NASA/TP-2019-220312

# Multilevel Monte Carlo Estimation of Unbiased Expectation via Sample Reuse and the Low Variance Estimation of Asymptotic Rates

*Timothy Barth*
*NASA Ames Research Center, Moffett Field, CA*

# Multilevel Monte Carlo Estimation of Unbiased Expectation via Sample Reuse and the Low Variance Estimation of Asymptotic Rates

Timothy Barth

NASA Ames Research Center, Moffett Field, California USA
*Timothy.J.Barth@nasa.gov*

**Abstract**

A new variant of the multilevel Monte Carlo estimator [5, 3, 9, 12] is presented for the estimation of expectation statistics that utilizes sample reuse in specified levels, explicitly removes approximation error bias associated with numerically computed output quantities of interest that have an asymptotic limit behavior, and permits a low variance estimate of the asymptotic rate of convergence to that limit. In addition, it is shown that this new multilevel Monte Carlo variant can yield a computational cost savings.

A review of Monte Carlo and multilevel Monte Carlo estimators is presented that includes analysis of expected value, expected mean squared error, and the calculation of optimized multilevel sample size parameters. The multilevel Monte Carlo estimator produces estimates of expectation for numerically approximated output quantities of interest that are biased by approximation error. When the quantity of interest can be modeled as the asymptotic limit of numerically approximated output quantities of interest, it is theoretically possible to remove this approximation error bias in the multilevel Monte Carlo estimator. In actual implementations, however, this procedure is unreliable due to statistical variability and inaccuracy in estimating the needed asymptotic limit. Analysis and numerical experiment show that the proposed variant of the multilevel Monte Carlo method greatly reduces (in some cases eliminates) the statistical variability in this limit estimation.

# 1 Introduction

Multilevel methods, including multigrid [1, 6] and multilevel domain decomposition methods [11, 2], have proven highly successful in drastically reducing the algorithmic complexity associated with solving discretized systems arising from the finite-dimensional approximation of deterministic PDEs. This improvement in algorithmic complexity has enabled the routine simulation of complex physical systems using numerically approximated PDEs.

More recently, multilevel approaches have been developed to accelerate the Monte Carlo (MC) sampling method originally developed by scientists at the Los Alamos Laboratory in the 1940s. The primary application area considered here is the Monte Carlo estimation of moment statistics for output quantities of interest derived from the numerical solution of PDEs with input sources of uncertainty. The original Monte Carlo sampling method applied to these applications becomes extremely expensive when very small errors in the desired moment statistics are required. To address this inefficiency, multilevel Monte Carlo estimators [5, 3, 9, 12] and multifidelity Monte Carlo estimators [10] have been developed that permit the efficient calculation of moment statistics with very small errors for this class of problems. These multilevel Monte Carlo estimators utilize a sequence of approximate quantity of interest models with increasing accuracy and cost. The expectation statistic is then calculated in incremental form. For example, let $J$ denote a quantity of interest and $\{J_0, J_1, \ldots, J_L\}$ a sequence of approximate models with increasing accuracy and evaluation cost. Linearity of the expectation functional permits writing the expectation of the high accuracy model, $\mathbb{E}[J_L]$, in terms of the expectation of the low accuracy model, $\mathbb{E}[J_0]$, (that is relatively inexpensive to estimate using MC sampling) and the sum of incremental corrections over levels

$$\mathbb{E}[J_L] = \mathbb{E}[J_0] + \mathbb{E}[J_1 - J_0] + \cdots + \mathbb{E}[J_L - J_{L-1}] \tag{1}$$

that can also be estimated using MC sampling but have small variance. Analysis shows that this strategy can yield a dramatic improvement in cost efficiency when compared to a single level Monte Carlo estimator. An outstanding issue with the multilevel Monte Carlo estimators described above is that these estimators produce expectation estimates biased by model approximation error. In the above example, $\mathbb{E}[J_L]$ can be estimated but the desired statistic is $\mathbb{E}[J]$. In principle, this approximation error bias can be removed [5, 3]. Unfortunately, for the PDE related problems considered here, the additional estimates needed for the bias correction are contaminated by sampling variability and generally unreliable. A contribution of the present work is a variant of the multilevel Monte Carlo estimator utilizing sample reuse that accommodates the efficient and robust (very low variance) calculation of approximation error bias corrections.

## 1.1 PDEs with random variable sources of uncertainty

Let $(\Omega, \Sigma, P)$ denote a probability space with event outcomes in $\Omega$, a $\sigma$-algebra $\Sigma$, and $P$ the assignment of events to probabilities. Our interest lies in the finite-dimensional approximation systems of random variable PDEs with $m$ dependent variables in $d+1$ space-time dimensions subject to random variable sources of uncertainty. Let $\mathcal{L}$ denote a nonlinear PDE operator with solution $u(x,t,\omega) : \mathbb{R}^d \times \mathbb{R}_+ \times \Omega \mapsto \mathbb{R}^m$ and forcing function $f(x,t,\omega) : \mathbb{R}^d \times \mathbb{R}_+ \times \Omega \mapsto \mathbb{R}^m$ such that

$$\mathcal{L}\, u(x,t,\omega) = f(x,t,\omega) \tag{2}$$

subject to suitable initial data and boundary conditions. The finite-dimensional approximation of this PDE system, the subsequent calculation of derived output quantities of interest denoted by $J(u(x,t,\omega))$, and the estimation of moment statistics is the motivation for this work. In fluid dynamics simulations, some output quantities of interest include

- integrated forces and force moments,
- graphs of pressure, temperature, velocity on specified space-time curves,
- pressure, temperature, velocity for space-time volume subsets.

In particular, we are interested in estimators for Lebesgue measurable expectation statistics of output quantities of interest

$$\mathbb{E}[J](x,t) = \int_\Omega J(u(x,t,\omega))\, p(\omega)\, d\omega \tag{3}$$

and variance statistics

$$\begin{aligned}\mathbb{V}\mathrm{ar}[J](x,t) &= \int_\Omega (J(u(x,t,\omega)) - \mathbb{E}[J](x,t))^2\, p(\omega)\, d\omega \\ &= \mathbb{E}[J^2] - \mathbb{E}^2[J] \end{aligned} \tag{4}$$

where $p(\omega)$ is the probability density associated with $P(\omega)$, i.e., $p(\omega)\, d\omega = dP(\omega)$.

## 1.2 Preview of the paper

This paper considers new variants of the multilevel Monte Carlo estimator [5, 3, 9, 12] for general quantities of interest $J(u(x,t,\omega))$. To provide additional guidance in the development of these new estimators, the following idealized quantity of interest is considered

$$J_l \equiv J(u_l(x,t,\omega); \Delta_0, \gamma, r, L) = p(x,t,\omega) + (\gamma^{-l}\Delta_0)^r q(x,t,\omega) \quad , \quad l = 0, \ldots, L \tag{5}$$

that includes a level-dependent model of PDE discretization error with mesh resolution $\Delta_0$, enrichment factor $\gamma$, rate parameter $r$, a maximum level parameter $L$, and random variable functions $p(x,t,\omega)$ and $q(x,t,\omega)$. This models output quantities of interest derived from numerically approximated PDEs using a sequence of $L+1$ successively refined space-time meshes. Each mesh level is refined from the previous level by increasing the number of degrees of freedom (e.g. mesh points) in each space-time dimension by the factor $\gamma$. This models quantities interest that are asymptotically approaching limit values with respect to mesh refinement.

The remainder of the paper is organized as follows:

- Section 2 reviews the Monte Carlo (MC) and multilevel Monte Carlo (MLMC) estimators. Development of the MLMC estimator and later variants consists of the following steps:

  1. Assume a multilevel calculation of the expectation statistic in incremental form (1) with right-hand-side terms estimated using Monte Carlo estimators. There is some flexibility on how this is done. This introduces algorithmic parameters that must be determined.

2. Calculate the expected value, expected mean squared error, and any systematic bias associated with the estimator.
3. Determine any unknown algorithm parameters by minimizing the expected mean squared error for a fixed algorithm cost [5, 3]. Other choices are available such as the equidistribution of error among levels [9, 12].
4. Design an algorithmic implementation with unbiased expected mean squared error control that accounts for any systematic bias.

The standard MLMC implementation is an incremental bootstrap algorithm that incrementally adds levels while updating estimates for optimal algorithm parameters. The algorithm terminates when the desired level of expected mean squared error is achieved.

- Section 3 discusses multilevel Monte Carlo estimators with sample reuse (MLMC-SR). Statistical event outcome samples are shared among levels of the MLMC-SR estimator. Reusing samples increases the expected mean squared error when compared to the MLMC estimator using the same number of samples. On the other hand, the number of quantity of interest samples required is reduced. This reduces the total cost, thus allowing more samples to be taken (which lowers the mean squared error). A more compelling motivation for sample reuse is discussed below and presented later in Sect. 4.2.

- Section 4 addresses the biased expectation statistic produced by the MLMC estimator. When output quantities of interest are computed from numerical PDE approximations, the statistical bias originates from finite-dimensional approximation error associated with the numerical PDE approximation. Whenever output quantities of interest, computed from this PDE approximation, have error that decreases at an asymptotic rate that can be estimated, the approximation error bias can be removed from the MLMC estimator. This unbiased Monte Carlo estimator is referred to as RMLMC. Unfortunately, the estimation of this asymptotic rate using MLMC level information is contaminated by statistical variability (variance) that increases the variance of the resulting RMLMC estimator.

- Section 4.2 addresses the biased expectation statistic produced by the MLMC-SR estimator. It is again observed that when the discretization error decreases at an asymptotic rate that can be estimated, the approximation error bias can be removed yielding the unbiased RMLMC-SR estimator. But unlike the RMLMC estimator, it is shown that sample reuse can be exploited to obtain reliable asymptotic rate estimates. In fact, when applied to the idealized quantity of interest (5), the calculation of the asymptotic rate is deterministic and exact.

- Section 5 provides cost and performance comparisons of the unbiased RMLMC and RMLMC-SR estimators.

# 2 Background

A starting point is the single level Monte Carlo estimator for the expectation statistic. This estimator is the underpinning of later multilevel estimators.

## 2.1 Monte Carlo (MC) estimator

Classic results for the Monte Carlo estimator are now presented. The results are used in the development of the multilevel Monte Carlo estimator.

**Definition 1 (MC Estimator)** *Let $J(u(\cdot,\omega))$ denote a random variable quantity of interest. Let $X^M = \{\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(M)}\}$ denote the set of $M$ independent and identically distributed (i.i.d.) event outcome samples that satisfy a given probability law $P$. Let $J_M$ denote the set of corresponding quantity of interest samples*

$$J_M = \{J^{(1)}, J^{(2)}, \ldots, J^{(M)}\} \tag{6}$$

*where the shorthand notation $J^{(i)} \equiv J(u(\cdot,\omega^{(i)}))$ has been used. The Monte Carlo (MC) estimated expected value of the quantity of interest $J(u(\cdot,\omega))$ with respect to the probability law $P$ is*

$$E_M[J] \equiv \frac{1}{M}\sum_{i=1}^{M} J^{(i)} \tag{7}$$

#### 2.1.1 MC expected value

The MC estimated expectation is itself a random variable. Thus, executing the MC algorithm many times for the same $J(u)$ produces a distribution of expectation estimates. Because the joint probability density associated with any particular MC calculation occurring is precisely known, $p(\omega^{(1)})p(\omega^{(2)})\dots p(\omega^{(M)})$, the expected value of the MC estimator can be readily calculated.

**Lemma 1 (MC Expected Value)** *Let $E_M[J(u)]$ denote the MC estimator described in Def.(1). $E_M[J]$ is a random variable with expected value $\mathbb{E}[J]$, i.e.,*

$$\mathbb{E}[E_M[J]] = \mathbb{E}[J] \tag{8}$$

**Proof:** Event outcome samples, $\omega^{(i)}$, $i = 1, \dots, M$, are i.i.d. random variables that each satisfy a probability law $P$. Consequently, the quantity of interest samples, $J^{(i)} \equiv J(u(\cdot, \omega^{(i)}))$, are also independent random variables. Since $E_M[J]$ is a summation over quantity of interest samples, it must also be a random variable. Application of the expectation formula (3) to a quantity of interest sample reveals that

$$\mathbb{E}[J^{(i)}] \equiv \mathbb{E}[J(u(\cdot, \omega^{(i)})] = \mathbb{E}[J] \ , \quad i = 1, \dots, M \tag{9}$$

Calculating the expected value of $E_M[J]$ then yields

$$\begin{aligned} \mathbb{E}[E_M[J]] &= \mathbb{E}\left[\frac{1}{M}\sum_{i=1}^{M} J^{(i)}\right] && \text{(10a)} \\ &= \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}\left[J^{(i)}\right] && \text{(10b)} \\ &= \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}\left[J\right] && \text{(10c)} \\ &= \mathbb{E}[J] && \text{(10d)} \end{aligned}$$

This establishes that $E_M[J]$ is a random variable with expected value $\mathbb{E}[J]$. ■

### 2.1.2 MC mean squared error

This previous lemma shows that the MC estimator is an unbiased statistic,

$$\text{Bias}[E_M[J]] = \mathbb{E}[E_M[J] - \mathbb{E}[J]] = 0$$

This does not imply that the mean squared error (MSE), $\mathbb{E}[(\mathbb{E}[J] - E_M[J])^2]$, is zero. The next lemma establishes the rate of convergence of the MSE for the MC estimator with respect to sample size $M$. A proof of this well known lemma is provided here because the steps used in the proof will be repeated later in multilevel MC analysis.

**Lemma 2 (MC Mean Squared Error)** *Let $J(u(\cdot, \omega))$ denote a random variable quantity of interest with bounded variance and $E_M[J(u)]$ denote the MC estimator described in Def. (1). The MC mean squared error (equal to the variance) depends only on $\mathbb{V}\text{ar}[J]$ and inversely on the number of samples $M$, i.e.,*

$$\mathbb{E}\left[\left(\mathbb{E}[J] - E_M[J]\right)^2\right] = \mathbb{V}\text{ar}[E_M[J]] = \frac{\mathbb{V}\text{ar}[J]}{M} \tag{11}$$

**Proof:** Application of Lemma 1 verifies that the mean squared error and variance are identical

$$\begin{aligned} \mathbb{E}\left[\left(\mathbb{E}[J] - E_M[J]\right)^2\right] &= \mathbb{E}\left[\left(\mathbb{E}[E_M[J]] - E_M[J]\right)^2\right] \\ &= \mathbb{V}\text{ar}[E_M[J]] \end{aligned} \tag{12}$$

Next, evaluate the expected mean squared error

$$\mathbb{E}\left[\left(\mathbb{E}[J] - E_M[J]\right)^2\right] = \frac{1}{M^2}\mathbb{E}\left[\sum_{i=1}^{M}(\mathbb{E}[J] - J^{(i)}) \sum_{j=1}^{M}(\mathbb{E}[J] - J^{(j)})\right] \tag{13a}$$

$$= \frac{1}{M^2}\mathbb{E}\left[\sum_{i=1}^{M}(\mathbb{E}[J] - J^{(i)})\big((\mathbb{E}[J] - J^{(i)}) + \sum_{\substack{j=1 \\ i\neq j}}^{M}(\mathbb{E}[J] - J^{(j)})\big)\right]$$

$$= \frac{1}{M^2}\mathbb{E}\left[\sum_{i=1}^{M}(\mathbb{E}[J] - J^{(i)})^2\right] \tag{13b}$$

$$= \frac{1}{M^2}\mathbb{E}\left[\sum_{i=1}^{M}((\mathbb{E}[J])^2 - 2\mathbb{E}[J]J^{(i)} + (J^{(i)})^2)\right] \tag{13c}$$

$$= \frac{1}{M}\left(\mathbb{E}[J^2] - (\mathbb{E}[J])^2\right) \tag{13d}$$

$$= \frac{\mathbb{V}\mathrm{ar}[J]}{M} \tag{13e}$$

Observe that the double summation appearing in (13a) collapses to a single summation in (13b) by exploiting the i.i.d. property (9) of the samples. The result in (13d) also follows from (9) and the generalization to $k$-moments. ■

The MC root mean squared error, $RMSE[J] = MSE^{\frac{1}{2}}[J]$ reveals the well known slow $\mathcal{O}(M^{-\frac{1}{2}})$ rate of convergence of the MC estimator.

## 2.2 Multilevel Monte Carlo (MLMC) estimator

The slow $\mathcal{O}(M^{-\frac{1}{2}})$ rate of convergence of the MC estimator makes the MC estimator costly when small RMS error is desired. Multilevel variants of the MC estimator have been introduced to reduce the estimation cost for a given RMS error level. A starting point for the multilevel Monte Carlo (MLMC) estimator is the introduction of $L + 1$ approximate models of $J(u(x, t, \omega))$

$$J^L = \{J_0, J_1, \ldots, J_L\} \ , \tag{14}$$

where the shorthand notation $J_l = J(u_l(x, t, \omega))$ has been used. Later in Sect. 2.2.6, we will assume these approximate models come from the finite-dimensional approximation of $u(x, t, \omega)$ with respect to the $(x, t)$ dependent variables, e.g., the numerical PDE solutions discretized on a sequence of meshes. Consequently, $J_0$ will be referred to as the "coarse scale" model, $J_L$ will be referred to as the "fine scale" model, and the subscript will be referred to as a "level" index.

Linearity of the expectation functional permits the calculation of $\mathbb{E}[J_L]$ in terms of the coarse scale problem $\mathbb{E}[J_0]$ and the sum of incremental corrections over levels

$$\mathbb{E}[J_L] = \mathbb{E}[J_0] + \mathbb{E}[J_1 - J_0] + \cdots + \mathbb{E}[J_L - J_{L-1}]$$

$$= \sum_{l=0}^{L} \mathbb{E}[J_l - J_{l-1}] \tag{15}$$

where the convention $J_{-1} = 0$ is imposed here and elsewhere. Given (15), the MLMC estimator is obtained by approximating the right-hand side expectations using MC estimators.

**Definition 2 (MLMC Estimator)** *Let $J(u(x,t,\omega))$ denote a random variable quantity of interest, $L$ a maximum levels parameter, and $J^L$ a set containing approximate models of $J(u(x,t,\omega))$*

$$J^L = \{J_0, J_1, \ldots, J_L\} \ , \tag{16}$$

*where the shorthand notation $J_l = J(u_l(x,t,\omega))$ for $l = 0,\ldots,L$ has been used. The Multilevel Monte Carlo (MLMC) estimator is*

$$E^L[J_L] = \sum_{l=0}^{L} E_{M_l}[J_l - J_{l-1}] \tag{17}$$

*where the convention $J_{-1} = 0$ is imposed. For each level $l$ in the right-hand-side summation, an MC estimator calculation (see Def. 1) with sample size $M_l$ is performed*

$$E_{M_l}[J_l - J_{l-1}] \equiv \frac{1}{M_l} \sum_{i=1}^{M_l} (J_l(\omega_l^{(i)}) - J_{l-1}(\omega_l^{(i)})) \tag{18}$$

*using independent and identically distributed (i.i.d.) random event outcome samples, $\{\omega_l^{(1)}, \omega_l^{(2)}, \ldots, \omega_l^{(M_l)}\}$ for level $l$, and corresponding quantity of interest samples $J_l(\omega_l^{(i)}) \equiv J(u_l(\cdot,\omega_l^{(i)}))$ and $J_{l-1}(\omega_l^{(i)}) \equiv J(u_{l-1}(\cdot,\omega_l^{(i)}))$.*

**Remark 1** *Note that because different i.i.d. event outcome samples are used in each evaluation of (18), quantities of interest $J_l$ and $J_{l-1}$ for each level $l$ must be calculated and can not be reused in other levels. This inefficiency is addressed in Sect. 3.*

### 2.2.1 MLMC expected value

The next lemma establishes that the MLMC estimator is a random variable with expected value $\mathbb{E}[J_L]$.

**Lemma 3 (MLMC Expected Value)** *Let $E^L[J_L]$ denote the MLMC estimator described in Def.(2). $E^L[J_L]$ is a random variable with expected value $\mathbb{E}[J_L]$,*

$$\mathbb{E}[E^L[J_L]] = \mathbb{E}[J_L] \tag{19}$$

**Proof:** From Lemma 1, each MC evaluation $E_{M_l}[J_l - J_{l-1}]$ is a random variable. Since $E^L[J_L]$ a sum over these MC evaluations, it then follows that $E^L[J_L]$ is a random variable. Note, for later use, that since there is no sharing of event outcome samples between levels, MLMC levels $E_{M_l}[J_l - J_{l-1}]$ are *independent* random variables. Calculating the expected value of $E^L[J_L]$ yields

$$\begin{aligned}
\mathbb{E}[E^L[J_L]] &= \mathbb{E}\left[\sum_{l=0}^{L} E_{M_l}[J_l - J_{l-1}]\right] && \text{(20a)}\\
&= \sum_{l=0}^{L} \mathbb{E}\left[E_{M_l}[J_l - J_{l-1}]\right] && \text{(20b)}\\
&= \sum_{l=0}^{L} \mathbb{E}[J_l - J_{l-1}] && \text{(20c)}\\
&= \mathbb{E}[J_L] && \text{(20d)}
\end{aligned}$$

The (20b) follows from linearity of the expectation functional. The (20c) result is minor revision of Lemma 1. This establishes that $E^L[J_L]$ is a random variable with expected value $\mathbb{E}[J_L]$. ∎

### 2.2.2 MLMC mean squared error for the $\mathbb{E}[J_L]$ statistic

The previous lemma shows that the expected value of MLMC estimator $E^L[J_L]$ is $\mathbb{E}[J_L]$ while the expectation statistic actually sought is $\mathbb{E}[J]$. This discrepancy is referred to as an approximation error bias. To reconcile this discrepancy, the mean squared error with respect to the $\mathbb{E}[J_L]$ statistic,

$\mathbb{E}[(\mathbb{E}[J_L] - E^L[J_L])^2]$, is first estimated in Lemma 4, followed by the addition of an approximation error bias correction in Lemma 5, thus yielding the desired estimate of the mean squared error, $\mathbb{E}[(\mathbb{E}[J] - E^L[J_L])^2]$.

**Lemma 4 (MLMC Mean Squared Error for $\mathbb{E}[J_L]$)** *Let $E^L[J_L]$ denote the MLMC estimator described in Def. (2). Further, assume that all models $J_l \equiv J(u_l(x,t,\omega)), l = 0,\ldots,L$ have bounded variance. The mean squared error for the MLMC estimator with respect to the $\mathbb{E}[J_L]$ statistic is given by*

$$\mathbb{E}\left[\left(\mathbb{E}[J_L] - E^L[J_L]\right)^2\right] = \mathbb{V}\mathrm{ar}[E^L[J_L]] = \sum_{l=0}^{L} \frac{\mathbb{V}\mathrm{ar}[J_l - J_{l-1}]}{M_l} \tag{21}$$

**Proof:**

$$\begin{aligned}
\mathbb{E}\left[\left(\mathbb{E}[J_L] - E^L[J_L]\right)^2\right] &= \mathbb{V}\mathrm{ar}[E^L[J_L]] && \text{(22a)}\\
&= \mathbb{V}\mathrm{ar}\left[\sum_{l=0}^{L} E_{M_l}[J_l - J_{l-1}]\right] && \text{(22b)}\\
&= \sum_{l=0}^{L} \mathbb{V}\mathrm{ar}[E_{M_l}[J_l - J_{l-1}]] && \text{(22c)}\\
&= \sum_{l=0}^{L} \frac{\mathbb{V}\mathrm{ar}[J_l - J_{l-1}]}{M_l} && \text{(22d)}
\end{aligned}$$

Equation (22a) follows from Lemma 3. The proof of Lemma 3 also establishes that $E_{M_l}[J_l - J_{l-1}]$ are each independent random variables for $l = 0,\ldots,L$. Consequently, $E^L[J_L]$ is the sum of independent random variables. Given $N$ independent random variables $\{X^{(1)}, X^{(2)}, \ldots, X^{(i)}\}$, the variance of the sum of these variables is the sum of the variances, i.e.,

$$\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} X^{(i)}\right] = \sum_{i=1}^{N} \mathbb{V}\mathrm{ar}[X^{(i)}] \tag{23}$$

which establishes the (22c) summation. Application of Lemma 2 to each term in the (22c) summation yields (22d) and the stated lemma. ∎

### 2.2.3 MLMC mean squared error for the $\mathbb{E}[J]$ statistic

Lemma 3 established that the expected value of the MLMC estimator is $\mathbb{E}[J_L]$ while $\mathbb{E}[J]$ is the expectation statistic that is actually sought. This difference is the approximation error bias associated with the MLMC estimator

$$\text{Bias}[E^L[J_L]] = \mathbb{E}[E^L[J_L] - \mathbb{E}[J]] = \mathbb{E}[J_L - J] \tag{24}$$

The next lemma adds a correction term to obtained the MSE, $\mathbb{E}[(\mathbb{E}[J] - E^L[J_L])^2]$.

**Lemma 5 (MLMC Mean Squared Error for $\mathbb{E}[J]$)** *Let $E^L[J_L]$ denote the MLMC estimator described in Def. (2). Further, assume that all models $J_l \equiv J(u_l(x,t,\omega)), l = 0,\ldots,L$ have bounded variance. The mean squared error for the MLMC estimator with respect to the $\mathbb{E}[J]$ statistic is given by*

$$\mathbb{E}\left[\left(\mathbb{E}[J] - E^L[J_L]\right)^2\right] = \sum_{l=0}^{L} \frac{\mathbb{V}\text{ar}[J_l - J_{l-1}]}{M_l} + \mathbb{E}^2[J - J_L] \tag{25}$$

**Proof:**

$$\begin{aligned}
\mathbb{E}[(\mathbb{E}[J] - E^L[J_L])^2] &= \mathbb{E}[(E^L[J_L])^2 - 2\mathbb{E}[J]E^L[J_L] + \mathbb{E}^2[J]] && \text{(26a)}\\
&= \mathbb{E}[(E^L[J_L])^2 - 2\mathbb{E}[J_L]E^L[J_L] + \mathbb{E}^2[J_L]] && \text{(26b)}\\
&\quad + \mathbb{E}[-2\mathbb{E}[J]E^L[J_L] + \mathbb{E}^2[J] + \mathbb{E}^2[J_L]] && \text{(26c)}\\
&= \mathbb{E}[(E^L[J_L] - \mathbb{E}[J_L])^2] + \mathbb{E}^2[J - J_L] && \text{(26d)}\\
&= \sum_{l=0}^{L} \frac{\mathbb{V}\text{ar}[J_l - J_{l-1}]}{M_l} + \mathbb{E}^2[J - J_L] && \text{(26e)}
\end{aligned}$$

Expanding the left-hand-side squared terms followed by addition-subtraction of identical terms yields (26d). Replacing the first term in the summation by the result of Lemma 4 yields the stated lemma. ■

### 2.2.4 Calculation of MLMC parameters

Observe that the MLMC formulation (17) introduces $L + 1$ undetermined sample size parameters, $\{M_0, M_1, \ldots, M_L\}$, that must be determined. The

MSE given in Lemma 5 contains variance terms that depend on these sample size parameters and an approximation error term that does not. Consequently, when a mean squared error of $\epsilon^2$ or less is desired, a heuristic approach for achieving this error control and determining the sample size parameters is based on the following splitting of $\epsilon^2$ and optimization strategy:

1. Require that the approximation error bias term in (25) is dominated by $\frac{\epsilon^2}{2}$
$$\mathbb{E}^2[J - J_L] \leq \frac{\epsilon^2}{2} \tag{27}$$
The left-hand-side can be made small by making $J - J_L$ small, i.e., making $J_L$ a good approximation to $J$. It is assumed here that $J - J_L$ can be monotonically decreased by increasing the maximum level parameter $L$.

2. Require that the variance terms in (25) are dominated by $\frac{\epsilon^2}{2}$
$$\sum_{l=0}^{L} \frac{\mathbb{V}\text{ar}[J_l - J_{l-1}]}{M_l} \leq \frac{\epsilon^2}{2} \tag{28}$$
The left-hand-side can be made small by minimizing the individual terms in the summation subject to a cost constraibnt.

   Let $C_l$ denote the computational cost of evaluating a quantity of interest sample at level $l$. The total cost of an MLMC calculation is
$$Cost = \sum_{l=0}^{L} M_l \; (C_l + C_{l-1}) \tag{29}$$
where the convention $C_{-1} = 0$ has been imposed. Next, minimize the left-hand-side of (28) subject to a fixed total cost (29). For notational brevity, let $V_l \equiv \mathbb{V}\text{ar}[J_l - J_{l-1}]$. Assuming $M$ a continuous variable, pose the Lagrange multiplier optimization problem with Lagrangian
$$L(M, \lambda) = \sum_{l=0}^{L} \frac{V_l}{M_l} + \lambda \, (\sum_{l=0}^{L} M_l (C_l + C_{l-1}) - Cost) \tag{30}$$

with optimality conditions

$$\frac{\partial L}{\partial M_l} = -\frac{V_l}{M_l^2} + \lambda \left( C_l + C_{l-1} \right) = 0 \ , \quad l = 0, \ldots, L \tag{31}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{l=0}^{L} M_l (C_l + C_{l-1}) - Cost = 0 \tag{32}$$

This determines the sample size parameters (rounded to the nearest integer in implementations)

$$M_l \approx \sqrt{\frac{V_l}{\lambda_L \left( C_l + C_{l-1} \right)}} \ , \quad l = 0, \ldots, L \tag{33}$$

with $\lambda_L$ chosen to satisfy (28)

$$\lambda_L = \left( \frac{\epsilon^2/2}{\sum_{l'=0}^{L} \sqrt{V_{l'} (C_{l'} + C_{l'-1})}} \right)^2 \tag{34}$$

Using this $\lambda_L$, the MLMC cost (29) then simplifies to

$$Cost = \frac{\epsilon^2/2}{\lambda_L} \tag{35}$$

#### 2.2.5 MLMC implementation

A bootstrap MLMC implementation starts from a single level $L = 0$ and incrementally adds levels until (27) and (28) are satisfied:

1. Initialize the maximum levels parameter, $L = 0$,

2. Estimate $\mathbb{V}\mathrm{ar}[J_L - J_{L-1}]$ using a variance estimator given an initial representative population of samples for $J_L - J_{L-1}$

3. Calculate $\lambda_L$ in (34) using previously estimated variances,

4. Calculate sample sizes $\{M_0, M_1, \ldots, M_L\}$ using (33) and enrich level sample populations using these revised sample sizes,

5. If $L > 0$, check the approximation error requirement (27), $E^2[J - J_L] \leq \frac{\epsilon^2}{2}$, by checking whether $|E_{M_L}[J_L - J_{L-1}]| \leq C_{approx} \frac{\epsilon}{\sqrt{2}}$ for a chosen $C_{approx}$ constant. If not satisfied, set $L = L + 1$ and go to 2.

A common criticism of the bootstrap implementation is the sequential addition of levels which inhibits a level of parallelism desirable on large scale computing platforms. Each time $\lambda_L$ is recomputed in Step 3 and sample size populations are adjusted in Step 4, additional quantity of interest calculations at all levels may be necessary. In the next section, we consider output quantities of interest derived from numerical PDE solutions obtained on a sequence of refined meshes. Using an idealized model that assumes an asymptotic behavior of a quantity of interest with respect to mesh refinement, it then becomes possible to predict optimal sample size populations for all $L$ levels given estimated variances from only coarse mesh solutions and an estimated convergence rate for the quantity of interest.

### 2.2.6 MLMC estimation using numerical PDE mesh hierarchies

Let $\{J_0, J_1, \ldots, J_L\}$ denote output quantities of interest derived from numerically approximated PDE solutions obtained on a sequence of refined meshes $\{m_0, m_1, \ldots, m_L\}$ with refinement between successive mesh levels achieved by increasing the number of mesh points in each of $d$ dimensions in space-time by a factor $\gamma$. The cost $C_l$ of evaluating a quantity of interest $J_l$ on a mesh $m_l$ is then assumed to be of the form

$$C_l = C_0 \gamma^{l\,d} \tag{36}$$

where $C_0$ denotes a reference cost value. Next, an idealized model of the quantity of interest with level-dependent discretization error model is given by

$$J_l \equiv J(u_l(x,t,\omega); \Delta_0, \gamma, r, L) = p(x,t,\omega) + (\gamma^{-l}\,\Delta_0)^r q(x,t,\omega) \ , \quad l = 0, \ldots, L \tag{37}$$

with mesh resolution $\Delta_0$, enrichment factor $\gamma$, rate parameter $r$, a maximum level parameter $L$, and random variable functions $p(x,t,\omega)$ and $q(x,t,\omega)$. Using the assumed forms (36) and (37), the optimal sample size parameters described in Sect. 2.2.4 can be determined for a given desired mean squared error $\epsilon^2$, rate parameter $r$, total number of levels $L$, space-time dimensions $d$, and coarse level variances $V_0 \equiv \mathbb{V}\mathrm{ar}[J_0]$ and $V_1 \equiv \mathbb{V}\mathrm{ar}[J_1 - J_0]$. The dependency on at least 2 variances (that must be estimated using an estimator) is unavoidable. As a practical matter, $V_0$ and $V_1$ have been chosen due to the relatively low cost in estimating them on coarse meshes. For simplicity, the formulas below assume an enrichment factor of $\gamma = 2$ (doubling mesh degrees of freedom in each dimension). Under these assumptions, the finite sums appearing in $\lambda_L$ in (34) can be evaluated in closed form yielding

$$(\lambda_L)^{-\frac{1}{2}} = \frac{2(V_0\, C_0)^{\frac{1}{2}}}{\epsilon^2} + \frac{2\left(V_1\, C_0\,(1+2^{-d})\right)^{\frac{1}{2}}}{\epsilon^2}\left(\frac{1-2^{-L(r-d/2)}}{2^{-d/2}-2^{-r}}\right) \tag{38}$$

A formula for the coarsest mesh sample size then reduces to

$$M_0 = \frac{2V_0}{\epsilon^2} + \frac{2\left(V_0\, V_1\,(1+2^{-d})\right)^{\frac{1}{2}}}{\epsilon^2}\left(\frac{1-2^{-L(r-d/2)}}{2^{-d/2}-2^{-r}}\right) \tag{39}$$

and a formula for the finest level sample size reduces to

$$M_L = \frac{2^{-L(r+d/2)+r}}{(1+2^{-d})^{\frac{1}{2}}}\left(\frac{V_1}{V_0}\right)^{\frac{1}{2}} M_0 \tag{40}$$

The remaining level sample sizes can then be simply calculated from

$$M_l = 2^{(L-l)(r+d/2)} M_L \ \ , \quad l = 1, \ldots, L-1 \tag{41}$$

which reveals the dependency on the rate parameter $r$ and the space-time dimension $d$.

**Remark 2** *The $\gamma = 2$ sample size formula (41) can be compared to the sample size formula obtained by Mishra and Schwab [8, 9] and Sukys [12]*

$$M_l = 2^{2(L-l)s} M_L \ \ , \quad l = 0, \ldots, L-1 \tag{42}$$

*In deriving this formula, they consider the finite-volume discretization of hyperbolic conservation laws and the MLMC estimation of norms of the solution expectation integrated over the spatial domain. The parameter $s$ denotes the convergence rate of the finite-volume method. The sample size parameter formula (42) has been obtained based on the equidistribution of mean squared error over levels of the MLMC estimator. The growth rate in sample size is approximately twice the rate in (40). This difference can be quite significant for even modest values of $s$.*

**Remark 3** *When solving many similar problems, where the rate of convergence $r$ is nearly constant and reliably estimated, it is often more convenient to perform an initial MLMC calculation using user specified values of $M_0$ and $M_L$ with $M_l$, $l = 1, \ldots, L-1$ obtained using (41). This strategy avoids the explicit estimation of the variances $V_0$ and $V_1$. An initial MLMC calculation can then be used as a basis for adjusting sample sizes (if needed) using either the formulas of Sect. 2.2.4 or else Eqns. (39)-(41).*

# 3 Multilevel Monte Carlo Estimator with Sample Reuse (MLMC-SR)

This section presents a variant of the MLMC estimator that reuses event outcome samples in calculating quantities of interest at levels greater than or equal to a specified level $l^*$. There are compelling reasons for sample reuse and the development of MLMC-SR that are deferred to Sect. 4.2.

**Definition 3 (MLMC-SR Estimator)** *Let $J(u(x,t,\omega))$ denote a random variable quantity of interest, $L$ a maximum levels parameter, $l^*$ a minimum sample reuse level parameter, $J^L$ a set containing approximate models of $J(u(x,t,\omega))$*

$$J^L = \{J_0, J_1, \ldots, J_L\} \ , \tag{43}$$

*where the shorthand notation $J_l = J(u_l(x,t,\omega))$ has been used, and $M^L$ a set containing decreasing integer level parameters*

$$M^L = \{M_0, M_1, \ldots, M_L\} \quad \text{with } M_j < M_i, \ \forall\, i < j \tag{44}$$

*Further, let $X^* = \{\omega_*^{(1)}, \omega_*^{(2)}, \ldots, \omega_*^{(M_{l^*})}\}$ denote an event outcome sample reuse set containing $M_{l^*}$ independent and identically distributed (i.i.d.) random event outcome samples that satisfy a given probability law $P$. The Multilevel Monte Carlo estimator with Sample Reuse (MLMC-SR) for the fine scale model $J_L$ is*

$$E_{l^*}^L[J_L] \;=\; \sum_{l=0}^{l^*-1} E_{M_l}[J_l - J_{l-1}] + \sum_{l=l^*}^{L} E_{M_l}^*[J_l - J_{l-1}] \tag{45}$$

*where $0 < l^* < L$ and the convention $J_{-1} = 0$ has been imposed. The first right-hand-side summand $E_{M_l}[J_l - J_{l-1}]$ is described in Def. 2 and the second right-hand-side summand is defined as*

$$E_{M_l}^*[J_l - J_{l-1}] \equiv \frac{1}{M_l} \sum_{i=1}^{M_l} (J_l(\omega_*^{(i)}) - J_{l-1}(\omega_*^{(i)})) \tag{46}$$

The introduction of the parameter $l^*$ in the MLMC-SR estimator permits retaining i.i.d. sampling and a standard MLMC calculation for levels less than $l^*$. As a practical matter, $l^*$ is always chosen greater than zero. Motivated by the analysis of Sect. 3.2, choosing $l^* > 0$ avoids possible large covariance contributions to the mean squared error when sample reuse includes the lowest level data, $J_0$.

### 3.1 MLMC-SR expected value

The next lemma shows that sample reuse in the MLMC-SR estimator does not change the resulting expected value.

**Lemma 6 (MLMC-SR Expected Value)** *Let $E^L_{l^*}[J_L]$ denote the MLMC-SR estimator described in Def.(3). $E^L_{l^*}[J_L]$ is a random variable with expected value $\mathbb{E}[J_L]$,*

$$\mathbb{E}[E^L_{l^*}[J_L]] = \mathbb{E}[J_L] \tag{47}$$

**Proof:** A proof follows from elements of the proof of Lemma 3 and is unaffected by sample reuse. ■

### 3.2 MLMC-SR mean squared error for the $\mathbb{E}[J_L]$ statistic

The MSE for the MLMC-SR estimator is now complicated by the use of sample reuse which introduces added covariance contributions.

**Lemma 7 (MLMC-SR Mean Squared Error for $\mathbb{E}[J_L]$)** *Let $E^L_{l^*}[J_L]$ denote the MLMC-SR estimator described in Def. (3). Further, assume that all models $J_l \equiv J(u_l(x,t,\omega), l = 0,\dots,L$ have bounded variance and covariance. The mean squared error for the MLMC-SR estimator with respect to the $\mathbb{E}[J_L]$ statistic is given by*

$$\mathbb{E}\left[\left(\mathbb{E}[J_L]-E^L_{l^*}[J_L]\right)^2\right] = \sum_{l=0}^{L}\frac{1}{M_l}\left(\mathbb{V}\mathrm{ar}[J_l-J_{l-1}] + 2\sum_{\substack{l'=l+1\\ l\geq l^*}}^{L}\mathbb{C}\mathrm{ov}[J_l-J_{l-1}, J_{l'}-J_{l'-1}]\right) \tag{48}$$

**Proof:** Due to sample reuse in the MLMC-SR estimator (45) for levels greater than or equal to $l^*$, the level summand terms $E^*_{M_l}[J_l - J_{l-1}]$ are no longer independent random variables which complicates the proof. The proof is simplified by first rewriting (45) as a sum over sample subsets that form independent random variables regardless of whether sample reuse is or is not used. Define the contiguous subinterval index sets

$$I_{M_k} = \{M_{L-k+1} + 1, M_{L-k+1} + 2, \ldots, M_{L-k}\} \ , \quad k = 0, \ldots, L \tag{49}$$

with the convention that $M_{L+1} = 0$. The index sets are non-overlapping, $I_{M_k} \cap_{k \neq k'} I_{M_{k'}} = \emptyset$, with unions

$$\cup_{k=0}^{l} I_{M_k} = \{1, \ldots, M_{L-l}\} \ , \quad l = 0, \ldots, L \tag{50}$$

and cardinality

$$|I_{M_k}| = M_{L-k} - M_{L-k+1} \tag{51}$$

As a strategy for the remainder of the proof, sample reuse will initially be assumed over all $L+1$ levels followed later by a reintroduction of the level parameter $l^*$ based on vanishing covariances. Rewrite the MLMC-SR estimator as a summation over the index subsets $I_{M_k}$

$$\begin{aligned}
E_{l^*}^{L}[J_L] &= \sum_{l=0}^{L} E_{M_l}^{*}[J_l - J_{l-1}] && \text{(52a)}\\
&= \sum_{l=0}^{L} \frac{1}{M_l} \sum_{i=1}^{M_l} (J_l(\omega_*^{(i)}) - J_{l-1}(\omega_*^{(i)})) && \text{(52b)}\\
&= \sum_{k=0}^{L} \sum_{i \in I_{M_k}} \sum_{l=0}^{L-k} \frac{1}{M_l} (J_l(\omega_*^{(i)}) - J_{l-1}(\omega_*^{(i)})) && \text{(52c)}\\
&= \sum_{k=0}^{L} T_k && \text{(52d)}
\end{aligned}$$

where in (52d)

$$T_k \equiv \sum_{i \in I_{M_k}} \sum_{l=0}^{L-k} \frac{1}{M_l} (J_l(\omega_*^{(i)}) - J_{l-1}(\omega_*^{(i)})) \tag{53}$$

are now independent random variables. Evaluating the MSE for the MLMC-SR estimator

$$\mathbb{E}[(\mathbb{E}[J_L] - E_{l^*}^L[J_L])^2] = \mathbb{V}\text{ar}[E_{l^*}^L[J_L]] \tag{54a}$$

$$= \mathbb{V}\text{ar}\left[\sum_{k=0}^{L} T_k\right] \tag{54b}$$

$$= \sum_{k=0}^{L} \mathbb{V}\text{ar}\left[T_k\right] \tag{54c}$$

$$= \sum_{k=0}^{L} (\mathbb{E}[T_k^2] - \mathbb{E}^2[T_k]) \tag{54d}$$

where Lemma 6 has been used in (54a) and the identity (23) has been used in (54c). Thus, the remaining tasks are the evaluation of $\mathbb{E}[T_k^2]$ and $\mathbb{E}^2[T_k]$. Evaluating $\mathbb{E}[T_k]$

$$\mathbb{E}[T_k] = \mathbb{E}\left[\sum_{i \in I_{M_k}} \sum_{l=0}^{L-k} \frac{1}{M_l}(J_l(\omega_*^{(i)}) - J_{l-1}(\omega_*^{(i)}))\right] \tag{55a}$$

$$= \sum_{l=0}^{L-k} \frac{1}{M_l} \sum_{i \in I_{M_k}} \mathbb{E}[J_l(\omega_*^{(i)}) - J_{l-1}(\omega_*^{(i)})] \tag{55b}$$

$$= \sum_{l=0}^{L-k} \frac{|I_{M_k}|}{M_l} \mathbb{E}[J_l - J_{l-1}] \tag{55c}$$

where linearity of the expectation functional has been used in (55b) and the i.i.d. property (9) has been used in the evaluation of (55c). Using this result, it then follows that

$$\mathbb{E}^2[T_k] = \sum_{l=0}^{L-k} \sum_{l'=0}^{L-k} \frac{|I_{M_k}|^2}{M_l M_{l'}} \mathbb{E}[J_l - J_{l-1}]\mathbb{E}[J_{l'} - J_{l'-1}] \tag{56}$$

Next, $\mathbb{E}[T_k^2]$ is evaluated

$$\begin{aligned}
\mathbb{E}[T_k^2] &= \sum_{l=0}^{L-k}\sum_{l'=0}^{L-k}\frac{1}{M_l M_{l'}}\left(\sum_{i\in I_{M_k}}(J_l(\omega_*^{(i)})-J_{l-1}(\omega_*^{(i)}))\right) \\
&\qquad \times\left(\sum_{i'\in I_{M_k}}(J_{l'}(\omega_*^{(i')})-J_{l'-1}(\omega_*^{(i')}))\right) &\text{(57a)}\\
&= \sum_{l=0}^{L-k}\sum_{l'=0}^{L-k}\frac{1}{M_l M_{l'}}\left(\sum_{i\in I_{M_k}}(J_l(\omega_*^{(i)})-J_{l-1}(\omega_*^{(i)}))\right) \\
&\qquad \times\left((J_{l'}(\omega_*^{(i)})-J_{l'-1}(\omega_*^{(i)}))+\sum_{\substack{i'\in I_{M_k}\\ i'\neq i}}(J_{l'}(\omega_*^{(i')})-J_{l'-1}(\omega_*^{(i')}))\right) \\
&= \sum_{l=0}^{L-k}\sum_{l'=0}^{L-k}\frac{|I_{M_k}|}{M_l M_{l'}}\mathbb{E}[(J_l-J_{l-1})(J_{l'}-J_{l'-1})] &\text{(57b)}\\
&\quad +\sum_{l=0}^{L-k}\sum_{l'=0}^{L-k}\frac{|I_{M_k}|(|I_{M_k}|-1)}{M_l M_{l'}}\mathbb{E}[J_l-J_{l-1}]\mathbb{E}[J_{l'}-J_{l'-1}] &\text{(57c)}
\end{aligned}$$

where the i.i.d. property of samples has been used in (57b) and (57c). Inserting (56), (57b), and (57c) into the MSE equation (54d) yields

$$\mathbb{E}[(\mathbb{E}[J_L]-E^L_{l^*}[J_L])^2] = \sum_{k=0}^{L}\sum_{l=0}^{L-k}\sum_{l'=0}^{L-k}\frac{|I_{M_k}|}{M_l M_{l'}}\mathbb{C}\text{ov}[J_l-J_{l-1},J_{l'}-J_{l'-1}] \tag{58a}$$

$$= \sum_{l=0}^{L}\sum_{l'=0}^{L}\sum_{k=0}^{L-\max(l,l')}\frac{|I_{M_k}|}{M_l M_{l'}}\mathbb{C}\text{ov}[J_l-J_{l-1},J_{l'}-J_{l'-1}] \tag{58b}$$

$$= \sum_{l=0}^{L}\sum_{l'=0}^{L}\frac{\min(M_l,M_{l'})}{M_l M_{l'}}\mathbb{C}\text{ov}[J_l-J_{l-1},J_{l'}-J_{l'-1}] \tag{58c}$$

$$= \sum_{l=0}^{L}\frac{\mathbb{V}\text{ar}[J_l-J_{l-1}]}{M_l} \tag{58d}$$

$$+ \sum_{l=0}^{L}\sum_{\substack{l'=0\\ l'\neq l}}^{L}\frac{\min(M_l,M_{l'})}{M_l M_{l'}}\mathbb{C}\text{ov}[J_l-J_{l-1},J_{l'}-J_{l'-1}] \tag{58e}$$

$$= \sum_{l=0}^{L}\frac{1}{M_l}\Big(\mathbb{V}\text{ar}[J_l-J_{l-1}] \tag{58f}$$

$$+2\sum_{l'=l+1}^{L}\mathbb{C}\text{ov}[J_l-J_{l-1},J_{l'}-J_{l'-1}]\Big) \tag{58g}$$

To perform the rearrangement of summations in (58b), it must be shown that the following identity holds for any summand depending on $(k,l,l')$

$$\sum_{k=0}^{L}\sum_{l=0}^{L-k}\sum_{l'=0}^{L-k} = \sum_{l=0}^{L}\sum_{l'=0}^{L}\sum_{k=0}^{L-max(l,l')}$$

A standard approach to showing this identity utilizes Iverson bracket notation [7] that reduces the task to the manipulation of elementary logic conjunctions, i.e.,

$$\sum_{k=0}^{L}\sum_{l=0}^{L-k}\sum_{l'=0}^{L-k} = \sum_{k}\sum_{l}\sum_{l'} [\mathbf{0 \leq k \leq L}]\cdot[\mathbf{0 \leq l \leq L-k}]\cdot[\mathbf{0 \leq l' \leq L-k}] \tag{59a}$$

$$= \sum_{k}\sum_{l}\sum_{l'} [\mathbf{0 \leq k}]\cdot[\mathbf{k \leq L}]\cdot[\mathbf{0 \leq l}]\cdot[\mathbf{l \leq L-k}]\cdot[\mathbf{0 \leq l'}]\cdot[\mathbf{l' \leq L-k}] \tag{59b}$$

$$= \sum_{l}\sum_{l'}\sum_{k} [\mathbf{0 \leq k}]\cdot[\mathbf{k \leq L}]\cdot[\mathbf{0 \leq l}]\cdot[\mathbf{k \leq L-l}]\cdot[\mathbf{0 \leq l'}]\cdot[\mathbf{k \leq L-l'}] \tag{59c}$$

$$= \sum_{l}\sum_{l'}\sum_{k} [\mathbf{0 \leq l}]\cdot[\mathbf{l \leq L}]\cdot[\mathbf{0 \leq l'}]\cdot[\mathbf{l' \leq L}]\cdot[\mathbf{0 \leq k}]\cdot[\mathbf{k \leq L} - \max(\mathbf{l},\mathbf{l'})] \tag{59d}$$

$$= \sum_{l}\sum_{l'}\sum_{k} [\mathbf{0 \leq l \leq L}]\cdot[\mathbf{0 \leq l' \leq L}]\cdot[\mathbf{0 \leq k \leq L} - \max(\mathbf{l},\mathbf{l'})] \tag{59e}$$

$$= \sum_{l=0}^{L}\sum_{l'=0}^{L}\sum_{k=0}^{L-\max(l,l')} \tag{59f}$$

where in (59e) the logic simplifications

$$[k \leq L]\cdot[k \leq L-l]\cdot[k \leq L-l'] \text{ implies } [k \leq L-\max(l,l')]$$

and

$$[l \leq L-k] \text{ implies } [l \leq L] \quad , \quad [l' \leq L-k] \text{ implies } [l' \leq L]$$

have been used. The simplification in (58c) results from application of (50) and (51). Invoking symmetry of both the covariance and $\frac{\min(M_l,M_{l'})}{M_l M_{l'}}$ terms yields (58g). As a final step, all covariances for levels $l < l^*$ vanish due to the assumption of i.i.d. sampling at those levels. ■

#### 3.2.1 MLMC-SR mean squared error for the $\mathbb{E}[J]$ statistic

Lemma 3 and Lemma 6 established that the expected values of the MLMC estimator and the MLMC-SR estimator are identical. Thus, the approximation error bias is also identical

$$\text{Bias}[E_{l^*}^{L}[J_L]] = \mathbb{E}[E_{l^*}^{L}[J_L] - \mathbb{E}[J]] = \mathbb{E}[J_L - J] \tag{60}$$

The next lemma adds a correction term to obtainb the MSE, $\mathbb{E}[(\mathbb{E}[J] - E_{l^*}^{L}[J_L])^2]$.

**Lemma 8 (MLMC-SR Mean Squared Error for $\mathbb{E}[J]$)** *Let $E^L_{l^*}[J_L]$ denote the MLMC-SR estimator described in Def. (3). Further, assume that all models $J_l \equiv J(u_l(x,t,\omega)), l = 0,\ldots,L$ have bounded variance and covariance. The mean squared error for the MLMC-SR estimator with respect to the $\mathbb{E}[J]$ statistic is given by*

$$\mathbb{E}\left[\left(\mathbb{E}[J]-E^L_{l^*}[J_L]\right)^2\right] = \sum_{l=0}^{L} \frac{1}{M_l}\left(\mathbb{V}\mathrm{ar}[J_l - J_{l-1}] + 2\sum_{\substack{l'=l+1 \\ l\geq l^*}}^{L} \mathbb{C}\mathrm{ov}[J_l - J_{l-1}, J_{l'} - J_{l'-1}]\right) + \mathbb{E}^2[J - J_L] \tag{61}$$

**Proof:** The proof closely follows the proof of Lemma 5 and is omitted. ■

#### 3.2.2 Calculation of MLMC-SR parameters

Determining parameters for the MLMC-SR estimator follows a path similar to the MLMC estimator. The MLMC-SR formulation (45) introduces $L+1$ undetermined sample size parameters, $\{M_0, M_1, \ldots, M_L\}$, that must be determined. The MSE given in Lemma 8 contains variance and covariance terms that depend on these sample size parameters and an approximation error that does not. Following the procedure of Sect. 2.2.4, when a mean squared error of $\epsilon^2$ or less is desired, a heuristic approach for achieving this error control and determining the sample size parameters is based on the following splitting of $\epsilon^2$ and optimization strategy:

1. Require that the approximation error bias term in (61) is dominated by $\frac{\epsilon^2}{2}$
$$\mathbb{E}^2[J - J_L] \leq \frac{\epsilon^2}{2} \tag{62}$$
It is assumed here that $J - J_L$ can be monotonically decreased by increasing the maximum level parameter $L$.

2. Require that the variance-covariance terms in (61) are dominated by

$\frac{\epsilon^2}{2}$

$$\sum_{l=0}^{L} \frac{1}{M_l} \left( \mathbb{V}\text{ar}[J_l - J_{l-1}] + 2 \sum_{\substack{l'=l+1 \\ l \geq l^*}}^{L} \mathbb{C}\text{ov}[J_l - J_{l-1}, J_{l'} - J_{l'-1}] \right) \leq \frac{\epsilon^2}{2} \tag{63}$$

The left-hand-side can be made small by minimizing the individual terms in the summation subject to a cost constraint.

Let $C_l$ denote the computational cost of evaluating and quantity of interest sample at level $l$. The total cost of an MLMC-SR calculation differs significantly from the MLMC cost (29) due to the reuse of samples at finer levels, i.e.,

$$Cost = \sum_{l=0}^{l^*} M_l (C_l + C_{l-1}) + \sum_{l=l^*+1}^{L} M_l \; C_l \tag{64}$$

where the convention $C_{-1} = 0$ has again been imposed. Next, minimize the left-hand-side of (63) subject to the fixed total cost (64). Let $\widehat{V}_l$ denote the combined variance-covariance terms

$$\widehat{V}_l \equiv \mathbb{V}\text{ar}[J_l - J_{l-1}] + 2 \sum_{\substack{l'=l+1 \\ l \geq l^*}}^{L} \mathbb{C}\text{ov}[J_l - J_{l-1}, J_{l'} - J_{l'-1}] \;\; , \quad l = 0, \ldots, L \tag{65}$$

Assuming $M$ a continuous variable, pose the Lagrange multiplier optimization problem with Lagrangian

$$L^{SR}(M, \lambda) = \sum_{l=0}^{L} \frac{\widehat{V}_l}{M_l} + \lambda \left( \sum_{l=0}^{l^*} M_l (C_l + C_{l-1}) + \sum_{l=l^*+1}^{L} M_l \; C_l - Cost \right) \tag{66}$$

with optimality conditions

$$\frac{\partial L^{SR}}{\partial M_l} = -\frac{\widehat{V}_l}{M_l^2} + \lambda \, (C_l + C_{l-1}) = 0 \;\; , \quad l = 0, \ldots, l^*, \tag{67}$$

$$\frac{\partial L^{SR}}{\partial M_l} = -\frac{\widehat{V}_l}{M_l^2} + \lambda \, C_l = 0 \;\; , \quad l = l^* + 1, \ldots, L, \tag{68}$$

and

$$\frac{\partial L^{SR}}{\partial \lambda} = \sum_{l=0}^{l^*} M_l(C_l + C_{l-1}) + \sum_{l=l^*+1}^{L} M_l\, C_l - Cost = 0 \tag{69}$$

This determines the sample size parameters (rounded to the nearest integer in implementations)

$$M_l \approx \begin{cases} \sqrt{\frac{\widehat{V}_l}{\lambda_L\,(C_l+C_{l-1})}} & \text{for } 0 \leq l \leq l^* \\ \\ \sqrt{\frac{\widehat{V}_l}{\lambda_L\, C_l}} & \text{for } l^* < l \leq L \end{cases} \tag{70}$$

with $\lambda_L$ chosen to satisfy (63)

$$\lambda_L = \left( \frac{\epsilon^2/2}{\sum_{l'=0}^{l^*} \sqrt{\widehat{V}_{l'}(C_{l'} + C_{l'+1})} + \sum_{l'=l^*+1}^{L} \sqrt{\widehat{V}_{l'} C_{l'}}} \right)^2 \tag{71}$$

The MLMC-SR cost (64) then simplifies to

$$Cost = \frac{\epsilon^2/2}{\lambda_L} \tag{72}$$

#### 3.2.3 MLMC-SR implementation

The bootstrap MLMC-SR implementation introduces new complications not present in the MLMC implementation of Sect. 2.2.5. The MLMC-SR bootstrap implementation starts from a single level $L = 0$ and incrementally adds levels until (62) and (63) are satisfied. New i.i.d. outcome event samples are generated if the current level $L < l^*$. Otherwise, an i.i.d. event outcome sample reuse set is established when $L = l^*$ and those event outcome samples are reused for $L \geq l^*$.

1. Initialize the maximum levels parameter, $L = 0$,

2. If $L = l^*$, establish a event outcome sample reuse population.

3. Estimate $\mathbb{V}\mathrm{ar}[J_L - J_{L-1}]$ given an initial population of new i.i.d. samples if $L < l^*$ or reused samples if $L \geq l^*$

4. Estimate $\mathbb{C}\mathrm{ov}[J_l - J_{l-1}, J_L - J_{L-1}]$ for $l^* \leq l < L-1$.

5. Calculate $\lambda_L$ in (71) using previously estimated variances and covariances.

6. Calculate sample sizes $\{M_0, M_1, \ldots, M_L\}$ using (70) and enrich level sample populations using these revised sample sizes,

7. If $L > 0$, check the approximation error requirement (62), $E^2[J - J_L] \leq \frac{\epsilon^2}{2}$, by checking whether $|E_{M_L}[J_L - J_{L-1}]| \leq C_{approx}\frac{\epsilon}{\sqrt{2}}$ for a chosen $C_{approx}$ constant. If not satisfied, set $L = L + 1$ and go to 2.

### 3.2.4 MLMC-SR estimation using numerical PDE mesh hierarchies

The calculation of optimal sample size parameters, assuming a PDE mesh hierarchy and idealized quantity of interest, follows the development of Sect. 2.2.6 for the MLMC estimator. Let $\{J_0, J_1, \ldots, J_L\}$ denote quantities of interest derived from numerically approximated PDE solutions obtained on a sequence of refined meshes $\{m_0, m_1, \ldots, m_L\}$ with refinement between successive mesh levels achieved by increasing the number of mesh points in each of $d$ dimensions in space-time by the factor $\gamma$. The cost $C_l$ of evaluating a quantity of interest $J_l$ on a mesh $m_l$ is then assumed to be of the form

$$C_l = C_0 \gamma^{l\,d} \tag{73}$$

where $C_0$ denotes a reference cost value. Next, assume the following idealized model of the quantity of interest with level-dependent discretization error model introduced previously

$$J_l \equiv J(u_l(x,t,\omega)) = p(x,t,\omega) + (\gamma^{-l}\,\Delta_0)^r q(x,t,\omega) \;\;,\quad l = 0, \ldots, L \tag{74}$$

with mesh resolution $\Delta_0$, enrichment factor $\gamma$, rate parameter $r$, a maximum level parameter $L$, and random variable functions $p(x,t,\omega)$ and $q(x,t,\omega)$. Using the assumed forms (73) and (74), the optimal sample size parameters described in Sect. 3.2.2 can be determined for a given desired mean squared error $\epsilon^2$, rate parameter $r$, total number of levels $L$, space-time dimensions $d$, and coarse level variances $V_0 \equiv \mathbb{V}\mathrm{ar}[J_0]$ and $V_1 \equiv \mathbb{V}\mathrm{ar}[J_1 - J_0]$. To simplify the presentation, $0 < l^* < L$ and $\gamma = 2$ has been assumed. Under these assumptions, $\lambda_L$ in (71) can be simplified to

$$(\lambda_L)^{-\frac{1}{2}} = \frac{2(V_0C_0)^{\frac{1}{2}}}{\epsilon^2} + \frac{2(V_1C_0)^{\frac{1}{2}}}{\epsilon^2}\Big(f_1(l^*,r,d) + f_2(l^*,L,r,d) + f_3(l^*,L,r,d)\Big) \tag{75}$$

using the functions

$$\begin{aligned}
f_1(l^*,r,d) &= (1+2^{-d})^{1/2}\left(\frac{1-2^{-(l^*-1)(r-d/2)}}{2^{-d/2}-2^{-r}}\right)\\
f_2(l^*,L,r,d) &= 2^r\,(1+2^{-d})^{1/2}2^{-l^*(r-d/2)}\left[1+2\left(\frac{1-2^{-(L-l^*)r}}{2^r-1}\right)\right]^{\frac{1}{2}}\\
f_3(l^*,L,r,d) &= 2^r\sum_{l=l^*+1}^{L}2^{-l(r-d/2)}\left[1+2\left(\frac{1-2^{-(L-l)r}}{2^r-1}\right)\right]^{\frac{1}{2}}
\end{aligned} \tag{76}$$

A formula for the lowest level sample size parameter then reduces to

$$M_0 = \frac{2V_0}{\epsilon^2} + \frac{2(V_0\,V_1)^{\frac{1}{2}}}{\epsilon^2}\Big(f_1(l^*,r,d) + f_2(l^*,L,r,d) + f_3(l^*,L,r,d)\Big) \tag{77}$$

and the remaining level sample sizes can then be calculated from

$$M_l = \begin{cases}
\frac{2^{-l(r+d/2)+r}}{(1+2^{-d})^{\frac{1}{2}}}\left(\frac{V_1}{V_0}\right)^{\frac{1}{2}} M_0 & \text{for } 0 < l < l^*\\
\frac{2^{-l(r+d/2)+r}}{(1+2^{-d})^{\frac{1}{2}}}\left[1+2\left(\frac{1-2^{-(L-l)r}}{2^r-1}\right)\right]^{\frac{1}{2}}\left(\frac{V_1}{V_0}\right)^{\frac{1}{2}} M_0 & \text{for } l = l^*\\
2^{-l(r+d/2)+r}\left[1+2\left(\frac{1-2^{-(L-l)r}}{2^r-1}\right)\right]^{\frac{1}{2}}\left(\frac{V_1}{V_0}\right)^{\frac{1}{2}} M_0 & \text{for } l^* < l \le L
\end{cases} \tag{78}$$

which reveals the dependency on the rate parameter $r$, space-time dimension $d$, and maximum number of levels $L$.

# 4 Unbiased RMLMC and RMLMC-SR Estimation of the $\mathbb{E}[J]$ Statistic

Once again, recall the idealized quantity of interest model that includes a level-dependent model of discretization error

$$J_l \equiv J(u_l(x,t,\omega)) = p(x,t,\omega) + (\gamma^{-l}\Delta_0)^r q(x,t,\omega)\ , \quad l = 0,\dots,L \tag{79}$$

When a rate $r > 0$ is given and $\Delta_0$ finite, an explicit recovery of the zero discretization error limit, $\lim_{\Delta_0 \to 0} J(u_l(x,t,\omega); \Delta_0, \gamma, r, L) = J(u)$, can still be obtained for $\gamma > 1$ via the extrapolation formula

$$J = J_L + \frac{1}{\gamma^r - 1}(J_L - J_{L-1}) \tag{80}$$

This equation is verified using (79). Exploiting linearity of the expectation functional, a similar extrapolation formula for expectation is obtained

$$\mathbb{E}[J] = \mathbb{E}[J_L] + \frac{1}{\gamma^r - 1}\mathbb{E}[J_L - J_{L-1}] \tag{81}$$

Replacing the right-hand-side of (81) with either the MLMC or MLMC-SR estimator yields an unbiased estimator for $\mathbb{E}[J]$ referred to later as RMLMC or RMLMC-SR, respectively. Analysis given below, that assumes the idealized quantity of interest (79) with known rate $r > 0$ and $\gamma > 1$, reveals that this modification changes the calculation of the optimal sample size at the finest level $L$ for both estimators.

This approach to removing MLMC approximation error bias is discussed in [5, 3]. Unfortunately, in practical implementations, the rate $r$ appearing in (81) is not known *a priori* and must be approximated as part of the RMLMC and RMLMC-SR estimators. Estimating this rate from RMLMC level information is convenient but is subject to statistical variability that increases the variance and degrades the overall reliability of the resulting RMLMC estimator. A new approach to rate estimation using the RMLMC-SR estimator with sample reuse reduces or eliminates variability in the rate estimation. This can significantly enhance the accuracy of the resulting RMLMC-SR estimator. This topic is addressed in Sects. 4.1.4 and 4.2.4 for the RMLMC and RMLMC-SR estimators, respectively.

## 4.1 Unbiased RMLMC estimation of $\mathbb{E}[J]$ for the idealized model

Motivated by the extrapolation formula (81), an unbiased MLMC estimator for $\mathbb{E}[J]$ is defined.

**Definition 4 (RMLMC Estimator for the Idealized Model)** *Let $J(u(x,t,\omega))$ denote a random variable quantity of interest and $J_l \equiv J(u_l(x,t,\omega))$*

*an idealized quantity of interest model with level-dependent discretization model*

$$J_l \equiv J(u_l(x,t,\omega)) = p(x,t,\omega) + (\gamma^{-l}\Delta_0)^r q(x,t,\omega) \ , \quad l = 0,\dots,L \tag{82}$$

*with mesh resolution $\Delta_0$, enrichment factor $\gamma$, rate parameter $r$, a maximum level parameter $L$, and random variable functions $p(x,t,\omega)$ and $q(x,t,\omega)$ Further, let $E^L[J_L]$ denote the MLMC estimator described in Def. 2. The unbiased RMLMC estimator for the $\mathbb{E}[J]$ statistic is*

$$RE^L[J_L] = E^L[J_L] + \frac{1}{\gamma^r - 1} E_{M_L}[J_L - J_{L-1}] \tag{83}$$

*where $E_{M_L}[J_L - J_{L-1}]$ is described in Def. 2.*

#### 4.1.1 RMLMC expected value for the idealized model

The next lemma verifies that the RMLMC estimator provides an unbiased estimate of $\mathbb{E}[J]$ for the idealized quantity of interest model.

**Lemma 9 (RMLMC Expected Value)** *Let $RE^L[J_L]$ denote the RMLMC estimator described in Def.(4). $RE^L[J_L]$ is a random variable with expected value $\mathbb{E}[J]$,*

$$\mathbb{E}[RE^L[J_L]] = \mathbb{E}[J] \tag{84}$$

**Proof:** From Lemmas 1 and 3, $RE^L[J_L]$ is the sum of random variables and therefore is a random variable. Next, calculate the expectation

$$\begin{aligned}
\mathbb{E}\left[RE^L[J_L]\right] &= \mathbb{E}\left[E^L[J_L] + \frac{1}{\gamma^r - 1} E_{M_L}[J_L - J_{L-1}]\right] && \text{(85a)}\\
&= \mathbb{E}\left[E^L[J_L]\right] + \frac{1}{\gamma^r - 1}\mathbb{E}\left[E_{M_L}[J_L - J_{L-1}]\right] && \text{(85b)}\\
&= \mathbb{E}[J_L] + \frac{1}{\gamma^r - 1}\mathbb{E}[J_L - J_{L-1}] && \text{(85c)}\\
&= \mathbb{E}\left[J_L + \frac{1}{\gamma^r - 1}(J_L - J_{L-1})\right] && \text{(85d)}\\
&= \mathbb{E}[J] && \text{(85e)}
\end{aligned}$$

where (85b) follows from linearity of the expectation functional, (85c) follows from Lemmas 1 and 3, (85d) again uses linearity of the expectation functional, and (85e) follows from (80). This verifies the stated lemma. ■

#### 4.1.2 RMLMC mean squared error for $\mathbb{E}[J]$ for the idealized model

The next lemma shows that the mean squared error for the RMLMC estimator differs from the MLMC mean squared error only at the finest level $L$.

**Lemma 10 (RMLMC Mean Square Error for $\mathbb{E}[J]$)** *Let $RE^L[J_L]$ denote the RMLMC estimator described in Def. (4). Further, assume that all models $J_l \equiv J(u_l(x,t,\omega))$, $l = 0,\dots,L$ have bounded variance. The unbiased mean squared error for the RMLMC estimator with respect to the $\mathbb{E}[J]$ statistic is given by*

$$\mathbb{E}\left[\left(\mathbb{E}[J] - RE^L[J_L]\right)^2\right] = \sum_{l=0}^{L-1} \frac{\mathbb{V}\text{ar}[J_l - J_{l-1}]}{M_l} + \left(\frac{1}{1-\gamma^{-r}}\right)^2 \frac{\mathbb{V}\text{ar}[J_L - J_{L-1}]}{M_L} \tag{86}$$

**Proof:** Begin by evaluating the mean squared error

$$\begin{aligned} \mathbb{E}\left[(\mathbb{E}[J] - RE^L[J_L])^2\right] &= \mathbb{E}\left[(RE^L[J_L])^2 + \mathbb{E}^2[J] - 2\mathbb{E}[J]\, RE^L[J_L]\right] && \text{(87a)}\\ &= \mathbb{E}\left[(RE^L[J_L])^2\right] - \mathbb{E}^2[J] && \text{(87b)} \end{aligned}$$

Expanding the left-hand-side squared terms and application of Lemma 9 yields the right-hand-side (87b). Using the extrapolation formula (81), $\mathbb{E}^2[J]$ in (87b) can be replaced by

$$\mathbb{E}^2[J] = \mathbb{E}^2[J_L] + \left(\frac{1}{\gamma^r - 1}\right)^2 \mathbb{E}^2[J_L - J_{L-1}] + \frac{2}{\gamma^r - 1}\mathbb{E}[J_L - J_{L-1}]\mathbb{E}[J_L] \tag{88}$$

Next, evaluate the first right-hand-side term in (87b)

$$\begin{aligned} \mathbb{E}\left[(RE^L[J_L])^2\right] &= \mathbb{E}\left[(E^L[J_L])^2\right] && \text{(89a)}\\ &\quad + \left(\frac{1}{\gamma^r - 1}\right)^2 \mathbb{E}\left[(E_{M_L}[J_L - J_{L-1}])^2\right] && \text{(89b)}\\ &\quad + \frac{2}{\gamma^r - 1}\mathbb{E}\left[E^L[J_L]\, E_{M_L}[J_L - J_{L-1}]\right] && \text{(89c)} \end{aligned}$$

Using Lemma 3 and Lemma 4, the first right-hand-side term (89a) reduces to

$$\mathbb{E}\left[(E^L[J_L])^2\right] = \sum_{l=0}^{L} \frac{\mathbb{V}\text{ar}[J_l - J_{l-1}]}{M_l} + \mathbb{E}^2[J_L] \tag{90}$$

Similarly, using Lemma 1 and Lemma 2, the second right-hand-side term (89b) expectation reduces to

$$\mathbb{E}\left[(E_{M_L}[J_L - J_{L-1}])^2\right] = \frac{\mathbb{V}\text{ar}[J_L - J_{L-1}]}{M_L} + \mathbb{E}^2[J_L - J_{L-1}] \tag{91}$$

Evaluation of the third right-hand-side term (89c) yields

$$\begin{aligned}
\mathbb{E}\left[E_{M_L}[J_L - J_{L-1}]E^L[J_L]\right] &= \mathbb{E}\left[E_{M_L}[J_L - J_{L-1}]\sum_{l=0}^{L} E_{M_l}[J_l - J_{l-1}]\right] && \text{(92a)}\\
&= \mathbb{E}\left[E^2_{M_L}[J_L - J_{L-1}]\right] && \text{(92b)}\\
&+ \mathbb{E}\left[E_{M_L}[J_L - J_{L-1}]\sum_{\substack{l=0\\ l\neq L}}^{L} E_{M_l}[J_l - J_{l-1}]\right] && \text{(92c)}\\
&= \mathbb{E}\left[E^2_{M_L}[J_L - J_{L-1}]\right] && \text{(92d)}\\
&+ \mathbb{E}[J_L - J_{L-1}]\sum_{\substack{l=0\\ l\neq L}}^{L} \mathbb{E}[J_l - J_{l-1}] && \text{(92e)}\\
&= \mathbb{E}\left[E^2_{M_L}[J_L - J_{L-1}]\right] - \mathbb{E}^2[J_L - J_{L-1}] && \text{(92f)}\\
&+ \mathbb{E}[J_L - J_{L-1}]\sum_{l=0}^{L} \mathbb{E}[J_l - J_{l-1}] && \text{(92g)}\\
&= \frac{\mathbb{V}\text{ar}[J_L - J_{L-1}]}{M_L} + \mathbb{E}[J_L - J_{L-1}]E^L[J_L] && \text{(92h)}
\end{aligned}$$

The fact that $E_{M_l}[J_l - J_{l-1}]$ are independent variables for $l = 0, \ldots, L$ (as noted in the proof of Lemma 3) has been used in obtaining (92e). Equation (91) has been used to obtain (92h). Finally, inserting (90), (91), and (92h) into (89a)-(89c) and collecting terms yields the stated lemma. ■

### 4.1.3 Calculation of RMLMC parameters for the idealized model

The RMLMC formulation (83) introduces $L + 1$ undetermined sample size parameters, $\{M_0, M_1, \ldots, M_L\}$, that must be determined. Recall that, assuming the idealized model (82), the RMLMC estimator provides an unbiased estimate of the $\mathbb{E}[J]$ statistic. There is no longer a need to account for an approximation error bias that was present in the MLMC estimator. Consequently, when a RMLMC mean squared error (86) of $\epsilon^2$ or less is desired, this can be accomplished by requiring that

$$\sum_{l=0}^{L-1} \frac{V_l}{M_l} + \left(\frac{1}{1-\gamma^{-r}}\right)^2 \frac{V_L}{M_L} \leq \epsilon^2 \tag{93}$$

with $V_l \equiv \mathbb{V}\mathrm{ar}[J_l - J_{l-1}]$. Following a parameter optimization process similar to that given in Sect. 2.2.4 for the MLMC estimator, the mean squared error is minimized subject to a fixed cost (29). For brevity, define

$$\widehat{V}_l \equiv \begin{cases} V_l & \text{for } \ 0 \leq l < L \\ \left(\frac{1}{1-\gamma^{-r}}\right)^2 V_l & \text{for } l = L \end{cases} \tag{94}$$

The optimization process yields sample size parameters (rounded to the nearest integer in implementations)

$$M_l = \sqrt{\frac{\widehat{V}_l}{\lambda_L \left(C_l + C_{l-1}\right)}} \tag{95}$$

with $\lambda_L$ chosen to satisfy (93)

$$\lambda_L = \left(\frac{\epsilon^2}{\sum_{l'=0}^{L} \sqrt{\widehat{V}_{l'}(C_{l'} + C_{l'-1})}}\right)^2 \tag{96}$$

Using this $\lambda_L$, the RMLMC cost (29) then simplifies to

$$Cost = \frac{\epsilon^2}{\lambda_L} \tag{97}$$

**Remark 4** *Due to the absence of an approximation error bias, there is no longer an explicit condition for selecting the maximum levels parameter $L$. A heuristic requirement, motivated by the approximation error bias condition (27) from the MLMC estimator, is*

$$\mathbb{E}^2[J - J_L] \leq \frac{\epsilon^2}{2} \tag{98}$$

*This additional requirement has no impact on the satisfaction (93), but affects the maximum number of levels and hence the cost (97) of the RMLMC estimator. The expectation quantity (98) can be estimated using*

$$\mathbb{E}[J - J_L] \approx \frac{1}{\gamma^r - 1} E_{M_L}[J_L - J_{L-1}] \tag{99}$$

*since from Lemma 1 and (81)*

$$\frac{1}{\gamma^r - 1}\mathbb{E}\left[E_{M_L}[J_L - J_{L-1}]\right] = \mathbb{E}[J - J_L] \tag{100}$$

**Remark 5** *Assuming that satisfaction of (27) and (98) results in both the MLMC and RMLMC estimators choosing the same maximum levels parameter $L$, the RMLMC estimator exhibits an overall cost savings when compared to the MLMC estimator. For example, in the limit of increasing rate parameter $r$ and identical mean squared error $\epsilon^2$*

$$\lim_{r\to\infty} \frac{Cost_{RMLMC}}{Cost_{MLMC}} = \frac{1}{2}$$

*This represents a considerable cost savings.*

#### 4.1.4 RMLMC rate estimation

Using the idealized model (82) with given enrichment factor $\gamma$, the rate parameter $r$ can be calculated from

$$\gamma^r = \frac{\mathbb{E}[J_{L-1} - J_{L-2}]}{\mathbb{E}[J_L - J_{L-1}]} \tag{101}$$

which motivates the approximate rate estimator

$$\gamma^{\widehat{r}} = \frac{E_{M_{L-1}}[J_{L-1} - J_{L-2}]}{E_{M_L}[J_L - J_{L-1}]} \tag{102}$$

where $\widehat{r}$ approximates the true rate $r$. Using the approximated rate, the RMLMC estimator can then be applied to problems that are well-represented by a discretization error depending on a discretization parameter $\Delta_0$ and rate parameter $r$, but do not have an explicit idealized model representation. The RMLMC implementation described in Sect. 4.1.5 uses this rate estimator. Unfortunately, rates estimated from (102) exhibit statistical variability resulting from the i.i.d. sampling. This degrades the accuracy of the RMLMC estimator as demonstrated in the numerical examples of Sect. 5. This variability can be greatly reduced or completely eliminated when sample reuse (see Sect. 3) is employed as discussed in Sect. 4.2.

#### 4.1.5 RMLMC implementation

A bootstrap RMLMC implementation starts from a single level $L = 0$ and incrementally adds levels until (98) and (93) are satisfied:

1. Initialize parameters, $L = 0$ and $\widehat{r} \gg 1.0$.
2. Estimate $\mathbb{V}\mathrm{ar}[J_L - J_{L-1}]$ using a variance estimator given an initial representative population of samples for $J_L - J_{L-1}$
3. If $L > 1$, calculate the estimated rate parameter $\widehat{r}$, e.g., $\gamma^{\widehat{r}} = \frac{E_{M_{L-1}}[J_{L-1}-J_{L-2}]}{E_{M_L}[J_L-J_{L-1}]}$.
4. Calculate $\lambda_L$ in (96) using previously estimated variances,
5. Calculate sample sizes $\{M_0, M_1, \ldots, M_L\}$ using (95) and enrich level sample populations using these revised sample sizes,
6. If $L > 1$, check the approximation error requirement (98), $E^2[J - J_L] \le \frac{\epsilon^2}{2}$, by checking whether $\frac{1}{\gamma^{\widehat{r}}-1}|E_{M_L}[J_L - J_{L-1}]| \le \frac{\epsilon}{\sqrt{2}}$. If not satisfied, set $L = L + 1$ and go to 2.

### 4.2 Unbiased RMLMC-SR estimation of $\mathbb{E}[J]$ for the idealized model

The RMLMC estimator provides unbiased estimates of the $\mathbb{E}[J]$ statistic but requires an estimate of the rate parameter $r$. Rate parameter estimates for the RMLMC estimator that use level data for this estimate suffer from statistical variability. This reduces accuracy of the estimator. The RMLMC-SR estimator described below reduces or eliminates this estimated rate variability by exploiting properties of sample reuse in the MLMC-SR estimator described in Sect. 3.

**Definition 5 (RMLMC-SR Estimator for the Idealized Model)** *Let $J(u(x,t,\omega))$ denote a random variable quantity of interest and $J_l \equiv J(u_l(x,t,\omega))$ an idealized quantity of interest model with level-dependent discretization model*

$$J_l \equiv J(u_l(x,t,\omega)) = p(x,t,\omega) + (\gamma^{-l}\Delta_0)^r q(x,t,\omega) \quad , \quad l = 0,\ldots,L \qquad (103)$$

*with mesh resolution $\Delta_0$, enrichment factor $\gamma$, rate parameter $r$, a maximum level parameter $L$, and random variable functions $p(x,t,\omega)$ and $q(x,t,\omega)$.*

*Further, let* $E^L_{l^*}[J_L]$ *denote the MLMC-SR estimator described in Def. 3 and assume* $0 < l^* < L$. *The unbiased RMLMC-SR estimator for the* $\mathbb{E}[J]$ *statistic utilizing sample reuse is*

$$RE^L_{l^*}[J_L] = E^L_{l^*}[J_L] + \frac{1}{\gamma^r - 1} E^*_{M_L}[J_L - J_{L-1}] \tag{104}$$

*where* $E^*_{M_L}[J_L - J_{L-1}]$ *is described in Def. 3.*

#### 4.2.1 RMLMC-SR expected value for the idealized model

The next lemma verifies that the RMLMC-SR estimator provides an unbiased estimate of $\mathbb{E}[J]$ for the idealized quantity of interest model.

**Lemma 11 (RMLMC-SR Expected Value)** *Let* $RE^L_{i^*}[J_L]$ *denote the RMLMC-SR estimator described in Def.(5).* $RE^L_{l^*}[J_L]$ *is a random variable with expected value* $\mathbb{E}[J]$,

$$\mathbb{E}[RE^L_{l^*}[J_L]] = \mathbb{E}[J] \tag{105}$$

**Proof:** The proof follows very closely the proof of Lemma 9 and is unaffected by sample reuse. ■

#### 4.2.2 RMLMC-SR mean squared error for $\mathbb{E}[J]$ for the idealized model

The next lemma shows that the mean squared error for the RMLMC-SR estimator differs from the MLMC-SR mean squared error only at the finest level $L$.

**Lemma 12 (RMLMC-SR Mean Square Error for $\mathbb{E}[J]$)** *Let* $RE^L_{l^*}[J_L]$ *denote the RMLMC-SR estimator described in Def. (5). Further, assume that all models* $J_l \equiv J(u_l(x,t,\omega)), l = 0,\dots,L$ *have bounded variance. The unbiased mean squared error for the RMLMC-SR estimator with respect to the* $\mathbb{E}[J]$ *statistic is given by*

$$\mathbb{E}\left[\left(\mathbb{E}[J] - RE^L_{l^*}[J_L]\right)^2\right] = \sum_{l=0}^{L-1} \frac{1}{M_l} \left( \mathbb{V}\mathrm{ar}[J_l - J_{l-1}] + 2 \sum_{\substack{l'=l+1 \\ l \geq l^*}}^{L} \mathbb{C}\mathrm{ov}[J_l - J_{l-1}, J_{l'} - J_{l'-1}] \right)$$

$$+\left(\frac{1}{1-\gamma^{-r}}\right)\frac{\mathbb{V}\mathrm{ar}[J_L - J_{L-1}]}{M_L} \tag{106}$$

**Proof:** The proof follows very closely the proof of Lemma 10 and is omitted. ■

### 4.2.3 Calculation of RMLMC-SR parameters for the idealized model

The RMLMC-SR formulation (104) introduces $L+1$ undetermined sample size parameters, $\{M_0, M_1, \ldots, M_L\}$, that must be determined. Recall that, assuming the idealized model (103), the RMLMC-SR estimator provides an unbiased estimate of the $\mathbb{E}[J]$ statistic. There is no longer a need to account for an approximation error bias that was present in the MLMC-SR estimator. Consequently, when a RMLMC-SR mean squared error (106) of $\epsilon^2$ or less is desired, this can be accomplished by requiring that

$$\sum_{l=0}^{L-1}\frac{1}{M_l}\left(V_l + 2\sum_{\substack{l'=l+1\\ l\geq l^*}} C_{l\,l'}\right) + \left(\frac{1}{1-\gamma^{-r}}\right)^2\frac{V_L}{M_L} \leq \epsilon^2 \tag{107}$$

with $V_l \equiv \mathbb{V}\mathrm{ar}[J_l - J_{l-1}]$ and $C_{l\,l'} = \mathbb{C}\mathrm{ov}[J_l - J_{l-1}, J_{l'} - J_{l'-1}]$. Following an optimization process similar to that given in Sect. 3.2.2 for the MLMC-SR estimator, the left-hand-side of (107) is minimized subject to the fixed RMLMC-SR cost

$$Cost = \sum_{l=0}^{l^*} M_l(C_l + C_{l-1}) + \sum_{l=l^*+1}^{L} M_l\; C_l \tag{108}$$

For brevity, define

$$\widehat{V}_l \equiv \begin{cases} V_l + 2\sum_{\substack{l'=l+1\\ l\geq l^*}} C_{l\,l'} & \text{for}\quad 0\leq l < L \\ \left(\frac{1}{1-\gamma^{-r}}\right)^2 V_l & \text{for } l = L \end{cases} \tag{109}$$

The optimization process yields sample size parameters (rounded to the nearest integer in implementations)

$$M_l \approx \begin{cases} \sqrt{\frac{\widehat{V}_l}{\lambda_L\,(C_l + C_{l-1})}} & \text{for } 0 \le l \le l^* \\ \sqrt{\frac{\widehat{V}_l}{\lambda_L\,C_l}} & \text{for } l^* < l \le L \end{cases} \tag{110}$$

with $\lambda_L$ chosen to satisfy (107)

$$\lambda_L = \left( \frac{\epsilon^2}{\sum_{l'=0}^{l^*} \sqrt{\widehat{V}_{l'}(C_{l'} + C_{l'+1})} + \sum_{l'=l^*+1}^{L} \sqrt{\widehat{V}_{l'} C_{l'}}} \right)^2 \tag{111}$$

The RMLMC-SR cost (108) then simplifies to

$$Cost = \frac{\epsilon^2}{\lambda_L} \tag{112}$$

**Remark 6** *Using the same arguments made in Sect. 4.1.3, an explicit condition for selecting the maximum levels parameter $L$ is given by*

$$\mathbb{E}^2[J - J_L] \le \frac{\epsilon^2}{2} \tag{113}$$

*This additional requirement has no impact on the satisfaction (107), but affects the maximum number of levels and hence the cost (108) of the RMLMC-SR estimator. The expectation quantity (113) can be estimated using*

$$\mathbb{E}[J - J_L] \approx \frac{1}{\gamma^r - 1} E^*_{M_L}[J_L - J_{L-1}] \tag{114}$$

**Remark 7** *Assuming that satisfaction of (62) and (113) results in both the MLMC-SR and RMLMC-SR estimators choosing the same maximum levels parameter $L$, the RMLMC-SR estimator exhibits an overall cost savings when compared to the MLMC estimator. For example, in the limit of increasing rate parameter $r$ and identical mean squared error $\epsilon^2$*

$$\lim_{r\to\infty} \frac{Cost_{RMLMC-SR}}{Cost_{MLMC-SR}} = \frac{1}{2}$$

*This represents a considerable cost savings.*

#### 4.2.4 RMLMC-SR rate estimation

Sample reuse in the RMLMC-SR estimator with $L - l^* \geq 2$ permits the construction of two related rate estimators. Lemma 13 then proves that these rate estimators exactly reproduce the rate $r$ when the idealized quantity of interest model (103) is assumed. The Type I rate estimator uses level data from the RMLMC-SR estimator.

**Definition 6 (Type I Rate Estimator)** *Let $J(u(x,t,\omega))$ denote a random variable quantity of interest, $L$ a maximum levels parameter, and $J^L$ a set containing approximate models of $J(u(x,t,\omega))$*

$$J^L = \{J_0, J_1, \ldots, J_L\} \ , \tag{115}$$

*where the shorthand notation $J_l = J(u_l(x,t,\omega))$ has been used. Further, let $RE_{l^*}^L[J_L]$ denote the RMLMC-SR estimator described in Def. 5 with $L - l^* \geq 2$. Let $\widehat{r}$ denote an estimated rate parameter for the RMLMC-SR estimator. The Type I rate estimator for $\widehat{r}$ is*

$$\gamma^{\widehat{r}} \equiv \frac{E^*_{M_L}[J_{L-1} - J_{L-2}]}{E^*_{M_L}[J_L - J_{L-1}]} \tag{116}$$

*where $E^*_{M_L}[J_L - J_{L-1}]$ is described in Def. 3 and $E^*_{M_L}[J_{L-1} - J_{L-2}]$ is a truncation of $E^{l^*}_{M_{L-1}}[J_{L-1} - J_{L-2}]$ for $M_L < M_{L-1}$ assumed in Def. 3.*

The Type II rate estimator estimates contributions to the rate $\widehat{r}$ on a sample-by-sample basis.

**Definition 7 (Type II Rate Estimator)** *Let $J(u(x,t,\omega))$ denote a random variable quantity of interest, $L$ a maximum levels parameter, and $J^L$ a set containing approximate models of $J(u(x,t,\omega))$*

$$J^L = \{J_0, J_1, \ldots, J_L\} \ , \tag{117}$$

*where the shorthand notation $J_l = J(u_l(x,t,\omega))$ has been used. Further, let $RE_{l^*}^L[J_L]$ denote the RMLMC-SR estimator described in Def. 5 with $L - l^* \geq 2$. Let $\widehat{r}$ denote an estimated rate parameter for the RMLMC-SR estimator. The Type II rate estimator for $\widehat{r}$ is*

$$\gamma^{\widehat{r}} \equiv E_{M_L}\left[\frac{J_{L-1}(\omega_*) - J_{L-2}(\omega_*)}{J_L(\omega_*) - J_{L-1}(\omega_*)}\right] = \frac{1}{M_L}\sum_{i=1}^{M_L} \frac{J_{L-1}(\omega_*^{(i)}) - J_{L-2}(\omega_*^{(i)})}{J_L(\omega_*^{(i)}) - J_{L-1}(\omega_*^{(i)})} \tag{118}$$

The next lemma shows that the Type I and Type II rate estimators result in an exact (deterministic) rate calculation when the idealized quantity of interest (103) is assumed.

**Lemma 13 (Type I and Type II determistic rate estimation)** *The RMLMC-SR Type I and Type II rate estimators described in Def. 6 and Def. 7 yield deterministic rate estimates equal to $\gamma^r$ when applied to the idealized quantity of interest function (103).*

**Proof:** Both rate estimators exploit sample reuse in RMLMC-SR estimation for $L - l^* \geq 2$. Starting with the Type I estimator

$$\begin{aligned} \widehat{\gamma^r} &\equiv \frac{E^*_{M_L}[J_{L-1} - J_{L-2}]}{E^*_{M_L}[J_L - J_{L-1}]} && \text{(119a)} \\ &= \frac{\sum_{i=1}^{M_L} J_{L-1}(\omega_*^{(i)}) - J_{L-2}(\omega_*^{(i)})}{\sum_{j=1}^{M_L} J_L(\omega_*^{(j)}) - J_{L-1}(\omega_*^{(jb)})} && \text{(119b)} \\ &= \frac{\gamma^r(1-\gamma^r)\gamma^{-l\ r}\Delta_0^r \frac{1}{M_L}\sum_{i=1}^{M_L} q(x,t,\omega_*^{(i)})}{(1-\gamma^r)\gamma^{-l\ r}\Delta_0^r \frac{1}{M_L}\sum_{j=1}^{M_L} q(x,t,\omega_*^{(j)})} && \text{(119c)} \\ &= \gamma^r && \text{(119d)} \end{aligned}$$

Inserting the idealized quantity of interest function (103) into (119b) yields (119c). Exact cancellation of identical random variable sums yields a exact (deterministic) rate calculation.

The Type II estimator directly exploits sample reuse and the assumed idealized quantity of interest function (103) on a sample-by-sample basis, i.e,

$$\gamma^r = \frac{J_{L-1}(\omega_*^{(i)}) - J_{L-2}(\omega_*^{(i)})}{J_L(\omega_*^{(i)}) - J_{L-1}(\omega_*^{(i)})} \ ,\ i = 1, \ldots, M_L \qquad (120)$$

Inserting into the Type II formula

$$\begin{aligned} \widehat{\gamma^r} &\equiv \frac{1}{M_L}\sum_{i=1}^{M_L} \frac{J_{L-1}(\omega_*^{(i)}) - J_{L-2}(\omega_*^{(i)})}{J_L(\omega_*^{(i)}) - J_{L-1}(\omega_*^{(i)})} && \text{(121a)} \\ &= \frac{1}{M_L}\sum_{i=1}^{M_L} \gamma^r && \text{(121b)} \\ &= \gamma^r && \text{(121c)} \end{aligned}$$

This verifies the stated lemma. ■

For general quantity of interest functions, statistical variance in both Type I and Type II rate estimators is expected. This is examined further in the numerical results Sect. 5.

#### 4.2.5 RMLMC-SR implementation

The implementation starts from a single level $L = 0$ and incrementally adds levels until (113), (107), and $L - l^* \geq 2$ are satisfied. New i.i.d. outcome event samples are generated if the current level $L < l^*$. Otherwise, an i.i.d. event outcome sample reuse set is established when $L = l^*$ and those event outcome samples are reused for $L \geq l^*$.

1. Initialize parameters, $L = 0$ and $\widehat{r} \gg 1.0$.

2. If $L = l^*$, establish a event outcome sample reuse population.

3. Estimate $\mathbb{V}\text{ar}[J_L - J_{L-1}]$ given an initial population of new i.i.d. samples if $L < l^*$ or reused samples if $L \geq l^*$

4. Estimate $\mathbb{C}\text{ov}[J_l - J_{l-1}, J_L - J_{L-1}]$ for $l^* \leq l < L - 1$.

5. If $L - l^* \geq 2$, calculate the estimated rate parameter $\widehat{r}$ using either Type I or Type II formulas, e.g., $\gamma^{\widehat{r}} = \frac{E^*_{M_L}[J_{L-1} - J_{L-2}]}{E^*_{M_L}[J_L - J_{L-1}]}$.

6. Calculate $\lambda_L$ in (111) using previously estimated variances and covariances.

7. Calculate sample sizes $\{M_0, M_1, \ldots, M_L\}$ using (110) and enrich level sample populations using these revised sample sizes,

8. If $L - l^* \geq 2$, check the approximation error requirement (113), $E^2[J - J_L] \leq \frac{\epsilon^2}{2}$, by checking whether $\frac{1}{\gamma^{\widehat{r}} - 1}|E_{M_L}[J_L - J_{L-1}]| \leq \frac{\epsilon}{\sqrt{2}}$. If not satisfied, set $L = L + 1$ and go to 2.

# 5 Numerical Examples

Example calculations are presented for random variable quantities of interest that are the asymptotic limit of approximate quantities of interest described by a level parameter. Two multilevel Monte Carlo estimators are then evaluated for these problems: the unbiased RMLMC estimator (83) with rate estimator (102) and the unbiased RMLMC-SR estimator (104) with Type I rate estimator (116) and sample reuse lower limit $l^* = 1$. For both examples, an enrichment factor of $\gamma = 2$ has been used.

The first example problem assumes the idealized quantity of interest form (5) with 10-dimensional random variable functions. Although a specific value of the rate parameter $r$ is used in calculating quantity of interest samples, the RMLMC and RMLMC-SR estimators are unaware of this chosen value and must estimate the rate from sample data using a rate estimator.

The second example problem utilizes a Gaussian process random field, $W(x, \omega)$, approximated by a truncated 10-term Karhunen-Loève expansion. The quantity of interest is $\|W^2(\cdot, \omega)\|_2^2$ with discretization error introduced by replacing the $L^2$ norm integral with a numerical quadrature. The discretization error for this quantity of interest decreases at a second order rate when the number of quadrature points is doubled between successive levels of the model. The RMLMC and RMLMC-SR estimators are unaware of this second order rate and must estimate the rate from the sample data using a rate estimator.

For each example, the following comparisons are made

1. Cost of the RMLMC and RMLMC-SR estimators (and others).

2. Variance of the RMLMC and RMLMC-SR estimated rate $\widehat{r}$. Comparison results, presented next, show degraded performance when the estimated rates have significant variance.

3. Variance of the RMLMC and RMLMC-SR approximation error bias estimators, $\frac{1}{2^{\widehat{r}}-1} E_{M_L}[J_L - J_{L-1}]$ and $\frac{1}{2^{\widehat{r}}-1} E^*_{M_L}[J_L - J_{L-1}]$, with $\widehat{r}$ calculated using a rate estimator. The expected value of these estimators is the expected approximation error bias, i.e., for the RMLMC approximation error bias estimator

$$\mathbb{E}[J - J_L] = \mathbb{E}\left[\frac{1}{2^{\widehat{r}} - 1} E_{M_L}[J_L - J_{L-1}]\right] \tag{122}$$

This result comes from rearrangement the RMLMC estimator

$$RE^L[J_L] - E^L[J_L] = \frac{1}{2^r - 1} E_{M_L}[J_L - J_{L-1}] \tag{123}$$

and application of Lemmas 9 and 11

$$\mathbb{E}[RE^L[J_L] - E^L[J_L]] = \mathbb{E}[J - J_L] = \mathbb{E}\left[\frac{1}{2^r - 1} E_{M_L}[J_L - J_{L-1}]\right] \quad (124)$$

Similar results are obtained for the RMLMC-SR estimator. In RMLMC and RLMLC-SR implementations, the rate $r$ in (124) is replaced by a rate estimator value $\widehat{r}$ which can significantly affect the the variance in the approximation error bias estimators.

4. Variance of the RMLMC and RMLMC-SR estimators.

## 5.1 Example 1

This example assumes the following idealized quantity of interest

$$J_l \equiv J(\omega; \Delta_0, r, l, L, a, N) = p(\omega; a, N) + C_{disc}\ (2^{-l}\Delta_0)^r q(\omega; a, N) \quad (125)$$

with mesh resolution $\Delta_0$, rate parameter $r$, level parameter $l$, maximum level parameter $L$, numerical discretization scaling constant $C_{disc}$, and random variable functions

$$p(\omega; a, N) = \prod_{i=1}^{N} \left(a \sin^2(\pi\omega_i) + 1 - \frac{a}{2}\right) \quad (126)$$

and

$$q(\omega; a, N) = \prod_{i=1}^{N} \left(a \cos^2(\pi\omega_i) + 1 - \frac{a}{2}\right) \quad (127)$$

with variance parameter $a$ and number of random variable dimensions $N$. The event outcome $\omega$ is an $N$-tuple, $[\omega_1, \ldots, \omega_N]$, with each $\omega_i$ uniformly distributed in the interval $[0, 1]$. Both random functions have unit expected value

$$\mathbb{E}[p] = \mathbb{E}[q] = 1 \quad (128)$$

and variance

$$\mathbb{V}\text{ar}[p] = \mathbb{V}\text{ar}[q] = \left(1 + \frac{a^2}{8}\right)^N - 1 \quad (129)$$

For a given number of random variable dimensions $N$, choosing a variance for $p$ determines the value $a$ and similarly for $q$. In the example calculations,

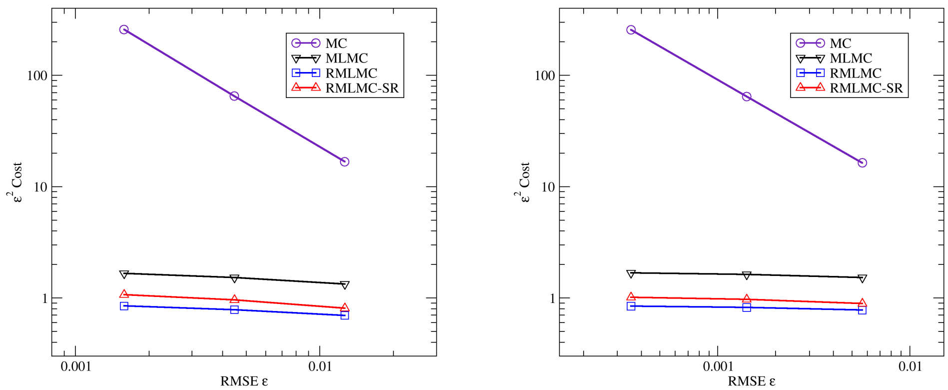


Figure 1: Example 1 comparison of $\epsilon^2\, Cost$ versus mean squared errors $\epsilon$ for the MC, MLMC, RMLMC, RMLMC-SR estimators. Shown are calculations for the output quantity of interest (125) with specified rates $r = 1.5$ (left) and $r = 2$ (right)

$\mathbb{V}\mathrm{ar}[p] = \frac{1}{4}$ and $\mathbb{V}\mathrm{ar}[q] = \frac{1}{4}$ have been chosen. Finally, choosing a level of discretization error, $C_{disc} = \frac{1}{10}$, and the number of random variable dimensions, $N = 10$, completes the model description.

Figure 1 compares the total cost (scaled by mean squared error $\epsilon^2$) associated with the MLMC, RMLMC, and RMLMC-SR estimators for 3 values of the root mean squared error corresponding to levels $L = 3, 4$, and 5 in the multilevel methods. For reference, a theoretical curve is also shown for the single level Monte Carlo estimator. Calculations shown in Figure 1 use quantity of interest samples with specified discretization rate parameter values of $r = \frac{3}{2}$ (left figure) and $r = 2$ (right figure). These graphs show the enormous cost benefit of the MLMC, RMLMC, and RMLMC-SR multilevel estimators when compared to the single level Monte Carlo estimator. These graphs also show a small increase in cost for the RMLMC-SR estimator when compared to the RMLMC estimator due to sample reuse and a significant overall decrease in cost when compared to the MLMC estimator. This latter decrease can be attributed to the absence of an approximation error bias contribution in the RMLMC and RMLMC-SR sample size calculation. This can be further understood by comparing Eqns. (28) and (93).

The graphs shown in Fig. 2 compare variances of the RMLMC and

RMLMC-SR estimated rates $\widehat{r}$. These graphs show a significant reduction in the RMLMC-SR rate estimator variance when compared to the RMLMC rate estimator. This improvement is completely expected since the RMLMC-SR

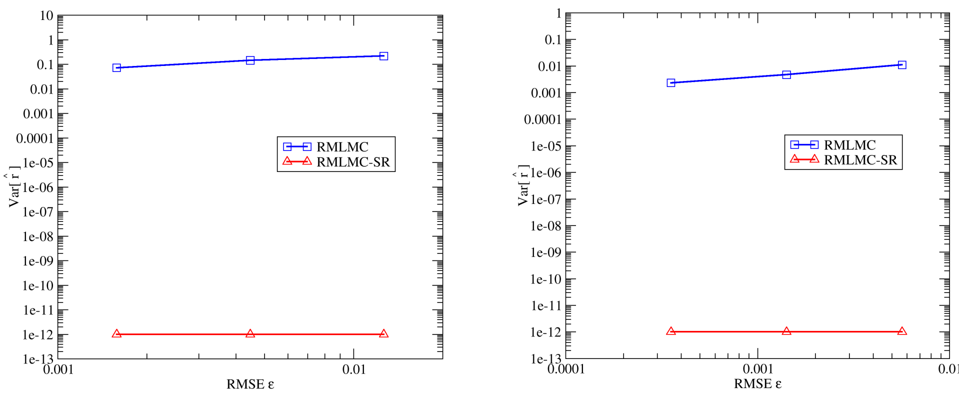


Figure 2: Example 1 comparison of $\mathbb{V}\text{ar}[\widehat{r}]$ versus mean squared errors $\epsilon$ using the RMLMC estimator with rate estimator (102) and the RMLMC-SR estimator with rate estimator (116), respectively. Shown are calculations for the output quantity of interest (125) with specified rates $r = 1.5$ (left) and $r = 2$ (right)

estimator and rate estimator formula were tailored to this form of idealized quantity of interest.

The Fig. 3 graphs compare variances in the RMLMC and RMLMC-SR approximation error bias estimators, $\frac{1}{2^{\widehat{r}}-1}E_{M_L}[J_L - J_{L-1}]$ and $\frac{1}{2^{\widehat{r}}-1}E^*_{M_L}[J_L - J_{L-1}]$. The expected value of these estimators is the expected approximation error bias, $\mathbb{E}[J - J_L]$. The graphs show a significant reduction in the variance of the RMLMC-SR approximation error bias estimator when compared to the RMLMC estimator. This can be directly attributed to the low variance RMLMC-SR rate estimation seen in Fig. 2.

Figure 4 compares the RMLMC and RMLMC-SR estimator variances. The graphs show some small overall reduction in variance of the RMLMC-SR estimator when compared to the RMLMC estimator. Further examination reveals that the amount of reduction decreases with decreasing root mean squared error $\epsilon$. The RMLMC-SR estimator variance reduction is substantially less than variance reduction observed for the approximation error bias

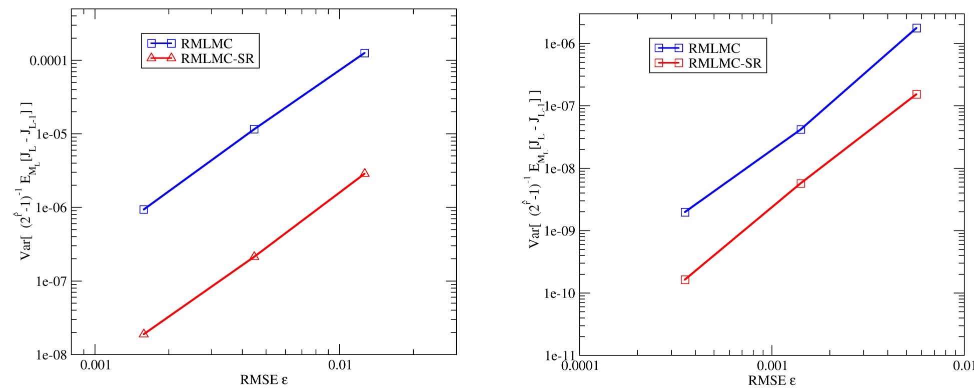


Figure 3: Example 1 comparison of $\mathbb{V}\text{ar}[\frac{1}{2^{\widehat{r}}-1}E_{M_L}[J_L - J_{L-1}]]$ and $\mathbb{V}\text{ar}[\frac{1}{2^{\widehat{r}}-1}E^*_{M_L}[J_L - J_{L-1}]]$ versus mean squared errors $\epsilon$ using the RMLMC estimator with rate estimator (102) and the RMLMC-SR estimator with rate estimator (116), respectively. Shown are calculations for the output quantity of interest (125) with specified rates $r = 1.5$ (left) and $r = 2$ (right)

estimator shown in Fig. 3. These observations are expected, since for the RMLMC-SR estimator (similarly for the RMLMC estimator)

$$RE^L_{l^*}[J_L] = E^L_{l^*}[J_L] + \underbrace{\frac{1}{2^{\widehat{r}}-1}E^{l^*}_{M_L}[J_L - J_{L-1}]}_{\text{approximation error bias}} \tag{130}$$

so the contribution of an improved rate estimate $\widehat{r}$ is confined to the right-hand-side correction term, which is small because the approximation error bias is small and gets smaller with increasing $\widehat{r}$.

## 5.2 Example 2

This example utilizes a truncated Karhunen-Loève (KL) expansion that represents a Gaussian process random field with covariance function

$$K_X(x_1, x_2) = e^{-|x_1 - x_2|/b} \tag{131}$$

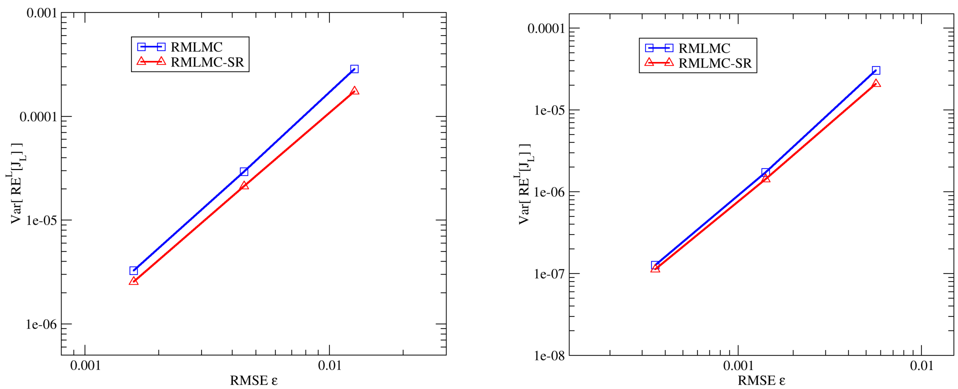


Figure 4: Example 1 comparison of $\mathbb{V}\text{ar}[RE^L[J_L]]$ and $\mathbb{V}\text{ar}[RE^L_{l^*}[J_L]]$ versus mean squared errors $\epsilon$ using the RMLMC estimator with rate estimator (102) and the RMLMC-SR estimator with rate estimator (116), respectively. Shown are calculations for the output quantity of interest (125) with specified rates $r = 1.5$ (left) and $r = 2$ (right)

where $b$ is a correlation length. The KL expansion takes the following form (see Ghamen and Spanos [4] for details)

$$W(x,\omega) = \sum_{n=1}^{\infty} \left( Z_n(\omega)\sqrt{\lambda_n} f_n(x) + Z_n^*(\omega)\sqrt{\lambda_n^*} f_n^*(x) \right) \tag{132}$$

where $(\lambda_n, f_n(x))$ and $(\lambda_n^*, f_n^*(x))$ are eigenpairs associated with the integral equation

$$\int_{-a}^{a} K_X(x_1,x_2) f_n(x_1)\, d\, x_1 = \lambda_n f_n(x_2) \tag{133}$$

A Gaussian process is simulated by choosing statistically independent random variables $Z_n$ and $Z_n^*$ with joint Gaussian distribution and unit variance. For convenience, let $c = 1/b$. For the specific choice of covariance function (131), the KL expansion eigenvalues

$$\lambda_n = \frac{2c}{\xi_n^2 + c^2} \quad , \quad \lambda_n^* = \frac{2c}{{\xi_n^*}^2 + c^2} \quad , \quad n = 1, \ldots, N \tag{134}$$

and the eigenfunctions

$$f_n(x) = \frac{\cos(\xi_n x)}{\sqrt{a + \frac{\sin(2\xi_n a)}{2\xi_n}}} \quad , \quad f_n^*(x) = \frac{\sin(\xi_n^* x)}{\sqrt{a - \frac{\sin(2\xi_n^* a)}{2\xi_n^*}}} \quad , \quad n = 1, \ldots, N \tag{135}$$

with $\xi_n$ and $\xi_n^*$ solutions of the transcendental equations

$$c - \xi \tan(\xi a) = 0 \quad \text{and} \quad \xi^* + c \tan(\xi^* a) = 0 \tag{136}$$

can be readily derived and computed [4].

The output quantity of interest chosen for this example is

$$J(\omega; K) = \|W^2(\cdot, \omega; K)\|^2_{L^2([-a,a])} = \int_{-a}^{a} W^2(x, \omega; K)\; d\,x \tag{137}$$

where $W(x, \omega; K)$ denotes a $K$-term truncation of (132). For this output quantity of interest, the exact expected value is known from the KL theory

$$\mathbb{E}[J(\omega; K)] = \sum_{n=1}^{K} \lambda_n + \lambda_n^* \tag{138}$$

Choosing $K = 5$ yields 10 random variable dimensions. A numerical discretization error is then introduced by replacing the integration in (137) with a $Q$-point numerical quadrature using a second order accurate midpoint rule. This results in the level-dependent final form

$$J_l(\omega; K) = \sum_{q=1}^{Q} w_q W^2(x_q, \omega; K) \;\; , \quad Q \equiv Q_0 2^l \tag{139}$$

where $w_q$ and $x_q$ denote quadrature point weights and locations. Finally, choosing $a = \frac{1}{2}$, $b = \frac{5}{100}$, and $Q_0 = 10$ completes a definition of the model.

The graphs shown in Fig. 5 compare $\epsilon^2 Cost$ (left) and variance of the estimated rate, $\mathbb{V}\text{ar}[\widehat{r}]$, (right) using the RMLMC estimator with rate estimator (102) and the RMLMC-SR estimator with rate estimator (116). Observe that the abscissa values of root mean squared error in this example are significantly smaller than those in Example 1. The cost comparisons in Fig. 5 (left) show insignificant cost differences between the RMLMC and RMLMC-SR estimators. This is due to small covariance contributions across all levels in the estimator for the (139) quantity of interest. Note that the RMLMC-SR rate estimator (116) is no longer exact given the sample data for this problem. Even so, the rate estimator variance shown in Fig. 5 (right) again shows a significant reduction in variance using the RMLMC-SR rate estimator.

The graphs shown in Fig. 6 compare approximation error bias estimator variances, $\mathbb{V}\text{ar}[\frac{1}{2^{\widehat{r}}-1} E_{M_L}[J_L - J_{L-1}]]$ and $\mathbb{V}\text{ar}[\frac{1}{2^{\widehat{r}}-1} E^*_{M_L}[J_L - J_{L-1}]]$, (left) as

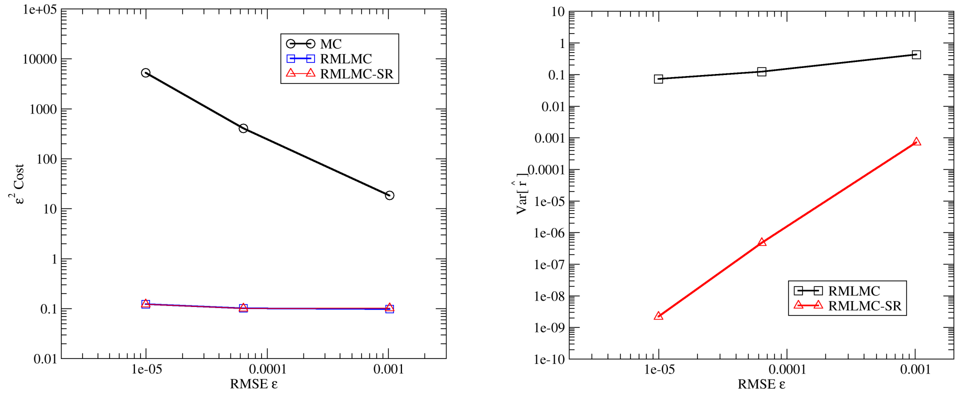


Figure 5: Example 2 comparison of $\epsilon^2\, Cost$ versus mean squared errors $\epsilon$ (left) and $\mathbb{V}\text{ar}[\widehat{r}]$ versus mean squared errors (right) using the RMLMC estimator with rate estimator (102) and the RMLMC-SR estimator with rate estimator (116), respectively.

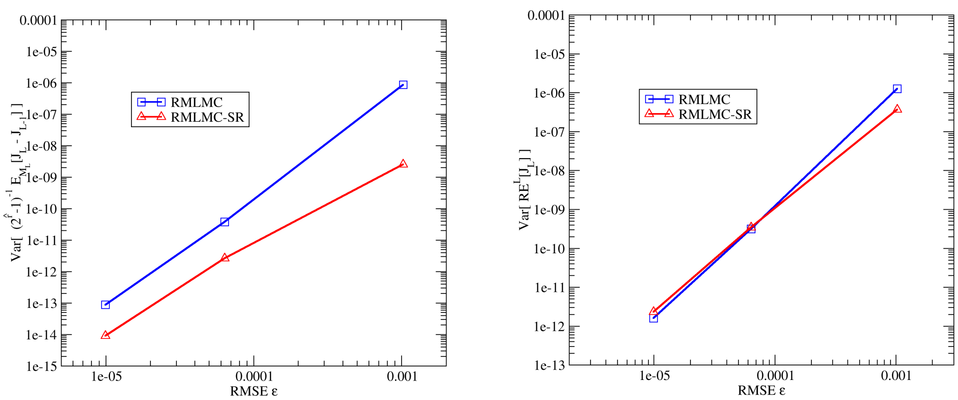


Figure 6: Example 2 comparison of $\mathbb{V}\text{ar}[\frac{1}{2^{\widehat{r}}-1} E_{M_L}[J_L - J_{L-1}]]$ and $\mathbb{V}\text{ar}[\frac{1}{2^{\widehat{r}}-1} E^*_{M_L}[J_L - J_{L-1}]]$ versus mean squared errors $\epsilon$ (left) as well as $\mathbb{V}\text{ar}[RE^L[J_L]]$ and $\mathbb{V}\text{ar}[RE^L_{l*}[J_L]]$ versus mean squared errors $\epsilon$ (right) using the RMLMC estimator with rate estimator (102) and the RMLMC-SR estimator with rate estimator (116), respectively.

well as estimator variances, $\mathbb{V}\mathrm{ar}[RE^L[J_L]]$ and $\mathbb{V}\mathrm{ar}[RE^L_{l^*}[J_L]]$, (right) corresponding to the RMLMC and RMLMC-SR estimators. Recall from (124) that $\mathbb{E}[\frac{1}{2^{\widehat{r}}-1}E_{M_L}[J_L - J_{L-1}]]$ and $\mathbb{E}[\frac{1}{2^{\widehat{r}}-1}E^*_{M_L}[J_L - J_{L-1}]]$ with estimated rate $\widehat{r}$ are estimates of the expected approximation error bias, $\mathbb{E}[J - J_L]$. The substantial reduction in variance using the RMLMC-SR estimator again implies that accurate estimates of $\mathbb{E}[J] - \mathbb{E}[J_L]$ can be obtained using fewer RMLMC-SR evaluations when compared to the RMLMC estimator. The graphs of estimator variance in Fig. 6 (right) show some small reduction in variance for larger values of the root mean squared error $\epsilon$ for the RMLMC-SR estimator when compared to the RMLMC estimator. This is again expected, using the same argument given in Example 1.

# 6 Summary

A modified form of the multilevel Monte Carlo estimator has been presented for the estimation of expectation statistics that utilizes sample reuse in specified levels, explicitly removes approximation error bias associated with numerically computed output quantities of interest that have an asymptotic limit behavior, and provides a low variance estimate of the asymptotic rate of convergence to that limit.

Numerical calculations demonstrate that the modified estimator is more cost efficient than the standard multilevel Monte Carlo estimator and is applicable to quantities of interest that have no explicit idealized quantity of interest form (5).

# References

[1] A. Brandt. Multi-level adaptive solutions to boundary value problems. *Math. Comp.*, 31:333–390, 1977.

[2] T. F. Chan, B. Smith, and J. Zou. Overlapping Schwarz methods on unstructured meshes using non-matching coarse grids. *Numer. Math.*, 73(2):149–167, 1996.

[3] K. Cliffe, M. Giles, R. Scheichl, and A. Techentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.

[4] R. Ghamen and P. Spanos. *Stochastic Finite Elements*. Dover Pub. Inc., Mineola, New York, 1991.

[5] M. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):981–986, 2008.

[6] W. Hackbusch. *Multi-Grid Methods and Applications*. Springer Pub, New York, 1985.

[7] D. Knuth. Two notes on notation. *Amer. Math. Monthly*, 99(5):403–422, 1992.

[8] S. Mishra and C. Schwab. Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws. Technical Report 2010-24, ETH Zurich, Zurich, 2010.

[9] S. Mishra and C. Schwab. Sparse tensor multi-level Monte Carlo finite volum methods for hyperbolic conservation laws with random initial data. *Math. Comp.*, 81:1979–2018, 2012.

[10] B. Peherstorfer, K. Willcox, and M. Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SISC*, 38(5):A3163–A2194, 2016.

[11] B. Smith, P. Bjorstad, and W. Gropp. *Domain decomposition: parallel multilevel methods for elliptic partial differential equations*. Cambridge University Press, 1996.

[12] J. Sukys. *Multi-level Monte Carlo Finite Volume Methods for Systems of Hyperbolic Conservation Laws With Random Input Data*. PhD thesis, ETH Zurich, Department of Applied Mathematics, 2014.