arXiv:2412.07723v4 [math.NA] 16 Jun 2025

MULTILEVEL RANDOMIZED QUASI-MONTE CARLO ESTIMATOR FOR NESTED INTEGRATION

ARVED BARTUSKA#,1, ANDRÉ GUSTAVO CARLON[1], LUIS ESPATH[3], SEBASTIAN KRUMSCHEID[5], & RAÚL TEMPONE[1,2,4]

Abstract. Nested integration problems arise in various scientific and engineering applications, including Bayesian experimental design, financial risk assessment, and uncertainty quantification. These nested integrals take the form $\int f\left(\int g(\boldsymbol{y},\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right)\,\mathrm{d}\boldsymbol{y}$, for nonlinear f, making them computationally challenging, particularly in high-dimensional settings. Although widely used for single integrals, traditional Monte Carlo (MC) methods can be inefficient when encountering complexities of nested integration. This work introduces a novel multilevel estimator, combining deterministic and randomized quasi-MC (rQMC) methods to handle nested integration problems efficiently. In this context, the inner number of samples and the discretization accuracy of the inner integrand evaluation constitute the level. We provide a comprehensive theoretical analysis of the estimator, deriving error bounds demonstrating significant reductions in bias and variance compared with standard methods. The proposed estimator is particularly effective in scenarios where the integrand is evaluated approximately, as it adapts to different levels of resolution without compromising precision. We verify the performance of our method via numerical experiments, focusing on estimating the expected information gain of experiments. When applied to Gaussian noise in the experiment, a truncation scheme ensures finite error bounds at the same computational complexity as in the bounded noise case up to multiplicative logarithmic terms. The results reveal that the proposed multilevel rQMC estimator outperforms existing MC and rQMC approaches, offering a substantial reduction in computational costs and offering a powerful tool for practitioners dealing with complex, nested integration problems across various domains.

Keywords: Nested integration · randomized quasi-Monte Carlo · multilevel quasi-Monte Carlo · expected information gain

AMS subject classifications: · 62F15 · 65C05 · 65D30 · 65D32 ·

Acknowledgments

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-2019-CRG8-4033, the Alexander von Humboldt Foundation, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 333849990/GRK2379 (IRTG Hierarchical and Hybrid Approaches in Modern Inverse Problems), and was partially supported by the Flexible Interdisciplinary Research Collaboration Fund at the University of Nottingham Project ID 7466664. The authors thank Prof. Abdul-Lateef Haji-Ali for his helpful and constructive comments.

Contents

Acknowledgments 1
1. Introduction 2
2. Overview of numerical integration methods 4

[1]Department of Mathematics, RWTH Aachen University, Gebäude-1953 1.OG, Pontdriesch 14-16, 161, 52062 Aachen, Germany

[2]King Abdullah University of Science & Technology (KAUST), Computer, Electrical and Mathematical Sciences & Engineering Division (CEMSE), Thuwal 23955-6900, Saudi Arabia

[3]School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, United Kingdom

[4]Alexander von Humboldt Professor in Mathematics for Uncertainty Quantification, RWTH Aachen University, Germany

[5]Scientific Computing Center, and Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology, Germany

E-mail address: #arved.bartuska@gmail.com.

Date: June 17, 2025.

2.1. Monte Carlo method 4
2.2. Quasi-Monte Carlo method 4
2.3. Randomized Quasi-Monte Carlo method 5
2.4. Double-loop Monte Carlo method for nested integrals 7
2.5. Double-loop randomized quasi-Monte Carlo method for nested integrals 8
2.6. Multilevel double-loop Monte Carlo method for nested integrals 8
2.7. Inexact sampling 9
3. Multilevel double-loop randomized quasi-Monte Carlo method for nested integrals with inexact sampling 9
4. Expected information gain estimation 24
5. Numerical results 29
5.1. Example 1: linear EIG example with exact sampling 29
5.2. Example 2: Poisson EIG example with inexact sampling 31
5.3. Example 3: Duffing oscillator EIG example with inexact sampling 36
6. Conclusion 40
7. Statements and Declarations 40
7.1. Conflict of interest 40
7.2. Data availability 40
Appendix A. Auxiliary results for the error bounds to estimate the EIG 40
Appendix B. Proof of Lemma 1 43
Appendix C. Proof of Lemma 3 44
References 49

1. Introduction

In this work, we address the computational challenges of nested integration problems of the form $\int f\left(\int g(\boldsymbol{y},\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right)\,\mathrm{d}\boldsymbol{y}$ where f is a nonlinear function. Such problems frequently arise in engineering, mathematical finance, and medical decision-making. Traditional quadrature methods are constrained by the curse of dimensionality, making them impractical as the dimension increases. In contrast, Monte Carlo (MC) methods offer a more scalable solution but may become less efficient when managing complex nested structures. Our goal is to improve the efficiency of these estimators by developing particularly tailored, hierarchical quasi-MC (QMC) techniques.

The MC estimators are widely used for standard (i.e., nonnested) integration problems due to their scalability. However, achieving a desired error tolerance $TOL > 0$ with standard MC comes with a computational cost of $\mathcal{O}(TOL^{-2})$. The QMC methods, utilizing low-discrepancy sequences, can provide improved convergence for smooth integrands, reducing computational effort for instance to almost $\mathcal{O}(TOL^{-1})$, where the multiplicative constant usually depends exponentially on the dimension of the integration domain. Dimension-independent constant terms are possible for certain integrands when weighted function spaces are employed [21]. Moreover, higher-order QMC methods permit even faster convergence for certain integrands [21]. Despite these improvements, challenges remain, especially when dealing with singularities. Indeed, standard error bounds for QMC methods via the Koksma–Hlawka inequality [24] become obsolete for singular integrands that exhibit infinite Hardy–Krause variation [15]. A useful error bound for such integrands was introduced by Owen [26] and later generalized by He et al. [27].

However, randomized QMC (rQMC) methods combine the efficiency of QMC with important properties of MC methods, including its unbiasedness and practical, variance-based error estimates. Nevertheless, for nested integration, the nonlinear function f that evaluates the inner integrand introduces a bias in the outer integral estimation, even if the inner integral is estimated using an unbiased method. Thus, the number of inner and outer samples of numerical estimators must be carefully controlled to reduce the overall error.

Ryan [1] introduced the double-loop MC (DLMC) estimator for the expected information gain (EIG) of an experiment, which is a comprehensive tool to evaluate the quality of an experimental setup, or design [5, 6]. Estimating the EIG efficiently and accurately enables design optimization in a Bayesian setting, and hence efficient data collection in real-world applications. We consider integrands whose approximate evaluation requires a numerical solution to a partial differential equation (PDE), for instance, via the finite

element method (FEM). The DLMC estimator was further developed in [3] with regard to additional bias error stemming from the FEM approximation of the experiment model. The computational cost of this estimator to achieve an error tolerance TOL is $\mathcal{O}(TOL^{-3})$ with no finite element approximation, and is $\mathcal{O}(TOL^{-3-\gamma/\eta_{\mathrm{w}}})$ with such an approximation, where the term γ/η_{w} relates to computational work and accuracy of the FEM and will be specified later.

A multilevel MC (MLMC) estimator for the EIG, based on the error analysis by Giles [10], was introduced in [4]. This estimator addressed the inner sampling and FEM discretization error. An MLMC estimator combined with antithetic sampling for EIG estimation was also proposed in [30], where the inexact evaluation of the inner integrand was not considered. Attempts to combine multilevel estimators with rQMC were made in [32] where dimension truncation via the Karhunen–Loéve expansion was also considered; however, there was no nested integration structure. The works [12] and [13] introduced estimators for nested integrals based on the MLMC method, where the number of samples in the inner QMC estimator defines the levels. On the other hand, [14] proposed a multilevel rQMC estimator where the levels are defined by the number of samples in an inner MC estimator. Neither of these studies addressed FEM discretizations. Estimators for nested integrals with more than one nesting were analyzed in [40] and [41] for general applications. The latter of these studies developed a randomized (i.e., unbiased) MLMC method, achieving a complexity independent of the number of nestings. Randomized MLMC estimators for financial applications were proposed in [42, 43] and for Bayesian experimental design applications in [31]. The recent work [37] proposed a multi-index MC (MIMC) estimator for EIG with discretized inner integrand that can achieve a computational cost of $\mathcal{O}(TOL^{-2})$. MIMC methods are an extension to MLMC methods, where an optimization among multi-indices of accuracies is performed [25]. This scheme allows for exploitation of anisotropies in the error convergence among different kinds of discretizations present in the integrand. The randomized double-loop QMC (rDLQMC) estimator was introduced in [2], where rQMC methods were applied to estimate both integrals. This estimator achieves an optimal computational cost of almost $\mathcal{O}(TOL^{-1.5-\gamma/\eta_{\mathrm{w}}})$. An importance sampling scheme for the inner integral based on the Laplace approximation to reduce the multiplicative error constant was also proposed in [2].

The two estimators based on rQMC methods with either a full tensor product structure or a sparse tensor product structure proposed in [35] were geared toward inner integrands that could be evaluated exactly, and achieved computational work of almost $\mathcal{O}(TOL^{-2})$ and $\mathcal{O}(TOL^{-1})$, respectively. Moreover, the FEM-based approximation was treated as the exact integrand in [35] whenever needed.

As our main contributions, we propose a novel estimator including a multilevel structure, combined with QMC and rQMC methods, analyze the resulting bias and statistical errors, and derive the optimal number of inner and outer samples and the discretization parameter. Moreover, we derive conditions for applying this estimator to the EIG of an experiment based on a truncation scheme introduced in a previous work [2]. We then demonstrate that our estimator achieves computational cost of nearly $\mathcal{O}(TOL^{-1-\gamma/\eta_{\mathrm{w}}})$ for EIG estimation in the presence of FEM discretization, and $\mathcal{O}(TOL^{-1})$ without such discretization. Whenever the evaluation of the integrand that is approximated via MC (or rQMC) methods requires additional discretization methods, such as the inner rQMC method in this setting, multilevel techniques employ a telescoping structure of increasing accuracy of the integrand approximation to reduce the variance of the combined estimator. The integral is approximated for a very coarse inner integrand, and the result is then corrected via increasing levels of difference estimators. The primary application of the estimator presented in this study is to approximate the EIG, where the nonlinear function separating the inner and outer integrals is the logarithm. The regularity assumptions stated in this work should be understood in this context; however, the analysis is conducted in general terms and can be adapted to more specific settings.

This work is structured as follows. In Section 2, we provide a brief overview of numerical approximation methods of integrals and introduce useful notation. As the primary contribution, we present the multilevel double-loop QMC (MLDLQMC) estimator with inexact sampling in Section 3, along with asymptotic error bounds of the bias and statistical error. This section concludes with a bound on the total work of this estimator for a given error tolerance based on a nearly optimal allocation of the inner and outer samples. In Section 4, we demonstrate how to apply the MLDLQMC estimator to the estimate the EIG of an experiment. This analysis involves a truncation of the Gaussian observation noise present in the experiment model. Section 5 provides numerical results to highlight the practical applicability and competitiveness of the proposed estimators for a Bayesian experimental design setting.

2. Overview of numerical integration methods

For the reader's convenience, we provide an overview of existing numerical methods to estimate integrals and nested integrals that are relevant to this work. Moreover, we provide error bounds and analyze the computational cost associated with these methods.

2.1. **Monte Carlo method.** Given an integral

$$I = \int_{[0,1]^d} \varphi(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z}, \tag{1}$$

where $\varphi : [0,1]^d \to \mathbb{R}$ is square-integrable and d is a positive integer, the MC estimator

$$I_{\mathrm{MC}} \coloneqq \frac{1}{M} \sum_{m=1}^{M} \varphi(\boldsymbol{z}^{(m)}) \tag{2}$$

uses independent and identically distributed (iid) samples $\boldsymbol{z}^{(m)}$, where $1 \le m \le M$, from the uniform distribution $\mathcal{U}\left([0,1]^d\right)$, to approximate I in (1). The central limit theorem (CLT) [19] reveals that the error of the MC estimator is proportional to the estimator variance $\mathbb{V}[I_{\mathrm{MC}}]$ in a probabilistic sense:

$$\mathbb{P}\left(|I_{\mathrm{MC}} - I| \le C_\alpha \sqrt{\mathbb{V}[I_{\mathrm{MC}}]}\right) \ge 1 - \alpha, \tag{3}$$

as $M \to \infty$, where $\alpha \in (0,1)$ is a parameter, $C_\alpha = \Phi^{-1}(1-\alpha/2)$, Φ^{-1} denotes the inverse cumulative distribution function (CDF) of the standard normal distribution, and $C_\alpha \to \infty$ as $\alpha \to 0$. The bound (3) motivates the use of variance reduction techniques to control the error. Independence of the samples in the MC estimator ensures that

$$\mathbb{V}[I_{\mathrm{MC}}] = \frac{\mathbb{V}[\varphi]}{M}. \tag{4}$$

Increasing the number of samples M is a straightforward way to reduce the variance (4). The cost of the MC estimator is proportional to the number of times the function φ is evaluated, i.e., M. Thus, for a linear increase in cost, we obtain a linear decrease in variance, and a sublinear decrease in the error (3), namely, at rate 1/2. A total number of samples proportional to TOL^{-2} is required to achieve an error less than a prescribed tolerance $TOL > 0$, i.e.,

$$C_\alpha \sqrt{\frac{\mathbb{V}[\varphi]}{M}} < TOL. \tag{5}$$

Thus, the total computational work of the MC estimator is

$$W^*_{MC} = \mathcal{O}\left(TOL^{-2}\right), \tag{6}$$

as $TOL \to 0$.

2.2. **Quasi-Monte Carlo method.** The convergence rate of the MC error is independent of the dimension d. However, the conventional rate of 1/2 is not large enough for many practical applications. One primary concern with iid samples is that they do not fill the integration domain well. The QMC method,

$$I_{\mathrm{Q}} \coloneqq \frac{1}{M} \sum_{m=1}^{M} \varphi(\boldsymbol{z}^{(m)}), \tag{7}$$

where

$$\boldsymbol{z}^{(m)} \coloneqq \boldsymbol{t}^{(m)}, \quad 1 \le m \le M, \tag{8}$$

instead uses deterministic points $\boldsymbol{t}^{(m)}$, where $1 \le m \le M$. These are known as low-discrepancy points with the following property:

$$\begin{aligned} D^*(\boldsymbol{t}^{(1)}, \ldots, \boldsymbol{t}^{(M)}) &\coloneqq \sup_{(s_1,\ldots,s_d)\in[0,1]^d} \left| \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{\{\boldsymbol{t}^{(m)} \in \prod_{i=1}^d [0,s_i)\}} - \prod_{i=1}^{d} s_i \right|, \\ &\le \frac{C_{\epsilon,d}}{M^{1-\epsilon}}, \end{aligned} \tag{9}$$

for any $\epsilon > 0$, where $C_{\epsilon,d} > 0$ is independent of M, and $C_{\epsilon,d} \to \infty$ as $\epsilon \to 0$, where D^* is the star-discrepancy [15, Def. (5.2)]. An important class of low-discrepancy sequences are digital sequences [16, 29, 17, 20]. We apply a digital sequence called the Sobol' sequence [22] throughout this work. However, other low-discrepancy sequences could also be used as long as the bound in (9) is satisfied. The approximation error of the QMC estimator is bounded by the Koksma–Hlawka inequality [24, 16, 15],

$$|I_{\mathrm{Q}} - I| \le D^*(\boldsymbol{t}^{(1)}, \dots, \boldsymbol{t}^{(M)}) V_{\mathrm{HK}}(\varphi), \tag{10}$$

where $V_{\mathrm{HK}}(\varphi)$ is the recursively defined Hardy–Krause variation [15, Def. (5.8)] for continuously differentiable φ:

$$V_{\mathrm{HK}}(\varphi) \coloneqq \int_{[0,1]^d} \left| \left(\prod_{i=1}^{d} \frac{\partial}{\partial z_i} \right) \varphi(\boldsymbol{z}) \right| \mathrm{d}\boldsymbol{z} + \sum_{i=1}^{d} V_{\mathrm{HK}}(\varphi(\boldsymbol{z})|_{z_i=1}), \tag{11}$$

with $\varphi(\boldsymbol{z})|_{z_i=1} : [0,1]^{d-1} \to \mathbb{R}$ representing the restriction of $\varphi(\boldsymbol{z})$ to $\varphi(z_1, \dots, z_i = 1, \dots, z_d)$. For $\varphi : [0,1] \to \mathbb{R}$, the Hardy–Krause variation is defined as follows:

$$V_{\mathrm{HK}}(\varphi) \coloneqq \int_{[0,1]} \left| \frac{\mathrm{d}}{\mathrm{d}z} \varphi(z) \right| \mathrm{d}z. \tag{12}$$

To achieve an error less than TOL, the QMC estimator consequently has computational work

$$W_Q^* = \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon}} \right) \tag{13}$$

as $TOL \to 0$ for integrands with bounded Hardy–Krause variation (11), based on the bounds in (9) and (10), where the multiplicative constant may depend on the integrand dimension d. The main concerns with the QMC method are, in no particular order, that:

- There may be a multiplicative constant in the error depending on the dimension d;
- The first point of many popular low-discrepancy sequences is the origin, i.e., $\boldsymbol{t}^{(1)} = \boldsymbol{0}$ including the Sobol' sequence);
- The Hardy–Krause variation in the error bound in (10) is usually more restrictive and more challenging to compute or estimate than the estimator variance arising in (3), which plays a crucial role in the error of the rQMC method;
- Many popular integrands exhibit infinite Hardy–Krause variation.

The first concern can be remedied for certain integrands and specialized QMC methods [21]. The second concern can be elegantly addressed by introducing randomization to the deterministic low-discrepancy points, as will be discussed in the following section. The third concern can also be partially addressed by randomization. The sample variance taken over multiple independent randomizations, as will be elaborated in the next section, serves as a practical error estimate through Chebyshev's inequality. Regarding the fourth concern, there are certain types of popular integrands, e.g., integrands weighted by a normal density rather than the uniform one, which must be mapped back to the interval $[0,1]^d$. This mapping results in infinite Hardy–Krause variation of the resulting integrand, even though the corresponding QMC error is finite.

2.3. **Randomized Quasi-Monte Carlo method.** The rQMC estimator for the integral (1) with R randomizations is defined as follows:

$$I_{\mathrm{rQ}}^{(R)} \coloneqq \frac{1}{R} \sum_{r=1}^{R} \frac{1}{M} \sum_{m=1}^{M} \varphi(\boldsymbol{z}^{(r,m)}), \tag{14}$$

where

$$\boldsymbol{z}^{(r,m)} \coloneqq \boldsymbol{\rho}^{(r)}(\boldsymbol{t}^{(m)}), \quad 1 \le r \le R,\ 1 \le m \le M. \tag{15}$$

This estimator uses independent randomizations $\boldsymbol{\rho}^{(r)} \in \Pi$, where $1 \le r \le R$, of the deterministic low-discrepancy sequence $\boldsymbol{t}^{(m)}$, where $1 \le m \le M$, and Π denotes the space of randomizations. Owen's scrambling for digital sequences [16, 29, 17] is employed throughout this work, where a random permutation

is applied to the base-two digits of $\boldsymbol{t}^{(m)}$, where $1 \le m \le M$. This randomization scheme preserves the low-discrepancy property with probability one (see [29, 20] for details on this method). It holds that

$$\begin{aligned}
\mathbb{V}[I_{\mathrm{rQ}}^{(R)}] &= \mathbb{V}\left[\frac{1}{R}\sum_{r=1}^{R}\frac{1}{M}\sum_{m=1}^{M}\varphi(\boldsymbol{z}^{(r,m)})\right], \\
&= \frac{1}{R}\mathbb{V}\left[\frac{1}{M}\sum_{m=1}^{M}\varphi(\boldsymbol{\rho}(\boldsymbol{t}^{(m)}))\right], \\
&= \frac{1}{R}\mathbb{E}\left[\left|\frac{1}{M}\sum_{m=1}^{M}\varphi(\boldsymbol{\rho}(\boldsymbol{t}^{(m)})) - \mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M}\varphi(\boldsymbol{\rho}(\boldsymbol{t}^{(m)}))\right]\right|^2\right], \\
&= \frac{1}{R}\mathbb{E}\left[\left|\frac{1}{M}\sum_{m=1}^{M}\varphi(\boldsymbol{\rho}(\boldsymbol{t}^{(m)})) - I\right|^2\right], \\
&\le \frac{1}{R}\mathbb{E}\left[\left|D^*\left(\boldsymbol{\rho}(\boldsymbol{t}^{(1)}),\dots,\boldsymbol{\rho}(\boldsymbol{t}^{(M)})\right)V_{\mathrm{HK}}(\varphi)\right|^2\right], \qquad (16)
\end{aligned}$$

where the last line follows from the Koksma–Hlawka inequality (10) and the variance and expectation are taken over the randomization $\boldsymbol{\rho}$. The variance of the rQMC estimator can thus be bounded as

$$\mathbb{V}[I_{\mathrm{rQ}}^{(R)}] \le \frac{C_{\epsilon,d}^2 (V_{\mathrm{HK}}(\varphi))^2}{RM^{2-2\epsilon}}, \qquad (17)$$

for any $\epsilon > 0$ and $C_{\epsilon,d} \to \infty$ as $\epsilon \to 0$ via (9) for all sequences that maintain the low-discrepancy property in $L^2(\Pi, \mathbb{R})$ under randomization, such as the Sobol' sequence with Owen's scrambling. It follows that increasing the number of QMC points is often more efficient than increasing the number of randomizations. There are two important concerns with the Hardy–Krause variation. First, several important classes of functions have infinite Hardy–Krause variation and finite variance. Thus, for such cases, the Koksma–Hlawka inequality does not provide a useful error bound. Second, it is challenging to obtain practical estimates of the Hardy–Krause variation, whereas the variance can be estimated using the sample variance. Regarding the latter point, the rQMC variance $\mathbb{V}[I_{\mathrm{rQ}}^{(R)}]$ can be estimated using R independent randomizations whereas the bound in (16) is only for theoretical purposes. This situation presents a trade-off: increasing the number of randomizations improves the reliability of the variance estimate but only improves the estimate of the integral, i.e., expectation, at the slower MC rate. A naive approximation based on the CLT for finite R can be problematic for certain integrands [18, 23]. Chebyshev's inequality serves as an alternative, at the expense of a larger confidence constant $C_\alpha = 1/\sqrt{\alpha}$. The number of randomizations R plays a lesser role in theoretical error bounds, and we thus introduce the rQMC estimator with one randomization as

$$I_{\mathrm{rQ}} \coloneqq I_{\mathrm{rQ}}^{(1)}. \qquad (18)$$

As for the first point mentioned above, the following result in (21) provides an error bound for a specific, yet common, type of integrands that have infinite Hardy–Krause variation. We begin by introducing Assumption 1 (from [26]):

Assumption 1 (Boundary growth condition)**.** *Let $\varphi : [0,1]^d \to \mathbb{R}$ and assume that there exists $0 < b < \infty$ and $A_i > 0$ for $1 \le i \le d$, where*

$$A_{\max} \coloneqq \max_{1\le i\le d} A_i < \frac{1}{2}, \qquad (19)$$

such that

$$\left|\left(\prod_{j\in u}\frac{\partial}{\partial z_j}\right)\varphi(\boldsymbol{z})\right| \le b\prod_{i=1}^{d}\min(z_i, 1-z_i)^{-A_i-\mathbb{1}_{\{i\in u\}}} \qquad (20)$$

for all $\boldsymbol{z} = (z_1,\dots,z_d) \in (0,1)^d$ and all $u \subseteq \{1,\dots,d\}$ with the convention that $(\prod_{j\in u}\partial/\partial z_j)\varphi \equiv \varphi$ for $u = \emptyset$.

Assumption 1 implies, among other things, that $\varphi \in L^2([0,1]^d)$, i.e., that $\varphi(\boldsymbol{Z})$, where $\boldsymbol{Z}$ is a uniform random variable, has finite variance. This allows for the application of standard error bounds for the MC method. Next, the following proposition is stated (from [27]):

Proposition 1 (Thm. 3.1 in [27])**.** *Given Assumption 1, the rQMC estimator* (18) *satisfies the following* $L^2([0,1]^d)$ *error bound:*

$$\mathbb{E}[|I_{\mathrm{rQ}} - I|^2]^{\frac{1}{2}} \leq bB_A C_{\epsilon,d} M^{-1+\epsilon+A_{\max}}, \tag{21}$$

for all $\epsilon > 0$*, where* $B_A \to \infty$ *as* $\min_{1\leq i\leq d} A_i \to 0$ *and* $C_{\epsilon,d} \to \infty$ *as* $\epsilon \to 0$*.*

This proposition is an extension of an earlier result from [26], where a similar bound was provided for the $L^1([0,1]^d)$ error. It follows that the computational work of the rQMC estimator to achieve a root mean square error less than TOL is

$$W^*_{rQ} = \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon}}\right) \tag{22}$$

as $TOL \to 0$ for integrands with bounded Hardy–Krause variation, and

$$W^*_{rQ} = \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}}\right) \tag{23}$$

as $TOL \to 0$ for integrands with singularities at the boundary satisfying Assumption 1, where $A_{\max} > 0$ as in (19) indicates the severity of the singularity. The condition that $A_{\max} < 1/2$ also implies that the rQMC error converges faster than the MC error.

2.4. **Double-loop Monte Carlo method for nested integrals.** Following [2], a nested integral is defined as follows:

Definition 1 (Nested integral)**.** *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a nonlinear function and let* $g : [0,1]^{d_1} \times [0,1]^{d_2} \to \mathbb{R}$ *define a nonlinear relation between* $\boldsymbol{x} \in [0,1]^{d_2}$ *and* $\boldsymbol{y} \in [0,1]^{d_1}$*, where* d_1, d_2 *are positive integers. Then, a nested integral is defined as follows:*

$$I = \int_{[0,1]^{d_1}} f\left(\int_{[0,1]^{d_2}} g(\boldsymbol{y},\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right)\mathrm{d}\boldsymbol{y}. \tag{24}$$

Depending on the context, it may be beneficial to write the nested integral as a nested expectation instead, that is,

$$I = \mathbb{E}[f(\mathbb{E}[g(\boldsymbol{y},\boldsymbol{x})|\boldsymbol{y}])], \tag{25}$$

where $\boldsymbol{y} \sim \mathcal{U}\left([0,1]^{d_1}\right)$ and $\boldsymbol{x} \sim \mathcal{U}\left([0,1]^{d_2}\right)$. Moreover, the notation $\bar{g} : [0,1]^{d_1} \mapsto \mathbb{R}$ is defined as follows:

$$\begin{aligned} \bar{g}(\boldsymbol{y}) &\coloneqq \mathbb{E}[g(\boldsymbol{y},\boldsymbol{x})|\boldsymbol{y}], \\ &= \int_{[0,1]^{d_2}} g(\boldsymbol{y},\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \end{aligned} \tag{26}$$

Ryan [1] introduced the DLMC estimator for nested integrals:

$$I_{\mathrm{DLMC}} \coloneqq \frac{1}{N}\sum_{n=1}^{N} f\left(\frac{1}{M}\sum_{m=1}^{M} g(\boldsymbol{y}^{(n)}, \boldsymbol{x}^{(n,m)})\right), \tag{27}$$

where $\boldsymbol{y}^{(n)}$, for $1 \leq n \leq N$, are iid samples from $\mathcal{U}\left([0,1]^{d_1}\right)$. Moreover, $\boldsymbol{x}^{(n,m)}$, for $1 \leq n \leq N$ and $1 \leq m \leq M$, are iid samples from $\mathcal{U}\left([0,1]^{d_2}\right)$ for the specific application of EIG estimation where f is the logarithm and g is the likelihood of the data observation. For large values of N and M, this estimator has a total error based on the CLT [1, 3] of

$$|I_{\mathrm{DLMC}} - I| \leq \frac{C_{\mathrm{DL,B}}}{M} + C_\alpha\sqrt{\frac{C^{(1)}_{\mathrm{DL,V}}}{N} + \frac{C^{(2)}_{\mathrm{DL,V}}}{NM}}, \quad \text{with probability at least } 1-\alpha, \tag{28}$$

for some $\alpha \in (0,1)$, where the first term is a bias stemming from the variance of the inner MC method, and the second term is derived from the total variance of the DLMC estimator, and $C_\alpha \to \infty$ as $\alpha \to 0$.

Moreover, $C_{\mathrm{DL,B}} > 0$ and $C^{(1)}_{\mathrm{DL,V}}, C^{(2)}_{\mathrm{DL,V}} > 0$ are constants independent of N and M. A total of $N \times M$ samples is required for this estimator, and thus the total work to achieve an error less than TOL is:

$$W^*_{DL} = \mathcal{O}\left(TOL^{-3}\right) \tag{29}$$

as $TOL \to 0$.

2.5. **Double-loop randomized quasi-Monte Carlo method for nested integrals.** The rDLQMC estimator as introduced in [2] is:

$$I^{(S,R)}_{\mathrm{rDLQ}} \coloneqq \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N}\sum_{n=1}^{N} f\left(\frac{1}{R}\sum_{r=1}^{R}\frac{1}{M}\sum_{m=1}^{M} g(\boldsymbol{y}^{(s,n)}, \boldsymbol{x}^{(s,n,r,m)})\right), \tag{30}$$

where the samples take the following form

$$\boldsymbol{y}^{(s,n)} \coloneqq \boldsymbol{\tau}^{(s)}(\boldsymbol{u}^{(n)}), \quad 1 \le s \le S, 1 \le n \le N. \tag{31}$$

Here, $\boldsymbol{\tau}^{(s)}$, with $1 \le s \le S$, are independent randomizations of the Sobol' sequence $\boldsymbol{u}^{(n)}$ in d_1, where $1 \le n \le N$, using Owen's scrambling, and

$$\boldsymbol{x}^{(s,n,r,m)} \coloneqq \boldsymbol{\rho}^{(s,n,r)}(\boldsymbol{t}^{(m)}), \quad 1 \le s \le S, 1 \le n \le N, 1 \le r \le R, 1 \le m \le M, \tag{32}$$

where $\boldsymbol{\rho}^{(s,n,r)}$, with $1 \le s \le S$, $1 \le n \le N$, and $1 \le r \le R$ are independent randomizations of the Sobol' sequence $\boldsymbol{t}^{(m)}$ in d_2, where $1 \le m \le M$, using Owen's scrambling. Assuming that the nested integrand $f(\bar{g})$ satisfies Assumption 1 with $\bar{g}$ as in (26), for $0 < A_{\max} < 1/2$, whereas the inner integrand g has bounded Hardy–Krause variation, this estimator has a total error based on Chebyshev's inequality [2] of

$$|I^{(1,1)}_{\mathrm{rDLQ}} - I| \le \frac{C_{\mathrm{DLQ,B}}}{M^{2-2\epsilon}} + C_\alpha\sqrt{\frac{C^{(1)}_{\mathrm{DLQ,V}}}{N^{2-2\epsilon-2A_{\max}}}} + \frac{C^{(2)}_{\mathrm{DLQ,V}}}{NM^{2-2\epsilon}}, \quad \text{with probability at least } 1-\alpha, \tag{33}$$

for some $\alpha \in (0,1)$, $C_\alpha \to \infty$ as $\alpha \to 0$, for $S = R = 1$ randomizations, where $C_{\mathrm{DLQ,B}} > 0$ and $C^{(1)}_{\mathrm{DLQ,V}}, C^{(2)}_{\mathrm{DLQ,V}} > 0$ are constants independent of N and M. The total work to achieve an error less than TOL based on this error bound is:

$$W^*_{rDLQ} = \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}-\frac{1}{2-2\epsilon}}\right) \tag{34}$$

as $TOL \to 0$. This setting is also analyzed in particular for the EIG estimation in [2].

2.6. **Multilevel double-loop Monte Carlo method for nested integrals.** The multilevel method for nested integrals [30] constructs a sequence of increasingly accurate MC estimators for the inner integral to harness variance reduction achieved by taking differences of adjacent levels evaluated for the same random variables. The general form of the MLMC estimator for a nonnested integral is given as follows:

$$I_{\mathrm{ML}} \coloneqq \frac{1}{N_0}\sum_{n=1}^{N_0}\varphi_0\left(\boldsymbol{z}^{(0,n)}\right) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left[\varphi_\ell\left(\boldsymbol{z}^{(\ell,n)}\right) - \varphi_{\ell-1}\left(\boldsymbol{z}^{(\ell,n)}\right)\right], \tag{35}$$

where $\boldsymbol{z}^{\ell,n}$ are iid samples from $\mathcal{U}([0,1]^d)$ for $1 \le n \le N_\ell$, and $1 \le \ell \le L$, $L \in \mathbb{N}$. The multilevel double-loop MC (MLDLMC) estimator introduced in [30] for nested integrals defines the level φ_ℓ as follows:

$$\varphi_\ell(\boldsymbol{z}^{(\ell,n)}) \coloneqq f\left(\frac{1}{M_\ell}\sum_{m=1}^{M_\ell} g\left(\boldsymbol{y}^{(\ell,n)}, \boldsymbol{x}^{(\ell,n,m)}\right)\right), \tag{36}$$

where $\boldsymbol{z}^{(\ell,n)} \coloneqq (\boldsymbol{y}^{(\ell,n)}, \boldsymbol{x}^{(\ell,n,m)})$ and $M_\ell = M_0 2^\ell$, for $M_0 > 0$. An antithetic construction first introduced in [38] to increase the coupling, and thus reduce the variance further, was used to obtain computational work:

$$W^*_{\mathrm{ML}} = \mathcal{O}\left(TOL^{-2}\right) \tag{37}$$

to achieve an error tolerance $TOL > 0$ as $TOL \to 0$ for estimating the EIG. Note that the asymptotic work presented in [30] was derived under the assumption that exact evaluations of g were available. Depending on the considered application, this assumption may be inadequate, because the considered experiment generating the data may involve numerical approximations.

2.7. **Inexact sampling.** Inner integrands g encountered in many practical applications cannot be evaluated exactly at a given point $\boldsymbol{z}$. Instead, an approximation g_h is introduced, where $g_h \to g$ as $h \to 0$ in a suitable sense, for a discretization parameter $h > 0$. This setting is referred to as inexact sampling and examples include finite difference methods, FEM, and time stepping methods. This discretization introduces bias to the estimators discussed in the previous sections, and thus affects the computational work required to achieve an error tolerance TOL. To account for this additional bias, the following assumptions are considered in this work:

Assumption 2 (Weak convergence rate)**.** *Let g_h be an approximation of $g : [0,1]^d \mapsto \mathbb{R}$, and $h > 0$ such that the weak error of g_h is given by*

$$|\mathbb{E}[g_h(\boldsymbol{z}) - g(\boldsymbol{z})]| \leq C_{\mathrm{w}} h^{\eta_{\mathrm{w}}}, \tag{38}$$

where $\eta_{\mathrm{w}} > 0$ is the weak h-convergence rate, and $C_{\mathrm{w}} > 0$ is a constant independent of h. The work of evaluating g_h is assumed to be of order $\mathcal{O}(h^{-\gamma})$, for $\gamma > 0$ uniformly in $[0,1]^d$.

Assumption 3 (Strong convergence rate)**.** *Let g_h be as in Assumption 2. Suppose the strong error of g_h is given by*

$$\mathbb{E}[|g_h(\boldsymbol{z}) - g(\boldsymbol{z})|^2]^{\frac{1}{2}} \leq C_{\mathrm{s}} h^{\eta_{\mathrm{s}}}, \tag{39}$$

where $\eta_{\mathrm{s}} > 0$ is the strong h-convergence rate, and $C_{\mathrm{s}} > 0$ is a constant independent of h.

In the inexact sampling case, an additional term $-\gamma/\eta_{\mathrm{w}}$ is added to the exponent of TOL in the total computational work of the MC, QMC, rQMC, DLMC, and rDLQMC estimators [3, 2], which is motivated by the EIG application. Note that only the inner integrand in the nested setting is assumed to require discretization throughout this work. The work [4] proposed an MLMC estimator for nested integration applied to the EIG, where the data generation involves approximating a PDE via the FEM. The accuracy level in the error bound proposed in this work was determined by the inner number of samples and the discretization parameter of the FEM. However, the total computational work is then bounded for a fixed inner number of samples as $\mathcal{O}(TOL^{-2})$ for large and intermediate error tolerances $TOL > 0$, and $\mathcal{O}(TOL^{-3})$ asymptotically as $TOL \to 0$. This change in complexity regime is due to the effects of Laplace-based importance sampling used in this estimator, which renders just one inner sample sufficient for TOL large enough.

Remark 1 (Laplace-based importance sampling)**.** *Laplace-based importance sampling is an efficient variance reduction technique for EIG estimation* [9, 3, 4, 2]*. However, combining this method with multilevel strategies and especially with rQMC methods is nontrivial* [28]*. Thus, we leave the study of this method for future work.*

3. Multilevel double-loop randomized quasi-Monte Carlo method for nested integrals with inexact sampling

As the primary focus of this work we analyzed the multilevel double-loop randomized quasi-Monte Carlo (MLDLQMC) estimator with inexact sampling. This estimator combines the multilevel technique with both deterministic QMC and rQMC methods. At all levels ℓ, where $0 \leq \ell \leq L$, the outer integral is estimated using the rQMC method, allowing for a practical error estimate via the sample variance. This estimate is theoretically justified by Proposition 1. The inner estimation at levels ℓ, where $0 \leq \ell \leq L-1$, is based on QMC methods. Due to the deterministic nature of these methods, the variances of the estimators for those levels depend only on the outer samples. The multilevel structure ensures that only the final level L contributes to the bias of the overall estimator. The rQMC method is applied for the inner integral approximation only at level L, allowing for practical error estimates of the bias and variance of the estimator.

Definition 2 (Multilevel double-loop randomized quasi-Monte Carlo estimator with inexact sampling)**.** *The MLDLQMC estimator with inexact sampling is defined as follows:*

$$
\begin{aligned}
I_{\mathrm{MLDLQ}}^{(S,R)} \coloneqq & \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_0}\sum_{n=1}^{N_0} f\left(\frac{1}{M_0}\sum_{m=1}^{M_0} g_{h_0}\left(\boldsymbol{y}^{(0,s,n)},\boldsymbol{x}^{(m)}\right)\right) \\
& + \sum_{\ell=1}^{L-1}\frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left[f\left(\frac{1}{M_\ell}\sum_{m=1}^{M_\ell} g_{h_\ell}\left(\boldsymbol{y}^{(\ell,s,n)},\boldsymbol{x}^{(m)}\right)\right) - f\left(\frac{1}{M_{\ell-1}}\sum_{m=1}^{M_{\ell-1}} g_{h_{\ell-1}}\left(\boldsymbol{y}^{(\ell,s,n)},\boldsymbol{x}^{(m)}\right)\right)\right] \\
& + \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_L}\sum_{n=1}^{N_L}\left[f\left(\frac{1}{R}\sum_{r=1}^{R}\frac{1}{M_L}\sum_{m=1}^{M_L} g_{h_L}\left(\boldsymbol{y}^{(L,s,n)},\boldsymbol{x}^{(s,r,m)}\right)\right)\right. \\
& \qquad \left. - f\left(\frac{1}{M_{L-1}}\sum_{m=1}^{M_{L-1}} g_{h_{L-1}}\left(\boldsymbol{y}^{(L,s,n)},\boldsymbol{x}^{(m)}\right)\right)\right]. \qquad (40)
\end{aligned}
$$

Here, g_{h_ℓ} is an approximation of g such that Assumption 2 and Assumption 3 hold for $0 \le \ell \le L$, and

$$
\boldsymbol{y}^{(\ell,s,n)} \coloneqq \boldsymbol{\tau}^{(\ell,s)}(\boldsymbol{u}^{(n)}), \quad 0 \le \ell \le L,\, 1 \le s \le S,\, 1 \le n \le N_\ell, \qquad (41)
$$

where $\boldsymbol{\tau}^{(\ell,s)}$ are iid Owen's scramblings of points in the Sobol' sequence $\boldsymbol{u}^{(n)}$ in dimension d_1, for $1 \le s \le S$, $1 \le n \le N_\ell$, and $0 \le \ell \le L$, such that

$$
\mathbb{E}[D^*(\boldsymbol{y}^{(\ell,s,1)},\dots,\boldsymbol{y}^{(\ell,s,N_\ell)})] \le \frac{C_{\epsilon,d_1}}{N_\ell^{1-\epsilon}}, \quad 0 \le \ell \le L,\, 1 \le s \le S, \qquad (42)
$$

for all $\epsilon > 0$, where $C_{\epsilon,d_1} > 0$ is independent of N_ℓ, where $0 \le \ell \le L$, and $C_{\epsilon,d_1} \to \infty$ as $\epsilon \to 0$. Moreover,

$$
\boldsymbol{x}^{(m)} \coloneqq \boldsymbol{t}^{(m)}, \quad 1 \le m \le M_\ell, \qquad (43)
$$

are points in the Sobol' sequence $\boldsymbol{t}^{(m)}$ in dimension d_2, for $1 \le m \le M_\ell$, $0 \le \ell \le L-1$, such that

$$
D^*(\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(M_\ell)}) \le \frac{C_{\epsilon,d_2}}{M_\ell^{1-\epsilon}}, \quad 0 \le \ell \le L, \qquad (44)
$$

for all $\epsilon > 0$, where $C_{\epsilon,d_2} > 0$ is independent of M_ℓ, where $0 \le \ell \le L$, and $C_{\epsilon,d_2} \to \infty$ as $\epsilon \to 0$. Finally, the following construction in applied for the randomized inner points at level L:

$$
\boldsymbol{x}^{(s,r,m)} \coloneqq \boldsymbol{\rho}^{(s,r)}(\boldsymbol{t}^{(m)}), \quad 1 \le s \le S,\, 1 \le r \le R,\, 1 \le m \le M_L, \qquad (45)
$$

where $\boldsymbol{\rho}^{(s,r)}$ are iid Owen's scramblings of points in the Sobol' sequence $\boldsymbol{t}^{(m)}$ in dimension d_2, for $1 \le s \le S$, $1 \le r \le R$, and $1 \le m \le M_L$, such that

$$
\mathbb{E}[D^*(\boldsymbol{x}^{(s,r,1)},\dots,\boldsymbol{x}^{(s,r,M_L)})] \le \frac{C_{\epsilon,d_2}}{M_L^{1-\epsilon}}, \quad 1 \le s \le S,\, 1 \le r \le R, \qquad (46)
$$

for all $\epsilon > 0$, where $C_{\epsilon,d_2} > 0$ is independent of M_L and $C_{\epsilon,d_2} \to \infty$ as $\epsilon \to 0$.

For now, we set:

$$
M_\ell \coloneqq M_0 2^\ell, \quad 1 \le \ell \le L, \qquad (47)
$$

where $M_0 \ge 1$, and

$$
h_\ell \coloneqq h_0 2^{-\ell}, \quad 1 \le \ell \le L, \qquad (48)
$$

where $h_0 > 0$. The choices (47) and (48) are not specific to the MLDLQMC estimator and are revisited later (see Remark 3). It is crucial for the quality of the estimator that all point sets used have the low-discrepancy property (9); however, it would be possible to use different points $\boldsymbol{x}^{(m)}$, where $1 \le m \le M_\ell$, at different levels $1 \le \ell \le L-1$, as long as the telescoping property is preserved. The following notations are defined for convenience:

$$
I_{\mathrm{MLDLQ}}^{(S,R)} = I_{\mathrm{MLDLQ}_0^0}^{(S,R)} + \sum_{\ell=1}^{L}\left[I_{\mathrm{MLDLQ}_\ell^\ell}^{(S,R)} - I_{\mathrm{MLDLQ}_\ell^{\ell-1}}^{(S,R)}\right], \qquad (49)
$$

where

$$I^{(S,R)}_{\mathrm{MLDLQ}^k_\ell} \coloneqq \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell} f\left(\frac{1}{M_k}\sum_{m=1}^{M_k} g_{h_k}\left(\boldsymbol{y}^{(\ell,s,n)}, \boldsymbol{x}^{(m)}\right)\right), \quad 1 \le \ell \le L,\, 1 \le k \le L-1, \tag{50}$$

and

$$I^{(S,R)}_{\mathrm{MLDLQ}^L_L} \coloneqq \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_L}\sum_{n=1}^{N_L} f\left(\frac{1}{R}\sum_{r=1}^{R}\frac{1}{M_L}\sum_{m=1}^{M_L} g_{h_L}\left(\boldsymbol{y}^{(L,s,n)}, \boldsymbol{x}^{(s,r,m)}\right)\right). \tag{51}$$

Here, two additional indices were introduced to distinguish between levels of the estimator and the two terms within each level difference. The MLDLQMC estimator with $S = R = 1$ randomizations is defined as follows:

$$I_{\mathrm{MLDLQ}} \coloneqq I^{(1,1)}_{\mathrm{MLDLQ}}, \tag{52}$$

with $I_{\mathrm{MLDLQ}^k_\ell}$ defined analogously for $1 \le \ell, k \le L$, and $\bar{g}_h$ is defined analogously to (26) for $\boldsymbol{y} \in [0,1]^{d_1}$ and $\boldsymbol{x} \in [0,1]^{d_2}$. We proceed by bounding the total error of the MLDLQMC estimator based on certain assumptions that are verified for the EIG application, where f is the logarithm and g is the likelihood combined with appropriate inverse CDFs. The total error is split into bias and statistical errors as follows:

$$|I_{\mathrm{MLDLQ}} - I| \le \underbrace{|\mathbb{E}[I_{\mathrm{MLDLQ}}] - I|}_{\text{bias error}} + \underbrace{|I_{\mathrm{MLDLQ}} - \mathbb{E}[I_{\mathrm{MLDLQ}}]|}_{\text{statistical error}}. \tag{53}$$

This study follows the work by [2] and restates the necessary assumptions for completeness to bound the bias of the MLDLQMC estimator.

Assumption 4 (Integrability assumption for the composition of the outer and inner integrand)**.** *There exists $0 < \delta_K < 1$ and $1 \le p \le \infty$ such that $f(K\bar{g}(\cdot)) : [0,1]^{d_1} \to \mathbb{R}$ is an element of the Sobolev space $W^{3,p}([0,1]^{d_1})$ for all $K \in (1-\delta_K, 1+\delta_K)$, where $\bar{g}$ is as in* (26)*.*

Assumption 5 (Inverse inequality for the inner integrand)**.** *Let $g(\cdot,\cdot) : [0,1]^{d_1} \times [0,1]^{d_2} \to \mathbb{R}$ and assume that there exists $0 \le k < \infty$ such that*

$$\sup_{\boldsymbol{y}\in[0,1]^{d_1}} \left|\frac{V_{\mathrm{HK}}(g(\boldsymbol{y},\cdot))}{\bar{g}(\boldsymbol{y})}\right| < k\,, \tag{54}$$

with V_{HK} as in (11) *and $\bar{g}$ as in* (26)*. Furthermore, if Assumption 4 holds, assume that there exists $1 \le q \le \infty$ with $1/p + 1/q = 1$ for p as in Assumption 4 such that $\bar{g} \in L^{3q}([0,1]^{d_1})$.*

Together, these assumptions ensure that the bias from the inner sampling in the MLDLQMC estimator remains bounded. The assumptions depend on the particular integrands f and g and must be verified for each application. For the EIG, it is demonstrated that these assumptions hold in [2] and also in Section 4 and Appendix A of this work, contingent on certain auxiliary assumptions on the experiment under consideration. The expectation of the MLDLQMC estimator is equal to the expectation of the corresponding single-level estimator at the finest level L, that is,

$$\mathbb{E}[I_{\mathrm{MLDLQ}}] = \mathbb{E}[I_{\mathrm{MLDLQ}^L_L}]. \tag{55}$$

The bound on the bias of the MLDLQMC estimator follows immediately from [2, Proposition 2] with $M = M_L$ and $h = h_L$ as stated below.

Proposition 2 (Bias of the MLDLQMC estimator with inexact sampling)**.** *Given Assumption 2 for $g_{h_L}(\cdot,\cdot) : [0,1]^{d_1} \times [0,1]^{d_2} \to \mathbb{R}$ and Assumptions 4 and 5, if $|f'''|$ is monotonic, the bias of the MLDLQMC estimator* (52) *with inexact sampling has the following upper bound:*

$$|\mathbb{E}[I_{\mathrm{MLDLQ}}] - I| \le C_{\mathrm{w}} h_L^{\eta_{\mathrm{w}}} + \frac{\mathbb{E}[|\bar{g}|^2\,|f''(\bar{g})|]k^2 C^2_{\epsilon,d_2}}{2M_L^{2-2\epsilon}} + \frac{\mathbb{E}[|\bar{g}|^3\,|f'''(K\bar{g})|]k^3 C^3_{\epsilon,d_2}}{6M_L^{3-3\epsilon}}, \tag{56}$$

for any $\epsilon > 0$, where $C_{\epsilon,d_2} \to \infty$ as $\epsilon \to 0$.

The statistical error of the MLDLQMC estimator is bounded in probability using Chebyshev's inequality as follows:

$$\mathbb{P}\left(|I_{\mathrm{MLDLQ}} - \mathbb{E}[I_{\mathrm{MLDLQ}}]| \le C_\alpha\sqrt{\mathbb{V}[I_{\mathrm{MLDLQ}}]}\right) \ge 1 - \alpha \tag{57}$$

for $\alpha \in (0,1)$, where $C_\alpha = 1/\sqrt{\alpha}$ and $\mathbb{V}[I_{\mathrm{MLDLQ}}]$ is the estimator variance. The randomizations of the outer samples are independent for each level; thus, the total variance of the MLDLQMC estimator is decomposed into the variances of the individual levels, that is,

$$\mathbb{V}\left[I_{\mathrm{MLDLQ}}\right] = \mathbb{V}\left[I_{\mathrm{MLDLQ}_0^0}\right] + \sum_{\ell=1}^{L-1} \mathbb{V}\left[I_{\mathrm{MLDLQ}_\ell^\ell} - I_{\mathrm{MLDLQ}_\ell^{\ell-1}}\right] + \mathbb{V}\left[I_{\mathrm{MLDLQ}_L^L} - I_{\mathrm{MLDLQ}_L^{L-1}}\right]. \tag{58}$$

The terms appearing in (58) are bounded individually in the following Propositions 3-5, leading to an upper bound of the total variance of the MLDLQMC estimator.

Assumption 6. *Let $\varphi^{(0)}_{h_0,M_0}$ be the auxiliary function*

$$\varphi^{(0)}_{h_0,M_0}(\cdot) \coloneqq f\left(\frac{1}{M_0}\sum_{m=1}^{M_0} g_{h_0}(\cdot, \boldsymbol{x}^{(m)})\right). \tag{59}$$

Moreover, let $h_0 > 0$, $M_0 \geq 1$ and $\boldsymbol{x}^{(m)}$ be as in (43)*, where $1 \leq m \leq M_0$. Then, assume that $\varphi^{(0)}_{h_0,M_0}$ satisfies Assumption 1, that is, there exists $0 < b^{(0)} < \infty$ and $A_i^{(0)} > 0$ for $1 \leq i \leq d_1$, where*

$$A^{(0)}_{\max} \coloneqq \max_{1\leq i\leq d_1} A_i^{(0)} < \frac{1}{2}, \tag{60}$$

such that

$$\left|\left(\prod_{j\in u}\frac{\partial^u}{\partial y_j}\right)\varphi^{(0)}_{h_0,M_0}(\boldsymbol{y})\right| \leq b^{(0)}\prod_{i=1}^{d_1}\min(y_i, 1-y_i)^{-A_i^{(0)}-\mathbb{1}_{\{i\in u\}}} \tag{61}$$

for all $\boldsymbol{y} = (y_1, \ldots, y_{d_1}) \in (0,1)^{d_1}$ and all $u \subseteq \{1, \ldots, d_1\}$.

The zeroth level of the MLDLQMC estimator may be regarded as an rQMC estimator (18), with $N = N_0$, for a particular choice of integrand. In view of Assumption 6, the bound on the variance contribution of the zeroth level of the MLDLQMC estimator follows immediately from Proposition 1 as stated next.

Proposition 3 (Variance of level 0 of the MLDLQMC estimator with inexact sampling)**.** *Let $\varphi^{(0)}_{h_0,M_0}$ as in* (59) *satisfy Assumption 6. The variance contribution of the zeroth level of the MLDLQMC estimator* (52) *with inexact sampling then has the following upper bound:*

$$\mathbb{V}\left[I_{\mathrm{MLDLQ}_0^0}\right] \leq \frac{(b^{(0)})^2 B^2_{A^{(0)}} C^2_{\epsilon,d_1}}{N_0^{2-2\epsilon-2A^{(0)}_{\max}}}, \tag{62}$$

for any $\epsilon > 0$, where $C_{\epsilon,d_1} \to \infty$ as $\epsilon \to 0$, and $B_{A^{(0)}} \to \infty$ as $\min_{1\leq i\leq d_1} A_i^{(0)} \to 0$.

The variance contributions of the levels ℓ, where $1 \leq \ell \leq L$, display a different behavior, as the multilevel structure introduces correlations. Nonetheless, these levels are similar in structure to the rQMC estimator (18) with $N = N_\ell$, for $1 \leq \ell \leq L-1$. The following assumptions are made to analyze these variances:

Assumption 7. *Let $\varphi^{(I)}_{h_\ell,M_\ell}$, for $0 \leq \ell \leq L$, be the auxiliary function*

$$\varphi^{(I)}_{h_\ell,M_\ell}(\cdot) \coloneqq f'\left(\bar{g}(\cdot)\right)\left(\frac{1}{M_\ell}\sum_{m=1}^{M_\ell} g_{h_\ell}(\cdot, \boldsymbol{x}^{(m)}) - \bar{g}(\cdot)\right). \tag{63}$$

Moreover, let $h_\ell > 0$, $M_\ell \geq 1$, $\bar{g}$ as in (26)*, and $\boldsymbol{x}^{(m)}$ be as in* (43)*, where $1 \leq m \leq M_\ell$, and $0 \leq \ell \leq L$. Then, assume that $\varphi^{(I)}_{h_\ell,M_\ell}$, for $0 \leq \ell \leq L$, satisfies Assumption 1, that is, there exists $0 < b^{(I)} < \infty$ and $A_i^{(I)} > 0$ for $1 \leq i \leq d_1$, where*

$$A^{(I)}_{\max} \coloneqq \max_{1\leq i\leq d_1} A_i^{(I)} < \frac{1}{2}, \tag{64}$$

such that

$$\left|\left(\prod_{j\in u}\frac{\partial^u}{\partial y_j}\right)\varphi^{(I)}_{h_\ell,M_\ell}(\boldsymbol{y})\right| \le \left(C_{\epsilon,d_2}M_\ell^{-1+\epsilon}+C_{\eta_{\mathrm{s},d_1}}h_\ell^{\eta_{\mathrm{s},d_1}}\right)b^{(I)}\prod_{i=1}^{d_1}\min(y_i,1-y_i)^{-A_i^{(I)}-\mathbb{1}_{\{i\in u\}}} \tag{65}$$

for all $\boldsymbol{y}=(y_1,\ldots,y_{d_1})\in(0,1)^{d_1}$ *and all* $u\subseteq\{1,\ldots,d_1\}$.

Assumption 8. *Let* $\varphi^{(II)}_{h_\ell,M_\ell}$, *for* $0\le\ell\le L$, *be the auxiliary function*

$$\varphi^{(II)}_{h_\ell,M_\ell}(\cdot)\coloneqq\left(\frac{1}{M_\ell}\sum_{m=1}^{M_\ell}g_{h_\ell}(\cdot,\boldsymbol{x}^{(m)})-\bar{g}(\cdot)\right)^2\int_0^1 f''\left(\bar{g}(\cdot)+s\left(\frac{1}{M_\ell}\sum_{m=1}^{M_\ell}g_{h_\ell}(\cdot,\boldsymbol{x}^{(m)})-\bar{g}(\cdot)\right)\right)(1-s)\,\mathrm{d}s. \tag{66}$$

Moreover, let $h_\ell>0$, $M_\ell\ge1$, $\bar{g}$ *as in* (26), *and* $\boldsymbol{x}^{(m)}$ *be as in* (43), *where* $1\le m\le M_\ell$, *and* $0\le\ell\le L$. *Then, assume that* $\varphi^{(II)}_{h_\ell,M_\ell}$, *for* $0\le\ell\le L$, *satisfies Assumption 1, that is, there exists* $0<b^{(II)}<\infty$ *and* $A_i^{(II)}>0$ *for* $1\le i\le d_1$, *where*

$$A^{(II)}_{\max}\coloneqq\max_{1\le i\le d_1}A_i^{(II)}<\frac{1}{2}, \tag{67}$$

such that

$$\left|\left(\prod_{j\in u}\frac{\partial^u}{\partial y_j}\right)\varphi^{(II)}_{h_\ell,M_\ell}(\boldsymbol{y})\right| \le \left(C^2_{\epsilon,d_2}M_\ell^{-2+2\epsilon}+C^2_{\eta_{\mathrm{s},d_1}}h_\ell^{2\eta_{\mathrm{s},d_1}}\right)b^{(II)}\prod_{i=1}^{d_1}\min(y_i,1-y_i)^{-A_i^{(II)}-\mathbb{1}_{\{i\in u\}}} \tag{68}$$

for all $\boldsymbol{y}=(y_1,\ldots,y_{d_1})\in(0,1)^{d_1}$ *and all* $u\subseteq\{1,\ldots,d_1\}$.

Proposition 4 (Variance of level ℓ for $1\le\ell\le L-1$ of the MLDLQMC estimator with inexact sampling)**.** *Let* $\varphi^{(I)}_{h_\ell,M_\ell}$ *as in* (63), *where* $0\le\ell\le L-1$, *satisfy Assumption 7 and* $\varphi^{(II)}_{h_\ell,M_\ell}$ *as in* (66), *where* $0\le\ell\le L-1$, *satisfy Assumption 8. The variance contribution of the* ℓ*th level of the MLDLQMC estimator* (52) *with inexact sampling, where* $1\le\ell\le L-1$, *then has the following upper bound:*

$$\begin{aligned}\mathbb{V}\left[I_{\mathrm{MLDLQ}^\ell_\ell}-I_{\mathrm{MLDLQ}^{\ell-1}_\ell}\right]\le{}&\frac{\left(\frac{C^2_{\epsilon,d_2}}{M_\ell^{2-2\epsilon}}+C^2_{\eta_{\mathrm{s},d_1}}h_\ell^{2\eta_{\mathrm{s},d_1}}\right)(b^{(I)})^2B^2_{A^{(I)}}C^2_{\epsilon,d_1}}{N_\ell^{2-2\epsilon-2A^{(I)}_{\max}}}\\&+\frac{\left(\frac{C^4_{\epsilon,d_2}}{M_\ell^{4-4\epsilon}}+C^4_{\eta_{\mathrm{s},d_1}}h_\ell^{4\eta_{\mathrm{s},d_1}}\right)(b^{(II)})^2B^2_{A^{(II)}}C^2_{\epsilon,d_1}}{N_\ell^{2-2\epsilon-2A^{(II)}_{\max}}},\end{aligned} \tag{69}$$

for any $\epsilon>0$, *where* $B_{A^{(I)}}\to\infty$ *as* $\min_{1\le i\le d_1}A_i^{(I)}\to0$, $B_{A^{(II)}}\to\infty$ *as* $\min_{1\le i\le d_1}A_i^{(II)}\to0$, *and* $C_{\epsilon,d_1},C_{\epsilon,d_2}\to\infty$ *as* $\epsilon\to0$.

Proof. The following derivations were performed for the choices of M_ℓ and h_ℓ detailed in (47) and (48), where $0\le\ell\le L-1$. We omitted the index $\ell\in\{0,\ldots,L-1\}$ for simplicity. The following Taylor expansion is considered for $1\le n\le N$:

$$\begin{aligned}f\left(\frac{1}{M}\sum_{m=1}^{M}g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)})\right)={}&f\left(\bar{g}(\boldsymbol{y}^{(n)})\right)+f'\left(\bar{g}(\boldsymbol{y}^{(n)})\right)\left(\frac{1}{M}\sum_{m=1}^{M}g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)})-\bar{g}(\boldsymbol{y}^{(n)})\right)\\&+\left(\frac{1}{M}\sum_{m=1}^{M}g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)})-\bar{g}(\boldsymbol{y}^{(n)})\right)^2\\&\times\int_0^1 f''\left(\bar{g}(\boldsymbol{y}^{(n)})+s\left(\frac{1}{M}\sum_{m=1}^{M}g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)})-\bar{g}(\boldsymbol{y}^{(n)})\right)\right)(1-s)\,\mathrm{d}s,\end{aligned} \tag{70}$$

and the expansion:

$$
\begin{aligned}
f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)})\right) &= f\left(\bar{g}(\boldsymbol{y}^{(n)})\right) + f'\left(\bar{g}(\boldsymbol{y}^{(n)})\right)\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right) \\
&\quad + \left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)^2 \\
&\quad \times \int_0^1 f''\left(\bar{g}(\boldsymbol{y}^{(n)}) + s\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)\right)(1-s)\,\mathrm{d}s. \qquad (71)
\end{aligned}
$$

Substituting the Taylor expansions (70) and (71) into the variance results in the following:

$$
\begin{aligned}
&\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N} f\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}\right)\right) - f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}\left(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}\right)\right)\right] \\
&= \mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N} f'\left(\bar{g}(\boldsymbol{y}^{(n)})\right)\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)\right. \\
&\quad - \frac{1}{N}\sum_{n=1}^{N} f'\left(\bar{g}(\boldsymbol{y}^{(n)})\right)\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) + \bar{g}(\boldsymbol{y}^{(n)})\right) \\
&\quad + \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)^2 \\
&\quad \times \int_0^1 f''\left(\bar{g}(\boldsymbol{y}^{(n)}) + s\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)\right)(1-s)\,\mathrm{d}s \\
&\quad - \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)^2 \\
&\quad \left.\times \int_0^1 f''\left(\bar{g}(\boldsymbol{y}^{(n)}) + s\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)\right)(1-s)\,\mathrm{d}s\right]. \qquad (72)
\end{aligned}
$$

Applying the Cauchy–Schwarz inequality (see [2, Appendix A]) to the above variance yields the following:

$$
\begin{aligned}
&\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N} f\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}\right)\right) - f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}\left(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}\right)\right)\right] \\
&\le 4\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N} f'\left(\bar{g}(\boldsymbol{y}^{(n)})\right)\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)\right] \\
&+ 4\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N} f'\left(\bar{g}(\boldsymbol{y}^{(n)})\right)\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) + \bar{g}(\boldsymbol{y}^{(n)})\right)\right] \\
&+ 4\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)^2\right. \\
&\qquad \left.\times\int_0^1 f''\left(\bar{g}(\boldsymbol{y}^{(n)}) + s\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)\right)(1-s)\,\mathrm{d}s\right] \\
&+ 4\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)^2\right. \\
&\qquad \left.\times\int_0^1 f''\left(\bar{g}(\boldsymbol{y}^{(n)}) + s\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}) - \bar{g}(\boldsymbol{y}^{(n)})\right)\right)(1-s)\,\mathrm{d}s\right]. \qquad (73)
\end{aligned}
$$

Taking

$$
\varphi_{h,M}^{(I)}(\cdot) \coloneqq f'\left(\bar{g}(\cdot)\right)\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}(\cdot)\right), \qquad (74)
$$

it follows from Assumption 7 that

$$
\left|\left(\prod_{j\in u}\frac{\partial^u}{\partial y_j}\right)\varphi_{h,M}^{(I)}(\boldsymbol{y})\right| \le \left(C_{\epsilon,d_2}M^{-1+\epsilon} + C_{\eta_{\mathrm{s},d_1}}h^{\eta_{\mathrm{s},d_1}}\right)\tilde{b}^{(I)}\prod_{i=1}^{d_1}\min(y_i,1-y_i)^{-A_i^{(I)}-\mathbb{1}_{\{i\in u\}}} \qquad (75)
$$

for all $\boldsymbol{y} = (y_1,\ldots,y_{d_1}) \in (0,1)^{d_1}$ and all $u \subseteq \{1,\ldots,d_1\}$. Proposition 1 then provides the following bound for the first term in (73):

$$
\begin{aligned}
4\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi_{h,M}^{(I)}(\boldsymbol{y}^{(n)})\right] &\le 4\left(C_{\epsilon,d_2}M^{-1+\epsilon} + C_{\eta_{\mathrm{s},d_1}}h^{\eta_{\mathrm{s},d_1}}\right)^2(\tilde{b}^{(I)})^2 B_{A^{(I)}}^2 C_{\epsilon,d_1}^2 N^{-2+2\epsilon+2A_{\max}^{(I)}}, \\
&\le 8\left(C_{\epsilon,d_2}^2M^{-2+2\epsilon} + C_{\eta_{\mathrm{s},d_1}}^2h^{2\eta_{\mathrm{s},d_1}}\right)(\tilde{b}^{(I)})^2 B_{A^{(I)}}^2 C_{\epsilon,d_1}^2 N^{-2+2\epsilon+2A_{\max}^{(I)}}, \qquad (76)
\end{aligned}
$$

for all $\epsilon > 0$, where $B_{A^{(I)}} \to \infty$ as $\min_{1\le i\le d_1} A_i^{(I)} \to 0$ and $C_{\epsilon,d_1}, C_{\epsilon,d_2} \to \infty$ as $\epsilon \to 0$. Taking

$$
\varphi_{2h,M/2}^{(I)}(\cdot) \coloneqq f'\left(\bar{g}(\cdot)\right)\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}(\cdot)\right), \qquad (77)
$$

a similar bound follows for the second term in (73):

$$
4\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi_{2h,M/2}^{(I)}(\boldsymbol{y}^{(n)})\right] \le 8\left(C_{\epsilon,d_2}^2 2^{2-2\epsilon}M^{-2+2\epsilon} + C_{\eta_{\mathrm{s},d_1}}^2 2^{2\eta_{\mathrm{s},d_1}}h^{2\eta_{\mathrm{s},d_1}}\right)(\tilde{b}^{(I)})^2 B_{A^{(I)}}^2 C_{\epsilon,d_1}^2 N^{-2+2\epsilon+2A_{\max}^{(I)}}. \qquad (78)
$$

Next, taking

$$\varphi_{h,M}^{(II)}(\cdot) \coloneqq \left(\frac{1}{M}\sum_{m=1}^{M} g_h(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}(\cdot)\right)^2 \int_0^1 f''\left(\bar{g}(\cdot) + s\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}(\cdot)\right)\right)(1-s)\,\mathrm{d}s, \tag{79}$$

it follows from Assumption 8 that

$$\left|\left(\prod_{j\in u}\frac{\partial^u}{\partial y_j}\right)\varphi_{h,M}^{(II)}(\boldsymbol{y})\right| \le \left(C_{\epsilon,d_2}^2 M^{-2+2\epsilon} + C_{\eta_{\mathrm{s},d_1}}^2 h^{2\eta_{\mathrm{s},d_1}}\right)\tilde{b}^{(II)}\prod_{i=1}^{d_1}\min(y_i,1-y_i)_i^{-A^{(II)}} - \mathbb{1}_{\{i\in u\}} \tag{80}$$

for all $\boldsymbol{y} = (y_1,\ldots,y_{d_1}) \in (0,1)^{d_1}$ and all $u \subseteq \{1,\ldots,d_1\}$. Proposition 1 then provides the following bound for the third term in (73):

$$4\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi_{h,M}^{(II)}(\boldsymbol{y}^{(n)})\right] \le 8\left(C_{\epsilon,d_2}^4 M^{-4+4\epsilon} + C_{\eta_{\mathrm{s},d_1}}^4 h^{4\eta_{\mathrm{s},d_1}}\right)(\tilde{b}^{(II)})^2 B_{A^{(II)}}^2 C_{\epsilon,d_1}^2 N^{-2+2\epsilon+2A_{\max}^{(II)}}, \tag{81}$$

and finally, taking

$$\varphi_{2h,M/2}^{(II)}(\cdot) \coloneqq \left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}(\cdot)\right)^2 \int_0^1 f''\left(\bar{g}(\cdot) + s\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}(\cdot)\right)\right)(1-s)\,\mathrm{d}s, \tag{82}$$

a similar bound applies to the fourth term in (73):

$$4\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N}\varphi_{2h,M/2}^{(II)}(\boldsymbol{y}^{(n)})\right] \le 8\left(C_{\epsilon,d_2}^4 2^{4-4\epsilon} M^{-4+4\epsilon} + C_{\eta_{\mathrm{s},d_1}}^4 2^{4\eta_{\mathrm{s},d_1}} h^{4\eta_{\mathrm{s},d_1}}\right)(\tilde{b}^{(II)})^2 B_{A^{(II)}}^2 C_{\epsilon,d_1}^2 N^{-2+2\epsilon+2A_{\max}^{(II)}}, \tag{83}$$

for all $\epsilon > 0$, where $B_{A^{(II)}} \to \infty$ as $\min_{1\le i\le d_1} A_i^{(II)} \to 0$, and $C_{\epsilon,d_1}, C_{\epsilon,d_2} \to \infty$ as $\epsilon \to 0$. Combining the above results yields the following:

$$\begin{aligned}\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N} f\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}\right)\right) - f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}} g_{2h}\left(\boldsymbol{y}^{(n)},\boldsymbol{x}^{(m)}\right)\right)\right] \\ \le \frac{\left(\frac{C_{\epsilon,d_2}^2}{M^{2-2\epsilon}} + C_{\eta_{\mathrm{s},d_1}}^2 h^{2\eta_{\mathrm{s},d_1}}\right)(b^{(I)})^2 B_{A^{(I)}}^2 C_{\epsilon,d_1}^2}{N^{2-2\epsilon-2A_{\max}^{(I)}}} \\ + \frac{\left(\frac{C_{\epsilon,d_2}^4}{M^{4-4\epsilon}} + C_{\eta_{\mathrm{s},d_1}}^4 h^{4\eta_{\mathrm{s},d_1}}\right)(b^{(II)})^2 B_{A^{(II)}}^2 C_{\epsilon,d_1}^2}{N^{2-2\epsilon-2A_{\max}^{(II)}}},\end{aligned} \tag{84}$$

where

$$(b^{(I)})^2 \coloneqq \frac{8(\tilde{b}^{(I)})^2}{\max\{2^{2-2\epsilon}, 2^{2\eta_{\mathrm{s},d_1}}\}+1} \tag{85}$$

and

$$(b^{(II)})^2 \coloneqq \frac{8(\tilde{b}^{(II)})^2}{\max\{2^{4-4\epsilon}, 2^{4\eta_{\mathrm{s},d_1}}\}+1}, \tag{86}$$

for all $\epsilon > 0$, where $B_{A^{(I)}} \to \infty$ as $\min_{1\le i\le d_1} A_i^{(I)} \to 0$, $B_{A^{(II)}} \to \infty$ as $\min_{1\le i\le d_1} A_i^{(II)} \to 0$, and $C_{\epsilon,d_1}, C_{\epsilon,d_2} \to \infty$ as $\epsilon \to 0$. □

A base other than two in (47) and (48) results in different multiplicative constants $b^{(I)}$ and $b^{(II)}$.

Remark 2 (Difference between terms in the variance bound)**.** *If it holds that $A_{\max}^{(I)} \le A_{\max}^{(II)}$, then the second term in the variance bound* (69) *is of higher order compared to the first term.*

For the bound of the variance of the MLDLQMC estimator at the final level L, the following additional assumptions are made regarding the variance from the inner randomization:

Assumption 9. *Let* $\varphi^{(III)}_{h_L,M_L}$ *be an auxiliary function, where*

$$\varphi^{(III)}_{h_L,M_L}(\cdot) \coloneqq f'\left(\bar{g}_{h_L}(\cdot)\right)\left(\frac{1}{M_L}\sum_{m=1}^{M_L} g_{h_L}(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}_h(\cdot)\right). \tag{87}$$

Moreover, let $h_L > 0$, $M_L \geq 1$,

$$\bar{g}_{h_L}(\cdot) \coloneqq \int_{[0,1]^{d_2}} g_{h_L}(\cdot,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{88}$$

and $\boldsymbol{x}^{(m)}$ *be as in* (43)*, where* $1 \leq m \leq M_L$*. Then, assume that there exists* $0 < b^{(III)} < \infty$ *and* $A_i^{(III)} > 0$ *for* $1 \leq i \leq d_1$*, where*

$$A^{(III)}_{\max} \coloneqq \max_{1\leq i\leq d_1} A_i^{(III)} < \frac{1}{2}, \tag{89}$$

such that

$$\left|\varphi^{(III)}_{h_L,M_L}(\boldsymbol{y})\right| \leq C_{\epsilon,d_2} M_L^{-1+\epsilon} b^{(III)} \prod_{i=1}^{d_1} \min(y_i, 1-y_i)^{-A_i^{(III)}} \tag{90}$$

for all $\boldsymbol{y} = (y_1,\ldots,y_{d_1}) \in (0,1)^{d_1}$*, and all* $u \subseteq \{1,\ldots,d_1\}$*, where* $b^{(III)}$ *implicitly depends on* h_L*, and* η_{s,d_1}*.*

Assumption 10. *Let* $\varphi^{(IV)}_{h_L,M_L}$ *be an auxiliary function, where*

$$\varphi^{(IV)}_{h_L,M_L}(\cdot) \coloneqq \left(\frac{1}{M_L}\sum_{m=1}^{M_L} g_{h_L}(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}_{h_L}(\cdot)\right)^2 \int_0^1 f''\left(\bar{g}_{h_L}(\cdot) + s\left(\frac{1}{M_L}\sum_{m=1}^{M_L} g_{h_L}(\cdot,\boldsymbol{x}^{(m)}) - \bar{g}_{h_L}(\cdot)\right)\right)(1-s)\,\mathrm{d}s. \tag{91}$$

Moreover, let $h_L > 0$*,* $M_L \geq 1$*,* $\bar{g}_{h_L}$ *as in* (88)*, and* $\boldsymbol{x}^{(m)}$ *be as in* (43)*, where* $1 \leq m \leq M_L$*. Then, assume that there exists* $0 < b^{(IV)} < \infty$ *and* $A_i^{(IV)} > 0$ *for* $1 \leq i \leq d_1$*, where*

$$A^{(IV)}_{\max} \coloneqq \max_{1\leq i\leq d_1} A_i^{(IV)} < \frac{1}{2}, \tag{92}$$

such that

$$\left|\varphi^{(IV)}_{h_L,M_L}(\boldsymbol{y})\right| \leq C^2_{\epsilon,d_2} M_L^{-2+2\epsilon} b^{(IV)} \prod_{i=1}^{d_1} \min(y_i, 1-y_i)^{-A_i^{(IV)}} \tag{93}$$

for all $\boldsymbol{y} = (y_1,\ldots,y_{d_1}) \in (0,1)^{d_1}$ *and all* $u \subseteq \{1,\ldots,d_1\}$*, where* $b^{(IV)}$ *implicitly depends on* h_L *and* η_{s,d_1}*.*

Proposition 5 (Variance of level L of the MLDLQMC estimator with inexact sampling)**.** *Let* $\varphi^{(I)}_{h_{L-1},M_{L-1}}$ *and* $\varphi^{(I)}_{h_L,M_L}$ *as in* (63) *satisfy Assumption 7,* $\varphi^{(II)}_{h_{L-1},M_{L-1}}$ *and* $\varphi^{(II)}_{h_L,M_L}$ *as in* (66) *satisfy Assumption 8,* $\varphi^{(III)}_{h_L,M_L}$ *as in* (87) *satisfy Assumption 9, and* $\varphi^{(IV)}_{h_L,M_L}$ *as in* (91) *satisfy Assumption 10. The variance contribution of the Lth level of the MLDLQMC estimator* (52) *with inexact sampling then has the following upper bound:*

$$\begin{aligned}\mathbb{V}\left[I_{\mathrm{MLDLQ}_L^L} - I_{\mathrm{MLDLQ}_L^{L-1}}\right] \leq{}& \frac{\left(\frac{C^2_{\epsilon,d_2}}{M_L^{2-2\epsilon}} + C^2_{\eta_{\mathrm{s},d_1}} h_L^{2\eta_{\mathrm{s},d_1}}\right)(b^{(I)})^2 B^2_{A^{(I)}} C^2_{\epsilon,d_1}}{N_L^{2-2\epsilon-2A^{(I)}_{\max}}} \\ &+ \frac{\left(\frac{C^4_{\epsilon,d_2}}{M_L^{4-4\epsilon}} + C^4_{\eta_{\mathrm{s},d_1}} h_L^{4\eta_{\mathrm{s},d_1}}\right)(b^{(II)})^2 B^2_{A^{(II)}} C^2_{\epsilon,d_1}}{N_L^{2-2\epsilon-2A^{(II)}_{\max}}} \\ &+ \frac{C^2_{\epsilon,d_2}(b^{(III)})^2 B^2_{A^{(III)}}}{M_L^{2-2\epsilon}} + \frac{C^4_{\epsilon,d_2}(b^{(IV)})^2 B^2_{A^{(IV)}}}{M_L^{4-4\epsilon}},\end{aligned} \tag{94}$$

for any $\epsilon > 0$*, where* $B_{A^{(I)}} \to \infty$ *as* $\min_{1\leq i\leq d_1} A_i^{(I)} \to 0$*,* $B_{A^{(II)}} \to \infty$ *as* $\min_{1\leq i\leq d_1} A_i^{(II)} \to 0$*,* $B_{A^{(III)}} \to \infty$ *as* $\min_{1\leq i\leq d_1} A_i^{(III)} \to 0$*,* $B_{A^{(IV)}} \to \infty$ *as* $\min_{1\leq i\leq d_1} A_i^{(IV)} \to 0$*, and* $C_{\epsilon,d_1}, C_{\epsilon,d_2} \to \infty$ *as* $\epsilon \to 0$*.*

Proof. The following derivations were performed for the specific choices of M_{L-1}, M_L, h_{L-1}, and h_L detailed in (47) and (48). We omitted the index L for simplicity. The law of total variance is applied to obtain

$$\begin{aligned}
&\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N}\left(f\left(\frac{1}{M}\sum_{m=1}^{M}g_h\left(\boldsymbol{\tau}(\boldsymbol{u}^{(n)}),\boldsymbol{\rho}(\boldsymbol{t}^{(m)})\right)\right)-f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}}g_{2h}\left(\boldsymbol{\tau}(\boldsymbol{u}^{(n)}),\boldsymbol{t}^{(m)}\right)\right)\right)\right]\\
&\quad\le\mathbb{V}\left[\mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}\left(f\left(\frac{1}{M}\sum_{m=1}^{M}g_h\left(\boldsymbol{\tau}(\boldsymbol{u}^{(n)}),\boldsymbol{\rho}(\boldsymbol{t}^{(m)})\right)\right)-f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}}g_{2h}\left(\boldsymbol{\tau}(\boldsymbol{u}^{(n)}),\boldsymbol{t}^{(m)}\right)\right)\right)\Bigg|\boldsymbol{\rho}\right]\right]\\
&\quad+\mathbb{E}\left[\mathbb{V}\left[\frac{1}{N}\sum_{n=1}^{N}\left(f\left(\frac{1}{M}\sum_{m=1}^{M}g_h\left(\boldsymbol{\tau}(\boldsymbol{u}^{(n)}),\boldsymbol{\rho}(\boldsymbol{t}^{(m)})\right)\right)-f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}}g_{2h}\left(\boldsymbol{\tau}(\boldsymbol{u}^{(n)}),\boldsymbol{t}^{(m)}\right)\right)\right)\Bigg|\boldsymbol{\rho}\right]\right],
\end{aligned}\tag{95}$$

where the randomization $\boldsymbol{\tau}$ maps the low-discrepancy points $\boldsymbol{u}^{(n)}$, where $1\le n\le N$, to a low-discrepancy point set with probability one, and the randomization $\boldsymbol{\rho}$ maps the low-discrepancy points $\boldsymbol{t}^{(m)}$, where $1\le m\le M$, to a low-discrepancy point set with probability one; thus, the variance bound of the second term in (95) is analogous to the bound for level ℓ where $1\le\ell\le L-1$. For the first term in (95), it holds that

$$\begin{aligned}
&\mathbb{V}\left[\mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}\left(f\left(\frac{1}{M}\sum_{m=1}^{M}g_h\left(\boldsymbol{\tau}(\boldsymbol{u}^{(n)}),\boldsymbol{\rho}(\boldsymbol{t}^{(m)})\right)\right)-f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}}g_{2h}\left(\boldsymbol{\tau}(\boldsymbol{u}^{(n)}),\boldsymbol{t}^{(m)}\right)\right)\right)\Bigg|\boldsymbol{\rho}\right]\right]\\
&\quad=\mathbb{V}\left[\int_{[0,1]^{d_1}}\left(f\left(\frac{1}{M}\sum_{m=1}^{M}g_h\left(\boldsymbol{y},\boldsymbol{\rho}(\boldsymbol{t}^{(m)})\right)\right)-f\left(\frac{1}{\frac{M}{2}}\sum_{m=1}^{\frac{M}{2}}g_{2h}\left(\boldsymbol{y},\boldsymbol{t}^{(m)}\right)\right)\right)\mathrm{d}\boldsymbol{y}\right],\\
&\quad=\mathbb{V}\left[\int_{[0,1]^{d_1}}f\left(\frac{1}{M}\sum_{m=1}^{M}g_h\left(\boldsymbol{y},\boldsymbol{\rho}(\boldsymbol{t}^{(m)})\right)\right)\mathrm{d}\boldsymbol{y}\right].
\end{aligned}\tag{96}$$

The following Taylor expansion is considered:

$$\begin{aligned}
f\left(\frac{1}{M}\sum_{m=1}^{M}g_h(\boldsymbol{y},\boldsymbol{x}^{(m)})\right)&=f\left(\bar{g}_h(\boldsymbol{y})\right)+f'\left(\bar{g}_h(\boldsymbol{y})\right)\left(\frac{1}{M}\sum_{m=1}^{M}g_h(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}_h(\boldsymbol{y})\right)\\
&\quad+\left(\frac{1}{M}\sum_{m=1}^{M}g_h(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}_h(\boldsymbol{y})\right)^2\\
&\quad\times\int_0^1 f''\left(\bar{g}_h(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M}g_h(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}_h(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s,
\end{aligned}\tag{97}$$

to obtain

$$
\begin{aligned}
&\mathbb{V}\left[\int_{[0,1]^{d_1}} f\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y},\boldsymbol{x}^{(m)}\right)\right)\mathrm{d}\boldsymbol{y}\right]\\
&=\mathbb{V}\left[\int_{[0,1]^{d_1}} f\left(\bar{g}_h(\boldsymbol{y})\right)+f'\left(\bar{g}_h(\boldsymbol{y})\right)\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y},\boldsymbol{x}^{(m)}\right)-\bar{g}_h(\boldsymbol{y})\right)\right.\\
&\qquad\left.+\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y},\boldsymbol{x}^{(m)}))-\bar{g}_h(\boldsymbol{y})\right)^2\int_0^1 f''\left(\bar{g}_h(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}_h(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s\,\mathrm{d}\boldsymbol{y}\right],\\
&\le 2\mathbb{V}\left[\int_{[0,1]^{d_1}} f'\left(\bar{g}_h(\boldsymbol{y})\right)\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y},\boldsymbol{x}^{(m)}\right)-\bar{g}_h(\boldsymbol{y})\right)\mathrm{d}\boldsymbol{y}\right]\\
&\quad+2\mathbb{V}\left[\int_{[0,1]^{d_1}}\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}_h(\boldsymbol{y})\right)^2\right.\\
&\qquad\left.\times\int_0^1 f''\left(\bar{g}_h(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}_h(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s\,\mathrm{d}\boldsymbol{y}\right].
\end{aligned}
\tag{98}
$$

Taking

$$
\varphi_{h,M}^{(III)}(\cdot)\coloneqq f'\left(\bar{g}_h(\cdot)\right)\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\cdot,\boldsymbol{x}^{(m)}\right)-\bar{g}_h(\cdot)\right),
\tag{99}
$$

it follows from the Cauchy–Schwarz inequality and Assumption 9 that the first term in (98) is bounded as follows:

$$
\begin{aligned}
2\mathbb{V}\left[\int_{[0,1]^{d_1}}\varphi_{h,M}^{(III)}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right]&\le 2\mathbb{E}\left[\left|\int_{[0,1]^{d_1}}\varphi_{h,M}^{(III)}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right|^2\right],\\
&\le 2\left|\int_{[0,1]^{d_1}} C_{\epsilon,d_2}M^{-1+\epsilon}\tilde{b}^{(III)}\prod_{i=1}^{d_1}\min(y_i,1-y_i)^{-A_i^{(III)}}\,\mathrm{d}\boldsymbol{y}\right|^2,\\
&= 2C_{\epsilon,d_2}^2M^{-2+2\epsilon}(\tilde{b}^{(III)})^2\left|\int_{[0,1]^{d_1}}\prod_{i=1}^{d_1}\min(y_i,1-y_i)^{-A_i^{(III)}}\,\mathrm{d}\boldsymbol{y}\right|^2,\\
&= \frac{C_{\epsilon,d_2}^2(b^{(III)})^2B_{A^{(III)}}^2}{M^{2-2\epsilon}},
\end{aligned}
\tag{100}
$$

where

$$
(b^{(III)})^2\coloneqq 2(\tilde{b}^{(III)})^2,
\tag{101}
$$

and

$$
B_{A^{(III)}}\coloneqq\int_{[0,1]^{d_1}}\prod_{i=1}^{d_1}\min(y_i,1-y_i)^{-A_i^{(III)}}\,\mathrm{d}\boldsymbol{y},
\tag{102}
$$

for any $\epsilon>0$, where $C_{\epsilon,d_2}\to\infty$ as $\epsilon\to 0$ and $B_{A^{(III)}}\to\infty$ as $\min_{1\le i\le d_1}A_i^{(III)}\to 0$. Finally, taking

$$
\varphi_{h,M}^{(IV)}(\cdot)\coloneqq\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\cdot,\boldsymbol{x}^{(m)})-\bar{g}_h(\cdot)\right)^2\int_0^1 f''\left(\bar{g}_h(\cdot)+s\left(\frac{1}{M}\sum_{m=1}^{M} g_h(\cdot,\boldsymbol{x}^{(m)})-\bar{g}_h(\cdot)\right)\right)(1-s)\,\mathrm{d}s,
\tag{103}
$$

it follows from the Cauchy–Schwarz inequality and Assumption 10 that the second term in (98) is bounded as follows:

$$\begin{aligned}
2\mathbb{V}\left[\int_{[0,1]^{d_1}} \varphi_{h,M}^{(IV)}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right] &\le 2\mathbb{E}\left[\left|\int_{[0,1]^{d_1}} \varphi_{h,M}^{(IV)}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right|^2\right], \\
&\le 2\left|\int_{[0,1]^{d_1}} C_{\epsilon,d_2}^2 M^{-2+2\epsilon}\tilde{b}^{(IV)} \prod_{i=1}^{d_1} \min(y_i, 1-y_i)^{-A_i^{(IV)}}\,\mathrm{d}\boldsymbol{y}\right|^2, \\
&= 2C_{\epsilon,d_2}^4 M^{-4+4\epsilon}(\tilde{b}^{(IV)})^2 \left|\int_{[0,1]^{d_1}} \prod_{i=1}^{d_1} \min(y_i, 1-y_i)^{-A_i^{(IV)}}\,\mathrm{d}\boldsymbol{y}\right|^2, \\
&= \frac{C_{\epsilon,d_2}^4 (b^{(IV)})^2 B_{A^{(IV)}}^2}{M^{4-4\epsilon}}, \qquad (104)
\end{aligned}$$

where

$$(b^{(IV)})^2 \coloneqq 2(\tilde{b}^{(IV)})^2, \qquad (105)$$

and

$$B_{A^{(IV)}} \coloneqq \int_{[0,1]^{d_1}} \prod_{i=1}^{d_1} \min(y_i, 1-y_i)^{-A_i^{(IV)}}\,\mathrm{d}\boldsymbol{y}, \qquad (106)$$

for any $\epsilon > 0$, where $C_{\epsilon,d_2} \to \infty$ as $\epsilon \to 0$ and $B_{A^{(IV)}} \to \infty$ as $\min_{1\le i\le d_1} A_i^{(IV)} \to 0$. □

A base other than two in (47) and (48) results in different multiplicative constants $b^{(I)}$, $b^{(II)}$, $b^{(III)}$, and $b^{(IV)}$.

Remark 3 (Optimal choices of M_ℓ and h_ℓ)**.** *It follows from Propositions 2, 4, and 5 that the choices (see* (47) *and* (48)*)*

$$M_\ell \coloneqq M_0 2^\ell, \quad 1 \le \ell \le L, \qquad (107)$$

and

$$h_\ell \coloneqq h_0 2^{-\ell}, \quad 1 \le \ell \le L, \qquad (108)$$

may be suboptimal if η_{w} differs from $1-\epsilon$. In that case, we select

$$M_\ell \coloneqq M_0 \left(2^{\frac{\eta_{\mathrm{w}}}{1-\epsilon}}\right)^\ell, \quad 1 \le \ell \le L. \qquad (109)$$

It must be ensured that M_ℓ, where $1 \le \ell \le L$, are powers of two by rounding up appropriately to maintain the properties of the Sobol' sequence.

Remark 4 (Extension of the rDLQMC estimator to the multilevel setting)**.** *It becomes apparent from Propositions 3–5 that the particular structure of the MLDLQMC estimator* (52) *(deterministic inner samples at levels $0 \le \ell \le L-1$, and randomized inner samples at level L) is crucial. Adapting the rDLQMC estimator* [2] *to the multilevel setting would deteriorate the convergence of the variances at levels $0 \le \ell \le L-1$ to rate N rather than rate $N^{2-2\epsilon-2A_{\max}}$. Instead, randomizing the inner samples in the last level does not affect the overall estimator complexity and provides a practical error estimate for the bias.*

Remark 5 (Antithetic sampling)**.** *The coupling in multilevel estimators can be increased by an antithetic construction as in the works* [38, 30]*; however, the MLDLQMC estimator does not benefit significantly from such an antithetic construction because of the combination of deterministic and randomized inner samples.*

Proposition 6 (Optimal work of the MLDLQMC estimator with inexact sampling)**.** *Given the assumptions of Propositions 2–5, the total computational work of the MLDLQMC estimator* (52) *with inexact sampling*

is given by

$$
\begin{aligned}
W^*_{\mathrm{MLDLQ}} &= \mathcal{O}\left(TOL^{-\left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_{\mathrm{w}}}\right)}\right) \\
&+ \begin{cases}
\mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}}\right), & \text{if} \quad \frac{\eta_{\mathrm{s},d_1}}{\eta_{\mathrm{w}}(1-\epsilon-A_{\max})} > \left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_{\mathrm{w}}}\right), \\
\mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}}\log_2\left(TOL^{-\max\left\{\frac{1}{\eta_{\mathrm{w}}},\frac{1}{1-\epsilon}\right\}}\right)\right), & \text{if} \quad \frac{\eta_{\mathrm{s},d_1}}{\eta_{\mathrm{w}}(1-\epsilon-A_{\max})} = \left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_{\mathrm{w}}}\right), \\
\mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}+\frac{\eta_{\mathrm{s},d_1}}{\eta_{\mathrm{w}}(1-\epsilon-A_{\max})}-\left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_{\mathrm{w}}}\right)}\right), & \text{if} \quad \frac{\eta_{\mathrm{s},d_1}}{\eta_{\mathrm{w}}(1-\epsilon-A_{\max})} < \left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_{\mathrm{w}}}\right),
\end{cases}
\end{aligned} \tag{110}
$$

where

$$
A_{\max} \coloneqq \max\left\{A^{(0)}_{\max}, A^{(I)}_{\max}, A^{(II)}_{\max}\right\}, \tag{111}
$$

with $A^{(0)}_{\max}$ *as in* (60), $A^{(I)}_{\max}$ *as in* (64), *and* $A^{(II)}_{\max}$ *as in* (67), *for any* $\epsilon > 0$ *as* $TOL \to 0$, *where the multiplicative constant in the* $\mathcal{O}$ *notation approaches infinity as* $\epsilon \to 0$.

Proof. The total computational work of the MLDLQMC estimator to estimate a nested integral I is

$$
W_{\mathrm{MLDLQ}} = \sum_{\ell=0}^{L} N_\ell M_\ell h_\ell^{-\gamma} + \sum_{\ell=0}^{L} M_\ell h_\ell^{-\gamma}. \tag{112}
$$

To achieve an error tolerance $TOL \geq 0$, it is required that

$$
|I_{\mathrm{MLDLQ}} - I| \leq TOL. \tag{113}
$$

From (53) and (57) the following constraints are derived:

$$
|\mathbb{E}[I_{\mathrm{MLDLQ}}] - I| \leq \frac{1}{2}TOL \tag{114}
$$

and

$$
\mathbb{V}\left[I_{\mathrm{MLDLQ}}\right] \leq \left(\frac{\frac{1}{2}TOL}{C_\alpha}\right)^2. \tag{115}
$$

We thus solve the optimization problem

$$
\begin{aligned}
(L^*, N_0^*, \ldots, N_{L^*}) = \underset{(L,N_0,\ldots,N_L)}{\arg\min} & \sum_{\ell=0}^{L} N_\ell M_\ell h_\ell^{-\gamma} + \sum_{\ell=0}^{L} M_\ell h_\ell^{-\gamma} \\
\text{subject to} & \begin{cases}
\frac{b^2 B_A^2 C^2_{\epsilon,d_1}}{N_0^{2-2\epsilon-2A_{\max}}} + \sum_{\ell=1}^{L-1} \frac{\left(\frac{C^2_{\epsilon,d_2}}{M_\ell^{2-2\epsilon}} + C^2_{\eta_{\mathrm{s},d_1}} h_\ell^{2\eta_{\mathrm{s},d_1}}\right)(b^{(I)})^2 B_A^2 C^2_{\epsilon,d_1}}{N_\ell^{2-2\epsilon-2A_{\max}}} + \frac{C^2_{\eta_{\mathrm{s},d_1}} h_L^{2\eta_{\mathrm{s},d_1}} (b^{(I)})^2 B_A^2 C^2_{\epsilon,d_1}}{N_L^{2-2\epsilon-2A_{\max}}} \leq \frac{1}{2}\left(\frac{\frac{1}{2}TOL}{C_\alpha}\right)^2, \\
\frac{C^2_{\epsilon,d_2}(b^{(III)})^2 B_A^2}{M_L^{2-2\epsilon}} \leq \frac{1}{2}\left(\frac{\frac{1}{2}TOL}{C_\alpha}\right)^2, \\
C_{\mathrm{w}} h_L^{\eta_{\mathrm{w}}} \leq \frac{1}{4}TOL, \\
\frac{\frac{1}{2}\mathbb{E}[|\bar{g}|^2|f''(\bar{g})|]k^2 C^2_{\epsilon,d_2}}{M_L^{2-2\epsilon}} \leq \frac{1}{4}TOL,
\end{cases}
\end{aligned} \tag{116}
$$

where higher-order terms in the bounds introduced by Propositions 2 to 5 were ignored. It immediately follows that the constraint

$$
\frac{\frac{1}{2}\mathbb{E}[|\bar{g}|^2|f''(\bar{g})|]k^2 C^2_{\epsilon,d_2}}{M_L^{2-2\epsilon}} \leq \frac{1}{4}TOL \tag{117}
$$

is satisfied whenever the constraint

$$
\frac{C^2_{\epsilon,d_2}(b^{(III)})^2 B_A^2}{M_L^{2-2\epsilon}} \leq \frac{1}{2}\left(\frac{\frac{1}{2}TOL}{C_\alpha}\right)^2 \tag{118}
$$

is satisfied as $TOL \to 0$ and the former is thus ignored in the following derivation. The Lagrangian

$$\begin{aligned}\mathcal{L}(L, N_0, \ldots, N_L, \lambda, \mu, \nu) = &\sum_{\ell=0}^{L} N_\ell W_\ell + \sum_{\ell=0}^{L} W_\ell + \lambda \left(\sum_{\ell=0}^{L} \frac{D_\ell}{N_\ell^{2-2\epsilon-2A_{\max}}} - \frac{1}{2}\left(\frac{\frac{1}{2}TOL}{C_\alpha}\right)^2 \right) \\ &+ \mu \left(\frac{C_{\epsilon,d_2}^2 (b^{(III)})^2 B_A^2}{M_L^{2-2\epsilon}} - \frac{1}{2}\left(\frac{\frac{1}{2}TOL}{C_\alpha}\right)^2 \right) + \nu \left(C_{\mathrm{w}} h_L^{\eta_{\mathrm{w}}} - \frac{1}{4}TOL \right), \end{aligned} \tag{119}$$

is introduced, where

$$W_\ell := M_\ell h_\ell^{-\gamma}, \quad 0 \le \ell \le L, \tag{120}$$

and

$$D_\ell := \begin{cases} b^2 B_A^2 C_{\epsilon,d_1}^2, & \ell = 0 \\ \left(\frac{C_{\epsilon,d_2}^2}{M_\ell^{2-2\epsilon}} + C_{\eta_{\mathrm{s},d_1}}^2 h_\ell^{2\eta_{\mathrm{s},d_1}} \right) (b^{(I)})^2 B_A^2 C_{\epsilon,d_1}^2, & 1 \le \ell \le L-1 \\ C_{\eta_{\mathrm{s},d_1}}^2 h_L^{2\eta_{\mathrm{s},d_1}} (b^{(I)})^2 B_A^2 C_{\epsilon,d_1}^2, & L. \end{cases} \tag{121}$$

Taking the derivatives with respect to μ and ν and setting them to zero yields the optimal L^* as follows:

$$L^* = \max \left\{ \log_2 \left(\left(\frac{4C_{\mathrm{w}} h_0^{\eta_{\mathrm{w}}}}{TOL} \right)^{\frac{1}{\eta_{\mathrm{w}}}} \right), \log_2 \left(\left(\frac{2\sqrt{2} b^{(III)} B_A C_\alpha \frac{C_{\epsilon,d_2}}{M_0^{1-\epsilon}}}{TOL} \right)^{\frac{1}{1-\epsilon}} \right) \right\}. \tag{122}$$

Next, taking the derivative with respect to N_0 yields

$$\lambda^* = N_0^{3-2\epsilon-2A_{\max}} \frac{W_0}{D_0(2-2\epsilon-2A_{\max})}. \tag{123}$$

Taking the derivatives with respect to N_ℓ, where $1 \le \ell \le L^*$, yields:

$$\begin{aligned} N_\ell^{3-2\epsilon-2A_{\max}} &= \lambda^* \frac{D_\ell (2-2\epsilon-2A_{\max})}{W_\ell}, \\ &= \frac{D_\ell W_0}{D_0 W_\ell} N_0^{3-2\epsilon-2A_{\max}}, \quad 1 \le \ell \le L^*. \end{aligned} \tag{124}$$

Taking the derivative with respect to λ results in

$$\begin{aligned} \frac{1}{2}\left(\frac{\frac{1}{2}TOL}{C_\alpha}\right)^2 &= \sum_{\ell=0}^{L^*} \frac{D_\ell}{N_\ell^{2-2\epsilon-2A_{\max}}}, \\ &= \sum_{\ell=0}^{L^*} \frac{D_\ell}{\left(\frac{D_\ell W_0}{D_0 W_\ell}\right)^{\frac{2-2\epsilon-2A_{\max}}{3-2\epsilon-2A_{\max}}} N_0^{2-2\epsilon-2A_{\max}}}, \\ &= \left(\frac{D_0}{W_0}\right)^{\frac{2-2\epsilon-2A_{\max}}{3-2\epsilon-2A_{\max}}} \sum_{\ell=0}^{L^*} \frac{\left(D_\ell W_\ell^{2-2\epsilon-2A_{\max}}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}}}{N_0^{2-2\epsilon-2A_{\max}}}. \end{aligned} \tag{125}$$

Rearranging yields

$$N_0^{*2-2\epsilon-2A_{\max}} = \left(\frac{\sqrt{8}C_\alpha}{TOL}\right)^2 \left(\frac{D_0}{W_0}\right)^{\frac{2-2\epsilon-2A_{\max}}{3-2\epsilon-2A_{\max}}} \sum_{\ell=0}^{L^*} \left(D_\ell W_\ell^{2-2\epsilon-2A_{\max}}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}}, \tag{126}$$

from which it follows that

$$N_0^* = \left(\frac{\sqrt{8}C_\alpha}{TOL}\right)^{\frac{1}{1-\epsilon-A_{\max}}} \left(\frac{D_0}{W_0}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}} \left(\sum_{\ell=0}^{L^*} \left(D_\ell W_\ell^{2-2\epsilon-2A_{\max}}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}} \right)^{\frac{1}{2-2\epsilon-2A_{\max}}}. \tag{127}$$

Substituting this into (124) yields

(128)
$$N_\ell^* = \left(\frac{\sqrt{8}C_\alpha}{TOL}\right)^{\frac{1}{1-\epsilon-A_{\max}}} \left(\frac{D_\ell}{W_\ell}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}} \left(\sum_{k=0}^{L^*} \left(D_k W_k^{2-2\epsilon-2A_{\max}}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}}\right)^{\frac{1}{2-2\epsilon-2A_{\max}}}, \quad 0 \le \ell \le L^*.$$

The first term in the total work is given as

$$\sum_{\ell=0}^{L^*} N_\ell^* W_\ell = \left(\frac{\sqrt{8}C_\alpha}{TOL}\right)^{\frac{1}{1-\epsilon-A_{\max}}} \sum_{\ell=0}^{L^*} W_\ell \left(\frac{D_\ell}{W_\ell}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}} \left(\sum_{k=0}^{L^*} \left(D_k W_k^{2-2\epsilon-2A_{\max}}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}}\right)^{\frac{1}{2-2\epsilon-2A_{\max}}},$$

(129)
$$= \left(\frac{\sqrt{8}C_\alpha}{TOL}\right)^{\frac{1}{1-\epsilon-A_{\max}}} \left(\sum_{\ell=0}^{L^*} \left(D_\ell W_\ell^{2-2\epsilon-2A_{\max}}\right)^{\frac{1}{3-2\epsilon-2A_{\max}}}\right)^{\frac{3-2\epsilon-2A_{\max}}{2-2\epsilon-2A_{\max}}}.$$

For appropriately chosen M_ℓ and h_ℓ, where $1 \le \ell \le L$ (see Remark 3), it follows from (121) that

$$\begin{aligned} D_{L^*} &= C_{\eta_{s,d_1}}^2 h_{L^*}^{2\eta_{s,d_1}} (b^{(I)})^2 B_A^2 C_{\epsilon,d_1}^2, \\ &= \mathcal{O}\left(TOL^{\frac{2\eta_{s,d_1}}{\eta_w}}\right) \end{aligned} \tag{130}$$

and from (120) that

$$W_{L^*}^{2-2\epsilon-2A_{\max}} = \mathcal{O}\left(TOL^{-\left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_w}\right)(2-2\epsilon-2A_{\max})}\right), \tag{131}$$

Moreover, it follows that

$$D_{L^*} W_{L^*}^{2-2\epsilon-2A_{\max}} = \mathcal{O}\left(TOL^{\frac{2\eta_{s,d_1}}{\eta_w}-\left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_w}\right)(2-2\epsilon-2A_{\max})}\right). \tag{132}$$

If the exponent in this term is positive, the sum in (129) converges as $L^* \to \infty$, and can be bounded by a constant in TOL. In which case, it follows that

$$\sum_{\ell=0}^{L^*} N_\ell^* W_\ell = \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}}\right). \tag{133}$$

In the case that the exponent in (132) is equal to zero, that is,

$$\frac{2\eta_{s,d_1}}{\eta_w} - \left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_w}\right)(2-2\epsilon-2A_{\max}) = 0, \tag{134}$$

each term in the sum in (129) is constant, and the sum is thus proportional to L^*, yielding

$$\sum_{\ell=0}^{L^*} N_\ell^* W_\ell = \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}} \max\left\{\log_2\left(TOL^{-\frac{1}{\eta_w}}\right), \log_2\left(TOL^{-\frac{1}{1-\epsilon}}\right)\right\}\right). \tag{135}$$

In the case that the exponent (132) is negative, and thus the sum in (129) diverges, it can be bounded by the largest term, i.e.,

$$\begin{aligned} \sum_{\ell=0}^{L^*} N_\ell^* W_\ell &= \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}} TOL^{\frac{2\eta_{s,d_1}}{\eta_w(2-2\epsilon-2A_{\max})}-\left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_w}\right)}\right), \\ &= \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}+\frac{\eta_{s,d_1}}{\eta_w(1-\epsilon-A_{\max})}-\left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_w}\right)}\right). \end{aligned} \tag{136}$$

Finally, the second term in the total work is bounded as follows:

$$\sum_{\ell=0}^{L^*} W_\ell = \mathcal{O}\left(TOL^{-\left(\frac{1}{1-\epsilon}+\frac{\gamma}{\eta_w}\right)}\right). \tag{137}$$

□

Remark 6 (Choice of splitting parameters in the optimal work derivation)**.** *Several error splitting parameters equal to* $1/2$ *are introduced in the derivation of the optimal computational work of the MLDLQMC estimator in the proof of Proposition 6. Optimizing the error splitting parameters in MLMC is outside the scope of this study; and we refer to the work* [39] *for a thorough investigation.*

Remark 7 (Impact of the dimension on the estimator work)**.** *The integration dimensions of the inner and outer integrands may affect the multiplicative error terms implied by the* $\mathcal{O}$ *notation in the above results. Investigating this dependence is not the focus of the present study; however,* [2] *successfully applied a nested rQMC estimator for EIG estimation up to a dimension of 18, and* [35] *estimated the EIG of an experiment at an even higher dimension using rQMC methods and proved independence of the inner integral estimation of the dimension.*

Corollary 1 (Optimal work of the MLDLQMC estimator with exact sampling)**.** *Under the same assumptions as Proposition 6, the total computational work of the MLDLQMC estimator* (52) *with exact sampling is given as follows:*

$$W^*_{\mathrm{MLDLQ}} = \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-A_{\max}}}\right) \tag{138}$$

for any $\epsilon > 0$ *as* $TOL \to 0$*.*

Proof. It holds that

$$\frac{1}{1-\epsilon-A_{\max}} > \frac{1}{1-\epsilon} \tag{139}$$

for any $\epsilon > 0$, where $0 < A_{\max} < 1/2$. The result then follows immediately from Proposition 6 in the limit as $h_\ell \to 0$, where $0 \le \ell \le L$, as $\gamma = 0$ and $\eta_{\mathrm{w}} = \eta_{\mathrm{s},d_1} = 1$ without loss of generality. □

Remark 8 (Comparison to other estimators)**.** *The result in Proposition 6 demonstrates the improved asymptotic efficiency of the MLDLQMC estimator with inexact sampling compared with the rDLQMC estimator in* [2] *and the MLDLMC esimator for EIG estimation proposed in* [4]*. Moreover, the result in Corollary 1 demonstrates the improved asymptotic efficiency of the MLDLQMC estimator with exact sampling compared with the MLDLMC estimator for EIG estimation analyzed in* [30] *and indicates comparable efficiency to the sparse tensor product estimator for EIG estimation analyzed in* [35]*.*

4. Expected information gain estimation

We now demonstrate how to apply the MLDLQMC estimator to estimate of the EIG of an experiment. This is an application from Bayesian experimental design, where uncertain parameters of interest $\boldsymbol{\theta}$ are modeled as random variables with a prior distribution $\pi(\boldsymbol{\theta})$. The selection of the prior reflects the a priori knowledge of the experimenter. The observation data is modeled as follows:

$$\boldsymbol{Y}(\boldsymbol{\xi}) = \boldsymbol{G}(\boldsymbol{\theta}_t, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}, \tag{140}$$

where $\boldsymbol{Y} \in \mathbb{R}^{d_y}$. The design $\boldsymbol{\xi}$ is omitted for conciseness, as it is not essential for the integral approximation that is the focus of the current work, and finding optimal designs is outside the scope of this study. Furthermore, $\boldsymbol{G} : \mathbb{R}^{d_\theta} \times \mathbb{R}^{d_\xi} \to \mathbb{R}^{d_y}$ is an appropriate model for the experiment evaluated at the unknown true parameter vector $\boldsymbol{\theta}_t$ and $\boldsymbol{\varepsilon} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$ is additive Gaussian noise with known covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \in \mathbb{R}^{d_y \times d_y}$ (i.e., symmetric and strictly positive definite). The noise $\boldsymbol{\varepsilon}$ should not be confused with the factor ϵ in the convergence rate of rQMC methods used throughout the remainder of this study. The EIG quantifies the amount of information that is expected to be gained from the data observation regarding the parameters of interest. It is defined as

$$EIG \coloneqq \int_{\mathcal{Y}} \int_{\boldsymbol{\Theta}} \log\left(\frac{\pi(\boldsymbol{\theta}|\boldsymbol{Y})}{\pi(\boldsymbol{\theta})}\right) \pi(\boldsymbol{\theta}|\boldsymbol{Y}) \,\mathrm{d}\boldsymbol{\theta} p(\boldsymbol{Y}) \,\mathrm{d}\boldsymbol{Y}, \tag{141}$$

where $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ is the posterior distribution of the parameters of interest given the data, and $p(\boldsymbol{Y})$ is the evidence, or marginal likelihood. The data model (140) implies the following likelihood:

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{G}(\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}}, \tag{142}$$

where $\|\cdot\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} \coloneqq \langle\cdot,\cdot\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}$, and $\langle\cdot,\cdot\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}$ is the $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}$-inner product, that is, $\langle\cdot,\cdot\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} \coloneqq \langle\cdot,\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\cdot\rangle$, and $\langle\cdot,\cdot\rangle$ is the standard L^2 inner product. This likelihood is used together with Bayes' theorem and marginalization to rewrite (141) as follows:

$$EIG = \int_{\boldsymbol{\Theta}}\int_{\mathcal{Y}} \left[\log\left(p(\boldsymbol{Y}|\boldsymbol{\theta})\right) - \log\left(\int_{\boldsymbol{\Theta}} p(\boldsymbol{Y}|\boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta})\,\mathrm{d}\boldsymbol{\vartheta}\right)\right] p(\boldsymbol{Y}|\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{Y}\,\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}, \tag{143}$$

where $\boldsymbol{\vartheta}$ indicates an auxiliary variable for integration. Substituting (140) into (142) yields

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{\theta}) &= \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{G}(\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}}, \\ &= \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\|\boldsymbol{G}(\boldsymbol{\theta})+\boldsymbol{\varepsilon}-\boldsymbol{G}(\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}}, \\ &= \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\|\boldsymbol{\varepsilon}\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}}, \end{aligned} \tag{144}$$

and

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{\vartheta}) &= \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{G}(\boldsymbol{\vartheta})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}}, \\ &= \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\|\boldsymbol{G}(\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}+\|\boldsymbol{\varepsilon}\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}+\|\boldsymbol{G}(\boldsymbol{\vartheta})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}+2\langle\boldsymbol{G}(\boldsymbol{\theta}),\boldsymbol{\varepsilon}\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}-2\langle\boldsymbol{G}(\boldsymbol{\theta}),\boldsymbol{G}(\boldsymbol{\vartheta})\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}-2\langle\boldsymbol{G}(\boldsymbol{\vartheta}),\boldsymbol{\varepsilon}\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}\right)}. \end{aligned} \tag{145}$$

From this, it follows that

$$\begin{aligned} EIG = \int_{\boldsymbol{\Theta}}\int_{\mathbb{R}^{d_y}} &\left[\frac{1}{2}\|\boldsymbol{G}(\boldsymbol{\theta})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \langle\boldsymbol{G}(\boldsymbol{\theta}),\boldsymbol{\varepsilon}\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}\right. \\ &\left. - \log\left(\int_{\boldsymbol{\Theta}} e^{-\frac{1}{2}\left(\|\boldsymbol{G}(\boldsymbol{\vartheta})\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}-2\langle\boldsymbol{G}(\boldsymbol{\theta}),\boldsymbol{G}(\boldsymbol{\vartheta})\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}-2\langle\boldsymbol{G}(\boldsymbol{\vartheta}),\boldsymbol{\varepsilon}\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}\right)}\pi(\boldsymbol{\vartheta})\,\mathrm{d}\boldsymbol{\vartheta}\right)\right] \\ &\times \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\|\boldsymbol{\varepsilon}\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}}\,\mathrm{d}\boldsymbol{\varepsilon}\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}. \end{aligned} \tag{146}$$

The first term in (146) is not nested and can thus be estimated using the standard rQMC or similar methods. The second term is symmetric around the origin and thus vanishes as it is integrated against the Gaussian measure for $\boldsymbol{\varepsilon}$. The remaining term constitutes a nested integral and is the main subject of this investigation. We introduce the notation $f \equiv \log$, and

$$g(\boldsymbol{y},\boldsymbol{x}) \coloneqq \exp\left(-\frac{1}{2}\left\|\boldsymbol{G}(F_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}))\right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle\boldsymbol{G}(F_{\boldsymbol{\theta}}^{-1}(\boldsymbol{y}_1)),\boldsymbol{G}(F_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}))\right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle\boldsymbol{G}(F_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x})),\Phi^{-1}(\boldsymbol{y}_2)\right\rangle_{\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}^{-\top}}\right), \tag{147}$$

where $F_{\boldsymbol{\theta}}^{-1}$ is the inverse CDF of the parameters of interest, $\boldsymbol{y} = (\boldsymbol{y}_1,\boldsymbol{y}_2)$, $F_{\boldsymbol{\theta}}^{-1}(\boldsymbol{y}_1) \coloneqq \boldsymbol{\theta}$, $F_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}) \coloneqq \boldsymbol{\vartheta}$, $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}\Phi^{-1}(\boldsymbol{y}_2) \coloneqq \boldsymbol{\varepsilon}$, and

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}^{\top}. \tag{148}$$

Moreover, we introduce the notation

$$\tilde{\boldsymbol{G}} \coloneqq \boldsymbol{G}\circ F_{\boldsymbol{\theta}}^{-1}. \tag{149}$$

The Gaussian noise present in typical data models together with the logarithm present in the Kullback–Leibler divergence [7, 8] introduce singularities at the integration domain boundary, rendering the standard error bound via the Koksma–Hlawka inequality (10) trivial, as the Hardy–Krause variation for such integrands is infinite. The behavior of these singularities must be carefully controlled to obtain the optimal rate of convergence of the rQMC error. Moreover, for the case where the experiment model $\tilde{\boldsymbol{G}}$ must be approximated as $\tilde{\boldsymbol{G}}_h$ by an appropriate method with discretization parameter $h > 0$, the inner integrand in (146) is instead defined as follows:

$$g_h(\boldsymbol{y},\boldsymbol{x}) \coloneqq \exp\left(-\frac{1}{2}\left\|\tilde{\boldsymbol{G}}_h(\boldsymbol{x})\right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle\tilde{\boldsymbol{G}}_h(\boldsymbol{y}_1),\tilde{\boldsymbol{G}}_h(\boldsymbol{x})\right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle\tilde{\boldsymbol{G}}_h(\boldsymbol{x}),\Phi^{-1}(\boldsymbol{y}_2)\right\rangle_{\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}^{-\top}}\right). \tag{150}$$

The following assumptions are made in this work:

Assumption 11 (Condition on the approximate experiment model)**.** *Let $\tilde{\boldsymbol{G}}_h : [0,1]^{d_2} \to \mathbb{R}^{d_1-d_2}$ be an approximation to $\tilde{\boldsymbol{G}}$ and assume that there exists $0 \le q < \infty$ independent of h such that*

$$\left\| \left(\prod_{j \in u} \frac{\partial}{\partial z_j} \right) \tilde{\boldsymbol{G}}_h(\boldsymbol{z}) \right\|_2 \le q, \tag{151}$$

for all $\boldsymbol{z} \in [0,1]^{d_2}$, all $h > 0$, and all $u \subseteq \{1, \ldots, d_2\}$.

Assumption 12 (Inverse inequality for the FEM rate)**.** *Let g_h be the approximate inner integrand in* (150)*, with the inner integral denoted as follows:*

$$\bar{g}_h(\boldsymbol{y}) \coloneqq \int_{[0,1]^{d_2}} \exp\left(-\frac{1}{2} \left\| \tilde{\boldsymbol{G}}_h(\boldsymbol{x}) \right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle \tilde{\boldsymbol{G}}_h(\boldsymbol{y}_1), \tilde{\boldsymbol{G}}_h(\boldsymbol{x}) \right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle \tilde{\boldsymbol{G}}_h(\boldsymbol{x}), \Phi^{-1}(\boldsymbol{y}_2) \right\rangle_{\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}^{-\top}} \right) \mathrm{d}\boldsymbol{x}, \tag{152}$$

where $\tilde{\boldsymbol{G}}_h$ as in Assumption 11, $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$, $\boldsymbol{y}_1, \boldsymbol{x} \in [0,1]^{d_2}$, $\boldsymbol{y}_2 \in [0,1]^{d_1-d_2}$, $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \in \mathbb{R}^{(d_1-d_2)\times(d_1-d_2)}$ a symmetric, strictly positive definite matrix, and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}^{\top}$ as in (148)*. Then, assume that*

$$\sup_{\boldsymbol{y} \in [0,1]^{d_1}} \left| \frac{\left(\prod_{j \in u} \frac{\partial}{\partial y_j} \right) \left(\bar{g}_h(\boldsymbol{y}) - \bar{g}(\boldsymbol{y}) \right)}{\left(\prod_{j \in u} \frac{\partial}{\partial y_j} \right) \bar{g}(\boldsymbol{y})} \right| < C_{\mathrm{s},d_1} h^{\eta_{\mathrm{s},d_1}}, \tag{153}$$

is the relative strong convergence rate depending on the regularity of $\bar{g}_h$ and $\bar{g}$ in the jth components, where $j \in u \subseteq \{1, \ldots, d\}$, and $0 < C_{\mathrm{s},d_1} < \infty$ is a constant independent of h, and $\eta_{\mathrm{s},d_1} > 0$.

The bound on the total error and computational work for estimating the EIG using the MLDLQMC estimator for any $A_i > 0$, where $1 \le i \le d_1$ and $d_1 = d_\theta + d_y$, is presented below.

Lemma 1 (Bound on the integrand at level 0 for the EIG)**.** *Let $f \equiv \log$ and g_h as in* (150) *for any $h > 0$. Moreover, let $\boldsymbol{x}^{(m)}$ be as in* (43)*, where $1 \le m \le M$. Then, provided that Assumption 11 holds, there exists $0 < b < \infty$ such that*

$$\left| \left(\prod_{j \in u} \frac{\partial}{\partial y_j} \right) f\left(\frac{1}{M} \sum_{m=1}^{M} g_h\left(\boldsymbol{y}, \boldsymbol{x}^{(m)} \right) \right) \right| \le b \prod_{i=1}^{d} \min(y_i, 1-y_i)^{-A_i - \mathbb{1}_{\{i \in u\}}} \tag{154}$$

for all $\boldsymbol{y} = (y_1, \ldots, y_{d_1}) \in (0,1)^{d_1}$, all $u \subseteq \{1, \ldots, d_1\}$, and any $\max_{1 \le i \le d_1} A_i > 0$.

Appendix C presents the proof of this result and Appendix A provides auxiliary results. A truncation of the Gaussian observation noise in the EIG, directly adapted from [2, Lemma 3], is employed to bound the variance of the levels $1 \le \ell \le L$:

Lemma 2 (Truncation of the Gaussian observation noise in the EIG)**.** *Let the EIG be as in* (146)*, that is,*

$$\begin{aligned} EIG = \int_{[0,1]^{d_2}} \int_{\mathbb{R}^{d_1-d_2}} \Bigg[& \frac{1}{2} \left\| \tilde{\boldsymbol{G}}(\boldsymbol{y}_1) \right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle \tilde{\boldsymbol{G}}(\boldsymbol{y}_1), \boldsymbol{\varepsilon} \right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} \\ & - \log\left(\int_{[0,1]^{d_2}} e^{-\frac{1}{2}\left(\left\| \tilde{\boldsymbol{G}}(\boldsymbol{x}) \right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} - 2\left\langle \tilde{\boldsymbol{G}}(\boldsymbol{y}_1), \tilde{\boldsymbol{G}}(\boldsymbol{x}) \right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} - 2\left\langle \tilde{\boldsymbol{G}}(\boldsymbol{x}), \boldsymbol{\varepsilon} \right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} \right)} \pi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right) \Bigg] \\ & \times \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\|\boldsymbol{\varepsilon}\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}} \, \mathrm{d}\boldsymbol{\varepsilon} \pi(\boldsymbol{y}_1) \, \mathrm{d}\boldsymbol{y}_1, \end{aligned} \tag{155}$$

and denote by

$$\begin{aligned} EIG^{\mathrm{tr}} = \int_{[0,1]^{d_2}} \int_{[-c(TOL), c(TOL)]^{d_1-d_2}} \Bigg[& \frac{1}{2} \left\| \tilde{\boldsymbol{G}}(\boldsymbol{y}_1) \right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle \tilde{\boldsymbol{G}}(\boldsymbol{y}_1), \boldsymbol{\varepsilon} \right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} \\ & - \log\left(\int_{[0,1]^{d_2}} e^{-\frac{1}{2}\left(\left\| \tilde{\boldsymbol{G}}(\boldsymbol{x}) \right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} - 2\left\langle \tilde{\boldsymbol{G}}(\boldsymbol{y}_1), \tilde{\boldsymbol{G}}(\boldsymbol{x}) \right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} - 2\left\langle \tilde{\boldsymbol{G}}(\boldsymbol{x}), \boldsymbol{\varepsilon} \right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} \right)} \pi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right) \Bigg] \\ & \times \frac{1}{\det(2\pi\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})^{\frac{1}{2}}} e^{-\frac{1}{2}\|\boldsymbol{\varepsilon}\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}} \, \mathrm{d}\boldsymbol{\varepsilon} \pi(\boldsymbol{y}_1) \, \mathrm{d}\boldsymbol{y}_1, \end{aligned} \tag{156}$$

the TOL−truncated EIG. Here,

$$c(TOL) \coloneqq \tilde{q}(2(1+p))^{\frac{1}{2}} \log(TOL^{-1})^{\frac{1}{2}}, \tag{157}$$

for some constant $0 < \tilde{q} < \infty$*, and any* $p > 0$*,* $TOL > 0$*, signifies a region of truncation. Then, it holds that*

$$\left|EIG - EIG^{\mathrm{tr}}\right| = o(TOL) \tag{158}$$

as $TOL \to 0$*.*

Remark 9 (Extension to general observation noise distributions)**.** *The truncation parameter* $c(TOL)$ *depends on the structure of the Gaussian observation noise and could differ for other noise distributions (see* [2, Remark 12]*).*

The result in Lemma 1 still applies to the truncated integrands g_h^{tr} and $\bar{g}_h^{\mathrm{tr}}$ and is not restated for brevity. The following notation is introduced:

$$\tilde{M}(TOL) \coloneqq \left(\frac{C_{\epsilon,d_2} c(TOL)(1 + C_{\mathrm{s},d_1} h^{\eta_{\mathrm{s},d_1}})}{(1 - C_{\mathrm{s},d_1} h^{\eta_{\mathrm{s},d_1}})}\right)^{\frac{1}{1-\epsilon}}, \tag{159}$$

where $c(TOL)$ as in (157), $TOL > 0$, $0 < h < C_{\mathrm{s},d_1}^{-1/\eta_{\mathrm{s},d_1}}$, $C_{\mathrm{s},d_1} > 0$, and $\eta_{\mathrm{s},d_1} > 0$.

Lemma 3 (Bound on the higher-order term in the Taylor expansion of the nested integrand)**.** *Let* $f \equiv \log$*,* g_h^{tr} *as in* (213)*,* $0 < h < C_{\mathrm{s},d_1}^{-1/\eta_{\mathrm{s},d_1}}$*,* $M > \tilde{M}(TOL)$*, where* $\tilde{M}(TOL)$ *as in* (159)*,* $TOL > 0$*, and* $\boldsymbol{x}^{(m)}$ *be as in* (43)*, where* $1 \le m \le M$*. Moreover, let* $\varphi_{h,M}$ *constitute an auxiliary function, where*

$$\varphi_{h,M}(\cdot) = \left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\cdot, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\cdot)\right)^2 \int_0^1 f''\left(\bar{g}^{\mathrm{tr}}(\cdot) + s\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\cdot, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\cdot)\right)\right)(1-s)\,\mathrm{d}s, \tag{160}$$

and $\bar{g}^{\mathrm{tr}} \coloneqq \lim_{h\to 0} \bar{g}_h^{\mathrm{tr}}$*. Then, given Assumptions 11 and 12, there exists* $0 < b < \infty$ *such that*

$$\left|\left(\prod_{j\in u}\frac{\partial^u}{\partial y_j}\right)\varphi_{h,M}(\boldsymbol{y})\right| \le \left(C_{\epsilon,d_2}^2 M^{-2+2\epsilon} + C_{\eta_{\mathrm{s},d_1}}^2 h^{2\eta_{\mathrm{s},d_1}}\right) b \prod_{i=1}^{d_1} \min(y_i, 1-y_i)^{-A_i - \mathbb{1}_{\{i\in u\}}} \tag{161}$$

for all $\boldsymbol{y} = (y_1, \ldots, y_{d_1}) \in (0,1)^{d_1}$*, all* $u \subseteq \{1, \ldots, d_1\}$*, and any* $A_i > 0$*, where* $1 \le i \le d_1$*.*

Appendix C presents the proof of this result. The condition that $M > \tilde{M}(TOL)$ for $TOL > 0$ and $h < C_{\mathrm{s},d_1}^{-1/\eta_{\mathrm{s},d_1}}$ in Lemma 3 is addressed by setting the following in the MLDLQMC estimator:

$$M_0 \coloneqq \left(\frac{C_{\epsilon,d_2}(1 + C_{\mathrm{s},d_1} h_0^{\eta_{\mathrm{s},d_1}})}{(1 - C_{\mathrm{s},d_1} h_0^{\eta_{\mathrm{s},d_1}})}\right)^{\frac{1}{1-\epsilon}} c(TOL)^{\frac{1}{1-\epsilon}}, \tag{162}$$

and

$$h_0 \coloneqq C_{\mathrm{s},d_1}^{-\frac{1}{\eta_{\mathrm{s},d_1}}}. \tag{163}$$

The bound (154) in Lemma 1 is independent of M and h and thus unaffected by these conditions. The multiplicative factor $c(TOL)^{1/(1-\epsilon)}$ is tracked throughout the remaining error bounds in the following corollary.

Corollary 2 (Total error of the MLDLQMC estimator for EIG estimation)**.** *Given Assumption 2, Assumption 11, and Assumption 12, for* $M_\ell > \tilde{M}(TOL)$ *and* $h_\ell < C_{\mathrm{s},d_1}^{-1/\eta_{\mathrm{s},d_1}}$*, where* $\tilde{M}(TOL)$ *is as in* (159) *and* $0 \le \ell \le L$*, there exist* $0 < \delta_K < 1/2$ *such that the total error of the MLDLQMC estimator for the EIG as in* (146) *is bounded as follows:*

$$\left|EIG - EIG^{\mathrm{tr}}\right| = o(TOL), \tag{164}$$

$$|\mathbb{E}[I_{\mathrm{MLDLQ}}] - EIG^{\mathrm{tr}}| \le C_{\mathrm{w}} h_L^{\eta_{\mathrm{w}}} + \frac{\mathbb{E}[|\bar{g}^{\mathrm{tr}}|^2 |f''(\bar{g}^{\mathrm{tr}})|] c(TOL)^2 C_{\epsilon,d_2}^2}{2M_L^{2-2\epsilon}} + \frac{\mathbb{E}[|\bar{g}^{\mathrm{tr}}|^3 |f'''(K(TOL)\bar{g}^{\mathrm{tr}})|] c(TOL)^3 C_{\epsilon,d_2}^3}{6M_L^{3-3\epsilon}}, \tag{165}$$

$$\mathbb{V}\left[I_{\mathrm{MLDLQ}}\right] \leq \frac{(b^{(0)})^2 B_A^2 C_{\epsilon,d_1}^2}{N_0^{2-2\epsilon-2A_{\max}}} + \sum_{\ell=1}^{L} \left(\frac{\left(\frac{C_{\epsilon,d_2}^2 c(TOL)^2}{M_\ell^{2-2\epsilon}} + C_{\eta_{\mathrm{s},d_1}}^2 h_\ell^{2\eta_{\mathrm{s},d_1}} \right) (b^{(I)})^2 B_A^2 C_{\epsilon,d_1}^2}{N_\ell^{2-2\epsilon-2A_{\max}}} \right.$$
$$\left. + \frac{\left(\frac{C_{\epsilon,d_2}^4 c(TOL)^4}{M_\ell^{4-4\epsilon}} + C_{\eta_{\mathrm{s},d_1}}^4 h_\ell^{4\eta_{\mathrm{s},d_1}} \right) (b^{(II)})^2 B_A^2 C_{\epsilon,d_1}^2}{N_\ell^{2-2\epsilon-2A_{\max}}} \right)$$
$$+ \frac{C_{\epsilon,d_2}^2 (b^{(III)})^2 B_A^2 c(TOL)^2}{M_L^{2-2\epsilon}} + \frac{C_{\epsilon,d_2}^4 (b^{(IV)})^2 B_A^2 c(TOL)^4}{M_L^{4-4\epsilon}}, \tag{166}$$

for any $\epsilon > 0$ and all $K(TOL) \in (\delta_K, 1-\delta_K)$ for an error tolerance $TOL > 0$, where $c(TOL)$ as in (157)*, where $C_{\epsilon,d_1}, C_{\epsilon,d_2} \to \infty$ as $\epsilon \to 0$, $B_A \to \infty$ as $\min_{1\leq i\leq d_1} A_i \to 0$, and any $A_{\max} \coloneqq \max_{1\leq i\leq d_1} A_i > 0$ where A_i as in* (154) *and* (161)*, where $1 \leq i \leq d_1$.*

Proof. The result in (164) follows from Lemma 2. Assumption 4 follows from Corollary 5 and Assumption 5 follows from Lemma 5 for k replaced with $c(TOL)$ and Corollary 4 as $\max_i A_i < 1/2$. It is verified that

$$|f'''(K(TOL)\bar{g}^{\mathrm{tr}})| = \left| \frac{2}{(K(TOL))^3 (\bar{g}^{\mathrm{tr}})^3} \right| \tag{167}$$

is monotonic for positive $K(TOL)\bar{g}^{\mathrm{tr}}$ and $f \equiv \log$. The result in (165) thus follows from Proposition 2. Assumption 6 follows from Lemma 1. Assumption 8 follows from Lemma 3 with the additional condition that $M_\ell > \tilde{M}(TOL)$ and $h_\ell < C_{\mathrm{s},d_1}^{-1/\eta_{\mathrm{s},d_1}}$, where $0 \leq \ell \leq L$. Assumption 7, Assumption 9, and Assumption 10 also follow from Lemma 3 with a similar condition on M_ℓ and h_ℓ. Finally, the result in (166) follows from Propositions 3 to 5, again with the condition that $M_L > \tilde{M}(TOL)$ and $h_L < C_{\mathrm{s},d_1}^{-1/\eta_{\mathrm{s},d_1}}$. □

Remark 10 (Total complexity of the MLDLQMC estimator for EIG estimation)**.** *The total computational complexity of the MLDLQMC estimator applied to the EIG differs from the result stated in Proposition 6 by a* log *factor as a result of the truncation of the observation noise. However, the factor $\epsilon > 0$ was introduced to hide logarithmic terms in the bound in* (9)*, hence, we omitted this dependence in the following result.*

The model $\tilde{\boldsymbol{G}}_{h_\ell}$ requires only $N_\ell + M_\ell$ evaluations at each level, rather than $N_\ell \times M_\ell$. Thus, the total computational work for estimating the EIG is

$$W_{\mathrm{MLQMC}} = \sum_{\ell=0}^{L} (N_\ell + M_\ell)\, h_\ell^{-\gamma} + \sum_{\ell=0}^{L} N_\ell M_\ell, \tag{168}$$

where the last term is due to the $N_\ell \times M_\ell$ required evaluations of the likelihood at each level. Tracking this reduced computational work throughout the proof of Proposition 6 results in the optimal work.

Corollary 3 (Optimal work of the MLDLQMC estimator with inexact sampling for EIG estimation)**.** *The total computational work of the MLDLQMC estimator* (52) *with inexact sampling for EIG estimation is given by*

$$W_{\mathrm{MLDLQ}}^* = \mathcal{O}\left(TOL^{-\left(\frac{1}{1-\epsilon} + \frac{\gamma}{\eta_{\mathrm{w}}}\right)} \right)$$
$$+ \begin{cases} \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-\max_i A_i}} \right), & \text{if } \frac{\eta_{\mathrm{s},d_1}}{\eta_{\mathrm{w}}(1-\epsilon-\max_i A_i)} > \max\left\{ \frac{1}{1-\epsilon} + \frac{\gamma}{\eta_{\mathrm{w}}} \right\}, \\ \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-\max_i A_i}} \log_2\left(TOL^{-\max\left\{\frac{1}{1-\epsilon} + \frac{\gamma}{\eta_{\mathrm{w}}}\right\}} \right) \right), & \text{if } \frac{\eta_{\mathrm{s},d_1}}{\eta_{\mathrm{w}}(1-\epsilon-\max_i A_i)} = \max\left\{ \frac{1}{1-\epsilon} + \frac{\gamma}{\eta_{\mathrm{w}}} \right\}, \\ \mathcal{O}\left(TOL^{-\frac{1}{1-\epsilon-\max_i A_i} + \frac{\eta_{\mathrm{s},d_1}}{\eta_{\mathrm{w}}(1-\epsilon-\max_i A_i)} - \max\left\{\frac{1}{1-\epsilon} + \frac{\gamma}{\eta_{\mathrm{w}}}\right\}} \right), & \text{if } \frac{\eta_{\mathrm{s},d_1}}{\eta_{\mathrm{w}}(1-\epsilon-\max_i A_i)} < \max\left\{ \frac{1}{1-\epsilon} + \frac{\gamma}{\eta_{\mathrm{w}}} \right\}. \end{cases} \tag{169}$$

for any $\epsilon > 0$ and any $A_i > 0$, $1 \leq i \leq d_1$, as $TOL \to 0$, where the multiplicative constant in the $\mathcal{O}$ notation approaches infinity as $\epsilon \to 0$ and as $\min_{1\leq i\leq d_1} A_i \to 0$.

This result implies that the total work is typically dominated by the term $\mathcal{O}(TOL^{-(\frac{1}{1-\epsilon} + \frac{\gamma}{\eta_{\mathrm{w}}})})$ when $\eta_{\mathrm{w}} = \eta_{\mathrm{s},d_1}$, which is a standard setting for random PDEs.

5. Numerical results

For the practical implementation of the MLDLQMC estimator applied to the EIG, the CLT confidence parameter $C_\alpha = \Phi^{-1}(1-\alpha/2)$ was selected for $\alpha = 0.05$ (i.e., $C_\alpha = 1.96$). This choice proved sufficient, and the more conservative Chebyshev's inequality was not needed (see Figure 2 (C) and Figure 5 (C)). Moreover, the derivation in (147) appears to be required for technical reasons only (see Appendix A), and the following inner integrand was applied in the implementation:

$$g(\boldsymbol{y}, \boldsymbol{x}) \coloneqq \exp\left(-\frac{1}{2}\left\|\boldsymbol{Y} - \boldsymbol{G}(\boldsymbol{\vartheta})\right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}}\right), \tag{170}$$

This integrand displayed preferable numerical behavior to that in (147). All numerical experiments were performed without truncation of the Gaussian observation noise, indicating that the MLDLQMC estimator is capable of addressing such integrands for the considered tolerances. Several pilot runs were employed to estimate the required constants and convergence rates using several randomizations S, R, independent of the error tolerance TOL. Thereafter, the optimal number of levels L^* and outer samples N_ℓ^*, where $0 \leq \ell \leq L^*$, were derived for a specified error tolerance $TOL > 0$ according to the expressions (122) and (128). A deterministic shift modulo 1 was applied to the inner points, that is,

$$\boldsymbol{\theta}^{(m)} = \left(\boldsymbol{t}^{(m)} + \frac{1}{\sqrt{2}}\right)_{\text{mod1}}, \quad 1 \leq m \leq M_\ell,\, 0 \leq \ell \leq L-1, \tag{171}$$

to ensure that the initial point $\boldsymbol{\theta}^{(1)}$ is not the origin $\mathbf{0}$, where $(\cdot)_{\text{mod1}}$ denotes the modulo 1 operator. The choice of $1/\sqrt{2}$ is arbitrary, and any irrational number would serve a similar purpose of moving the Sobol' points away from the origin.

5.1. **Example 1: linear EIG example with exact sampling.**

5.1.1. *Setting and notation.* The first example we consider is given as follows:

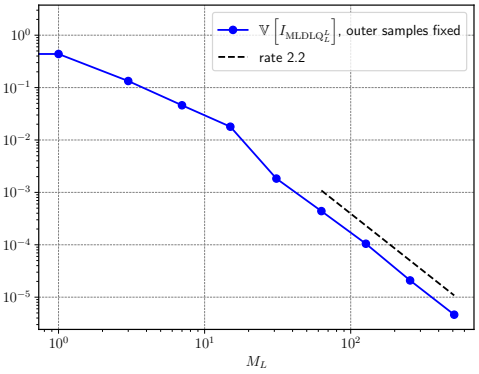
$$\boldsymbol{G}(\boldsymbol{\theta}) \coloneqq \boldsymbol{A}\boldsymbol{\theta}, \tag{172}$$

where $\boldsymbol{\theta} \overset{iid}{\sim} \mathcal{U}([0,1]^{d_\theta})$, for $d_\theta = 4$, and $\boldsymbol{A} \coloneqq \boldsymbol{I}_{d_\theta \times d_\theta}$ is the identity matrix in $\mathbb{R}^{d_\theta}$. Moreover,

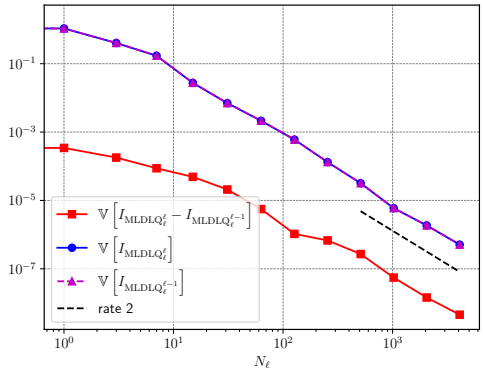
$$\boldsymbol{Y} = \boldsymbol{G}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}, \tag{173}$$

where $\boldsymbol{\varepsilon} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_{d_y \times d_y})$, for $\sigma^2 = 0.25$ and $\boldsymbol{I}_{d_y \times d_y}$ is the identity matrix in dimension $d_y = 4$. Thus, in the nested integration setting, $d_1 = d_\theta + d_y = 8$ and $d_2 = d_\theta = 4$. A design $\boldsymbol{\xi}$ is not specified for this example to focus on the demonstration of applicability of the MLDLQMC estimator to the EIG of an experiment. Optimizing the EIG regarding a design is outside the scope of this study. It is easily verified that Assumption 11 holds for $\boldsymbol{G}_h \equiv \boldsymbol{G}$ in expression (172). The EIG for this linear example separates into four independent inner integrals in one dimension, allowing for efficient estimation via deterministic quadrature methods. The MLDLQMC estimator with exact sampling was nevertheless able to efficiently approximate the four-dimensional integration problem, as evidenced in the following sections.

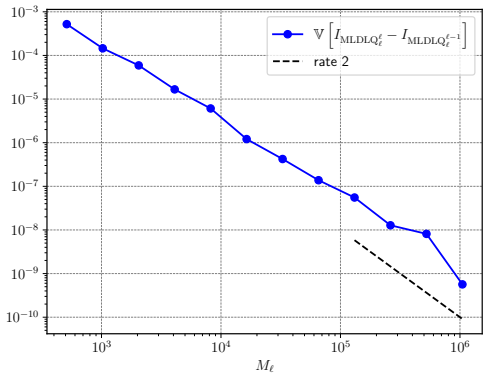
5.1.2. *Pilot runs.* We employed pilot runs to verify that the convergence rates predicted via the theoretical analysis were achieved. This verification also allowed us to estimate the constant terms in the error bounds appearing in Corollary 2 required for the optimal setting specified in (169). First, we verified that the variance contribution from the inner samples at the final level L converged at a rate of 2.2 and was thus slightly higher than the projected rate of 2 (Figure 1 (A)). The rQMC method with scrambled Sobol' points is known to display improved convergence rate for some integrands [21, Chapter 2]. This estimation was carried out using $N_L = 1$ outer sample and $S = 1000$, $R = 1$ randomizations. The convergence of the variance of the level differences as a function of N_ℓ at the expected rate of 2 was demonstrated in Panel (B), where $M_{\ell-1} = 2^8$, $M_\ell = 2^9$ samples and $S = 1000$ randomizations were used. The variance of the single-level estimators converged at the same rate with a larger multiplicative constant and was independent of M_ℓ, $M_{\ell-1}$. To demonstrate the convergence of the variance of the level difference as a function of M_ℓ, we kept $N_\ell = 1$ fixed. This variance converged at a rate of 2, as seen in Panel (C) for $S = 1000$ randomizations.



(A) The observed convergence rate of the variance of the inner samples at the final level L of 2.2 (slightly higher than predicted by the theory).



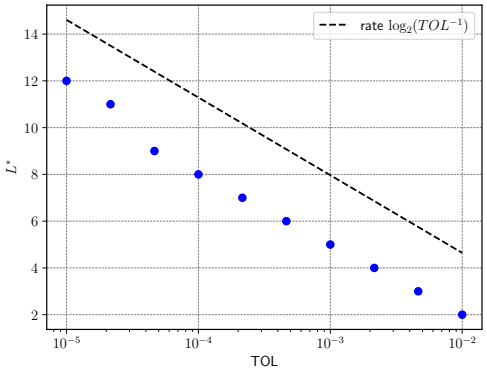
(B) The observed convergence rate of the variance at a rate of 2 with a reduced multiplicative factor for the difference estimator.



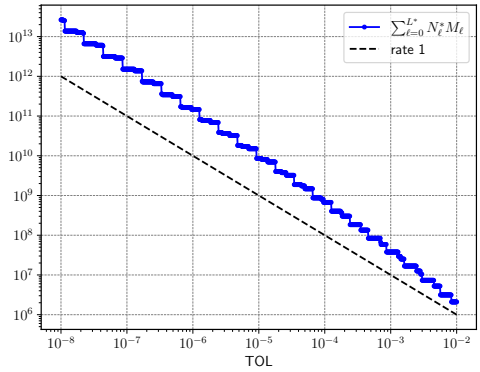
(C) The variance of the level difference converged at a rate of 2 as a function of M_ℓ for $N_\ell = 1$ fixed.

FIGURE 1. Example 1: Pilot runs for the MLDLQMC estimator with exact sampling. Each observed variance contribution converged at least at the rate predicted by the theoretical analysis.

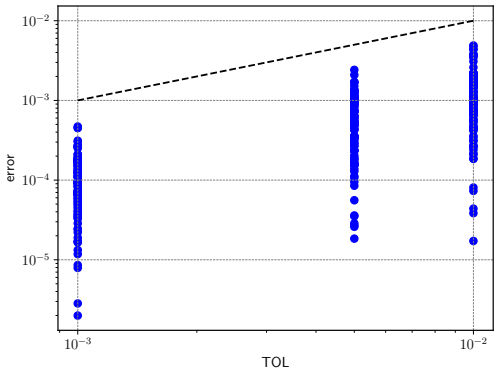
5.1.3. *Consistency and work analysis.* We then determined the optimal final level L^* of the MLDLQMC estimator with exact sampling for a given tolerance $TOL > 0$. The number of levels is displayed in Figure 2 (A) and the rate of increase is $\log_2(TOL^{-1})$, as expected. Finally, the accuracy and efficiency of the MLDLQMC estimator for optimal outer samples N_ℓ^*, where $1 \leq \ell \leq L^*$, was analyzed for different error tolerances TOL. In Panel (B), the total work of the MLDLQMC estimator, depending on the optimal L^* and N_ℓ^*, where $0 \leq \ell \leq L^*$, is presented as a function of the tolerance TOL. The projected rate of increase of $1/(1-\epsilon) \approx 1$ is only observed for $TOL < 10^{-4}$. In Panel (C), 100 runs of the MLDLQMC estimator with $S = R = 1$ randomizations at each level for $TOL = 10^{-2}$, $TOL = 5 \times 10^{-3}$, and $TOL = 10^{-3}$ are presented, each of them staying below the error tolerance projected by the CLT for the given L^* and N_ℓ^*, where $0 \leq \ell \leq L^*$ and $M_0 = 2^8$. An average of 100 runs for $TOL = 5 \times 10^{-4}$ was applied as a reference solution.



(A) The optimal number of levels L^* increased at a rate of $\log_2(TOL^{-1})$ as a function of the error tolerance $TOL > 0$.



(B) Total computational work of the MLDLQMC estimator as a function of the error tolerance TOL. The projected rate of 1 was only attained for low tolerances.



(C) Error vs. tolerance (TOL) for 100 runs of the MLDLQMC estimator. Each run resulted in an error less than the tolerance projected by the central limit theorem.

FIGURE 2. Example 1: The MLDLQMC estimator with exact sampling resulted in error less than the projected tolerance TOL (C) at a cost increasing at the projected rate of 1 asymptotically as $TOL \to 0$ (B).

5.2. **Example 2: Poisson EIG example with inexact sampling.**

5.2.1. *Setting and notation.* The second example we consider is given as follows:

$$\boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{\xi}) \coloneqq \big(u(\xi_1), u(\xi_2)\big)\,, \tag{174}$$

where $u \in H_0^1(\mathcal{D})$ is the solution to the Poisson equation in one dimension, that is,

$$\begin{aligned} -u'' &= \psi \quad \text{in } \mathcal{D}, \\ u &= 0, \quad \text{on } \{0,1\}. \end{aligned} \tag{175}$$

Here, $\mathcal{D} = (0,1)$ and

$$\psi(x, \boldsymbol{\theta}) \coloneqq \theta_1 \sin(4x) + \theta_2 \cos(4x) + \theta_3 \sin(8x) + \theta_4 \cos(8x), \quad x \in \mathcal{D}, \tag{176}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4) \overset{iid}{\sim} \mathcal{U}([0,1]^{d_\theta})$, for $d_\theta = 4$. Moreover, $\boldsymbol{\xi} \in (0,1)^{d_\xi}$, where $d_\xi = 2$, and

$$\boldsymbol{Y} = \boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}, \tag{177}$$

where $\boldsymbol{\varepsilon} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_{d_y \times d_y})$, for $\sigma^2 = 10^{-4}$ and $\boldsymbol{I}_{d_y \times d_y}$ is the identity matrix in dimension $d_y = 2$. Thus, in the nested integration setting, $d_1 = d_\theta + d_y = 6$ and $d_2 = d_\theta = 4$. We consider the discretization

$$\boldsymbol{G}_h(\boldsymbol{\theta}, \boldsymbol{\xi}) \coloneqq \left(u_h(\xi_1), u_h(\xi_2)\right), \tag{178}$$

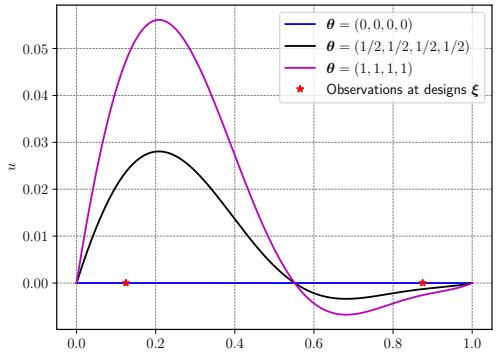
where $u_h(x) : [0,1] \mapsto \mathbb{R}$ is an FEM approximation of u, where $\lceil 1/h \rceil$ specifies the number of elements for $h > 0$, $\lceil \cdot \rceil$ is the ceiling function. It is easily verified that Assumption 11 holds for $\boldsymbol{G}_h$ in expression (178). Verifying Assumption 12, concerning the strong convergence of mixed derivatives of the inner integrand in a relative sense is contingent on the particular discretization method and will thus be reserved for future research. However, the convergence of the estimator bias as a function of the FEM discretization parameter was demonstrated numerically (in the weak sense) in Figure 4 (A), and the convergence of the estimator variance (in the strong sense) was indicated in Panel (E). For this example, evaluating the experiment model $\boldsymbol{G}_h$ involves application of an FEM discretization.

The Poisson equation (175) can be solved exactly, however, for the sake of demonstration, we use a naive FEM implementation. The exact solution u is presented in Figure 3 (A) for three choices of $\boldsymbol{\theta}$. The design is considered fixed at $\boldsymbol{\xi} = (1/8, 7/8)$. Using this solution, we were able to verify that the bias of this FEM discretization converged at an observed rate of 2 (Figure 3 (B)). Moreover, this method has a computational cost of $\mathcal{O}(h^{-\gamma})$, where γ was estimated to be 1 (Figure 3 (C)). To balance the contribution of the inner number of samples and mesh discretization parameter, the following choice was selected for this example:

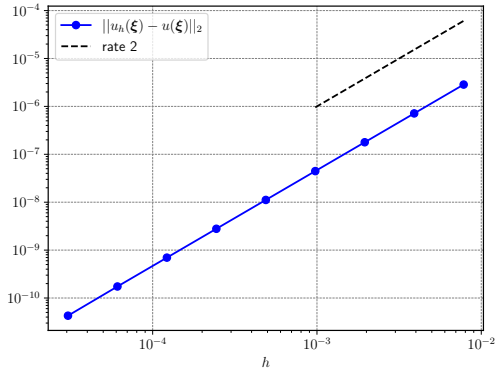
$$M_\ell = M_0 4^\ell, \quad 1 \leq \ell \leq L. \tag{179}$$

5.2.2. *Pilot runs.* We again verified the convergence rates predicted by the theoretical analysis using pilot runs. This verification enabled us to estimate the constant terms in the error bounds appearing in Corollary 2 required for the optimal setting specified (169). First, we verified that the bias induced by the FEM approximation converged at a rate of $\eta_{\mathrm{w}} \approx 2$ by fixing the inner number of samples M_L at the final level L as presented in Figure 4 (A). For this estimation, $N_L = 1$, $M_L = 32$ samples, and $S = 100$, $R = 1$ randomizations were used, and the bias was estimated using $h_L = 2^{-16}$ as a reference solution. To demonstrate generality in our estimations, we preferred this reference solution to the exact solution as the latter may not be available in general. Next, we confirmed that the variance contribution from the inner samples at the final level L converged at a rate of 2.23 and was slightly higher than the projected rate of 2 (Panel (B)), similar to the previous example. This estimation was conducted using $N_L = 1$ outer sample, $h_L = 2^{-4}$, and $S = 100$, $R = 1$ randomizations. The convergence of the variance of the level differences as a function of N_ℓ at the expected rate of 2 is illustrated in Panel (C), where $M_{\ell-1} = 4^4$, $M_\ell = 4^5$ samples, $h_{\ell-1} = 2^{-4}$, $h_\ell = 2^{-5}$, and $S = 100$ randomizations were applied. The variances of the single-level estimators converged at the same rate with a larger multiplicative constant and were independent of $M_{\ell-1}$, M_ℓ and $h_{\ell-1}$, h_ℓ. To demonstrate the convergence of the variance of the level difference as a function of M_ℓ, we kept $N_\ell = 1$ and $h_\ell = h_{\ell-1} = 2^{-4}$ fixed. This variance converged at rate 2, as seen in Panel (D) for $S = 100$ randomizations. Similarly, we kept $M_\ell = M_{\ell-1} = 8$ fixed to demonstrate the convergence of the variance of the level difference as a function of h_ℓ in Panel (E). This variance converged at rate of $2\eta_{\mathrm{s},d_1} \approx 4$, implying that for this example, $\eta_{\mathrm{w}} = \eta_{\mathrm{s},d_1} \approx 2$.

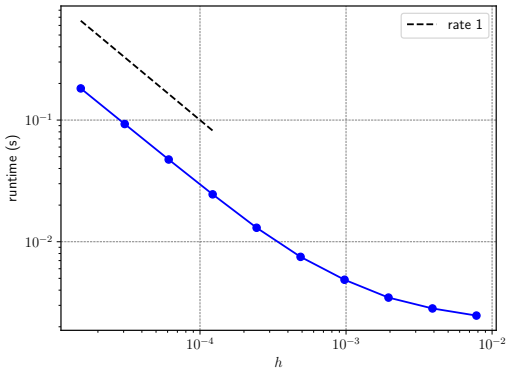
5.2.3. *Consistency and work analysis.* We determined the optimal final level L^* of the MLDLQMC estimator with inexact sampling for a given tolerance $TOL > 0$. The number of levels is displayed in Figure 5 (A) and the rate of increase was $\log_2(TOL^{-1/2})$, as expected. Finally, the accuracy and efficiency of the MLDLQMC estimator for optimal outer samples N_ℓ^*, where $1 \leq \ell \leq L^*$, was analyzed for various error tolerances TOL. In Panel (B), the total work of the MLDLQMC estimator, depending on the optimal L^* and N_ℓ^*, where $0 \leq \ell \leq L^*$, is presented as a function of the tolerance TOL. The projected rate of increase of $1/(1-\epsilon) + \gamma/\eta_{\mathrm{w}} \approx 1.5$ is only observed for $TOL < 10^{-10}$, indicating that the outer number of samples, for which the computational effort increases at a rate of 1, dominate the overall cost for higher tolerances. The cost is then dominated by the effort to estimate the final level L^* accurately, using M_{L^*} randomized inner samples and mesh discretization parameter h_{L^*}. This low tolerance does not correspond to a practically feasible computational cost and is only understood as a verification that the projected cost is attained as



(A) Exact solution u of the Poisson equation (175) in one dimension as a function of $\boldsymbol{\theta}$ and locations of the observations $\boldsymbol{\xi}$.



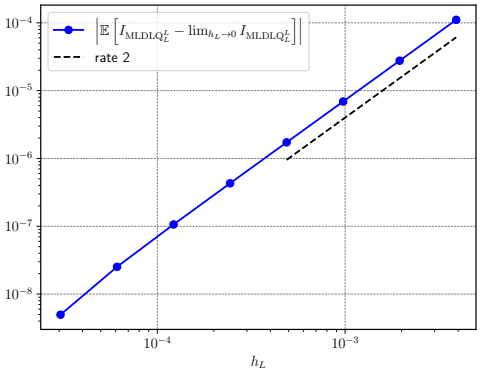
(B) The observed weak convergence rate $\eta_{\mathrm{w}} = 2$ of the FEM solution u_h of the Poisson equation (175) in one dimension with linear elements as a function of the discretization parameter h.



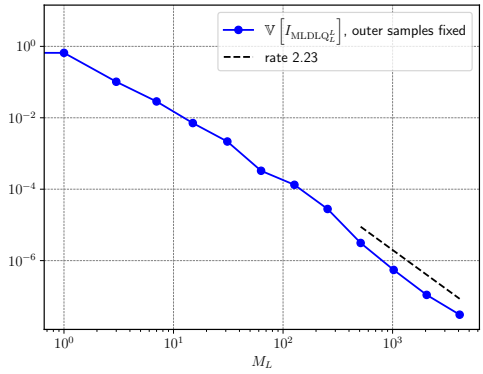
(C) Runtime in seconds to obtain the FEM solution u_h of the Poisson equation (175) yields the work rate $\gamma \approx 1$ as a function of the discretization parameter h (averaged over 20 runs).

FIGURE 3. Example 2: FEM approximation of the Poisson equation (175) in one dimension with linear elements, where the number of elements is specified as $\lceil 1/h \rceil$.

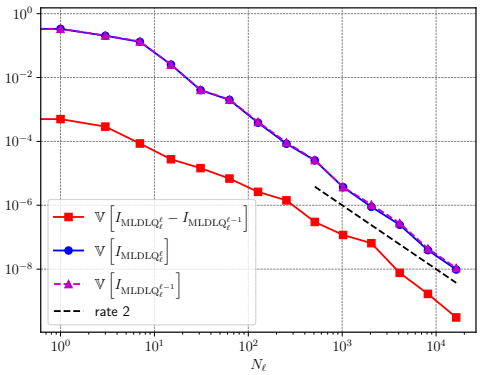
$TOL \to 0$. In Panel (C), 100 runs of the MLDLQMC estimator with $S = R = 1$ randomizations at each level for $TOL = 10^{-2}$, $TOL = 5 \times 10^{-3}$, and $TOL = 10^{-3}$ are presented, each remaining below the error tolerance projected by the CLT for the given L^* and N_ℓ^*, where $0 \le \ell \le L^*$ and $M_0 = 4^4$, $h_0 = 2^{-4}$. An average of 100 runs for $TOL = 5 \times 10^{-4}$ was used as a reference solution.



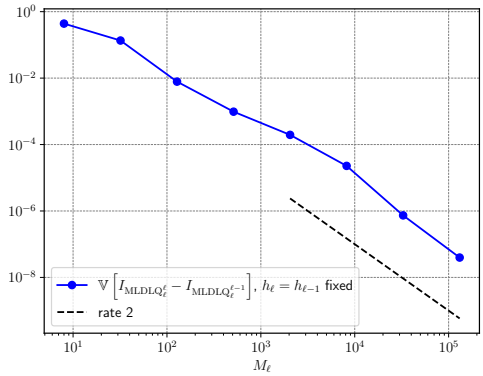
(A) The observed weak convergence rate of the discretization bias of the EIG estimation was $\eta_{\mathrm{w}} = 2$.



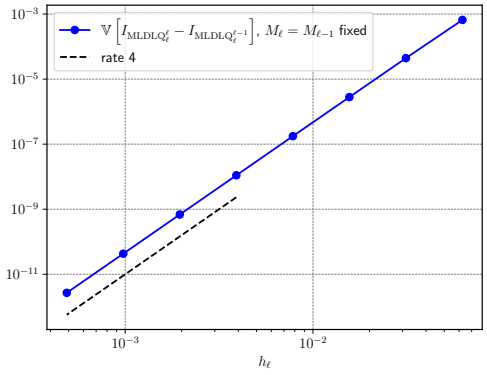
(B) The observed convergence rate of the variance of the inner samples at the final level L: 2.23.



(C) The observed convergence rate of the variance of 2, with a reduced multiplicative factor for the difference estimator.

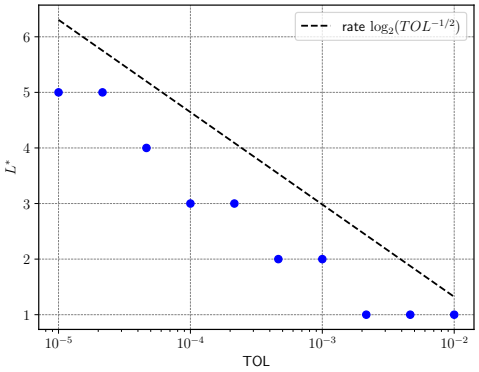


(D) Convergence of the level difference as a function of M_ℓ for N_ℓ and h_ℓ fixed at a rate of 2.

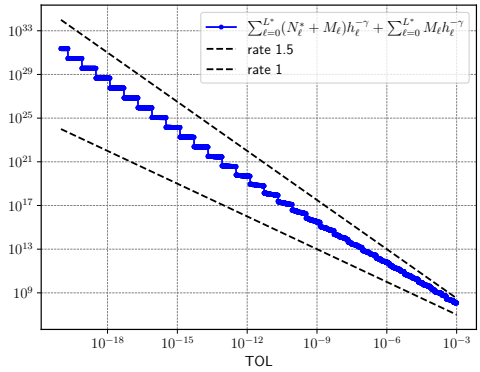


(E) Convergence of the level difference at a rate of $2\eta_{\mathrm{s},d_1} \approx 4$ as a function of h_ℓ for fixed values of N_ℓ and M_ℓ.

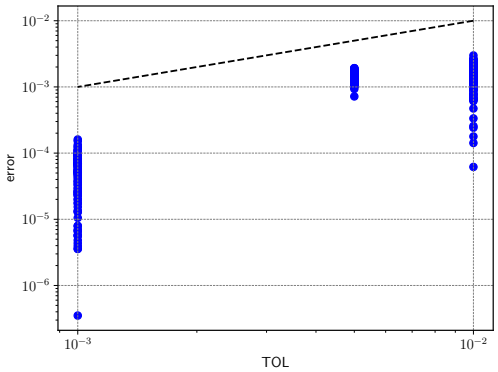
FIGURE 4. Example 2: Pilot runs for the MLDLQMC estimator with inexact sampling.



(A) The optimal number of levels L^* increased at a rate of $\log_2(TOL^{-1/2})$ as a function of the error tolerance $TOL > 0$.



(B) Total computational work of the MLDLQMC estimator as a function of the error tolerance TOL. The projected rate of 1.5 was only attained for low tolerances.



(C) Error vs. tolerance (TOL) for 100 runs of the MLDLQMC estimator. Each run resulted in an error less than the tolerance projected by the central limit theorem.

FIGURE 5. Example 2: The MLDLQMC estimator with inexact sampling resulted in error less than the projected tolerance TOL (C) at a cost increasing at the projected rate of 1.5 asymptotically as $TOL \to 0$ (B).

5.3. **Example 3: Duffing oscillator EIG example with inexact sampling.** We consider a three-degree-of-freedom spring–mass–damper system in which each spring's stiffness is augmented by a quadratic term in its relative displacement, yielding a Duffing-type oscillator. In this formulation, the effective stiffness of each spring grows nonlinearly with the square of the displacement difference between its endpoints. In our experiment, we observe the three mass positions at a single time T and use these observations to infer the unknown quadratic stiffness coefficients for each spring. A diagram describing the system is depicted in Figure 6.

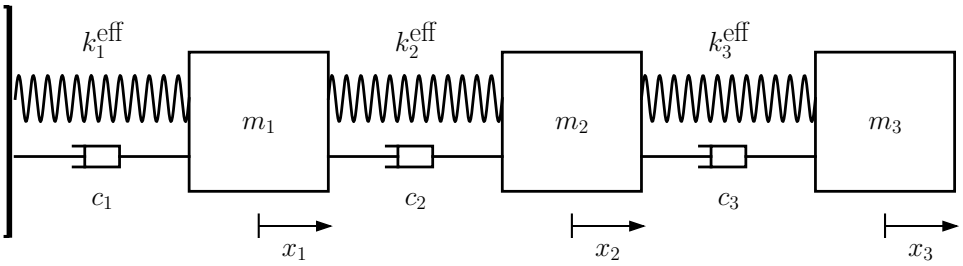


FIGURE 6. Example 3: diagram representing the spring–mass–damper system. Three masses (m_1, m_2, m_3) are connected by springs with stiffness coefficients $(k_1^{\text{eff}}, k_2^{\text{eff}}, k_3^{\text{eff}})$ and dampers with coefficients (c_1, c_2, c_3). Observations of the displacements (x_1, x_2, x_3) are used to infer the unknown quadratic contributions to the stiffness coefficients.

5.3.1. *Setting and notation.* We consider

$$\boldsymbol{G}(\boldsymbol{\theta}, \xi) = \boldsymbol{u}(\xi), \tag{180}$$

where $\boldsymbol{u} : [0, T] \mapsto \mathbb{R}^3$ is the state vector $\boldsymbol{u} = \boldsymbol{x} = (x_1, x_2, x_3)^\top$ and the mechanics of the system are governed by the nonlinear second-order differential equation

$$\boldsymbol{M}\ddot{\boldsymbol{x}} + \boldsymbol{C}\dot{\boldsymbol{x}} + \boldsymbol{K}(\boldsymbol{x})\,\boldsymbol{x} = \boldsymbol{0}, \quad \text{for all } 0 < t \le T, \tag{181}$$

$$\boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{182}$$

$$\dot{\boldsymbol{x}}(0) = \dot{\boldsymbol{x}}_0. \tag{183}$$

For this problem, we assume a diagonal mass matrix

$$\boldsymbol{M} = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}, \tag{184}$$

a damping matrix

$$\boldsymbol{C} = \begin{pmatrix} c_1 + c_2 & -c_2 & 0 \\ -c_2 & c_2 + c_3 & -c_3 \\ 0 & -c_3 & c_3 \end{pmatrix}, \tag{185}$$

and a state-dependent stiffness matrix

$$\boldsymbol{K}(\boldsymbol{x}) = \begin{pmatrix} k_1^{\text{eff}} + k_2^{\text{eff}} & -k_2^{\text{eff}} & 0 \\ -k_2^{\text{eff}} & k_2^{\text{eff}} + k_3^{\text{eff}} & -k_3^{\text{eff}} \\ 0 & -k_3^{\text{eff}} & k_3^{\text{eff}} \end{pmatrix}, \tag{186}$$

where

$$k_1^{\text{eff}} = k_1 + \theta_1\, x_1^2, \tag{187}$$

$$k_2^{\text{eff}} = k_2 + \theta_2\, (x_2 - x_1)^2, \tag{188}$$

$$k_3^{\text{eff}} = k_3 + \theta_3\, (x_3 - x_2)^2. \tag{189}$$

We moreover assume initial conditions $\boldsymbol{x}_0 = (100, 50, -50)^\top$mm and $\dot{\boldsymbol{x}}_0 = (0, 0, 0)^\top$mm/s. The fixed model parameters are presented in Table 1. The prior distribution of the nonlinear stiffness coefficients is modeled

i	m_i (kg)	k_i (N/mm)	c_i (N s/mm)
1	20	0.5	0.01
2	50	0.75	0.007
3	30	1	0.005

TABLE 1. Example 3: Duffing oscillator fixed model parameters.

as $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \overset{iid}{\sim} \mathcal{U}([6 \cdot 10^{-5}, 7 \cdot 10^{-5}]^{d_\theta})\text{N/mm}^3$, constituting the parameters of interest; hence $d_\theta = 3$. We only consider one observation at final time $\xi = T = 5\text{s}$; thus, the design ξ is in dimension $d_\xi = 1$. The system (181) describes three masses connected by springs and dampers, where x_i is the displacement of the ith mass for $1 \leq i \leq 3$. Moreover, the data observations are

$$\boldsymbol{Y} = \boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{\xi}) + \boldsymbol{\varepsilon}, \tag{190}$$

where $\boldsymbol{\varepsilon} \overset{iid}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{d_y \times d_y})$, for $\sigma^2 = 10^2\text{mm}^2$ and $\boldsymbol{I}_{d_y \times d_y}$ is the identity matrix in dimension $d_y = 3$, yielding $d_1 = d_\theta + d_y = 6$ and $d_2 = d_\theta = 3$ in the nested integration setting. We consider a discretization

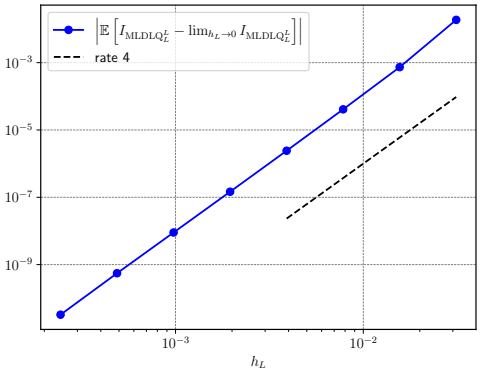
$$\boldsymbol{G}_h(\boldsymbol{\theta}, \xi) = \boldsymbol{u}_h(\xi), \tag{191}$$

where $\boldsymbol{u}_h$ is obtained via the Runge–Kutta 4 method with weak error convergence rate $\eta_\text{w} = 4$ and work rate $\gamma = 1$; thus $\gamma/\eta_\text{w} = 0.25$. The following choice was selected to balance the contribution of the inner number of samples M and the time stepping parameter h:

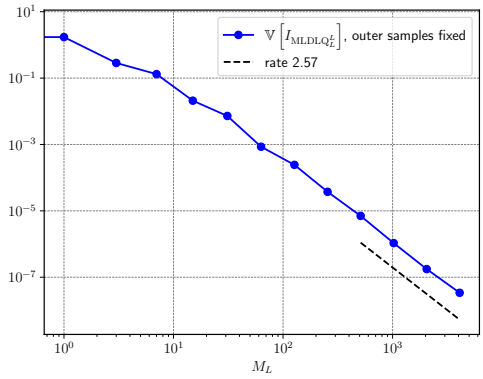
$$M_\ell = M_0 2^{4\ell} = M_0 16^\ell, \quad 1 \leq \ell \leq L. \tag{192}$$

5.3.2. *Pilot runs.* We verified that the bias induced by the Runge–Kutta 4 method converged at a rate of $\eta_\text{w} \approx 4$ by fixing the inner number of samples M_L at the final level L as presented in Figure 7 (A). For this estimation, $N_L = 1$, $M_L = 8$ samples, and $S = 100$, $R = 1$ randomizations were used, and the bias was estimated using $h_L = 2^{-12}$ as a reference solution. Next, we confirmed that the variance contribution from the inner samples at the final level L converged at a rate of 2.57, which was again higher than the projected rate of 2 (Panel (B)). This estimation was conducted using $N_L = 1$ outer sample, $h_L = 2^{-8}$, and $S = 100$, $R = 1$ randomizations. The convergence of the variance of the level differences as a function of N_ℓ at the expected rate of 2 is illustrated in Panel (C), where $M_{\ell-1} = 4 \cdot 16^2$, $M_\ell = 4 \cdot 16^3$ samples, $h_{\ell-1} = 2^{-4}$, $h_\ell = 2^{-5}$, and $S = 100$ randomizations were applied. The variances of the single-level estimators converged at the same rate with a larger multiplicative constant. To demonstrate the convergence of the variance of the level difference as a function of M_ℓ, we kept $N_\ell = 1$ and $h_\ell = h_{\ell-1} = 2^{-4}$ fixed. This variance converged at rate 2, as seen in Panel (D) for $S = 100$ randomizations. Similarly, we kept $M_\ell = M_{\ell-1} = 16^2$ fixed to demonstrate the convergence of the variance of the level difference as a function of h_ℓ in Panel (E). This variance converged at rate of $2\eta_{\text{s},d_1} \approx 8$, implying that for this example, $\eta_\text{w} = \eta_{\text{s},d_1} \approx 4$, as expected.

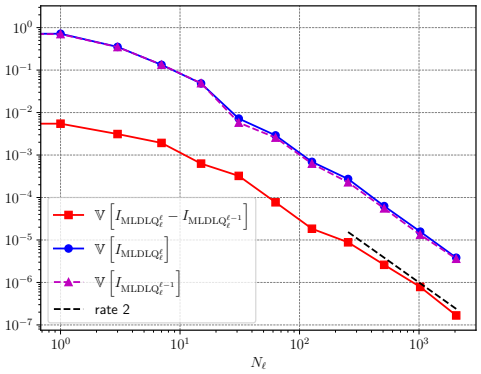
5.3.3. *Optimal work analysis.* The optimal final level L^* of the MLDLQMC estimator with inexact sampling for a given tolerance $TOL > 0$ is displayed in Figure 8 (A), where the rate of increase was $\log_2(TOL^{-1/\eta_\text{w}})$, for $\eta_\text{w} = 4$. In Panel (B), the total work of the MLDLQMC estimator is presented as a function of the tolerance TOL, where the projected rate of increase of $1/(1-\epsilon) + \gamma/\eta_\text{w} \approx 1.25$ is observed.



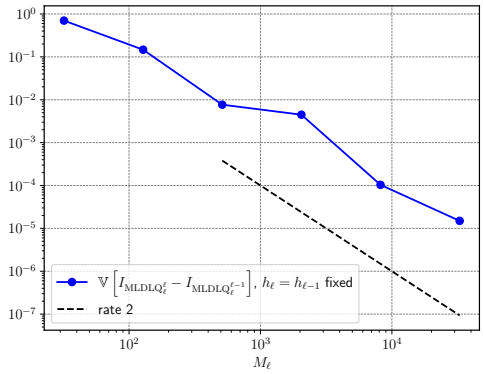
(A) The observed weak convergence rate of the discretization bias of the EIG estimation was $\eta_{\mathrm{w}} \approx 4$.



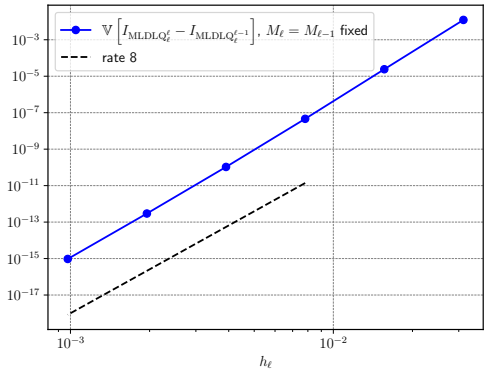
(B) The observed convergence rate of the variance of the inner samples at the final level L: 2.57.



(C) The observed convergence rate of the variance of approximately 2, with a reduced multiplicative factor for the difference estimator.

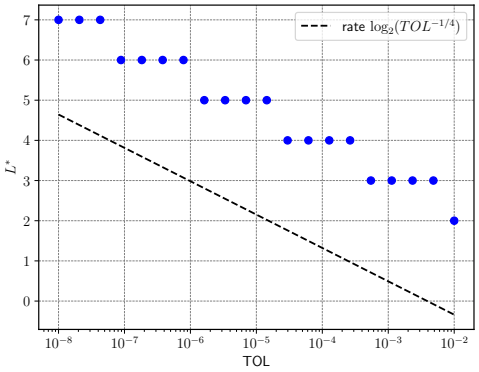


(D) Convergence of the level difference as a function of M_ℓ for N_ℓ and h_ℓ fixed at a rate of approximately 2.

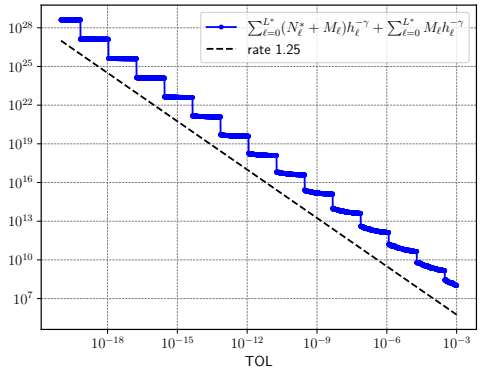


(E) Convergence of the level difference at a rate of $2\eta_{\mathrm{s},d_1} \approx 8$ as a function of h_ℓ for fixed values of N_ℓ and M_ℓ.

FIGURE 7. Example 3: Pilot runs for the MLDLQMC estimator with inexact sampling.



(A) The optimal number of levels L^* increased at a rate of $\log_2(TOL^{-1/4})$ as a function of the error tolerance $TOL > 0$.



(B) Total computational work of the MLDLQMC estimator as a function of the error tolerance TOL.

FIGURE 8. Example 3: The MLDLQMC estimator with inexact sampling resulted in a cost increasing at the projected rate of 1.25 asymptotically as $TOL \to 0$ (B).

6. Conclusion

This work proposes the multilevel double-loop randomized quasi-Monte Carlo estimator for nested integrals with inexact sampling. This estimator combines the multilevel strategy with deterministic and randomized quasi-Monte Carlo (QMC) methods to achieve high approximation accuracy when the inner integrand of a nested integration problem must itself be approximated. The bias and statistical errors are bounded asymptotically, yielding an estimate of the total number of inner and outer samples and the discretization parameter required to achieve a specified error tolerance. This estimator is then applied to expected information gain estimation, where singularities render standard QMC error bounds via the Koksma–Hlawka inequality pointless. This is attained via a truncation scheme of the Gaussian observation noise in the experiment model with error bounds introduced by Owen [26] and extended by He et al. [27]. The efficiency of the method is then demonstrated on three numerical examples.

7. Statements and Declarations

7.1. **Conflict of interest.** The authors have no conflicts to disclose.

7.2. **Data availability.** The data that support the findings of this study are available from the corresponding author upon reasonable request.

Appendix A. Auxiliary results for the error bounds to estimate the EIG

Lemma 4 (Bound on the Hardy–Krause variation of $g_h(\boldsymbol{y},\cdot)$)**.** *Let $g_h(\boldsymbol{y},\cdot)$ be as in* (150)*. Then, given Assumption 11, there exists $0 < b < \infty$ such that*

$$V_{\mathrm{HK}}\left(\left(\prod_{j\in u}\frac{\partial}{\partial y_j}\right) g_h(\boldsymbol{y},\cdot)\right) \le b\prod_{i=1}^{d_1}\min(y_i, 1-y_i)^{-A_i-\mathbb{1}_{\{i\in u\}}} \tag{193}$$

for all $\boldsymbol{y}\in[0,1]^{d_1}$, all $u\subseteq\{1,\dots,d_1\}$, and any $A_i > 0$, where $1\le i\le d_1$.

Proof. For the case where $d_1 = 2$, $d_2 = 1$, and

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \coloneqq \sigma^2, \tag{194}$$

for $\sigma > 0$, it follows that

$$g_h(\boldsymbol{y},x) = \exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x))^2 + \frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x) + \frac{1}{\sigma}\tilde{G}_h(x)\Phi^{-1}(y_2)\right). \tag{195}$$

As a first step, we demonstrate that there exists $0 < b < \infty$ such that

$$V_{\mathrm{HK}}\left(g_h(\boldsymbol{y},\cdot)\right) \le b\min(y_2, 1-y_2)^{-A_2}, \tag{196}$$

for any $A_2 > 0$. The Hardy–Krause variation is bounded as

$$\begin{aligned}
V_{\mathrm{HK}}\left(g_h(\boldsymbol{y},\cdot)\right) &= \int_{[0,1]}\left|\frac{\partial}{\partial x}\exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x))^2 + \frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x) + \frac{1}{\sigma}\tilde{G}_h(x)\Phi^{-1}(y_2)\right)\right|\mathrm{d}x,\\
&= \int_{[0,1]}\left|\frac{\mathrm{d}}{\mathrm{d}x}\tilde{G}_h(x)\left(-\frac{1}{\sigma^2}\tilde{G}_h(x) + \frac{1}{\sigma^2}\tilde{G}_h(y_1) + \frac{1}{\sigma}\Phi^{-1}(y_2)\right)\right.\\
&\qquad\left.\times\exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x))^2 + \frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x) + \frac{1}{\sigma}\tilde{G}_h(x)\Phi^{-1}(y_2)\right)\right|\mathrm{d}x,\\
&\le \int_{[0,1]}\left|\frac{\mathrm{d}}{\mathrm{d}x}\tilde{G}_h(x)\right|\left(\frac{1}{\sigma^2}\left|\tilde{G}_h(x)\right| + \frac{1}{\sigma^2}\left|\tilde{G}_h(y_1)\right| + \frac{1}{\sigma}\left|\Phi^{-1}(y_2)\right|\right)\\
&\qquad\times\exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x))^2 + \frac{1}{\sigma^2}|\tilde{G}_h(y_1)||\tilde{G}_h(x)| + \frac{1}{\sigma}|\tilde{G}_h(x)||\Phi^{-1}(y_2)|\right)\mathrm{d}x.
\end{aligned} \tag{197}$$

For the case in which $y_2 \to 0^+$, the approximation in [26, 36] is employed:

$$\Phi^{-1}(y_2) = -\sqrt{2}\sqrt{\log(y_2^{-1})} + o(1), \tag{198}$$

for which it holds that

$$\left|\Phi^{-1}(y_2)\right| \leq \sqrt{2}\sqrt{\log(y_2^{-1})} + P, \tag{199}$$

for any $P > 0$ and all $y_2 \leq y_{2,0}$ for some $y_{2,0} \in (0,1)$ as $y_2 \to 0^+$. Moreover, it follows that

$$\begin{aligned} \sqrt{2}\sqrt{\log(y_2^{-1})} &\leq \tilde{A}_2 \log(y_2^{-1}), \\ &= \log(y_2^{-\tilde{A}_2}) \end{aligned} \tag{200}$$

for any $\tilde{A}_2 > 0$ as $y_2 \to 0^+$, yielding, together with Assumption 11, that

$$\begin{aligned} &\int_{[0,1]} \left| \frac{\mathrm{d}}{\mathrm{d}x} \tilde{G}_h(x) \right| \left(\frac{1}{\sigma^2} \left| \tilde{G}_h(x) \right| + \frac{1}{\sigma^2} \left| \tilde{G}_h(y_1) \right| + \frac{1}{\sigma} \left| \Phi^{-1}(y_2) \right| \right) \\ &\qquad \times \exp\left(-\frac{1}{2\sigma^2} (\tilde{G}_h(x))^2 + \frac{1}{\sigma^2} |\tilde{G}_h(y_1)||\tilde{G}_h(x)| + \frac{1}{\sigma} |\tilde{G}_h(x)||\Phi^{-1}(y_2)| \right) \mathrm{d}x, \\ &\leq \int_{[0,1]} q \left(\frac{2q}{\sigma^2} + \frac{1}{\sigma} \log(y_2^{-\tilde{A}_2}) + \frac{1}{\sigma} P \right) \exp\left(\frac{1}{2\sigma^2} q^2 + \frac{1}{\sigma} q \log(y_2^{-\tilde{A}_2}) + \frac{1}{\sigma} qP \right) \mathrm{d}x, \\ &\leq q \left(\frac{2q}{\sigma^2} + \frac{1}{\sigma} \log(y_2^{-\tilde{A}_2}) + \frac{1}{\sigma} P \right) y_2^{-\frac{q\tilde{A}_2}{\sigma}} \exp\left(\frac{1}{2\sigma^2} q^2 + \frac{1}{\sigma} qP \right), \\ &\leq b y_2^{-A_2}, \end{aligned} \tag{201}$$

for any $A_2 > 0$ and an appropriately chosen constant $0 < b < \infty$ independent of y_2. For the case in which $y_2 \to 1^-$, the approximation in [26, 36] is employed:

$$\Phi^{-1}(y_2) = -\sqrt{2}\sqrt{\log((1-y_2)^{-1})} + o(1), \tag{202}$$

for which it holds that

$$\left|\Phi^{-1}(y_2)\right| \leq \sqrt{2}\sqrt{\log((1-y_2)^{-1})} + Q, \tag{203}$$

for any $Q > 0$ and all $y_2 \geq y_{2,1}$ for some $y_{2,1} \in (0,1)$, where $y_{2,1} > y_{2,0}$, as $y_2 \to 1^-$. It follows immediately that the same bound as in (201) applies for any $A_2 > 0$ and possibly different $0 < b < \infty$. The Hardy–Krause variation $V_{\mathrm{HK}}\left(g_h(\boldsymbol{y}, \cdot)\right)$ is trivially bounded independently of y_2 in between the values $0 < y_{2,0} < y_{2,1} < 1$, again for an appropriately chosen constant $0 < b < \infty$. Thus, we arrive at the following bound:

$$V_{\mathrm{HK}}\left(g_h(\boldsymbol{y}, \cdot)\right) \leq b \min(y_2, 1-y_2)^{-A_2}, \tag{204}$$

for any $A_2 > 0$. As a second step, we demonstrate that there exists $0 < b < \infty$ such that

$$V_{\mathrm{HK}}\left(\frac{\partial}{\partial y_2} g_h(\boldsymbol{y}, \cdot) \right) \leq b \min(y_2, 1-y_2)^{-A_2-1} \tag{205}$$

for any $A_2 > 0$. For this bound, we utilize the following result:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}y_2} \Phi^{-1}(y_2) &= \left(\Phi'(\Phi^{-1}(y_2)) \right)^{-1}, \\ &= \sqrt{2\pi} e^{\frac{1}{2}(\Phi^{-1}(y_2))^2}, \\ &\leq \sqrt{2\pi} e^{\left(\log(y_2^{-1}) + 2P\sqrt{\log(y_2^{-1})} + P^2 \right)}, \\ &\leq \sqrt{2\pi} e^{\left(\log(y_2^{-1}) + \tilde{A}_2 \log(y_2^{-1}) + P^2 \right)}, \\ &= \sqrt{2\pi} e^{P^2} y_2^{-\tilde{A}_2 - 1}, \end{aligned} \tag{206}$$

as $y_2 \to 0^+$ to bound

$$
\begin{aligned}
&V_{\mathrm{HK}}\left(\frac{\partial}{\partial y_2} g_h(\boldsymbol{y},\cdot)\right) \\
&\quad= \int_{[0,1]} \left| \frac{\partial}{\partial x}\frac{\partial}{\partial y_2} \exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x))^2 + \frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x) + \frac{1}{\sigma}\tilde{G}_h(x)\Phi^{-1}(y_2)\right)\right| \mathrm{d}x, \\
&\quad\le \int_{[0,1]} \left| \frac{\mathrm{d}}{\mathrm{d}y_2}\Phi^{-1}(y_2)\frac{1}{\sigma}\frac{\mathrm{d}}{\mathrm{d}x}\tilde{G}_h(x) \exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x))^2 + \frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x) + \frac{1}{\sigma}\tilde{G}_h(x)\Phi^{-1}(y_2)\right)\right| \\
&\qquad + \left| \frac{\mathrm{d}}{\mathrm{d}y_2}\Phi^{-1}(y_2)\frac{1}{\sigma}\tilde{G}_h(x)\frac{\mathrm{d}}{\mathrm{d}x}\tilde{G}_h(x)\left(-\frac{1}{\sigma^2}\tilde{G}_h(x) + \frac{1}{\sigma^2}\tilde{G}_h(y_1) + \frac{1}{\sigma}\Phi^{-1}(y_2)\right)\right. \\
&\qquad\quad \left.\times \exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x))^2 + \frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x) + \frac{1}{\sigma}\tilde{G}_h(x)\Phi^{-1}(y_2)\right)\right| \mathrm{d}x, \\
&\quad\le \tilde{b} y_2^{-\tilde{A}_2-1} \qquad (207)
\end{aligned}
$$

as $y_2 \to 0^+$ for some $0 < \tilde{b} < \infty$ and any $\tilde{A}_2 > 0$. From an analogous argument as $y_2 \to 0^-$, it follows that there exists $0 < b < \infty$ such that

$$
V_{\mathrm{HK}}\left(\frac{\partial}{\partial y_2} g_h(\boldsymbol{y},\cdot)\right) \le b \min(y_2, 1-y_2)^{-A_2-1} \qquad (208)
$$

for any $A_2 > 0$. The partial derivative with respect to y_1 follows trivially. The general case follows from Faà di Bruno's formula. □

Corollary 4 (Bound on $\bar{g}_h(\boldsymbol{y})$)**.** *Let $\bar{g}_h$ be as in* (152)*. Then, given Assumption 11, there exists $0 < b < \infty$ such that*

$$
\left| \left(\prod_{j\in u} \frac{\partial}{\partial y_j}\right) \bar{g}_h(\boldsymbol{y}) \right| \le b \prod_{i=1}^{d_1} \min(y_i, 1-y_i)^{-A_i - \mathbb{1}_{\{i\in u\}}}, \qquad (209)
$$

for all $\boldsymbol{y} \in [0,1]^{d_1}$, all $u \subseteq \{1,\dots,d_1\}$, and any $A_i > 0$, where $1 \le i \le d_1$.

Proof. For the case where $d_1 = 2$, $d_2 = 1$, and g_h as in (195), the result follows from the Leibniz integral rule and a similar argument as in (207). The general case follows from Faà di Bruno's formula as well. □

Corollary 5 (Bound on the nested integrand)**.** *Given Assumption11, for $f \equiv \log$ and g_h as in* (150)*, there exists $0 < \delta_K < 1/2$ and $0 < b < \infty$ such that*

$$
\left| f^{(2+d_1)}\left(K\bar{g}_h(\boldsymbol{y})\right)\right| \le b \prod_{i=d_2+1}^{d_1} \min(y_i, 1-y_i)^{-A_i} \qquad (210)
$$

for any $A_i > 0$, where $1 \le i \le d_1$, and all $K \in (\delta_K, 1-\delta_K)$. Here, $f^{(d)}$ denotes the dth derivative of f.

Proof. The following holds for $f \equiv \log$:

$$
\left| f^{(2+d_1)}\left(K\bar{g}_h(\boldsymbol{y})\right)\right| = (1+d_1)! K^{-2-d_1} (\bar{g}_h(\boldsymbol{y}))^{-2-d_1}. \qquad (211)
$$

In Corollary 4, $\bar{g}_h$ was bounded for the cases when $\Phi^{-1}(\boldsymbol{y}) \to -\infty$ and $\Phi^{-1}(\boldsymbol{y}) \to +\infty$; thus, similar bounds also apply to $(\bar{g}_h)^{-1}$. □

The following result is a direct consequence of [2, Lemma 2].

Lemma 5 (Inverse inequality for the inner integrand with truncated noise)**.** *Let $TOL > 0$ and $c(TOL)$ as in* (157)*. Moreover, let*

$$
F^{-1}_{c(TOL)}(\boldsymbol{y}) \coloneqq \Phi^{-1}\Big((2\Phi(c(TOL)) - 1)\boldsymbol{y} + (1 - \Phi(c(TOL)))\Big) \qquad (212)
$$

denote the inverse of the multivariate standard normal distribution truncated to the interval $[-c(TOL), c(TOL)]^{d_1-d_2}$. Then, it holds for

$$
g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}) \coloneqq \exp\left(-\frac{1}{2}\left\|\tilde{\boldsymbol{G}}_h(\boldsymbol{x})\right\|^2_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle \tilde{\boldsymbol{G}}_h(\boldsymbol{y}_1), \tilde{\boldsymbol{G}}_h(\boldsymbol{x})\right\rangle_{\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}} + \left\langle \tilde{\boldsymbol{G}}_h(\boldsymbol{x}), F^{-1}_{c(TOL)}(\boldsymbol{y}_2)\right\rangle_{\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}^{-\top}}\right) \qquad (213)
$$

and

$$\bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) \coloneqq \int_{[0,1]^{d_2}} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \tag{214}$$

that

$$\sup_{\boldsymbol{y} \in [0,1]^{d_1}} \left| \frac{V_{\mathrm{HK}}(g_h^{\mathrm{tr}}(\boldsymbol{y}, \cdot))}{\bar{g}_h^{\mathrm{tr}}(\boldsymbol{y})} \right| < c(TOL), \tag{215}$$

with V_{HK} *as in* (11).

Appendix B. Proof of Lemma 1

Proof of Lemma 1. We again consider the case where $d_1 = 2$, $d_2 = 1$, and g_h as in (195). As a first step, we demonstrate that there exists $0 < b < \infty$ such that

$$\left| f\left(\frac{1}{M} \sum_{m=1}^{M} g_h\left(\boldsymbol{y}, x^{(m)}\right) \right) \right| \leq b \min(y_2, 1 - y_2)^{-A_2} \tag{216}$$

for any $A_2 > 0$. For $f \equiv \log$ and g_h as in (195), it follows that

$$\begin{aligned} &\left| f\left(\frac{1}{M} \sum_{m=1}^{M} g_h\left(\boldsymbol{y}, x^{(m)}\right) \right) \right| \\ &\quad = \left| \log\left(\frac{1}{M} \sum_{m=1}^{M} \exp\left(-\frac{1}{2\sigma^2} (\tilde{G}_h(x^{(m)}))^2 + \frac{1}{\sigma^2} \tilde{G}_h(y_1) \tilde{G}_h(x^{(m)}) + \frac{1}{\sigma} \tilde{G}_h(x^{(m)}) \Phi^{-1}(y_2) \right) \right) \right|. \end{aligned} \tag{217}$$

For now, we consider the case where $\tilde{G}_h$ is nonnegative for all $x^{(m)}$, where $1 \leq m \leq M$, that is,

$$\min_{1 \leq m \leq M} \tilde{G}_h(x^{(m)}) \geq 0. \tag{218}$$

Then, as $y_2 \to 0^+$, it follows from Assumption 11 and the bound in (198) that there exists $0 < \tilde{b} < \infty$ such that

$$\begin{aligned} &\exp\left(-\frac{1}{2\sigma^2} (\tilde{G}_h(x^{(m)}))^2 + \frac{1}{\sigma^2} \tilde{G}_h(y_1) \tilde{G}_h(x^{(m)}) + \frac{1}{\sigma} \tilde{G}_h(x^{(m)}) \Phi^{-1}(y_2) \right) \\ &\quad \geq \exp\left(-\frac{3}{2\sigma^2} q^2 \right) \exp\left(-\frac{1}{\sigma} q \left| \Phi^{-1}(y_2) \right| \right), \\ &\quad \geq \tilde{b} y_2^{\tilde{A}_2}, \quad 1 \leq m \leq M, \end{aligned} \tag{219}$$

for any $\tilde{A}_2 > 0$. From this result, it follows that

$$\left| \log\left(\tilde{b} y_2^{\tilde{A}_2} \right) \right| \leq b y_2^{-A_2} \tag{220}$$

for some appropriately chosen $0 < b < \infty$ and any $A_2 > 0$ as $y_2 \to 0^+$. Now, we consider the case that $\tilde{G}_h$ is negative for at least one $x^{(m)}$, where $1 \leq m \leq M$, that is,

$$\min_{1 \leq m \leq M} \tilde{G}_h(x^{(m)}) < 0, \tag{221}$$

and furthermore, let

$$m^* \coloneqq \operatorname*{arg\,min}_{1 \leq m \leq M} \tilde{G}_h(x^{(m)}). \tag{222}$$

Thus, it follows that there exists $0 < \tilde{b} < \infty$ such that

$$\begin{aligned}
&\frac{1}{M}\sum_{m=1}^{M}\exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x^{(m)}))^2+\frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x^{(m)})+\frac{1}{\sigma}\tilde{G}_h(x^{(m)})\Phi^{-1}(y_2)\right)\\
&\quad\le \exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x^{(m^*)}))^2+\frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x^{(m^*)})+\frac{1}{\sigma}\tilde{G}_h(x^{(m^*)})\Phi^{-1}(y_2)\right),\\
&\quad\le \exp\left(\frac{1}{2\sigma^2}q^2\right)\exp\left(\frac{1}{\sigma}q\left|\Phi^{-1}(y_2)\right|\right),\\
&\quad\le \tilde{b}y_2^{-\tilde{A}_2}, \qquad (223)
\end{aligned}$$

for any $\tilde{A}_2 > 0$, and hence,

$$\left|\log\left(\tilde{b}y_2^{-\tilde{A}_2}\right)\right| \le by_2^{-A_2} \qquad (224)$$

for some appropriately chosen $0 < b < \infty$ and any $A_2 > 0$ as $y_2 \to 0^+$. The case where $y_2 \to 1^-$ follows by a similar argument and the bound (202). Thus, it is demonstrated that

$$\left|f\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y}, x^{(m)}\right)\right)\right| \le b\min(y_2, 1-y_2)^{-A_2} \qquad (225)$$

for any $A_2 > 0$. As a second step, it follows that

$$\begin{aligned}
&\left|\frac{\partial}{\partial y_2}f\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y}, \boldsymbol{x}^{(m)}\right)\right)\right| \qquad (226)\\
&= \left|\frac{\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\mathrm{d}}{\mathrm{d}y_2}\Phi^{-1}(y_2)\frac{1}{\sigma}\tilde{G}_h(x^{(m)})\right)\exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x^{(m)}))^2+\frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x^{(m)})+\frac{1}{\sigma}\tilde{G}_h(x^{(m)})\Phi^{-1}(y_2)\right)}{\frac{1}{M}\sum_{m=1}^{M}\exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x^{(m)}))^2+\frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x^{(m)})+\frac{1}{\sigma}\tilde{G}_h(x^{(m)})\Phi^{-1}(y_2)\right)}\right|.
\end{aligned}$$

For the numerator in equation (226), there exists $0 < \tilde{b} < \infty$ such that

$$\begin{aligned}
&\left|\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\mathrm{d}}{\mathrm{d}y_2}\Phi^{-1}(y_2)\frac{1}{\sigma}\tilde{G}_h(x^{(m)})\right)\exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x^{(m)}))^2+\frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x^{(m)})+\frac{1}{\sigma}\tilde{G}_h(x^{(m)})\Phi^{-1}(y_2)\right)\right| \qquad (227)\\
&\le \tilde{b}y_2^{-\tilde{A}_2-1},
\end{aligned}$$

for any $\tilde{A}_2 > 0$ by a similar derivation and the bound (206). For the denominator in equation (226), it holds that

$$\left|\frac{1}{M}\sum_{m=1}^{M}\exp\left(-\frac{1}{2\sigma^2}(\tilde{G}_h(x^{(m)}))^2+\frac{1}{\sigma^2}\tilde{G}_h(y_1)\tilde{G}_h(x^{(m)})+\frac{1}{\sigma}\tilde{G}_h(x^{(m)})\Phi^{-1}(y_2)\right)\right| \le \tilde{b}y_2^{-\tilde{A}_2}, \qquad (228)$$

again by a similar derivation. Thus, it follows that there exists $0 < b < \infty$ such that

$$\left|\frac{\partial}{\partial y_2}f\left(\frac{1}{M}\sum_{m=1}^{M} g_h\left(\boldsymbol{y}, \boldsymbol{x}^{(m)}\right)\right)\right| \le by_2^{-A_2-1} \qquad (229)$$

for any $A_2 > 0$. The general case follows similarly by Faà di Bruno's formula. □

Appendix C. Proof of Lemma 3

Proof of Lemma 3. We introduce the following notations:

$$\varphi_{h,M}(\boldsymbol{y}) = (\lambda(\boldsymbol{y}))^2\,\mu(\boldsymbol{y}), \qquad (230)$$

where

$$\lambda(\boldsymbol{y}) \coloneqq \frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}), \qquad (231)$$

and

$$\mu(\boldsymbol{y}) \coloneqq \int_0^1 f''\left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) + s\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s. \tag{232}$$

For the case $d_1 = 2$, it must be verified that the following inequalities hold:

$$|\varphi_{h,M}(y_1,y_2)| \le \left(C^2_{\epsilon,d_2} M^{-2+2\epsilon} + C^2_{\eta_{\mathrm{s},d_1}} h^{2\eta_{\mathrm{s},d_1}}\right) b \min(y_1, 1-y_1)^{-A_1} \min(y_2, 1-y_2)^{-A_2}, \tag{233}$$

$$\left|\frac{\partial}{\partial y_1}\varphi_{h,M}(y_1,y_2)\right| \le \left(C^2_{\epsilon,d_2} M^{-2+2\epsilon} + C^2_{\eta_{\mathrm{s},d_1}} h^{2\eta_{\mathrm{s},d_1}}\right) b \min(y_1, 1-y_1)^{-A_1-1} \min(y_2, 1-y_2)^{-A_2}, \tag{234}$$

$$\left|\frac{\partial}{\partial y_2}\varphi_{h,M}(y_1,y_2)\right| \le \left(C^2_{\epsilon,d_2} M^{-2+2\epsilon} + C^2_{\eta_{\mathrm{s},d_1}} h^{2\eta_{\mathrm{s},d_1}}\right) b \min(y_1, 1-y_1)^{-A_1} \min(y_2, 1-y_2)^{-A_2-1}, \tag{235}$$

$$\left|\frac{\partial^2}{\partial y_1 \partial y_2}\varphi_{h,M}(y_1,y_2)\right| \le \left(C^2_{\epsilon,d_2} M^{-2+2\epsilon} + C^2_{\eta_{\mathrm{s},d_1}} h^{2\eta_{\mathrm{s},d_1}}\right) b \min(y_1, 1-y_1)^{-A_1-1} \min(y_2, 1-y_2)^{-A_2-1}. \tag{236}$$

We start with the Condition (233), for which it follows that

$$\begin{aligned} |\varphi_{h,M}(y_1,y_2)| &= \left|(\lambda(y_1,y_2))^2\,\mu(y_1,y_2)\right| \\ &= (\lambda(y_1,y_2))^2\,|\mu(y_1,y_2)|. \end{aligned} \tag{237}$$

The bound for the first term in (237) is

$$\begin{aligned} (\lambda(\boldsymbol{y}))^2 &= \left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)^2 \\ &= \left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) + \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)^2, \\ &\le 2\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y})\right)^2 + 2\left(\bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)^2. \end{aligned} \tag{238}$$

By the Koksma–Hlawka inequality (10) and Lemma 4, there exists $0 < \tilde{b} < \infty$ such that the first term in (238) is bounded as follows:

$$\begin{aligned} 2\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y})\right)^2 &\le 2V_{\mathrm{HK}}(g_h^{\mathrm{tr}}(\boldsymbol{y},\cdot))^2 C^2_{\epsilon,d_2} M^{-2+2\epsilon}, \\ &\le 2C^2_{\epsilon,d_2} M^{-2+2\epsilon} \tilde{b}^2 \min(y_1, 1-y_1)^{-2\tilde{A}_1} \min(y_2, 1-y_2)^{-2\tilde{A}_2} \end{aligned} \tag{239}$$

for any $\tilde{A}_1, \tilde{A}_2 > 0$. By Assumption 12 and Corollary 4, there exists $0 < \tilde{b} < \infty$ such that the second term in (238) is bounded as follows:

$$\begin{aligned} 2\left(\bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)^2 &\le 2(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}))^2 C^2_{\eta_{\mathrm{s}}} h^{2\eta_{\mathrm{s}}}, \\ &\le 2C^2_{\eta_{\mathrm{s}}} h^{2\eta_{\mathrm{s}}} \tilde{b}^2 \min(y_1, 1-y_1)^{-2\tilde{A}_1} \min(y_2, 1-y_2)^{-2\tilde{A}_2} \end{aligned} \tag{240}$$

for any $\tilde{A}_1, \tilde{A}_2 > 0$. Thus, the bound for the first term in (237) follows as

$$(\lambda(\boldsymbol{y}))^2 \le \left(C^2_{\epsilon,d_2} M^{-2+2\epsilon} + C^2_{\eta_{\mathrm{s}}} h^{2\eta_{\mathrm{s}}}\right) 2\tilde{b}^2 \min(y_1, 1-y_1)^{-2\tilde{A}_1} \min(y_2, 1-y_2)^{-2\tilde{A}_2}. \tag{241}$$

To bound the second term in (237), it follows directly from the definition (150) that

$$\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) + s\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right) > 0. \tag{242}$$

For $f \equiv \log$, it is thus verified that $|f''|$ is monotonically decreasing for positive inputs. It follows that

$$\begin{aligned}
|\mu(\boldsymbol{y})| &= \left| \int_0^1 f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) + s \left(\frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) (1-s) \, \mathrm{d}s \right| \\
&\leq \int_0^1 \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) + s \left(\frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right| (1-s) \, \mathrm{d}s, \\
&\leq \int_0^1 \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - s \left| \frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \right) \right| (1-s) \, \mathrm{d}s, \\
&\leq \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left| \frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \right) \right| \int_0^1 (1-s) \, \mathrm{d}s, \\
&= \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left| \frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \right) \right|. \qquad (243)
\end{aligned}$$

From the Koksma–Hlawka inequality (10) and Assumption 12 it follows that

$$\begin{aligned}
&\frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left| \frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \right) \right| \\
&= \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left| \frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) + \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \right) \right|, \\
&\leq \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(\left| \frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) \right| + \left| \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \right) \right) \right|, \\
&\leq \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(C_{\epsilon,d_2} M^{-1+\epsilon} |V_{\mathrm{HK}}(g_h^{\mathrm{tr}}(\boldsymbol{y}, \cdot))| + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right|. \qquad (244)
\end{aligned}$$

Next, Lemma 5 provides that

$$\begin{aligned}
&\frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(C_{\epsilon,d_2} M^{-1+\epsilon} |V_{\mathrm{HK}}(g_h^{\mathrm{tr}}(\boldsymbol{y}, \cdot))| + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right| \\
&\leq \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(C_{\epsilon,d_2} M^{-1+\epsilon} c(TOL) \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right|. \qquad (245)
\end{aligned}$$

Again from Assumption 12 it follows that

$$\begin{aligned}
&\frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(C_{\epsilon,d_2} M^{-1+\epsilon} c(TOL) \bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right| \\
&= \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(C_{\epsilon,d_2} M^{-1+\epsilon} c(TOL) |\bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) + \bar{g}^{\mathrm{tr}}(\boldsymbol{y})| + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right|, \\
&\leq \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(C_{\epsilon,d_2} M^{-1+\epsilon} c(TOL) (|\bar{g}_h^{\mathrm{tr}}(\boldsymbol{y}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})| + \bar{g}^{\mathrm{tr}}(\boldsymbol{y})) + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right|, \\
&\leq \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(C_{\epsilon,d_2} M^{-1+\epsilon} c(TOL) (1 + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}}) \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right|. \qquad (246)
\end{aligned}$$

Introducing the notation

$$k(TOL) \coloneqq \frac{C_{\epsilon,d_2} c(TOL)(1 + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}}) \bar{g}^{\mathrm{tr}}(\boldsymbol{y})}{M^{1-\epsilon}} + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}}, \qquad (247)$$

it holds for

$$h < C_{\mathrm{s}}^{-\frac{1}{\eta_{\mathrm{s}}}} \qquad (248)$$

and

$$M > \left(\frac{C_{\epsilon,d_2} c(TOL)(1 + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}})}{(1 - C_{\mathrm{s}} h^{\eta_{\mathrm{s}}})} \right)^{\frac{1}{1-\epsilon}}, \tag{249}$$

that $0 \leq k(TOL) < 1$. Moreover, for

$$K(TOL) \coloneqq 1 - k(TOL), \tag{250}$$

it holds that $0 < K(TOL) \leq 1$. Thus, we write

$$\begin{aligned} \frac{1}{2} \left| f'' \left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) - \left(C_{\epsilon,d_2} M^{-1+\epsilon} c(TOL)(1 + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}}) \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right) \right| &= \frac{1}{2} \left| f'' \left((1 - k(TOL)) \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right|, \\ &= \frac{1}{2} \left| f'' \left(K(TOL) \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right|, \end{aligned} \tag{251}$$

and from Corollary 5, it follows that there exists $0 < \tilde{b} < \infty$ such that

$$\frac{1}{2} \left| f'' \left(K(TOL) \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right) \right| \leq \frac{1}{2} \tilde{b} \min(y_2, 1 - y_2)^{-\tilde{A}_2} \tag{252}$$

for any $\tilde{A}_2 > 0$. Thus, the bound for the second term in (237) follows as

$$|\mu(\boldsymbol{y})| \leq \frac{1}{2} \tilde{b} \min(y_2, 1 - y_2)^{-\tilde{A}_2}. \tag{253}$$

It follows that for $h < C_{\mathrm{s}}^{-1/\eta_{\mathrm{s}}}$ and

$$M > \left(\frac{C_{\epsilon,d_2} c(TOL)(1 + C_{\mathrm{s}} h^{\eta_{\mathrm{s}}})}{(1 - C_{\mathrm{s}} h^{\eta_{\mathrm{s}}})} \right)^{\frac{1}{1-\epsilon}}, \tag{254}$$

there exists $0 < b < \infty$ such that

$$|\varphi_{h,M}(y_1, y_2)| \leq \left(C_{\epsilon,d_2}^2 M^{-2+2\epsilon} + C_{\eta_{\mathrm{s}}}^2 h^{2\eta_{\mathrm{s}}} \right) b \min(y_1, 1 - y_1)^{-A_1} \min(y_2, 1 - y_2)^{-A_2} \tag{255}$$

for any $A_1, A_2 > 0$, and thus the verification of Condition (233). For Condition (234), it follows that

$$\begin{aligned} \left| \frac{\partial}{\partial y_1} \varphi_{h,M}(\boldsymbol{y}) \right| &= \left| \frac{\partial}{\partial y_1} \left((\lambda(\boldsymbol{y}))^2 \mu(\boldsymbol{y}) \right) \right|, \\ &\leq \left| \frac{\partial}{\partial y_1} (\lambda(\boldsymbol{y}))^2 \right| |\mu(\boldsymbol{y})| + (\lambda(\boldsymbol{y}))^2 \left| \frac{\partial}{\partial y_1} \mu(\boldsymbol{y}) \right|. \end{aligned} \tag{256}$$

The bound on $(\lambda(\boldsymbol{y}))^2$ follows from (241) and the bound on $|\mu(\boldsymbol{y})|$ follows from (253). Next, it follows that

$$\left| \frac{\partial}{\partial y_1} (\lambda(\boldsymbol{y}))^2 \right| = 2 \left| \frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \left| \frac{1}{M} \sum_{m=1}^{M} \frac{\partial}{\partial y_1} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \frac{\partial}{\partial y_1} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right|. \tag{257}$$

The Bound on the first term in (257) follows by a similar derivation as the bound in (241) as

$$2 \left| \frac{1}{M} \sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \leq \left(C_{\epsilon,d_2} M^{-1+\epsilon} + C_{\eta_{\mathrm{s}}} h^{\eta_{\mathrm{s}}} \right) 2\tilde{b} \min(y_1, 1 - y_1)^{-\tilde{A}_1} \min(y_2, 1 - y_2)^{-\tilde{A}_2}. \tag{258}$$

A similar bound applies to the second term in (257),

$$\left| \frac{1}{M} \sum_{m=1}^{M} \frac{\partial}{\partial y_1} g_h^{\mathrm{tr}}(\boldsymbol{y}, \boldsymbol{x}^{(m)}) - \frac{\partial}{\partial y_1} \bar{g}^{\mathrm{tr}}(\boldsymbol{y}) \right| \leq \left(C_{\epsilon,d_2} M^{-1+\epsilon} + C_{\eta_{\mathrm{s},1}} h^{\eta_{\mathrm{s},1}} \right) \tilde{b} \min(y_1, 1 - y_1)^{-\tilde{A}_1 - 1} \min(y_2, 1 - y_2)^{-\tilde{A}_2}, \tag{259}$$

except for an extra factor $\min(y_1, 1 - y_1)$ and a possibly different strong rate $\eta_{\mathrm{s},1} \leq \eta_{\mathrm{s}}$ and associated constant $C_{\eta_{\mathrm{s},1}}$ because of the partial derivative. It follows from the inequality of arithmetic and geometric means that

$$\left| \frac{\partial}{\partial y_1} (\lambda(\boldsymbol{y}))^2 \right| \leq \left(C_{\epsilon,d_2}^2 M^{-2+2\epsilon} + C_{\eta_{\mathrm{s},1}}^2 h^{2\eta_{\mathrm{s},1}} \right) 4\tilde{b}^2 \min(y_1, 1 - y_1)^{-2\tilde{A}_1 - 1} \min(y_2, 1 - y_2)^{-2\tilde{A}_2}. \tag{260}$$

Moreover, for the remaining term in (256), it follows from the Leibniz integral rule that

$$\begin{aligned}\left|\frac{\partial}{\partial y_1}\mu(\boldsymbol{y})\right| &= \left|\frac{\partial}{\partial y_1}\int_0^1 f''\left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s\right| \\ &= \left|\int_0^1 \frac{\partial}{\partial y_1} f''\left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s\right|, \\ &= \left|\int_0^1 f'''\left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)\right. \\ &\qquad \left.\times\left(\frac{\partial}{\partial y_1}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M}\frac{\partial}{\partial y_1} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)})-\frac{\partial}{\partial y_1}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s\right|. \end{aligned} \tag{261}$$

A similar derivation as for the bound in (253) yields that for

$$h < C_{\mathrm{s},1}^{-\frac{1}{\eta_{\mathrm{s},1}}} \tag{262}$$

and

$$M > \left(\frac{C_{\epsilon,d_2}c(TOL)(1+C_{\mathrm{s},1}h^{\eta_{\mathrm{s},1}})}{(1-C_{\mathrm{s},1}h^{\eta_{\mathrm{s},1}})}\right)^{\frac{1}{1-\epsilon}}, \tag{263}$$

there exists $0<\tilde{b}<\infty$ such that

$$\begin{aligned}&\left|\int_0^1 f'''\left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)})-\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)\right. \\ &\qquad \left.\times\left(\frac{\partial}{\partial y_1}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})+s\left(\frac{1}{M}\sum_{m=1}^{M}\frac{\partial}{\partial y_1} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)})-\frac{\partial}{\partial y_1}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s\right|, \\ &\le \frac{1}{2}\left|f'''\left(K(TOL)|\bar{g}^{\mathrm{tr}}(\boldsymbol{y})|\right)\right|\left|K(TOL)\frac{\partial}{\partial y_1}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right|, \\ &\le \frac{1}{2}\tilde{b}^2\min(y_1,1-y_1)^{-\tilde{A}_1-1}\min(y_2,1-y_2)^{-\tilde{A}_2} \end{aligned} \tag{264}$$

for any $\tilde{A}_1,\tilde{A}_2>0$. Combining the above results yields that there exists $0<b<\infty$ such that

$$\left|\frac{\partial}{\partial y_1}\varphi_{h,M}(y_1,y_2)\right| \le \left(C_{\epsilon,d_2}^2 M^{-2+2\epsilon}+C_{\eta_{\mathrm{s},1}}^2 h^{2\eta_{\mathrm{s},1}}\right) b\min(y_1,1-y_1)^{-A_1-1}\min(y_2,1-y_2)^{-A_2} \tag{265}$$

for any $A_1,A_2>0$. A similar bound applies to the Condition (235), that is,

$$\left|\frac{\partial}{\partial y_2}\varphi_{h,M}(y_1,y_2)\right| \le \left(C_{\epsilon,d_2}^2 M^{-2+2\epsilon}+C_{\eta_{\mathrm{s},2}}^2 h^{2\eta_{\mathrm{s},2}}\right) b\min(y_1,1-y_1)^{-A_1}\min(y_2,1-y_2)^{-A_2-1}. \tag{266}$$

Finally, for Condition (236), it follows that

$$\frac{\partial^2}{\partial y_1\partial y_2}\varphi_{h,M}(\boldsymbol{y}) = \frac{\partial^2}{\partial y_1\partial y_2}\left(\lambda(\boldsymbol{y})\right)^2\mu(\boldsymbol{y})+\frac{\partial}{\partial y_1}\left(\lambda(\boldsymbol{y})\right)^2\frac{\partial}{\partial y_2}\mu(\boldsymbol{y})+\frac{\partial}{\partial y_2}\left(\lambda(\boldsymbol{y})\right)^2\frac{\partial}{\partial y_1}\mu(\boldsymbol{y})+\left(\lambda(\boldsymbol{y})\right)^2\frac{\partial^2}{\partial y_1\partial y_2}\mu(\boldsymbol{y}). \tag{267}$$

The following bound applies:

$$\begin{aligned}
\left|\frac{\partial^2}{\partial y_1 \partial y_2}\left(\lambda(\boldsymbol{y})\right)^2\right| &= \left|\frac{\partial^2}{\partial y_1 \partial y_2}\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)^2\right| \\
&= 2\left|\frac{\partial}{\partial y_1}\left(\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\left(\frac{1}{M}\sum_{m=1}^{M} \frac{\partial}{\partial y_2} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \frac{\partial}{\partial y_2}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)\right|, \\
&\leq 2\left|\left(\frac{1}{M}\sum_{m=1}^{M} \frac{\partial}{\partial y_1} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \frac{\partial}{\partial y_1}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right|\left|\left(\frac{1}{M}\sum_{m=1}^{M} \frac{\partial}{\partial y_2} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \frac{\partial}{\partial y_2}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right|, \\
&\quad + 2\left|\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right|\left|\left(\frac{1}{M}\sum_{m=1}^{M} \frac{\partial^2}{\partial y_1 \partial y_2} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \frac{\partial^2}{\partial y_1 \partial y_2}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right|. \qquad (268)
\end{aligned}$$

The bound on the first term in (268) follows by a derivation similar to that in (260):

$$\begin{aligned}
&2\left|\left(\frac{1}{M}\sum_{m=1}^{M} \frac{\partial}{\partial y_1} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \frac{\partial}{\partial y_1}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right|\left|\left(\frac{1}{M}\sum_{m=1}^{M} \frac{\partial}{\partial y_2} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \frac{\partial}{\partial y_2}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right| \\
&\leq 4\left(C_{\epsilon,d_2}^2 M^{-2+2\epsilon} + C_{\eta_{\mathrm{s},d_1}}^2 h^{2\eta_{\mathrm{s},d_1}}\right)\tilde{b}^2 \min(y_1, 1-y_1)^{-2\tilde{A}_1-1}\min(y_2, 1-y_2)^{-2\tilde{A}_2-1} \qquad (269)
\end{aligned}$$

for

$$h < C_{\mathrm{s},d_1}^{-\frac{1}{\eta_{\mathrm{s},d_1}}} \qquad (270)$$

and

$$\begin{aligned}
M &> \left(\frac{C_{\epsilon,d_2} c(TOL)(1 + C_{\mathrm{s},d_1} h^{\eta_{\mathrm{s},d_1}})}{(1 - C_{\mathrm{s},d_1} h^{\eta_{\mathrm{s},d_1}})}\right)^{\frac{1}{1-\epsilon}}, \\
&= \tilde{M}(TOL). \qquad (271)
\end{aligned}$$

The same bound also applies to the second term in (268). Moreover, it follows from the Leibniz integral rule that

$$\begin{aligned}
\left|\frac{\partial^2}{\partial y_1 \partial y_2}\mu(\boldsymbol{y})\right| &= \left|\frac{\partial^2}{\partial y_1 \partial y_2}\int_0^1 f''\left(\bar{g}^{\mathrm{tr}}(\boldsymbol{y}) + s\left(\frac{1}{M}\sum_{m=1}^{M} g_h^{\mathrm{tr}}(\boldsymbol{y},\boldsymbol{x}^{(m)}) - \bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\right)(1-s)\,\mathrm{d}s\right| \\
&\leq \frac{1}{2} f''''\left(K(TOL)\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\left|K(TOL)\frac{\partial}{\partial y_1}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right|\left|K(TOL)\frac{\partial}{\partial y_2}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right| \\
&\quad + \frac{1}{2} f'''\left(K(TOL)\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right)\left|K(TOL)\frac{\partial^2}{\partial y_1 \partial y_2}\bar{g}^{\mathrm{tr}}(\boldsymbol{y})\right|, \\
&\leq \tilde{b}^2 \min(y_1, 1-y_1)^{-2\tilde{A}_1-1}\min(y_2, 1-y_2)^{-2\tilde{A}_2-1}. \qquad (272)
\end{aligned}$$

Combining the previous results, it is demonstrated that for $h < C_{\mathrm{s},d_1}^{-1/\eta_{\mathrm{s},d_1}}$ and $M > \tilde{M}(TOL)$, there exists $0 < b < \infty$ such that

$$\left|\frac{\partial^2}{\partial y_1 \partial y_2}\varphi_{h,M}(y_1, y_2)\right| \leq \left(C_{\epsilon,d_2}^2 M^{-2+2\epsilon} + C_{\eta_{\mathrm{s},d_1}}^2 h^{2\eta_{\mathrm{s},d_1}}\right) b \min(y_1, 1-y_1)^{-A_1-1}\min(y_2, 1-y_2)^{-A_2-1} \qquad (273)$$

for any $A_1, A_2 > 0$. The general case for an arbitrary dimension is a consequence of Faà di Bruno's formula. □

References

[1] Ryan KJ. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12:585–603, 2003.

[2] Bartuska A, Carlon AG, Espath L, Krumscheid S, and Tempone R. Double-loop randomized quasi-Monte Carlo estimator for nested integration. *arXiv preprint arXiv:2302.14119*, 2024.

[3] Beck J, Dia BM, Espath LF, Long Q, and Tempone R. Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain. *Computer Methods in Applied Mechanics and Engineering*, 334:523–553, 2018.

[4] Beck J, Dia BM, Espath LF, and Tempone R. Multilevel double loop Monte Carlo and stochastic collocation methods with importance sampling for Bayesian optimal experimental design. *International Journal for Numerical Methods in Engineering*, 121:3482–3503, 2020.
[5] Chaloner K and Verdinelli I. Bayesian experimental design: a review. *Statistical Sciences*, 10:273–304, 1995.
[6] Lindley DV. On a measure of information provided by an experiment. *Annals of Mathematical Statistics*, 27:986–1005, 1956.
[7] Kullback S and Leibler RA. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.
[8] Kullback S. *Information theory and statistics*. Wiley, 1959.
[9] Long Q, Scavino M, Tempone R, and Wang S. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39, 2013.
[10] Giles MB. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
[11] Tsilifis P, Ghanem RG, and Hajali P. Efficient Bayesian experimentation using an expected information gain lower bound. *SIAM/ASA Journal on Uncertainty Quantification*, 5:30–62, 2017.
[12] Goda T, Daisuke M, Kei T, and Kozo S. Decision-theoretic sensitivity analysis for reservoir development under uncertainty using multilevel quasi-Monte Carlo methods. *Computational Geosciences*, 22:1009–1020, 2018.
[13] Xu Z, He Z, and Wang X. Efficient risk estimation via nested multilevel quasi-Monte Carlo simulation. *Journal of Computational and Applied Mathematics*, 443:115745, 2024.
[14] Fang W, Wang Z, Giles MB, Jackson CH, Welton NJ, Andrieu C, and Thom H. Multilevel and quasi Monte Carlo methods for the calculation of the expected value of partial perfect information. *Medical Decision Making*, 42(2):168–181, 2022.
[15] Caflisch RE. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 1–49, 1998.
[16] Niederreiter H. Random number generation and quasi-Monte Carlo methods. *SIAM*, Philadelphia, 1992.
[17] Owen AB. Quasi-Monte Carlo sampling. *In: Monte Carlo Ray Tracing: SIGGRAPH*, 69–88, 2003.
[18] Tuffin B. Randomization of quasi-Monte Carlo methods for error estimation: survey and normal approximation. *In: Monte Carlo Methods and Applications*, De Gruyter, 10(3-4):617–628, 2004.
[19] Durrett R. Probability: Theory and examples. *Cambridge University Press*, 2019.
[20] Dick J and Pillichshammer F. Digital nets and sequences: discrepancy theory and quasi-Monte Carlo integration. *Cambridge University Press*, 2010
[21] Dick J, Kuo FY, and Sloan IH. High-dimensional integration: the quasi-Monte Carlo way. *Acta Numerica*, 133–288, 2013.
[22] Sobol' IM. Distribution of points in a cube and approximate evaluation of integrals. *Ž. Vyčisl. Mat. i Mat. Fiz. (in Russian)*, 7:784–802, 1967.
[23] L'Ecuyer P, Munger D, and Tuffin B. On the distribution of integration error by randomly-shifted lattice rules. *Electronic Journal of Statistics*, 4:950–993, 2010.
[24] Hlawka E. Funktionen von beschränkter Variation in der Theorie der Gleichverteilung. *Annali di Matematica Pura ed Applicata*, 54:325–333, 1961.
[25] Haji-Ali A-L, Nobile F, Tamellini, L, and Tempone R. Multi-index stochastic collocation convergence rates for random PDEs with parametric regularity. *Foundations of Computational Mathematics*, 16:1555–1605, 2016.
[26] Owen AB. Halton sequences avoid the origin. *SIAM Review*, 48:487–503, 2006.
[27] He Z, Zheng Z, and Wang X. On the error rate of importance sampling with randomized quasi-Monte Carlo. *SIAM Journal on Numerical Analysis*, 61(2):10.1137/22M1510121, 2023.
[28] Ouyang D, Wang X, and He Z. Quasi-Monte Carlo for unbounded integrands with importance sampling. *arXiv preprint arXiv:2310.00650*, 2023.
[29] Owen AB. Randomly permuted (t,m,s)-nets and (t,s)-sequences. *in: Niederreiter H and Shiue PJS (Eds.), Monte Carlo and Quasi-Monte Carlo methods in Scientific Computing. Lecture Notes in Statistics 106*, New York, Springer, 1995.
[30] Goda T, Hironaka T, and Iwamoto T. Multilevel Monte Carlo estimation of expected information gains. *Stochastic Analysis and Applications*, 38(4):581–600, 2020.
[31] Goda T, Hironaka T, Kitade W, and Foster A. Unbiased MLMC stochastic gradient-based optimization of Bayesian experimental designs. *SIAM Journal on Scientific Computing*, 44(1):A286–A311, 2022.
[32] Kuo FY, Schwab C, and Sloan IH. Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients. *Foundations of Computational Mathematics*, 15:411–449, 2015.
[33] Owen AB. Scrambled net variance for integrals of smooth functions. *The Annals of Statistics*, 25:1541–1562, 1997.
[34] Pan Z, and Owen AB. The nonzero gain coefficients of Sobol's sequences are always powers of two. *Journal of Complexity*, 75:101700, 2023.
[35] Kaarnioja V, and Schillings C. Quasi-Monte Carlo for Bayesian design of experiment problems governed by parametric PDEs. *arXiv preprint arXiv:2405.03529 [v2]*, 2024.
[36] Patel JK, and Read CB. Handbook of the normal distribution. *2nd ed.*, New York, Marcel Dekker, 1996.
[37] Du X, and Wang H. Efficient estimation of expected information gain in Bayesian experimental design with multi-index Monte Carlo. *Statistics and Computing*, 34:200, 2024.
[38] Giles MB, and Szpruch L. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Annals of Applied Probability*, 24(4):1585–1620, 2014.
[39] Haji-Ali A-L, Nobile F, von Schwerin E, and Tempone R. Optimization of mesh hierarchies in multilevel Monte Carlo samplers. *Stochastics and Partial Differential Equations Analysis and Computations*, 4(1):76–112, 2014.
[40] Rainforth T, Cornish R, Yang H, Warrington A, and Wood F. On nesting Monte Carlo estimators. *Proceedings of the* 35th *International Conference on Machine Learning, PLMR 80* Stockholm, 2008.
[41] Syed Y, and Wang G. Optimal randomized multilevel Monte Carlo for repeatedly nested expectations. *Proceedings of the 40th International Conference on Machine Learning*, 2023.

[42] Spence J. Hierarchical and adaptive methods for accurate and efficient risk estimation. *Ph.D. thesis, Department of Mathematics, Heriot-Watt University*, 2023.

[43] Giles MB, Haji-Ali A-L, and Spence J. Efficient risk estimation for the credit valuation adjustment. *arXiv preprint arXiv:2301.05886 [v2]*, 2024.