Adaptive Multilevel Subset Simulation with Selective Refinement[*]

D. Elfverson[†], R. Scheichl[‡], S. Weissmann[§], and F. A. Diaz De La O[¶]

Abstract. In this work we propose an adaptive multilevel version of subset simulation to estimate the probability of rare events for complex physical systems. Given a sequence of nested failure domains of increasing size, the rare event probability is expressed as a product of conditional probabilities. The proposed new estimator uses different model resolutions and varying numbers of samples across the hierarchy of nested failure sets. In order to dramatically reduce the computational cost, we construct the intermediate failure sets such that only a small number of expensive high-resolution model evaluations are needed, whilst the majority of samples can be taken from inexpensive low-resolution simulations. A key idea in our new estimator is the use of a posteriori error estimators combined with a selective mesh refinement strategy to guarantee the critical subset property that may be violated when changing model resolution from one failure set to the next. The efficiency gains and the statistical properties of the estimator are investigated both theoretically via shaking transformations, as well as numerically. On a model problem from subsurface flow, the new multilevel estimator achieves gains of more than a factor 60 over standard subset simulation for a practically relevant relative error of 25%.

Key words. rare event probabilities, adaptive model hierarchies, high-dimensional problems, Markov chain Monte Carlo, shaking transformations

MSC codes. 65N30, 65C05, 65C40, 35R60

DOI. 10.1137/22M1515240

1. Introduction. Estimating the probability of rare events is one of the most important and computationally challenging tasks in science and engineering. By definition, the probability that a rare event occurs is very small. However, its effect could be catastrophic, e.g., when it is associated with some critical system failure such as the breakthrough of pollutants into a water reservoir or the structural failure of an airplane wing. An efficient and reliable estimator is of utmost importance in such situations.

[†]Formerly at Department of Mathematics and Mathematical statistics, Umeå University, SE-901 87 Umeå, Sweden (daniel.elfverson@gmail.com).
[‡]Institute for Applied Mathematics & Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, D-69120 Heidelberg, Germany (r.scheichl@uni-heidelberg.de).
[§]Institute of Mathematics, University of Mannheim, D-68138 Mannheim, Germany (simon.weissmann@uni-mannheim.de).
[¶]Clinical Operational Research Unit, Department of Mathematics, University College London, London WC1H 0BT, UK (alex.diaz@ucl.ac.uk).

We are interested in estimating rare event probabilities occurring in mathematical models of physical systems described by stochastic differential equations (SDEs) or partial differential equation (PDEs) with uncertain data, with possibly large stochastic dimension. Standard Monte Carlo (MC) methods are infeasible due to the huge amount of samples needed to produce even a single rare event. Remedies for this are offered by different variance reduction techniques, such as *importance sampling* [37, 38, 12], *multilevel MC methods* [24, 7, 18, 23, 13, 31], and *subset simulation* [5, 6, 34, 39]. In the statistics and probability literature [30], subset simulation is also known as *splitting* and can be interpreted as a sequential MC method [10, 16, 28, 15, 11]. There are also Bayesian versions of subset simulation using Gaussian process emulators to reduce the number of model evaluations [9]. Moreover, a cross-entropy-based importance sampling method for rare events and a failure-informed dimension reduction through the connection to Bayesian inverse problems has been addressed in [40].

In this work, we propose a multilevel version of the classical subset simulation approach of Au and Beck [5] for rare event estimation in the context of complex models. More specifically, we combine the ideas of incorporating multilevel model resolutions across the subsets proposed in [39] with the selective mesh refinement strategy proposed in [23]. As a result, our method does not violate the subset property, critical in subset simulation, when changing the model resolution. Furthermore, we formulate the subset simulation based on shaking transformations to ensure asymptotic convergence of the resulting estimator, as shown in [26].

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space with σ-algebra Σ and probability measure $\mathbb{P}$, defined over some nonempty set Ω. Given a mathematical model that describes the behavior of a physical system, the sample space Ω will form the input space for any sources of uncertainty or randomness. Unless otherwise stated, we refer to a random variable X with state space $\mathbb{R}^d$ as a Σ-$\mathcal{B}(\mathbb{R}^d)$-measurable mapping. The expected value of a measurable functional $F : \mathbb{R}^d \to \mathbb{R}^m$ of such a random variable X with density π is denoted by

$$\mathbb{E}[F(X)] = \int_\Omega F(X(\omega))\,\mathbb{P}(\mathrm{d}\omega) = \int_{\mathbb{R}^d} F(x)\pi(\mathrm{d}x) =: \mathbb{E}_\pi[F(X)].$$

If it is clear from the context, we will suppress the subscript π and simply write $\mathbb{E}[F(X)]$.

Let $F \in \Sigma$ be a *rare event* that can be associated to system failure, that is, the demand exceeding the capacity of the system under study. Given the above probability space, subset simulation splits the estimation of $\mathbb{P}(F)$ into a product of conditional probabilities by introducing a sequence of nested events

$$(1.1) \qquad F = F_K \subset F_{K-1} \subset \cdots \subset F_1 \subset F_0 = \Omega, \quad F_j \in \Sigma, \ j = 0, \ldots, K,$$

corresponding to larger failure probabilities as the subscript $j \to 0$. The rare event probability $\mathbb{P}(F)$ is usually prohibitive to estimate directly by MC sampling, but it can be expressed as the product of conditional probabilities

$$(1.2) \qquad \mathbb{P}(F) = \prod_{j=1}^K \mathbb{P}(F_j | F_{j-1}),$$

where each $\mathbb{P}(F_j | F_{j-1})$ is larger than $\mathbb{P}(F)$ and hence easier to estimate. The conditional probabilities in (1.2) can be estimated effectively using Markov Chain Monte Carlo (MCMC).

Computer simulations to study physical systems very often take the form of SDE or PDE models. Making use of a hierarchy of discretizations of these underlying models, we exploit ideas from multilevel Monte Carlo (MLMC) methods [24, 7, 18, 23] to reduce the overall cost of subset simulation. We refer to [29, 41] for different variants of MLMC estimators of rare event probabilities. In principle, our approach is similar to the approach presented in [39]. However, in contrast to that work and more closely related to the ideas in MLMC methods, we design the sequence of nested events in (1.1) such that most samples can be computed with inexpensive, coarse models. This reduces the total cost significantly compared to classical subset simulation, leading to more significant gains than the multilevel variant in [39].

Following the ideas in [22, 23], another key advance in our new method is the use of a posteriori error estimators to guarantee the critical subset property, which may be violated when changing the model resolution from one intermediate failure set to the next. It also allows a selective, sample-dependent choice of model resolution. Finally, these three advances lead to significant speed-up over classical subset simulation for our multilevel estimator.

In summary, this paper makes the following contributions: (i) The subset simulation method is formulated via shaking transformations, incorporating a selective mesh refinement strategy. The resulting algorithm is based on an MCMC method, where each sample involves an adaptive mesh resolution based on its limit state function value. As a result, high accuracy is only needed for samples with a rather small limit state function, whereas for most samples low resolution evaluations are sufficiently accurate when the state is far away from failure. (ii) Under certain assumptions, it is shown that the proposed selective refinement strategy does not violate the subset property. Moreover, a detailed complexity analysis quantifies the gains due to the selective refinement strategy. (iii) A novel, adaptive multilevel subset simulation method is proposed where the accuracy increases over the defined subsets. Through a selective refinement strategy and appropriately chosen intermediate threshold values, the failure sets satisfy the critical subset property. Our complexity analysis shows significant improvement through the proposed adaptive multilevel subset choice and through the additional application of selective mesh refinement.

This paper is organized as follows. In section 2, we formulate the problem and define hierarchies of discretizations via two abstract, sample-wise assumptions. In section 3, we explore in detail the estimation of rare event probabilities, first using standard MC and classical subset simulation, before proposing two improved estimators based on selective refinement and adaptive multilevel subset selection. Section 4 contains a concrete implementation of the algorithms via shaking transformations, as well as an asymptotic convergence result. The complexity analysis of the novel approaches is provided in section 5, while section 6 demonstrates their performance on a series of numerical experiments, starting with two toy problems and finishing with a Darcy flow model. Finally, section 7 offers some conclusions.

2. Problem formulation and model hierarchies. We consider a (linear or nonlinear) model $\mathcal{M}$ on an infinite-dimensional function space V, e.g., a PDE, which is subject to uncertainty, or an SDE. The solution is modelled as a random field on the probability space $(\Omega, \Sigma, \mathbb{P})$ with values in V. For any $\omega \in \Omega$, we denote by $u(\omega) \in V$ the solution of

$$(2.1) \qquad\qquad \mathcal{M}(\omega, u(\omega)) = 0.$$

Given a quantity of interest $\Phi : V \to \mathbb{R}$, i.e., a functional of the model solution u, we are interested in computing the probability that a so-called *rare event* occurs. We let $G = G(\omega) := \Phi^* - \Phi(u(\omega))$ be the associated limit state function, which is negative when the quantity of interest exceeds a critical value Φ^*. Thus, we want to compute the probability that $\omega \in \Omega$ is in the failure set

$$F := \{\omega \in \Omega : G(\omega) \leq 0\}.$$

For simplicity, we will assume that G is a real valued random variable with probability density function (pdf) (with respect to Lebesgue measure) $\pi : \mathbb{R} \to \mathbb{R}_+$, which is assumed to be unknown. If $\mathbb{1}_F$ denotes the indicator function of F, i.e., $\mathbb{1}_F(\omega) = 1$ if $\omega \in F$ and $\mathbb{1}_F(\omega) = 0$ otherwise, then the failure probability can also be expressed as the integral

$$(2.2) \qquad P = \mathbb{P}(F) = \int_\Omega \mathbb{1}_F(\omega)\,\mathbb{P}(\mathrm{d}\omega) = \int_\mathbb{R} \mathbb{1}_{(-\infty,0]}(x)\pi(x)\,\mathrm{d}x = \int_{-\infty}^0 \pi(x)\,\mathrm{d}x$$

which, for very small $P \ll 1$, will be classified as a rare event. Note that the integral in (2.2) is equivalent to the expected value $\mathbb{E}[\mathbb{1}_F]$. We are interested in applications where not only the dimension of V, but also the dimension of the underlying sample space Ω is high (or infinite), e.g., in subsurface flow simulations where the permeability is described by a spatially correlated random field. To estimate P, both V and the high-dimensional integral in (2.2) need to be approximated in practice.

We return to the approximation of the above integral in section 3 and finish this section by formulating some abstract assumptions on the numerical approximation of the model $\mathcal{M}$ and of the limit state function G, for any given $\omega \in \Omega$. To this end, we introduce a hierarchy of numerical approximations to G with increasing accuracy, namely G_ℓ, for $\ell = 1, \ldots, L$.

Assumption 2.1.
(a) We assume that the cdf of G is Lipschitz continuous, i.e., there exists a constant $c_{\mathrm{Lip}} > 0$ such that for any $x, y \in \mathbb{R}$

$$(2.3) \qquad |\mathbb{P}(G \leq y) - \mathbb{P}(G \leq x)| \leq c_{\mathrm{Lip}}|y - x|.$$

(b) Let $\gamma \in (0,1)$ and $q \geq 0$. Then, for all $\ell \in \mathbb{N}$, we assume that

$$(2.4) \qquad |G(\omega) - G_\ell(\omega)| \leq \gamma^\ell, \quad \mathbb{P}\text{-a.s. in } \omega,$$

and that there exists a constant $c_0 > 0$ such that

$$(2.5) \qquad \mathbb{E}[\mathcal{C}\,[G_\ell]] \leq c_0 \gamma^{-\ell q}.$$

Note that to estimate $\mathbb{P}(F)$, high accuracy is only needed for samples $\omega \in \Omega$ where the limit state function G is small. Thus, given a family of random variables $(G_\ell, \ell \in \mathbb{N})$ satisfying Assumption 2.1, the following *selective refinement* strategy was introduced in [22, 23], in fact for the more general task of approximating $\mathbb{P}(G \leq y)$ for any $y \in \mathbb{R}$.

It has been shown in [23, Lems. 5.1 and 5.2] that the construction in Algorithm 2.1 leads to a (y-dependent) random variable $G_\ell^y : \Omega \to \mathbb{R}$ which satisfies the following assumption.

Algorithm 2.1 Selective refinement strategy introduced in [22, 23].

1: Let $y \in \mathbb{R}$, $\ell \in \mathbb{N}$, and $\omega \in \Omega$.
2: Compute $G_1(\omega)$ and initialize $k = 1$ and $G_\ell^y(\omega) = G_1(\omega)$.
3: **while** $k < \ell$ and $|G_k(\omega) - y| < \gamma^k$ **do**
4:     refine the accuracy by setting $k \leftarrow k + 1$ and compute $G_k(\omega)$.
5:     Set $G_\ell^y(\omega) = G_k(\omega)$.
6: **end while**

Assumption 2.2. Suppose that Assumption 2.1 (a) holds, and let $\gamma \in (0, 1)$ and $q \geq 0$. Then, for fixed $y \in \mathbb{R}$ and for all $\ell \in \mathbb{N}$, we assume that

$$(2.6) \qquad |G(\omega) - G_\ell^y(\omega)| \leq \max\left(\gamma^\ell, |G_\ell^y(\omega) - y|\right), \quad \mathbb{P}\text{-a.s. in } \omega,$$

and that there exists a constant $c_0 > 0$ such that

$$\mathbb{E}[\mathcal{C}\left[G_\ell^y\right]] \leq c_0\left(1 + \gamma^{(1-q)\ell}\right).$$

It was also shown in the proof of [23, Lem. 5.2] that under Assumption 2.2 the bias for approximating $\mathbb{P}(G \leq y)$ remains of order γ^ℓ, i.e., there exists a constant $c_y > 0$ independent of γ and ℓ such that

$$\left|\mathbb{P}(G_\ell^y \leq y) - \mathbb{P}(G \leq y)\right| \leq c_y \gamma^\ell.$$

In order to verify that G_ℓ^y indeed is a random variable on $(\Omega, \Sigma, \mathbb{P})$, we introduce (as in [20]) the stopping time $\omega \mapsto \tau_y(\omega) := \inf\{k \in \{1, \ldots, \ell\} \mid |G_k(\omega) - y| \geq \gamma^\ell\}$ with respect to the natural filtration $\mathcal{F}_k = \sigma(G_s, s \leq k)$ and define $G_\ell^y := G_{\ell \wedge \tau_y}$ as the final iteration of the stopped discrete time stochastic process $(G_{k \wedge \tau_y}, k = 1, \ldots, \ell)$ which is measurable with respect to $\mathcal{F}_L \subset \Sigma$, since all G_k are assumed to be $\Sigma/\mathcal{B}(R)$-measurable.

Note that Assumption 2.2 holds naturally for the choice $y = \infty$, since any finite random variable G_ℓ satisfying Assumption 2.1 also satisfies (2.6) and $\mathbb{1}_{G_\ell \leq \infty} = 1$ almost surely.

Remark 2.3. For the complexity analysis of subset simulation and of the standard MC method, we will consider a fixed $y \in \mathbb{R}$ and thus suppress the dependence of G_ℓ^y on y. Under Assumption 2.2 we then refer to the random variable G_ℓ^y satisfying (2.6) simply by G_ℓ.

In a PDE setting, we typically have approximations $\tilde{G}_h$ of G associated with some numerical discretization method and error bounds of the type

$$(2.7) \qquad |G(\omega) - \tilde{G}_h(\omega)| \leq C(\omega)h^\alpha,$$

where h is a discretization parameter, e.g., mesh size, and α is a convergence rate. The constant $C(\omega)$ may depend on ω. For SDEs or PDEs with random coefficients, α and $C(\omega)$ can be estimated using adjoint methods or hierarchical error estimators; see, e.g., [14, 25, 33].

For sufficiently well behaved models $\mathcal{M}$ and sufficiently smooth functionals Φ, there exists a constant $C_{\max} < \infty$ such that $C(\omega) \leq C_{\max}$ $\mathbb{P}$-almost surely in ω. Thus, considering,

for example, numerical approximations of G obtained on a sequence of uniformly refined meshes $\mathcal{T}_\ell$, $\ell \in \mathbb{N}_0$, with $h_\ell = h_0 2^{-\ell}$ and $\mathcal{T}_0$ some fixed coarsest mesh, the bound in (2.4) can be achieved for every $\ell \in \mathbb{N}$ by choosing $G_\ell = \tilde{G}_{h_\ell}$ and $\gamma = 2^{-\alpha}$, provided $h_0 \leq C_{\max}^{-1/\alpha}$. The constant q in (2.5) depends on the (physical) dimension of the problem, the order of accuracy of the underlying numerical method, and the choice of deterministic solver. The constant c_0 depends on the distribution of the constant $C(\omega)$. A similar choice is possible in the case of finite element methods with locally adaptive mesh refinement driven by an a posteriori error estimator. In that case, the bound in (2.7) can be replaced by a sharper bound in terms of the number of mesh elements; cf. [8].

Finally, we mention that both Assumption 2.1 and Assumption 2.2 are formulated as $\mathbb{P}$-almost sure conditions, and we are aware of the fact that in practical scenarios this may be very restrictive. It is of great interest to relax this assumption and to allow numerical error bounds that hold only with high probability or in an L^p-sense instead of $\mathbb{P}$-almost surely. However, this relaxation would lead to nontrivial changes in the complexity analysis presented in section 5. Therefore, we will leave this extension for future work.

3. Adaptive subset simulation and selective refinement. In this section, we explore in detail the estimation of the failure probability P given by (2.2), and propose two new extensions of the classical subset simulation approach.

Let $\widehat{P}$ denote an estimator of P. There are several sources of error in this estimator. These include the following: systematic error due to the choice of mathematical model, numerical error due to model approximation, and statistical error due to finite sampling size. Here, we will assume the model is exact and will only consider how to control the numerical and statistical errors. We measure the quality of the estimator $\widehat{P}$ by using the relative root mean squared error (rRMSE)

$$
(3.1) \qquad \delta(\widehat{P}) = \frac{1}{P}\sqrt{\mathbb{E}[(\widehat{P} - P)^2]},
$$

and we let $\mathrm{TOL} > 0$ be the desired accuracy for $\delta(\widehat{P})$. In subset simulation, the rRMSE is often referred to as the coefficient of variation (c.o.v.); see, e.g., [5].

3.1. Estimating rare event probabilities by standard Monte Carlo. The most basic way of estimating the probability in (2.2) to an accuracy TOL is to use a standard MC method where all samples are computed with a numerical approximation to accuracy $\mathcal{O}(\mathrm{TOL})$. Let $F^L := \{\omega \in \Omega : G_L(\omega) \leq 0\}$ be the approximate failure set on a fixed numerical discretisation level L, and consider the standard MC estimator

$$
(3.2) \qquad \widehat{P}^{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{F^L}(\omega_i) = \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{(-\infty,0]}(G_L^{(i)}),
$$

where $\omega_i \in \Omega$ are independent and identically distributed (i.i.d.) samples and each $G_L^{(i)} = G_L(\omega_i)$ is computed to accuracy γ^L. This is an unbiased estimator for

$$
(3.3) \qquad P_L = \mathbb{P}\left(F^L\right) = \int_\Omega \mathbb{1}_{F^L}(\omega)\mathbb{P}(\mathrm{d}\omega) = \mathbb{E}\left[\widehat{P}^{\mathrm{MC}}\right],
$$

and we can expand

$$
\begin{aligned}
\delta(\widehat{P}^{\mathrm{MC}})^2 &= \frac{\mathbb{E}\left[(\widehat{P}^{\mathrm{MC}})^2\right] - 2\mathbb{E}\left[\widehat{P}^{\mathrm{MC}}\right]P + P^2}{P^2} \\
&= \frac{\mathbb{E}\left[(\widehat{P}^{\mathrm{MC}})^2\right] - \mathbb{E}\left[\widehat{P}^{\mathrm{MC}}\right]^2 + \mathbb{E}\left[\widehat{P}^{\mathrm{MC}}\right]^2 - 2\mathbb{E}\left[\widehat{P}^{\mathrm{MC}}\right]P + P^2}{P^2} \\
&= \left(\frac{\mathbb{E}\left[\widehat{P}^{\mathrm{MC}} - P\right]}{P}\right)^2 + \frac{\mathbb{V}\left[\widehat{P}^{\mathrm{MC}}\right]}{P_L^2}\frac{P_L^2}{P^2} = \left(1 - \frac{P_L}{P}\right)^2 + \frac{1 - P_L}{N P_L}\frac{P_L^2}{P^2},
\end{aligned}
\tag{3.4}
$$

which includes a bias error due to the numerical approximation $G_L \approx G$. The expectation and the variance of $\widehat{P}^{\mathrm{MC}}$ in (3.4) are with respect to the joint distribution of the samples $G_L^{(i)}$, in the i.i.d. case here, the N-fold product measure of the distribution of G_L. As shown in [23, sect. 3], it follows from Assumption 2.2 that there exists a constant $c_1 > 0$ such that

$$
\left(1 - \frac{P_L}{P}\right)^2 \leq c_1^2 \gamma^{2L}.
$$

To ensure that the first term on the right hand side of (3.4) is less than $\mathrm{TOL}^2/2$ it suffices that

$$
L \geq \log\left(\sqrt{2}c_1 \mathrm{TOL}^{-1}\right).
\tag{3.5}
$$

We will assume this throughout this paper.

We note that the constant c_1 does, in general, depend on the underlying distribution of G and G_L, in particular, on the rareness of the event and on the gradient of G near the boundary of the failure domain. However, L only grows logarithmically with c_1 and this contribution to the rRMSE is the same for all methods considered. We refer to [42] for a detailed analysis of approximation errors for rare event probabilities in the context of PDE based models.

The main challenge in achieving the required accuracy for $\delta(\widehat{P}^{\mathrm{MC}})$, is to ensure that the second term on the right hand side of (3.4) is sufficiently small such that

$$
\frac{1 - P_L}{N P_L}\frac{P_L^2}{P^2} \leq \frac{\mathrm{TOL}^2}{2}.
\tag{3.6}
$$

A sufficient condition for this to hold is

$$
N \propto \mathrm{TOL}^{-2} P_L^{-1} (P_L/P)^2.
\tag{3.7}
$$

Hence, the number of samples needs to be proportional to the inverse of the rare event probability. For realistic applications, where the cost $\mathcal{C}[G_L] \gg 1$ and $P_L \approx P \ll 1$, this is completely infeasible. Under Assumption 2.1 and choosing L as in (3.5), the total cost of the standard MC estimator would be

$$
\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\mathrm{MC}}]\right] = N\mathbb{E}\left[\mathcal{C}[G_L]\right] \leq Nc_0\gamma^{-Lq} \leq c_2 \mathrm{TOL}^{-(2+q)} P_L^{-1}\frac{P_L^2}{P^2},
\tag{3.8}
$$

for some constant $c_2 > 0$ which is independent of TOL. It is possible to improve this through importance sampling techniques [37, 38, 12], but that requires some a priori knowledge of the distribution of G_L, which we typically do not have. Further note that the factor P_L/P is close to 1, since L was chosen large enough such that $P_L/P \in (1 - \mathrm{TOL}/\sqrt{2}, 1 + \mathrm{TOL}/\sqrt{2})$.

3.2. Subset simulation. In engineering applications, subset simulation [5, 6, 34] is one of the most widespread variance reduction techniques to design an efficient estimator for P in (2.2). It has been successfully used in many different contexts and for different applications, which include engineering reliability analysis [4], robust design [2], topology optimization [35], multiobjective optimization [43], Bayesian inference [21], and model calibration through history matching [27].

The main idea is to define a sequence of nested failure sets that contain the target failure set F, as in (1.1). This is accomplished using a sequence of intermediate failure thresholds

$$(3.9) \qquad\qquad 0 = y_K < y_{K-1} < \ldots < y_0 = \infty.$$

That way, each intermediate failure set is defined as

$$F_j := \{\omega \in \Omega : G(\omega) \le y_j\} = \{G \in (-\infty, y_j]\}, \quad \text{for} \quad j = 1, \ldots, K.$$

The failure set F and some intermediate failure sets that contain it are illustrated in Figure 3.1 via level curves of $G(\omega)$. As stated in (1.2), the rare event probability $\mathbb{P}(F)$ can be expressed as a product of conditional probabilities, i.e., $\mathbb{P}(F) = \prod_{j=1}^{K} \mathbb{P}(F_j|F_{j-1})$, where each $\mathbb{P}(F_j|F_{j-1})$ is by construction larger than $\mathbb{P}(F)$ and thus less rare—significantly so, if K is sufficiently large and the y_j are chosen appropriately.

To estimate $\mathbb{P}(F)$ from the product of conditional probabilities we need to compute

$$(3.10) \qquad \mathbb{P}(F_j|F_{j-1}) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty, y_j]}(x)\pi_{j-1}(x)\,\mathrm{d}x, \quad \text{for} \quad j = 1, \ldots, K,$$

where π_{j-1} is the pdf of G conditioned on the event $G \le y_{j-1}$, i.e.,

$$(3.11) \qquad \pi_{j-1}(x) = \frac{\pi(x)\mathbb{1}_{(-\infty, y_{j-1}]}(x)}{Z} \quad \text{with} \quad Z := \int_{-\infty}^{y_{j-1}} \pi(x)\,\mathrm{d}x.$$
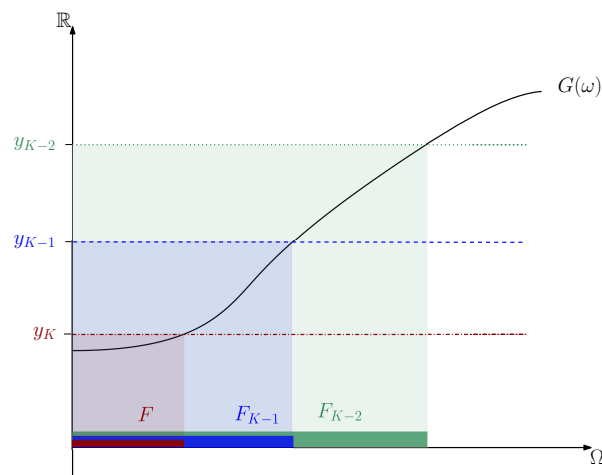

Figure 3.1. *A cartoon of the limit state function $\omega \mapsto G(\omega)$ and the resulting failure sets.*

The standard MC estimator for the integral in (3.10) is given by

$$(3.12) \qquad \mathbb{P}(F_j | F_{j-1}) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{(-\infty, y_j]}(G^{(i)}), \quad G^{(i)} \sim \pi_{j-1}.$$

In general, we cannot generate i.i.d. samples from π_{j-1} directly, at least not efficiently. In order to circumvent this, MCMC methods are commonly employed; see section 4 below.

3.3. Subset simulation with selective refinement. In practice we also need to take into account the numerical approximations of the failure domains. Instead of having a fixed computational mesh for all samples, which is the typical approach in the literature, we follow [23] and use a sample-dependent approximation $G_L(\omega)$ that guarantees instead that the error in the limit state function satisfies either the bound (2.4) in Assumption 2.1 or the weaker bound (2.6) in Assumption 2.2.

Firstl, in the case of Assumption 2.1, we use the sequence of intermediate failure sets

$$(3.13) \qquad F_j^L := \{\omega : G_L(\omega) \leq y_j\}, \quad \text{for} \ \ \infty = y_0 > \ldots > y_{K-1} > y_K = 0,$$

which obviously fulfill the critical subset property

$$(3.14) \qquad F^L = F_K^L \subset F_{K-1}^L \subset \cdots \subset F_1^L \subset F_0^L = \Omega.$$

In the case of the selective refinement strategy, i.e., under Assumption 2.2, the following sequence of intermediate failure sets is chosen:

$$(3.15) \qquad F_j^L := \{\omega : G_L^{y_j}(\omega) \leq y_j\}, \quad \text{for} \ \ \infty = y_0 > \ldots > y_{K-1} > y_K = 0.$$

The subset property can be guaranteed again if the failure thresholds are sufficiently far apart.

Lemma 3.1. *Let $j \in \{1, \ldots, K\}$. Suppose that the random variable G_L satisfies Assumption 2.2 and that $y_{j-1} - y_j \geq 2\gamma^L$. Then*

$$F_j^L = \{\omega : G_L^{y_j}(\omega) \leq y_j\} \subset \{\omega : G_L^{y_{j-1}}(\omega) \leq y_{j-1}\} = F_{j-1}^L.$$

Proof. First, note that by convention $F_0 = \Omega$ in which case the subset property is always satisfied. Next, we assume that $j \geq 2$ and we will give a proof by contradiction. Fix $\omega \in \Omega$ such that $G_L^{y_j}(\omega) \leq y_j$ and $G_L^{y_{j-1}}(\omega) > y_{j-1}$.
Case 1: Assume that $|G_L^{y_j}(\omega) - G(\omega)| > \gamma^L$, i.e.,

$$|G_L^{y_j}(\omega) - G(\omega)| \leq \max(|G_L^{y_j}(\omega) - y_j|, \gamma^L) = y_j - G_L^{y_j}(\omega).$$

If $|G_L^{y_{j-1}}(\omega) - G(\omega)| \leq \gamma^L$, then

$$\begin{aligned} G_L^{y_{j-1}}(\omega) = G_L^{y_{j-1}}(\omega) - G(\omega) + G(\omega) &\leq \gamma^L + G(\omega) \\ &= \gamma^L + G(\omega) - G_L^{y_j}(\omega) + G_L^{y_j}(\omega) \\ &\leq \gamma^L + |G(\omega) - G_L^{y_j}(\omega)| + G_L^{y_j}(\omega) \\ &\leq \gamma^L + y_j - G_L^{y_j}(\omega) + G_L^{y_j}(\omega) \\ &= \gamma^L + y_j \leq y_{j-1}, \end{aligned}$$

which contradicts $G_L^{y_{j-1}}(\omega) > y_{j-1}$. On the other hand, if $|G_L^{y_{j-1}}(\omega) - G(\omega)| > \gamma^L$, then

$$|G_L^{y_{j-1}}(\omega) - G(\omega)| \le \max(|G_L^{y_{j-1}}(\omega) - y_{j-1}|, \gamma^L) = G_L^{y_{j-1}}(\omega) - y_{j-1},$$

and therefore,

$$\begin{aligned}
G_L^{y_{j-1}}(\omega) = G_L^{y_{j-1}}(\omega) - G(\omega) + G(\omega) &\le |G_L^{y_{j-1}}(\omega) - G(\omega)| + G(\omega) \\
&\le G_L^{y_{j-1}}(\omega) - y_{j-1} + G(\omega) \\
&= G_L^{y_{j-1}}(\omega) - y_{j-1} + G(\omega) - G_L^{y_j}(\omega) + G_L^{y_j}(\omega) \\
&\le G_L^{y_{j-1}}(\omega) - y_{j-1} + y_j < G_L^{y_{j-1}}(\omega),
\end{aligned}$$

which is a contradiction in itself.

Case 2: Now, assume that $|G_L^{y_j}(\omega) - G(\omega)| \le \gamma^L$, i.e.,

$$|G_L^{y_j}(\omega) - G(\omega)| \le \max(|G_L^{y_j}(\omega) - y_j|, \gamma^L) = \gamma^L.$$

Now, if $|G_L^{y_{j-1}}(\omega) - G(\omega)| \le \gamma^L$, then

$$G_L^{y_{j-1}}(\omega) = G_L^{y_{j-1}}(\omega) - G(\omega) + G(\omega) - G_L^{y_j}(\omega) + G_L^{y_j}(\omega) \le 2\gamma^L + y_j \le y_{j-1},$$

which contradicts $G_L^{y_{j-1}}(\omega) > y_{j-1}$. If again $|G_L^{y_{j-1}}(\omega) - G(\omega)| > \gamma^L$, then

$$\begin{aligned}
G_L^{y_{j-1}}(\omega) = G_L^{y_{j-1}}(\omega) - G(\omega) + G(\omega) - G_L^{y_j}(\omega) + G_L^{y_j}(\omega) \\
\le G_L^{y_{j-1}}(\omega) - y_{j-1} + \gamma^L + y_j < G_L^{y_{j-1}}(\omega),
\end{aligned}$$

due to the assumed upper bound on $y_j - y_{j-1}$, which is again a contradiction in itself. ∎

Hence, in both cases the numerical approximation of the rare event probability $P_L = \mathbb{P}(F^L)$ on level L can be written as a product of intermediate failure set probabilities as

$$(3.16) \qquad P_L = \prod_{j=1}^{K} \mathbb{P}(F_j^L \mid F_{j-1}^L).$$

Finally, given estimators $\widehat{P}_j$ for $\mathbb{P}(F_j^L \mid F_{j-1}^L)$, we define the subset simulation estimator as

$$(3.17) \qquad \widehat{P}^{\text{SuS}} = \prod_{j=1}^{K} \widehat{P}_j.$$

In general, this is a biased estimator for P_L, but it can be shown that it is asymptotically unbiased [5]. We will return to this in section 4.

As stated already in Remark 2.3, for simplicity we will omit the dependence on y_j and write in the following $G_\ell(\omega)$ instead of $G_\ell^{y_j}(\omega)$, also in the case of selective refinement when using Assumption 2.2.

3.4. Adaptive multilevel subset simulation. We will now go one step further and consider the sequence of failure sets

$$F_j^{\text{ML}} = \{\omega : G_{\ell_j}(\omega) \le y_j\}, \quad j = 1, \dots, K,$$

where each G_{ℓ_j} is computed only to tolerance γ^{ℓ_j} with $\ell_j \leq L$, and y_j and ℓ_j are adaptively chosen. Typically, we assume $\ell_j \geq \ell_{j-1}$ and $\ell_K = L$. We consider both the cases without and with selective refinement, i.e., $G_{\ell_j}(\omega)$ or $G^{y_j}_{\ell_j}(\omega)$, respectively.

Such a multilevel strategy was also at the heart of [39], but there the thresholds y_j were chosen to roughly balance the contributions from the individual subsets, i.e., $\widehat{P}_1 \approx \widehat{P}_2 \approx \ldots \approx \widehat{P}_K$, as in classical subset simulation. This reduces the number of expensive fine resolution samples only by a linear factor $\mathcal{O}(K)$. Furthermore, without any further conditions the crucial subset property $F_j^{\mathrm{ML}} \subset F_{j-1}^{\mathrm{ML}}$ cannot be guaranteed. Thus, one focus of [39] was to estimate the correction factor that arises due to the loss of the subset property.

To circumvent this problem, in the following lemma we propose a method whereby the thresholds y_j are chosen adaptively, to maintain the subset property. Crucially, we exploit here the samplewise error bounds in Assumptions 2.1 and 2.2.

Lemma 3.2. *Consider the sequence F_j^{ML}, $j = 1, \ldots, K$, where either Assumption 2.1 or Assumption 2.2 is satisfied. Let $y_K = 0$ and $y_0 = \infty$, and choose*

$$y_j = y_{j+1} + (\gamma^{\ell_j} + \gamma^{\ell_{j+1}}) > 0, \quad for \;\; j = K-1, \ldots, 1.$$

Then, the subset property $F_{j+1}^{ML} \subset F_j^{ML}$ holds.

Proof. First, suppose that Assumption 2.1 holds. Then

$$(3.18) \qquad |G_{\ell_j}(\omega) - G_{\ell_{j+1}}(\omega)| \leq |G(\omega) - G_{\ell_j}(\omega)| + |G(\omega) - G_{\ell_{j+1}}(\omega)| \leq (\gamma^{\ell_j} + \gamma^{\ell_{j+1}}),$$

and hence, if $G_{\ell_{j+1}}(\omega) \leq y_{j+1}$, then

$$(3.19) \qquad G_{\ell_j}(\omega) \leq G_{\ell_{j+1}}(\omega) + (\gamma^{\ell_j} + \gamma^{\ell_{j+1}}) \leq y_{j+1} + (\gamma^{\ell_j} + \gamma^{\ell_{j+1}}) = y_j,$$

which concludes the proof for Assumption 2.1.

The proof for Assumption 2.2 then follows directly since $y_j - y_{j-1} \geq 2\gamma^{\ell_j}$. ∎

For simplicity, we assume that $K = L$ and that $\ell_j = j$ for all $j = 1, \ldots, K$. The rare event probability can then be written as

$$(3.20) \qquad \mathbb{P}(F_L^{\mathrm{ML}}) = \prod_{\ell=1}^{L} \mathbb{P}(F_\ell^{\mathrm{ML}} \mid F_{\ell-1}^{\mathrm{ML}}),$$

where $F_\ell^{\mathrm{ML}} = \{\omega : G_\ell(\omega) \leq y_\ell\}$. However, this does not preclude us from using more than L subsets. If the first intermediate failure set F_1^{ML} is still a rare event, we can estimate $\mathbb{P}(F_1^{\mathrm{ML}}|F_0^{\mathrm{ML}}) = \mathbb{P}(F_1^{\mathrm{ML}})$ using classical subset simulation with an additional K_1 subsets instead of plain MC, but with all evaluations of the limit state function on those additional K_1 subsets only computed to an accuracy of γ^1 in Assumptions 2.1 and 2.2. This may also be necessary for intermediate failure probabilities $P(F_j|F_{j-1})$ if they happen to be very small.

Let $\widehat{P}_1$ be an estimator for $\mathbb{P}(F_1^{\mathrm{ML}}) = \mathbb{P}(F_1^{\mathrm{ML}}|F_0^{\mathrm{ML}})$ using standard MC or classical subset simulation on discretization level $\ell = 1$. For $\ell > 1$ we assume as above that we are given estimators $\widehat{P}_\ell$ for $\mathbb{P}\left(F_\ell^{\mathrm{ML}} \mid F_{\ell-1}^{\mathrm{ML}}\right)$ that will be constructed by MCMC sampling in the following. We define the multilevel subset simulation estimator by

$$(3.21) \qquad \widehat{P}^{\mathrm{ML}} = \widehat{P}_1 \prod_{\ell=2}^{L} \widehat{P}_\ell.$$

In practice, we propose applying classical subset simulation for the estimation of $\widehat{P}_1$.

   An illustration of the idea behind Lemma 3.2 is given in Figure 3.2. An additional benefit of the choice of thresholds in Lemma 3.2 is that the difference in the intermediate failure thresholds y_ℓ shrinks geometrically as ℓ increases. See Figure 3.3 for an illustration. Thus, by continuity of $\mathbb{P}$ it follows that $\mathbb{P}(F_\ell) \to \mathbb{P}(F)$, and therefore, also $\mathbb{P}(F_\ell^{\mathrm{ML}} \mid F_{\ell-1}^{\mathrm{ML}}) \to 1$ as $\ell \to \infty$. This effect reduces the variance on the latter subsets that have to be computed to higher accuracy. As a consequence, significantly fewer samples have to be computed at high accuracy reducing the overall complexity of the estimator dramatically. We will return to this point in section 5.

   The three considered algorithms, namely subset simulation, subset simulation with selective refinement, and multilevel subset simulation, are summarized in Appendix A. To estimate the conditional probabilities in (3.16) and (3.20) an MCMC algorithm is used, which will be described in the following section. The different types of subset simulation only differ in the construction of the considered subsets F_j. We have further included an adaptive choice for N_ℓ in Algorithm A.2 in order to verify the required tolerance with respect to the rRMSE $\delta(\widehat{P})$. Sufficient bounds on $\delta(\widehat{P})$ are derived in section 5.1. A similar choice can also be included in Algorithm A.1.
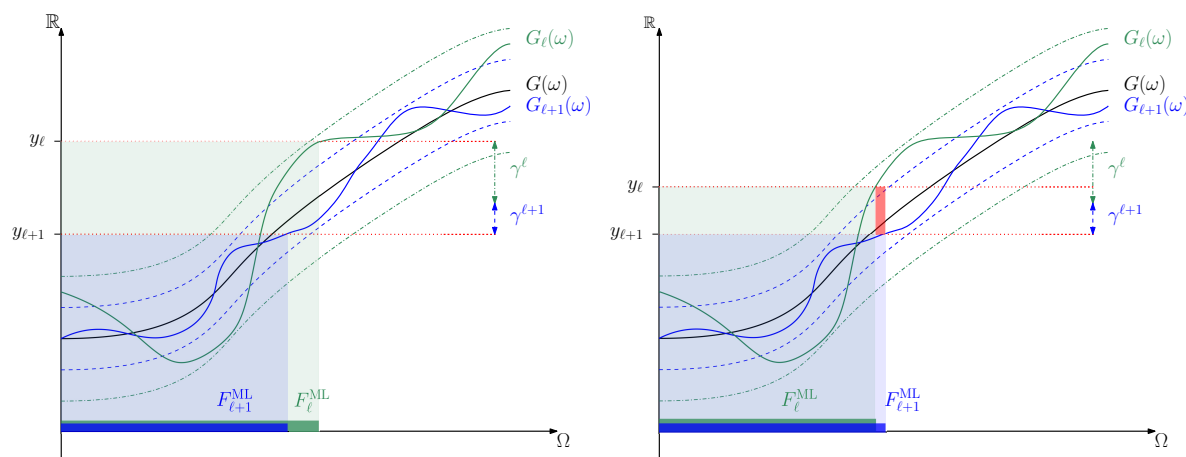

Figure 3.2. *Illustration of the choice of failure thresholds in Lemma 3.2, which guarantee the subset property $F_{\ell+1}^{ML} \subset F_\ell^{ML}$. It holds, when $y_\ell - y_{\ell+1} = \gamma^{\ell+1} + \gamma^\ell$ and $|G_\ell(\omega) - G_{\ell+1}(\omega)| \le \gamma^{\ell+1} + \gamma^\ell$, as chosen in Lemma 3.2 (left plot). If $y_\ell - y_{\ell+1} < \gamma^{\ell+1} + \gamma^\ell$, then Lemma 3.2 does not apply. In particular, for too small choices of $y_\ell - y_{\ell+1}$ we may find $\omega \in \Omega$ such that $G_{\ell+1}(\omega) \le y_{\ell+1}$, but $G_\ell(\omega) > y_\ell$, as illustrated by the red zone (in the right plot).*
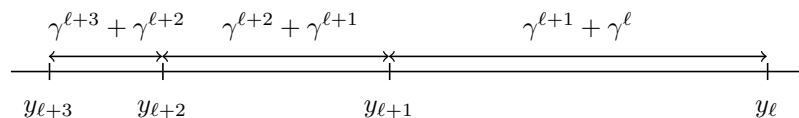

Figure 3.3. *Illustration of the shrinkage of the subsets F_ℓ^{ML} with increasing ℓ.*

4. Shaking transformations and asymptotic convergence. From the above discussion we see that a crucial component of any subset simulation is an efficient estimator for the conditional probabilities $\mathbb{P}(F_j|F_{j-1})$, for $j > 1$. In other words, we need to generate samples from the pdf given in (3.11). Due to the small probability of failure, standard MC sampling is infeasible. Instead, several authors have developed MCMC algorithms [5, 34, 3]. Here, we use the general, parallel, one-path (POP) algorithm based on shaking transformations introduced in [26] and further developed in [1]. The key idea is to build one Markov chain, exploring the space through shaking transformations and moving from the coarsest subset to the finest via a sequence of rejection operators.

The shaking transformation of a random variable X with respect to a random variable W, acting on measurable spaces $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{W}, \mathcal{W})$, respectively, can be defined as a measurable mapping $S : \mathbb{X} \times \mathbb{W} \to \mathbb{X}$ that satisfies

$$(4.1) \qquad\qquad (X, S(X, W)) = (S(X, W), X)$$

in distribution. Assuming that the rare event and the corresponding intermediate subsets can be written as events of X, i.e.,

$$F = \{X \in A\}, \quad F_j = \{X \in A_j\}, \quad \text{such that} \quad A, A_j \in \mathcal{X}, \quad j = 1, \dots, K,$$

the shaking transformation will act as the proposal in the constructed MCMC algorithm. In order to force the Markov chain to explore the subset F_j, we define the rejection operator

$$M_j^S : \mathbb{X} \times \mathbb{W} \to \mathbb{X}, \quad \text{with} \quad M_j^S(x, w) = S(x, w)\mathbb{1}_{S(x,w) \in A_j} + x\mathbb{1}_{S(x,w) \notin A_j}.$$

The proposed MCMC algorithm in the earlier multilevel subset simulation paper [39] is based on a preconditioned Crank–Nicholson (pCN) [32, 19] proposal which can, in fact, be written as a shaking transformation and hence the theoretical results in [26] can be applied. The property of being a shaking transformation is related to the detailed balance condition. This observation has been made and verified already in [1, Thm. 8]. The general POP algorithm is given in Algorithm 4.1.

Assuming that the resulting Markov chains $(X_{j,i})_{i \geq 0}$ are π_j-irreducible and Harris recurrent under a small set condition, the POP algorithm converges in the sense that for every j there exists a constant $C_j > 0$ such that

$$\mathbb{E}\left[|\widehat{P}_j - \mathbb{P}(X \in A_{j+1} \mid X \in A_j)|^2\right] \leq \frac{C_j}{N_j}.$$

In [1] the authors include a model for X via Gaussian transformations which has also been considered in [34]. We consider X modelled through a measurable transformation T as

$$X(\omega) = T(\Theta(\omega)), \ \omega \in \Omega, \ T : \mathbb{R}^d \to \mathbb{X} \text{ measurable},$$

where $\Theta \sim \mathcal{N}(0, I_{\mathbb{R}^d})$. The underlying rare event can then be formulated as

$$A = \{\theta \in \mathbb{R}^d \mid \varphi(\theta, \bar{y}) \leq 0\} \subset \mathcal{B}(\mathbb{R}^d), \quad \varphi : \mathbb{R}^d \times \mathbb{R} \cup \{\infty\} \to \mathbb{R}$$

Algorithm 4.1 Parallel One-Path (POP) algorithm.

1: Given a seed $X_{0,0} \sim X$
2: **for** $j = 0, \ldots, K-1$ **do**
3:     **for** $i = 0, \ldots, N_{j+1} - 1$ **do**
4:         Generate $W_{j,i} \sim W$,
5:         Shake and accept/reject $X_{j,i+1} = M_j^S(X_{j,i}, W_{j,i})$
6:     **end for**
7:     Estimate the probability of subset F_{j+1} by

$$\widehat{P}_{j+1} = \frac{1}{N_{j+1}} \sum_{k=0}^{N_{j+1}-1} \mathbb{1}_{A_{j+1}}(X_{j,k}),$$

8:     $i_j = \arg\min\{k \mid X_{j,k} \in A_{j+1}\}$
9:     Define initial state for next level $X_{j+1,0} = X_{j,i_j}$
10: **end for**
11: **Result:** $\widehat{P} = \prod_{j=1}^{K} \widehat{P}_j$

---

where φ is nonincreasing in the second component in the sense that $\varphi(\theta, y) \geq \varphi(\theta, y')$, for any $\theta \in \mathbb{R}^d$ and $y \geq y'$, and the convention $\varphi(\theta, \infty) := -\infty$ is assumed. Further, we assume that φ is measurable in the first component. The nested subsets are built through a sequence of level parameters $\bar{y} := y_K < \ldots < y_k < \ldots < y_0 \leq \infty$ by

(4.2) $$A_j = \{\theta \in \mathbb{R}^d \mid \varphi(\theta, y_j) \leq 0\} \in \mathcal{B}(\mathbb{R}^d),$$

such that

$$A := A_K \subset \cdots \subset A_j \subset \cdots \subset A_0 := \mathbb{R}^d.$$

In fact, the special case $\varphi(\theta, y) := T(\theta) - y$ for a measurable model functional $T : \mathbb{R}^d \to \mathbb{R}$ connects the present approach to the setting considered in the previous sections. The sequence of subsets F_j considered above can then be written as $F_j = \{\Theta \in A_j\} \in \Sigma$.

Consider the shaking transformation

(4.3) $$S_\eta(\theta, w) = \sqrt{1 - \eta^2}\,\theta + \eta w,$$

which obviously satisfies (4.1), and let W be an independent copy of Θ. The shaking transformation (with rejection) is now defined as

(4.4) $$M_j^{S_\eta} : \begin{cases} \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d, \\ (\theta, w) \mapsto S_\eta(\theta, w)\mathbb{1}_{S_\eta(\theta,w) \in A_j} + x\mathbb{1}_{S_\eta(\theta,w) \notin A_j}. \end{cases}$$

We note that this is, in fact, equivalent to using the pCN approach, which is in detailed balance with the prior. This has the advantage that the acceptance rate does not decrease with the dimension, which is the case, e.g., for traditional random walk proposals. The POP algorithm is then of very similar form as one of the algorithms proposed in [34] to generate samples from the conditional pdf π_{j-1}. For completeness, details are provided in Algorithm 4.2. Under certain assumptions geometric convergence of the resulting estimator can be verified.

Algorithm 4.2 POP algorithm based on Gaussian transformation.

1: Given a seed $\Theta_{0,0} \sim \Theta$ and a correlation parameter $\eta \in [0,1]$
2: **for** $j = 0, \ldots, K-1$ **do**
3:     **for** $i = 0, \ldots, N_{j+1}-1$ **do**
4:         Generate $W_{j,i} \sim \Theta$,
5:         Shake and accept/reject $\Theta_{j,i+1} = M_j^{S_\eta}(\Theta_{j,i}, W_{j,i})$
6:     **end for**
7:     Estimate the probability of subsets F_{j+1} by

$$\widehat{P}_{j+1} = \frac{1}{N_{j+1}} \sum_{k=0}^{N_{j+1}-1} \mathbb{1}_{A_{j+1}}(\Theta_{j,k}),$$

8:     $i_j = \arg\min\{k \mid \Theta_{j,k} \in A_{j+1}\}$
9:     Define initial state for next level $\Theta_{j+1,0} = \Theta_{j,i_j}$
10: **end for**
11: **Result:** $\widehat{P} = \prod_{j=1}^{K} \widehat{P}_j$

Theorem 4.1 (see [1, Thm. 8]). *Let $j \in \{0, \ldots, K-1\}$ be fixed and consider the Markov chain $(\Theta_{j,i})_{i \geq 0}$ resulting from Algorithm* 4.2. *Assume that*

(4.5)
$$\sup_{\theta \in A_j} \mathbb{P}(S(\theta, W) \notin A_j) = \delta_1 < 1,$$

and let $c \geq 2$ and $q \geq 2$. Define $V(\theta) = \exp(c \sum_{k=1}^{d} \theta_k)$ and assume that the initial condition $\Theta_{j,0}$ is independent of the future evolution of the Markov chain and such that $\mathbb{E}\left[V(\Theta_{j,0})\right] < +\infty$. Then, there exists $C > 0$ and a geometric rate $r \in (0,1)$ such that for any measurable function $g : \mathbb{R}^d \to \mathbb{R}$ with

$$\sup_{\theta \in \mathbb{R}^d} \frac{|g(\theta)|}{V^{1/q}(\theta)} < +\infty$$

it holds true that

$$\mathbb{E}\left[\left|\frac{1}{N} \sum_{i=1}^{N} g(\Theta_{j,i}) - \mathbb{E}[g(\Theta) \mid \Theta \in A_j]\right|^q\right] \leq CN_j^{-q/2},$$

$$and \qquad |\mathbb{E}[g(\Theta_{j,N}) - \mathbb{E}[g(\Theta) \mid \Theta \in A_j]| \leq Cr^n.$$

In this setting, the estimator $\widehat{P}_j$ of $\mathbb{P}(F_j|F_{j-1})$ is then given by

(4.6)
$$\widehat{P}_j = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{\Theta_{j-1,i} \in A_j\}} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{(-\infty,0]}(\varphi(\Theta_{j-1,i}, y_j)).$$

Note that by using a measurable transformation T_ℓ such that $G_\ell(\omega) = T_\ell(\Theta(\omega))$, the limit state function can also be defined pointwise in y_ℓ, i.e., $\varphi(\theta, y_\ell) := T_\ell(\theta) - y_\ell$. In that case,

the subset property follows from Lemma 3.2 and Theorem 4.1 can be applied to the adaptive multilevel subset simulation algorithm as well.

A brief discussion of the assumptions is in order. To ensure condition (4.5) the acceptance rate of the algorithm needs to be uniformly bounded from below. If

$$\sup_{\theta \in A_j} \mathbb{P}(S(\theta, W) \notin A_j) = 1,$$

the Markov chain may get stuck in some points. The condition is problem dependent and, in general, difficult to verify. However, this is a common problem with MCMC algorithms.

Since the sequences of samples $\{\Theta_{j-1,i}\}_{i=1}^N$ are dependent, we use convergence diagnostics to estimate the autocorrelation factor in the sequence. A common measure of the dependence in the sequence is the autocorrelation factor. Following [5], we use multiple chains to compute an estimate of the autocorrelation factor ϕ_j. The total number of samples N to achieve a certain accuracy of the estimator in (4.6) is again determined by the rRMSE in (3.1):

$$\delta(\widehat{P}_j) = \frac{1}{\mathbb{P}(F_j|F_{j-1})} \sqrt{\mathbb{E}\left[(\widehat{P}_j - \mathbb{P}(F_j|F_{j-1}))^2\right]}.$$

Expanding the square of the rRMSE, we can estimate

$$(4.7) \qquad \delta(\widehat{P}_j)^2 = \frac{1 - \mathbb{E}\left[\widehat{P}_j\right]}{N\mathbb{P}(F_j|F_{j-1})} = \frac{1 - \mathbb{P}(F_j|F_{j-1})}{N\mathbb{P}(F_j|F_{j-1})}(1 + \phi_j),$$

with autocorrelation factor $\phi_j = 0$ for i.i.d. samples and $\phi_j > 0$ for dependent samples. We note that for each j the expectation in (4.7) is with respect to the joint distribution of the $\Theta_{j-1,i}$, $i = 1, \ldots, N$.

5. Complexity analysis. We will now bound the complexity of the new estimators.

5.1. Complexity of subset simulation with selective refinement. Analogous to (3.4), the square $\delta(\widehat{P}^{\mathrm{SuS}})^2$ of the rRMSE for the subset simulation estimator can be expanded into the sum of a bias term—identical to the one in (3.4) for standard MC—and the relative variance $\mathbb{V}[\widehat{P}^{\mathrm{SuS}}]/P_L^2$. The bias term can again be made less than $\mathrm{TOL}^2/2$ by choosing L as before in (3.5). Moreover, applying results derived in [5] it turns out that

$$(5.1) \qquad \mathbb{E}\left[\frac{(\widehat{P}^{\mathrm{SuS}} - P_L)^2}{P_L^2}\right] \leq \frac{\mathrm{TOL}^2}{2} \quad \text{if} \quad \delta(\widehat{P}_j)^2 = \mathbb{E}\left[\frac{(\widehat{P}_j - P_j)^2}{P_j^2}\right] \propto K^{-s} \mathrm{TOL}^2,$$

where $s = 1$ if the estimators $\widehat{P}_j$, $j = 1, \ldots, K$, are uncorrelated and $s = 2$ if they are fully correlated. To simplify notation, the biased densities defined by F_{j-1}^L instead of F_{j-1} in (3.11) are again denoted by π_{j-1} suppressing the dependence on L.

Let us now estimate the cost for the particular case where $\widehat{P}_j$ is chosen to be the MCMC estimator for $\mathbb{P}(F_j \mid F_{j-1})$ described in section 4 with N_j samples. As it has been shown in [5] the rRMSE can be computed as

$$(5.2) \qquad \delta(\widehat{P}_j)^2 = \frac{1 - \mathbb{P}(F_j \mid F_{j-1})}{N_j \mathbb{P}(F_j \mid F_{j-1})}(1 + \phi_j),$$

where ϕ_j is the autocorrelation factor of the Markov chain produced by Algorithm 4.2. We note that (5.2) crucially depends on the fact that the underlying Markov chain generated through the MCMC algorithm is ergodic. The ergodicity of the Markov chain can be obtained from Theorem 4.1.

In the following complexity analysis, we assume that (5.2) is satisfied. Moreover, we assume an adaptive selection of the failure sets has been applied, e.g., as described in [5], where the values of y_j are chosen in the course of the algorithm, such that $\widehat{P}_j \approx p_0$, for all $j = 1, \ldots, K$. It was shown in [44] that the best performance is achieved for values of the constant $p_0 \in [0.1, 0.3]$. For simplicity, we assume (without loss of generality) that $\mathbb{P}(F_j \mid F_{j-1}) = p_0$ for all $j = 1, \ldots, K$.

Theorem 5.1. *Suppose the following:*
1. *Assumption* 2.1 *is satisfied for some* $q \geq 0$ *and* $\gamma \in (0, 1)$;
2. *the subsets are chosen such that* $\mathbb{P}(F_j \mid F_{j-1}) = p_0 := P_L^{1/K}$, *for all* $j = 1, \ldots, K - 1$; *and*
3. *there exists a constant* $\phi < \infty$ *such that* $\phi_j \leq \phi$ *in* (5.2) *for all* $j = 1, \ldots, K$.

Then, for any $TOL > 0$ *the maximum discretisation level* L *and the numbers of samples* $\{N_j\}$ *can be chosen such that*

$$\mathbb{E}\left[\frac{|\widehat{P}^{\mathrm{SuS}} - P|^2}{P^2}\right] \leq TOL^2$$

with a cost that is bounded by

$$\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\mathrm{SuS}}]\right] \leq c_4 \frac{P_L^2}{P^2} TOL^{-(2+q)}\left(\log\left(P_L^{-1}\right)\right)^{s+1}$$

for some constant $c_4 > 0$. *Here,* $s = 1$ *if the estimators* $\widehat{P}_j$, $j = 1, \ldots, K$, *are uncorrelated and* $s = 2$ *if they are fully correlated.*

If, in addition, Assumption 2.2 *holds and the conditions* $y_{j-1} - y_j \geq 2\gamma^L$ *are satisfied for all* $j = 1, \ldots K$, *then the cost can be bounded by*

$$\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\mathrm{SuS}}]\right] \leq c_4 \frac{P_L^2}{P^2} TOL^{-\max(2, 1+q)}\left(\log\left(P_L^{-1}\right)\right)^{s+1}.$$

Proof. We first split

$$(5.3) \qquad \mathbb{E}\left[\frac{|\widehat{P}^{\mathrm{SuS}} - P|^2}{P^2}\right] = \left(\frac{\mathbb{E}[\widehat{P}^{\mathrm{SuS}} - P]}{P}\right)^2 + \frac{P_L^2}{P^2}\mathbb{E}\left[\frac{(\widehat{P}^{\mathrm{SuS}} - P_L)^2}{P_L^2}\right],$$

where the first term can be bounded by

$$(5.4) \qquad \left(\frac{\mathbb{E}[\widehat{P}^{\mathrm{SuS}} - P]}{P}\right)^2 \leq 2\left(\frac{P_L - P}{P}\right)^2 + 2\frac{P_L^2}{P^2}\left(\frac{\mathbb{E}[\widehat{P}^{\mathrm{SuS}} - P_L]}{P_L}\right)^2.$$

To bound the bias error we choose $L \propto \log(TOL^{-1})$ as in (3.5). Due to the assumptions of the theorem we can choose $N_j = N$ uniformly across all subsets. The expressions

$$A := \left(\frac{\mathbb{E}[\widehat{P}^{\mathrm{SuS}} - P_L]}{P_L}\right)^2 \quad \text{and} \quad B := \mathbb{E}\left[\frac{(\widehat{P}^{\mathrm{SuS}} - P_L)^2}{P_L^2}\right]$$

have been analyzed in [5], where it has been shown that

$$A \leq \sum_{i,j=1,\ j>i}^{K} \delta(\widehat{P}_i)\delta(\widehat{P}_j) + o\left(\frac{1}{N}\right) \quad \text{and} \quad B \leq \sum_{i,j=1}^{K} \delta(\widehat{P}_i)\delta(\widehat{P}_j) + o\left(\frac{1}{N}\right)$$

if the estimators $\{\widehat{P}_j\}$ are correlated, whereas

$$A \leq \sum_{j=1}^{K} \delta(\widehat{P}_j)^2 \quad \text{and} \quad B \leq \sum_{j=1}^{K} \delta(\widehat{P}_j)^2$$

if the estimators $\{\widehat{P}_j\}$ are uncorrelated. If N is now chosen such that

$$(5.5) \qquad \delta(\widehat{P}_j)^2 \leq c_3 K^{-s} \frac{\text{TOL}^2}{5} \frac{P^2}{P_L^2},$$

it follows that

$$(5.6) \qquad A \leq c_3 \frac{\text{TOL}^2}{5} \frac{P^2}{P_L^2} \quad \text{and} \quad B \leq c_3 \frac{\text{TOL}^2}{5} \frac{P^2}{P_L^2}.$$

Taking into account the choice of L in (3.5) and combining (5.3), (5.4), and (5.6), finally leads to

$$\mathbb{E}\left[\frac{|\widehat{P}^{\text{SuS}} - P|^2}{P^2}\right] \leq \text{TOL}^2.$$

Now, using the third assumption of the theorem, i.e., that $\phi_j \leq \phi$, a sufficient condition on the number of samples to ensure that the bound in (5.5) holds is

$$N \propto \frac{P_L^2}{P^2} K^s \text{TOL}^{-2},$$

where the proportionality constant depends on p_0, ϕ, and c_3, but is independent of TOL and K. Then, recalling that $P_L = p_0^K$ and thus $K \propto \log\left(P_L^{-1}\right)$ it follows that

$$N \propto \left(\log\left(P_L^{-1}\right)\right)^{s+1} \frac{P_L^2}{P^2} \text{TOL}^{-2}.$$

Hence, combining this with the assumed bound on the cost per sample in Assumption 2.1 and using the fact that $L \propto \log(\text{TOL}^{-1})$ implies $\gamma^{-qL} \propto \text{TOL}^q$, the total cost to compute the subset simulation estimator can be bounded by

$$\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\text{SuS}}]\right] \leq KN c_0 \gamma^{-Lq} \leq c_4 \frac{P_L^2}{P^2} \text{TOL}^{-(2+q)} \left(\log\left(P_L^{-1}\right)\right)^{s+1},$$

for some constant $c_4 > 0$. Under the selective refinement Assumption 2.2 and for $q > 1$, the computational cost can be bounded by

$$\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\text{SuS}}]\right] \leq KN c_0 (1 + \gamma^{(1-q)L}) \leq c_4 \frac{P_L^2}{P^2} \text{TOL}^{-\max(2,1+q)} \left(\log\left(P_L^{-1}\right)\right)^{s+1}. \qquad \blacksquare$$

Remark 5.2. We note that following Lemma 3.1, the assumption $y_{j-1} - y_j \geq 2\gamma^L$, for $j = 1, \ldots, K$, is needed to ensure the subset property in the case of Assumption 2.2 (i.e., for subsets defined through $F_j = \{\omega \in \Omega : G_L^{y_j}(\omega) \leq y_j\}$). Although this condition may fail in practical scenarios, it is easy to check on the fly and then to apply the selective refinement strategy only on subsets where it is satisfied. Alternatively, since the choice of p_0 is somewhat arbitrary, it could also be chosen level-dependent so that the condition $y_{j-1} - y_j \geq 2\gamma^L$ is satisfied on all subsets (at the expense of potentially increasing the number of subsets K). The presented complexity analysis in Theorem 5.1 can then be viewed as the best-case scenario. In the worst case, the complexity is the same as the cost of classical subset simulation.

5.2. Complexity of adaptive multilevel subset simulation. We now turn our attention to the complexity of the adaptive multilevel subset estimator defined in (3.21). We will make use of the following convergence property in the analysis below, assuming that we can control the ratio of the subset probabilities sufficiently well.

Lemma 5.3. *Suppose Assumption 2.1 is satisfied and y_ℓ is defined as in Lemma 3.2. Furthermore, we assume that*

$$\text{(5.7)} \qquad \frac{\mathbb{P}(G \leq y)}{\mathbb{P}(G \leq x)} \leq 1 + \tilde{c}|y - x| \log(P^{-1}),$$

for $y > x > 0$ and $\tilde{c} > 0$ independent of P. Then the difference between the intermediate probabilities can be bounded in the following way:

$$\frac{1 - \mathbb{P}(F_\ell^{ML} \mid F_{\ell-1}^{ML})}{\mathbb{P}(F_\ell^{ML} \mid F_{\ell-1}^{ML})} = \frac{1 - \mathbb{P}(G_\ell \leq y_l \mid G_{\ell-1} \leq y_{\ell-1})}{\mathbb{P}(G_\ell \leq y_l \mid G_{\ell-1} \leq y_{\ell-1})} \leq c_5 \gamma^\ell \log(P^{-1}).$$

Proof. Applying the subset property we obtain

$$\frac{1 - \mathbb{P}(G_\ell \leq y_l \mid G_{\ell-1} \leq y_{\ell-1})}{\mathbb{P}(G_\ell \leq y_\ell \mid G_{\ell-1} \leq y_{\ell-1})} = \frac{\mathbb{P}(G_{\ell-1} \leq y_{\ell-1})}{\mathbb{P}(G_\ell \leq y_\ell)} - 1 = \frac{\mathbb{P}(G_{\ell-1} \leq y_{\ell-1}) - \mathbb{P}(G_\ell \leq y_\ell)}{\mathbb{P}(G_\ell \leq y_\ell)}.$$

First, note that by Assumption 2.1 we have $G_\ell \leq G + \gamma^\ell$ and $G - \gamma^{\ell-1} \leq G_{\ell-1}$, and thus

$$\mathbb{P}(G_\ell \leq y_\ell) \geq \mathbb{P}(G + \gamma^\ell \leq y_\ell) = \mathbb{P}(G \leq y_\ell - \gamma^\ell),$$
$$\mathbb{P}(G_{\ell-1} \leq y_{\ell-1}) \leq \mathbb{P}(G - \gamma^{\ell-1} \leq y_{\ell-1}) = \mathbb{P}(G \leq y_{\ell-1} + \gamma^{\ell-1}).$$

By definition we have $y_{\ell-1} + \gamma^\ell = y_\ell + (\gamma^\ell + \gamma^{\ell-1}) + \gamma^\ell$, which together with (5.7) implies

$$0 \leq \frac{\mathbb{P}(G_{\ell-1} \leq y_{\ell-1}) - \mathbb{P}(G_\ell \leq y_\ell)}{\mathbb{P}(G_\ell \leq y_\ell)} \leq \frac{\mathbb{P}(G \leq y_{\ell-1} + \gamma^{\ell-1}) - \mathbb{P}(G \leq y_\ell - \gamma^\ell)}{\mathbb{P}(G \leq y_\ell - \gamma^\ell)}$$
$$\leq \tilde{c}(2\gamma^\ell + 2\gamma^{\ell-1}) \log(P^{-1}),$$

where we have used that $\mathbb{P}(G_{\ell-1} \leq y_{\ell-1}) \geq \mathbb{P}(G_\ell \leq y_\ell)$ due to Lemma 3.2. ∎

We note that the dependence on $\log(P^{-1})$ in (5.7) is the weakest assumption we can take in order to obtain an improvement through our proposed multilevel subset simulation strategy. However, the bound (5.7) crucially depends on the underlying limit state function G and one might drop the dependence on $\log(P^{-1})$ for certain models.

As in the classical subset simulation, we estimate the failure sets such that

$$(5.8) \qquad \delta(\widehat{P}_\ell)^2 = \frac{1 - \mathbb{P}\left(F_\ell^{\mathrm{ML}} \mid F_{\ell-1}^{\mathrm{ML}}\right)}{\mathbb{P}\left(F_\ell^{\mathrm{ML}} \mid F_{\ell-1}^{\mathrm{ML}}\right)}(1 + \phi_\ell)N_\ell^{-1} \leq L^{-s}\mathrm{TOL}^2,$$

which guarantees that the relative variance of the multilevel estimator is bounded by TOL^2.

We finish by stating and proving the main theoretical result of the paper on the complexity of the proposed adaptive multilevel subset simulation method, with and without selective refinement, under similar assumptions made for the single-level complexity result in Theorem 5.1. Recall that for simplicity we have set $K = L$ and $\ell = \ell_j = j$ for $j = 1, \ldots, K$.

Theorem 5.4. *Suppose that*
1. *Assumption 2.1 is satisfied for some $q \geq 0$ and $\gamma \in (0, 1)$ and that (5.7) holds;*
2. *the level thresholds y_ℓ are defined as in Lemma 3.2; and*
3. *there exists a constant $\phi < \infty$ such that $\phi_\ell \leq \phi$ in (5.8), for all $\ell = 1, \ldots, L$.*

Then, for any $TOL > 0$, the maximum discretization level L and the numbers $\{N_\ell\}$ of samples on each level can be chosen such that

$$\mathbb{E}\left[\frac{|\widehat{P}^{\mathrm{ML}} - P|^2}{P^2}\right] \leq TOL^2$$

with a cost that is bounded by

$$\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\mathrm{ML}}]\right] \leq \begin{cases} c_7\mathrm{TOL}^{-\max(2,(1+q))}(1 + \phi)\log(P^{-1}), & s = 1, \\ c_7\mathrm{TOL}^{-\max(2,(1+q))}\log(\mathrm{TOL}^{-1})(1 + \phi)\log(P^{-1}), & s = 2, \end{cases}$$

for some constant $c_7 > 0$. Here, $s = 1$ if the estimators $\widehat{P}_\ell$, $\ell = 1, \ldots, L$, are uncorrelated and $s = 2$ if they are fully correlated.

If, in addition, Assumption 2.2 is satisfied, the cost can be bounded by

$$\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\mathrm{ML}}]\right] \leq \begin{cases} c_7\mathrm{TOL}^{-\max(2,q)}(1 + \phi)\log(P^{-1}), & s = 1, \\ c_7\mathrm{TOL}^{-\max(2,q)}\log(\mathrm{TOL}^{-1})(1 + \phi)\log(P^{-1}), & s = 2. \end{cases}$$

Proof. The result follows similarly as the proof of Theorem 5.1 for single-level subset simulation. Note that the resulting number of samples

$$N_\ell \propto \frac{P_L^2}{P^2}\frac{1 - \mathbb{P}\left(F_\ell^{\mathrm{ML}} \mid F_{\ell-1}^{\mathrm{ML}}\right)}{\mathbb{P}\left(F_\ell^{\mathrm{ML}} \mid F_{\ell-1}^{\mathrm{ML}}\right)}(1 + \phi_\ell)L^s\mathrm{TOL}^{-2}$$

are now level dependent due to the fact that the probabilities $\mathbb{P}(F_\ell^{\mathrm{ML}} \mid F_{\ell-1}^{\mathrm{ML}})$ differ in ℓ. Hence, applying Lemma 5.3 the total computational costs result in

$$\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\mathrm{ML}}]\right] \leq \mathrm{TOL}^{-2}L^{s-1}(1 + \phi)\sum_{\ell=1}^{L}\mathbb{P}(F_\ell^{\mathrm{ML}}|F_{\ell-1}^{\mathrm{ML}})^{-1}(1 - \mathbb{P}(F_\ell^{\mathrm{ML}}|F_{\ell-1}^{\mathrm{ML}}))\gamma^{-\ell q},$$

$$\leq \mathrm{TOL}^{-2}L^{s-1}(1 + \phi)\sum_{l=1}^{L}c_5\gamma^{(1-q)\ell}\log(P^{-1})$$

$$\leq c_6\mathrm{TOL}^{-2}L^{s-1}(1 + \phi)(1 + \gamma^{(1-q)L})\log(P^{-1})$$

$$\leq c_7\mathrm{TOL}^{-\max(2,1+q)}L^{s-1}(1 + \phi)\log(P^{-1}),$$

for some $c_5, c_6, c_7 > 0$. If, in addition, Assumption 2.2 holds the bound can be improved to

$$\mathbb{E}\left[\mathcal{C}[\widehat{P}^{\mathrm{ML}}]\right] \leq c_7 \mathrm{TOL}^{-\max(2,q)} L^{s-1}(1+\phi)\log(P^{-1}). \qquad \blacksquare$$

Clearly, the asymptotic complexity is significantly improved over classical subset simulation. In addition to the gains due to the level-dependent cost for each sample and to the variance reduction on the rarer subsets, an additional cost reduction in practice comes from the fact that the accept/reject step for $\ell > 0$ in Algorithm 4.2 is computed to tolerance $\gamma^{\ell-1}$ and only accepted samples are then computed also to tolerance γ^ℓ. The intermediate failure thresholds and thus the failure sets F_ℓ^{ML} are defined a priori based on the value of γ. Thus, the probabilities $\mathbb{P}(F_\ell^{\mathrm{ML}}|F_{\ell-1}^{\mathrm{ML}})$ are problem dependent and—as in MLMC [24]—optimal sample sizes $\{N_\ell\}$ are difficult to compute. This would be an interesting area for future investigation.

6. Numerical results. In the following, we consider three numerical examples with increasing difficulty. We start the experiments with a one-dimensional toy example where we can ensure Assumption 2.2. In this example, it is possible to verify the complexity results expected from Theorems 5.1 and 5.4, respectively. In our second example we consider a rare event estimation problem based on a Brownian motion. In this case, Assumption 2.2 does not hold almost surely but only in L^p. The results of our numerical experiments remain promising and we observe a significant improvement through our multilevel subset simulation and the incorporation of selective refinement. The last experiment is based on an elliptic PDE model and represents a more realistic scenario of application.

6.1. Example 1: Toy experiment. We start by verifying our derived complexity results on a simplified toy model. We assume that $G \sim \mathcal{N}(0,1)$ and define the pointwise approximation $G_\ell(\omega) := G(\omega) + \kappa(\omega)\gamma^\ell$ with $\kappa \sim \mathcal{U}(\{-1,1\})$ such that (2.4) is obviously satisfied. Furthermore, we let $\gamma = 1/2$ and assume that $q = 2$ for the expected costs in (2.5).

The aim is to estimate the failure probability

$$(6.1) \qquad \mathbb{P}(G \leq -3.8) \approx 7.23 \cdot 10^{-5}.$$

Applying Algorithm 2.1 [23, Alg. 1] allows one to simulate G_ℓ^y satisfying Assumption 2.2. For any accuracy level ℓ, the selective refinement algorithm starts with the coarsest approximation $k = 1$ and successively refines the accuracy by increasing k until $|G_k(\omega) - y| \geq \gamma^k$ or $|G_k(\omega) - G(\omega)| \leq \gamma^\ell$ is satisfied. Note that in more realistic applications, estimates of the error $|G_k(\omega) - G(\omega)|$ are needed; cf. section 6.3 for more details. In this simplified example we increase the accuracy until $|G_k(\omega) - y| \geq \gamma^k$ or until $k = \ell$, since we know that $|G_\ell(\omega) - G(\omega)| \leq \gamma^\ell$ by definition. We then set $G_\ell^y(\omega) = G_k(\omega)$.

We compare standard MC, classical subset simulation with and without selective refinement, as well as our proposed adaptive multilevel subset simulation algorithm, choosing $L \propto \log(\mathrm{TOL}^{-1})$ in all cases, where TOL is the required tolerance. For subset simulation, we use $K = 5$ subsets such that all $\mathbb{P}(F_j \mid F_{j-1}) \approx [0.1, 0.2]$, where we choose the threshold values $(y_5, y_4, y_3, y_2, y_1, y_0) = (-3.8, -3.3, -2.8, -2, -1.3, \infty)$. In contrast, for multilevel subset simulation we choose y_ℓ following Lemma 3.2 and let the number of subsets increase depending
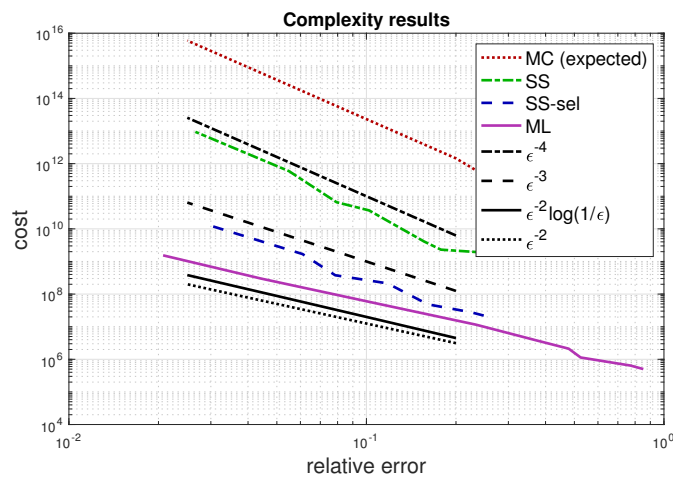
Figure 6.1. *Mean computational cost plotted against the estimated values of $\delta(\widehat{P})$ (with tolerance* TOL*) for the toy problem in Example 1 using standard MC (expected), classical subset simulation, as well as subset simulation and multilevel subset simulation with selective refinement.*

on the size of L. For the estimation of $\widehat{P}_1$ in (3.21) we apply a standard MC estimate. The highest accuracy level L and the numbers of samples on each subset (N_j and N_ℓ, resp.) are chosen according to the assumptions in Theorems 5.1 and 5.4, respectively.

Figure 6.1 shows the estimated rRMSE $\delta(\widehat{P})$ using the true reference probability of (6.1) plotted against the computational cost for the different applied estimators. We have used 100 runs for building the estimates of $\delta(\widehat{P})$ for each algorithm.

6.2. Example 2: Brownian motion. In our second numerical example, we consider estimating the probability that a standard Brownian motion drops below a threshold value within a certain time interval. To be more precise, let $(B_t)_{t\geq 0}$ be a Brownian motion. We are interested in the estimation of

$$\mathbb{P}(F) = \mathbb{P}(\min_{t\in[0,1]} B_t \leq -4),$$

where we can compute the reference value of $\mathbb{P}(F)$ by

$$\mathbb{P}(\min_{t\in[0,1]} B_t \leq -4) = 2 \cdot \mathbb{P}(B_1 \leq -4) \approx 6.3 \cdot 10^{-5}.$$

We define the limit state function pointwise by

$$G(\omega) := \min_{t\in[0,1]} B_t(\omega) + 4,$$

and introduce approximations of the limit state function such that

$$\tilde{G}_k(\omega) := \min_{t\in T_k} B_t(\omega) + 4, \quad T_k := \left\{ \frac{i}{2^k} \mid i = 0, \ldots, 2^k \right\}.$$

For all $1 \leq p < \infty$, the resulting approximation error can be bounded in L^p [36], i.e., there exists a constant $C_p > 0$ such that

$$(6.2) \qquad \|G - \tilde{G}_k\|_{L^p(\Omega)} = \mathbb{E}\left[|\min_{t\in[0,1]} B_t(\omega) - \min_{t\in T_k} B_t(\omega)|^p \right]^{1/p} \leq C_p 2^{-k/2}.$$
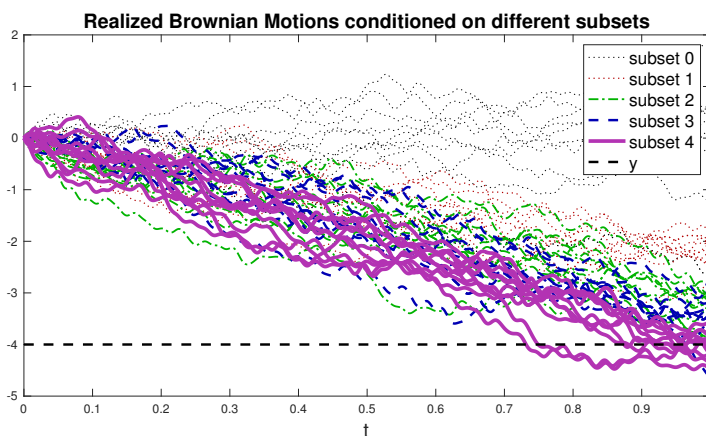
Figure 6.2. *Paths of the Brownian motion conditioned on the different subsets F_j.*

The Brownian motion is generated pathwise through a Karhunen–Loéve expansion

$$B_t(\omega) = \sum_{i=1}^{\infty} \xi_i(\omega)\varphi_i(t),$$

with $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{(i-1/2)^2})$ and $\varphi_i(t) = \frac{\sqrt{2}}{\pi}\sin((i-1/2)\pi t)$. See Figure 6.2 for various realizations of the Brownian motion, conditioned on the different chosen subsets, i.e., for different level thresholds y_j.

As in Example 1, on each accuracy level ℓ and for each sample ω, the selective refinement algorithm starts on the coarsest level $k = 1$ and successively refines the accuracy by increasing k until $|\tilde{G}_k(\omega) - y| \geq \gamma^k$ or $|\tilde{G}_k(\omega) - G(\omega)| \leq \gamma^\ell$. We then set $G_\ell^y(\omega) := \tilde{G}_k(\omega)$. Based on the error bound (6.2), we consider $\gamma = \frac{1}{\sqrt{2}}$ in our numerical experiments. Then $q = 2$ and the latter of the two conditions above is satisfied at least in an L^p-sense for $k = \ell + 2\log_2(C_p)$. Unfortunately, it is not possible to satisfy this condition in an almost sure sense.

Depending on the required tolerance value TOL we set again $L \propto \log(\text{TOL}^{-1})$ and fix the number of subsets in the classical subset simulation to $K_{\text{SL}} = 5$. In the multilevel formulation, $K_{\text{ML}} = \max(L-1, 8)$ subsets are considered and the threshold values y_j are chosen again according to Lemma 3.2, with the following values of ℓ_j for the failure sets $F_j = \{\omega : G_{\ell_j}^{y_j}(\omega) \leq y_j\}$:

j	1	2	3	4	5	...	$L-1$
ℓ_j	4	4	4	5	6	...	L

The MC estimates for the multilevel estimator ML are built using 100 paths, resulting in

$$\mathbb{E}[\widehat{P}^{\text{ML}}] \approx 5.11 \cdot 10^{-5} \quad \text{and} \quad \delta(\widehat{P}^{\text{ML}}) \approx 0.0534 \leq \text{TOL} =: 0.1 \,.$$

To compare the proposed multilevel method with classical subset simulation in Figure 6.3 we compare the resulting costs for various choices of TOL. Note that we have estimated the expected number of samples for the classical MC estimator as $N = (\text{TOL})^{-2}P^{-1}$.
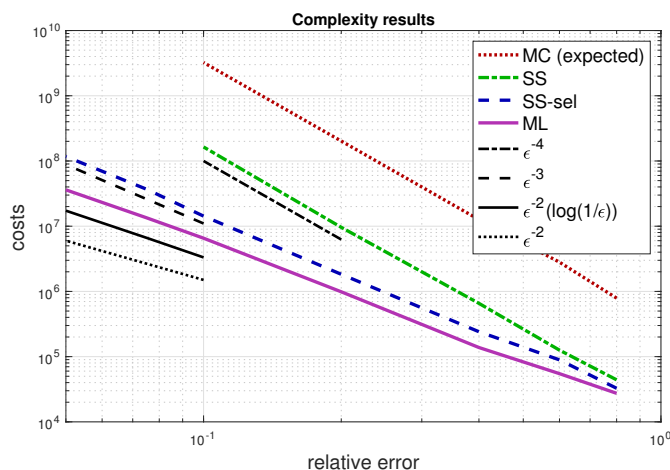
Figure 6.3. *Mean computational cost plotted against the estimated values of $\delta(\widehat{P})$ (with tolerance* TOL*) for the Brownian motion in Example* 2 *using standard MC (expected), classical subset simulation, as well as subset simulation and multilevel subset simulation with selective refinement.*

6.3. Example 3: Elliptic PDE. Finally, we consider the following diffusion equation, which is used, e.g., to model stationary Darcy flow:

$$-\nabla \cdot A\nabla u = 0 \quad \text{in } \mathcal{D},$$
(6.3)
$$\text{subject to} \quad u = 0 \quad \text{on } \Gamma_1, \quad u = 1 \quad \text{on } \Gamma_2, \quad \nu \cdot A\nabla u = 0 \quad \text{on } \Gamma_3 \cup \Gamma_4,$$

where $\mathcal{D} = (0,1) \times (0,1)$ is a unit square and Γ_1, Γ_2, Γ_3, Γ_4 are the left, right, upper, and lower boundaries, respectively. The permeability $A(x,\omega)$ is a log-normal random field. In particular, $\log(A(x,\omega))$ is a stationary, zero mean Gaussian field $\mathcal{N}(0,\mathcal{C})$ with covariance operator $\mathcal{C} = (-\Delta+\tau^2\cdot\text{id})^{-\alpha}$, where Δ denotes the Laplacian operator equipped with Neumann boundary conditions and we set $\tau = 0.1$ and $\alpha = 1$. The random field has been generated path-wise via a truncated Karhunen–Loéve expansion; see Figure 6.4 for two realizations of the random field and the associated PDE solutions.

The functional to define the limit state function is chosen to be

$$\mathcal{G}(u) = y - \frac{1}{|B|} \int_B u(x)\mathrm{d}x,$$
(6.4)

for $B = [0.4, 0.6] \times [0.9, 0.99] \subset \mathcal{D}$, i.e., "failure" occurs when the mean of u over the subdomain B exceeds y, with $|\cdot|$ denoting the area of B. For the numerical experiments below we choose $y = 0.92$.

Now, given a sample $\omega \in \Omega$ and defining

$$H_D^1(\mathcal{D}) = \{v \in H^1(\mathcal{D}) : v|_{\Gamma_1} = 0,\ v|_{\Gamma_2} = 1\} \quad \text{and} \quad H_0^1(\mathcal{D}) = \{v \in H^1(\mathcal{D}) : v|_{\Gamma_1\cup\Gamma_2} = 0\},$$

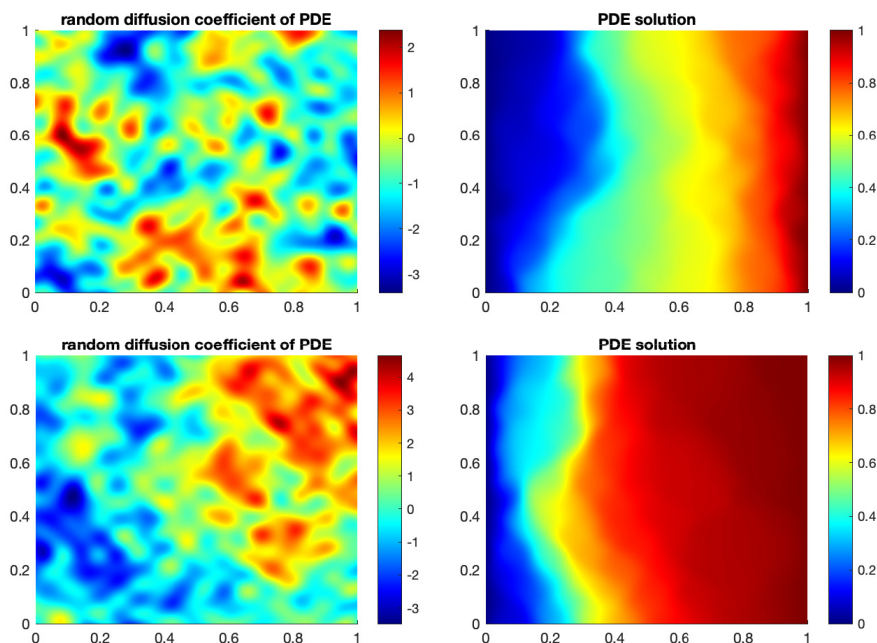the weak form of (6.3) is equivalent to finding $u \in H_D^1(\mathcal{D})$ such that

Figure 6.4. *Two realizations of the log-normal permeability field $A(x, \omega)$ (left) and the corresponding solution for the Darcy flow problem* (6.3) *(right) in Example* 3. *The first realization is sampled from the whole probability space, whereas the second one is a rare event realization.*

$$(6.5) \qquad a(\omega; u, v) = \int_{\mathcal{D}} A(x, \omega) \nabla u \cdot \nabla v \, dx = 0 \quad \text{for all } v \in H_0^1(\mathcal{D}).$$

The value of the exact limit state function at ω is set to be $G(\omega) = \mathcal{G}(u)$.

We approximate the weak formulation and the limit state function with a finite element (FE) method. Let $\mathcal{T}_h$ be a uniform triangulation of $\mathcal{D}$, and suppose $\mathcal{V}_h$ denotes the associated P_1-Lagrange FE space. The FE approximation of the solution u of (6.5) on $\mathcal{T}_h$ is then defined to be the $u_h \in \mathcal{V}_h \cap H_D^1(\mathcal{D})$ satisfying

$$(6.6) \qquad a(\omega; u_h, v_h) = 0 \quad \text{for all } v_h \in \mathcal{V}_h \cap H_0^1(\mathcal{D}).$$

The FE approximation of the limit state function is defined as $\tilde{G}_h(\omega) = \mathcal{G}(u_h)$.

Given $\gamma \in (0, 1)$, we then compute $G_\ell(\omega)$ by iterating over a sequence of uniformly refined meshes $\mathcal{T}_h$ with $h \to 0$, starting from some $h_0 > 0$, until $|G(\omega) - \tilde{G}_h(\omega)| \leq \gamma^\ell$. For selective refinement we start with accuracy $k = 1$ and reduce h until $|G(\omega) - \tilde{G}_h(\omega)| \leq \gamma^k$ or $|\tilde{G}_h(\omega) - y| \geq \gamma^k$. If the second condition is satisfied, we stop and set $G_\ell^y(\omega) = \tilde{G}_h(\omega)$. Otherwise, we increment $k \to k + 1$ and repeat the process until the second condition is satisfied for some k or until $k = \ell$.

To estimate the error in the limit state function $\tilde{G}_h$ computed on mesh $\mathcal{T}_h$ we use a hierarchical a posteriori error estimator. In particular, for the chosen random field and limit state function we may assume that

$$(6.7) \qquad ch^2 \leq |G(\omega) - \tilde{G}_h(\omega)| \leq Ch^2, \quad \text{for some } 0 < c \leq C < \infty.$$
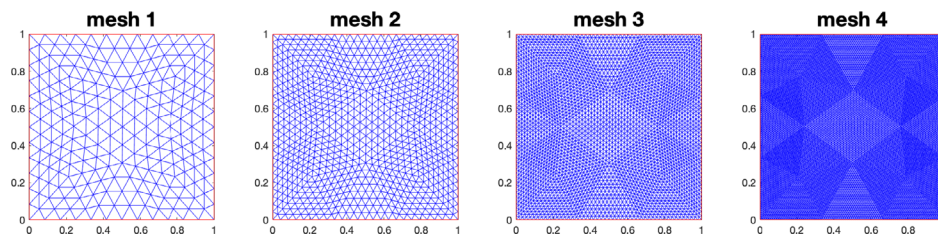
Figure 6.5. *Different levels of refinement of the FE mesh for the Darcy flow problem* (6.3) *in Example* 3.

Table 6.1

Chosen level (ℓ_j), mean probability (P_j), and mean acceptance rates (α_j) for each subset j after 100 runs of adaptive multilevel subset simulation (left) and single-level subset simulation (right), both with selective refinement.

j	1	2	3	4	5		j	1	2	3	4
ℓ_j	2	2	2	3	4		ℓ_j	4	4	4	4
P_j	0.05	0.13	0.17	0.38	0.83		P_j	0.13	0.1	0.16	0.07
α_j	—	0.45	0.32	0.22	0.18		α_j	—	0.5	0.38	0.29

Provided $r := C/c < 4$ it follows from those bounds and the triangle inequality that

$$|G(\omega) - \tilde{G}_h(\omega)| \leq \frac{1}{1 - r/4} |\tilde{G}_{h/2}(\omega) - \tilde{G}_h(\omega)| =: \eta_h .$$

The `PDE Toolbox` within `MATLAB` is used in our implementation. The FE meshes that were constructed are shown in Figure 6.5. For the numerical experiments we choose $\gamma = 1/4$ and use the sufficient condition $\eta_h \leq \gamma^\ell$ to bound the FE error on $\mathcal{T}_h$ (with $r = 2$). In the experiments, the estimate η_h of the FE error on the finest mesh depicted in Figure 6.5 was always below γ^4.

For the proposed single-level and multilevel estimators, we choose L, N_j, N_ℓ according to Theorems 5.1 and 5.4, with a fixed number of subsets $K = 4$ for the classical (single level) subset simulation. The number of subsets for our multilevel subset simulation on the highest evaluated accuracy, together with the corresponding accuracy levels ℓ_j, are presented in Table 6.1 (left). Moreover, the table shows the mean probability for each level and the mean acceptance rate for the Markov chains. The details for single-level subset simulation are presented in Table 6.1 (right).

After 100 runs of the multilevel estimator ML in this setting with $L = 4$, the estimated mean rare event probability and the relative variance are

$$(6.8) \qquad \mathbb{E}[\widehat{P}^{\mathrm{ML}}] \approx 1.8 \cdot 10^{-4} \quad \text{and} \quad \delta(\widehat{P}^{\mathrm{ML}}) \approx 0.172,$$

while for the classical subset simulation estimator SuS they are

$$(6.9) \qquad \mathbb{E}[\widehat{P}^{\mathrm{SuS}}] \approx 1.5 \cdot 10^{-4} \quad \text{and} \quad \delta(\widehat{P}^{\mathrm{SuS}}) \approx 0.192.$$

For each level ℓ in the multilevel subset simulation, the mean number of samples computed on each mesh level k are presented in Table 6.2 (left). In the selective refinement process,

Table 6.2

Mean number of samples computed on mesh level k for subset j, for adaptive multilevel subset simulation (left) and single-level subset simulation (right), both with selective refinement, estimated from 100 runs with $L = 4$ and summed up over all mesh levels. The tables also show the total number of samples and the (normalized) theoretical cost per sample c_k on each mesh level assuming $q = 3/2$.

$j \backslash k$	1	2	3	4	$j \backslash k$	1	2	3	4
1	3602	610.7	0	0	1	3602	1179.8	381.4	66.4
2	3602	1902.9	0	0	2	3602	2518.6	1272.1	247.6
3	3602	1564	0	0	3	3602	2373.3	1271.1	252.1
4	2002	1071.6	326.2	0	4	3602	2220.3	1066.8	181.8
5	502	281	108.5	15.2					
Total	13310	5430.2	434.7	15.2	Total	14404	8291.9	3991.4	747.9
c_k	1	8	64	512	c_k	1	8	64	512

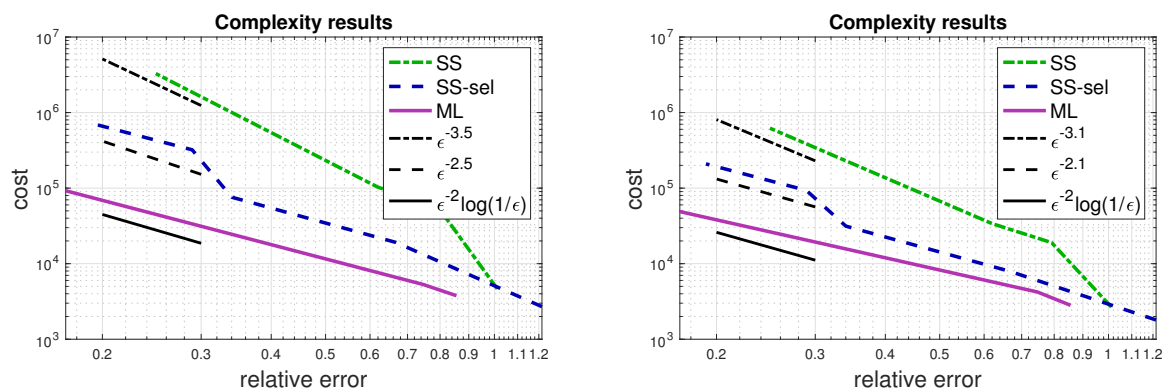**Figure 6.6.** *Mean computational cost plotted against the estimated values of $\delta(\widehat{P})$ for Example 3, the Darcy problem in (6.3), obtained by varying TOL for $q = 3/2$ (left) and $q = 1.1$ (right) using classical subset simulation, as well as subset simulation and adaptive multilevel subset simulation with selective refinement.*

the computation for each sample always starts on mesh level $k = 1$ and the numbers on the lower mesh levels are the cumulative ones. The vast majority of samples are generated from inexpensive low-resolution simulations, or put it another way, our multilevel estimator requires the equivalent cost of ~ 175 PDE solves on the finest accuracy level to estimate a rare event of $\mathcal{O}(10^{-4})$ with a relative variance of $\delta(\widehat{P}^{\mathrm{ML}}) \approx 0.17$. Since the chosen regime is not yet optimally balancing the cost, there is still potential for significant further gains.

The table also shows the normalized theoretical cost c_k for one sample on mesh level k (including the cost for the error estimation). Here, we have assumed a sparse direct solver such as CHOLMOD [17] (i.e., backslash in MATLAB) to solve each of the arising FE systems. Theoretically, the cost for a sparse direct solver like CHOLMOD applied to a two-dimensional FE system grows like $\mathcal{O}(n_{\mathrm{FE}}^{3/2})$, while the number of unknowns grows like $n_{\mathrm{FE}} = \mathcal{O}(4^k)$ under uniform refinement. Hence, we have chosen $q = 3/2$ in the definition of the normalized theoretical cost c_k. Based on this assumption, a comparison of the total normalized cost for all the estimators is presented in Figure 6.6 (left) for various choices of the tolerance TOL. The vastly superior performance of the proposed multilevel estimator is again apparent, e.g., for an estimate with $\delta(\widehat{P}) \approx 0.25$ there is a more than 10-fold gain in efficiency compared to

subset simulation with selective refinement and a more than 60-fold gain compared to standard subset simulation with all samples being computed to accuracy γ^L. For a fair comparison, standard subset simulation is also ran with the hierarchical error estimator to guarantee the required accuracy for each sample. Asymptotically, the cost does appear to grow as predicted in Theorem 5.4, even though (as in Example 2) Assumption 2.2 does not hold uniformly in ω here. Note that as predicted in Theorem 5.1, the selective refinement approach alone also leads to clear gains over standard subset simulation and to a better asymptotic rate.

In practice, the cost of CHOLMOD applied to a two-dimensional FE system is often observed to grow only like $\mathcal{O}(n_{\text{FE}}^{1.1})$ (at least in the preasymptotic regime). Thus, in Figure 6.6 (right) we also plot the gains for $q = 1.1$. Even in this case, we still see a 20-fold gain in efficiency of the proposed multilevel estimator compared to standard subset simulation. The gains for both the single- and the multilevel method with selective refinement are even more dramatic when $q > 3/2$, e.g., for three-dimensional problems or for rougher random coefficients.

7. Conclusions. In this paper, we propose a new multilevel subset simulation estimator of the probability of rare events. By constructing a hierarchy of numerical approximations to a model of a complex physical process, and by using a posteriori error estimation, the subset property of the intermediate failure domains is preserved. The estimator was tested in a Darcy flow problem, for which it reduced the cost compared to the classical subset simulation estimator by more than a factor of 60 for a practically relevant relative error tolerance of 25%. Given the wide applicability of subset simulation, several problems beyond the simulation of rare events may also benefit from this dramatic increase in efficiency.

Appendix A. Algorithmic details. We provide here some details on our versions of subset simulation, subset simulation with selective refinement, and the adaptive multilevel subset simulation algorithm based on shaking transformations and the POP Algorithm 4.2, as presented in section 4. The limit state function G, as well as the corresponding numerical approximations G_ℓ are modelled via measurable Gaussian transformations T and T_ℓ such that

$$G(\omega) = T(\Theta(\omega)) \quad \text{and} \quad G_\ell(\omega) = T_\ell(\Theta(\omega)),$$

where $T, T_\ell : \mathbb{R}^d \to \mathbb{R}$ are measurable mappings and $\Theta \sim \mathcal{N}(0, I_{\mathbb{R}^d})$. They only differ in the construction of the subsets $F_j = \{\Theta \in A_j\} \subset \Sigma$ and $A_j \subset \mathcal{B}(\mathbb{R}^d)$.

Subset simulation. In order to apply shaking transformations within subset simulation, the subsets are represented in the form of (4.2) with a fixed level of accuracy L and

$$F_j = \{\omega \in \Omega : G_L(\omega) \leq y_j\}, \quad \text{for} \quad j = 1, \dots, K.$$

We define $\varphi(\theta, y) = T_L(\theta) - y$ and apply the specific variant of POP based on a Gaussian transformation presented in Algorithm A.1.

Subset simulation with selective refinement. In order to incorporate the selective refinement strategy into subset simulation, we need to define the intermediate failure sets

$$F_j = \{\omega \in \Omega : G_L^{y_j}(\omega) \leq y_j\}, \quad \text{for} \quad j = 1, \dots, K.$$

Algorithm A.1 Subset simulation based on POP—with and without selective refinement.

1: Given an accuracy level L, a value for $\gamma \in (0,1)$, a correlation parameter $\eta \in [0,1]$, and a seed $\Theta_{0,0} \sim \Theta$, define subsets $F_j = \{\Theta \in A_j\}$ for all $j = 1, \dots, K$ such that

$$A_j = \begin{cases} \{\theta \in \mathbb{R}^d \mid T_L(\theta) \leq y_j\} & \text{for classical subset simulation,} \\ \{\theta \in \mathbb{R}^d \mid T_L^{y_j}(\theta) \leq y_j\} & \text{for selective refinement if } y_{j-1} - y_j \geq 2\gamma^L, \end{cases}$$

as well as a shaking transformation $S_\eta(\theta, w)$ and a rejection operator $M_j^{S_\eta}$ according to (4.3) and (4.4), respectively.

2: **for** $j = 0, \dots, K-1$ **do**

3:     **for** $i = 0, \dots, N_{j+1} - 1$ **do**

4:         Generate $W_{j,i} \sim \Theta$.

5:         Shake and accept/reject $\Theta_{j,i+1} = M_j^{S_\eta}(\Theta_{j,i}, W_{j,i})$.

6:     **end for**

7:     Estimate the probability of subsets F_{j+1} by $\widehat{P}_{j+1} = \frac{1}{N_{j+1}} \sum_{k=0}^{N_{j+1}-1} \mathbb{1}_{A_{j+1}}(\Theta_{j,k})$.

8:     Set $i_j = \arg\min\{k \mid \Theta_{j,k} \in A_{j+1}\}$.

9:     Define initial state for next level $\Theta_{j+1,0} = \Theta_{j,i_j}$.

10: **end for**

11: **Result:** $\widehat{P} = \prod_{j=1}^{K} \widehat{P}_j$.

Following Lemma 3.1, these sets satisfy the subset property provided $y_{j-1} - y_j \geq 2\gamma^L$. However, the classical strategy to choose the threshold values y_j such that $\mathbb{P}(F_j \mid F_{j-1}) \approx p_0$ for some constant value $p_0 \in [0.1, 0.3]$ will not always ensure $y_{j-1} - y_j \geq 2\gamma^L$. On the other hand, it is easy to check in practice on each of the failure sets whether it is satisfied in order to decide whether to apply selective refinement on that set or not. To formulate the algorithm via POP, we introduce measurable Gaussian transformations $T_L^{y_j} : \mathbb{R}^d \to \mathbb{R}$ such that the $G_L^{y_j}(\omega) = T_L^{y_j}(\Theta(\omega))$ satisfy Assumption 2.2. This small variant of the above subset simulation algorithm is also incorporated in Algorithm A.1.

Adaptive multilevel subset simulation. To formulate this algorithm in such a way that the subset property is satisfied we choose $y_1, \dots, y_K$ as defined in Lemma 3.2 and choose the subsets

$$F_j^{\mathrm{ML}} = \{\omega \in \Omega : G_{\ell_j}^{y_j}(\omega) \leq y_j\}, \quad \text{for} \quad j = 1, \dots, K.$$

The proposed adaptive multilevel subset simulation algorithm with a standard MC estimator for the first subset is summarized in Algorithm A.2. In addition, an adaptive stopping criterion based on a statistical estimation of the rRMSE as presented in (4.7) is introduced in Algorithm A.2. Note that for simplicity we set $K = L$ and $\ell_j = j$, for all $j = 1, \dots, L$.

To employ classical subset simulation for the estimation of $\mathbb{P}(F_1^{\mathrm{ML}})$ instead, it suffices to replace lines 2–7 by a single-level subset simulation estimator as defined in Algorithm A.1, but on the coarsest accuracy level $\ell = 1$.

Algorithm A.2 Adaptive multilevel subset simulation.

1: Given an accuracy L, a value for $\gamma \in (0,1)$, and correlation parameter $\eta \in [0,1]$, define
  - subsets $F_\ell^{\mathrm{ML}} = \{\Theta \in A_\ell\}$ with $A_\ell = \{\theta \in \mathbb{R}^d \mid T_\ell^{y_\ell}(\theta) \le y_\ell\}$, for $\ell = 0, \ldots, L$, and the sequence of failure thresholds

$$y_L = 0 \quad \text{and} \quad y_\ell = (\gamma^\ell + \gamma^{\ell+1}) + y_{\ell+1}, \quad \ell = 1, \ldots, L-1,$$

  - a shaking transformation $S_\eta(\theta, w)$ as defined in (4.3), and
  - a rejection operator $M_j^{S_\eta}$ as defined in (4.4).

2: Set $i = 0$, $N_1 = 1$, $\widehat{\delta}(\widehat{P}_1) = \infty$.
3: **while** $\widehat{\delta}(\widehat{P}_1) > \mathrm{TOL}$ **do**
4:      Generate i.i.d. $\Theta_{0,i} \sim \Theta$.
5:      Estimate the probability of subset F_1^{ML} by $\widehat{P}_1 = \frac{1}{N_1} \sum_{k=0}^{N_1-1} \mathbb{1}_{A_1}(\Theta_{0,k})$.
6:      Estimate $\delta(\widehat{P}_1) \approx \widehat{\delta}(\widehat{P}_1)$ and increase $i = i+1$, $N_1 = N_1 + 1$.
7: **end while**
8: Set $i_0 = \arg\min\{k \mid \Theta_{0,k} \in A_1\}$ and define initial state $\Theta_{1,0} = \Theta_{0,i_0}$ for next level.
9: **for** $\ell = 1, \ldots, L-1$ **do**
10:      Set $i = 0$, $N_{\ell+1} = 1$, $\widehat{P}_{\ell+1} = \mathbb{1}_{A_{\ell+1}}(\Theta_{\ell,i})$.
11:      **while** $\widehat{\delta}(\widehat{P}_{\ell+1}) > \mathrm{TOL}$ **do**
12:          Generate $W_{\ell,i} \sim \Theta$.
13:          Shake and accept/reject $\Theta_{\ell,i+1} = M_\ell^{S_\eta}(\Theta_{\ell,i}, W_{\ell,i})$.
14:          Increase $i = i+1$, $N_{\ell+1} = N_{\ell+1} + 1$.
15:          Estimate the subset probability $F_{\ell+1}^{\mathrm{ML}}$ by $\widehat{P}_{\ell+1} = \frac{1}{N_{\ell+1}} \sum_{i=0}^{N_{\ell+1}-1} \mathbb{1}_{A_{\ell+1}}(\Theta_{\ell,i})$.
16:          Estimate $\delta(\widehat{P}_{\ell+1}) \approx \widehat{\delta}(\widehat{P}_{\ell+1})$.
17:      **end while**
18:      Set $i_\ell = \arg\min\{k \mid \Theta_{\ell,k} \in A_{\ell+1}\}$ and define initial state $\Theta_{\ell+1,0} = \Theta_{\ell,i_\ell}$ for next level.
19: **end for**
20: **Result:** $\widehat{P}^{\mathrm{ML}_1} = \prod_{\ell=1}^{L} \widehat{P}_\ell$.

REFERENCES

[1] A. Agarwal, S. De Marco, E. Gobet, and G. Liu, *Rare Event Simulation Related to Financial Risks: Efficient Estimation and Sensitivity Analysis*, HAL preprint, hal-01219616, 2017, https://hal-polytechnique.archives-ouvertes.fr/hal-01219616.

[2] S. Au, *Reliability-based design sensitivity by efficient simulation*, Comput. Struct., 83 (2005), pp. 1048–1061.

[3] S. Au and E. Patelli, *Rare event simulation in finite-infinite dimensional space*, Reliab. Eng. Syst. Safety, 148 (2016), pp. 67–77.

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

[5] S.-K. Au and J. L. Beck, *Estimation of small failure probabilities in high dimensions by subset simulation*, Probab. Eng. Mech., 16 (2001), pp. 263–277.

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

[7] A. Barth, C. Schwab, and N. Zollinger, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161.

[8] R. Becker and R. Rannacher, *A feed-back approach to error control in finite element methods: Basic analysis and examples*, East-West J. Numer. Math., 4 (1996), pp. 237–264.

[9] J. Bect, L. Li, and E. Vazquez, *Bayesian subset simulation*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 762–786, https://doi.org/10.1137/16M1078276.

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

[11] C.-E. Bréhier, T. Lelièvre, and M. Rousset, *Analysis of adaptive multilevel splitting algorithms in an idealized case*, ESAIM Probab. Stat., 19 (2015), pp. 361–394.

[12] C. G. Bucher, *Adaptive sampling — An iterative fast Monte Carlo procedure*, Struct. Saf., 5 (1988), pp. 119–126, https://doi.org/10.1016/0167-4730(88)90020-3.

[13] C. G. Bucher, *Asymptotic sampling for high-dimensional reliability analysis*, Probab. Eng. Mech., 24 (2009), pp. 504–510.

[14] T. Butler, C. Dawson, and T. Wildey, *A posteriori error analysis of stochastic differential equations using polynomial chaos expansions*, SIAM J. Sci. Comput., 33 (2011), pp. 1267–1291, https://doi.org/10.1137/100795760.

[15] V. Caron, A. Guyader, M. M. Zuniga, and B. Tuffin, *Some recent results in rare event estimation*, ESAIM Proc. Surveys, 44 (2014), pp. 239–259, https://doi.org/10.1051/proc/201444015.

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

[17] Y. Chen, T. A. Davis, W. W. Hager, and S. Rajamanickam, *Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate*, ACM Trans. Math. Software, 35 (2008), 22.

[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. Vis. Sci., 14 (2011), pp. 3–15.

[19] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White, *MCMC methods for functions: Modifying old algorithms to make them faster*, Statist. Sci., 28 (2013), pp. 424–446, https://doi.org/10.1214/13-STS421.

[20] G. Detommaso, T. Dodwell, and R. Scheichl, *Continuous level Monte Carlo and sample-adaptive model hierarchies*, SIAM/ASA J. Uncertain. Quantif., 7 (2019), pp. 93–116, https://doi.org/10.1137/18M1172259.

[21] F. DiazDelaO, A. Garbuno-Inigo, S. Au, and I. Yoshida, *Bayesian updating and model class selection with subset simulation*, CMAME, (2017), pp. 1102–1121.

[22] D. Elfverson, D. Estep, F. Hellman, and A. Målqvist, *Uncertainty quantification for approximate-quantiles for physical models with stochastic inputs*, SIAM/ASA J. Uncertain. Quantif., 2 (2014), pp. 826–850, https://doi.org/10.1137/140967039.

[23] D. Elfverson, F. Hellman, and A. Målqvist, *A multilevel Monte Carlo method for computing failure probabilities*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 312–330, https://doi.org/10.1137/140984294.

[24] M. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[25] M. B. Giles and E. Suli, *Adjoint methods for pdes: A posteriori error analysis and postprocessing by duality*, Acta Numer., 11 (2002), pp. 145–236.

[26] E. Gobet and G. Liu, *Rare event simulation using reversible shaking transformations*, SIAM J. Sci. Comput., 37 (2015), pp. A2295–A2316, https://doi.org/10.1137/14098418X.

[27] Z. T. Gong, F. A. DiazDelaO, P. O. Hristov, and M. Beer, *History matching with subset simulation*, Int. J. Uncertain. Quantif., 11 (2021), pp. 19–38.

[28] A. Guyader, N. Hengartner, and E. Matzner-Løber, *Simulation and estimation of extreme quantiles and extreme probabilities*, Appl. Math. Optim., 64 (2011), pp. 171–196, https://doi.org/10.1007/s00245-011-9135-z.

[29] A.-L. Haji-Ali, J. Spence, and A. Teckentrup, *Adaptive Multilevel Monte Carlo for Probabilities*, preprint, https://arxiv.org/abs/2107.09148.

[30] H. Kahn and T. Harris, *Estimation of particle transmission by random sampling*, Bureau Standard Appl. Math. Ser., 12 (1951), pp. 27–30.

[31] P.-S. Koutsourelakis, *Accurate uncertainty quantification using inaccurate computational models*, SIAM J. Sci. Comput., 31 (2009), pp. 3274–3300, https://doi.org/10.1137/080733565.

[32] R. NEAL, *Regression and classification using Gaussian process priors*, Bayesian Statistics, 6 (1998), pp. 475–501.

[33] T. ODEN, C. DAWSON, AND T. WILDEY, *Adaptive multiscale predictive modelling*, Acta Numer., 27 (2018), pp. 353–450, https://doi.org/10.1017/S096249291800003X.

[34] I. PAPAIOANNOU, W. BETZ, K. ZWIRGLMAIER, AND D. STRAUB, *MCMC algorithms for subset simulation*, Probab. Eng. Mech., 41 (2015), pp. 89–103.

[35] W. QI, Z. LU, AND C. ZHOW, *New topology optimization method for wing leading-edge ribs*, J. Aircraft, 48 (2011), pp. 1741–1748.

[36] K. RITTER, *Approximation and optimization on the Wiener space*, J. Complexity, 6 (1990), pp. 337–364, https://doi.org/10.1016/0885-064X(90)90027-B.

[37] G. SCHUËLLER AND R. STIX, *A critical appraisal of methods to determine failure probabilities*, Struct. Saf., 4 (1987), pp. 293–309.

[38] M. SHINOZUKA, *Basic analysis of structural safety*, J. Struct. Eng., 109 (1983), pp. 721–740.

[39] E. ULLMANN AND I. PAPAIOANNOU, *Multilevel estimation of rare events*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 922–953, https://doi.org/10.1137/140992953.

[40] F. URIBE, I. PAPAIOANNOU, Y. M. MARZOUK, AND D. STRAUB, *Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation*, SIAM/ASA J. Uncertain. Quantif., 9 (2021), pp. 818–847, https://doi.org/10.1137/20M1344585.

[41] F. WAGNER, J. LATZ, I. PAPAIOANNOU, AND E. ULLMANN, *Multilevel sequential importance sampling for rare event estimation*, SIAM J. Sci. Comput., 42 (2020), pp. A2062–A2087, https://doi.org/10.1137/19M1289601.

[42] F. WAGNER, J. LATZ, I. PAPAIOANNOU, AND E. ULLMANN, *Error analysis for probabilities of rare events with approximate models*, SIAM J. Numer. Anal., 59 (2021), pp. 1948–1975, https://doi.org/10.1137/20M1359808.

[43] S. XIN-SHI, Y. XIONG-QING, AND L. HONG-SHUANG, *Subset simultion for multi-objective optimisation*, Appl. Math. Model., 44 (2017), pp. 425–445.

[44] K. ZUEV, J. BECK, S. AU, AND L. KATAFYGIOTIS, *Bayesian post-processor and other enhancements of subset simulation for estimating failure probabilities in high dimensions*, Comput. Struct., 92 (2012), pp. 283–296.