_{i,n}h + \sigma_d \Delta W_{d_i,n}, \quad \hat{X}_{i,0} = 0, i = 1, 2,
\]

\[
\hat{r}_{f,n+1} = \sum_{i=1}^{2} \hat{Y}_{i,n} + \gamma_{f,n}, \quad \hat{Y}_{i,n+1} = \hat{Y}_{i,n} - (\kappa_f \hat{Y}_{i,n} + \rho_{S,f} \hat{Y}_{i,n} \sqrt{\hat{\nu}_n})h + \sigma_f \Delta W_{f_i,n},
\]

\[
\hat{Y}_{i,0} = 0, i = 1, 2,
\]

\[
\hat{\nu}_{n+1} = \frac{\hat{\nu}_n + \kappa \hat{\nu} h + \sigma \sqrt{\hat{\nu}_n} \Delta W_{n,n} + 0.25 \sigma^2 ((\Delta W_{\nu,n})^2 - h)}{1 + h \kappa}.
\]

Here, \( \gamma_{i,n}(t) = (r_i,0 - \theta_i)e(-\kappa_i n h) + \theta_i, \quad i \in \{ d, f \} \). Details of the antithetic mlMC technique for multi-dimensional problems discretized by the Milstein scheme, such as (20), are discussed in [7], and hence omitted here. We note that, although the coefficients of the variance process are not Lipschitz continuous, and hence the assumptions in [7] are not satisfied, the numerical tests show that the anti-mlMC performs well, and is able to achieve \( \lambda \hat{\nu} = O(h^2) \). Similar convergence results are reported in [7] for the Heston model.

For the numerical experiments, we use the domestic and foreign term structures in [10]; \( r_d(0) = 0.02, \kappa_{d1} = 0.03, \kappa_{d2} = 0.03, \sigma_{d1} = 0.03, \sigma_{d2} = 0.03, \theta_d = 0.02, \) and \( r_f(0) = 0.05, \kappa_{f1} = 0.03, \kappa_{f2} = 0.03, \sigma_{f1} = 0.012, \sigma_{f2} = 0.012, \) and \( \theta_f = 0.05 \). The correlations are \( \rho_{S,d1} = 0.08, \rho_{S,d2} = 0.08, \rho_{S,f1} = 0.08, \rho_{S,f2} = 0.08, \rho_{S,\nu} = -0.02, \rho_{d2,\nu} = 0.12, \rho_{d1,f1} = 0.12, \rho_{d1,f2} = 0.12, \rho_{d2,f1} = 0.12, \rho_{d2,f2} = 0.12, \rho_{d2,\nu} = 0.15, \rho_{f1,\nu} = -0.70, \rho_{f1,\nu} = -0.70, \rho_{f2,\nu} = 0.15, \rho_{f2,\nu} = 0.15 \). We also use \( S(0) = 10, K = 10 \). For the variance, we use \( \kappa_{\nu} = 0.5, \nu(0) = 0.9, \sigma_{\nu} = 0.05, \nu(0) = 0.9, \) which are taken from [7]. We use long maturity with \( T = 20 \) (years), which is highly challenging for practical applications.

In our subsequent discussions, we compare three MC methods, namely ml-drMC, drMC, anti-mlMC. Here, drMC is the non-multilevel counterpart of ml-drMC. In addition, following [5], we define the computational cost for each method as the total number of random numbers.
generated. More specifically, for each of the multilevel based methods, namely ml-drMC and anti-mlMC, the computational cost is computed as

\[ M_0 + \sum_{\ell=1}^{L} M_\ell (N_\ell + N_{\ell-1}) \]

where \( L \) is the number of levels required by the method, \( M_\ell \) and \( N_\ell = 2^\ell \) respectively are the number of samples required by the method, and number of timesteps of level \( \ell, \ell = 0, \ldots, L \). The computational cost of the non-multilevel method, namely drMC, is computed as \( \sum_{\ell=0}^{L} M_*^\ell N_\ell \), where \( M_*^\ell = 2\epsilon^{-2}V[\hat{P}_\ell] \), so that the variance bound is also \( \epsilon^2 / 2 \) as with its multilevel counterpart [5].

In Table 1, to illustrate the accuracy of the ml-drMC method, we present the option prices obtained by these three methods, and the corresponding standard derivation for the case \( \epsilon = 10^{-3} \). We observed that the option prices obtained by all methods agree well.

<table>
<thead>
<tr>
<th>Method</th>
<th>Option Price</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ml-drMC</td>
<td>12.563512</td>
<td></td>
</tr>
<tr>
<td>dr-MC</td>
<td>12.563405</td>
<td></td>
</tr>
<tr>
<td>anti-mlMC</td>
<td>12.563221</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Option prices obtained by different methods when \( \epsilon = 10^{-3} \).

In Figure 1 (a), we investigate the convergence behavior of the variance \( \mathbb{V}_\ell = \mathbb{V}[\hat{P}_\ell - \hat{P}_{\ell-1}] \) as a function of the level of approximation when \( \epsilon = 10^{-3} \). These values were estimated using \( 10^6 \) samples, so the sampling error is negligible. We make following observations. Both multilevel-based MC methods, namely ml-drMC and anti-mlMC, result in lines having slope -2, which indicates that \( \mathbb{V}_\ell = O(h_\ell^2) \). This indicates that the improved efficiency expected from Milstein scheme has been achieved for the ml-drMC method. Moreover, the \( \mathbb{V}_\ell \) of the ml-drMC method is about 50 times smaller than the \( \mathbb{V}_\ell \) of the anti-mlMC method. This is expected, due to the (possible) variance reduction in the drMC approach. We also note that the multilevel-based methods are substantially more accurate than their non-multilevel-based counterparts. On level \( \ell = 2 \), which has just 4 timesteps, \( \mathbb{V}_\ell \) of ml-drMC is already more than 1000 times smaller than the variance of drMC (which is essentially \( \mathbb{V}[\hat{P}_\ell] \)).

In Figure 1 (b), we investigate the computational costs of the three methods. We compute the cost required by each method for different values of \( \epsilon \). It is clear that the ml-drMC are significantly more efficient than the drMC and anti-mlMC methods. In particular, the ml-drMC
method is about 80 times more efficient than the anti-mlMC method. The results from these two figures indicate that ml-drMC can achieve the same second-order rate of convergence as the anti-mlMC method of [7], but significantly more efficiently.

6 Conclusion

In this paper, we develop a highly efficient multilevel and dimension reduction MC method for pricing plain-vanilla European options under multi-dimensional one-way coupled models in finance. The method is based on two steps. First, by applying the drMC method, we can reduce the dimension of an $N$-dim problem from $N$ to 1. Then, we apply the multilevel technique with the Milstein discretization on the stochastic factor we condition on, which is essentially an application of the multilevel technique on a 1-dim problem. We have demonstrated numerically that it is possible to achieve a second-order rate of convergence for the multilevel estimator with Milstein discretization in the context of the drMC method. In addition, while having the same rate of convergence as the antithetic mlMC of [7], the ml-drMC is significantly more efficient.

There are several major directions for future research. The first is to establish the theoretical analysis of the ml-drMC method presented in this paper for European options, and possibly for discontinuous payoffs, such as those arise that in digital options. The second direction is to extend the ml-drMC approach to include treatment of exotic features, such as early exercise and barriers. The third direction is to handle model extensions, such as Lévy jump-diffusion models. While mlMC methods for jump-diffusion models are available in the literature, an important advantage of the ml-drMC over these mlMC methods is that the jumps can be handled separately and very efficiently, via the Fourier transform technique applied to the conditional Partial Integral-Differential Equations.

References