Hamburger Beiträge
zur Angewandten Mathematik

Multilevel Estimation of Rare Events

Elisabeth Ullmann and Iason Papaioannou

Nr. 2014-24
October 2014

Multilevel Estimation of Rare Events*

Elisabeth Ullmann† and Iason Papaioannou‡

October 24, 2014

Abstract

This work is motivated by the need to estimate the probability of rare events in engineering systems with random inputs. We introduce a multilevel estimator which is based on and generalises the idea of subset simulation. The novel estimator employs a hierarchy of approximations to the system response computed with different resolutions. This leads to reduced computational costs compared to subset simulation. We study the statistical properties and implementation details of the proposed estimator. Markov chain Monte Carlo runs are required within the estimator and we demonstrate that the nestedness of the associated multilevel failure domains enables a perfect MCMC simulation without burn-in. We show that nestedness follows from certain simple one-dimensional failure domains. In high dimensions we propose a modification of the multilevel estimator which uses level-dependent stochastic input dimensions. We report on numerical experiments in 1D and 2D physical space; in particular, we estimate rare events arising from a Darcy flow problem with random permeability.

Keywords: reliability, subset simulation, Markov chain Monte Carlo, random porous media, fluid flow, breakthrough time

Mathematics Subject Classification: 65N30, 65C05, 65C40, 60H35, 35R60

1 Introduction

The estimation of failure probabilities is a fundamental problem in reliability analysis and risk management of engineering systems with uncertain inputs. Specifically, we focus on systems described by partial differential equations (PDEs) with random coefficients. We are interested in the estimation of small failure probabilities associated with *rare* events.

Consider, for example, the simulation of highly uncertain subsurface flows for the safety assessment of proposed long-term radioactive waste repositories. In this problem it is important to study the time it takes radionuclides, in case of an accidental damage of the waste repository, to travel from the repository to the boundary of a well-defined safety zone. Failure occurs if the travel time of the radionuclides falls below a certain threshold. We are interested in short travel times after which the radionuclides could still be highly

*The major part of this work has been carried out while the first author worked in the Department of Mathematical Sciences, University of Bath, UK, supported by the EPSRC Grant EP/H051503/1.

†Fachbereich Mathematik, Universität Hamburg, Bundesstr. 55, 20146 Hamburg, Germany (elisabeth.ullmann@uni-hamburg.de).

‡Engineering Risk Analysis Group, TU München, Theresienstr. 90, 80333 München, Germany (iason.papaioannou@tum.de).

active and harmful would they get back to the human environment. The associated failure probability will be small; the event of interest occurs infrequently, but could have a high impact.

The permeability of the random porous medium is often modeled as a correlated, log-normal random field; this is consistent with field data (see e.g. [23, 31]). If the correlation length is small then the input sample space is very high-dimensional. To date only Monte Carlo (MC) based approaches achieve convergence independently of the dimension of the input space. Unfortunately, the number of required samples for crude MC is proportional to the inverse of the failure probability. Hence a large number of samples is needed to estimate small probabilities with Monte Carlo. Moreover, one sample of the travel time requires the solution of discretised groundwater flow and transport equations in 2D or 3D physical space. This yields a very high total computational cost of Monte Carlo which might exceed a given computational budget.

Many alternatives to crude Monte Carlo have been developed, and we will not give a complete overview here. We refer to [40, 51, 55] for a discussion of existing methods. A straightforward idea is to construct an approximation or surrogate for the system output or limit state function; this gives an approximation to the probability of failure. For example, the well known FORM/SORM methods are based on first/second order Taylor series expansions of the limit-state surface at the design point [36]. In high dimensions this approach can be expensive since it requires the solution of an optimisation problem. Moreover, the approximation error can be considerably large [55, 64].

Hybrid approaches, combining Monte Carlo with surrogates, can reduce the computational costs without compromising the accuracy. For example, surrogates based on generalised polynomial chaos are studied in [41, 42, 43]. Specifically, [41] proposes a surrogate-based importance sampling algorithm which allows the estimation of very small failure probabilities. However, polynomial chaos approaches suffer from the curse of dimensionality and are not really practical for very high-dimensional input sample spaces. Hybrid and goal-oriented adaptive reduced basis methods are investigated in [17]; the benchmark studies presented in this work are associated with relatively large failure probabilities. In [21] a surrogate model based on kriging is combined with importance sampling; an unbiased estimate of the failure probability is derived that solves efficiently problems with up to 100 random variables.

Another option is to enhance the efficiency of Monte Carlo by variance reduction. A standard technique for this is importance sampling (IS), however, its success depends highly on the choice of an appropriate IS density. Typical choices are unimodal sampling densities based on results from a preliminary FORM analysis [60] or from an initial sampling step [12]. Alternatively, multimodal densities have also been applied [6, 39]. However, the choice of a suitable IS density for application to high dimensional problems is not straightforward [7, 35]. We mention a recent suggestion in [44] in the context of material failure where the underlying PDE model is identical to the Darcy flow problem with a lognormal coefficient.

Recently, a number of advanced variance reduction schemes for application in high dimensions have been developed. These include subset simulation [1], line sampling [38] and asymptotic sampling [13]. Among these, subset simulation has been shown in numerous applications to be a robust technique for accurate and efficient estimation of rare events in high dimensions. It is now widely used for reliability analyses and estimation of failure probabilities, e.g. in earthquake engineering [2, 3], geotechnical engineering [3, 53, 65], mechanical engineering and fatigue [9], spacecraft engineering [54], and nuclear engineering

[66]. An enhanced version of the method was proposed in [67]. Moreover, subset simulation has been combined with surrogates based on machine learning theory in [11, 50]. In the statistics and probability theory literature the idea of subset simulation is known under the names splitting and sequential Monte Carlo. The earliest reference is [34] where a splitting method is used to estimate dynamic rare events associated with neutron transmissions. More recently, a generalised splitting method for dynamic as well as steady-state simulations has been proposed and analysed in [10]. Sequential Monte Carlo estimators for rare events are proposed and investigated in [14, 33, 48].

Subset simulation relies on a decomposition of the sample space into a sequence of nested, partial failure domains. The expensive Monte Carlo estimator is then replaced by a product of conditional estimators with respect to the partial failure domains. Variance reduction, and thus a reduction of the total number of samples, is achieved by conditioning. However, the physical discretisation of the desired system response, e.g. by finite elements, is fixed in each failure domain and the cost of the conditional sampling is still large. The novel contribution of this work is a multilevel approach to subset simulation. We define more flexible failure domains by using system responses computed on a hierarchy of finite element discretisations; this reduces the cost per sample.

The idea is inspired by multilevel Monte Carlo (MLMC) which has been introduced by Heinrich [30] for high-dimensional parameter-dependent integrals, and has been extended by Giles [26] for stochastic differential equations. More recently it has been used to estimate the statistics of output quantities arising from PDEs with random inputs, see e.g. the pioneering works [8, 18]. In further developments MLMC is used to approximate the distribution function and density of univariate random variables [25] and to estimate failure probabilities [22]. These works, however, are not concerned with rare events. Moreover, the variance reduction is achieved by the estimation of certain corrections, not by conditioning.

The success of MLMC is based on a simple principle for tackling complex, computing-intensive systems with random inputs. First, we decompose the original problem into a set of subproblems each of which has (i) a smaller variance, and (ii) a smaller cost per sample. Then, we combine the solutions of the individual subproblems to get a solution of the original problem. Crucially, the total cost to solve all subproblems is in general much smaller compared to the original problem because the reduced variance leads to a smaller number of samples, and because of the reduced cost per sample. MLMC as described e.g. in [8, 18] is not the only possibility to achieve this. We mention [37] where the *law of total expectation* is used to combine inexpensive models to estimate the statistics of expensive outputs. Subset simulation satisfies property (i); it is a variance reduction technique. The novel multilevel estimator generalises subset simulation; it satisfies properties (i) and (ii). It can be used in a variety of situations where it is possible to compute the response of interest with different resolutions or models. In this study we focus on system responses arising from PDEs with random coefficients.

The paper is organised as follows. In §2 we describe the problem of rare event estimation and review subset simulation. We introduce the multilevel estimator in §3 and discuss implementation, computational cost, and statistical properties. In §4 we study the nestedness of the multilevel failure domains, and in §5 we modify the multilevel estimator for level-dependent stochastic input dimensions. In §6 we test the implementation of the estimator for a 2D flow cell problem with random permeability. We summarise the discussion in §7.

2 Background

2.1 Problem setting

We consider some engineering system or model subject to uncertain inputs. Specifically, let $\boldsymbol{\theta} = (\theta_m)_{m=1}^M \subseteq \mathbb{R}^M$ denote the random input vector of the system with associated probability density function (PDF) f. Let $Q = Q(\boldsymbol{\theta}) \in \mathbb{R}$ denote the (scalar) system response of interest. For example, the model response could be the displacement of the tip of a cantilever beam with random Young's modulus. Another example is the breakthrough time of particles in subsurface flows in random porous media. Note that Q is often given implicitly as functional of the solution of a partial differential equation (PDE) subject to certain initial and/or boundary conditions. The uncertainty in the system inputs propagates to the response and Q is uncertain as well. We are thus interested in statistical properties of Q.

To carry out a reliability analysis we consider *failure events* associated with Q that occur for some $\boldsymbol{\theta} \in F$. The set $F \subseteq \mathbb{R}^M$ in the input parameter space is termed *failure domain*. Formally, we assume that we may parameterise F using a so called *limit state function* $G\colon \mathbb{R}^M \to \mathbb{R}$. We define

$$F := \{\boldsymbol{\theta} \in \mathbb{R}^M \colon G(\boldsymbol{\theta}) \leq 0\} \ . \tag{2.1}$$

For example, $G(\boldsymbol{\theta}) = c - Q(\boldsymbol{\theta})$ describes the failure event $Q \geq c$, i.e. the system response exceeds the threshold c. We now wish to calculate the *probability of failure*

$$P_F := \mathrm{Prob}(\boldsymbol{\theta} \in F) = \int_{\boldsymbol{\theta} \in \mathbb{R}^M} I_F(\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{2.2}$$

where I_F denotes the indicator function: $I_F(\boldsymbol{\theta}) = 1$ if $\boldsymbol{\theta} \in F$ and $I_F(\boldsymbol{\theta}) = 0$ otherwise. For brevity we will write $\mathrm{Prob}(F)$ in place of $\mathrm{Prob}(\boldsymbol{\theta} \in F)$ in the remainder of this paper.

In most practical applications the geometry of F is complicated and/or the dimension of the input parameter space is very high. Thus closed form expressions for P_F are not available in general and an estimator for P_F must be used. Moreover, the evaluation of the indicator function I_F, or, equivalently, the limit state function G, requires the system response $Q(\boldsymbol{\theta})$ which is usually unavailable in closed form as well. Instead, we assume that we can approximate Q by Q_h. The parameter $h > 0$ refers to a certain approximation level, e.g. the mesh size of a finite element discretisation or the time step size used to obtain Q_h. This leads us to approximate the limit state function G by G_h, the associated failure region F in (2.1) by

$$F_h := \{\boldsymbol{\theta} \in \mathbb{R}^M \colon G_h(\boldsymbol{\theta}) \leq 0\}, \tag{2.3}$$

and the probability of failure P_F by P_{F_h}, respectively. In summary, our goal is to estimate

$$P_{F,h} := \mathrm{Prob}(F_h) = \int_{\boldsymbol{\theta} \in \mathbb{R}^M} I_{F_h}(\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta} \ . \tag{2.4}$$

Remark 2.1 A failure event described by a limit-state function of the form $G(\boldsymbol{\theta}) = c - Q(\boldsymbol{\theta})$ is termed component failure event, because it describes a particular failure mode of the system. It is also possible to consider more complex failure events which are intersections and/or unions of component failure events. Such failure events can be expressed without loss of generality with a single limit-state function G within the framework of (2.1). We refer to [1, §5.2] for the construction of appropriate limit state functions.

2.2 Monte Carlo estimator

The standard Monte Carlo (MC) estimator for $P_{F,h}$ is

$$\widehat{P}_{F,h}^{MC} := \frac{1}{N}\sum_{i=1}^{N} I_{F_h}(\boldsymbol{\theta}^{(i)}), \tag{2.5}$$

where $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^{N}$ are independent, identically distributed (i.i.d.) samples of the random input vector $\boldsymbol{\theta}$. This approach is flexible and easy to implement even for complex engineering systems. It requires N independent system analyses (to obtain Q_h) that can be carried out in parallel. Subsequently, $P_{F,h}$ is approximated by the sample mean of the indicator function I_{F_h} evaluated at the input samples.

The MC estimator is unbiased, that is, $\mathbb{E}[\widehat{P}_{F,h}^{MC}] = P_{F,h}$, and its variance is given by

$$\mathbb{V}[\widehat{P}_{F,h}^{MC}] = N^{-1}\mathbb{V}[I_{F_h}] = N^{-1}P_{F,h}(1-P_{F,h}) \ .$$

A typical relative error measure for rare event estimators is δ, the so called *coefficient of variation* (c.o.v.). δ is the ratio of standard deviation and expected value of an estimator. For the MC estimator in (2.5) we obtain

$$\delta^2(\widehat{P}_{F,h}^{MC}) = \frac{\mathbb{V}[\widehat{P}_{F,h}^{MC}]}{\mathbb{E}[\widehat{P}_{F,h}^{MC}]^2} = \frac{1-P_{F,h}}{NP_{F,h}} \ .$$

To achieve $\delta(\widehat{P}_{F,h}^{MC}) < \varepsilon$ for a given tolerance $\varepsilon > 0$ requires $N > \varepsilon^{-2}P_{F,h}^{-1}(1-P_{F,h})$ system analyses. As is typical for MC estimators, the number of required samples N depends on ε^{-2}. Much more problematic is the fact that it also depends on the inverse $P_{F,h}^{-1}$ which will be very large if the failure probability $P_{F,h} \ll 1$ is small. In our setting where a single system analysis requires the solution of a discretised PDE (possibly in three space dimensions on a fine FE mesh) the standard Monte Carlo estimator for $P_{F,h}$ is thus too expensive.

2.3 Subset simulation

To overcome the limits of Monte Carlo we employ subset simulation. We discuss the basic idea and implementation; for more details we refer to the recent monograph [4]. We will see that subset simulation is the single-level version of the multilevel estimator in §3.

2.3.1 Basic idea

Consider the task of estimating the failure probability P_F in (2.2) associated with a failure domain F. If P_F is small then the MC estimator requires an excessive amount of samples since the c.o.v. of the estimator is very large (cf. §2.2). To overcome this obstacle we decompose the original problem into a sequence of subproblems each of which has a smaller c.o.v. The key idea is to express the failure domain F as intersection of $L > 0$ intermediate or partial failure domains

$$F = \bigcap_{\ell=1}^{L} F_\ell \ .$$

Importantly, we assume that $F_1 \supset F_2 \supset \cdots \supset F_{L-1} \supset F_L = F$, i.e. the partial failure domains are *nested*. It is easy to see that P_F can then be written as product of conditional probabilities

$$P_F = \text{Prob}(F_L) = \text{Prob}(F_1) \prod_{\ell=2}^{L} \text{Prob}(F_\ell | F_{\ell-1}) \ . \tag{2.6}$$

Crucially, we select the failure levels for the limit state function G in (2.1) adaptively and define the intermediate failure domains such that $\text{Prob}(F_1)$ and $\text{Prob}(F_\ell|F_{\ell-1})$, $\ell = 2, \ldots, L-1$ are much larger than $\text{Prob}(F)$. This reduces the number of samples substantially. Consider, e.g., the intermediate failure domain F_1. Intuitively, since F_1 is a larger region in the parameter space compared to $F_L = F$ it is easier to find failure points in F_1. Hence we need only a small number of system analyses to achieve $\delta(\widehat{P}_{F_1}^{MC}) < \varepsilon$ for a given accuracy $\varepsilon > 0$. A similar argument applies to the conditional failure probabilities in (2.6).

2.3.2 Implementation

Consider the factorisation of P_F in (2.6). The first, unconditional failure probability $\text{Prob}(F_1)$ is estimated by Monte Carlo. The estimator $\widehat{P}_{F_1}^{MC}$ is defined analogously to (2.5). The subsequent failure probabilities $\text{Prob}(F_\ell|F_{\ell-1})$ require input samples conditioned on $F_{\ell-1}$ with associated conditional PDFs

$$f(\cdot|F_{\ell-1}) = \frac{f(\cdot) I_{F_{\ell-1}}(\cdot)}{\text{Prob}(F_{\ell-1})}, \quad \ell = 2, \ldots, L \ . \tag{2.7}$$

A Monte Carlo estimator for $\text{Prob}(F_\ell|F_{\ell-1})$ based on rejection sampling could be used. However, this approach is in general not efficient because the acceptance probability of rejection sampling is proportional to $\text{Prob}(F_{\ell-1})$. Given that we have found a failure point $\boldsymbol{\theta} \in F_{\ell-1}$ it is reasonable to expect that more failure points are located nearby. Therefore we use an estimator based on Markov Chain Monte Carlo (MCMC) to obtain conditional samples. The estimator for $\text{Prob}(F_\ell|F_{\ell-1})$ is then defined as follows,

$$\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC} := \frac{1}{N} \sum_{i=1}^{N} I_{F_\ell}(\boldsymbol{\theta}^{(i)}), \tag{2.8}$$

where $\boldsymbol{\theta}^{(i)} \sim f(\cdot|F_{\ell-1})$, $i = 1, \ldots, N$ are generated by a Markov chain whose stationary distribution is equal to the desired conditional distribution.

The MCMC samples are in general not statistically independent. The crucial component of MCMC is the transition from the current state $\boldsymbol{\theta}^{(n)}$ to the next state $\boldsymbol{\theta}^{(n+1)}$ of the chain. A standard implementation of this transition is given by the Metropolis-Hastings (M-H) algorithm [29, 47, 57]. There the transition $\boldsymbol{\theta}^{(n)} \to \boldsymbol{\theta}^{(n+1)}$ is performed in two steps. Using a proposal distribution based on $\boldsymbol{\theta}^{(n)}$ a candidate $\boldsymbol{\theta}'$ for the next state is proposed. Then the candidate is either accepted or rejected with a certain acceptance probability $\alpha = \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')$ where $\alpha \in [0, 1]$. Importantly, α is chosen such that the transitional PDF satisfies a reversibility condition. This ensures that the stationary distribution of the chain is equal to the desired target distribution.

For the remainder of this paper we make the following assumption on the random input data.

A1. The components of $\boldsymbol{\theta} \in \mathbb{R}^M$ are statistically independent, standard Gaussian random variables.

This implies $f(\boldsymbol{\theta}) \equiv \varphi(\boldsymbol{\theta})$, where φ denotes the M-variate, standard Gaussian probability density function.

Remark 2.2 Assumption A1 is not very restrictive since the stochastic inputs could be functions of some independent Gaussian random variables or Gaussian processes. In other situations it might be possible to represent a vector $\boldsymbol{\xi} \in \mathbb{R}^M$ of uncorrelated *non-Gaussian* random inputs as transformations $\xi_n = \xi_n(\theta_1, \ldots, \theta_M)$, $n = 1, \ldots, M$, of independent Gaussian random variables $\{\theta_m\}_{m=1}^M$. A standard technique for this is the Rosenblatt transformation [58].

For use in subset simulation we wish to generate conditional samples $\boldsymbol{\theta}^{(i)} \sim \varphi(\cdot|F_{\ell-1})$. A Metropolis-Hastings-type transition for sampling from $\varphi(\cdot|F_{\ell-1})$ can be done in two steps. First a transition with stationary distribution $\varphi(\cdot)$ is performed with acceptance probability $\tilde{\alpha}$ that depends on the choice of the proposal distribution. Then, the sample is accepted if it lies in $F_{\ell-1}$, otherwise it is rejected. Hence the combined acceptance probability of the two steps is $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \tilde{\alpha}(\boldsymbol{\theta}, \boldsymbol{\theta}') \cdot I_{F_{\ell-1}}(\boldsymbol{\theta}')$. In [1], the first step of the MCMC transition is performed by a component-wise Metropolis-Hastings that generates each candidate state component-by-component. The acceptance probability of each component $\tilde{\alpha}_m$ depends on the proposal distribution. However, because of the component-wise generation, the probability of having repeated candidates simultaneously in all components decreases geometrically with increasing number of random variables M. Hence, this approach is suitable for application to high dimensional problems. Here, we perform the first step using a specific proposal density which greatly simplifies the computation. The implementation is summarised in Algorithm 1.

Algorithm 1 M-H type MCMC for sampling from the target distribution $\varphi(\cdot|F_{\ell-1})$

Choose a correlation parameter $\gamma \in [0, 1]$.
Given a seed $\boldsymbol{\theta}^{(1)}$, repeat for $n = 1, 2, \ldots, :$

(1) Generate a candidate $\boldsymbol{\theta}'$

For $i = 1, \ldots, M$: $\quad \theta_i' = \gamma\theta_i^{(n)} + \sqrt{1-\gamma^2}Z, \;\; Z \sim N(0, 1)$

(2) Accept or reject $\boldsymbol{\theta}'$

$$\boldsymbol{\theta}^{(n+1)} = \begin{cases} \boldsymbol{\theta}', & \boldsymbol{\theta}' \in F_{\ell-1} \ , \\ \boldsymbol{\theta}^{(n)}, & \boldsymbol{\theta}' \notin F_{\ell-1} \ . \end{cases}$$

It can be shown that under Assumption A1 the acceptance probability for $\boldsymbol{\theta}'$ in part (1) of Algorithm 1 is $\tilde{\alpha} = 1$. To put it another way, the (Gaussian) prior distribution is preserved under the proposal. Therefore we only need to check if $\boldsymbol{\theta}' \in F_{\ell-1}$ in part (2) of Algorithm 1. An additional acceptance/rejection step in part (1) based on $\tilde{\alpha}$ is not required. Algorithm 1 is proposed in [52] for use with subset simulation. The correlation parameter γ controls the statistical dependence of consecutive states of the Markov chain. We mention that it is possible to choose γ adaptively to avoid a large rejection rate of

candidates or a large correlation between the states of the Markov chain (see [52] for details). In our computations we fix $\gamma = 0.8$.

Now we define the subset simulation estimator for the failure probability P_F. Recall that $\widehat{P}^{MCMC}_{F_\ell|F_{\ell-1}}$ denotes the MCMC estimator for $\mathrm{Prob}(F_\ell|F_{\ell-1})$, $\ell = 2, \dots, L$, as defined in (2.8). Then

$$\widehat{P}^{SubS}_F := \widehat{P}^{MC}_{F_1} \prod_{\ell=2}^{L} \widehat{P}^{MCMC}_{F_\ell|F_{\ell-1}} \tag{2.9}$$

is the estimator for P_F generated by subset simulation. Note that the intermediate failure domains can be defined analogously to F in (2.1). Precisely, let

$$F_\ell := \{\boldsymbol{\theta} \in \mathbb{R}^M : G(\boldsymbol{\theta}) \le c_\ell\}, \qquad \ell = 1, \dots, L, \tag{2.10}$$

where, importantly, $c_1 > c_2 > \cdots > c_{L-1} > c_L = 0$ to ensure the nestedness of $F_1, F_2, \dots, F_L$. Moreover, by choosing the sequence $c_1, \dots, c_{L-1}$ "on the fly" we can guarantee that $\widehat{P}^{MC}_{F_1} = p_0$ and $\widehat{P}^{MCMC}_{F_\ell|F_{\ell-1}} = p_0$, $\ell = 2, \dots, L-1$ for some fixed $0 < p_0 < 1$. The expression for the estimate of P_F then simplifies to

$$\widehat{P}^{SubS}_F = p_0^{L-1} \widehat{P}^{MCMC}_{F_L|F_{L-1}} \,. \tag{2.11}$$

A typical value used in engineering applications is $p_0 = 0.1$. In this case, the estimation of a failure probability $P_F \approx 10^{-6}$ by subset simulation would require $L \approx 6$ intermediate failure domains/levels. We summarise the (standard) subset simulation in Algorithm 2.

Algorithm 2 Subset simulation

Input: p_0, N

(1) Generate N i.i.d. samples $\boldsymbol{\theta}^{(i)} \sim \varphi(\cdot)$ for use in Monte Carlo.
Determine a failure level $c_1 > 0$ s.t. $\widehat{P}^{MC}_{F_1} = p_0$.

(2) For $\ell = 2, \dots, L$:
Use the N_0 failure points in $F_{\ell-1}$ as seeds and generate $N - N_0$ samples $\boldsymbol{\theta}^{(i)} \sim \varphi(\cdot|F_{\ell-1})$ with Algorithm 1.
If $\ell < L$, determine a failure level $c_\ell > 0$ s.t. $\widehat{P}^{MCMC}_{F_\ell|F_{\ell-1}} = p_0$.

(3) Evaluate $\widehat{P}^{MCMC}_{F_L|F_{L-1}}$ as in (2.8) and return $\widehat{P}^{SubS}_F$ in (2.11).

Remark 2.3 Typically, since not all seeds follow the desired target distribution, a burn-in of the Markov chains is required. By this we mean that the first few samples generated by the chain are not used for estimation. However, thanks to the nestedness of the (intermediate) failure domains the seeds in part (2) of Algorithm 2 are already distributed according to the target distribution. Hence all subsequent samples generated by the Markov chain follow the target distribution, and it is justified to write $\boldsymbol{\theta}^{(i)} \sim \varphi(\cdot|F_{\ell-1})$ in part (2) of Algorithm 2. A burn-in of the chains is *not* needed. This situation is termed *perfect simulation* in the literature, see e.g. [57, §8.6]. In the context of subset simulation this is discussed in [5, 67] and [4, §5].

Remark 2.4 Replacing F by F_h in (2.3) it is straightforward to use subset simulation for the estimation of $P_{F_h} \approx P_F$ in (2.4).

3 Multilevel estimator

Consider the estimation of the failure probability $P_{F,h}$ in (2.4) by subset simulation as discussed in §2.3. Recall that we approximate the system response Q by Q_h. The limit state function G and the partial failure domains F_ℓ in (2.10) then depend on h as well, and we would work with the *single-level failure domains*

$$F_\ell^{SL} := \{\boldsymbol{\theta} \in \mathbb{R}^M \colon G_h(\boldsymbol{\theta}) \le c_\ell\}, \qquad \ell = 1, \dots, L. \tag{3.1}$$

Again, we assume that $c_1 > c_2 > \cdots > c_{L-1} > c_L = 0$ holds for the sequence of failure levels. This implies nestedness, that is, $F_\ell^{SL} \subset F_{\ell-1}^{SL}$, $\ell = 2, \dots, L$. However, each evaluation of G_h requires Q_h which can be very expensive in many practical applications. To demonstrate this we make the following assumption.

A2. The cost to obtain one sample of the system response Q_h is $\mathcal{C}(Q_h) = \mathcal{O}(h^{-m})$, where $m > 0$ is a constant independent of h.

For example, Q_h might require the solution of a discretised elliptic PDE with smooth coefficients in a domain $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, using standard piecewise linear finite elements on a triangulation with mesh size h. If the sampling of the coefficients and the solution of the discretised PDE can be done in optimal, linear cost, then $\mathcal{C}(Q_h) = \mathcal{O}(h^{-d})$ and thus $m = d$ in Assumption A2. Hence we see that if a small mesh size h is required for accuracy, then $\mathcal{C}(Q_h)$ is very large in 2D and 3D models.

This motivates us to introduce and study a generalised version of the subset simulation estimator (2.9). To reduce the computational cost we define the failure domains on a hierarchy of discretisations with decreasing associated mesh sizes (or, equivalently, increasing resolution in the physical space) and increasing computational cost.

Consider the *multilevel failure domains*

$$F_\ell^{ML} := \{\boldsymbol{\theta} \in \mathbb{R}^M \colon G_{h_\ell}(\boldsymbol{\theta}) \le c_\ell\}, \tag{3.2}$$

where $h_\ell = 2^{-\ell} h_0$, $\ell = 1, \dots, L$ and $h_0 > 0$. Observe that under Assumption A2 the cost of obtaining the system response $Q_\ell := Q_{h_\ell}$ is $\mathcal{C}(Q_\ell) = \mathcal{O}(2^{\ell m})$. Hence using F_ℓ^{ML} in place of F_ℓ^{SL} in the factorisation (2.6) brings obvious computational benefits since the cost of obtaining Q_{h_ℓ} is small on (coarse) levels where h_ℓ is large. Unfortunately, if we change the failure level *and* the discretisation parameter at the same time in the transition from F_ℓ^{ML} to $F_{\ell+1}^{ML}$ then $F_\ell^{ML} \not\subset F_{\ell-1}^{ML}$, i.e. the multilevel failure domains are not nested in general. Hence we cannot use the factorisation in (2.6). We address this problem by deriving a generalised version of (2.6) which is also valid for non-nested failure domains.

3.1 Basic idea

Let us consider three partial failure domains F_j, F_k, and F_ℓ in the input parameter space. Application of Bayes' formula gives

$$\mathrm{Prob}(F_k|F_\ell) \times \mathrm{Prob}(F_\ell) = \mathrm{Prob}(F_k \cap F_\ell) = \mathrm{Prob}(F_\ell|F_k) \times \mathrm{Prob}(F_k)$$

and thus

$$\mathrm{Prob}(F_\ell) = \frac{\mathrm{Prob}(F_\ell|F_k)}{\mathrm{Prob}(F_k|F_\ell)} \times \mathrm{Prob}(F_k) \ . \tag{3.3}$$

Using Bayes' formula again for the pair F_j and F_k we obtain

$$\text{Prob}(F_k) = \frac{\text{Prob}(F_k|F_j)}{\text{Prob}(F_j|F_k)} \times \text{Prob}(F_j) \ . \tag{3.4}$$

Inserting (3.4) into (3.3) we see that

$$\text{Prob}(F_\ell) = \frac{\text{Prob}(F_\ell|F_k)}{\text{Prob}(F_k|F_\ell)} \times \frac{\text{Prob}(F_k|F_j)}{\text{Prob}(F_j|F_k)} \times \text{Prob}(F_j) \ .$$

It is clear that this technique can be applied to any countable, finite collection of partial failure domains $\{F_\ell\}_{\ell=1}^L$. Assuming that $F_L = F$ is the target failure domain this gives the factorisation

$$P_F = \text{Prob}(F_L) = \text{Prob}(F_1) \prod_{\ell=2}^{L} \frac{\text{Prob}(F_\ell|F_{\ell-1})}{\text{Prob}(F_{\ell-1}|F_\ell)} \ . \tag{3.5}$$

Note that (3.5) does not require the nestedness of the partial failure domains. In particular, (3.5) holds for the multilevel failure domains $F_\ell = F_\ell^{ML}$ defined in (3.2). If, however, the failure domains are nested, then it is easy to see that all denominators in (3.5) are equal to one. Hence, for the collection of nested, single-level failure domains $F_\ell = F_\ell^{SL}$ defined in (3.2) we see that (3.5) reduces to the factorisation (2.6). We have therefore generalised the basic idea of subset simulation to a wider class of failure domains which are not necessarily nested.

3.2 Implementation

Based on (3.5) we define the multilevel estimator (MLE) for P_F as follows,

$$\widehat{P}_F^{ML} := \widehat{P}_{F_1}^{MC} \prod_{\ell=2}^{L} \frac{\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC}}{\widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC}}, \tag{3.6}$$

where $\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC}$ is the MCMC estimator for $\text{Prob}(F_\ell|F_{\ell-1})$ (see (2.8)), and $\widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC}$ is the MCMC estimator for $\text{Prob}(F_{\ell-1}|F_\ell)$, respectively.

Note that the estimators in the numerator of (3.6) can be evaluated exactly as in subset simulation. To estimate the denominators we introduce an additional step in the transition from level ℓ to $\ell+1$. Having estimated $\text{Prob}(F_\ell|F_{\ell-1})$ we can use the samples in F_ℓ as seeds for Markov chains to generate conditional samples $\boldsymbol{\theta} \sim \varphi(\cdot|F_\ell)$ for use in the estimation of $\text{Prob}(F_{\ell-1}|F_\ell)$. The implementation of the multilevel estimator is summarised in Algorithm 3. It is a straightforward extension of subset simulation (see Algorithm 2). Recall that we can choose $0 < p_0 < 1$ and construct the intermediate failure levels $c_1, c_2, \ldots, c_{L-1}$ such that $\widehat{P}_{F_1}^{MC} = p_0$ and $\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC} = p_0$, $\ell = 2, \ldots, L-1$. Hence we can simplify (3.6) to obtain

$$\widehat{P}_F^{ML} = \frac{p_0^{L-1} \widehat{P}_{F_L|F_{L-1}}^{MCMC}}{\prod_{\ell=2}^{L} \widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC}} \ . \tag{3.7}$$

Algorithm 3 Multilevel estimator

Input: p_0, N, L_b

(1) Generate N i.i.d. samples $\boldsymbol{\theta}^{(i)} \sim \varphi(\cdot)$ for use in Monte Carlo.
Determine a failure level $c_1 > 0$ s.t. $\widehat{P}_{F_1}^{MC} = p_0$.

(2) For $\ell = 2, \ldots, L$:
If $\ell < 3$, set $n = 0$, else set $n = L_b$.
Use the N_0 failure points in $F_{\ell-1}$ as seeds and generate $N + (n-1)N_0$ samples $\boldsymbol{\theta}^{(i)}$ with Algorithm 1 with target distribution $\varphi(\cdot|F_{\ell-1})$. Discard the first n samples in each Markov chain.
If $\ell < L$, determine a failure level $c_\ell > 0$ s.t. $\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC} = p_0$.
Then, use all N_0 failure points in F_ℓ as seeds and generate $N + (n-1)N_0$ samples $\boldsymbol{\theta}^{(k)}$ with Algorithm 1 with target distribution $\varphi(\cdot|F_\ell)$. Discard the first n samples in each Markov chain. Evaluate $\widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC}$ as in (2.8).

(3) Evaluate $\widehat{P}_{F_L|F_{L-1}}^{MCMC}$ as in (2.8) and return $\widehat{P}_F^{ML}$ in (3.7).

3.3 Burn-in

The setting of the MLE is more general than subset simulation; this brings new issues. For the MLE we do not assume nestedness of the multilevel failure domains. Thus not all seeds will follow the target distribution, and a burn-in of the Markov chains is needed. Hence we discard the first $L_b \geq 0$ samples of each Markov chain in Algorithm 3. We mention that the burn-in length $L_b = L_{b,\ell}$ might depend on the current level ℓ, and might be smaller on levels with a higher resolution of the system response Q_{h_ℓ} (cf. §5.3).

The burn-in effect can be reduced by application of enhanced MCMC samplers that incorporate regeneration steps. Regeneration techniques [27, 49] are based on restarting the Markov chain at random stopping times. Regenerative MCMC algorithms can also be used to construct better estimates of the variance of quantities of interest. We leave the application of such techniques within the MLE estimator for future studies.

If a burn-in of the Markov chains in Algorithm 3 is required, then the computational cost of the estimator increases (cf. §3.4). Moreover, if the burn-in length is insufficient, then this can increase the bias of the MLE (cf. §3.5). Of course, by using nested multilevel failure domains we can avoid the burn-in. In §4 we investigate conditions under which nestedness applies. If the system response arises from the solution of a PDE with random coefficients, then we show in §5 that we can overcome the burn-in problem by a level-dependent estimator. By this we mean that the number of input random variables $M = M_\ell$ depends on the failure level.

3.4 Computational cost

Let us now predict the savings in terms of computational costs that we can expect when using the multilevel estimator in place of subset simulation. Consider a sequence of approximate system responses $\{Q_\ell\}_{\ell=1}^L$, where $Q_\ell = Q_{h_\ell}$ is associated with the discretisation parameter $h_\ell = 2^{-\ell} h_0$, $h_0 > 0$. Under Assumption A2 the cost to obtain Q_ℓ on level ℓ is $\mathcal{C}(Q_\ell) = c_{\mathrm{A2}} \times h_\ell^{-m} = \widetilde{c} \times 2^{\ell m}$, $\ell = 1, \ldots, L$, where $\widetilde{c}$ is a constant independent of ℓ. The goal is to estimate the failure probability $P_{F,h}$ in (2.4) on level L with $h = h_L$. Algorithm 2

with the single-level failure domains (3.1) requires $N + N(1-p_0)(L-1)$ system responses Q_L. Hence the total cost of subset simulation is

$$\mathcal{C}(\widehat{P}_{F,h}^{SubS}) = \widetilde{c} \times N(1 + (1-p_0)(L-1)) \times 2^{Lm} \ .$$

On the other hand, Algorithm 3 with the multilevel failure domains (3.2) requires $3N - N_0$ system responses Q_1, $2N + (L_b - 1)N_0$ responses Q_2, and $N + (L_b - 1)N_0 + N$ responses on Q_ℓ and $Q_{\ell-1}$, resp., on levels $\ell = 3, \ldots, L$. Hence, assuming that $N_0 = p_0 N$ in Algorithm 3, the total cost of the MLE is

$$\mathcal{C}(\widehat{P}_{F,h}^{ML}) = \widetilde{c}N \times ((3-p_0)2^m + (2 + (L_b-1)p_0)2^{2m} + (2 + (L_b-1)p_0) \sum_{\ell=3}^{L} (2^{m\ell} + 2^{m(\ell-1)})) \ .$$

In Figure 3.1 we plot the predicted cost reduction of the MLE compared to subset simulation versus a different total number of levels. The burn-in length is $L_b = 0$ (left) and $L_b = 10$ (right). We assume $p_0 = 0.1$ fixed in Algorithms 2 and 3. Moreover, we choose $m = d$, which is the minimal cost of a system response arising from typical discretised PDE models in d-dimensional physical space. We see that in the case of no burn-in the MLE will be cheaper than subset simulation in 1D models for $L \geq 6$ levels. In the best case the computational cost will be only about 50% of the cost of subset simulation. For 2D and 3D models we can expect savings for $L \geq 4$ levels, and costs of about 20%–30% of the cost of subset simulation in the best case. This might sound like a modest improvement, but we would like to point out that the MLE improves a method that is already much more efficient than the standard MC estimator (cf. [1]). If burn-in applies, then, of course, the cost of the MLE increases. However, for $L_b = 10$ and 3D models we expect savings for $L \geq 5$ levels.

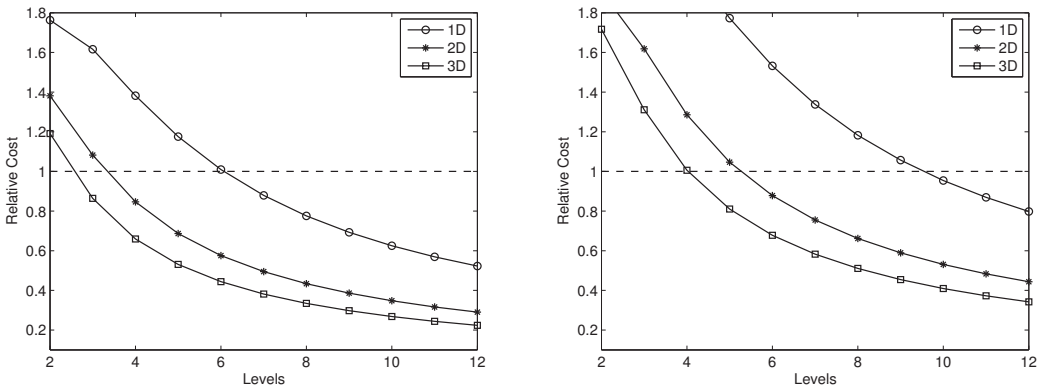


Figure 3.1: Cost reduction of the MLE compared to subset simulation for $p_0 = 0.1$ fixed. The cost to obtain one sample on level ℓ is $\mathcal{C}(Q_\ell) = \mathcal{O}(2^{m\ell})$, where $m = d$ is the dimension of the physical space in the model. The burn-in length is $L_b = 0$ (left) and $L_b = 10$ (right).

Remark 3.1 In Algorithm 3 we could use the exact same conditional samples for the estimation of $\text{Prob}(F_{\ell-1}|F_\ell)$ in the denominator and $\text{Prob}(F_{\ell+1}|F_\ell)$ in the numerator. We do not consider this option here since it introduces additional correlations between the MCMC estimators.

Remark 3.2 The cost of the MLE could be further reduced by choosing level-dependent partial failure probabilities $p_0 = p_{0,\ell}$. This allows us to *prescribe* a certain variance reduction between levels to combat the increase of the cost per sample from level ℓ to $\ell + 1$. Observe that in multilevel Monte Carlo (MLMC) methods it is not possible to prescribe the variance reduction a priori. In contrast, the variance reduction in MLMC depends on the convergence rate of the FE approximation to the system response, see e.g. [8, 15, 22, 63].

3.5 Statistical properties

Now we investigate the statistical properties of the multilevel estimator (3.6) when used together with the multilevel failure domains in (3.2). Consider the bias $\mathbb{E}[\widehat{P}_F^{ML}] - P_F$ of the multilevel estimator. We obtain the decomposition

$$\mathbb{E}[\widehat{P}_F^{ML}] - P_F = (\mathbb{E}[\widehat{P}_F^{ML}] - P_{F,h}) + (P_{F,h} - P_F) = \text{(I)} + \text{(II)} \ . \tag{3.8}$$

The term (II) is controlled by the discretisation parameter h. We expect that $P_{F,h} - P_F$ can be made arbitrarily small for sufficiently accurate approximations $Q_h \approx Q$ of the system response. The term (I) is not equal to zero, because the multilevel estimator as implemented in Algorithm 3 is biased, $\mathbb{E}[\widehat{P}_F^{ML}] \neq P_{F,h}$. This is mainly due to the fact that we use statistically dependent seeds in the transition from level ℓ to level $\ell + 1$ in Algorithm 3, and because the samples generated by MCMC are correlated. We expect, however, that the term (I) is controlled by the sample size N and decreases as N increases, since this can be shown for subset simulation [1, Prop. 1]. For the MC estimator (2.5) the term (I) is equal to zero.

Let us study (I) in (3.8) in more detail. To this end we take a closer look at the $2L-1$ individual estimators that make up the multilevel estimator,

$$\widehat{P}_F^{ML} = \widehat{P}_{F_1}^{MC} \prod_{\ell=2}^{L} \frac{\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC}}{\widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC}} \ .$$

Clearly, the MC estimator $\widehat{P}_{F_1}^{MC}$ is unbiased, $\mathbb{E}[\widehat{P}_{F_1}^{MC}] = \text{Prob}(F_1)$. What can we say about the remaining $2L - 2$ estimators? Recall that these estimators use statistically dependent samples generated by Markov chains. If these samples are, however, identically distributed according to the target conditional distribution, then, by taking the expected value on both sides of (2.8), we see that each of the MCMC based estimators in (3.6) is unbiased. That is, $\mathbb{E}[\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC}] = \text{Prob}(F_\ell|F_{\ell-1})$, and $\mathbb{E}[\widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC}] = \text{Prob}(F_{\ell-1}|F_\ell)$, for $\ell = 2, \ldots, L$, resp. To ensure unbiasedness of the individual MCMC based estimators, we formulate the following assumption.

A3. All samples $\boldsymbol{\theta}^{(i)}$ in Algorithm 3 that are generated by MCMC are distributed according to the associated target distribution $\varphi(\cdot|F_{\ell-1})$, for $\ell = 2, \ldots, L$.

In subset simulation, where we use the single-level failure domains (3.1), it is clear that Assumption A3 is satisfied, because the failure domains F_ℓ^{SL} are nested (see Remark 2.3). In the MLE, the failure domains F_ℓ^{ML} in (3.2) are not necessarily nested. However, Assumption A3 still holds if we use a sufficiently large burn-in length $L_b > 0$ for the Markov chains in Algorithm 3.

Proposition 3.3 *Under Assumption A3 the bias of the multilevel estimator* $\widehat{P}_F^{ML}$ *satisfies*

$$|\mathbb{E}[\widehat{P}_F^{ML}] - P_{F,h}| = \mathcal{O}(N^{-1}) \ . \tag{3.9}$$

Moreover, the variance of $\widehat{P}_F^{ML}$ *satisfies*

$$\mathbb{V}[\widehat{P}_F^{ML}] = \mathcal{O}(N^{-1}) \ . \tag{3.10}$$

Proof. The proof, which is given in Appendix A, follows the proof of Prop. 1 and Prop. 2 in [1]. □

The mean-square error of an estimator is the bias error squared plus the variance of the estimator. Squaring the bias decomposition in (3.8) we observe that, asymptotically for large N, the square of the bias error (3.9) is dominated by the variance error (3.10), and is not relevant in the mean-square error. Note that an analogous result to Proposition 3.3 is shown in [1, Prop. 1] and [1, Prop. 2] for the bias and the variance of the subset simulation estimator, respectively.

To summarise, under Assumption A3 the proposed multilevel estimator (3.6) has the same key statistical properties (bias and variance) as the subset simulation estimator (2.9). The important difference is, however, the computational cost of the estimators. As discussed in §3.4, the multilevel estimator is cheaper to evaluate, in particular, for 2D and 3D physical models associated with the system response.

4 Nestedness of failure domains

Consider a collection of failure domains

$$F_\ell^{ML} = \{\boldsymbol{\theta} \in \mathbb{R}^M : G_{h_\ell}(\boldsymbol{\theta}) \leq c_\ell\}, \quad \ell = 1, \ldots, L,$$

for use in the multilevel estimator. Here, the failure levels c_ℓ and discretisation parameters h_ℓ can be distinct on every level ℓ. When can we expect nestedness of these sets? This question is relevant for the following reasons:

- The multilevel estimator (3.6) is based on the decomposition (3.5). If the failure domains are nested, then the denominators in (3.5) are equal to one, and thus need not be estimated. Instead, the subset simulation (see Algorithm 2) can be used together with the failure domains $F_\ell = F_\ell^{ML}$ in (3.2). This reduces the absolute cost for the estimator $\widehat{P}_F^{ML}$ by 50% (see §3.4).
- Nestedness of the failure domains implies that Assumption A3 is satisfied. Then, the bias and variance of the multilevel estimator are asymptotically of the same order w.r.t. the number of samples N as in subset simulation (see Proposition 3.3).
- Nestedness of the failure domains enables a perfect simulation within MCMC; a burn-in of the Markov chains in Algorithm 3 is not required (see Remark 2.3).

In the remainder of this section we introduce a property termed *complementarity*. We will see that this property ensures nestedness of the failure domains within the multilevel estimation process. Complementarity arises from certain unbounded *one-dimensional* failure domains, that is $M = 1$. Unfortunately, for random inputs $\boldsymbol{\theta} \in \mathbb{R}^M$ with $M > 1$, and specifically in high dimensions, we cannot hope for complementarity, and thus nestedness of the failure domains F_ℓ^{ML} cannot be guaranteed easily.

4.1 Complementarity

Consider a situation where we have one input random variable θ and one constraint which defines a failure domain. Specifically, we assume that the failure domains are of the form

$$F_\ell^{rb} := \{\theta \in \mathbb{R} \colon \theta \leq b_\ell\}, \tag{4.1}$$

with real and finite bounds b_ℓ, $\ell = 1, \ldots, L$. We make the following observation.

Proposition 4.1 (Complementarity) *Let $M = 1$ and consider a pair of failure domains (F_j^{rb}, F_k^{rb}) defined in* (4.1)*. Under Assumption A1, either*
$Prob(F_k^{rb}|F_j^{rb}) < 1$ and $Prob(F_j^{rb}|F_k^{rb}) = 1$, or, alternatively, $Prob(F_k^{rb}|F_j^{rb}) = 1$ and $Prob(F_j^{rb}|F_k^{rb}) < 1$, or, finally, $Prob(F_k^{rb}|F_j^{rb}) = 1$ and $Prob(F_j^{rb}|F_k^{rb}) = 1$.

Proof. Depending on the position of the bounds b_k and b_j we distinguish three cases:
(a) If $b_k < b_j$, then $\text{Prob}(F_k^{rb}|F_j^{rb}) = \text{Prob}(F_k^{rb})/\text{Prob}(F_j^{rb}) < 1$, and $\text{Prob}(F_j^{rb}|F_k^{rb}) = \text{Prob}(F_k^{rb})/\text{Prob}(F_k^{rb}) = 1$.
(b) If $b_k > b_j$, then $\text{Prob}(F_k^{rb}|F_j^{rb}) = \text{Prob}(F_j^{rb})/\text{Prob}(F_j^{rb}) = 1$, and $\text{Prob}(F_j^{rb}|F_k^{rb}) = \text{Prob}(F_j^{rb})/\text{Prob}(F_k^{rb}) < 1$.
(c) If $b_k = b_j$, then $\text{Prob}(F_k^{rb}|F_j^{rb}) = \text{Prob}(F_j^{rb}|F_k^{rb}) = 1$. □

Remark 4.2 An analogous result to Proposition 4.1 can be proved for failure domains of the form $F_\ell^{lb} := \{\theta \in \mathbb{R} \colon \theta \geq b_\ell\}$ with real and finite bounds b_ℓ, $\ell = 1, \ldots, L$.

Corollary 4.3 *Consider the multilevel estimator for some failure probability $0 < P_F < 1$ as implemented in Algorithm 3 with $0 < p_0 < 1$. Assume that A1 holds and that all failure domains $\{F_\ell\}_{\ell=1}^L$ produced by Algorithm 3 are of the form* (4.1)*. Then these failure domains are nested for $\ell = 1, \ldots, L-1$, that is, $F_{L-1} \subseteq F_{L-2} \subseteq \cdots \subseteq F_2 \subseteq F_1$.*

Proof. Let L be the total number of failure domains at which Algorithm 3 terminates. Consider the estimation step when we go from level $\ell - 1$ to ℓ, where $1 \leq \ell < L$. In Algorithm 3 we would choose the new failure level c_ℓ and hence the bound b_ℓ in (4.1) s.t. $\text{Prob}(F_\ell|F_{\ell-1}) = p_0 < 1$. Using Proposition 4.1 this implies $\text{Prob}(F_{\ell-1}|F_\ell) = 1$. By assumption, all failure domains are of the form (4.1). Hence $\text{Prob}(F_{\ell-1}|F_\ell) = 1$ is only possible if $b_\ell \leq b_{\ell-1}$, and thus $F_\ell \subseteq F_{\ell-1}$. Since this argument holds for all $1 \leq \ell < L$ the result follows. □

Remark 4.4 In the last step of Algorithm 3 with $\ell = L$ we would obtain a failure domain F_L s.t. $\text{Prob}(F_L|F_{L-1})$ is not necessarily equal to $p_0 < 1$. In fact, $\text{Prob}(F_L|F_{L-1}) = 1$ is possible. Therefore, we cannot use Proposition 4.1 to conclude that $\text{Prob}(F_{L-1}|F_L) = 1$. Without further assumptions on the bound b_L (or, equivalently, the failure level c_L and the discretisation parameter h_L) we can guarantee nestedness of the multilevel failure domains only up to level $\ell = L - 1$.

It is important to note that the result in Corollary 4.3 holds independently of the approximation Q_{h_ℓ} that we use for the system response Q. Of course, the bounds b_ℓ in (4.1) will depend on the choice of Q_{h_ℓ}, the discretisation parameter h_ℓ, and the failure level c_ℓ. Regardless, any pair of multilevel failure domains $\{F_\ell^{ML}\}_{\ell=1}^L$ of the form (4.1) will have the complementarity property. Thus all multilevel failure domains constructed by Algorithm 3 up to level $L-1$ will be nested, even if we change h_ℓ and the failure level c_ℓ at the same time within $F_{\ell-1}^{ML}$ and F_ℓ^{ML}. We demonstrate this using the following simple example.

Example 1 Consider the initial value problem (IVP)

$$\frac{dU}{dt} = -\theta U, \qquad U(t_0) = u_0, \tag{4.2}$$

where u_0 is a constant, and θ is a standard Gaussian random variable. We define the limit state function $G(\theta) := u_{\max} - U(t)$. The associated failure event is a situation where the solution of (4.2) exceeds a given threshold $u_{\max}$ at time $t > t_0$. We approximate the exact solution U using the *forward Euler method* with time steps of size $h > 0$. We obtain the approximation $U_h(t) = u_0(1-\theta h)^{(t-t_0)/h}$ to $U(t)$, and the approximate failure probability

$$P_{F,h} = \text{Prob}(U_h(t) \geq u_{\max}) = \Phi((1 - (u_{\max}/u_0)^{h/(t-t_0)})/h). \tag{4.3}$$

Now consider the multilevel failure domains of the form (3.2) associated with Example 1. For every level ℓ we select a time step size $h_\ell > 0$ and a failure level c_ℓ s.t. $c_1 > c_2 > \cdots > c_{L-1} > c_L = 0$. Looking at (4.3) it is clear that

$$F_\ell^{ML} = \{\theta \colon u_{\max} - U_h(t) \leq c_\ell\} = \{\theta \colon \theta \leq (1 - ((u_{\max} - c_\ell)/u_0)^{h_\ell/(t-t_0)})/h_\ell\} \ .$$

Hence the multilevel failure domains are indeed of the desired form (4.1) with bounds

$$b_\ell = (1 - ((u_{\max} - c_\ell)/u_0)^{h_\ell/(t-t_0)})/h_\ell, \quad \ell = 1, \ldots, L.$$

Unfortunately, it is easy to see that we cannot establish a complementarity property as in Proposition 4.1 for more general one-dimensional failure domains that require more than one constraint, e.g. left- and right-bounded intervals. In higher dimensions with $M > 1$ it is also not possible to prove a complementarity result, e.g. for rectangular failure domains.

4.2 High-dimensional stochastic inputs

Now we study a more challenging problem with a large number of stochastic inputs $M \gg 1$. The system response Q is implicitly given as solution of a PDE with a random coefficient. Problems of this form have been studied extensively in the last decade (see e.g. [28, 46, 62] and the references therein), and are key building blocks for uncertainty quantification in computational science and engineering applications. To be able to conduct numerical experiments in a reasonable time we study a boundary value problem in a 1D physical domain.

Example 2 Consider the diffusion equation $-(au')' = 1$ for a function $u = u(x)$, $0 \leq x \leq 1$, subject to the (deterministic) boundary conditions $u(0) = 0$, and $u'(1) = 0$. The coefficient a is a lognormal random field with constant mean value $\mu_a \equiv 1$ and standard deviation $\sigma_a \equiv 0.1$. Note that $\log(a)$ is a Gaussian random field, and its mean μ and standard deviation σ are uniquely determined by μ_a and σ_a. The covariance function of $\log(a)$ is of exponential type, $c(x, y) = \exp(-|x - y|/\lambda)$, $0 \leq x, y \leq 1$, with correlation length $\lambda = 0.01$. The spatial discretisation of the boundary value problem is done by piecewise linear continuous finite elements. The system response is $Q = u(1)$ which we approximate by the finite element solution $Q_h = u_h(1)$. We are interested in the failure event $u_h(1) > u_{\max}$, where $u_{\max} = 0.535$. The associated failure probability is $P_{F,h} \approx 1.6 \times 10^{-4}$.

In Example 2 we represent $\log(a)$ by a truncated Karhunen-Loève (KL) expansion (see e.g. [45] for details) of the form

$$\log(a)(x) = \mu + \sigma \sum_{m=1}^{M} \sqrt{\nu_m} a_m(x) \theta_m, \tag{4.4}$$

where $\{\theta_m\}_{m=1}^{M}$ are statistically independent, standard Gaussian random variables. The KL eigenpairs (ν_m, a_m) are known analytically for the exponential-type correlation function (see [24, p. 26ff]). Assuming that the eigenvalues ν_m are arranged in decreasing order, we retain only the M leading terms in the KL expansion of $\log(a)$. The correlation length $\lambda = 0.01$ is very small compared to the size of the domain of interest and thus requires a large number of terms to ensure a small KL truncation error. In our experiments we use $M = 150$ which captures 87% of the variability of $\log(a)$.

Let us now compare the output of the subset simulation as implemented in Algorithm 2 to the output of the multilevel estimator as implemented in Algorithm 3. For subset simulation we use the single-level failure domains F_ℓ^{SL} in (3.1) with $h = h_{\min} := 1/512$ fixed in every F_ℓ^{SL}. For the multilevel estimator we use the failure domains F_ℓ^{ML} in (3.2) with $h_1 = 1/4$ on the first level, and $h_\ell = h_{\ell-1}/2$ until we reach the finest mesh size $h_{\min}$. Note that it is possible for Algorithm 3 to arrive at the desired failure level c_ℓ for some level $\ell \geq 1$ *before* we have actually arrived at the finest mesh size, that is, $h_\ell > h_{\min}$. In this case, we continue the simulation for $k \geq 1$ levels with $c_{\ell+k} = c_\ell$ fixed until $h_{\ell+k} = h_{\min}$. Unless stated otherwise we use $N = 10^3$ samples and $p_0 = 0.1$ fixed on every level in both algorithms.

The estimator outputs are random variables, and we compare their empirical cumulative distribution functions (c.d.f.s) obtained with 500 individual runs of each estimator. For comparison we use the so-called two-sample Kolmogorov-Smirnov (K-S) test [61]. The null hypothesis is that both data sets follow the same distribution. If it is rejected, then there is enough statistical evidence to conclude that the outputs of the subset simulation and the multilevel estimator are different. Given a certain significance level $\alpha > 0$, the K-S test incorrectly rejects the null hypothesis with probability α.

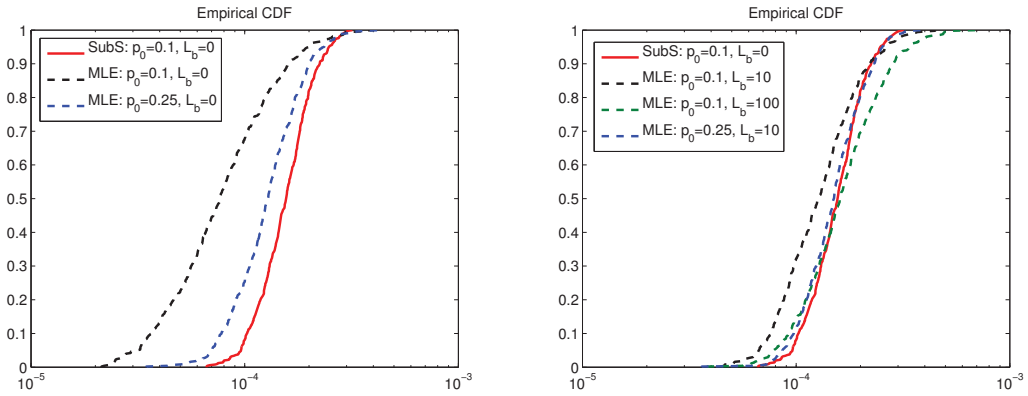


Figure 4.1: Diffusion problem (Example 2): Empirical CDFs of $\text{Prob}(u_h(1) > u_{\max})$ estimated with 500 runs of subset simulation (SubS) and the multilevel estimator (MLE). The coarsest mesh has 4 elements, the final mesh has 512 elements. We vary p_0 and/or the burn-in length L_b of the Markov chains in the MLE. We include 150 KL modes.

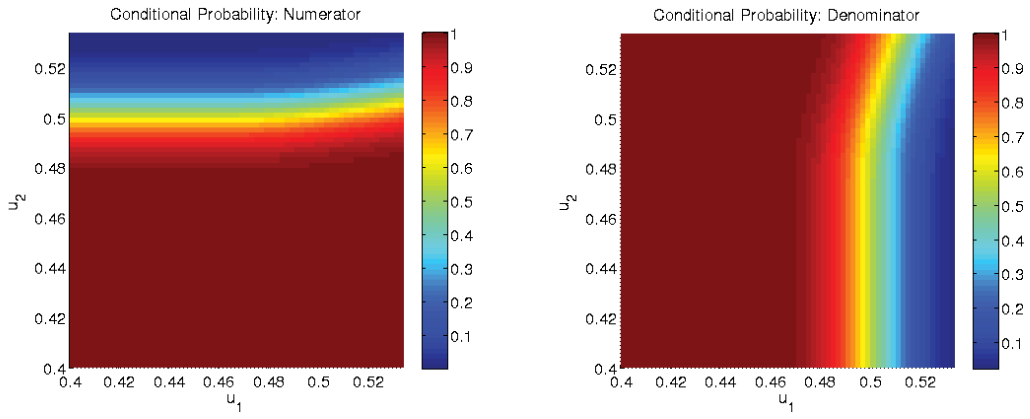


Figure 4.2: Diffusion problem (Example 2): $\text{Prob}(u_h(1) > u_2 | u_{2h}(1) > u_1)$ (left) and $\text{Prob}(u_{2h}(1) > u_1 | u_h(1) > u_2)$ (right) with $h = 1/8$. We include 150 KL modes.

In Figure 4.1 we plot the empirical c.d.f.s obtained with subset simulation and the multilevel estimator (MLE). We run 100 Markov chains of length 10 in parallel on each level; this gives a total of $N = 10^3$ samples per level. In the left panel of Figure 4.1 the burn-in-length $L_b = 0$, that is, all 10 samples in each chain are used in the estimator. However, we vary the size of the partial failure probabilities p_0. It can happen that we obtain more than 100 failure points per level, e.g. if $p_0 > 0.1$. In this case we randomly select 100 points and use those as seeds for the Markov chains in the next level. We do this to maintain a comparable correlation between the MCMC samples. Meaning that we get empirical CDFs of similar form with subset simulation even if we vary p_0. Note further that the computational cost of the MLE does not differ for $p_0 = 0.1$ and $p_0 = 0.25$ since we always use 8 finite element meshes/levels in total even though the number of intermediate failure levels is different.

In the left panel of Figure 4.1 we can clearly see that the c.d.f.s of the multilevel estimators differ significantly from the c.d.f. of the subset simulation. In fact, no data set obtained with one of the MLEs passes the K-S test when compared to subset simulation (the significance level is $\alpha = 0.01$). In the right panel of Figure 4.1 the burn-in length of the Markov chains is $L_b > 0$, and we vary again the size of p_0 in Algorithm 3. This time, the data set for $p_0 = 0.25$ and $L_b = 10$ passes the K-S comparison test with the reference set obtained by subset simulation (the significance level is $\alpha = 0.01$). It is also clear from the figure that the associated c.d.f.s match quite well. In contrast, burn-in for $p_0 = 0.1$ does not give data sets which pass the K-S comparison test with subset simulation.

We conclude from these experiments that the idea of the multilevel estimator is valid in the sense that the novel decomposition of the failure probability in (3.5) is correct. By choosing an appropriate partial failure probability p_0 and by burn-in of the Markov chains in Algorithm 3 we eventually obtain the same result as subset simulation. However, without burn-in or with p_0 too small this is not possible in Example 2. This tells us that the actual *implementation* of the MLE in (3.6) requires some care.

Assumption A3 holds for nested failure domains and ensures that we don't need to burn-in the Markov chains. In §4.1 we showed that complementarity guarantees the nestedness of simple one-dimensional failure domains independently of the physical discretisation. In a stochastically high-dimensional setting as in Example 2 this is in general not possible. To demonstrate this, we plot in Figure 4.2 the conditional probabilities

Prob($F_\ell^{ML}|F_{\ell-1}^{ML}$) and Prob($F_{\ell-1}^{ML}|F_\ell^{ML}$) for two consecutive failure domains in Example 2. The associated mesh sizes are $h_{\ell-1} = 1/4$ and $h_\ell = 1/8$, respectively. The conditional probabilities have been estimated with Monte Carlo using $N = 10^6$ samples. Note that the left plot in Figure 4.2 shows probabilities in the numerator, and the right plot shows probabilities in the denominator of the decomposition (3.5). Complementarity does not hold here, because there are pairs of failure levels (u_1, u_2) in the upper right corner of these plots where the numerator is smaller than one, and the corresponding denominator is smaller than one as well. If complementarity and thus nestedness do not hold, then burn-in of the Markov chains used in Algorithm 3 is required; this is clearly demonstrated by the results in Figure 4.1.

Why is the choice of p_0 relevant in Example 2? We see in Figure 4.2 that two consecutive failure domains on coarse meshes are nested if the failure level u_1 is small. That is, for small u_1 the denominators in the right panel of Figure 4.2 are equal to one. In Example 2 small failure levels occur for large values of the target intermediate probability p_0. Hence, as the value of p_0 increases so are the values of the denominators in (3.6). Having large values of the denominators is beneficial for two reasons. First, large denominators imply that the domain $F_\ell \cap F_{\ell-1}^C$, i.e. the domain for which $F_\ell \not\subset F_{\ell-1}$, is small. Therefore, the number of seeds used for sampling conditional on F_ℓ which do not follow the target distribution will be smaller. Hence, a shorter burn-in period can be applied. This is consistent with the right panel of Figure 4.1, where the choice $p_0 = 0.25$ and $L_b = 10$ gives a good agreeement with subset simulation. Second, large denominators will decrease the variance of the estimator $\widehat{P}_F^{\mathrm{ML}}$. This can again be observed in the right panel of Figure 4.1. Choosing $p_0 = 0.1$ and $L_b = 100$ gives nearly unbiased estimates, but the dispersion of their distribution is larger than the one of $\widehat{P}_F^{\mathrm{SubS}}$.

To obtain nested failure domains we need to make sure that the combination of approximate system responses Q_{h_ℓ} and failure levels c_ℓ leads to failure domains F_ℓ^{ML} in (3.2) such that failure on level $\ell - 1$ implies failure on level ℓ. In Example 2 this is difficult since the correlation length $\lambda = 0.01$ is very small and the then required highly oscillatory KL modes cannot be resolved on a coarse spatial mesh with size $h_1 = 1/4$. (We will come back to this in §5.) Alternatively, we can start the multilevel simulation on a finer mesh which will give a better initial approximation to the system response. Figure 4.3 shows again the empirical c.d.f.s of the subset simulation and MLE; the initial mesh for the MLE is $h_1 = 1/32$. We can clearly see that the empirical c.d.f.s match quite well this time; both data sets pass the K-S comparison test ($\alpha = 0.01$). Contrast this with Figure 4.1. In addition, in Figure 4.4 we plot the conditional probabilities Prob($F_\ell^{ML}|F_{\ell-1}^{ML}$) and Prob($F_{\ell-1}^{ML}|F_\ell^{ML}$) for two consecutive failure domains with associated mesh sizes $h_{\ell-1} = 1/32$ and $h_\ell = 1/64$, respectively. We can see that complementarity now holds for nearly all pairs (u_1, u_2), in contrast to Figure 4.2. Unfortunately, starting on a relatively fine initial mesh increases the cost of the MLE; we will not be able to benefit from really inexpensive simulations on coarse meshes. Alternatively, we now suggest a modified MLE with level-dependent stochastic input dimensions.

5 Level-dependent stochastic input dimension

5.1 Motivation and choice of dimension

Assumption A1 states that the random input vector $\boldsymbol{\theta} \in \mathbb{R}^M$ of our system contains M independent, standard Gaussian random variables. A typical model problem in this

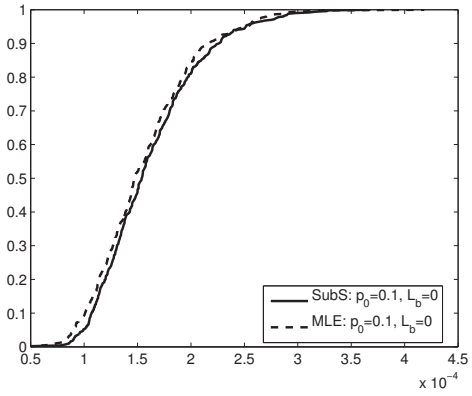


Figure 4.3: Diffusion problem (Example 2): Empirical CDFs of $\text{Prob}(u_h(1) > u_{\max})$ estimated with 500 runs of subset simulation (SubS) and the MLE. The coarsest mesh has 32 elements, the final mesh has 512 elements. We include 150 KL modes.

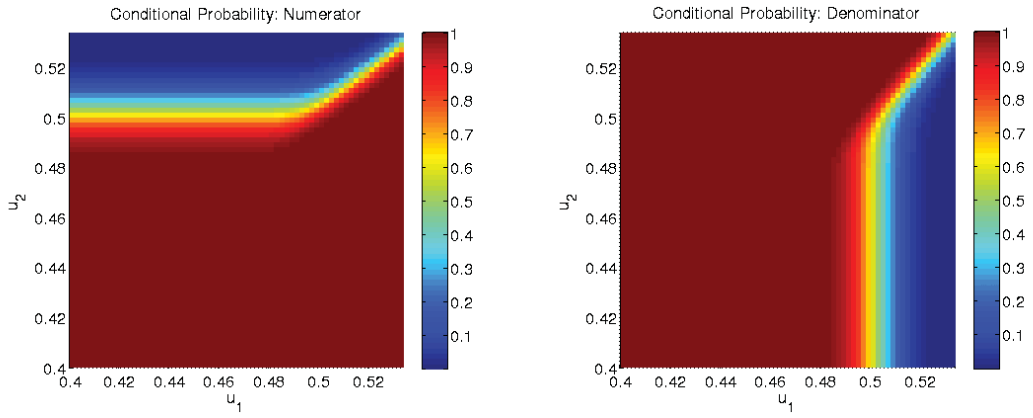


Figure 4.4: Diffusion problem (Example 2): $\text{Prob}(u_h(1) > u_2 | u_{2h}(1) > u_1)$ (left) and $\text{Prob}(u_{2h}(1) > u_1 | u_h(1) > u_2)$ (right) with $h = 1/64$. We include 150 KL modes.

setting is given in Example 2. This is a boundary value problem with a random diffusion coefficient which is represented by a truncated KL expansion. We retain the leading M terms in the KL expansion, and for small correlation lengths λ we require $M \gg 1$. In §4.2 we learned that this is a potential problem for the multilevel estimator. Let h_ℓ denote the mesh size of the finite element discretisation of the boundary value problem. If we start the multilevel simulation on a very coarse mesh, then the failure domains

$$F_\ell^{ML} = \{\boldsymbol{\theta} \in \mathbb{R}^M : G_{h_\ell}(\boldsymbol{\theta}) \leq c_\ell\}, \quad \ell = 1, \ldots, L,$$

are not nested. Hence Assumption A3 does not hold and we cannot use the implementation in Algorithm 3 without burn-in of the Markov chains. How can we overcome this problem?

Recall the basic idea of the multilevel estimator: We use different approximations to the system response Q in two consecutive failure domains $F_{\ell-1}^{ML}$ and F_ℓ^{ML}. To ensure nestedness of those we need to choose appropriate approximations $Q_\ell := Q_{h_\ell}$ and failure levels c_ℓ s.t. failure on level $\ell - 1$ implies failure on level ℓ. One way to achieve this is by making sure that $Q_\ell \approx Q_{\ell-1}$ does not differ dramatically between two consecutive levels.

In other words, we require a small variance $\mathbb{V}[Q_\ell - Q_{\ell-1}]$. Error estimates of the form

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] \leq |||Q_\ell - Q_{\ell-1}|||^2 \leq c h_\ell^\beta$$

for some constants $\beta > 0$, $c > 0$, and a norm $|||\cdot|||$ can be obtained for various system responses Q and finite element approximations arising from PDEs with random inputs (see, e.g., [15, 32, 63] for details.) This means that if the mesh size h_ℓ is small, then $\mathbb{V}[Q_\ell - Q_{\ell-1}]$ is small as well. However, such error estimates hold only asymptotically for h_ℓ sufficiently small. If we want to work on coarse spatial meshes, then it is well known, that $\mathbb{V}[Q_\ell - Q_{\ell-1}]$ can be very large if the correlation length of the random coefficient is small and we use a large, fixed number of leading terms in the KL expansion of it. This phenomenon was already observed in [18, §4.1] and can be cured by using a *level-dependent* number of KL terms, that is, $M = M_\ell$ (see [63, §4] for details).

Coming back to Example 2 we demonstrate this by comparing $\mathbb{V}[Q_\ell - Q_{\ell-1}]$ for the original setup of a fixed number of $M = 150$ KL modes on each level to the modified setup, where we have $M_1 = 10$, $M_2 = 20$, $M_3 = 40$, $M_4 = 80$, and then $M_5 = \cdots = M_L = 150$ KL modes on the resp. level, until we arrive at the final level L. The results in Figure 5.1 confirm the observations in the literature. We can clearly see that $\mathbb{V}[Q_\ell - Q_{\ell-1}]$ is very large on coarse meshes for the original, fixed number of modes setup. It can be reduced drastically by up to two orders of magnitude by using the modified, variable number of modes setup. The expected value $\mathbb{E}[|Q_\ell - Q_{\ell-1}|]$ shows a similar behavior; it is large for large h_ℓ and fixed KL modes on each level and decreases with the modified setup (see the right panel of Figure 5.1).

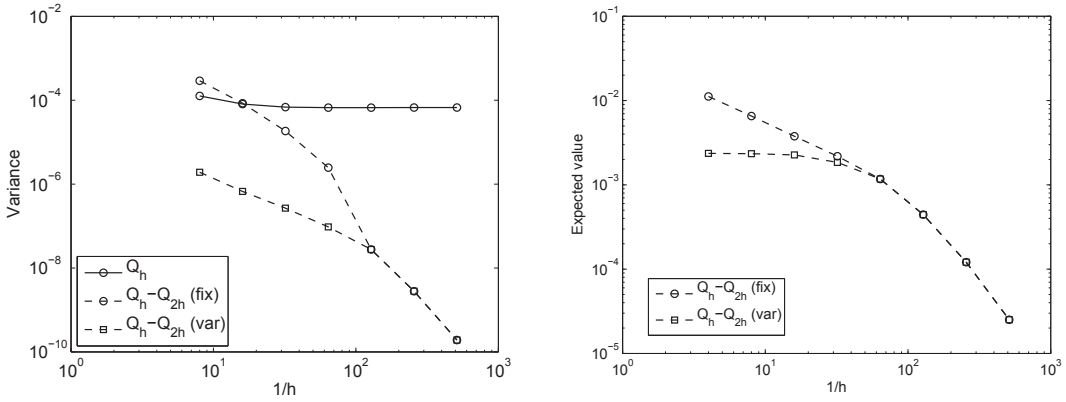


Figure 5.1: Diffusion problem (Example 2): Variance of $Q_h - Q_{2h}$ (left) and expected value of $|Q_h - Q_{2h}|$ (right) for fixed and variable KL modes setup for $Q_h = u_h(1)$.

What can we learn from this experiment? For fixed KL modes the variance $\mathbb{V}[Q_\ell - Q_{\ell-1}]$ is quite large on coarse levels, and we should not expect nestedness of the multilevel failure domains. However, the variance decreases and is small on fine meshes (see Figure 5.1). This is interesting for two reasons. First, we expect that on sufficiently fine meshes it is easier to achieve nested failure domains; this is consistent with our observations in §4.2. Second, this tells us that the burn-in length of the Markov chains is likely to depend on the level; it will be large on coarse meshes and will decrease on finer meshes (cf. §3.3). A detailed study of this observation is beyond the scope of this paper. Finally, for variable KL modes, since the variance $\mathbb{V}[Q_\ell - Q_{\ell-1}]$ is quite small on all levels, it seems possible to avoid the burn-in altogether. We will demonstrate this in §5.3.

Now we discuss our choice of the sequence $\{M_\ell\}_{\ell=1}^L$. In Example 2 $\log(a)$ is given in terms of a truncated KL expansion with eigenpairs (ν_m, a_m), $m = 1, \ldots, M$ (see (4.4)). The eigenfunctions $a_m(x)$ are sine and cosine waves with a certain angular frequency ω_m (see [24, p. 26ff]). If the associated eigenvalues ν_m are arranged in decreasing order then ω_m increases with m. Hence the modes with large m become highly oscillatory and we will not be able to represent those on coarse FE meshes.

The resolution is determined by the quadrature nodes we use to approximate the entries in the FE stiffness matrix. For example, if we use the composite trapezoidal rule, then the quadrature nodes are the vertices of the FE mesh. On a mesh with $n = 1/h$ elements this gives $n + 1$ quadrature nodes and a (spatial) sample frequency $f_s = n$. In signal theory it is well known that the sample frequency used to read a signal should be larger than the *Nyquist frequency* which is twice the frequency of the signal (see e.g. [16, Ch. 7]). For the eigenfunctions in Example 2 the associated frequency $f_m = \omega_m/2\pi$. This gives the Nyquist frequency $2f_m = \omega_m/\pi$. Hence the spacing of the spatial sample points should be at most π/ω_m for a KL mode with index m.

This theory holds for uniformly distributed sample points, however, we can also use it as a guideline for other quadrature rules. In Example 2 we use 3 Gauss Legendre nodes per element. Consideration of the Nyquist frequency leads us to a sequence of KL modes with $M_{\ell+1} \approx 2M_\ell$. We mention that this regime has been suggested in [63, §4] to balance the FE approximation/quadrature and the KL truncation error. The analysis in [63, §4] holds asymptotically for h_ℓ sufficiently small and is of limited use in the pre-asymptotic range with large mesh sizes. The Nyquist frequency provides a practical guideline in these situations.

To illustrate the idea we plot in Figure 5.2(a)–(b) one realisation of the coefficient a evaluated at the vertices and midpoints of a FE mesh with 4 and 512 elements; note that the sample frequencies are $f_s = 8$ and $f_s = 1024$, resp. In the left panel $M = 150$ gives a highly oscillatory realisation; in the right panel $M = 10$ which yields a smooth realisation. We can clearly see that it is not possible to represent the highly oscillatory realisation with only 8 sample points; the sample frequency is too small in this case. This is also confirmed by Figure 5.2(c)–(d) where we plot the amplitude spectrum of the realisations and the associated approximations in Figure 5.2(a)–(b). For $M = 150$ we observe *aliasing*; the amplitude of small frequencies in the signal is amplified in its approximation. The high frequencies in the signal are not represented. In contrast, for $M = 10$ the amplitude spectrum of the signal and its approximation agree quite well.

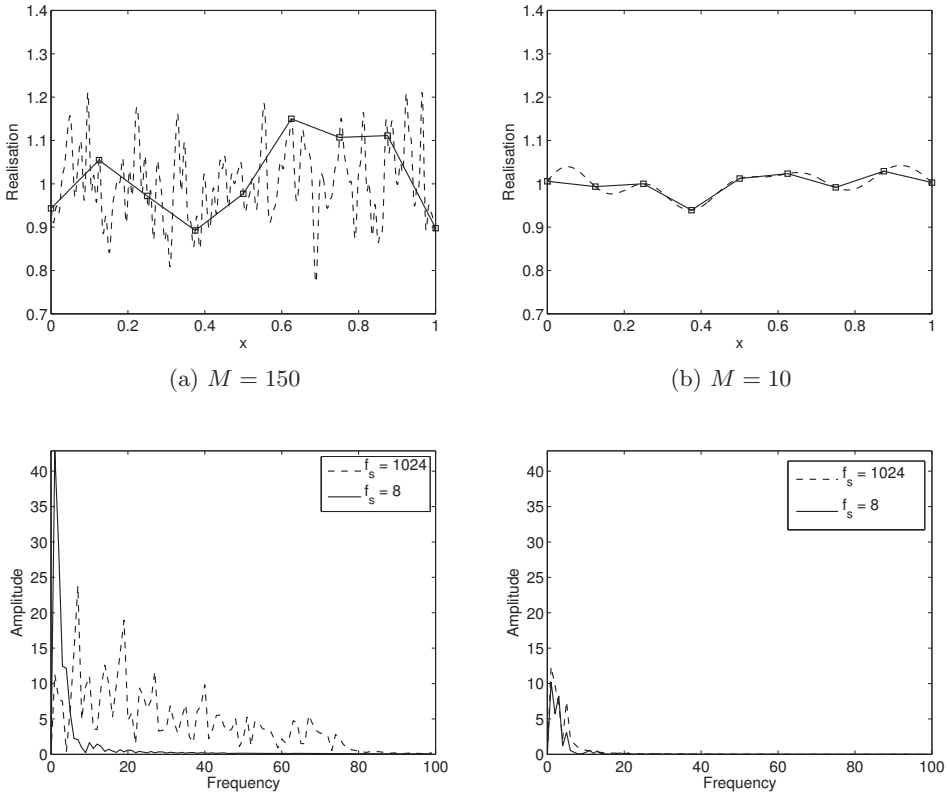
The observations described here motivate us to introduce multilevel failure domains of the form

$$F_\ell^{var} = \{\boldsymbol{\theta} \in \mathbb{R}^{M_\ell} : G_{h_\ell}(\boldsymbol{\theta}) \leq c_\ell\}, \tag{5.1}$$

where the dimension of the random input vector $\boldsymbol{\theta}$ depends on the level ℓ, that is $M = M_\ell$. To be able to use these domains, we first explain how we can modify and generalise Algorithm 3 to accommodate level-dependent stochastic input dimensions.

5.2 Modified multilevel estimator

Assume that we have a sequence of dimensions $1 \leq M_1 \leq M_2 \leq \cdots \leq M_{L-1} \leq M_L$. Let φ_{M_ℓ} denote the M_ℓ-variate standard Gaussian probability density function. On level ℓ we want to use random input vectors $\boldsymbol{\theta}_\ell \in \mathbb{R}^{M_\ell}$. We assume that the components of $\boldsymbol{\theta}_\ell$ are ordered so that $[\boldsymbol{\theta}_\ell]_{1:M_k}$ is a valid random input vector in $\mathbb{R}^{M_k}$ for $1 \leq k \leq \ell$.



(a) $M = 150$ (b) $M = 10$

(c) $M = 150$ (d) $M = 10$

Figure 5.2: Diffusion problem (Example 2): Realisation of the coefficient evaluated at the vertices and midpoints of a FE mesh with 4 elements and corresponding amplitude vs frequency plot where the spatial mean has been removed. The sample frequency is denoted by f_s.

The implementation of part (1) in Algorithm 3 is straightforward. Here $\ell = 1$ and we construct unconditional samples $\boldsymbol{\theta}_1^{(i)} \sim \varphi_{M_1}(\cdot)$ with M_1 components for use in Monte Carlo. However, the modification of part (2) in Algorithm 3 requires some care since a failure point $\boldsymbol{\theta}_{\ell-1} \in F_{\ell-1}^{var}$ has only $M_{\ell-1} \leq M_\ell$ components. How can we construct suitable points $\boldsymbol{\theta}_\ell \sim \varphi_{M_\ell}(\cdot|F_{\ell-1}^{var})$ with M_ℓ components?

Assume that we have generated a sample $\boldsymbol{\theta}_{\ell-1} \sim \varphi_{M_{\ell-1}}(\cdot|F_{\ell-1}^{var})$ on level $\ell-1$ with Algorithm 1 where we have replaced φ by $\varphi_{M_{\ell-1}}$. We then generate an *unconditional* sample $\boldsymbol{\Theta}_\ell \sim \varphi_{\Delta_\ell}(\cdot)$ where $\Delta_\ell := M_\ell - M_{\ell-1}$. Putting together the two samples we finally define $\boldsymbol{\theta}_\ell := [\boldsymbol{\theta}_{\ell-1}, \boldsymbol{\Theta}_\ell]^\top$. Note that $\boldsymbol{\theta}_\ell$ has indeed $M_{\ell-1} + \Delta_\ell = M_\ell$ components. Importantly, the surplus components in $\boldsymbol{\Theta}_\ell$ are *statistically independent* of $\boldsymbol{\theta}_{\ell-1}$. In summary, by going from level $\ell-1$ to ℓ we retain the components of the failure points $\boldsymbol{\theta}_{\ell-1} \in F_{\ell-1}^{var}$ and generate $M_\ell - M_{\ell-1}$ additional components independently of the components from level $\ell-1$. It is easy to see that for samples $\boldsymbol{\theta}_{\ell-1} \sim \varphi_{M_{\ell-1}}(\cdot|F_{\ell-1}^{var})$ this construction actually yields samples $\boldsymbol{\theta}_\ell \sim \varphi_{M_\ell}(\cdot|F_{\ell-1}^{var})$.

Finally, going from level ℓ to $\ell-1$ in part (3) of Algorithm 3 is clear: Given a failure point $\boldsymbol{\theta}_\ell \in F_\ell^{var}$ with M_ℓ components we simply discard the last $M_\ell - M_{\ell-1}$ components and hence obtain a sample with $M_{\ell-1}$ components as required. We summarise the imple-

mentation of the modified multilevel estimator termed MLE-var in Algorithm 4.

Algorithm 4 MLE-var estimator

Input: p_0, N, L_b, $\{M_\ell\}_{\ell=1}^L$

(1) Generate N i.i.d. samples $\boldsymbol{\theta}_1^{(i)} \sim \varphi_{M_1}(\cdot)$ for use in Monte Carlo.
Determine a failure level $c_1 > 0$ s.t. $\widehat{P}_{F_1}^{MC} = p_0$.

(2) For $\ell = 2, \ldots, L$:
If $\ell < 3$, set $n = 0$, else set $n = L_b$.
Use the N_0 failure points in $F_{\ell-1}^{var}$ as seeds and generate $N + (n-1)N_0$ samples $\boldsymbol{\theta}_{\ell-1}^{(i)}$ with Algorithm 1 with target distribution $\varphi(\cdot|F_{\ell-1}^{var})$. Discard the first n samples in each Markov chain.
Then, generate N samples $\boldsymbol{\Theta}_\ell^{(i)} \sim \varphi_{\Delta_\ell}(\cdot)$, $\Delta_\ell = M_\ell - M_{\ell-1}$, and set $\boldsymbol{\theta}_\ell^{(i)} = [\boldsymbol{\theta}_{\ell-1}^{(i)}, \boldsymbol{\Theta}_\ell^{(i)}]^\top$.
If $\ell < L$, determine a failure level $c_\ell > 0$ s.t. $\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC} = p_0$.
Then, use all N_0 failure points in F_ℓ^{var} as seeds and generate $N + (n-1)N_0$ samples $\boldsymbol{\theta}_\ell^{(k)}$ with Algorithm 1 with target distribution $\varphi(\cdot|F_\ell^{var})$. Discard the first n samples in each Markov chain. Evaluate $\widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC}$ as in (2.8); discard components of $\boldsymbol{\theta}_\ell^{(k)}$ where necessary.

(3) Evaluate $\widehat{P}_{F_L|F_{L-1}}^{MCMC}$ as in (2.8) and return $\widehat{P}_F^{ML}$ in (3.7).

5.3 Experiments

Now we test the implementation of the MLE-var estimator given in Algorithm 4. We continue the discussion started in §4.2 and §5.1 for Example 2. On the coarsest FE mesh we use $h_1 = 1/4$, and on the final mesh we use $h_L = 1/512$. As in §4.2 we run 100 Markov chains of length 10 each on every level in Algorithm 4. In the multilevel estimators we use $p_0 = 0.25$. This is because we observed in §4.2 that the empirical c.d.f.s of subset simulation and the MLE with burn-in length $L_b = 10$ match well in this case (see Figure 4.1). Now we would like to achieve this without burn-in. By including $M_1 = 10$, $M_2 = 20$, $M_3 = 40$, $M_4 = 80$, and then $M_5 = \cdots = M_L = 150$ KL modes we obtained a significant reduction of the variance of $Q_\ell - Q_{\ell-1}$ compared to the fixed setup with $M = 150$ on every level (see Figure 5.1). Thus we expect nearly nested multilevel failure domains (5.1) which should not require burn-in of the Markov chains in parts (2)–(3) of Algorithm 4.

In Figure 5.3 we plot the empirical c.d.f.s obtained with subset simulation, and the MLE with fixed and variable number of KL modes on every level, resp. No burn-in of the Markov chains is applied. We can see that this time the c.d.f.s of subset simulation and the MLE-var estimator match up to statistical fluctuations; both data sets pass the K-S comparison test with $\alpha = 0.01$. In addition, we have checked $\text{Prob}(u_h(1) > u_2 | u_{2h}(1) > u_1)$ and $\text{Prob}(u_{2h}(1) > u_1 | u_h(1) > u_2)$ on the two coarsest pairs of FE meshes with $h = 1/8$ for the MLE-var estimator. We have obtained a similar picture to Figure 4.4; we do not include it here for brevity. This confirms that we can achieve complementarity and thus nestedness for the multilevel failure domains in (5.1) even if we start the multilevel estimation on very coarse FE meshes.

The decomposition of the failure probability in (3.5) holds without further assumptions on the collection of partial failure domains. There is a lot of flexibility in how we define these domains and which parameters and/or model specifications change between levels. We have shown here how we can use this to construct nested or nearly nested failure domains. All observations have been made for failure events of interest arising from a PDE with random coefficients in 1D physical space. The ideas outlined here can be applied in other settings (see §6.2 for a test problem in 2D physical space). In general, however, the appropriate definition of the partial failure domains will be highly problem-dependent.

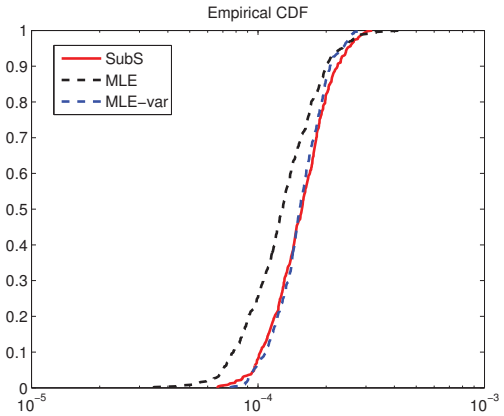


Figure 5.3: Diffusion problem (Example 2): Empirical CDFs of $\text{Prob}(u_h(1) > u_{\max})$ estimated with 500 runs of subset simulation (SubS), the MLE with 150 KL modes, and the modified MLE-var. The coarsest mesh has 4 elements, the final mesh has 512 elements.

6 Application: Rare events in a 2D flow cell

Consider the system of PDEs

$$a^{-1}\boldsymbol{q} + \nabla u = \mathbf{0}, \tag{6.1}$$

$$\nabla \cdot \boldsymbol{q} = 0 \tag{6.2}$$

in the unit square domain $D = (0,1) \times (0,1)$. In steady-state flow problems in a porous medium $\boldsymbol{q}$ is the Darcy velocity, u is the hydrostatic pressure, and a is the permeability. The empirical relation between pressure and velocity (6.1) is known as Darcy's law, and (6.2) is the law of conservation of mass. We consider a standard situation of flow from left to right, this is often termed "flow cell". The horizontal boundaries are no-flow boundaries, that is, $\nu \cdot \boldsymbol{q} \equiv 0$. We have $u \equiv 1$ along the western (inflow) and $u \equiv 0$ along the eastern (outflow) boundary. The physical discretisation of (6.1)–(6.2) is done with lowest order Raviart-Thomas mixed finite elements (see e.g. [56]) for the Darcy velocity $\boldsymbol{q}$ and piecewise constant elements for the pressure u on a uniform mesh of $n \times n$ squares each of which is divided into two triangles. We implement the Raviart-Thomas elements using a divergence-free reduction technique (see [19, 59] for details).

The permeability $a = a(\boldsymbol{x}, \boldsymbol{\theta})$ is a lognormal random field with a random input vector $\boldsymbol{\theta} \in \mathbb{R}^M$. Specifically, $\log(a)$ is a mean-zero Gaussian random field with constant variance

$\sigma^2 \equiv 1$ and a specific covariance function ρ. Here we use the exponential covariance

$$\rho(r) = \rho_{\exp}(r) := \sigma^2 \ \exp(-r/\lambda), \tag{6.3}$$

where $r = \|\boldsymbol{x} - \boldsymbol{y}\|_1$ is the distance of $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ measured in the 1-norm and $\lambda > 0$ denotes the correlation length. We generate samples of $\log(a)$ using a truncated KL expansion of the form (4.4) where only a certain number M of the leading eigenpairs is retained. For the exponential covariance (6.3) with 1-norm distance the eigenpairs in a rectangular domain can be constructed by tensorisation of the analytically known 1D eigenpairs.

6.1 Probability of breakthrough

Coming back to our motivating example in §1 we now estimate the statistics of the time it takes particles to travel from a location in the computational domain D to the boundary of it. To do this we use a simple particle tracking model. Having computed the mixed FE approximation to the Darcy velocity $\boldsymbol{q}_h$ we compute the time $\tau_h \in [0, \infty)$ when the particle hits the boundary, that is, $\boldsymbol{x}(\tau_h) \in \partial D$ for the first time. The particle path satisfies the initial value problem

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{q}_h, \quad \boldsymbol{x}(0) = \boldsymbol{x}_0 \ , \tag{6.4}$$

where $\boldsymbol{x}_0 \in D$ denotes the starting point. The Picard-Lindelöf Theorem tells us that problem (6.4) has a unique solution which can be computed element by element over the FE mesh of D. By following the particle through the domain D and summing up the travel times in each element we obtain the FE approximation τ_h to the actual travel time τ for each realisation of the random permeability. Particles are released at $\boldsymbol{x}_0 = [0, 0.5]^\top$.

We wish to estimate the probability $P_F = \text{Prob}(\tau_h < \tau_0)$ for a given breakthrough time $\tau_0 \in [0, \infty)$. The associated limit state function is $G_h(\boldsymbol{\theta}) = \tau_h(\boldsymbol{\theta}) - \tau_0$. We are interested in rare events where τ_0 is very small. This means that particles leave the computational domain after a very short amount of time. In our example the maximal travel time is $\tau_0 = 0.03$ which gives a failure probability $P_{F,h} \approx 10^{-6}$. The correlation length of $\log(a)$ is $\lambda = 0.5$ and we truncate its KL expansion after $M = 150$ leading terms.

In the MLE we choose $p_0 = 0.1$ and $N = 10^3$ samples. Normally, this gives $Np_0 = 100$ seeds for Markov chains of length 10 in the estimation of each factor. If, however, the number of seeds is too large and thus the length of each chain would be smaller than 10, then we only use 10 seeds for chains of length 100 each. No burn-in is applied in the MCMC runs.

We compare the subset simulation with mesh size $h_\ell = 1/128$ on each level ℓ to the MLE where the coarsest mesh has 4 elements in each spatial direction. The results are depicted in the left panel of Figure 6.1. The data sets generated by the MLE and subset simulation pass the K-S comparison test with significance level $\alpha = 0.01$. The MLE saves 31% of the computational cost compared to subset simulation.

6.2 Exceedence of outflow through boundary

As a proof of concept we consider a situation with short correlation length $\lambda = 0.05$. It requires a large number of KL modes of $\log(a)$ and thus very high-dimensional stochastic inputs. This time we use the primal formulation of (6.1)–(6.2) where we have eliminated the flux $\boldsymbol{q}$. We consider the system response

$$g = -\int_D a \nabla u \cdot \nabla \psi d\boldsymbol{x}, \tag{6.5}$$

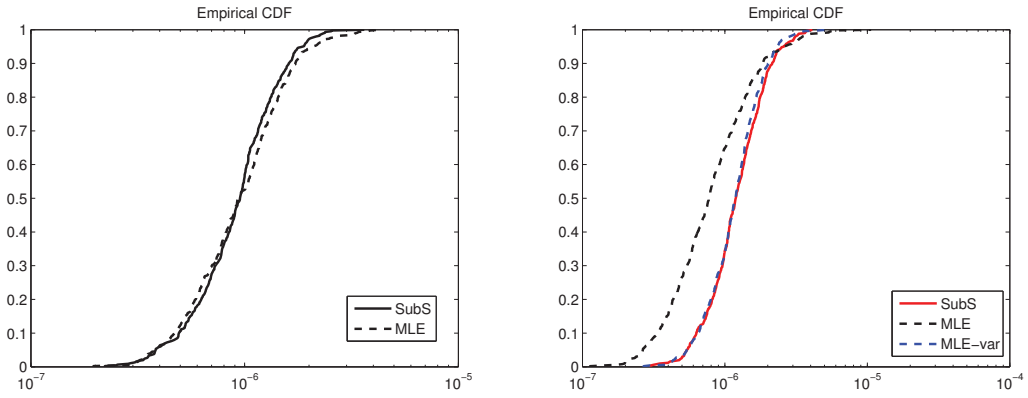


Figure 6.1: Empirical CDFs of failure probabilities in a 2D flow cell. Left: Probability of breakthrough $\text{Prob}(\tau_h < 0.03)$ obtained with 450 runs. Right: Probability of exceedence of the outflow $\text{Prob}(g_h > 0.63)$ obtained with 500 runs. The estimators are subset simulation (SubS), the multilevel estimator (MLE), and the MLE-var estimator.

where $\psi|_{\Gamma_{\text{out}}} \approx 1$ and $\psi|_{D\setminus\Gamma_{\text{out}}} \approx 0$. It can be shown that g in (6.5) approximates the outflow through the boundary Γ_{out} of D (see e.g. [20]). We approximate the pressure u by standard piecewise linear FEs. Replacing u by u_h in (6.5) gives the approximation g_h to the outflow.

We estimate the probability $P_{F,h} = \text{Prob}(g_h > g_{\min})$ for a given lower bound $g_{\min}$. The associated limit state function is $G_h(\boldsymbol{\theta}) = g_{\min} - g_h(\boldsymbol{\theta})$. In our example the minimal outflow is $g_{\min} = 0.63$ which gives again a failure probability $P_{F,h} \approx 10^{-6}$. We compare the subset simulation with mesh size $h_\ell = 1/256$ on each level ℓ to the MLE where the coarsest mesh has 4 elements in each spatial direction. For the MLE we include $M = 1024$ modes in the KL of $\log(a)$. In addition we use the modified ML-var estimator introduced in §5.2. The sequence of KL modes per level is $(4, 8, 16, 64, 256, 512, 1024, 1024, \dots)$. The MCMC setup is the same as in §6.1. In particular, no burn-in is applied.

We give the results in the right panel of Figure 6.1. The data sets generated by the MLE-var and subset simulation pass the K-S comparison test with significance level $\alpha = 0.01$. The MLE-var saves 48% of the computational cost compared to subset simulation. As expected the empirical c.d.f. of the MLE does not agree with the c.d.f. of the subset simulation estimator. This confirms our observations in §5.3 for a more challenging application in 2D physical space.

7 Conclusions

We have implemented a novel multilevel estimator (MLE) for the probability P_F of rare events. It is a generalisation of the well-known subset simulation procedure. Specifically, we have studied failure events arising from PDEs with random coefficients given in terms of a truncated KL expansion. For a 2D flow cell problem with $P_F \approx 10^{-6}$ the MLE saves 31% and 48 % of the cost compared to subset simulation. If no burn-in applies we expect more savings of up to 70 % for smaller P_F. The MLE requires MCMC runs, and the nestedness of the associated multilevel failure domains is crucial for a good performance of MCMC. Ideally, nestedness enables a perfect MCMC simulation without burn-in of the

chains. We have shown that nestedness follows from the complementarity of certain simple unbounded, one-dimensional failure domains. To maintain nestedness for high-dimensional stochastic inputs it is necessary to use a level-dependent number of KL modes to construct the failure domains.

Acknowledgements

We thank Robert Scheichl, Oliver Ernst, and Wolfgang Betz for useful comments and discussions. The computations in §6 were carried out on the University of Bath's High Performance Computing Facility, time on which was funded by the Engineering and Physical Sciences Research Council, UK, under grant EP/H051503/1.

A Proof of Proposition 3.3

To analyse the bias we define random variables associated with the individual estimators in the MLE (3.6):

$$\begin{aligned} Z_1 &:= (\widehat{P}_{F_1}^{MC} - \mathrm{Prob}(F_1))/\sigma_{1,Z}, \quad \sigma_{1,Z} := (\mathbb{V}[\widehat{P}_{F_1}^{MC}])^{1/2}, \\ Z_\ell &:= (\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC} - \mathrm{Prob}(F_\ell|F_{\ell-1}))/\sigma_{\ell,Z}, \quad \sigma_{\ell,Z} := (\mathbb{V}[\widehat{P}_{F_\ell|F_{\ell-1}}^{MCMC}])^{1/2}, \\ Y_\ell &:= (\widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC} - \mathrm{Prob}(F_{\ell-1}|F_\ell))/\sigma_{\ell,Y}, \quad \sigma_{\ell,Y} := (\mathbb{V}[\widehat{P}_{F_{\ell-1}|F_\ell}^{MCMC}])^{1/2}, \quad \ell = 2,\ldots,L. \end{aligned} \tag{A.1}$$

By construction or by Assumption A3 we have $\mathbb{E}[Z_1] = 0$, $\mathbb{E}[Z_\ell] = 0$, and $\mathbb{E}[Y_\ell] = 0$, $\ell = 2,\ldots,L$. Moreover, by construction,

$$\mathbb{E}[Z_1^2] = 1, \quad \mathbb{E}[Z_\ell^2] = 1, \ \text{ and } \ \mathbb{E}[Y_\ell^2] = 1, \ \ell = 2,\ldots,L. \tag{A.2}$$

Let $\delta_{1,Z}$, $\delta_{\ell,Z}$, and $\delta_{\ell,Y}$, $\ell = 2,\ldots,L$, denote the c.o.v.s of the random variables in (A.1). Note that $\delta_{1,Z} = \mathcal{O}(N^{-1/2})$ (see §2.2). Assuming that the samples generated by different Markov chains in the subset simulation procedure are independent, it is demonstrated in [1, §6] that the c.o.v.s of the remaining variables show the same asymptotic behaviour with respect to N. In summary,

$$\delta_{1,Z} = \mathcal{O}(N^{-1/2}), \ \delta_{\ell,Z} = \mathcal{O}(N^{-1/2}), \ \delta_{\ell,Y} = \mathcal{O}(N^{-1/2}), \ \ell = 2,\ldots,L. \tag{A.3}$$

Now, if $\delta_{\ell,Y}$ is sufficiently small (or, equivalently, N is sufficiently large), we obtain

$$\frac{1+\delta_{\ell,Z}Z_\ell}{1+\delta_{\ell,Y}Y_\ell} = 1 + \delta_{\ell,Z}Z_\ell - \delta_{\ell,Y}Y_\ell - \delta_{\ell,Z}\delta_{\ell,Y}Z_\ell Y_\ell + \delta_{\ell,Y}^2 Y_\ell^2 + R_\ell, \qquad \ell = 2,\ldots,L,$$

where R_ℓ is a random variable with $R_\ell \in \mathrm{span}\{Z_\ell^n Y_\ell^m\}_{m+n\geq 3}$. Let $A_1 := \delta_{1,Z}Z_1$, and $A_\ell := \delta_{\ell,Z}Z_\ell - \delta_{\ell,Y}Y_\ell - \delta_{\ell,Z}\delta_{\ell,Y}Z_\ell Y_\ell + \delta_{\ell,Y}^2 Y_\ell^2$, $\ell = 2,\ldots,L$. Then, the relative bias of $\widehat{P}_F^{ML}$ satisfies

$$\begin{aligned} \frac{\widehat{P}_F^{ML} - P_{F,h}}{P_{F,h}} &= (1+\delta_{1,Z}Z_1)\frac{\prod_{\ell=2}^L(1+\delta_{\ell,Z}Z_\ell)}{\prod_{\ell=2}^L(1+\delta_{\ell,Y}Y_\ell)} - 1 = \prod_{\ell=1}^L(1+A_\ell) - 1 + B \\ &= \sum_{\ell=1}^L A_\ell + \sum_{\ell>k} A_\ell A_k + \sum_{\ell>k>j} A_\ell A_k A_j + \cdots + \prod_{\ell=1}^L A_\ell + B\ , \end{aligned} \tag{A.4}$$

where B is a random variable with $B \in \text{span}\{Z_\ell^n Y_\ell^m\}_{m+n\geq 3, \ell=2,\dots,L}$. Taking the expectation on both sides of the equation above (assuming that the variables Z_ℓ and Y_ℓ have sufficiently many moments) we arrive at

$$\frac{\mathbb{E}[\widehat{P}_F^{ML}] - P_{F,h}}{P_{F,h}} = \sum_{\ell=1}^{L} \mathbb{E}[A_\ell] + \sum_{\ell>k} \mathbb{E}[A_\ell A_k] + \sum_{\ell>k>j} \mathbb{E}[A_\ell A_k A_j] + \cdots + \mathbb{E}[\prod_{\ell=1}^{L} A_\ell] + \mathbb{E}[B] \ .$$

Note that $\mathbb{E}[A_1] = 0$, and $\mathbb{E}[A_\ell] = \delta_{\ell,Y}^2 \mathbb{E}[Y_\ell^2] - \delta_{\ell,Z}\delta_{\ell,Y}\mathbb{E}[Z_\ell Y_\ell]$, $\ell = 2, \dots, L$. Let $Y_1 := 0$. Using again the asymptotics in (A.3) we obtain

$$\begin{aligned}\frac{\mathbb{E}[\widehat{P}_F^{ML}] - P_{F,h}}{P_{F,h}} &= \sum_{\ell=2}^{L} \delta_{\ell,Y}^2 \mathbb{E}[Y_\ell^2] - \delta_{\ell,Z}\delta_{\ell,Y}\mathbb{E}[Z_\ell Y_\ell] + \sum_{\ell>k} \mathbb{E}[A_\ell A_k] + \mathbb{E}[B] + \mathcal{O}(N^{-1}) \\ &= \sum_{\ell\geq k} \delta_{\ell,Y}\delta_{k,Y}\mathbb{E}[Y_\ell Y_k] + \sum_{\ell>k} \delta_{\ell,Z}\delta_{k,Z}\mathbb{E}[Z_\ell Z_k] - \sum_{k,\ell} \delta_{\ell,Y}\delta_{k,Y}\mathbb{E}[Z_k Y_\ell] + \mathbb{E}[B] + \mathcal{O}(N^{-1}) \ .\end{aligned}$$

Using the triangle inequality, the Cauchy-Schwarz inequality, (A.2), and (A.3) we arrive at

$$\left|\frac{\mathbb{E}[\widehat{P}_F^{ML}] - P_{F,h}}{P_{F,h}}\right| \leq \sum_{\ell\geq k} \delta_{\ell,Y}\delta_{k,Y} + \sum_{\ell>k} \delta_{\ell,Z}\delta_{k,Z} + \sum_{k,\ell} \delta_{\ell,Y}\delta_{k,Z} + |\mathbb{E}[B]| + \mathcal{O}(N^{-1}) = \mathcal{O}(N^{-1}) \ .$$

Hence the term (I) in the bias decomposition (3.8) satisfies (3.9). To bound the variance of the MLE we use the estimate

$$\begin{aligned}\mathbb{V}[\widehat{P}_F^{ML}] &\leq \mathbb{V}[\widehat{P}_F^{ML}] + (\mathbb{E}[\widehat{P}_F^{ML}] - P_{F,h})^2 = \mathbb{E}[\widehat{P}_F^{ML} - \mathbb{E}[\widehat{P}_F^{ML}] + \mathbb{E}[\widehat{P}_F^{ML}] - P_{F,h}]^2 \\ &= \mathbb{E}[\widehat{P}_F^{ML} - P_{F,h}]^2 = P_{F,h}^2 \mathbb{E}\left[(\widehat{P}_F^{ML} - P_{F,h})/P_{F,h}\right]^2 \ .\end{aligned}$$

Recalling (A.4) and following the line of arguments to bound the bias it is easy to see that $\mathbb{E}[(\widehat{P}_F^{ML} - P_{F,h})/P_{F,h}]^2 = \mathcal{O}(N^{-1})$ and thus the variance of the MLE satisfies (3.10).

References

[1] S.-K. Au and J. L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Prob. Eng. Mech.*, 16:263–277, 2001.

[2] S.-K. Au and J.L. Beck. Subset simulation and its application to seismic risk based on dynamic analysis. *J. Eng. Mech., ASCE*, 129(8):901–917, 2003.

[3] S.-K. Au, J. Ching, and J.L. Beck. Application of subset simulation methods to reliability benchmark problems. *Struct. Saf.*, 29:183–193, 2007.

[4] S.-K. Au and Y. Wang. *Engineering Risk Assessment with Subset Simulation.* Wiley, Singapore, 2014.

[5] S.-K. Au, J.L. Beck K.M. Zuev, and L.S. Katafygiotis. Discussion of paper by F. Miao and M. Ghosn "Modified Subset Simulation method for reliability analysis of structural systems". *Struct. Saf.*, 34:379–380, 2011.

[6] S.K. Au and J.L. Beck. A new adaptive importance sampling scheme. *Struct. Saf.*, 21(2):135–158, 1999.

[7] S.K. Au and J.L. Beck. Importance sampling in high dimensions. *Struct. Saf.*, 25(2):139–163, 2003.

[8] A. Barth, Ch. Schwab, and N. Zollinger. Multi-level Monte Carlo finite element method for elliptic PDE's with stochastic coefficients. *Numer. Math.*, 119:123–161, 2011.

[9] P. Beaurepaire, M.A. Valdebenito, G.I. Schuëller, and H.A. Jensen. Reliability-based optimization of maintenance scheduling of mechanical components under fatigue. *Comput. Methods Appl. Mech. Engrg.*, 221–222:24–40, 2012.

[10] Z. I. Botev and D. P. Kroese. Efficient Monte Carlo simulation via the generalized splitting method. *Stat. Comput.*, 22:1–16, 2012.

[11] J.-M. Bourinet, F. Deheeger, and M. Lemaire. Assessing small failure probabilities by combined subset simulation and supprot vector machines. *Struct. Saf.*, 33(6):343–353, 2011.

[12] Ch. Bucher. Adaptive sampling – an iterative fast Monte Carlo procedure. *Struct. Saf.*, 5(2):119–126, 1988.

[13] Ch. Bucher. Asymptotic sampling for high-dimensional reliability analysis. *Prob. Eng. Mech.*, 24(4):504–510, 2009.

[14] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Stat. Comput.*, 22:795–808, 2012.

[15] J. Charrier, R. Scheichl, and A.L. Teckentrup. Finite Element Error Analysis of Elliptic PDEs with Random Coefficients and its Application to Multilevel Monte Carlo Methods. *SIAM J. Numer. Anal.*, 51:322–352, 2013.

[16] C. Chatfield. *The Analysis of Time Series: Theory and Practice*. Chapman and Hall, 1975.

[17] P. Chen and A. Quarteroni. Accurate and efficient evaluation of failure probability for partial differential equations with random input data. *Comput. Methods Appl. Mech. Engrg.*, 267:233–260, 2013.

[18] K.A. Cliffe, M.B. Giles, R. Scheichl, and A.L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Comput. Visual. Sci.*, 14(1):3–15, 2011.

[19] K.A. Cliffe, I.G. Graham, R. Scheichl, and L. Stals. Parallel Computation of Flow in Heterogeneous Media Modelled by Mixed Finite Elements. *J. Comput. Phys.*, 164:258–282, 2000.

[20] J. Douglas, T. Dupont, and M.F. Wheeler. A Galerkin procedure for approximating the flux on the boundary for elliptic and parabolic boundary value problems. *ESAIM, Math. Model. Numer. Anal.*, 8:47–59, 1974.

[21] V. Dubourg, B. Sudret, and F. Deheeger. Metamodel-based importance sampling for structural reliability analysis. *Prob. Eng. Mech.*, 33:47–57, 2013.

[22] D. Elfverson, F. Hellman, and A. Måhlquist. A multilevel Monte Carlo method for computing failure probabilities. Available from arXiv:1408.6856.

[23] R.A. Freeze. A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resour. Res.*, 11(5):725–741, 1975.

[24] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, New York, 1991.

[25] M. Giles, T. Nagapetyan, and K. Ritter. Multilevel Monte Carlo approximation of distribution function and densities. Technical Report 157, DFG SPP 1324. Available from http://www.dfg-spp1324.de/download/preprints/preprint157.pdf.

[26] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[27] W.R. Gilks, G.O. Roberts, and S.K. Sahu. Adaptive Markov chain Monte Carlo through regeneration. *J. Am. Statist. Assoc.*, 93(443):1045–1054, 1998.

[28] M.D. Gunzburger, C.G. Webster, and G. Zhang. Stochastic finite element methods for partial differential equations with random input data. *Acta Numerica*, 23:521–650, 2014.

[29] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[30] S. Heinrich. Multilevel Monte Carlo methods. In S. Margenov, J. Wasniewski, and P. Yalamov, editors, *LSSC 2001 Revised Papers*, Lecture Notes in Sci. Comput. 2179, pages 58–67. Springer-Verlag, 2001.

[31] R. Hoeksema and P.K. Kitanidis. Analysis of the spatial structure of properties of selected water aquifers. *Water Resour. Res.*, 21(4):536–572, 1985.

[32] I.G.Graham, R. Scheichl, and E. Ullmann. Mixed Finite Element Analysis of Lognormal Diffusion and Multilevel Monte Carlo Methods. Available from arXiv:1312.6047.

[33] A.M. Johansen, P. Del Moral, and A. Doucet. Sequential Monte Carlo samplers for rare events. In *Proc. 6th International Workshop on Rare Event Simulation, Bamberg*, pages 256–267, 2006.

[34] H. Kahn and T.E. Harris. Estimation of Particle Transmission by Random Sampling. *National Bureau of Standards Appl. Math. Series*, 12:27–30, 1951.

[35] L.S. Katafygiotis and K.M. Zuev. Geometric insight into the challenges of solving high-dimensional reliability problems. *Prob. Eng. Mech.*, 23(2–3):208–218, 2008.

[36] A. Der Kiureghian. *Engineering Design Reliability Handbook*, chapter First- and second- order reliability methods. CRC Press, Boca Raton, FL, 2005.

[37] P.-S. Koutsourelakis. Accurate uncertainty quantification using inaccurate models. *SIAM J. Sci. Comput.*, 31(5):3274–3300, 2009.

[38] P.S. Koutsourelakis, H.J. Pradlwarter, and G.I. Schuëller. Reliability of structures in high-dimensions, part I: algorithms and applications. *Prob. Eng. Mech.*, 19(4):409–417, 2004.

[39] N. Kurtz and J. Song. Cross-entropy-based adaptive importance sampling using Gaussian mixture. *Struct. Saf.*, 42:35–44, 2013.

[40] M. Lemaire. *Structural Reliability*. Wiley, Hoboken, NJ, 2009.

[41] J. Li, J. Li, and D. Xiu. An efficient surrogate-based method for computing rare failure probability. *J. Comput. Phys.*, 230:8683–8697, 2011.

[42] J. Li and D. Xiu. Evaluation of failure probability via surrogate models. *J. Comput. Physics*, 229:8966–8980, 2010.

[43] J. Li and D. Xiu. Surrogate based method for evaluation of failure probability under multiple constraints. *SIAM J. Sci. Comput.*, 36(2):A828–A845, 2014.

[44] J. Liu, J. Lu, and X. Zhou. Efficient rare event simulation for failure problems in random media. Available from arXiv:1404.4225.

[45] M. Loève. *Probability Theory*, volume II. Springer-Verlag, New York, Heidelberg, Berlin, 4th edition, 1978.

[46] H. G. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 194:1295–1331, 2005.

[47] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J.Chem.Phys.*, 21:1087–1092, 1953.

[48] P. Del Moral and P. Lezaud. Branching and interacting particle interpretation of rare events. In H. Blom and J. Lygeros, editors, *Stochastic Hybrid Systems: Theory and Safety Critical Applications*, Lecture Notes in Control and Information Sciences 337, pages 277–323, 2006.

[49] P. Mykland, L. Tierney, and B. Yu. Regeneration in Markov chain samplers. *J. Am. Statist. Assoc.*, 90(429):233–241, 1995.

[50] V. Papadopoulos, D.G. Giovanis, N.D. Lagaros, and M. Papadrakakis. Accelerated subset simulation with neural networks for reliability analysis. *Comput. Methods Appl. Mech. Engrg.*, 223–224:70–80, 2012.

[51] I. Papaioannou. *Non-intrusive Finite Element Reliability Analysis*. SVH-Verlag, Saarbrücken, 2013.

[52] I. Papaioannou, W. Betz, K. Zwirglmaier, and D. Straub. MCMC algorithms for subset simulation. Available from `www.era.bgu.tum.de`.

[53] I. Papaioannou and D. Straub. Reliability updating in geotechnical engineering including spatial variability of soil. *Computers and Geotechnics*, 42:44–51, 2012.

[54] M.F. Pellissetti, G.I. Schuëller, H.J. Pradlwarter, A. Calvi, S. Fransen, and M. Klein. Reliability analysis of spacecraft structures under static and dynamic loading. *Computers and Structures*, 84:1313–1325, 2006.

[55] R. Rackwitz. Reliability analysis – a review and some perspectives. *Struct. Saf.*, 23(4):365–395, 2001.

[56] P.-A. Raviart and J. M. Thomas. A mixed finite element method for second order elliptic problems. In I. Galligani and E. Magenes, editors, *Mathematical Aspects of the Finite Element Method*, Lecture Notes in Math. 606, pages 292–315. Springer-Verlag, 1977.

[57] Ch. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 1999.

[58] M. Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23:470–472, 1952.

[59] R. Scheichl. *Iterative Solution of Saddle-Point Problems using Divergence-free Finite Elements with Applications to Groundwater Flow*. PhD thesis, University of Bath, 2000.

[60] G.I. Schuëller and R. Stix. A critical appraisal of methods to determine failure probabilities. *Struct. Saf.*, 4(4):239–309, 1987.

[61] N.V. Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow, Univ., Bull. Math., (série internationale),* v. 2, 1939, fasc. 2, 16p.

[62] G. Stefanou. The stochastic finite element method: Past, present and future. *Comput. Methods Appl. Mech. Engrg.*, 198(9–12):1031–1051, 2009.

[63] A.L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numer.Math.*, 125(3):569–600, 2013.

[64] M.A. Valdebenito, H.J. Pradlwarter, and G.I. Schuëller. The role of the design point for calculating failure probabilities in view of dimensionality and structural nonlinearities. *Struct. Saf.*, 32(2):101–111, 2010.

[65] Y. Wang, Z. Cao, and S.-K. Au. Practical reliability analysis of slope stability by advanced Monte Carlo simulations in a spreadsheet. *Can. Geotech. J.*, 48(1):162–172, 2011.

[66] E. Zio and N. Pedroni. Estimation of the functional failure probability of a thermal-hydraulic passive system by subset simulation. *Nucl. Eng. Des.*, 239(3):580–599, 2009.

[67] K.M. Zuev, J.L. Beck, S.-K. Au, and L.S. Katafygiotis. Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions. *Computers and Structures*, 92–93:283–296, 2012.