Received: 13 June 2019 | Revised: 14 November 2019 | Accepted: 26 November 2019

DOI: 10.1002/nla.2281

SPECIAL ISSUE PAPER

WILEY

Enhanced multi-index Monte Carlo by means of multiple semicoarsened multigrid for anisotropic diffusion problems

Pieterjan Robbe | **Dirk Nuyens** | **Stefan Vandewalle**

Department of Computer Science, KU Leuven, Leuven, Belgium

Correspondence
Pieterjan Robbe, Department of Computer Science, Celestijnenlaan 200A, Box 2402, 3001 Leuven, Belgium.
Email: pieterjan.robbe@kuleuven.be

Funding information
Fonds Wetenschappelijk Onderzoek, Grant/Award Number: IWT-140068

Summary

In many models used in engineering and science, material properties are uncertain or spatially varying. For example, in geophysics and porous media flow, in particular, the uncertain permeability of the material is modeled as a random field. These random fields can be highly anisotropic. Efficient solvers, such as the multiple semicoarsened multigrid (MSG) method are required to compute solutions for various realizations of the uncertain material. The MSG method is an extension of the classic multigrid method, which uses additional coarse grids that are coarsened in only a single coordinate direction. In this sense, it closely resembles the extension of multilevel Monte Carlo to multi-index Monte Carlo (MIMC). We present an unbiased MIMC method that reuses the MSG coarse solutions. Our formulation of the estimator can be interpreted as the problem of learning the unknown distribution of the number of samples across all indices and unifies the previous work on adaptive MIMC and unbiased estimation. We analyze the cost of this new estimator theoretically and present numerical experiments with various anisotropic random fields, where the unknown coefficients in the covariance model are considered as hyperparameters. We illustrate its robustness and superiority over unbiased MIMC without sample reuse.

KEYWORDS

anisotropic diffusion problems, multi-index Monte Carlo, multiple semicoarsened multigrid

1 | INTRODUCTION

Consider the elliptic PDE with random coefficients

$$\begin{aligned} -\nabla \cdot (a(\boldsymbol{x},\omega)\nabla u(\boldsymbol{x},\omega)) &= h(\boldsymbol{x}) \quad && \text{with } \boldsymbol{x} \in D, \omega \in \Omega, \text{and} \\ u(\boldsymbol{x},\omega) &= 0 && \text{on } \partial D, \end{aligned} \tag{1}$$

where D is a bounded domain in $\mathbb{R}^d$, $d = 1, 2, 3$, with boundary ∂D, and Ω is the sample space of a probability space $(\Omega, \mathcal{F}, P)$. We assume that the diffusion coefficient $a : D \times \Omega \to \mathbb{R}$ is given as a lognormal random field, that is, $a(\boldsymbol{x}, \omega) = \exp(Z(\boldsymbol{x}, \omega))$, where Z is a zero-mean Gaussian random field with given covariance function. In this article, we consider the so-called Matérn covariance function[9]

$$C(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{2^{1-\nu}}{\Gamma(\nu)}(2\sqrt{\nu}\, r(\boldsymbol{x}_1, \boldsymbol{x}_2))^{\nu} K_{\nu}(2\sqrt{\nu}\, r(\boldsymbol{x}_1, \boldsymbol{x}_2)), \tag{2}$$

with smoothness parameter ν, and where the distance function $r(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is given by

$$r(\boldsymbol{x}_1, \boldsymbol{x}_2) = \|T(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2, \tag{3}$$

with T a suitable transformation matrix. In what follows, we are interested in the two-dimensional case ($d = 2$), and assume the linear transformation matrix T is a concatenation of a scaling and a rotation, that is,

$$T = \sqrt{\Lambda} R \text{ with } \quad \sqrt{\Lambda} = \begin{pmatrix} 1/\sqrt{\eta\lambda} & 0 \\ 0 & 1/\sqrt{\lambda} \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \tag{4}$$

where λ is the length scale, η is the anisotropic ratio, and θ is the rotation angle. Some realizations of the random field $Z(\boldsymbol{x}, \omega)$, for relevant combinations of the parameters in Matérn covariance model (2), are shown in Figure 1.

These problems find their application in groundwater hydrology, where Equation (1) describes the steady-state single-phase flow through a porous medium.[10] In this setting, $a(\boldsymbol{x}, \omega)$ is a random field representing the permeability of the porous medium. In stochastic modeling of groundwater flow, it is well known that a lognormal random field may accurately represent the permeability of a naturally occurring porous medium.[11] Typical values of interest for the Matérn

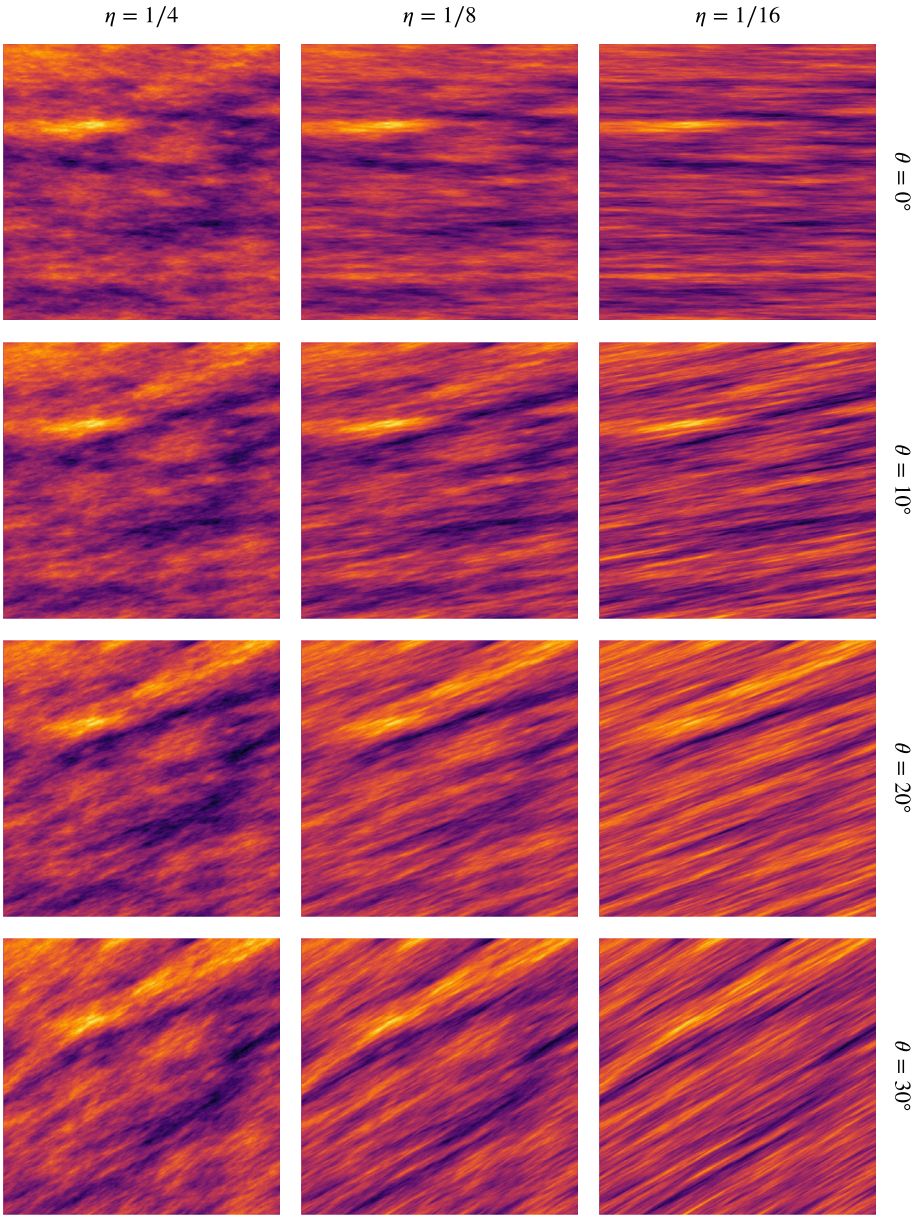

FIGURE 1 Realizations of a Gaussian random field with Matérn covariance function with parameters $\lambda = 1/4$, $\nu = 1/2$, $\eta \in \{1/4, 1/8, 1/16\}$ (left to right) and $\theta \in \{0°, 10°, 20°, 30°\}$ (top to bottom)

covariance parameters in hydrology applications are a smoothness $\nu < 1$, length scale $\lambda \ll \text{size}(D)$, and anisotropic ratio $\eta \ll 1$. The solution $u(\boldsymbol{x}, \omega)$ of Equation (1) is a random field that represents the fluid pressure at equilibrium.

Depending on the characteristics of the random field, and the way it is represented, many uncertain parameters may be required to accurately model its variability. Sampling-based methods, such as the Monte Carlo (MC) method, are the preferred tool to deal with these many uncertainties. They require the solution of a deterministic PDE for every realization $\omega_i \in \Omega, i = 1, \ldots, N$, with N the number of samples. From this sample set, conclusions are drawn about the statistics of a quantity of interest $Q = F(u(\boldsymbol{x}, \omega))$, such as the expected value, variance, or higher order moments. Here, F is a functional applied to the solution $u(\boldsymbol{x}, \omega)$ of the PDE, for example, a point evaluation or an average over a subdomain. The classic MC method is often viewed as impractical due to the large number of expensive realizations required. It is a well known but notorious fact that the root mean square error decays as slowly as $1/\sqrt{N}$ for estimating an expected value, nevertheless independent of the number of uncertain parameters.

So-called multilevel methods have been proposed to lower the MC cost, using samples of the PDE on coarser grids. These methods are known as multilevel Monte Carlo (MLMC) methods (see Giles[4]) or multi-index Monte Carlo (MIMC) methods (see Haji-Ali et al.[5]). When using a particular iterative method, called full multigrid (FMG), to solve the deterministic PDE underlying every sample of Equation (1), coarse solutions returned by the solver can be reused as samples in the multilevel estimator. This idea was first proposed in the context of MLMC by Kumar et al.,[6] and, in the present work, we would like to extend this concept to the MIMC setting.

The problem at hand can be decomposed into two different subproblems that will be addressed accordingly in the remainder of this article.

a. We require efficient and robust deterministic multilevel solvers that compute a solution of problem (1) for every possible realization $\omega \in \Omega$.
b. We require more efficient multilevel methods to compute statistics of a quantity of interest derived from the solution of Equation (1) that overcome the inefficiency of standard MC.

2 | MULTIPLE SEMICOARSENED MULTIGRID

The convergence rate of multigrid (MG) algorithms based on point relaxation smoothers, such as damped Jacobi or Gauss-Seidel, degenerates on problems with strong anisotropies (see Hackbusch[12] or Trottenberg et al.[13]) Robustness can be improved by using a more powerful smoother, such as an incomplete LU (ILU)-type smoother or line ($d = 2$) or plane ($d = 3$) smoothers. However, these smoothers are typically more expensive, and implementation on parallel machines is nontrivial.[14] An alternative way to recover good multigrid convergence rates, which avoids line and plane relaxations altogether, is to use multiple coarse grids formed by semicoarsening in each of the coordinate directions. This method is known as multiple semicoarsened multigrid (MSG). Although first described as a nonlinear full-approximation scheme (FAS)-type algorithm by Mulder,[1] a linear MSG correction scheme was introduced and analyzed in Naik and Van Rosendale[2] and Oosterlee and Wesseling.[3] See Figure 2 for a graphical comparison of standard and semicoarsened coarse grids. In this article, we use MSG for robustness with respect to the random input diffusion coefficient.

We suppose, for simplicity, that the domain of the model problem is the unit square, that is, $D = [0, 1]^2$. Define a sequence of coarse grids as

$$G^{(p,q)} = \{(i \cdot \delta x, j \cdot \delta y) : i = 0, 1, \ldots, 2^p, \; j = 0, 1, \ldots, 2^q\}, \tag{5}$$

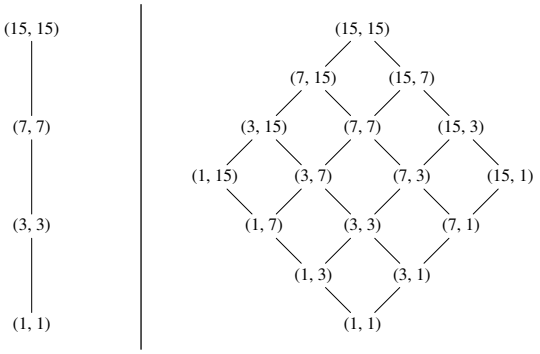

FIGURE 2 A comparison of standard coarse grids (left) and multiple semicoarsened grids (right) for $d = 2$. Number of degrees of freedom in each coordinate direction for $\overline{p} = \overline{q} = 4$

with $\delta x = 2^{-p}$ and $\delta y = 2^{-q}$, $p = 1, \ldots, \overline{p}$, and $q = 1, \ldots, \overline{q}$. The coarsest grid is thus $G^{(1,1)}$, and the finest grid is $G^{(\overline{p},\overline{q})}$. Note that a standard coarsening uses only grids with values $p = q$. On this hierarchy of grids, we discretize (1) using finite differences, resulting in a discrete version of the PDE,

$$A^{(p,q)}\boldsymbol{u}^{(p,q)} = \boldsymbol{b}^{(p,q)}, \tag{6}$$

on grid $G^{(p,q)}$. Important for the remainder of this article is that the coarse matrices $A^{(p,q)}$ are obtained from direct discretization of the differential equation and not in variational form (as based on the Galerkin condition).

We now assume that we have, at our disposal, intergrid transfer operators between the different grids, defined in the following way:

$$\mathcal{R}_x^{(p,q)} : G^{(p+1,q)} \mapsto G^{(p,q)} \quad \text{and} \quad \mathcal{R}_y^{(p,q)} : G^{(p,q+1)} \mapsto G^{(p,q)} \quad \text{(restriction), and} \tag{7}$$

$$\mathcal{P}_x^{(p,q)} : G^{(p-1,q)} \mapsto G^{(p,q)} \quad \text{and} \quad \mathcal{P}_y^{(p,q)} : G^{(p,q-1)} \mapsto G^{(p,q)} \quad \text{(prolongation).} \tag{8}$$

These transfer operators are simple one-dimensional operators. For example, the restriction operator can be chosen as the well-known full-weighting procedure. In stencil notation, these operators are written as

$$\mathcal{R}_x^{(p,q)} = \frac{1}{4}\begin{bmatrix}1 & 2 & 1\end{bmatrix}_{p+1,q}^{p,q} \quad \text{and} \quad \mathcal{R}_y^{(p,q)} = \frac{1}{4}\begin{bmatrix}1\\2\\1\end{bmatrix}_{p,q+1}^{p,q}. \tag{9}$$

Similarly, the prolongation operators $\mathcal{P}_x^{(p,q)}$ and $\mathcal{P}_y^{(p,q)}$ could be chosen as the usual linear interpolation operators.

Algorithm 1. Multiple semicoarsened multigrid μ-cycle (recursive definition)

input: level parameter L
output: approximations $\boldsymbol{v}^{(p,q)}$ for the solution of $A^{(p,q)}\boldsymbol{u}^{(p,q)} = \boldsymbol{b}^{(p,q)}, p + q = L$

procedure MSG-μ-CYCLE(L)
  if $L = 0$ **then**
    $\boldsymbol{v}^{(0,0)} \leftarrow \boldsymbol{b}^{(0,0)}/A^{(0,0)}$
  else
    for each grid $G^{(p,q)}$ **where** $p + q = L$ **do**
      call SMOOTH($\boldsymbol{v}^{(p,q)}, A^{(p,q)}, \boldsymbol{b}^{(p,q)}$) ν_1 times
      $\boldsymbol{r}^{(p,q)} \leftarrow \boldsymbol{b}^{(p,q)} - A^{(p,q)}\boldsymbol{v}^{(p,q)}$
    end for
    for each grid $G^{(p,q)}$ **where** $p + q = L - 1$ **do**
      $\boldsymbol{b}^{(p,q)} \leftarrow$ the restricted residual from (10)
      $\boldsymbol{v}^{(p,q)} \leftarrow \boldsymbol{0}$
    end for
    call MSG-μ-CYCLE($L - 1$) μ times
    for each grid $G^{(p,q)}$ **where** $p + q = L$ **do**
      compute the weight factors $\boldsymbol{\kappa}_x^{(p,q)}$ and $\boldsymbol{\kappa}_y^{(p,q)}$ using (15)
      $\boldsymbol{v}^{(p,q)} \leftarrow \boldsymbol{v}^{(p,q)}+$ the interpolated correction from (11)
      call SMOOTH($\boldsymbol{v}^{(p,q)}, A^{(p,q)}, \boldsymbol{b}^{(p,q)}$) ν_2 times
    end for
  end if
end procedure

A multigrid cycle with $\overline{p} + \overline{q} - 1$ levels is now performed in the usual way (see Algorithm 1). In each step of the algorithm, multiple semicoarsened grids are involved (one in each coordinate direction). In the two-dimensional case

considered here, essentially two semicoarsened grids should be treated. Hence, we must specify how to transfer and combine the information from these two grids. We first describe our approach for restriction and subsequently for prolongation.

The residual $\boldsymbol{r}^{(p,q)} = \boldsymbol{b}^{(p,q)} - A^{(p,q)}\boldsymbol{v}^{(p,q)}$ is restricted to grid $G^{(p,q)}$ simply by taking the average of the restricted residuals from both finer grids $G^{(p+1,q)}$ and $G^{(p,q+1)}$, if they exist, that is,

$$\begin{cases} \mathcal{R}_x^{(p,q)}\boldsymbol{r}^{(p+1,q)} & \text{if } q = \overline{q}, \\ \mathcal{R}_y^{(p,q)}\boldsymbol{r}^{(p,q+1)} & \text{if } p = \overline{p}, \\ \frac{1}{2}\mathcal{R}_x^{(p,q)}\boldsymbol{r}^{(p+1,q)} + \frac{1}{2}\mathcal{R}_y^{(p,q)}\boldsymbol{r}^{(p,q+1)} & \text{otherwise.} \end{cases} \tag{10}$$

Recall that only semicoarsening in the direction of strongest coupling can overcome the ineffectiveness of a smoother based on point relaxation.[12] In the case of grid-aligned anisotropies, only half of the MSG grids will be effective in reducing the high-frequency components of the error. Hence, it is crucial that a weighted average of interpolated corrections from both coarser grids $G^{(p-1,q)}$ and $G^{(p,q-1)}$ is used to update the solution on grid $G^{(p,q)}$, if both exist, that is,

$$\begin{cases} \mathcal{P}_x^{(p,q)}\boldsymbol{v}^{(p-1,q)} & \text{if } q = 1, \\ \mathcal{P}_y^{(p,q)}\boldsymbol{v}^{(p,q-1)} & \text{if } p = 1, \\ \boldsymbol{\kappa}_x^{(p,q)}\mathcal{P}_x^{(p,q)}\boldsymbol{v}^{(p-1,q)} + \boldsymbol{\kappa}_y^{(p,q)}\mathcal{P}_y^{(p,q)}\boldsymbol{v}^{(p,q-1)} & \text{otherwise,} \end{cases} \tag{11}$$

where $\boldsymbol{\kappa}_x^{(p,q)}$ and $\boldsymbol{\kappa}_y^{(p,q)}$ are appropriate weight factors. The naive approach uses weight factors $\boldsymbol{\kappa}_x^{(p,q)} = \boldsymbol{\kappa}_y^{(p,q)} = 1/2$. However, these weight factors are not effective in the case of strong alignment along one of the coordinate directions, since the appropriate grids get only half of the necessary information. For our model problem, where the diffusion coefficient is given as a random field, the optimal choice for the weight factors will vary over the domain, and matrix-dependent prolongation is required to achieve acceptable convergence rates. We follow the approach from Naik and Van Rosendale.[2]

Define

$$f_{i,j} = \cos(i\pi) \quad \text{and} \quad g_{i,j} = \cos(j\pi) \quad \text{for } 0 \le i \le 2^p, 0 \le j \le 2^q. \tag{12}$$

These are two high-frequency Fourier modes, one oscillatory in the x-direction, the other oscillatory in the y-direction, that locally look like

$$f = \begin{matrix} 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{matrix} \quad \text{and} \quad g = \begin{matrix} 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \end{matrix}. \tag{13}$$

Appropriate weight factors can be computed by applying the discrete operator $A^{(p,q)}$ to f and g, that is, compute

$$\lambda_x = A^{(p,q)}f \quad \text{and} \quad \lambda_y = A^{(p,q)}g, \tag{14}$$

and define the weight factors

$$\boldsymbol{\kappa}_x^{(p,q)} := \frac{\lambda_x^2}{\lambda_x^2 + \lambda_y^2} \quad \text{and} \quad \boldsymbol{\kappa}_y^{(p,q)} := \frac{\lambda_y^2}{\lambda_x^2 + \lambda_y^2}. \tag{15}$$

It can be shown that the convergence factor of the MSG method, using these weight factors, can be made arbitrarily small if sufficient relaxation steps are performed.[2] Alternative forms for the weight factors $\boldsymbol{\kappa}_x^{(p,q)}$ and $\boldsymbol{\kappa}_y^{(p,q)}$ can be found in References 3, 15, and 16.

The recursive procedure in Algorithm 1 is started by constructing the hierarchy of grids given in Equation (5), setting an initial guess $\boldsymbol{v}^{(\overline{p},\overline{q})} = \boldsymbol{0}$ and calling the procedure with $L = \overline{p} + \overline{q}$. When $L = 0$, there is only one unknown, and the system can be solved simply by inverting the matrix $A^{(0,0)}$, which is then simply a scalar. A multigrid V-cycle is obtained when $\mu = 1$, a W-cycle is obtained when $\mu = 2$.

As in standard MG, it is possible to combine MSG with nested iteration. In nested iteration, an initial approximation for an iterative method on a fine grid is provided by the computation and subsequent interpolation of solutions on coarser grids. A good initial guess for the solution on the fine grid means that just a few iterations will be made with a multigrid V- or W-cycle in order to converge the solution to discretization accuracy.[12] Combined with MG, this method is called FMG, or, in the MSG context, full-multiple semicoarsened multigrid (FMSG) (see Algorithm 2). Some remarks concerning this algorithm will be given next. First, in line 6 of the algorithm, it is the right-hand side that is restricted, not the residual. However, Equation (10) can still be used when the residuals $\boldsymbol{r}^{(p+1,q)}$ and $\boldsymbol{r}^{(p,q+1)}$ are replaced by right-hand side vectors $\boldsymbol{b}^{(p+1,q)}$ and $\boldsymbol{b}^{(p,q+1)}$, respectively. Second, in line 11, it is customary to use a higher order interpolation scheme to get improved initial conditions for the subsequent V- or W-cycles. In our numerical experiments, we will use cubic interpolation. For example, the stencil for cubic interpolation in the x-direction is given by

$$\mathcal{P}_x^{(p,q)} = \frac{1}{16}\left[-1\ \ 0\ \ 9\ \ 16\ \ 9\ \ 0\ \ -1\right]_{p,q}^{p+1,q} \quad \text{or at the left-most boundary } \mathcal{P}_x^{(p,q)} = \frac{1}{16}\left[10\ \ 16\ \ 9\ \ 0\ \ -1\right]_{p,q}^{p+1,q} \tag{16}$$

and similar for the right-most boundary. Third and finally, in the FMG algorithm, it is common to use an adaptive number of V- or W-cycles ν_0. This number is then increased until $\|\boldsymbol{r}^{(p,q)}\|/\|\boldsymbol{b}^{(p,q)}\| \leq \varepsilon$, where ε is a tolerance provided by the user.

Algorithm 2. Multiple semicoarsening multigrid F-cycle (recursive definition)

input: level parameter L
output: approximations $\boldsymbol{v}^{(p,q)}$ for the solution of $A^{(p,q)}\boldsymbol{u}^{(p,q)} = \boldsymbol{b}^{(p,q)}, p+q=L$

```
1:  procedure MSG-F-CYCLE(L)
2:      if L = 0 then
3:          v^(0,0) ← 0
4:      else
5:          for each grid G^(p,q) where p + q = L − 1 do
6:              b^(p,q) ← the restricted right-hand side from (10)
7:          end for
8:          call MSG-F-CYCLE(L − 1)
9:          for each grid G^(p,q) where p + q = L do
10:             compute the weight factors κ_x^(p,q) and κ_y^(p,q) using (15)
11:             v^(p,q) ← the interpolated solution from (11)
12:         end for
13:     end if
14:     call MSG-μ-CYCLE(L) ν_0 times
15: end procedure
```

We test our implementation of MSG on the model elliptic PDE with lognormal coefficients given in Equation (1), defined on the unit square $D = [0,1]^2$, where we assume a Matérn covariance function for the random field, and set the source term $h(\boldsymbol{x}) = 1$. In Figure 3, we plot the evolution of the norm of the residual for classic MG and MSG over 50 multigrid W-cycles, for 100 samples of the random field. We used two pre- and postsmoothing steps with red-black Gauss–Seidel ($\nu_1 = \nu_2 = 2$) and found numerically that a small damping factor before the coarse grid correction improves the overall convergence rate.[3] W-cycles ($\mu = 2$) are used because they are known to be generally more stable than V-cycles ($\mu = 1$).[13] Notice the expected failure of the multigrid method with standard coarsening (MG) as the anisotropic ratio η decreases. The MSG method, on the other hand, offers fast convergence for a wide range of parameters. As the anisotropy is less grid-aligned (larger θ) and for a fixed anisotropic ratio η (as shown in the middle column of Figure 3), the convergence of classic MG is improved, a result also reported by Oosterlee and Wesseling.[3]

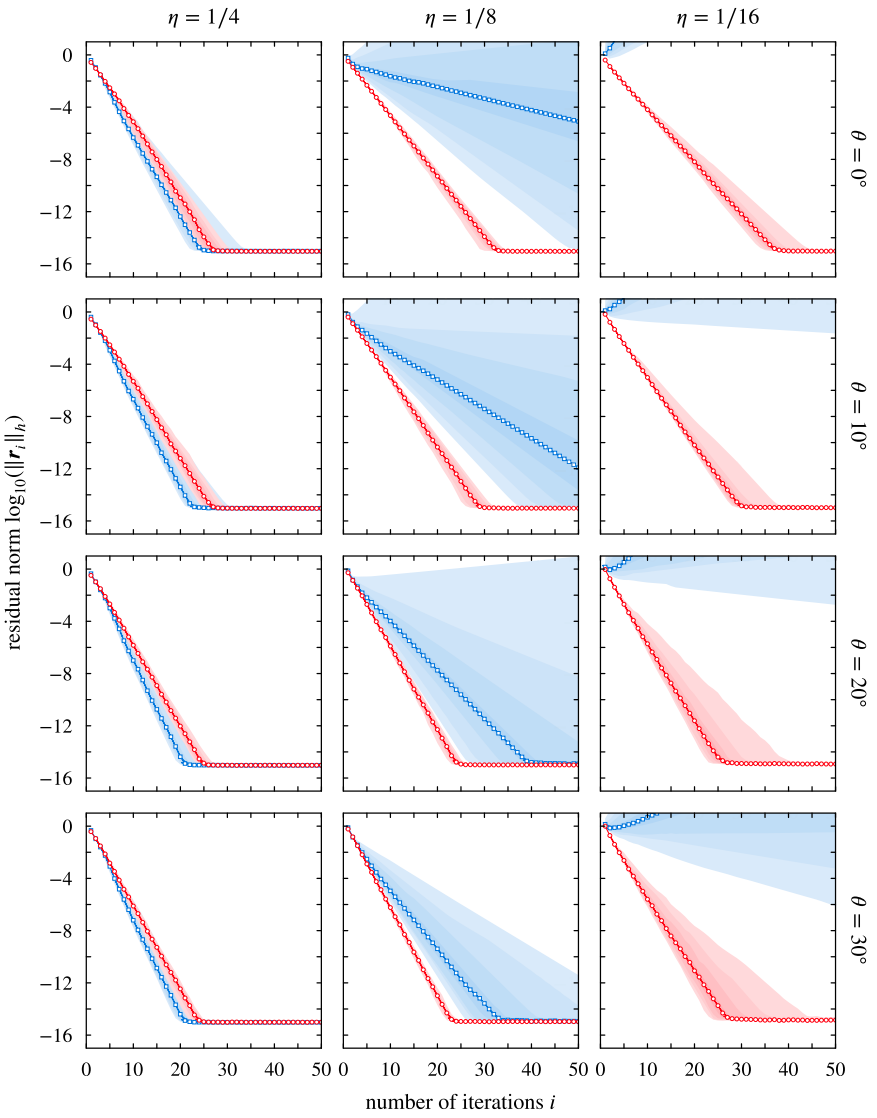

FIGURE 3 20%, 40%, … , 80% quantiles of the norm of the residual after 50 multigrid W(2, 2)-cycles on a grid with $p = q = 6$ (63×63 degrees of freedom) for Matérn parameters $\lambda = 1/4$, $\nu = 1/2$, $\eta \in \{1/4, 1/8, 1/16\}$ (left to right), and $\theta \in \{0°, 10°, 20°, 30°\}$ (top to bottom) based on 100 samples. Median values are shown for MG (—▫—) and MSG (—▫—)

We investigate the dependence on θ further in Figure 4, where we plot the averaged convergence factor for different cycling strategies using both MG and MSG. This convergence factor is computed as

$$\phi_k := \sqrt[k]{\frac{\|\boldsymbol{r}_k\|}{\|\boldsymbol{r}_0\|}},$$

where the number of iterations k is chosen such that $\|\boldsymbol{r}_k\|/\|\boldsymbol{r}_0\| < 10^{-12}$. Notice how the classic MG method requires a large number of smoothing steps to reach converge on finer grids. With $\theta = 0°$, no convergence is reached for grids beyond 31×31 unknowns, even when five pre- and postsmoothing steps are used. The MSG method, on the other hand, retains an acceptable convergence factor in all cases considered.

Numerical results for various other parameter settings can be found online at https://people.cs.kuleuven.be/~pieterjan.robbe/copper2019. Finally, we refer to Kumar et al.[17] for a local Fourier analysis (LFA) of the model problem with isotropic Matérn covariance.

3 | MULTI-INDEX MONTE CARLO

Recall that our aim is to compute statistics of a quantity of interest Q derived from the solution of problem (1). Let us denote the quantity of interest computed on grid $G^{(p,q)}$ by Q_ℓ, where $\ell = (p - p_0, q - q_0)$ is a multi-index with $p_0 \leq p \leq \overline{p}$

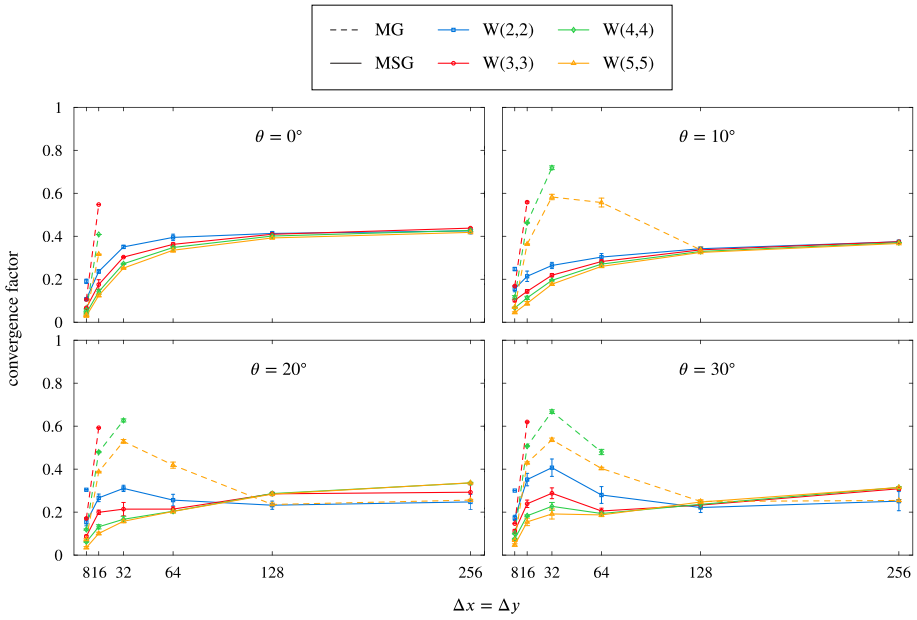

FIGURE 4 Expected convergence factors for MG (dashed line) and MSG (solid line) using different cycling strategies for Matérn parameters $\lambda = 1/4$, $\nu = 1/2$, $\eta = 1/16$, and $\theta \in \{0°, 10°, 20°, 30°\}$ (left to right and top to bottom) based on 100 samples. Missing values indicate that some of the samples did not convergence

and $q_0 \leq q \leq \overline{q}$. Index $\boldsymbol{\ell} = (0, 0)$ thus corresponds to an approximation on the coarse grid $G^{(p_0,q_0)}$, which is not necessarily the coarsest grid in the hierarchy defined by Equation (5). An MC estimator for the expected value of the quantity of interest Q is just the sample average, that is,

$$\mathbb{E}[Q] \approx Q_N := \frac{1}{N} \sum_{n=1}^{N} Q_{\boldsymbol{L}}^{(n)}, \tag{17}$$

where $Q_{\boldsymbol{L}}^{(n)}$ is the nth sample of Q computed on the finest grid $G^{(\overline{p},\overline{q})}$.

Instead of estimating the quantity of interest directly on the finest grid, the multi-index construction, for a generic number of dimensions d, starts from a tensor product of single-direction differences defined as

$$\boldsymbol{\Delta} Q_{\boldsymbol{\ell}} := \left(\bigotimes_{i=1}^{d} \Delta_i \right) Q_{\boldsymbol{\ell}} \quad \text{with} \quad \Delta_i Q_{\boldsymbol{\ell}} = \begin{cases} Q_{\boldsymbol{\ell}} - Q_{\boldsymbol{\ell} - \boldsymbol{e}_i} & \text{if } \ell_i > 0 \\ Q_{\boldsymbol{\ell}} & \text{if } \ell_i = 0 \end{cases}, \tag{18}$$

where $\boldsymbol{e}_i$ is the unit vector in direction i. For example, with $d = 2$ and $\boldsymbol{\ell} = (3, 4)$, we have that

$$\begin{aligned} \boldsymbol{\Delta} Q_{(3,4)} &= \Delta_2(\Delta_1 Q_{(3,4)}) \\ &= \Delta_2(Q_{(3,4)} - Q_{(2,4)}) \\ &= Q_{(3,4)} - Q_{(2,4)} - Q_{(3,3)} + Q_{(2,3)}. \end{aligned}$$

In general, taking a sample of $\boldsymbol{\Delta} Q_{\boldsymbol{\ell}}$ requires the solution of a PDE on 2^d different grids. However, when using the FMSG method for our two-dimensional model problem, we get free solutions on all coarser grids $G^{(p' q')}$ where $p' < p$ or $q' < q$. Hence, a sample of $\boldsymbol{\Delta} Q_{\boldsymbol{\ell}}$ can be computed at the same cost of computing a sample of $Q_{\boldsymbol{\ell}}$.

In order to obtain an efficient multi-index estimator, it is crucial that the difference $\boldsymbol{\Delta} Q_{\boldsymbol{\ell}}$ is computed from a quantity of interest that is based on a discretization of the PDE with the same underlying sample of the random field $Z(\boldsymbol{x}, \omega)$. This will ensure that the quantities that constitute the multi-index difference are strongly positively correlated, and, hence, their difference will be small. As a consequence, the variance of the difference is heavily reduced. This means that, to reach the same mean square error, fewer samples of the difference are required compared with sampling the quantity of interest directly. Furthermore, as $\boldsymbol{\ell} \to +\infty$ in every coordinate, we expect the quantity of interest to converge toward the true quantity of interest, that is, $Q_{\boldsymbol{\ell}} \to Q$, and thus $\boldsymbol{\Delta} Q_{\boldsymbol{\ell}} \to 0$, so that the variance of the difference decreases rapidly with $\boldsymbol{\ell}$. This means that fewer and fewer (more expensive) samples are needed on finer grids.

The multi-index Monte Carlo estimator, proposed by Haji-Ali et al.,[5] is the sum of sample averages of multi-index differences, that is,

$$\mathcal{Q}_{L,\{N_\ell\}} = \sum_{\ell \in \mathcal{I}(L)} \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \mathbf{\Delta} Q_\ell^{(n)}, \tag{19}$$

with $\mathcal{I}(L) \subseteq \mathbb{N}_0^d$ being a suitable set of indices, where L governs the size of the set, and N_ℓ is the number of samples on index ℓ. In its current form, the multi-index estimator does not allow sample reuse. When computing the variance of Equation (19), necessary to control the root mean square error, unwanted correlations appear, which are hard to estimate reliably.

Our unbiased multi-index Monte Carlo estimator, inspired by the work by Rhee and Glynn[8] and Detomasso et al.,[18] does allow for sample reuse. The estimator can be written as

$$\mathcal{Q}_{\infty,N} = \frac{1}{N} \sum_{n=1}^{N} \sum_{\mathbf{0} \le \ell \le \boldsymbol{L}^{(n)}} \frac{1}{p_\ell} \mathbf{\Delta} Q_\ell^{(n)}, \tag{20}$$

where for every sample n, we draw a sample $\boldsymbol{L}^{(n)}$ of a d-variate discrete random variable $\boldsymbol{L}$ according to some nonzero probability mass function $\pi_{\boldsymbol{L}} : \mathbb{N}_0^d \mapsto [0,1]$ with $\pi_{\boldsymbol{L}}(\ell) \ge 0$, componentwise, and $\sum_{\ell \in \mathbb{N}_0^d} \pi_{\boldsymbol{L}}(\ell) = 1$, independent from the random samples of $\mathbf{\Delta} Q_\ell$. The

$$p_\ell := \Pr[\mathbf{0} \le \ell \le \boldsymbol{L}] = \Pr[0 \le \ell_1 \le L_1 \cup 0 \le \ell_2 \le L_2 \cup \ldots \cup 0 \le \ell_d \le L_d] \tag{21}$$

are the probability that $\boldsymbol{L}$ is at least ℓ, componentwise. These probabilities are related with the cumulative distribution function $\Pi_{\boldsymbol{L}} : \mathbb{N}_0^d \mapsto [0,1]$ of the multivariate random variable $\boldsymbol{L}$ since $\Pi_{\boldsymbol{L}}(\ell) := \Pr[\boldsymbol{L} \le \ell] = 1 - p_{\ell+1}$.

The argument inside the multiple sum in Equation (20) will be nonzero up to only a finite subset of $\mathbb{N}_0^d$. If the probability mass $\pi_{\boldsymbol{L}}(\ell)$ is decreasing with ℓ, componentwise, then the product Np_ℓ is mimicking the decreasing multivariate sequence of sample sizes $\{N_\ell\}$ in the classic MIMC estimator from Equation (19). This observation will be used in the derivation of p_ℓ below.

We can proof unbiasedness of the MIMC estimator (20) by noting that

$$\begin{aligned}
\mathbb{E}[\mathcal{Q}_{\infty,N}] &= \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}\sum_{\ell \in \mathbb{N}_0^d} \frac{1}{p_\ell} \mathbf{\Delta} Q_\ell^{(n)} \chi_{[\mathbf{0},\boldsymbol{L}^{(n)}]}(\ell)\right] \\
&= \sum_{\ell \in \mathbb{N}_0^d} \frac{1}{p_\ell} \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} \mathbf{\Delta} Q_\ell^{(n)} \chi_{[\mathbf{0},\boldsymbol{L}^{(n)}]}(\ell)\right] \\
&= \sum_{\ell \in \mathbb{N}_0^d} \frac{1}{p_\ell} \mathbb{E}[\mathbf{\Delta} Q_\ell] \, \mathbb{E}\left[\chi_{[\mathbf{0},\boldsymbol{L}^{(n)}]}(\ell)\right] \\
&= \sum_{\ell \in \mathbb{N}_0^d} \mathbb{E}[\mathbf{\Delta} Q_\ell] = \mathbb{E}[Q].
\end{aligned} \tag{22}$$

Here, we used the multivariate indicator function, defined as the tensor product of one-dimensional indicator functions. That is, $\chi_{[\mathbf{0},\boldsymbol{L}^{(n)}]}(\ell) = \chi_{[0,L_1^{(n)}]}(\ell_1) \cdot \ldots \cdot \chi_{[0,L_d^{(n)}]}(\ell_d)$. In the last step, we used the telescoping property of the multi-index differences.

The randomization of the index $\boldsymbol{L}$ is the key difference with the classic multi-index estimator. The finite index set $\mathcal{I}(L)$ in Equation (19) introduces an additional bias term in the expression for the root mean square error of the estimator, which needs to be controlled. Although adaptive approaches have been proposed to tackle this problem,[7] obtaining a stable and accurate bound for the bias term is nontrivial. The formulation in Equation (20) avoids this problem altogether, and the root mean square error of the estimator consists solely of a statistical error term, that is,

$$\mathrm{RMSE}(\mathcal{Q}_{\infty,N}) := \sqrt{\mathbb{E}\left[\left(\mathcal{Q}_{\infty,N} - \mathbb{E}[Q]\right)^2\right]} = \sqrt{\mathbb{V}\left[\mathcal{Q}_{\infty,N}\right]}. \tag{23}$$

The variance of the estimator, $\mathbb{V}\left[\mathcal{Q}_{\infty,N}\right]$, can be approximated using the sample variance of

$$Y^{(n)} = \sum_{\mathbf{0}\le\boldsymbol{\ell}\le\boldsymbol{L}^{(n)}} \frac{1}{p_{\boldsymbol{\ell}}} \boldsymbol{\Delta} Q_{\boldsymbol{\ell}}^{(n)}, \tag{24}$$

if every nth sample from $\boldsymbol{L}$ and $\boldsymbol{\Delta}Q_{\boldsymbol{\ell}}$ is i.i.d. In particular, we have that

$$\mathbb{E}\left[\mathcal{Q}_{\infty,N}\right] \approx E := \frac{1}{N}\sum_{n=1}^{N} Y^{(n)}, \text{ and} \tag{25}$$

$$\mathbb{V}\left[\mathcal{Q}_{\infty,N}\right] \approx V := \frac{1}{N(N-1)}\sum_{n=1}^{N}\left(Y^{(n)} - E\right)^2. \tag{26}$$

At this point, it is important to stress that a single sample of $Y^{(n)}$ can be computed from only one deterministic PDE solution when using FMSG, since coarse solutions are returned for free by the multigrid solver. We will call the MIMC estimator that reuses samples, multiple semicoarsened multigrid multi-index Monte Carlo (MSG-MIMC), and denote it by $\mathcal{Q}_{\infty}^{\mathfrak{r}}, N$. This is opposed to $\mathcal{Q}_{\infty,N}$, the MIMC estimator that does not reuse samples. A basic algorithm for MSG-MIMC simulation takes N draws of $\boldsymbol{L}$ according to a given distribution $\pi_{\boldsymbol{L}}$ and performs one deterministic PDE solve using FMSG for every $\boldsymbol{L}^{(n)}$, albeit on different grids. Note that this ensures the so-called downward-closed condition of the set of all indices that are added in the estimator.[5] This condition incorporates, among others, that the index set does not contain gaps and ensures the validity of the telescoping sum identity used to prove unbiasedness. Algorithm 3 provides the key instructions to implement our multi-index estimator.

Algorithm 3. Multiple semicoarsened multigrid multi-index Monte Carlo

input: a tolerance ε for the RMSE
output: an approximation E for the mean of the quantity of interest Q, and an error estimate $\sqrt{V}$

procedure MSG-MIMC(ε)
  $n \leftarrow 1$
  $\pi_{\boldsymbol{L}} \leftarrow \text{GEOMETRIC}^d(\log(2)(1+4)/2)$
  repeat
    sample $\boldsymbol{L}^{(n)}$ according to the probability mass function $\pi_{\boldsymbol{L}}$
    solve the PDE using FMSG to obtain $\boldsymbol{\Delta}Q_{\boldsymbol{\tau}}^{(n)}, \mathbf{0} \le \boldsymbol{\tau} \le \boldsymbol{L}^{(n)}$
    compute $Y^{(n)}$ using (24)
    compute E using (25) and V using (26)
    update the probability mass function $\pi_{\boldsymbol{L}}$
    $n \leftarrow n+1$
  until $\sqrt{V} \le \varepsilon$
end procedure

It remains to be determined how to make a favorable choice of $\pi_{\boldsymbol{L}}$. We rewrite our multi-index estimator as follows:

$$\mathcal{Q}_{\infty,N}^{\mathfrak{r}} = \sum_{n=1}^{N}\sum_{\mathbf{0}\le\boldsymbol{\ell}\le\boldsymbol{L}^{(n)}} \frac{1}{Np_{\boldsymbol{\ell}}}\boldsymbol{\Delta}Q_{\boldsymbol{\ell}}^{(n)} = \sum_{n=1}^{N}\sum_{\mathbf{0}\le\boldsymbol{\ell}\le\boldsymbol{L}^{(n)}} \frac{1}{N'_{\boldsymbol{\ell}}}\boldsymbol{\Delta}Q_{\boldsymbol{\ell}}^{(n)}, \tag{27}$$

where $N'_{\boldsymbol{\ell}}$ denotes a discrete density of sample sizes, similar to the decreasing sequence of sample sizes $N_{\boldsymbol{\ell}}$ in standard MIMC (see Equation (19)). In the latter, it is a well-known result that the optimal number of samples on each index is

$$N_{\boldsymbol{\ell}} \simeq \sqrt{\frac{V_{\boldsymbol{\ell}}}{C_{\boldsymbol{\ell}}}}, \tag{28}$$

where $V_\ell := \mathbb{V}[\mathbf{\Delta} Q_\ell]$ and C_ℓ is the cost to compute a single sample of $\mathbf{\Delta} Q_\ell$.[4,5] Hence, knowing the discrete distributions of V_ℓ and C_ℓ across all ℓ allows one to learn the unknown probability p_ℓ by simple normalization of the N_ℓ. In a practical implementation, one can resort to either an on-the-fly computed empirical probability mass function using the (square root of the) ratio of V_ℓ over C_ℓ or a proposed model for their respective distributions. In our numerical experiments later on, we will opt for the first approach, whereas, for the remainder of this section, we choose the latter.

In particular, and as is customary in a multi-index setting, assume the expected value, variance and cost of $\mathbf{\Delta} Q_\ell$ follow a product structure:

$$E_\ell := |\mathbb{E}[\mathbf{\Delta} Q_\ell]| \leq c_E \prod_{j=1}^{d} 2^{-\alpha_j \ell_j}, \tag{A1}$$

$$V_\ell = \mathbb{V}[\mathbf{\Delta} Q_\ell] \leq c_V \prod_{j=1}^{d} 2^{-\beta_j \ell_j}, \text{and} \tag{A2}$$

$$C_\ell = \text{cost}(\mathbf{\Delta} Q_\ell) \leq c_C \prod_{j=1}^{d} 2^{\gamma_j \ell_j}, \tag{A3}$$

where c_E, c_V, c_C and α_j, β_j, $\gamma_j, j = 1, \dots, d$ are positive constants independent of ℓ.[5] Using these assumptions, the optimal distribution for the number of samples N_ℓ in Equation (28) can be written as

$$N_\ell \simeq \prod_{j=1}^{d} 2^{-\frac{1}{2}(\gamma_j+\beta_j)\ell_j} \simeq \prod_{j=1}^{d} \exp(-r_j \ell_j), \tag{29}$$

where $r_j = \frac{1}{2}\log(2)(\gamma_j + \beta_j)$, for $j = 1, \dots, d$. Hence, $\boldsymbol{\pi_L}$ must be chosen as a multivariate geometric distribution, that is, the discrete equivalent of an exponential distribution, and

$$p_\ell = \prod_{j=1}^{d} \exp(-r_j \ell_j). \tag{30}$$

This explains the choice for the initial distribution in line 3 in Algorithm 3, where we picked $\beta_j = 4$ and $\gamma_j = 1$, $j = 1, 2, \dots, d$.

There are several ways in which one could update the probability mass function $\boldsymbol{\pi_L}$ in line 9 in Algorithm 3. For example, one could assume the multivariate geometric distribution from Equation (30), where the unknown parameters $r_j, j = 1, \dots, d$ must be fitted from the available samples. Alternatively, one can use the empirical probability mass function directly. Hybrid approaches, where missing or unreliable data—due to a lack of samples—are replaced by estimates for $\boldsymbol{\pi_L}$ are also possible. In our experiments, we used the empirically determined probability mass function. We note that updating the probability mass function $\boldsymbol{\pi_L}$ in every iteration of the algorithm does not ruin the unbiasedness of our estimator. In particular, Equation (25) still holds when all samples of $Y^{(n)}$ are independent (but not identically distributed) using the Lindeberg–Feller central limit theorem.

We are now in a position to formulate a theorem that provides the expected cost reduction of our new MIMC estimator with sample reuse, compared with MIMC where no samples are reused, based on assumptions (A1)–(A3).

Theorem 1. *Let assumptions (A1)–(*A3*) hold. The cost reduction factor of the MSG-MIMC estimator is given by*

$$\frac{\mathcal{Q}^{\mathbf{r}_{\infty,N}}}{\mathcal{Q}_{\infty,N}} = 1 - \sum_{\substack{\mathbf{u} \subseteq \{1,\dots,d\} \\ \mathbf{u} \neq \emptyset}} (-1)^{|\mathbf{u}|+1} \prod_{j \in \mathbf{u}} 2^{-\frac{1}{2}(\gamma_j+\beta_j)}, \tag{31}$$

*where β_j is the rate of the decrease in variance in direction j, see (*A2*), and γ_j is the rate of the increase in cost in direction j, see (*A3*).*

Proof. Suppose the algorithm requires N_ℓ samples to be taken on index ℓ. In the MSG-MIMC method, when recycling samples, the amount of samples that remain to be taken on index ℓ is given by

$$\tilde{N}_\ell := N_\ell - \sum_{\substack{\mathbf{u}\subseteq\{1,\dots,d\}\\ \mathbf{u}\neq\emptyset}} (-1)^{|\mathbf{u}|+1} N_{\ell+\boldsymbol{e}_\mathbf{u}}. \tag{32}$$

From Equation (28), we find that

$$N_{\ell+\boldsymbol{e}_\mathbf{u}} = \left(\prod_{j\in\mathbf{u}} 2^{-(\gamma_j+\beta_j)/2}\right) N_\ell, \tag{33}$$

and hence

$$\tilde{N}_\ell = \left(1 - \sum_{\substack{\mathbf{u}\subseteq\{1,\dots,d\}\\ \mathbf{u}\neq\emptyset}} (-1)^{|\mathbf{u}|+1} \prod_{j\in\mathbf{u}} 2^{-(\gamma_j+\beta_j)/2}\right) N_\ell. \tag{34}$$

Therefore,

$$\mathrm{cost}\left(\mathcal{Q}^{\mathbf{r}_{\infty,N}}\right) = \sum_{\ell\in\mathbb{N}_0^d} \tilde{N}_\ell C_\ell \tag{35}$$

$$= \left(1 - \sum_{\substack{\mathbf{u}\subseteq\{1,\dots,d\}\\ \mathbf{u}\neq\emptyset}} (-1)^{|\mathbf{u}|+1} \prod_{j\in\mathbf{u}} 2^{-(\gamma_j+\beta_j)/2}\right) \sum_{\ell=0}^{\infty} N_\ell C_\ell \tag{36}$$

$$= \left(1 - \sum_{\substack{\mathbf{u}\subseteq\{1,\dots,d\}\\ \mathbf{u}\neq\emptyset}} (-1)^{|\mathbf{u}|+1} \prod_{j\in\mathbf{u}} 2^{-(\gamma_j+\beta_j)/2}\right) \mathrm{cost}\left(\mathcal{Q}_{\infty,N}\right). \tag{37}$$

This proves the theorem. ∎

The sample reuse in the MSG-MIMC method yields only a constant factor in cost reduction, compared with unbiased MIMC without sample reuse. In applications, usually, the rate of increase in cost, $\gamma_j, j = 1, \dots, d$, is fixed, and the benefit of the sample reuse will be more pronounced when the rate of decrease in variance, $\beta_j, j = 1, \dots, d$, is small, reflecting a slow decay of the number of samples as ℓ increases. This is in agreement with previous results obtained for the ML(Q)MC setting.[19]

4 | NUMERICAL RESULTS

In this section, we present some numerical results for the model problem in Equation (1), with $D = [0,1]^2$ and $h(\boldsymbol{x}) = 1$. We consider three quantities of interest:

1. A point evaluation of the solution at $\boldsymbol{x} = (1/2, 1/2)$:

$$Q_1 := u((1/2,1/2),\cdot). \tag{38}$$

2. The average value of the solution over the subdomain $D_{\text{sub}} = [1/4, 1/2]^2$:

$$Q_2 := \frac{1}{|D_{\text{sub}}|} \int_{D_{\text{sub}}} u(\boldsymbol{x}, \cdot)\, d\boldsymbol{x}, \tag{39}$$

where the integral is approximated using a two-dimensional trapezoidal rule.

3. The flux through the rightmost side of the domain:

$$Q_3 := -\int_0^1 a((1, y), \cdot) \frac{\partial u}{\partial x}((1, y), \cdot)\, dy, \tag{40}$$

where the integral is approximated using a trapezoidal rule, and the derivative is approximated using first-order finite differences.

In all test cases, the uncertain diffusion coefficient $a(\boldsymbol{x}, \omega)$ is modeled as a lognormal Gaussian random field with Matérn covariance (2) with smoothness $\nu = 1/2$, length scale $\lambda = 1/4$, and where the anisotropic ratio and rotation are considered as hyperparameters, uniformly distributed in $[1/16, 1/4]$ and $[-30°, 30°]$, respectively. Exact samples of the random field are computed using circulant embedding, where the field is sampled on a cube, that is,

$$1 + 2\left\lceil \sqrt{\nu} \log_2(\max(2^{p_0}, 2^{q_0})) \right\rceil \tag{41}$$

TABLE 1 Fitted values for the rates α_j, β_j and γ_j, $j = 1, 2$ in assumptions (A1)–(A3) for each quantity of interest Q_1, Q_2, and Q_3

	(α_1, α_2)	(β_1, β_2)	(γ_1, γ_2)
Q_1	(2.02, 1.97)	(4.73, 4.47)	(1.27, 1.21)
Q_2	(0.43, 0.30)	(2.21, 2.57)	(1.21, 1.34)
Q_3	(0.33, 0.40)	(1.36, 1.51)	(1.49, 1.26)

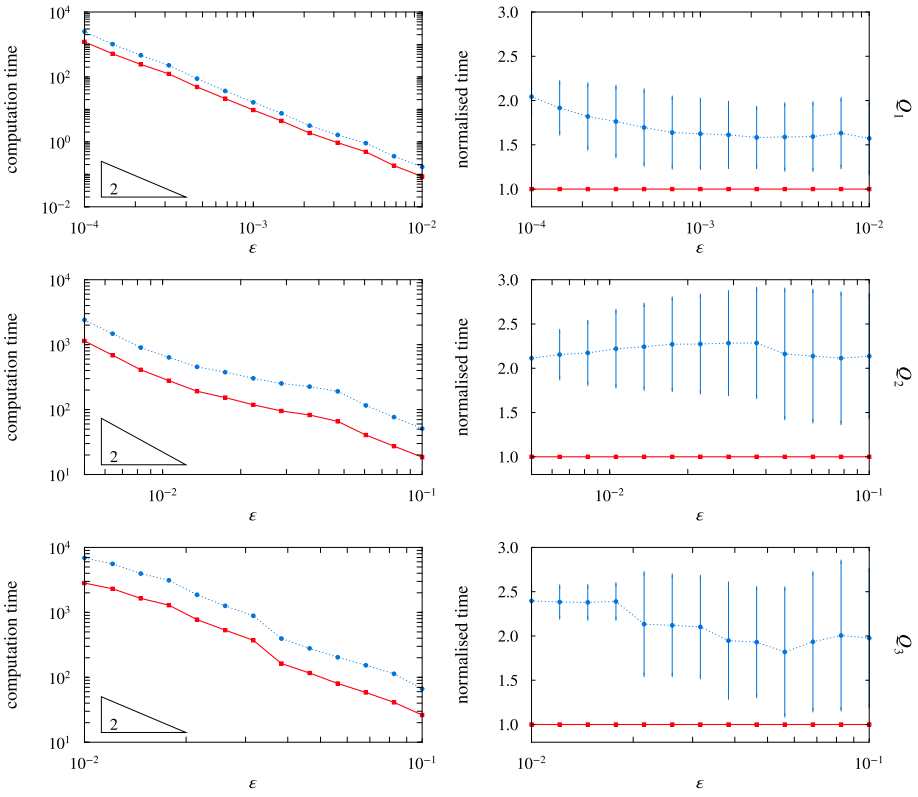

FIGURE 5 Performance of the MSG-MIMC method with (—■—) and without (·····●·····) recycling for different quantities of interest: point evaluation of the solution (Q_1, top), average value of the solution over a subdomain (Q_2, middle), flux through a boundary (Q_3, bottom)

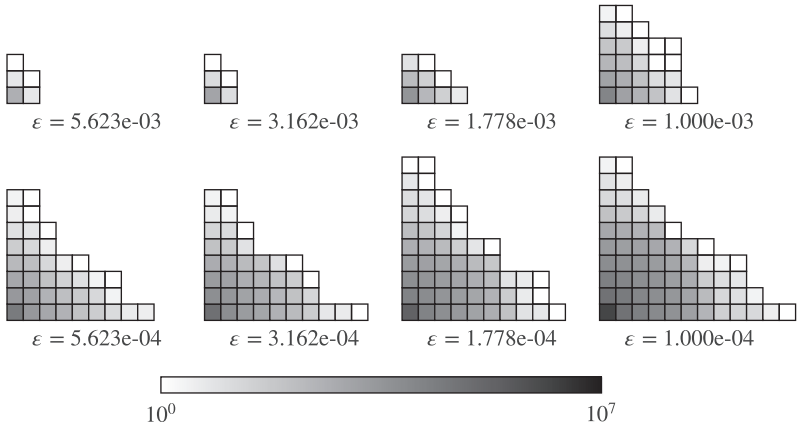

FIGURE 6 Shape of the index set for the first quantity of interest Q_1 (point evaluation)

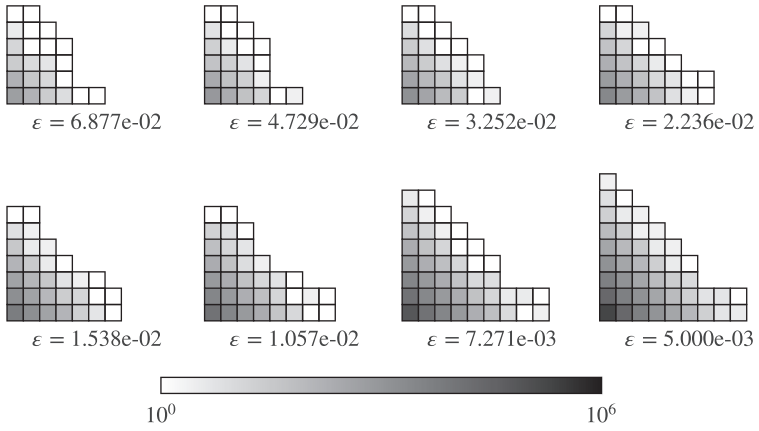

FIGURE 7 Shape of the index set for the second quantity of interest Q_2 (average value)

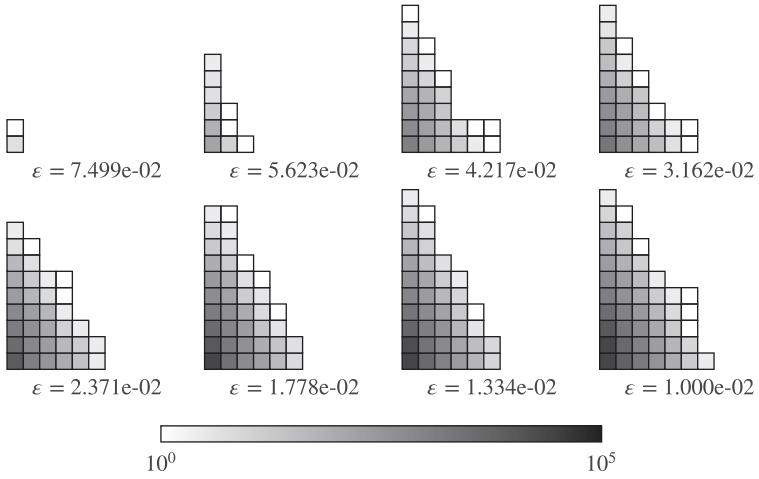

FIGURE 8 Shape of the index set for the third quantity of interest Q_3 (flux)

times larger in every direction, compared with the size of the physical domain, to ensure positive definiteness of the covariance matrix (see Graham et al.[20]). We should mention that, in the case of hyperparameters, the eigenvalues of the covariance matrix must be recomputed in every sample. Fortunately, they can be computed rapidly ($O\left(2^{p+q}\log(2^{p+q})\right)$ time) using an FFT routine. See also Latz et al.[21] for an alternative method to efficiently generate samples from a Gaussian random field with hyperparameters, based on reduced basis surrogate modeling.

A coarsest mesh with $p_0 = q_0 = 4$ was used for Q_1, and $p_0 = q_0 = 16$ for Q_2 and Q_3, and the deterministic PDE is solved using FMSG with an adaptive number of W(2,2)-cycles ν_0 (see Section 2). The finest mesh on which we were able to solve the PDE consists of $2^{18} = 262\,144$ degrees of freedom ($\overline{p} + \overline{q} = 18$). The expected rates α_j, β_j, and γ_j, $j = 1, 2$ in assumptions (A1)–(A3) are summarized in Table 1, for each quantity of interest. The rate of increase in computational cost, $\gamma_j, j = 1, 2$, remains approximately constant for all quantities of interest considered. The rate of decrease in variance, $\beta_j, j = 1, 2$, diminishes, going from Q_1 to Q_2 to Q_3. Hence, according to Theorem 1, we expect the largest benefit of sample reuse for Q_3, followed by Q_2, and the smallest benefit for the first quantity of interest, Q_1.

We compare our implementation of MSG-MIMC based on Algorithm 3 with unbiased MIMC, where no samples are reused. We did not actually perform a simulation for the latter, but just added to the total computational cost, the additional cost inferred by computing the recycled samples using FMSG. All simulations are performed on a parallel computer with 24 logical cores. The results are outlined in Figure 5. On the left in the figure is the average run time for five different random number generator seeds, plotted against a decreasing sequence of tolerances on the RMSE. On the right is the corresponding normalized run time. By recycling samples, we can achieve a cost reduction factor of 2 or more. Furthermore, the results are in agreement with our theoretical analysis: The largest speedup is observed for the third quantity of interest Q_3, at the bottom of the figure, because the rate of decrease in variance, $\beta_j, j = 1, 2$ is smallest. The least improvement is observed for the first quantity of interest, Q_1, at the top of the figure, because the rate of decrease in variance, $\beta_j, j = 1, 2$ is highest.

The shape of the index set for selected tolerances ε is shown in Figures 6–8. Superimposed on the figure, using logarithmic color codes, is the number of samples taken on each of these indices. By Equation (27), this corresponds to a view of the learned probability mass function $\pi_{\boldsymbol{L}}$. Notice the effect of the anisotropic field, combined with a nonsymmetric quantity of interest, Q_3, on the final shape of the index set in Figure 8. It is apparent that most refinement is needed in the y-direction.

Finally, we should mention that the MSG-MIMC estimator for Q_3 is no longer unbiased for $\varepsilon < 1.334\text{e}{-2}$: the algorithm requested a sample of the multi-index difference on a grid with $p = 0$ and $q = 19$, which is beyond the capabilities of our current hardware, due to memory constraints. We argue, however, that this bias would be unavoidable, even when using the classic MIMC estimator.

ACKNOWLEDGMENTS

This research was funded by project IWT/SBO EUFORIA: "Efficient Uncertainty Quantification For Optimization in Robust design of Industrial Applications" (IWT-140068) of the Agency for Innovation by Science and Technology, Flanders, Belgium.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Pieterjan Robbe https://orcid.org/0000-0002-6254-8245

REFERENCES

1. Mulder WA. A new multigrid approach to convection problems. J Comput Phys. 1989;83(2):303–323.
2. Naik NH, Van Rosendale J. The improved robustness of multigrid elliptic solvers based on multiple semicoarsened grids. SIAM J Numer Anal. 1993;30(1):215–229.
3. Oosterlee CW, Wesseling P. On the robustness of a multiple semi-coarsened grid method. Z Angew Math Mech. 1995;75(4):251–257.
4. Giles MB. Multilevel Monte Carlo path simulation. Oper Res. 2008;56(3):607–617.
5. Haji-Ali AL, Nobile F, Tempone R. Multi-index Monte Carlo: When sparsity meets sampling. Numer Math. 2016;132(4):767–806.
6. Kumar P, Oosterlee CW, Dwight RP. A multigrid multilevel Monte Carlo method using high-order finite-volume scheme for lognormal diffusion problems. Int J Uncertain Quantif. 2017;7(1):57–81.

7. Robbe P, Nuyens D, Vandewalle S. A dimension-adaptive multi-index Monte Carlo method applied to a model of a heat exchanger. In: Owen AB, Glynn P, editors. Monte Carlo and quasi-Monte Carlo methods (MCQMC 2016). Proceedings in Mathematics & Statistics. Volume 241. Cham: Springer, 2018; p. 429–445.
8. Rhee C, Glynn PW. Unbiased estimation with square root convergence for SDE models. Oper Res. 2015;63(5):1026–1043.
9. Matérn B. Spatial variation – Stochastic models and their application to some problems in forest survey sampling investigations. Rep Forest Res Inst Sweden. 1960;49:1–144.
10. Hiscock KM, Bense VF. Hydrogeology: Principles and practice. 2nd ed. Hoboken, New Jersey: Wiley-Blackwell, 2014.
11. Delhomme J. Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. Water Resour Res. 1979;15(2):269–280.
12. Hackbusch W. Multi-grid methods and applications. Berlin Heidelberg: Springer-Verlag, 1985.
13. Trottenberg U, Oosterlee CW, Schuller A. Multigrid. Cambridge, Massachusetts: Academic Press, 2000.
14. Chow E, Falgout RD, Hu JJ, Tuminaro RS, Yang UM. A survey of parallelization techniques for multigrid solvers. In: Heroux MA, Raghavan P, Simon HD, editors. Parallel Processing for Scientific Computing. Philadelphia: SIAM, 2006; p. 179–201.
15. De Zeeuw PM. Matrix-dependent prolongations and restrictions in a blackbox multigrid solver. J Comput Appl Math. 1990;33(1):1–27.
16. Dendy JE. Black box multigrid. J Comput Phys. 1982;48(3):366–386.
17. Kumar P, Rodrigo C, Gaspar F, Oosterlee C. On local fourier analysis of multigrid methods for PDEs with jumping and random coefficients. SIAM J Sci Comput. 2019;41(3):A1385–A1413.
18. Detommaso G, Dodwell T, Scheichl R. Continuous level Monte Carlo and sample-adaptive model hierarchies. SIAM/ASA J Uncertain Quantif. 2019;7(1):93–116.
19. Robbe P, Nuyens D, Vandewalle S. Recycling samples in the multigrid multilevel (Quasi-)Monte Carlo method. SIAM J Sci Comput. 2019;41(5):S37–S60.
20. Graham I, Kuo F, Nuyens D, Scheichl R, Sloan I. Analysis of circulant embedding methods for sampling stationary random fields. SIAM J Numer Anal. 2018;56(3):1871–1895.
21. Latz J, Eisenberger M, Ullmann E. Fast sampling of parameterised Gaussian random fields. Comput Methods Appl Mech Eng. 2019;348:978–1012.

How to cite this article: Robbe P, Nuyens D, Vandewalle S. Enhanced multi-index Monte Carlo by means of multiple semicoarsened multigrid for anisotropic diffusion problems. *Numer Linear Algebra Appl*. 2021;28:e2281. https://doi.org/10.1002/nla.2281