
Optimization of mesh hierarchies in Multilevel Monte Carlo samplers

Abdul-Lateef Haji-Ali · Fabio Nobile ·
Erik von Schwerin · Raúl Tempone

Abstract We perform a general optimization of the parameters in the Multilevel Monte Carlo (MLMC) discretization hierarchy based on uniform discretization methods with general approximation orders and computational costs. We optimize hierarchies with geometric and non-geometric sequences of mesh sizes and show that geometric hierarchies, when optimized, are nearly optimal and have the same asymptotic computational complexity as non-geometric optimal hierarchies. We discuss how enforcing constraints on parameters of MLMC hierarchies affects the optimality of these hierarchies. These constraints include an upper and a lower bound on the mesh size or enforcing that the number of samples and the number of discretization elements are integers. We also discuss the optimal tolerance splitting between the bias and the statistical error contributions and its asymptotic behavior. To provide numerical grounds for our theoretical results, we apply these optimized hierarchies together with the Continuation MLMC Algorithm [11]. The first example considers a three-dimensional elliptic partial differential equation with random inputs. Its space discretization is based on continuous piecewise trilinear finite elements and the corresponding linear system is solved by either a direct or an iterative solver. The second example considers a one-dimensional Itô stochastic differential equation discretized by a Milstein scheme.

Keywords Multilevel Monte Carlo, Monte Carlo, Partial Differential Equations with random data, Stochastic Differential Equations, Optimal discretization

Mathematics Subject Classification (2000) 65C05 · 65N30 · 65N22

A. Haji-Ali (E-mail: abdullateef.hajiali@kaust.edu.sa) · R. Tempone (E-mail: raul.tempone@kaust.edu.sa)
Applied Mathematics and Computational Sciences, KAUST, Thuwal, Saudi Arabia.

F. Nobile
MATHICSE-CSQI, EPF de Lausanne, Switzerland.

E. von Schwerin
Department of Mathematical Sciences, University of Delaware, Newark, USA

1 Introduction

The history of Multilevel Monte Carlo methods can be traced back to Heinrich et al. [16, 17], where it was introduced in the context of parametric integration. Keibaier [22] then used similar ideas for a two-level Monte Carlo (MC) method to approximate weak solutions to stochastic differential equations (SDEs) in mathematical finance. The basic idea of the two-level MC method is to reduce the number of samples on the fine mesh by using a control variate that is obtained by approximating the solution on a coarser mesh. In [13], Giles extended this idea to more than two levels and dubbed his extension the Multilevel Monte Carlo (MLMC) method. Giles introduced a hierarchy of discretizations with geometrically decreasing mesh sizes. His work also included an optimization of the number of samples on each level that reduced the computational complexity to $\mathcal{O}(\text{TOL}^{-2}(\log \text{TOL})^2)$ when applied to SDEs with Euler-Maruyama discretization, compared to $\mathcal{O}(\text{TOL}^{-3})$ of the standard Euler-Maruyama MC method. In [12], Giles further reduced the computational complexity of approximating weak solutions of a one-dimensional SDE to $\mathcal{O}(\text{TOL}^{-2})$ by using the Milstein scheme instead of the Euler-Maruyama scheme to discretize the SDE. MLMC has also been extended and applied in many contexts, including equations with jump diffusions [29], partial differential equations (PDEs) with stochastic coefficients [6, 8, 9, 27] and stochastic partial differential equations (SPDEs) [5, 14], to compute scalar quantities of interest that are functionals of the solutions. In [27, Theorem 2.3], an optimal convergence rate is derived for general rates of strong and weak convergence and the computational complexity associated with generating a single sample of the quantity of interest. It is shown that if the strong convergence is sufficiently fast, the computational complexity can be of the optimal rate, $\mathcal{O}(\text{TOL}^{-2})$.

Several points can be investigated in this standard MLMC setting. For instance, the standard MLMC uses uniform mesh sizes on each level and across the levels the mesh sizes follow a geometric sequence in which the ratio between mesh sizes of subsequent levels is a constant, β , henceforth referred to as level separation. However, it is not clear if this is an optimal choice. Moreover, in the literature, the derivation of the optimal number of samples on each level assumed an equal, fixed splitting of accuracy between statistical and bias error contributions. In [11], the authors used a more efficient splitting that improved the running time of MLMC by a constant factor, but no analysis of the splitting parameter was provided. In this work, we show that, in certain cases, the optimal level separation is not a constant and depends on several parameters, including the level index, ℓ . Moreover, when restricted to geometric, but not nested, hierarchies, we optimize for the constant level separation parameter, β , by using some heuristics and show that using this optimized value the computational complexity of the geometric hierarchies is close to the computational complexity of the optimized non-geometric hierarchies. We also show that the computational complexity of both hierarchies are the same in the limit as $\text{TOL} \rightarrow 0$. In addition, we analyze the optimal splitting parameter, θ , and note its asymptotic behavior as $\text{TOL} \rightarrow 0$. Several issues arise in a practical implementation of MLMC. One of these issues is that the hierarchies generated by optimality theorems are usually not applicable due to constraints on either mesh sizes (for instance due to CFL stability limitations) or the number of samples; the constraint on the latter being an integer, for example. We analyze these issues and note their effect on the optimality of the MLMC hierarchies. Other issues include

the stopping criteria [7] and the estimation of variances in the case of a small number of samples, a feature that is inherent to MLMC and is always present in the deepest levels of the MLMC hierarchies. To this end, we here apply these optimized hierarchies together with the Continuation MLMC algorithm (CMLMC) [11] and show the effectiveness of the resulting algorithm in several examples. The use of *a posteriori* error estimates and related adaptive algorithms, as introduced first in [18], is beyond the scope of this work, which focuses instead on optimizing *a priori* defined parametric families to create the discretization hierarchies.

This work is organized as follows. Section 2.1 recalls the MLMC sampling framework and states the hierarchy optimization problem. Several approximation steps lead to an analytically treatable problem. Section 2.2 presents the solution for the case of unconstrained optimal mesh sizes, including the number of samples per level and the splitting accuracy parameter; these optimal mesh sizes do not form geometric sequences in general. Then, Section 2.3 presents the optimal hierarchies if they are restricted to geometric sequences of mesh sizes. Finally, Section 3 illustrates the theoretical results with numerical examples, which include three-dimensional PDEs with random inputs and Itô SDEs, and Section 4 draws conclusions and proposes future extensions of this work. To avoid cluttering the presentation, the technical derivations of the formulas included in this work are outlined in the appendix.

2 Optimal MLMC hierarchies

Here we state the problem of optimizing the mesh hierarchies in MLMC and present the mesh hierarchies resulting from a theoretical optimization, first allowing very general sequences of mesh sizes and then for comparison restricting ourselves to geometric sequences.

In Section 2.1 we introduce the MLMC hierarchy, the parameters that we consider free to optimize in the hierarchy, and the models of the computational work and of the weak and strong errors that define the general, discrete and non-convex, optimization problem. Simplifying assumptions then lead to an analytically treatable continuous optimization problem in Sections 2.2–2.3.

2.1 Problem setting

Let $g(u)$ denote a scalar quantity of interest, which is a function of the solution u of an underlying stochastic model. Our goal is to approximate the expected value, $E[g(u)]$, to a given accuracy TOL with a high probability of success. We assume that individual outcomes of the underlying solution, u , and the evaluation of $g(u)$ are approximated by a discretization-based numerical scheme characterized by a mesh size¹, h . The following examples are adapted from [11] with some modification:

Example 2.1 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and \mathcal{D} be a open, bounded and convex polygonal domain in \mathbb{R}^d . Find $u : \mathcal{D} \times \Omega \rightarrow \mathbb{R}$ that solves

¹ We consider uniform meshes, but the extension to certain non-uniform meshes is immediate; see Remark 2.2 in Section 2.2.

\mathbb{P} -almost surely (a.s.) the following equation:

$$\begin{aligned} -\nabla \cdot (a \nabla u) &= f, & \text{in } \mathcal{D}, \\ u &= 0, & \text{on } \partial \mathcal{D}, \end{aligned}$$

where the value of the diffusion coefficient and the forcing are represented by random fields, yielding a random solution. We wish to compute $E[g(u)]$ for some deterministic functional g which is globally Lipschitz satisfying $|g(u) - g(v)| \leq G\|u - v\|_{H^1(\mathcal{D})}$ for some constant $G > 0$ and all $u, v \in H^1(\mathcal{D})$. Following [27], we also make the following assumptions

- $a_{\min} = \min_{\mathbf{x} \in \mathcal{D}} a(\mathbf{x}) > 0$ a.s. and $1/a_{\min} \in L^p_{\mathbb{P}}(\Omega)$, for all $p \in (0, \infty)$.
- $a \in L^p_{\mathbb{P}}(\Omega, C^1(\overline{\mathcal{D}}))$, for all $p \in (0, \infty)$.
- $f \in L^p_{\mathbb{P}}(\Omega, L^2(\mathcal{D}))$ for some $p^* > 2$.

Here, $L^p_{\mathbb{P}}(\Omega, \mathcal{B})$ is the Bochner space of p -th integrable \mathcal{B} -valued random fields, where the p -th integrability is with respect to the probability measure \mathbb{P} . On the other hand, $C^1(\overline{\mathcal{D}})$ is the space of continuously differentiable functions with the usual norm [8]. Note that with these assumptions and since \mathcal{D} is bounded, one can show that $\max_{\mathbf{x} \in \mathcal{D}} a(\mathbf{x}) < \infty$ a.s. A standard approach to approximate the solution of the previous problem is to use finite elements on regular triangulations. In such a setting, the parameter $h > 0$ refers to either the maximum element diameter or another characteristic length and the corresponding approximate solution is denoted by u_h . For piecewise linear or piecewise d -multilinear continuous finite element approximations, and with the previous assumptions, it can be shown [27, Corollary 3.1] that asymptotically as $h \rightarrow 0$:

- $|E[g(u) - g(u_h)]| = \mathcal{O}(h^2)$.
- $\text{Var}[g(u) - g(u_h)] = \mathcal{O}(h^4)$.

Example 2.2 Here we study the weak approximation of Itô stochastic differential equations (SDEs). Let $(\Omega, \mathcal{F}, \mathbb{P})$ again denote a complete probability space and let

$$du(t) = a(t, u(t))dt + b(t, u(t))dB(t), \quad 0 < t < T, \quad (2.2)$$

where $u(t)$ is a stochastic process in \mathbb{R}^d , with randomness generated by a k -dimensional Wiener process with independent components, $B(t)$, cf. [21, 26], and $a(t, u) \in \mathbb{R}^d$ and $b(t, u) \in \mathbb{R}^{d \times k}$ are the drift and diffusion fluxes, respectively. For any given sufficiently well-behaved function, $g : \mathbb{R}^d \rightarrow \mathbb{R}$, our goal is to approximate the expected value, $E[g(u(T))]$. A typical application is to compute option prices in mathematical finance, cf. [20, 15], and other related models based on stochastic dynamics.

When one uses a standard Milstein scheme based on uniform time steps of size h to approximate (2.2), the following rates of approximation hold: $E[g(u(T)) - g(u_h(T))] = \mathcal{O}(h)$ and $E[(g(u(T)) - g(u_h(T)))^2] = \mathcal{O}(h^2)$. For suitable assumptions on the functions a , b , and g , we refer to [23].

To avoid cluttering the notation, we omit the reference to the underlying solution from now on, simply denoting the quantity of interest as g . Following the standard MLMC approach, we introduce a hierarchy of $L + 1$ meshes defined by decreasing mesh sizes $\{h_\ell\}_{\ell=0}^L$ and we denote the resulting approximation of g

using mesh size h_ℓ by g_ℓ , or by $g_\ell(z)$ when we want to stress the dependence on an outcome on some underlying random variable z . Then, the expected value of the finest approximation, g_L , can be expressed as

$$\mathbb{E}[g_L] = \mathbb{E}[g_0] + \sum_{\ell=1}^L \mathbb{E}[g_\ell - g_{\ell-1}],$$

and the MLMC estimator is obtained by approximating the expected values in the telescoping sum by sample averages as

$$\mathcal{A} = \frac{1}{M_0} \sum_{m=1}^{M_0} g_0(z_{0,m}) + \sum_{\ell=1}^L \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} (g_\ell(z_{\ell,m}) - g_{\ell-1}(z_{\ell,m})), \quad (2.3)$$

where, for every ℓ , $\{z_{\ell,m}\}_{m=1}^{M_\ell}$ denotes independent identically distributed (i.i.d.) random variables representing the underlying, mesh-independent, stochastic model. In addition, the random variables in the union of all these sets are independent. We note that, given the model for g_ℓ , the MLMC estimator is defined by the triplet $\mathbf{H} = (L, \{h_\ell\}_{\ell=0}^L, \{M_\ell\}_{\ell=0}^L)$, which we also refer to as the MLMC hierarchy. Depending on the numerical discretization method, possible mesh sizes will be restricted to a discrete set of positive real numbers, which we denote by \mathfrak{H} . For instance, for uniform meshes in the domain $(0,1)^d$, the number of subdivisions in each dimension has to be a positive integer, resulting in the constraint $h^{-1} \in \mathbb{N}_+$. We do not, however, introduce any other restriction on the mesh sizes but allow the MLMC hierarchy to use any decreasing sequence of attainable mesh sizes. Moreover, the number of samples on any level is a positive integer, $M_\ell \in \mathbb{N}_+$, while L is a non-negative integer, $L \in \mathbb{N}$.

If W_ℓ is the average cost associated with generating one sample of the difference, $g_\ell - g_{\ell-1}$, or simply g_0 if $\ell = 0$, then the cost of the estimator (2.3) is

$$W(\mathbf{H}) = \sum_{\ell=0}^L M_\ell W_\ell. \quad (2.4)$$

We assume that the work required to generate one sample of mesh size h is proportional to $h^{-d\gamma}$, where d is the dimension of the computational domain and $\gamma > 0$ represents the complexity of generating one sample with respect to the number of degrees of freedom. Thus, we model the average cost on level ℓ as

$$W_\ell = h_\ell^{-d\gamma}, \quad (2.5)$$

and consequently use the representation

$$W(\mathbf{H}) = \sum_{\ell=0}^L \frac{M_\ell}{h_\ell^{d\gamma}} \quad (2.6)$$

for the total work to evaluate the MLMC estimator (2.3). This can be motivated in two ways. Namely, we are simply neglecting the work to generate the coarser variable in each realization pair $(g_\ell, g_{\ell-1})$ or, we are bounding the work to generate the pair by a constant factor, which is clearly less than or equal to twice the work to generate the finest variable in each realization pair.

For example, if each sample evaluation is the approximation of an Itô stochastic differential equation by a time stepping scheme, then $d = \gamma = 1$. If, instead, the underlying differential equation is an elliptic partial differential equation with a stochastic coefficient field, then a numerical method based on an ideal multigrid solver will still have $\gamma = 1$ up to a logarithmic factor, while a naive implementation of Gaussian elimination based on full matrices leads to $\gamma = 3$.

We want to find a hierarchy, \mathbf{H} , which, with a prescribed failure probability, $0 < \alpha \ll 1$, satisfies

$$\mathbb{P}[|E[g] - \mathcal{A}| > \text{TOL}] \leq \alpha,$$

while minimizing the work, $W(\mathbf{H})$. Here, we aim to meet this accuracy requirement by controlling the bias and statistical error separately as

$$|E[g - \mathcal{A}]| \leq (1 - \theta)\text{TOL} \quad \text{and} \quad \mathbb{P}[|E[\mathcal{A}] - \mathcal{A}| > \theta \text{TOL}] \leq \alpha. \quad (2.7)$$

This splitting of the error introduces a new parameter, $0 < \theta < 1$, which we are free to choose. We will later see that the choice of θ that minimizes the work is not obvious, and does not reduce to any simple rule of thumb. Motivated by the Lindeberg-Feller Central Limit Theorem in the limit $\text{TOL} \rightarrow 0$; see [11, Lemma A.2], the probabilistic constraint in (2.7) can be replaced by a constraint on the variance of the estimator as follows:

$$\text{Var}[\mathcal{A}] \leq \left(\frac{\theta \text{TOL}}{C_\alpha} \right)^2, \quad (2.8)$$

where C_α satisfies $\Phi(C_\alpha) = 1 - \frac{\alpha}{2}$; here, Φ is the standard normal cumulative distribution function.

By construction of the estimator, $E[\mathcal{A}] = E[g_L]$ and using the notation

$$V_\ell = \begin{cases} \text{Var}[g_0] & \ell = 0, \\ \text{Var}[g_\ell - g_{\ell-1}] & \ell > 0, \end{cases}$$

and by independence we have $\text{Var}[\mathcal{A}] = \sum_{\ell=0}^L V_\ell M_\ell^{-1}$. The requirements (2.7) and (2.8) therefore become

$$|E[g - g_L]| \leq (1 - \theta)\text{TOL}, \quad (2.9a)$$

$$\sum_{\ell=0}^L V_\ell M_\ell^{-1} \leq \left(\frac{\theta \text{TOL}}{C_\alpha} \right)^2. \quad (2.9b)$$

We now assume that the numerical approximation of g_ℓ leads to weak convergence of order q_1 and strong convergence of order $q_2 \leq 2q_1$ as $h \rightarrow 0$, and we further assume that the variance on the coarsest level is approximately independent of its corresponding mesh size. Note that the condition that is usually assumed for MLMC is $\min(q_2, d\gamma) \leq 2q_1$, c.f. [27, Theorem 2.3], to ensure that the cost of MLMC is not dominated by the cost of a single sample on each level. We assume instead the slightly more restrictive condition $q_2 \leq 2q_1$, since it does not depend on the dimensionality of the problem, d . Using these assumptions and neglecting

all higher order terms in h_ℓ , we postulate for some constants, $0 < Q_W, Q_S < \infty$ the following models for the bias and variances:

$$\begin{aligned} |\mathbb{E}[g - g_L]| &= Q_W h_L^{q_1}, \\ V_\ell &= Q_S h_{\ell-1}^{q_2}, \quad \text{for } \ell > 0. \end{aligned} \quad (2.10)$$

These models are reasonable for many cases (such as those listed in Section 3) and hold asymptotically as $\ell \rightarrow \infty$. We limit ourselves to these cases, while acknowledging that there are others, such as those in [25, 28], that do not follow the model (2.10). We observe that the problem of finding $\mathbf{H} = (L, \{h_\ell\}_{\ell=0}^L, \{M_\ell\}_{\ell=0}^L) \in \mathbb{N} \times \mathfrak{H}^{L+1} \times \mathbb{N}_+^{L+1}$ minimizing $W(\mathbf{H})$ in (2.6) while satisfying the constraints (2.9) is a difficult discrete optimization problem. Hence, we make a further simplification by temporarily removing the constraints on h_ℓ and M_ℓ to let $\mathbf{H} \in \mathbb{N} \times \mathbb{R}_+^{L+1} \times \mathbb{R}_+^{L+1}$. The simplified variance model (2.10) is valid for example for nested geometric sequences of mesh sizes, but in our more general setting in this paper it can instead be seen as a penalty on closely spaced meshes, where it overestimates the resulting variance.

The simplified models for the bias and the variance of the MLMC estimator are then

$$|\mathbb{E}[g - \mathcal{A}]| = Q_W h_L^{q_1}, \quad (2.11a)$$

$$\text{Var}[\mathcal{A}] = \frac{V_0}{M_0} + Q_S \sum_{\ell=1}^L \frac{h_{\ell-1}^{q_2}}{M_\ell}, \quad (2.11b)$$

with problem- and method-specific positive constants, Q_W , Q_S , and V_0 . We note that neglecting the higher-order terms in h_ℓ is usually justified in the model of the bias, which only depends on the finest mesh. On the other hand, it may be that the contribution of the higher-order terms on coarse meshes makes the model (2.11b) inaccurate, causing the hierarchies derived in this work to be suboptimal. Dealing with such non-asymptotic behavior is beyond the scope of this work and we leave it for future work.

2.2 General mesh size sequences

Here, we present the optimal hierarchy, \mathbf{H} , using the continuous, convex, model of the previous subsection, which solves:

Problem 2.1 Find $\mathbf{H} = (L, \{h_\ell\}_{\ell=0}^L, \{M_\ell\}_{\ell=0}^L) \in \mathbb{N} \times \mathbb{R}_+^{L+1} \times \mathbb{R}_+^{L+1}$ such that

$$W(\mathbf{H}) = \sum_{\ell=0}^L \frac{M_\ell}{h_\ell^{d\gamma}}, \quad (2.12a)$$

is minimized while satisfying the constraints

$$Q_W h_L^{q_1} \leq (1 - \theta) \text{TOL}, \quad (2.12b)$$

$$\frac{V_0}{M_0} + Q_S \sum_{\ell=1}^L \frac{h_{\ell-1}^{q_2}}{M_\ell} \leq \left(\frac{\theta \text{TOL}}{C_\alpha} \right)^2, \quad (2.12c)$$

for some $\theta \in (0, 1)$.

Note that, even though the parameter θ is not part of the hierarchy \mathbf{H} defining the MLMC estimator, determining θ is still an important part of the optimization. Initially, we treat the parameters θ and L as given and optimize first with respect to $\{M_\ell\}_{\ell=0}^L$ and then $\{h_\ell\}_{\ell=0}^L$. From a Lagrangian formulation of the problem of minimizing the general work model (2.4) under the constraint (2.9b), it is straightforward to obtain the optimal number of samples,

$$M_\ell = \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2 \sqrt{\frac{V_\ell}{W_\ell}} \sum_{k=0}^L \sqrt{W_k V_k}, \quad (2.13)$$

in terms of general work estimates, $\{W_\ell\}_{\ell=0}^L$, and variance estimates, $\{V_\ell\}_{\ell=0}^L$; see Section A.1 for more details on this and the following steps. The finest mesh size is determined by the bias constraint (2.12b), for any given choice of θ . The optimality conditions then lead to a linear difference equation, which can easily be solved for the remaining mesh sizes. In the idealized situation, where the coarsest mesh size is treated as an unconstrained variable in the optimization, we can analytically minimize the computational complexity with respect to θ to obtain the optimal hierarchy for any fixed L . Introducing the two model- and method-dependent parameters,

$$\eta = \frac{q_1}{d\gamma} \quad \text{and} \quad \chi = \frac{q_2}{d\gamma}, \quad (2.14)$$

we can summarize the result derived in Sections A.1.1 and A.1.2 in the following theorems for the two cases: $\chi = 1$ and $\chi \neq 1$.

Theorem 2.1 (On the optimal hierarchies when $\chi = 1$) *For any fixed $L \in \mathbb{N}$, with $\chi = 1$, the optimal sequences $\{h_\ell\}_{\ell=0}^L$ and $\{M_\ell\}_{\ell=0}^L$ in Problem 2.1 are given by*

$$h_\ell = \beta^{L-\ell} \left(\frac{(1-\theta)\text{TOL}}{Q_W} \right)^{\frac{1}{q_1}}, \quad \text{for } \ell = 0, 1, 2, \dots, L, \quad (2.15a)$$

$$M_\ell = \beta^{-q_2 \ell} V_0 (L+1) \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2, \quad \text{for } \ell = 0, 1, 2, \dots, L, \quad (2.15b)$$

where the level separation $\beta \in (1, \infty)$ is independent of ℓ ,

$$\beta = \left\{ \left(\frac{Q_W}{(1-\theta)\text{TOL}} \right)^{\frac{1}{q_1}} \left(\frac{V_0}{Q_S} \right)^{\frac{1}{q_2}} \right\}^{\frac{1}{L+1}}, \quad (2.15c)$$

and the optimal choice of the splitting parameter

$$\theta(1, \eta, L) = \left(1 + \frac{1}{2\eta} \frac{1}{L+1} \right)^{-1}. \quad (2.15d)$$

Lemma 2.1 *For the case $\chi = 1$ and the optimal hierarchies in Theorem 2.1, the optimal number of levels, L , satisfies*

$$1 < \frac{2\eta(L+1)}{\log(\text{TOL}^{-1} Q_W V_0^\eta Q_S^{-\eta})} \leq \frac{1}{\exp(1) - 1} \left(\exp(1) + \frac{1}{\log(\text{TOL}^{-1} Q_W V_0^\eta Q_S^{-\eta})} \right), \quad (2.16)$$

for any $\text{TOL} < Q_W V_0^\eta Q_S^{-\eta}$, and asymptotically

$$\lim_{\text{TOL} \rightarrow 0} \frac{L+1}{\log \text{TOL}^{-1}} = \frac{1}{2\eta}. \quad (2.17)$$

Corollary 2.1 *For the case $\chi = 1$ and the optimal hierarchies in Theorem 2.1 and using L in (2.17), the total work (2.6) satisfies*

$$\frac{W(\mathbf{H})}{\text{TOL}^{-2}(\log \text{TOL})^2} \rightarrow C_\alpha^2 \exp(2) Q_S \left(\frac{1}{2\eta} \right)^2, \quad \text{as } \text{TOL} \searrow 0. \quad (2.18)$$

Theorem 2.2 (On the optimal hierarchies when $\chi \neq 1$) *For any fixed $L \in \mathbb{N}$, with $\chi \neq 1$, the optimal sequences, $\{h_\ell\}_{\ell=0}^L$ and $\{M_\ell\}_{\ell=0}^L$, in Problem 2.1 are given by*

$$h_\ell(\theta, L) = \left(\frac{(1-\theta) \text{TOL}}{Q_W} \right)^{\frac{1}{q_1} \frac{1-\chi^{\ell+1}}{1-\chi^{L+1}}} \left(\frac{V_0}{Q_S} \right)^{\frac{1}{d_\gamma} \frac{\chi^\ell - \chi^L}{1-\chi^{L+1}}} \\ \cdot \chi^{-\frac{1}{d_\gamma} \frac{2}{1-\chi} \left(\frac{\chi^{L+1} - \chi^{\ell+1}}{1-\chi^{L+1}} + \frac{L(1-\chi^{\ell+1}) - \ell(1-\chi^{L+1})}{1-\chi^{L+1}} \right)}, \quad (2.19a)$$

$$M_\ell(\theta, L) = \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2 ((1-\theta) \text{TOL})^{\frac{\chi}{\eta} \frac{1-\chi^\ell}{1-\chi^{L+1}}} V_0^{\frac{\chi^\ell - \chi^{L+1}}{1-\chi^{L+1}}} \\ \cdot \left(\frac{Q_S^{1/\chi}}{Q_W^{1/\eta}} \right)^{\frac{\chi(1-\chi^\ell)}{1-\chi^{L+1}}} \frac{1-\chi^{L+1}}{\chi^L(1-\chi)} \chi^{\left\{ -\frac{2\chi}{1-\chi} \frac{1-\chi^\ell}{1-\chi^{L+1}} (L+1) + \frac{1+\chi^\ell}{1-\chi} \right\}}, \quad (2.19b)$$

where the optimal choice of the splitting parameter is

$$\theta(\chi, \eta, L) = \left(1 + \frac{1}{2\eta} \frac{1-\chi}{1-\chi^{L+1}} \right)^{-1}. \quad (2.19c)$$

Lemma 2.2 *For the case $\chi \neq 1$ and the optimal hierarchies in Theorem 2.2, the optimal number of levels, L , satisfies*

$$\frac{1}{c_2} \left(1 + \frac{c_1 + \log(1+2\eta)}{\log(\text{TOL}^{-1})} \right) < \frac{L+1}{\log(\text{TOL}^{-1})} < \begin{cases} \frac{1}{c_2} \left(1 + \frac{c_1 + \log\left(1 + \frac{2\eta}{1-\chi}\right)}{\log(\text{TOL}^{-1})} \right), & \chi \in (0, 1), \\ \frac{\chi}{c_2} \left(1 + \frac{c_1 + \log\left(\frac{2\eta}{\chi-1}\right)}{\log(\text{TOL}^{-1})} \right), & \chi \in (1, \infty), \end{cases} \quad (2.20)$$

where

$$c_1 = \log \left(\frac{V_0^{\eta/\chi}}{Q_S^{\eta/\chi}} Q_W \right) \quad \text{and} \quad c_2 = \log(\chi) \frac{2\eta}{\chi-1} > 0, \quad (2.21)$$

and asymptotically

$$\frac{1}{2\eta} \frac{\chi-1}{\log \chi} \leq \liminf_{\text{TOL} \rightarrow 0} \frac{L+1}{\log(\text{TOL}^{-1})} \leq \limsup_{\text{TOL} \rightarrow 0} \frac{L+1}{\log(\text{TOL}^{-1})} \leq \frac{\max\{1, \chi\}}{2\eta} \frac{\chi-1}{\log \chi}. \quad (2.22)$$

Corollary 2.2 *For the case $\chi \neq 1$ and the optimal hierarchies in Theorem 2.2 and using the upper bound on L in (2.20), the total work (2.6) satisfies*

$$\frac{W(\mathbf{H})}{\text{TOL}^{-2\left(1+\frac{1-\chi}{2\eta}\right)}} \rightarrow C_1, \quad \text{as } \text{TOL} \searrow 0 \text{ for } \chi \in (0, 1), \text{ and} \quad (2.23a)$$

$$\frac{W(\mathbf{H})}{\text{TOL}^{-2}} \rightarrow C_2, \quad \text{as } \text{TOL} \searrow 0 \text{ for } \chi > 1, \quad (2.23b)$$

with known constants of proportionality,

$$C_1 = C_\alpha^2 Q_S Q_W^{\left\{\frac{1-\chi}{\eta}\right\}} \chi^{\left\{-\frac{2\chi}{1-\chi}\right\}} \left(\frac{1}{2\eta}\right)^2 \left(1 + \frac{2\eta}{1-\chi}\right)^{2\left(1+\frac{1-\chi}{2\eta}\right)}, \quad (2.24a)$$

$$C_2 = C_\alpha^2 V_0^{\left\{\frac{\chi-1}{\chi}\right\}} Q_S^{\left\{\frac{1}{\chi}\right\}} \chi^{2\left\{\frac{\chi}{\chi-1}\right\}} (\chi-1)^{-2}. \quad (2.24b)$$

Note that the parameter θ controlling the split between the statistical and discretization errors depends non-trivially on the problem parameters. The above theorem shows that the choice of $\theta = 1/2$, used for example in the initial works by Giles [13, 12] and by some of the authors of the present work in [19, 18] for adaptive MLMC, is increasingly suboptimal as the number of levels increases. To further understand the splitting parameter, θ , we consider the asymptotic behavior as $L(\text{TOL}) \rightarrow \infty$ and see that

$$\theta(\chi, \eta, L) \rightarrow 1, \quad \text{as } L \rightarrow \infty, \text{ if } \chi \geq 1, \quad (2.25a)$$

$$\theta(\chi, \eta, L) \rightarrow \frac{1}{1 + \frac{1-\chi}{2\eta}}, \quad \text{as } L \rightarrow \infty, \text{ if } \chi < 1. \quad (2.25b)$$

The qualitative observations here are: 1) if the strong convergence is sufficiently fast, that is $\chi \geq 1$, almost all the tolerance is allocated to the statistical error (forcing the discretization to be fine), and 2) for slower strong convergence, $\chi < 1$, the tolerance can be shifted either towards the statistical error or towards the bias according to

$$\begin{aligned} \lim_{L \rightarrow \infty} \theta(\chi, \eta, L) &> \frac{1}{2} \text{ (stat. error larger) }, & \text{if } \chi < 1 < \chi + 2\eta, \\ \lim_{L \rightarrow \infty} \theta(\chi, \eta, L) &< \frac{1}{2} \text{ (stat. error smaller) }, & \text{if } \chi + 2\eta < 1. \end{aligned}$$

Finally, we note that the value of the optimal splitting parameter, θ , in (2.15d) for $L = 0$ is consistent with the single level adaptive Monte Carlo analysis in [24].

Since the above theorems give the optimal $\{h_\ell\}_{\ell=0}^L$ and $\{M_\ell\}_{\ell=0}^L$ for any given $L \in \mathbb{N}$, it is easy to find the optimal L by doing an extensive search over a finite range of integer values. In typical cases, for computationally feasible tolerances, L is a small non-negative integer, $0 \leq L \leq 10$; we can also use the obtained bounds on the optimal value of L to delimit the range of possible integer values. Moreover, using the optimal sequences $\{h_\ell\}_{\ell=0}^L$ and $\{M_\ell\}_{\ell=0}^L$ for any given L , we have observed that the total computational complexity is usually rather insensitive to the value of L near the optimum.

We observe that the rates in the asymptotic complexity in Corollaries 2.1 and 2.2 are the same ones obtained with more restrictive assumptions on the

sequences of mesh sizes; see for instance [9, Theorem 1] and Section 2.3. With the optimal number of levels, the optimized hierarchies minimize the multiplicative constants in the complexity without improving the rate. In Corollary 2.2, the blow up of the constants C_1 and C_2 as $\chi \rightarrow 1$ corresponds to the need for including the $\log(\text{TOL}^{-1})^2$ factor that appears in the complexity of MLMC when $\chi = 1$ as Corollary 2.1 shows.

The ratio between two successive mesh sizes in Theorem 2.2 has the following complicated, non-constant expression:

$$\frac{h_{\ell+1}}{h_\ell} = \left(\frac{V_0}{Q_S} \right)^{-\frac{(\chi-1)\chi^\ell}{d_\gamma(\chi^{L+1}-1)}} \chi^{\frac{2}{d_\gamma} \left(\frac{1}{1-\chi} + \frac{(L+1)\chi^{\ell+1}}{\chi^{L+1}-1} \right)} \left(\frac{(1-\theta)\text{TOL}}{Q_W} \right)^{\frac{(\chi-1)\chi^{\ell+1}}{q_1(\chi^{L+1}-1)}}. \quad (2.26)$$

Remark 2.1 (On relation to geometric hierarchies) Clearly, when $\chi \neq 1$, the optimal mesh sequences are not geometric in general. On the other hand, according to Theorem 2.1, the optimal mesh sequences are indeed geometric when $\chi = 1$. We further note that an asymptotic analysis when $\text{TOL} \rightarrow 0$, using optimal θ and L , shows that for L sufficiently large, the ratios (2.26) are approximately the constant $\chi^{\frac{2}{d_\gamma(\chi-1)}}$ over most of the range of ℓ -values. In case $\chi < 1$, this holds for $\ell \gg 0$, and in case $\chi > 1$, for $\ell \ll L$. In both cases, these are the levels where most of the computational work would be spent using geometrically spaced meshes. This suggests that a geometric hierarchy with this level separation constant can be nearly optimal. We show that this is the case in Section 2.3.

Remark 2.2 (On non-uniform meshes) The optimization and the resulting optimal hierarchies do not depend on the assumption that the discretizations were uniform. Indeed, h_ℓ can also be interpreted as a more general mesh parameter that defines a mesh size, Δx_ℓ , of the underlying discretization as

$$\Delta x_\ell = r(h_\ell, x),$$

for some mesh grading function $r(h_\ell, x)$, allowing for example, for local a priori refinement of meshes close to known singularities in the computational domain. As long as approximate models (2.6) and (2.11) can be provided in terms of the mesh parameter, the expressions for the optimal hierarchies in Theorems 2.1 and 2.2 can still be applied. As mentioned previously, the construction of MLMC hierarchies based on the use of a posteriori error estimates and related adaptive algorithms, as introduced first in [18], is out of the scope of the present work.

Remark 2.3 (On a lower bound on possible mesh sizes) Since equations (2.15a)–(2.15c) and (2.19a)–(2.19b) are expressed in terms of general θ and L , they remain valid when an additional constraint is imposed on the smallest possible mesh sizes. If for example the available computer memory dictates a lower limit on the practical mesh sizes, $h_\ell \geq h_{\min}$, then the optimal splitting for given L is

$$\theta(\chi, \eta, L) = \begin{cases} \min \left\{ 1 - \frac{Q_W h_{\min}^{q_1}}{\text{TOL}}, \left(1 + \frac{1}{2\eta} \frac{1-\chi}{1-\chi^{L+1}} \right)^{-1} \right\}, & \text{if } \chi \neq 1, \\ \min \left\{ 1 - \frac{Q_W h_{\min}^{q_1}}{\text{TOL}}, \left(1 + \frac{1}{2\eta(L+1)} \right)^{-1} \right\}, & \text{if } \chi = 1, \end{cases} \quad (2.27)$$

where tolerances $\text{TOL} \leq Q_W h_{\min}^{q_1}$ are out of reach of the computation. Such an extra constraint can in turn cause the optimal number of levels to be smaller

than the lower bound in (2.20) or (2.16), but it can still easily be found by an extensive search over a small integer set; the asymptotic bounds (2.17) and (2.22) are obviously not relevant then.

Remark 2.4 (On an upper bound on possible mesh sizes) If the coarsest meshes in (2.19a) or (2.15a) are unfeasibly large for the given method of discretization, for instance due to CFL stability constraints, or the asymptotic models that we assumed are only valid for small enough h_0 , then we should treat the largest mesh size, h_0 , as fixed. We briefly analyze this case at the end of Section A.1.2 for the case $\chi \neq 1$. There, we can still express all remaining mesh sizes in terms of h_0 and h_L by (A.16), and use (2.13) for the optimal number of samples on the resulting sequence of mesh sizes. However, we no longer have an explicit expression for the optimal splitting parameter, but only bounds from below and above in (A.29). Since L varies over a finite integer range, we can easily obtain the optimal θ and L in a two-stage numerical optimization.

Remark 2.5 The optimized h_ℓ in (2.15a) and (2.19a) do not necessarily belong to \mathfrak{H} and might be unusable in an actual computation. We instead use the closest element in \mathfrak{H} to each h_ℓ . For example, for uniform meshes in the domain $(0, 1)^d$ where h_ℓ^{-1} is the number of elements along every dimension, we can simply round h_ℓ^{-1} up to the nearest integer. Similarly, M_ℓ in (2.15b) and (2.19b) or equivalently (2.13) is not necessarily an integer and we round these expression up to the nearest integer to get an integer number of samples that can be used in actual computations; see also Remark 2.7.

2.3 Geometric mesh size sequences

In the optimal hierarchies of Problem 2.1 presented above, the mesh sizes do not form a geometric sequence except for the case $\chi = 1$. In this section, we optimize MLMC hierarchies with the more restrictive assumption that the mesh sizes *do* form a geometric sequence; that is, $h_\ell = h_0 \beta^{-\ell}$ for some positive value $\beta > 1$ and a given h_0 . The work and variance models in this case become

$$V_\ell = \begin{cases} V_0 & \ell = 0, \\ Q_S h_0^{q_2} \beta^{q_2} \beta^{-q_2 \ell} & \ell > 0, \end{cases} \quad (2.28a)$$

$$W_\ell = h_0^{-d\gamma} \beta^{d\gamma \ell}. \quad (2.28b)$$

We do *not* force β to be a positive integer corresponding to successive refinements of existing meshes but instead propose the following value of $\beta \in (1, \infty)$:

$$\beta = \begin{cases} \chi^{\frac{2}{d\gamma(\chi-1)}}, & \text{if } \chi \in \mathbb{R}_+ \setminus \{1\}, \\ \exp\left(\frac{2}{q_2}\right), & \text{if } \chi = 1, \end{cases} \quad (2.29)$$

We get this value using the asymptotic analysis in Remark 2.1 or a heuristic optimization that treats L as a real parameter (cf. Section A.2). The following corollary shows the asymptotic computational complexity of these geometric hierarchies.

Corollary 2.3 *Consider geometric hierarchies, $h_\ell = h_0\beta^{-\ell}$, for a given h_0 , and the optimal number of samples M_ℓ in (2.13) and the work and variance models (2.28). Moreover, assume that we choose β in (2.29) and choose the number of levels, L , to be the lower bound of (A.30). In other words, choose*

$$L = \left\lceil \frac{\log(h_0) - \frac{1}{q_1} \log\left(\frac{(1-\theta)\text{TOL}}{Q_w}\right)}{\log(\beta)} \right\rceil. \quad (2.30)$$

We distinguish between two cases:

- If $\chi = 1$, the optimal θ goes to 1 as $L \rightarrow \infty$, and the total work satisfies (2.18).
- Otherwise, if $\chi \neq 1$, the optimal θ satisfies (2.25) and the total work satisfies (2.23) with C_1 as defined in (2.24a) and

$$C_2 = C_\alpha^2 h_0^{d\gamma(\chi-1)} \left(\sqrt{V_0} h_0^{\frac{-q_2}{2}} + \sqrt{Q_S} \frac{\chi^{\frac{x}{\chi-1}}}{\chi-1} \right)^2. \quad (2.31)$$

Moreover, if we choose

$$h_0 = \left(\frac{V_0}{Q_S} \right)^{\frac{1}{q_2}} \chi^{\frac{2}{d\gamma(1-\chi)}}, \quad (2.32)$$

then C_2 simplifies to (2.24b). Notice that (2.32) is the limit of h_0 in Theorem 2.2 when $\text{TOL} \rightarrow 0$.

Remark 2.6 Corollary 2.3 shows that, asymptotically as $\text{TOL} \rightarrow 0$, the work and optimal splitting of the geometric hierarchies with optimal β (2.29) is exactly the same as the work and optimal splitting of the optimized hierarchies as stated in Corollaries 2.1 and 2.2.

Remark 2.7 Just as a hierarchy $\mathbf{H}_1 \in \mathbb{N} \times \mathbb{R}_+^{L+1} \times \mathbb{R}_+^{L+1}$ solving Problem 2.1 must be adjusted to satisfy the practical constraints of the discretization, $\mathbf{H}_1 \approx \mathbf{H} \in \mathbb{N} \times \mathfrak{H}^{L+1} \times \mathbb{N}_+^{L+1}$, so must a hierarchy that is geometric with a general β . Hence, the restriction to general geometric sequences of mesh sizes, without the true constraint $\{h_0\beta^{-\ell}\}_{\ell=0}^L \in \mathfrak{H}^{L+1}$, offers no practical improvement over the more general optimization in Section 2.2; we merely include the comparison here to point out that one can often find geometric hierarchies that are close to optimal hierarchies. Figure 2.1 shows the effect of applying these domain constraints to the number of elements and number of samples on the optimality of the hierarchies. This figure compares the work (2.6) of five hierarchies:

1. The “real-valued” optimized hierarchy with h_ℓ defined by (2.19a) and M_ℓ defined by (2.13),
2. The “integer-valued” hierarchy obtained by ceiling M_ℓ in (2.13) and h_ℓ^{-1} in (2.19a) to obtain an integer number of samples and an integer number of elements, respectively,
3. The real-valued geometric hierarchy with β as defined by (2.29), $h_0 = 0.5$ and M_ℓ again as defined by (2.13),
4. The integer-valued geometric hierarchy obtained by using the ceiling of both M_ℓ in (2.13) and the previous h_ℓ^{-1} .

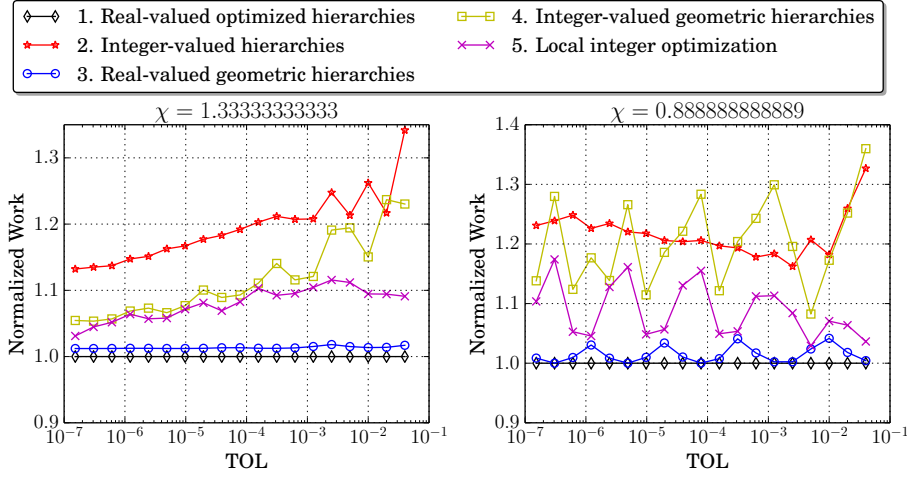


Fig. 2.1 Ex.1: Work, according to (2.6), of different hierarchies normalized by the work estimate of the “real-valued” optimized hierarchy. Taking the ceiling of h_ℓ^{-1} and M_ℓ seems to produce near-optimal hierarchies. To generate these hierarchies we used the parameters of **Ex.1** in Table 3.1 (See Remark 2.7).

5. Finally, a hierarchy obtained by performing a limited brute-force integer optimization in the neighboring integer space around the optimized h_ℓ^{-1} and M_ℓ in this work.

In all cases, we use the parameters of **Ex.1** in Table 3.1. Similar plots can be produced with different values. On the other hand, the number of levels, L , was numerically optimized and chosen according to Figure 2.2. These plots show that simply taking the ceiling of the number of samples and number of elements produces a hierarchy that is nearly optimal. Notice also in Figure 2.2 that the optimal L of the optimized real-valued hierarchies is well within the developed bounds (2.20), up to an integer rounding. However, the bounds no longer hold when considering integer-valued hierarchies.

Remark 2.8 (When optimal hierarchies become geometric) Recall that for $\chi \neq 1$, given L , h_0 , and h_L the optimal intermediate mesh sizes satisfy (A.16) which corresponds to the ratios

$$\frac{h_\ell}{h_{\ell+1}} = \left(\frac{h_0}{h_L} \right)^{\chi^\ell \frac{1-\chi}{1-\chi^L}} \chi^{\frac{2}{d\gamma} \left(\frac{L\chi^\ell}{1-\chi^L} - \frac{1}{1-\chi} \right)} \quad (2.23)$$

between successive levels. Keep the optimal choices of $h_0(\theta, L)$ and $h_L(\theta, L)$ in (2.19a), let β satisfy (2.29) and pick θ to solve the equation

$$h_L(\theta, L) = h_0(\theta, L)\beta^{-L} \quad (2.34)$$

given L . This means choosing the error splitting parameter

$$\theta = 1 - Q_W \text{TOL}^{-1} \beta^{-q_1(L+1)} \left(\frac{V_0}{Q_S} \right)^{\frac{\eta}{x}}$$

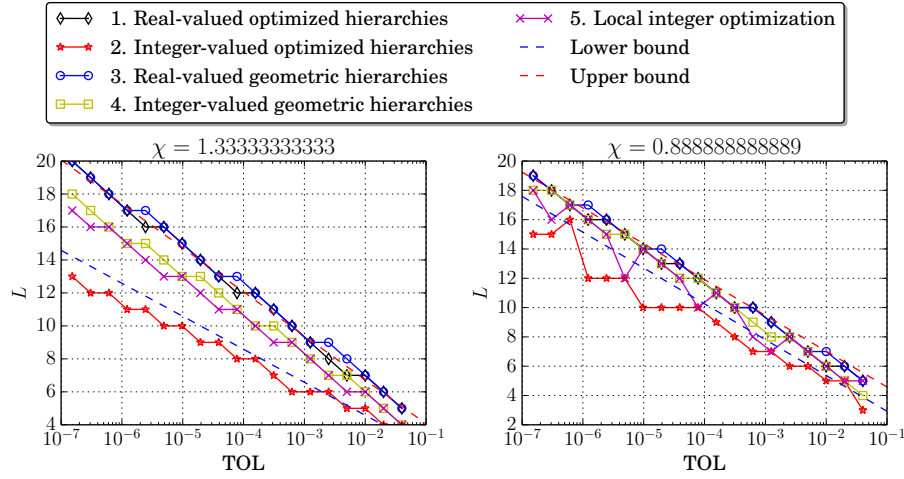


Fig. 2.2 Ex.1: Optimal L of different hierarchies. Here the bounds are from (2.20). To generate these hierarchies we used the parameters of **Ex.1** in Table 3.1 (See Remark 2.7).

for any $L \in \mathbb{N}$ sufficiently large to make the bias smaller than the total error tolerance, that is $\theta > 0$. Now, by (2.34) the ratios (2.33) reduces to $h_\ell/h_{\ell+1} = \beta$. In other words, given TOL, to any sufficiently large L there corresponds at least one (suboptimal) θ such that the corresponding optimized hierarchy is geometric.

3 Numerical results

In this section, we first introduce the two test problems: a geometric Brownian motion SDE for which $\chi > 1$ and a random PDE for which $\chi < 1$ or $\chi > 1$, depending on the linear solver used to solve the corresponding linear system. We then describe several implementation details and finally conclude by presenting the actual numerical results. We do not show results for the case $\chi = 1$ since we proved that geometric hierarchies are optimal in this case and similar results can be found in the standard work of Giles [13].

3.1 Overview of examples

We consider two numerical examples for which we can compute a reference solution.

3.1.1 Ex.1

This problem is based on Example 2.1 in Section 2.1 with some particular choices that satisfy the assumptions therein. First, we choose $\mathcal{D} = (0, 1)^3$ and assume that the forcing is

$$f(\mathbf{x}) = f_0 + \hat{f} \sum_{i=0}^K \sum_{j=0}^K \sum_{k=0}^K \Phi_{ijk}(\mathbf{x}) Z_{ijk},$$

where

$$\Phi_{ijk}(\mathbf{x}) = \sqrt{\lambda_i \lambda_j \lambda_k} \phi_i(x_1) \phi_j(x_2) \phi_k(x_3),$$

and

$$\phi_i(x) = \begin{cases} \cos\left(\frac{5\Lambda i}{2}\pi x\right), & i \text{ is even,} \\ \sin\left(\frac{5\Lambda(i+1)}{2}\pi x\right), & i \text{ is odd,} \end{cases}$$

$$\lambda_i = (2\pi)^{\frac{7}{6}} \Lambda^{\frac{11}{6}} \begin{cases} \frac{1}{2}, & i = 0, \\ \exp\left(-2\left(\pi \frac{i}{2}\Lambda\right)^2\right), & i \text{ is even,} \\ \exp\left(-2\left(\pi \frac{i+1}{2}\Lambda\right)^2\right), & i \text{ is odd,} \end{cases}$$

for given parameters Λ , positive, and K , positive integer, and $\mathbf{Z} = \{Z_{ijk}\}$, a set of $(K+1)^3$ i.i.d. standard normal random variables. Moreover, we choose the diffusion coefficient to be a function of two random variables as follows:

$$a(\mathbf{x}) = a_0 + \exp\left(4Y_1\Phi_{121}(\mathbf{x}) + 40Y_2\Phi_{877}(\mathbf{x})\right). \quad (3.1)$$

Here, $\mathbf{Y} = \{Y_1, Y_2\}$ is a set of i.i.d. standard normal random variables, also independent of \mathbf{Z} . Finally, we choose the quantity of interest, g , as a localized average around a point \mathbf{x}_0 ,

$$g = \left(2\pi\sigma^2\right)^{-\frac{3}{2}} \int_{\mathcal{D}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_0\|_2^2}{2\sigma^2}\right) u(\mathbf{x}) d\mathbf{x},$$

and select the parameters $a_0 = 0.01, f_0 = 50, \hat{f} = 10, \Lambda = \frac{0.2}{\sqrt{2}}, K = 10, \sigma^2 = 0.02622863$ and $\mathbf{x}_0 = [0.5026695, 0.26042876, 0.62141498]$. Since the diffusion coefficient, a , is independent of the forcing, f , a reference solution can be calculated to sufficient accuracy by scaling and taking expectation of the weak form with respect to \mathbf{Z} to obtain a formula with constant forcing for the conditional expectation with respect to \mathbf{Y} . We then use stochastic collocation, [2], with a sufficiently accurate quadrature to produce the reference value, $E[g]$. Using this method, the reference value 1.6026 is computed with an error estimate of 10^{-4} .

3.1.2 Ex.2

The second example is a one-dimensional geometric Brownian motion based on Example 2.2 where we make the following choices:

$$\begin{aligned} T &= 1, \\ a(t, u) &= 0.05u, \\ b(t, u) &= 0.2u, \\ g(u) &= 10 \max(u(1) - 1, 0). \end{aligned}$$

The exact solution can be computed using a change of variables and Itô's formula. For the selected parameters, the solution is 1.04505835721856.

3.2 Implementation and runs

To test the different hierarchies presented in this work we extend the CMLMC algorithm [11] to optimal hierarchies and implement it in the C programming language. The CMLMC algorithm solves problems with tolerances larger than the requested TOL to cheaply get increasingly accurate estimates of the constants Q_W and Q_S and the variances V_ℓ for all $\ell = 0, 1 \dots L$. This is achieved in a Bayesian setting to incorporate the generated samples with the models (2.11).

We stress that with these numerical results we aim to illustrate what happens when the hierarchies of Section 2 are used in a practically viable algorithm which approximates several relevant parameters during the computations. This means that we do not directly observe the optimal hierarchies as derived in the continuous optimization.

Unless θ is specified explicitly, our CMLMC algorithm uses a *computational* splitting parameter, θ , calculated based on the expected bias as

$$\theta = 1 - \frac{Q_W h_L^{q_1}}{\text{TOL}}, \quad (3.2)$$

to relax the statistical error constraint. If h_L satisfies (2.15a) or (2.19a), this is the same as the optimal splitting parameter defined by (2.15c) or (2.19c), respectively.

For implementing the solver for the PDEs in test problem **Ex.1**, we use PetIGA [10]. While the primary intent of this framework is to provide high-performance B-spline-based finite element discretizations, it is also useful in applications where the domain is topologically square and subject to uniform refinements. As its name suggests, PetIGA is designed to tightly couple to PETSc [4]. The framework can be thought of as an extension of the PETSc library, which provides methods for assembling matrices and vectors that result from integral equations. We use uniform meshes with a standard trilinear basis to discretize the weak form of the model problem, integrating it with eight quadrature points. We also generate results for two linear solvers for which PETSc provides an interface. The first solver is an **Iterative** GMRES solver that solves a linear system in almost linear time with respect to the number of degrees of freedom for the mesh sizes of interest; in other words, in this case $\gamma = 1$ and $\chi > 1$. The second solver is the **Direct** solver MUMPS [1]. For the mesh sizes of interest, the running time of MUMPS varies from quadratic to linear in the total number of degrees of freedom. The best fit turns out to be $\gamma = 1.5$ in this case, which gives $\chi < 1$. From Corollary 2.2 (or Corollary 2.3), the complexity for all the examples is expected to be $\mathcal{O}(\text{TOL}^{-s})$, where s depends on q_1, q_2 , and $d\gamma$. These and other problem parameters are summarized in Table 3.1 for the different examples. Also included in this table is the optimal level separation constant β , which we used when computing with geometric hierarchies. Obviously, as mentioned in Remark 2.5, the “real-valued” hierarchies we derived cannot always be used in practice and we follow the strategies outlined in that remark to produce “integer-valued” hierarchies that can be used.

We run each setting 100 times and show in plots in the next section the medians with vertical bars spanning from the 5% percentile to the 95% percentile. Finally, all results were generated on the same machine with 52 gigabytes of memory to ensure that no overhead is introduced due to hard disk access during swapping that could occur when solving the three-dimensional PDEs with a fine mesh. We use the parameters listed in Table 3.2 for the CMLMC algorithm [11].

	Ex.1	Ex.1	Ex.2
d	3		1
q_1	2		1
q_2	4		2
Estimated V_0	0.0565		1.7805
Estimated Q_W	1.3653		0.0307
Estimated Q_S	0.1519		0.2630
Solver	GMRES	MUMPS	Milstein
γ	1	1.5	1
χ	4/3	8/9	2
η	2/3	4/9	1
s	2	2.25	2
Optimal β	1.7778	1.6018	4
Work to seconds constant	10^{-4}	3×10^{-6}	9×10^{-8}

Table 3.1 Summary of problem parameters.

Parameter	Purpose	Value for Ex.1	Value for Ex.2
κ_0 and κ_1	Confidence parameter for the weak and strong error models	0.1 for both	0.1 for both
TOL_{\max}	The maximum tolerance with which to start the algorithm.	0.5	0.1
r_1 and r_2	Controls computational burden to calibrate the problem parameters.	2 and 1.1, respectively	2 and 1.1, respectively
Initial hierarchy	The initial hierarchy to start the CMLMC algorithm.	$L = 2$ and $h_\ell^{-1} = \{4, 6, 8\}$ and $M_\ell = 10$ for all ℓ .	$L = 2$ and $h_\ell^{-1} = \{1, 2, 4\}$ and $M_\ell = 10$ for all ℓ .
L_{inc}	Maximum number of values to consider when optimizing for L .	2	2
\mathfrak{L}	Maximum number of levels used to compute estimates of Q_W and Q_S .	3	5
C_α	Parameter related to the confidence in the statistical constraint	2	2

Table 3.2 Summary of parameter values used in the CMLMC algorithm in our numerical tests. This table is reproduced from [11] where more information is available.

3.3 Results

We start by presenting the results of **Ex.1**. We show in Figure 3.1 the total running time of the CMLMC algorithm and its last iteration when using optimal hierarchies, after taking the ceiling of the optimal number of elements h_ℓ^{-1} in (2.19a) and the optimal number of samples M_ℓ in (2.13). Using the parameters in 3.1, we also show in this figure the expected running time when using the optimal, unconstrained hierarchy defined by (2.19) and the expected asymptotic work according to Corollary (2.2). This figure shows good agreement between the expected theoretical results and the actual final running time. Figure 3.2 shows that the true error that was computed using the reference solution when using optimal hierarchies is less than the required tolerance with the required confidence of 95%, in accordance with the chosen value of $C_\alpha = 2$ and [11, Lemma A.2]. Figure 3.3 compares the computational complexity of optimal hierarchies to geometric hierarchies for different values of θ . This figure shows numerical confirmation that optimal hierarchies do not give significant improvement over geometric hierarchies,

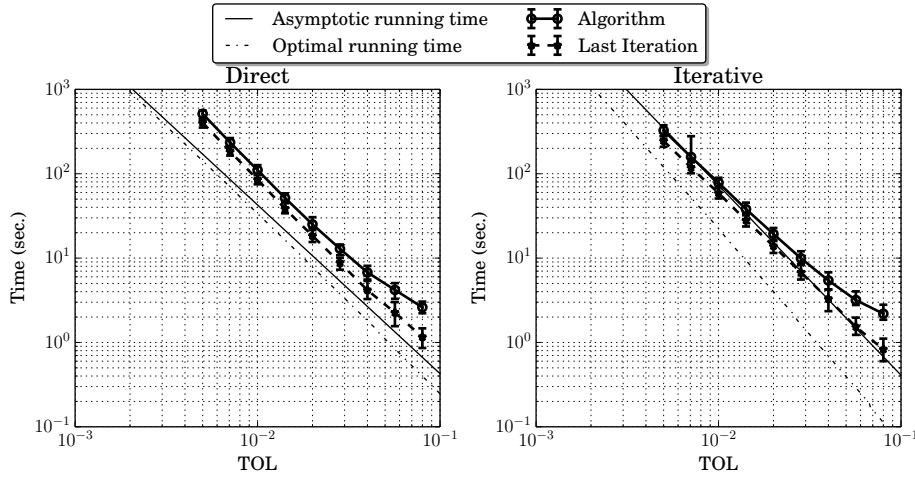


Fig. 3.1 Ex.1: The running time of the CMLMC algorithm and its last iteration when using optimal hierarchies. Using the parameters in 3.1, we also show in this figure the expected running time when using the optimal, unconstrained hierarchy defined by (2.19) and the expected asymptotic work according to Corollary (2.2).

especially for optimal values of θ . In other words, the improvement of the running time is mainly due the better choice of θ as discussed in [11]. Figure 3.5 shows the optimal splitting, θ , as defined by (2.19c). Compare this figure to Figure 3.4, which shows the used number of levels, L , for different tolerances, and notice the dependence of θ on the number of levels, L . On the other hand, Figure 3.6 shows the computational splitting used in the CMLMC algorithm. Notice that θ follows a similar pattern in both Figure 3.5 and Figure 3.6. The continuous change in the latter is due to differences in the estimation of Q_W for different runs of the algorithm. For comparison, Figure 3.7 shows that the computational splitting parameter produced when using geometric hierarchies is different from the computational splitting parameter produced when using optimal hierarchies. However, as $\text{TOL} \rightarrow 0$, they both seem to be not too far from the limit in (2.25). Finally, even though [11, Lemma A.2] assumes a geometric sequence, Figure 3.8 shows that the lemma still holds for non-geometric hierarchies; i.e., that the cumulative density function (CDF) of the true error when suitably normalized is well approximated by a standard normal CDF.

Next, we focus on **Ex.2** where $\chi = 2$ using the Milstein scheme. Since we showed previously that geometric hierarchies are near-optimal, we only present the results when using geometric hierarchies in this case. The optimal geometric constant, β , is 4 in this case according to (2.29). Figure 3.9 shows that the actual running time of the CMLMC algorithm has the expected rate TOL^{-2} , again as predicted in Corollary 2.3, or indeed for these nested geometric hierarchies already in [12]. Figure 3.10 shows that the true errors for different tolerances are less than the required tolerance with the required confidence of 95%.

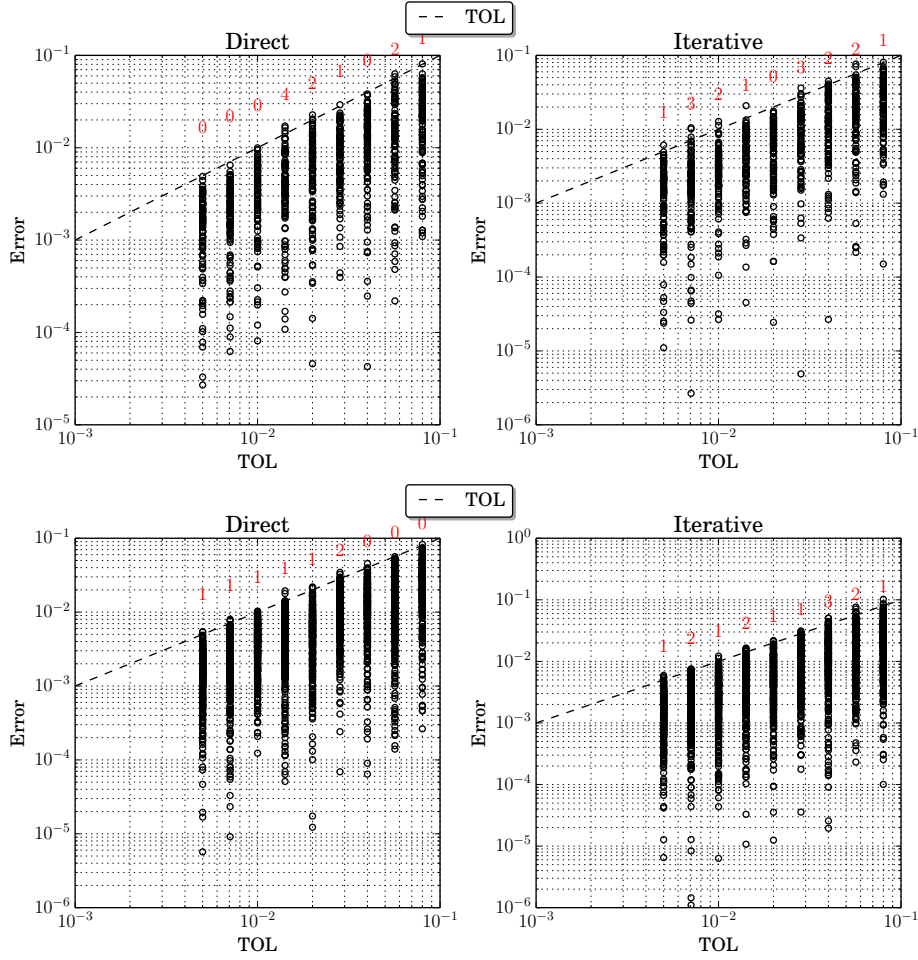


Fig. 3.2 Ex.1: The true errors calculated using optimal hierarchies in the top plot, and using geometric hierarchies in the bottom one. Each point shows the true error of an independent run of MLMC. The numbers on top of the TOL line are the percentage of algorithm runs that produced a larger error than the required tolerance. Notice that the choice $C_\alpha = 2$ gives a confidence of a 95% in the error bound, as predicted in [11, Lemma A.2].

4 Conclusions

MLMC sampling methods are becoming increasingly popular due to their robustness and simplicity. In this work, in Theorems 2.1 and 2.2 and Corollary 2.3, we have developed optimal non-geometric and geometric hierarchies for MLMC by assuming certain asymptotic models on the weak and strong convergence and the average computational cost per sample. While it is important to optimize the geometric level separation parameter, β , and the tolerance splitting parameter, θ , to obtain significant computational savings, we have shown, in Remark 2.6, that with these optimized parameters, geometric hierarchies are nearly optimal and that,

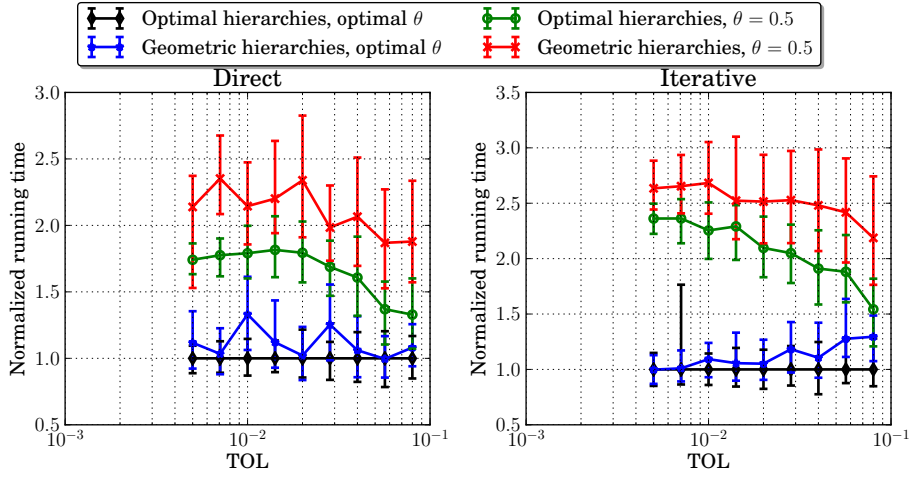


Fig. 3.3 Ex.1: Actual running time of the CMLMC algorithm when using optimal and geometric hierarchies with different tolerance splitting, normalized by the average running time of the algorithm when using optimal hierarchies. Compare this figure to Figure 2.1, where the latter is based on the theoretical results. Observe that most of the gain in computational complexity is due to the choice (3.2) of θ and that using optimal hierarchies does not significantly improve the running time over geometric hierarchies.

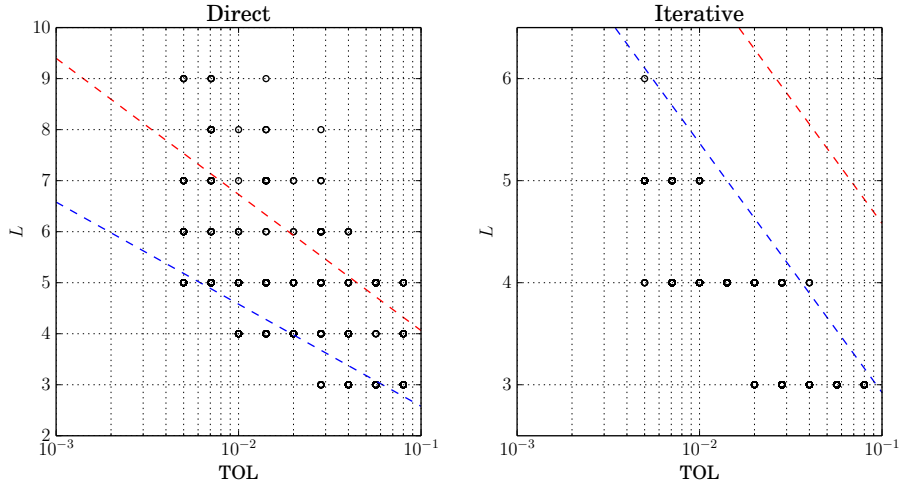


Fig. 3.4 Ex.1: The used number of levels, L , for different tolerances in the last iteration of the CMLMC algorithm when using optimal hierarchies and ceiling h_ℓ^{-1} and M_ℓ . Here, different circles correspond to independent runs of CMLMC. Compare this figure to the Figure 2.2, where the latter is based on the theoretical results. The bounds are taken from (2.20). The L values used by the CMLMC algorithm fall outside the predicted bounds because the bounds are valid for the real-valued optimal hierarchies only. On the other hand, CMLMC is constrained to the discrete sets of feasible hierarchies and further limits the increments of L across iterations.

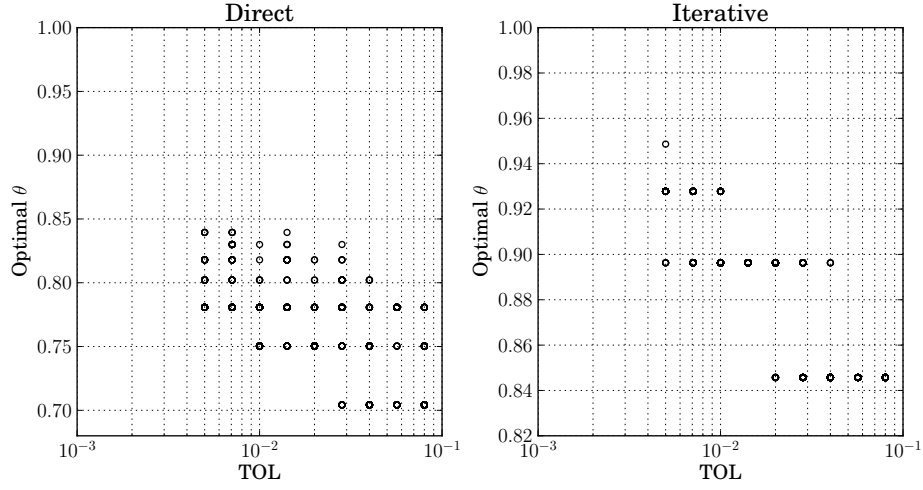


Fig. 3.5 Ex.1: The error splitting parameter, θ , as defined by (2.19c). Recall that θ depends on L and notice the correspondence to L values in Figure 3.4

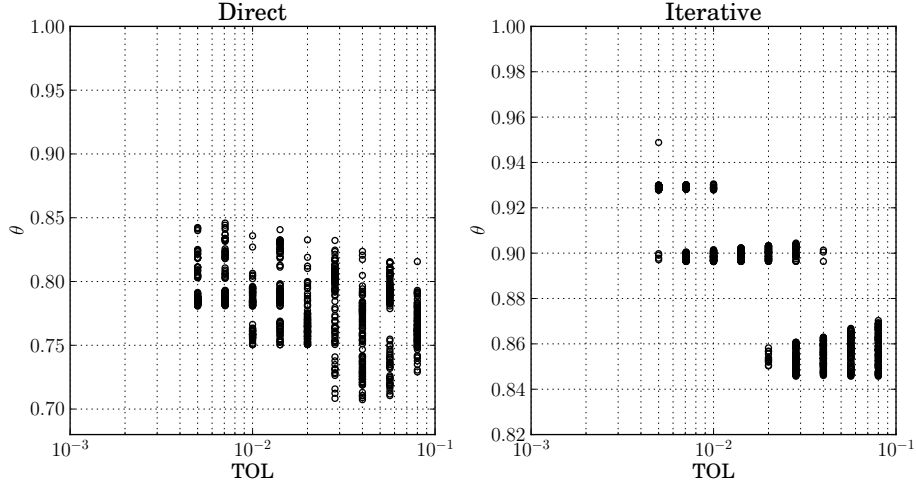


Fig. 3.6 Ex.1: The computational splitting parameter, θ , as defined in (3.2) for the CMLMC algorithm for optimal hierarchies. Though these θ values correspond to the value in Figure 3.5, the differences are due to enforcing constraints on h_L .

asymptotically, their computational complexity is the same as the non-geometric optimal hierarchies. Moreover, we have analyzed the asymptotic behavior of the optimal tolerance splitting parameter, θ , between the bias and the statistical error contribution. We have also discussed how enforcing constraints on parameters of MLMC hierarchies affects the optimality of these hierarchies. These constraints include an upper and lower bound on the mesh size or enforcing that the number of samples and the number of discretization elements are integers. Our numerical

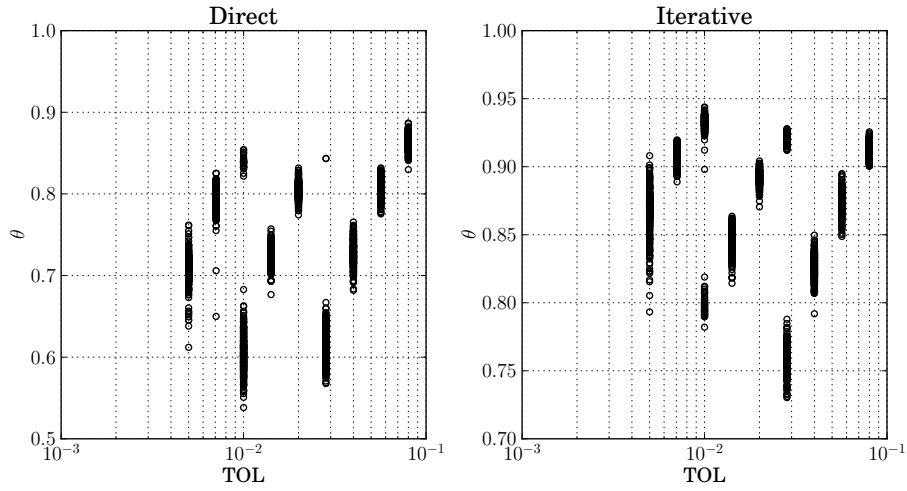


Fig. 3.7 Ex.1: The computational splitting parameter, θ , in the CMLMC algorithm for optimal geometric hierarchies.

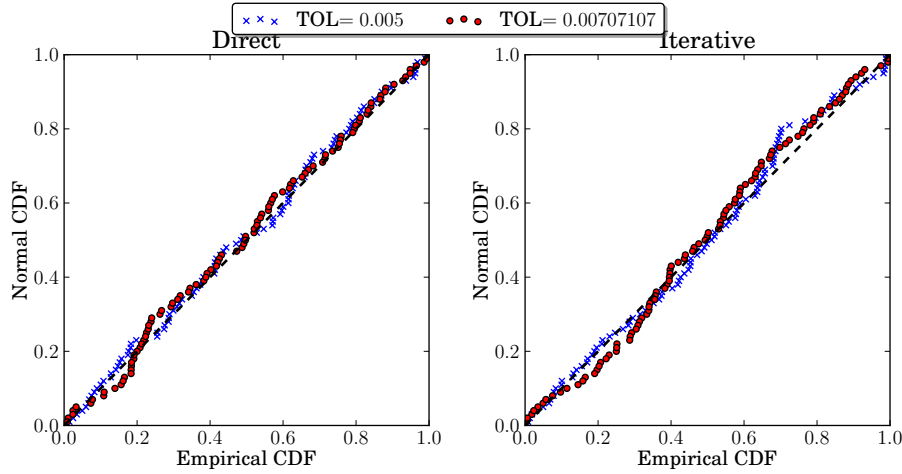


Fig. 3.8 Ex.1: A QQ-plot indicating that, even when using non-geometric hierarchies, the distribution of the normalized statistical error is well approximated by the standard normal density. The work [11, Lemma A.2] proved such results for geometric hierarchies.

results show remarkable agreement between our theory of optimal hierarchies and their asymptotic behavior and the performance of the CMLMC algorithm.

In future work, it is possible to improve the efficiency of the MLMC method by including certain non-asymptotic terms in the models for the weak and strong convergence or the computational complexity. Moreover, since the asymptotic dependence of the computational complexity on the different problem constants is clearly shown in Corollaries 2.1 and 2.2, one can devise methods to combine with

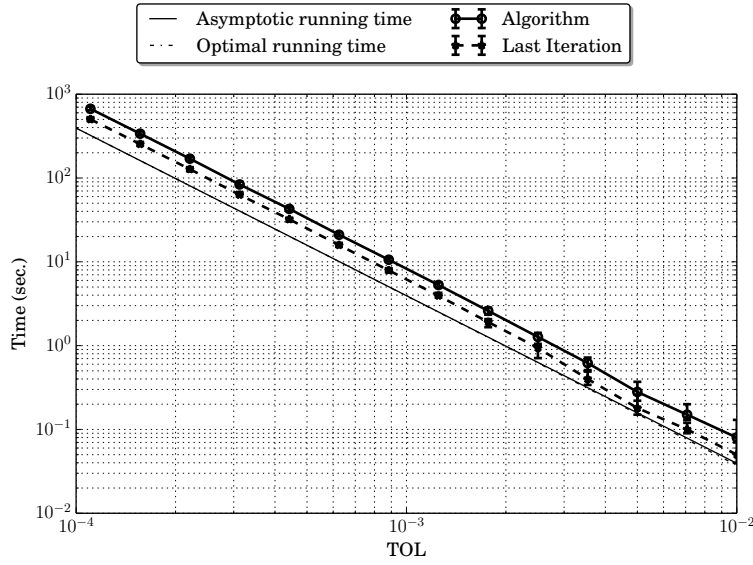


Fig. 3.9 Ex.2: The running time of the CMLMC algorithm. The reference dashed line is $\mathcal{O}(\text{TOL}^{-2})$ as predicted in [27, Theorem 2.3].

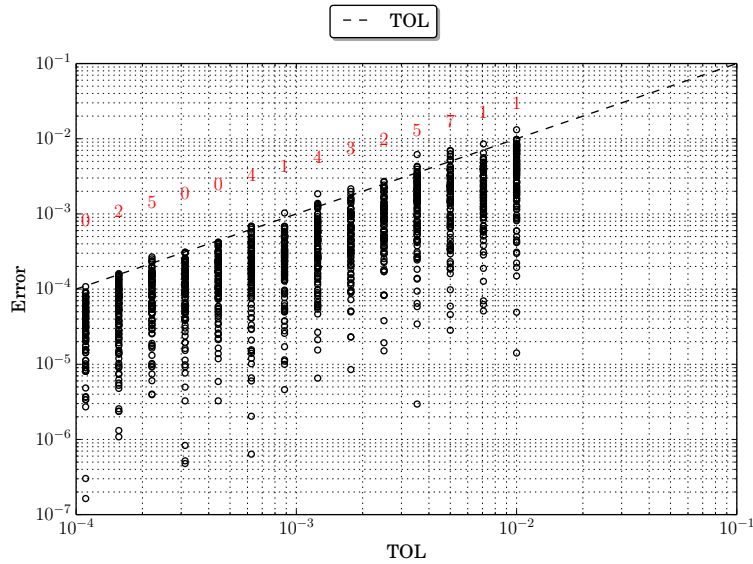


Fig. 3.10 Ex.2: The true errors calculated using geometric hierarchies with $\beta = 4$. The numbers on top of the TOL line are the percentage of algorithm runs that produced a larger error than the required tolerance. Remember that $C_\alpha = 2$ gives a confidence of a 95% in the error bound, as predicted in [11, Lemma A.2].

MLMC to reduce the total computational complexity by affecting these constants, for example by reducing the variance, V_0 , for the case where $\chi > 1$.

Acknowledgments

R. Tempone is a member of the Research Center on Uncertainty Quantification (SRI-UQ), division of Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) at King Abdullah University of Science and Technology (KAUST). The authors would like to recognize the support of the following KAUST and University of Texas at Austin AEA projects: Round 2, “Predictability and Uncertainty Quantification for Models of Porous Media”, and Round 3, “Uncertainty quantification for predictive modeling of the dissolution of porous and fractured media”. F. Nobile has been partially supported by the Swiss National Science Foundation under the Project No. 140574 “Efficient numerical methods for flow and transport phenomena in heterogeneous random porous media” and by the Center for ADvanced MOdeling Science (CADMOS). E. von Schwerin has been partially supported by the aforementioned SRI-UQ and CADMOS. We would also like to acknowledge the following open source software packages that made this work possible: PETSc [3], PetIGA [10].

A Derivations and proofs

A.1 Optimal hierarchies given h_0 , θ , and L

Here we solve Problem 2.1 of Section 2.2 for the optimal hierarchy for any fixed value of L . We initially treat the parameter θ as given, postponing its optimization until later, and proceed in two steps to find the optimal $\{M_\ell\}_{\ell=0}^L$ and $\{h_\ell\}_{\ell=0}^L$. Assuming general work estimates $\{W_\ell\}_{\ell=0}^L$ in (2.4) and general variance estimates of $\{V_\ell\}_{\ell=0}^L$, we assume equality in (2.9b) and introduce the Lagrange multiplier λ to obtain the Lagrangian

$$\mathcal{L}(\{M_\ell\}_{\ell=0}^L, \lambda) = \sum_{\ell=0}^L M_\ell W_\ell + \lambda \left\{ \sum_{\ell=0}^L \frac{V_\ell}{M_\ell} - \left(\theta \frac{\text{TOL}}{C_\alpha} \right)^2 \right\}.$$

The requirement that the variation of the Lagrangian with respect to M_ℓ is zero, gives $M_\ell = \sqrt{\lambda \frac{V_\ell}{W_\ell}}$. Solving for λ in the variance constraint (2.9b) and substituting back leads to (2.13). Substituting this optimal M_ℓ in the total work (2.4) yields

$$W(\mathbf{H}) = \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2 \left(\sum_{\ell=0}^L \sqrt{W_\ell V_\ell} \right)^2. \quad (\text{A.1})$$

We proceed to find the optimal $\{h_\ell\}_{\ell=0}^L$ under the particular models (2.12). The total work (A.1) is minimized when

$$\sum_{\ell=0}^L \sqrt{W_\ell V_\ell} = \sqrt{\frac{V_0}{h_0^{d_\gamma}}} + \sqrt{Q_S} \sum_{\ell=1}^L \sqrt{\frac{h_{\ell-1}^{q_2}}{h_\ell^{d_\gamma}}}, \quad (\text{A.2})$$

is minimized. Here the finest mesh, h_L , is given by the bias constraint (2.12b) as

$$h_L = \left(\frac{(1-\theta) \text{TOL}}{Q_W} \right)^{\frac{1}{q_1}}, \quad (\text{A.3})$$

independently of the multilevel construction. Now, treat the coarsest mesh, h_0 , as given and find the optimal h_1, \dots, h_{L-1} that minimize

$$\frac{1}{\sqrt{Q_S}} \sum_{\ell=1}^L \sqrt{W_\ell V_\ell} = \sum_{\ell=1}^L \sqrt{\frac{h_{\ell-1}^{q_2}}{h_\ell^{d_\gamma}}}. \quad (\text{A.4})$$

The requirement that the derivative of this sum with respect to h_ℓ equals zero, for $\ell = 1, \dots, L-1$, leads to the optimality condition

$$q_2 h_\ell^{\left(\frac{q_2+d_\gamma}{2}\right)} = d_\gamma h_{\ell-1}^{\left(\frac{q_2}{2}\right)} h_{\ell+1}^{\left(\frac{d_\gamma}{2}\right)},$$

which after taking the logarithm and using χ defined in (2.14), leads to

$$-\log(h_{\ell+1}) + (1+\chi) \log(h_\ell) - \chi \log(h_{\ell-1}) = -\frac{2}{d_\gamma} \log(\chi). \quad (\text{A.5})$$

This is a second order linear difference whose solution depends on χ .

A.1.1 For $\chi = 1$

This section provides proofs of Theorem 2.1, Lemma 2.1, and Corollary 2.1. The solution of the difference equation (A.5) for the case $\chi = 1$ is the geometric sequence

$$h_\ell = h_0 \beta^{-\ell}, \quad \text{with } \beta = \left(\frac{h_0}{h_L} \right)^{1/L}. \quad (\text{A.6})$$

In other words, all h_ℓ are defined in terms of h_0 and h_L , where the latter is determined by θ through (A.3) and we solve for the former by setting the derivative of (A.2) with respect to h_0 equal to zero. This optimality condition becomes (for $q_2 = d\gamma$)

$$h_1 = \left(\frac{Q_S}{V_0} \right)^{\frac{1}{q_2}} h_0^2.$$

Combining this expression with (A.6) for $\ell = 1$ and solving for h_0 yields

$$h_0 = h_L^{\frac{1}{L+1}} \left(\frac{V_0}{Q_S} \right)^{\frac{L}{q_2(L+1)}}. \quad (\text{A.7})$$

Substituting this expression and (A.3) in the expression for β in (A.6) we obtain (2.15c). Moreover, substituting (2.15c) and (A.6) and (2.10) and (2.5) in (2.13) yields (2.15b). Next, we substitute (2.15b) and (2.15a) in (2.6) to obtain the optimal work for $q_2 = d\gamma$

$$W = \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2 \left(\sqrt{V_0} h_0^{\frac{-q_2}{2}} + \sqrt{Q_S} \beta^{\frac{q_2}{2}} L \right)^2. \quad (\text{A.8})$$

Using (A.6) and (A.7), we obtain

$$W = \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2 h_L^{\frac{-q_2}{L+1}} V_0^{\frac{1}{L+1}} Q_S^{\frac{L}{L+1}} (1+L)^2. \quad (\text{A.9})$$

Substituting for h_L from (A.3)

$$W = \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2 \left(\frac{Q_W}{(1-\theta) \text{TOL}} \right)^{\frac{1}{\eta(L+1)}} V_0^{\frac{1}{L+1}} Q_S^{\frac{L}{L+1}} (1+L)^2. \quad (\text{A.10})$$

Optimizing for θ yields (2.15d). Substituting back gives the work as a function of L

$$W(L) = C_\alpha^2 \text{TOL}^{-2(1+\epsilon(L))} Q_W^{2\epsilon(L)} V_0^{2\eta\epsilon(L)} Q_S^{-2\eta\epsilon(L)} Q_S \left(\frac{1}{2\eta} \right)^2 \left(1 + \frac{1}{\epsilon(L)} \right)^{2(1+\epsilon(L))}, \quad (\text{A.11})$$

where $\epsilon(L) = \frac{1}{2\eta(L+1)}$. Treating L as a continuous variable and differentiating with respect to L yields

$$W'(L) = 2W(L)\epsilon'(L)(C - y + \log(1+y)), \quad (\text{A.12})$$

where $y = 2\eta(L+1) > 0$ and $C = \log \left(\text{TOL}^{-1} Q_W V_0^\eta Q_S^{-\eta} \right)$. Setting (A.12) to zero gives the equation

$$y - \log(1+y) = C. \quad (\text{A.13})$$

It follows that $\lim_{C \rightarrow \infty} \frac{y}{C} = 1$ which leads to (2.17) for the value of L and (2.18) for the work. Since $y > 0$, equation (A.13) implies

$$1 < yC^{-1} \quad (\text{A.14})$$

for any $C > 0$. Furthermore, for any $y > 0$, it holds

$$\left(\frac{\exp(1) - 1}{\exp(1)} \right) (1+y) - C \leq 1 + y - \log(1+y) - C,$$

which together with (A.13) gives

$$yC^{-1} \leq \frac{\exp(1)}{\exp(1) - 1} + \frac{1}{(\exp(1) - 1)C}, \quad (\text{A.15})$$

for any $C > 0$. Inequalities (A.14) and (A.15) are (2.16).

A.1.2 For $\chi \neq 1$

This section provides proofs of Theorem 2.2, Lemma 2.2, and Corollary 2.2. The solution of the difference equation (A.5) for the case $\chi \neq 1$ is

$$h_\ell = h_0 \left(\frac{\chi^\ell - \chi^L}{1 - \chi^L} \right) h_L \left(\frac{1 - \chi^\ell}{1 - \chi^L} \right) \chi^{-\frac{2}{d\gamma} \left(\frac{L(1 - \chi^\ell) - \ell(1 - \chi^L)}{(1 - \chi)(1 - \chi^L)} \right)}. \quad (\text{A.16})$$

We now distinguish between two different cases for h_0 : either we are free to choose the optimal $h_0 \in \mathbb{R}_+$, or we have an upper bound on the coarsest mesh h_0 . The first, idealized, situation will allow us to obtain explicit expressions for the optimal splitting parameter θ and the asymptotic work, and we start by considering this case. We return to the other case at the end of this section.

Unconstrained optimization of h_0 We take h_1, \dots, h_L given by (A.16) and (A.3) and set the derivative of (A.2) with respect to h_0 equal to zero. This optimality condition becomes (after some straightforward simplifications)

$$-\frac{d\gamma}{2} \frac{\sqrt{V_0}}{h_0^{1+d\gamma/2}} + \frac{q_2}{2} \sqrt{Q_S} \frac{h_0^{q_2/2-1}}{h_1^{d\gamma/2}} = 0,$$

which, since all parameters are positive, is equivalent to

$$h_1 = \left(\frac{\chi^2 Q_S}{V_0} \right)^{\frac{1}{d\gamma}} h_0^{1+\chi}.$$

Combining this expression for h_1 with the one in (A.16) and solving for h_0 gives

$$h_0 = h_L^{\frac{1-\chi}{1-\chi^{L+1}}} \left(\frac{V_0}{Q_S} \right)^{\frac{1}{d\gamma} \frac{1-\chi^L}{1-\chi^{L+1}}} \chi^{-\frac{2}{d\gamma} \frac{1-\chi}{1-\chi} \left(L \frac{1-\chi}{1-\chi^{L+1}} - \chi \frac{1-\chi^L}{1-\chi^{L+1}} \right)},$$

which after substituting back into (A.16) and using (A.3) yields (2.19a). Finally substituting these optimal mesh sizes into (2.13) yields (2.19b).

Optimal splitting parameter θ Now the sequences $\{h_\ell\}_{\ell=0}^L$ and $\{M_\ell\}_{\ell=0}^L$ are determined in terms of the still not optimized L and θ as well as measurable model parameters. The work per level in (2.6) becomes

$$\begin{aligned} \frac{M_\ell}{h_\ell^{d\gamma}} &= \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2 \left(\frac{Q_W}{(1-\theta) \text{TOL}} \right)^{\frac{1}{\eta} \frac{1-\chi}{1-\chi^{L+1}}} V_0 \left(\frac{Q_S}{V_0} \right)^{\left\{ \frac{1-\chi^L}{1-\chi^{L+1}} \right\}} \\ &\quad \cdot \frac{1-\chi^{L+1}}{1-\chi} \chi^{\left\{ -\frac{2\chi}{1-\chi} \frac{1-\chi^L}{1-\chi^{L+1}} + L \frac{1+\chi^{L+1}}{1-\chi^{L+1}} \right\}} \chi^{-\ell}. \end{aligned}$$

Since the only ℓ -dependent factor in the right hand side is the last one, $\chi^{-\ell}$, and using $\sum_{\ell=0}^L \chi^{-\ell} = \chi^{-L} (1 - \chi^{L+1}) / (1 - \chi)$, the total work in (2.6) becomes

$$W(L, \theta, \text{TOL}) = w_1(L, \text{TOL}) w_2(L) f(L, \theta) \left(\frac{1 - \chi^{L+1}}{1 - \chi} \right)^2, \quad (\text{A.17})$$

with

$$w_1(L, \text{TOL}) = \text{TOL}^{-\left(2 + \frac{1}{\eta} \frac{1-\chi}{1-\chi^{L+1}}\right)}, \quad (\text{A.18a})$$

$$\begin{aligned} w_2(L) &= C_\alpha^2 V_0 \left(\frac{Q_S}{V_0} \right)^{\left\{ \frac{1-\chi^L}{1-\chi^{L+1}} \right\}} Q_W^{\left\{ \frac{1}{\eta} \frac{1-\chi}{1-\chi^{L+1}} \right\}} \\ &\quad \cdot \chi^{\left\{ -\frac{2\chi}{1-\chi} \frac{1-\chi^L}{1-\chi^{L+1}} + 2L \frac{\chi^{L+1}}{1-\chi^{L+1}} \right\}}, \end{aligned} \quad (\text{A.18b})$$

$$f(L, \theta) = \frac{1}{\theta^2 (1 - \theta)^{\frac{1}{\eta} \frac{1-\chi}{1-\chi^{L+1}}}}. \quad (\text{A.18c})$$

Thus given the value of L the dependence on the splitting parameter θ is straightforward, and the minimal work for a given L is obtained with the minimizer of (A.18c), namely (2.19c). With this optimal splitting parameter θ in (A.17) the total work as a function of the yet to be determined parameter L and the tolerance is

$$W(L, \text{TOL}) = w_1(L, \text{TOL}) w_2(L) w_3(L), \quad (\text{A.19})$$

with

$$w_3(L) = \left(\frac{1}{2\eta}\right)^2 \left(1 + 2\eta \frac{1 - \chi^{L+1}}{1 - \chi}\right)^{2\left(1 + \frac{1}{2\eta} \frac{1 - \chi}{1 - \chi^{L+1}}\right)}. \quad (\text{A.20})$$

Optimal number of levels The optimal integer L seems impossible to find analytically. In practical computations we instead perform an extensive search over a small range of integer values. In the analysis below we treat L as a real parameter to obtain the bounds (2.20) that delimit the range of integer values that must be tested, and allow a complexity analysis as $\text{TOL} \rightarrow 0$ without an exactly determined L .

Treating L as a real parameter, we differentiate the work (A.19) with respect to L to obtain

$$\frac{\partial W}{\partial L} = \frac{\partial w_1}{\partial L} w_2 w_3 + w_1 \frac{\partial w_2}{\partial L} w_3 + w_1 w_2 \frac{\partial w_3}{\partial L},$$

where, introducing the shorthand

$$\xi(L) = 2\eta \frac{1 - \chi^{L+1}}{1 - \chi} \quad \text{for } L \in [0, \infty), \quad (\text{A.21})$$

and using the constants c_1 and c_2 in (2.21) we write

$$\frac{\partial w_1}{\partial L} = w_1(L, \text{TOL}) \frac{\log(\chi) \chi^{L+1}}{1 - \chi^{L+1}} \frac{2}{\xi(L)} \log(\text{TOL}^{-1}), \quad (\text{A.22a})$$

$$\frac{\partial w_2}{\partial L} = w_2(L) \frac{\log(\chi) \chi^{L+1}}{1 - \chi^{L+1}} \frac{2}{\xi(L)} (c_1 + -c_2(L+1) + \xi(L)), \quad (\text{A.22b})$$

$$\frac{\partial w_3}{\partial L} = w_3(L) \frac{\log(\chi) \chi^{L+1}}{1 - \chi^{L+1}} \frac{2}{\xi(L)} (\log(1 + \xi(L)) - \xi(L)), \quad (\text{A.22c})$$

so that

$$\frac{\partial W}{\partial L}(L, \text{TOL}) = u(L, \text{TOL}) v(L, \text{TOL}), \quad (\text{A.23})$$

with

$$\begin{aligned} u(L, \text{TOL}) &= W(L, \text{TOL}) \frac{\log(\chi) \chi^{L+1}}{1 - \chi^{L+1}} \frac{2}{\xi(L)}, \\ v(L, \text{TOL}) &= \log(\text{TOL}^{-1}) + c_1 + -c_2(L+1) + \log(1 + \xi(L)). \end{aligned}$$

Clearly $u(L, \text{TOL}) < 0$ for all $\chi \in \mathbb{R}_+ \setminus \{1\}$ so the sign of $\partial W / \partial L$ is the opposite of the sign of $v(L, \text{TOL})$. For a fixed $\chi \in \mathbb{R}_+ \setminus \{1\}$ we have

$$v(L, \text{TOL}) > 0 \Leftrightarrow L + 1 < \frac{1}{c_2} (\log(\text{TOL}^{-1}) + c_1 + \log(1 + \xi(L))),$$

and, since $\xi(L) \geq \xi(0) = 2\eta$,

$$L + 1 < \frac{1}{c_2} (\log(\text{TOL}^{-1}) + c_1 + \log(1 + 2\eta)) \Rightarrow v(L, \text{TOL}) > 0 \Leftrightarrow \frac{\partial W}{\partial L} < 0. \quad (\text{A.24})$$

For the opposite inequality,

$$v(L, \text{TOL}) < 0 \Leftrightarrow L + 1 > \frac{1}{c_2} \left(\log(\text{TOL}^{-1}) + c_1 + \log(1 + \xi(L)) \right),$$

we distinguish between the cases $0 < \chi < 1$ and $1 < \chi$. When $0 < \chi < 1$ we have the upper bound $\xi(L) < \frac{2\eta}{1-\chi}$ and consequently

$$L + 1 > \frac{1}{c_2} \left(\log(\text{TOL}^{-1}) + c_1 + \log\left(1 + \frac{2\eta}{1-\chi}\right) \right) \Rightarrow \frac{\partial W}{\partial L} > 0, \quad \chi \in (0, 1). \quad (\text{A.25})$$

In contrast $\xi(L)$ is unbounded when $1 < \chi$ but, since the definitions of χ and η and the relation between strong and weak convergence orders implies that $2\eta \geq \chi$, we have

$$\log(1 + \xi(L)) < \log\left(\frac{2\eta}{\chi - 1}\right) + (L + 1) \log \chi,$$

and

$$c_2 \geq \frac{\chi}{\chi - 1} \log \chi,$$

which gives the bound

$$\frac{1}{c_2} \log(1 + \xi(L)) < \frac{\chi - 1}{\chi} (L + 1) + \frac{1}{c_2} \log\left(\frac{2\eta}{\chi - 1}\right).$$

Hence

$$L + 1 - \frac{1}{c_2} \log(1 + \xi(L)) > \frac{L + 1}{\chi} - \frac{1}{c_2} \log\left(\frac{2\eta}{\chi - 1}\right),$$

and it follows that

$$L + 1 > \frac{\chi}{c_2} \left(\log(\text{TOL}^{-1}) + c_1 + \log\left(\frac{2\eta}{\chi - 1}\right) \right) \Rightarrow \frac{\partial W}{\partial L} > 0, \quad \chi \in (1, \infty). \quad (\text{A.26})$$

Combining (A.24) with (A.25) and (A.26), we obtain the bounds (2.20).

Optimal hierarchies with an upper bound on h_0 Practical computations will impose an upper limit on the mesh sizes, $h_0 \leq h_{\max}$. If the mesh sizes (2.19a) violate such a bound, we must modify our analysis slightly. We now consider h_0 given as one of the coarsest mesh sizes that can be realized in the given discretization, and analyze the case $L \geq 1$. Using the optimal mesh sizes (A.16) yields

$$\sqrt{\frac{h_{\ell-1}^{q_2}}{h_{\ell}^{d_\gamma}}} = h_0^{\frac{d_\gamma}{2} \chi^L \frac{1-\chi}{1-\chi^L}} h_L^{-\frac{d_\gamma}{2} \frac{1-\chi}{1-\chi^L}} \chi^{\left(\frac{L}{1-\chi^L} - \frac{\chi}{1-\chi} - \ell\right)},$$

where the only ℓ -dependent factor in the right hand side is the last one, $\chi^{-\ell}$, so that the sum in (A.4) is

$$\sum_{\ell=1}^L \sqrt{\frac{h_{\ell-1}^{q_2}}{h_{\ell}^{d_\gamma}}} = \left(\frac{h_0^{(\chi^L)}}{h_L} \right)^{\frac{d_\gamma}{2} \frac{1-\chi}{1-\chi^L}} \chi^{\left(\frac{L\chi^L}{1-\chi^L} - \frac{\chi}{1-\chi}\right)} \frac{1-\chi^L}{1-\chi}.$$

In this sum only h_L depends on θ through (A.3). Keeping L fixed we wish to minimize the total work, which by (A.1)–(A.2) is

$$W(\mathbf{H}) = \left(\frac{C_\alpha}{\theta \text{TOL}} \right)^2 \left(\sqrt{\frac{V_0}{h_0^{d_\gamma}}} + \sqrt{Q_S} \left(\frac{h_0^{(\chi^L)}}{h_L} \right)^{\frac{d_\gamma}{2} \frac{1-\chi}{1-\chi^L}} \chi^{\left(\frac{L\chi^L}{1-\chi^L} - \frac{\chi}{1-\chi}\right)} \frac{1-\chi^L}{1-\chi} \right)^2,$$

with respect to θ . Letting

$$\Delta = \frac{1}{2\eta} \frac{1-\chi}{1-\chi^L},$$

and

$$C = \sqrt{\frac{Q_S}{V_0}} h_0^{\frac{d\gamma}{2}} \frac{1-\chi^{L+1}}{1-\chi^L} \chi^{\left(\frac{L\chi^L}{1-\chi^L} - \frac{\chi}{1-\chi}\right)} \frac{1-\chi^L}{1-\chi} \left(\frac{Q_W}{\text{TOL}}\right)^\Delta,$$

we obtain

$$W(\mathbf{H}) \propto \tilde{f}(\theta, L, h_0) = \frac{1}{\theta^2} \left(1 + \frac{C}{(1-\theta)^\Delta}\right)^2,$$

with the optimality condition

$$\frac{\partial \tilde{f}}{\partial \theta} = \frac{2}{\theta^2} \left(1 + \frac{C}{(1-\theta)^\Delta}\right) \left(\frac{C\Delta}{(1-\theta)^{\Delta+1}} - \frac{1}{\theta} \left(1 + \frac{C}{(1-\theta)^\Delta}\right)\right) = 0,$$

where

$$\frac{2}{\theta^2} \left(1 + \frac{C}{(1-\theta)^\Delta}\right) > 0.$$

In this case when h_0 is constrained we no longer have an explicit expression for the optimal θ . However, using

$$\frac{C\Delta}{(1-\theta)^{\Delta+1}} - \frac{1}{\theta} \left(1 + \frac{C}{(1-\theta)^\Delta}\right) < \frac{C}{(1-\theta)^\Delta} \left(\frac{\Delta}{1-\theta} - \frac{1}{\theta}\right),$$

and that

$$\frac{\Delta}{1-\theta} - \frac{1}{\theta} = 0 \Leftrightarrow \theta = \frac{1}{1+\Delta},$$

we conclude that the optimal θ satisfies

$$\frac{1}{1+\Delta} \leq \theta. \quad (\text{A.27})$$

Similarly, from the inequality

$$\frac{C\Delta}{(1-\theta)^{\Delta+1}} - \frac{1}{\theta} \left(1 + \frac{C}{(1-\theta)^\Delta}\right) > \frac{1}{(1-\theta)^\Delta} \left(\frac{C\Delta}{1-\theta} - \frac{1+C}{\theta}\right),$$

and the relation

$$\frac{C\Delta}{1-\theta} - \frac{1+C}{\theta} = 0 \Leftrightarrow \theta = \frac{1+C}{1+C+\Delta},$$

we obtain an upper bound for θ , namely

$$\theta \leq \frac{1+C}{1+C+C\Delta}. \quad (\text{A.28})$$

Finally, combining (A.27) and (A.28) we have the following bounds for the optimal θ :

$$\left(1 + \frac{1}{2\eta} \frac{1-\chi}{1-\chi^L}\right)^{-1} \leq \theta \leq \left(1 + \frac{1}{2\eta} \frac{1-\chi}{1-\chi^L} \frac{C}{1+C}\right)^{-1}, \quad (\text{A.29})$$

where the upper bound has a non-trivial dependence on TOL and L through C .

A.2 Heuristic optimization of geometric hierarchies

This section motivates the results in Section 2.3 and Corollary 2.3 where we optimized geometric hierarchies defined by $h_\ell = h_0 \beta^{-\ell}$ for given h_0 and $\beta > 1$. In this case, the work and variance models are in (2.28) and L must satisfy the bias constraint

$$L \geq \frac{\log(h_0) - \frac{1}{q_1} \log\left(\frac{(1-\theta)\text{TOL}}{Q_W}\right)}{\log(\beta)}. \quad (\text{A.30})$$

We distinguish between two cases:

- $\chi = 1$: Or equivalently $q_2 = d\gamma$. In this case, the total work defined in (A.1) simplifies to

$$W = \left(\frac{C_\alpha}{\theta \text{TOL}}\right)^2 \left(\sqrt{V_0} h_0^{-\frac{q_2}{2}} + L \sqrt{Q_S} \beta^{\frac{q_2}{2}}\right)^2, \quad (\text{A.31})$$

We make the simplification of treating L as a real parameter and substitute the lower bound of (A.30) in (A.31) and optimize with respect to β to get $\beta = \exp\left(\frac{2}{q_2}\right)$. Substituting this choice and (2.30), the total work satisfies

$$\frac{W}{\text{TOL}^{-2} (\log \text{TOL})^2} \rightarrow \theta^{-2} C_\alpha^2 Q_S \exp(2) \left(\frac{1}{2\eta}\right)^2, \quad \text{as } \text{TOL} \rightarrow 0.$$

Optimizing for θ suggests that $\theta \rightarrow 1$ as $\text{TOL} \rightarrow 0$ and (2.18) follows.

- $\chi \neq 1$: In this case, the total work defined in (A.1) simplifies to

$$W = \left(\frac{C_\alpha}{\theta \text{TOL}}\right)^2 h_0^{d\gamma(\chi-1)} \left(\sqrt{V_0} h_0^{-\frac{q_2}{2}} + \sqrt{Q_S} \frac{\left(1 - \beta^{\frac{L(d\gamma-q_2)}{2}}\right)}{\beta^{-\frac{d\gamma}{2}} - \beta^{-\frac{q_2}{2}}}\right)^2, \quad (\text{A.32})$$

for a given L, h_0 and θ . Again, we make the simplification of treating L as a real parameter and substitute the lower bound (A.30) to obtain

$$\beta^{\frac{L(d\gamma-q_2)}{2}} = \left(\frac{(1-\theta)\text{TOL}}{Q_W}\right)^{\frac{\chi-1}{2\eta}} h_0^{\frac{d\gamma(1-\chi)}{2}},$$

for any β . Substituting back in (A.32) and optimizing with respect to β to minimize the work gives (2.29). Substituting this optimal β in (A.32) yields

$$W = \left(\frac{C_\alpha}{\theta \text{TOL}}\right)^2 h_0^{d\gamma(\chi-1)} \left(\sqrt{V_0} h_0^{-\frac{q_2}{2}} + \sqrt{Q_S} \frac{\chi^{\frac{\chi-1}{2\eta}}}{\chi-1} (1 - \chi^{-L})\right)^2. \quad (\text{A.33})$$

Asymptotically, using (2.30) as $\text{TOL} \rightarrow 0$ yields (2.23) with the following constants

$$C_1 = (1-\theta)^{\frac{\chi-1}{\eta}} \theta^{-2} C_\alpha^2 Q_W^{\frac{1-\chi}{\eta}} Q_S \left(\frac{\chi^{\frac{\chi-1}{2\eta}}}{\chi-1}\right)^2, \quad (\text{A.34a})$$

$$C_2 = \theta^{-2} C_\alpha^2 h_0^{d\gamma(\chi-1)} \left(\sqrt{V_0} h_0^{-\frac{q_2}{2}} + \sqrt{Q_S} \frac{\chi^{\frac{\chi-1}{2\eta}}}{\chi-1}\right)^2. \quad (\text{A.34b})$$

Optimizing these constants with respect to θ yields (2.25) and substituting this and (2.32) back yields (2.24a) and (2.31) for C_1 and C_2 , respectively. This, as Remark 2.6 mentions, shows that the asymptotic computational complexities of optimal non-geometric and geometric hierarchies are the same.

References

1. Amestoy, P.R., Duff, I.S., L'Excellent, J.Y., Koster, J.: A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.* **23**, 15–41 (2001). DOI 10.1137/S0895479899358194. URL <http://portal.acm.org/citation.cfm?id=587708.587825>
2. Babuška, I., Nobile, F., Tempone, R.: A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM review* **52**(2), 317–355 (2010)
3. Balay, S., Brown, J., Buschelman, K., Gropp, W.D., Kaushik, D., Knepley, M.G., McInnes, L.C., Smith, B.F., Zhang, H.: PETSc Web page (2013). URL <http://www.mcs.anl.gov/petsc>
4. Balay, S., Gropp, W.D., McInnes, L.C., Smith, B.F.: Efficient management of parallelism in object oriented numerical software libraries. In: Arge, E., Bruaset, A.M., Langtangen, H.P. (eds.) *Modern Software Tools in Scientific Computing*, pp. 163–202. Birkhäuser Press (1997)
5. Barth, A., Lang, A., Schwab, C.: Multilevel Monte Carlo method for parabolic stochastic partial differential equations. *BIT Numerical Mathematics* **53**(1), 3–27 (2013)
6. Barth, A., Schwab, C., Zollinger, N.: Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik* **119**(1), 123–161 (2011)
7. Bayer, C., Hoel, H., von Schwerin, E., Tempone, R.: On nonasymptotic optimal stopping criteria in monte carlo simulations. *SIAM Journal on Scientific Computing* **36**(2), A869–A885 (2014). DOI 10.1137/130911433. URL <http://dx.doi.org/10.1137/130911433>
8. Charrier, J., Scheichl, R., Teckentrup, A.: Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods. *SIAM Journal on Numerical Analysis* **51**(1), 322–352 (2013)
9. Cliffe, K., Giles, M., Scheichl, R., Teckentrup, A.: Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science* **14**(1), 3–15 (2011)
10. Collier, N., Dalcin, L., Calo, V.: PetIGA: High-performance isogeometric analysis. arxiv (1305.4452) (2013). URL <http://arxiv.org/abs/1305.4452>
11. Collier, N., Haji-Ali, A.L., Nobile, F., von Schwerin, E., Tempone, R.: A continuation multilevel monte carlo algorithm. *BIT Numerical Mathematics* pp. 1–34 (2014). DOI 10.1007/s10543-014-0511-3. URL <http://dx.doi.org/10.1007/s10543-014-0511-3>
12. Giles, M.: Improved multilevel Monte Carlo convergence using the Milstein scheme. In: *Monte Carlo and quasi-Monte Carlo methods 2006*, pp. 343–358. Springer (2008)
13. Giles, M.: Multilevel Monte Carlo path simulation. *Operations Research* **56**(3), 607–617 (2008)
14. Giles, M., Reisinger, C.: Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance. *SIAM Journal of Financial Mathematics* **3**(1), 572–592 (2012)
15. Glasserman, P.: *Monte Carlo methods in financial engineering, Applications of Mathematics (New York)*, vol. 53. Springer-Verlag, New York (2004). *Stochastic Modelling and Applied Probability*
16. Heinrich, S.: Monte Carlo complexity of global solution of integral equations. *Journal of Complexity* **14**(2), 151–175 (1998)
17. Heinrich, S., Sindambiwe, E.: Monte Carlo complexity of parametric integration. *Journal of Complexity* **15**(3), 317–341 (1999)
18. Hoel, H., von Schwerin, E., Szepessy, A., Tempone, R.: Adaptive multilevel Monte Carlo simulation. In: Engquist, B., Runborg, O., Tsai, Y.H. (eds.) *Numerical Analysis of Multiscale Computations*, no. 82 in *Lecture Notes in Computational Science and Engineering*, pp. 217–234. Springer (2012)
19. Hoel, H., von Schwerin, E., Szepessy, A., Tempone, R.: Implementation and analysis of an adaptive multilevel Monte Carlo algorithm. *Monte Carlo Methods and Applications* **20**(1), 1–41 (2014)
20. Jouini, E., Cvitanic, J., Musiela, M. (eds.): *Option pricing, interest rates and risk management*. Handbooks in Mathematical Finance. Cambridge University Press, Cambridge (2001)
21. Karatzas, I., Shreve, S.E.: *Brownian motion and stochastic calculus, Graduate Texts in Mathematics*, vol. 113. Second edn. Springer-Verlag, New York (1991)
22. Kebaier, A.: Statistical Romberg extrapolation: a new variance reduction method and applications to options pricing. *Annals of Applied Probability* **14**(4), 2681–2705 (2005)

23. Milstein, G.N., Tretyakov, M.V.: Stochastic numerics for mathematical physics. Springer (2004)
24. Moon, K.S., Szepessy, A., Tempone, R., Zouraris, G.E.: Convergence rates for adaptive weak approximation of stochastic differential equations. *Stoch. Anal. Appl.* **23**(3), 511–558 (2005)
25. Moraes, A., Tempone, R., Vilanova, P.: Multilevel hybrid chernoff tau-leap. Accepted for publication in *BIT Numerical Mathematics* (2015)
26. Øksendal, B.: Stochastic differential equations. Universitext, fifth edn. Springer-Verlag, Berlin (1998)
27. Teckentrup, A., Scheichl, R., Giles, M., Ullmann, E.: Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik* **125**(3), 569–600 (2013)
28. Tesei, F., Nobile, F.: A multi level monte carlo method with control variate for elliptic pdes with log-normal coefficients. Tech. rep. (2014)
29. Xia, Y., Giles, M.: Multilevel path simulation for jump-diffusion SDEs. In: Plaskota, L., Woźniakowski, H. (eds.) *Monte Carlo and Quasi-Monte Carlo Methods 2010*, pp. 695–708. Springer (2012)