Accepted for publication in "Computer Methods in Applied Mechanics and Engineering"

An efficient framework for optimal robust stochastic system design using stochastic simulation

Alexandros A. Taflanidis and James L. Beck*
California Institute of Technology, Engineering and Applied Science Division

Abstract

The knowledge about a planned system in engineering design applications is never complete. Often, a probabilistic quantification of the uncertainty arising from this missing information is warranted in order to efficiently incorporate our partial knowledge about the system and its environment into their respective models. This leads to a robust stochastic design framework where probabilistic models of excitation uncertainties and system modeling uncertainties can be introduced; the design objective is then typically related to the expected value of a system performance measure, such as reliability or expected life-cycle cost. For complex system models, this expected value can rarely be evaluated analytically and so it is often calculated using stochastic simulation techniques, which involve an estimation error and significant computational cost. An efficient framework, consisting of two stages, is presented here for the optimization in such stochastic design problems. The first stage implements a novel approach, called Stochastic Subset Optimization (SSO), for iteratively identifying a subset of the original design space that has high plausibility of containing the optimal design variables. The second stage adopts some other stochastic optimization algorithm to pinpoint the optimal design variables within that subset. The focus is primarily on the theory and implementation issues for SSO but also on topics related to the combination of the two different stages for overall enhanced efficiency. An illustrative example is presented that shows the efficiency of the proposed methodology; it considers the optimization of the reliability of a base-isolated structure considering future near-field ground motions.
KEYWORDS: Optimal stochastic design, stochastic optimization, stochastic subset optimization, reliability-based design, common random numbers, stochastic simulation.

1. Introduction

In engineering design, the knowledge about a planned system is never complete. First, it is not known in advance which design will lead to the best system performance in terms of a specified metric; it is therefore desirable to optimize the performance measure over the space of design variables that define the set of acceptable designs. Second, modeling uncertainty arises because no mathematical model can capture perfectly the behavior of a real system and its environment (future excitations). In practice, the designer chooses a model that he or she feels will adequately represent the behavior of the built system as well as its future excitation; however, there is always uncertainty about which values of the model parameters will give the best representation of the constructed system and its environment, so this parameter uncertainty should be quantified. Furthermore, whatever model is chosen, there will always be an uncertain prediction error between the model and system responses. For an efficient engineering design, all uncertainties, involving future excitation events as well as the modeling of the system, must be explicitly accounted for. A probabilistic approach provides a rational and consistent framework for quantifying all of these uncertainties [1]. In this case, this process is often called *robust stochastic system design*.

In this context, consider some controllable parameters that define the system design, referred to herein as *design variables*, $\boldsymbol{\varphi} = [\varphi_1\ \varphi_2\ \ldots\ \varphi_{n_\varphi}] \in \Phi \subset \mathbb{R}^{n_\varphi}$, where Φ denotes the bounded admissible design space. Also consider a model class that is chosen to represent a system design and its future excitation, where each model in the class is specified by an n_θ-dimensional vector $\boldsymbol{\theta} = [\theta_1\ \theta_2\ \ldots\ \theta_{n_\theta}]$ lying in $\Theta \subset \mathbb{R}^{n_\theta}$, the set of possible values for the model parameters. Because there is uncertainty in which model best represents the system behavior, a PDF (probability density function) $p(\boldsymbol{\theta}|\boldsymbol{\varphi})$, which incorporates available knowledge about the system, is assigned to these parameters. The objective function for a robust-to-uncertainties design is, then, expressed as:

$$E_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi},\boldsymbol{\theta})] = \int_{\Theta} h(\boldsymbol{\varphi},\boldsymbol{\theta})\, p(\boldsymbol{\theta}\,|\,\boldsymbol{\varphi})\, d\boldsymbol{\theta} \tag{1}$$

where $E_{\boldsymbol{\theta}}[.]$ denotes expectation with respect to the PDF for $\boldsymbol{\theta}$ and $h(\boldsymbol{\varphi},\boldsymbol{\theta}) : \mathbb{R}^{n_\varphi} \mathrm{x} \mathbb{R}^{n_\theta} \rightarrow \mathbb{R}$ denotes the performance measure of the system. In engineering applications, stochastic design problems are many times posed by adopting deterministic objective functions and using constraints related to stochastic integrals like (1) to characterize the admissible design space -such an approach is common, for example, in the context of Reliability-Based Design Optimization (RBDO) where

reliability constraints are adopted [2, 3]. In this study we focus on design problems that entail as objective function a stochastic integral of the form (1). The *optimal stochastic design problem* in this case takes the form:

$$\begin{aligned} &\text{minimize } E_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi},\boldsymbol{\theta})] \\ &\text{subject to } \mathbf{f}_c(\boldsymbol{\varphi}) \geq 0 \end{aligned} \tag{2}$$

where $\mathbf{f}_c(\boldsymbol{\varphi})$ corresponds to a vector of constraints. Such optimization problems, arising in decision making under uncertainty, are typically referred to as stochastic optimization problems (e.g. [4, 5]). The constraints in optimization (2) can be taken into account by appropriate definition of the admissible design space Φ; the stochastic design problem is then equivalently formulated as:

$$\boldsymbol{\varphi}^* = \arg\min_{\boldsymbol{\varphi}\in\Phi} E_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi},\boldsymbol{\theta})] \tag{3}$$

For this optimization, the integral in (1) must be evaluated. For complex systems this integral can rarely be calculated, or even efficiently approximated, analytically and so it is commonly evaluated through stochastic simulation techniques. In this setting, an unbiased estimate of the expected value in (1) can be obtained using a finite number, N, of random samples of $\boldsymbol{\theta}$, drawn from $p(\boldsymbol{\theta}|\boldsymbol{\varphi})$:

$$\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)] = \frac{1}{N}\sum_{i=1}^{N} h(\boldsymbol{\varphi},\boldsymbol{\theta}_i) \tag{4}$$

where $\boldsymbol{\Omega}_N=[\boldsymbol{\theta}_1 \ \ldots \ \boldsymbol{\theta}_N]$ is the sample set of the model parameters with vector $\boldsymbol{\theta}_i$ denoting the sample of these parameters used in the i^{th} simulation. This estimate of $E_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi},\boldsymbol{\theta})]$ involves an unavoidable error $e_N(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)$ which is a complex function of both the sample set $\boldsymbol{\Omega}_N$ as well as the current system model configuration. The optimization in (3) is then approximated by:

$$\boldsymbol{\varphi}_N^* = \arg\min_{\boldsymbol{\varphi}\in\Phi} \hat{E}_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)] \tag{5}$$

If the stochastic simulation procedure is a consistent one, then as $N\to\infty$, $\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)] \to E_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi},\boldsymbol{\theta})]$ and $\boldsymbol{\varphi}_N^* \to \boldsymbol{\varphi}^*$ under mild regularity conditions for the optimization algorithms used [5]. The existence of the estimation error $e_N(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)$, which may be considered as noise in the objective function, contrasts with classical deterministic optimization where it is assumed that one has perfect information..

Figure 1 illustrates the difficulties associated with $e_N(\boldsymbol{\Omega}_N,\boldsymbol{\varphi})$. The curves corresponding to simulation-based evaluation of the objective function have non-smooth characteristics, a feature which makes application of gradient-based algorithms challenging. Also, the estimated optimum depends on the exact influence of the estimation error, which is not the same for all evaluations. Another source of difficulty, especially when complex system models are considered, is the high computational cost associated with the estimation in (5) since N system analyses must be performed for each objective function evaluation. Even though recent advanced stochastic optimization algorithms (see section 3) can efficiently address the first two aforementioned problems this latter one remains challenging for many engineering design applications. Specialized, approximate approaches have been proposed in various engineering fields for reduction of the computational cost (e.g. [2, 3, 6] for reliability-based optimal design problems). These approaches may work satisfactorily under certain conditions, but are not proved to always converge to the solution of the original design problem. For this reason such approaches are avoided in this current study. Optimization problem (5) is directly solved so that $\boldsymbol{\varphi}_N^* \approx \boldsymbol{\varphi}^*$.

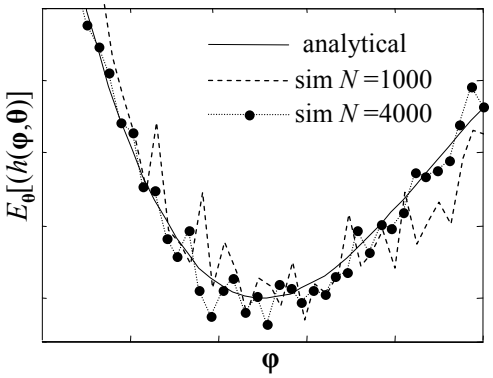

Figure 1: (a) analytical and simulation-based (sim) evaluation of an objective function

An efficient framework, consisting of two stages, is discussed in the following sections for a computational efficient solution to this optimization. The first stage implements a novel approach, called Stochastic Subset Optimization (SSO) [7, 8], for efficiently exploring the global sensitivity of the objective function to the design variables and for iteratively converging to a subset of the original design space that has high plausibility of containing the optimal design variables. The second stage adopts some appropriate stochastic optimization algorithm to pinpoint the optimal design variables within the set identified in the first stage. The focus is primarily on the theory and implementation issues for SSO but also on topics related to the combination of the two different stages for enhanced overall efficiency.

2. Stochastic Subset Optimization

Stochastic Subset Optimization (SSO) was initially suggested for reliability-based optimization problems (for a proper definition of such problems see Section 5.1 later on) in [9] and has been recently [8] extended to address general stochastic design problems, such as the one in (2). The basic features of the algorithm are summarized next.

2.1 Augmented problem and subset analysis

Consider, initially, the modified positive function, $h_s(\boldsymbol{\varphi},\boldsymbol{\theta}) : \mathbb{R}^{n_\varphi} \mathrm{x} \mathbb{R}^{n_\theta} \to \mathbb{R}^+$ defined as

$$h_s(\boldsymbol{\varphi},\boldsymbol{\theta}) = h(\boldsymbol{\varphi},\boldsymbol{\theta}) - s \quad \text{where} \quad s < \min_{\boldsymbol{\varphi},\boldsymbol{\theta}} h(\boldsymbol{\varphi},\boldsymbol{\theta}) \tag{6}$$

and note that $E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})] = E_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi},\boldsymbol{\theta})] - s$. Since the two expected values differ only by a constant, optimization of the expected value of $h(.)$ is equivalent, in terms of the optimal design choice, to optimization for the expected value for $h_s(.)$. In the SSO setting we focus on the latter optimization.

The basic idea in SSO is the formulation of an augmented problem, a general concept initially discussed in [10] for reliability-based design problems, where the design variables are artificially considered as uncertain with distribution $p(\boldsymbol{\varphi})$ over the design space Φ. In the setting of this augmented stochastic design problem, define the auxiliary PDF:

$$\pi(\boldsymbol{\varphi},\boldsymbol{\theta}) = \frac{h_s(\boldsymbol{\varphi},\boldsymbol{\theta})\, p(\boldsymbol{\varphi},\boldsymbol{\theta})}{E_{\boldsymbol{\varphi},\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]} \propto h_s(\boldsymbol{\varphi},\boldsymbol{\theta})\, p(\boldsymbol{\varphi},\boldsymbol{\theta}) \tag{7}$$

where $p(\boldsymbol{\varphi},\boldsymbol{\theta}) = p(\boldsymbol{\varphi}) p(\boldsymbol{\theta}|\boldsymbol{\varphi})$. The normalizing constant in the denominator is defined as:

$$E_{\boldsymbol{\varphi},\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})] = \int_{\Phi}\int_{\Theta} h_s(\boldsymbol{\varphi},\boldsymbol{\theta})\, p(\boldsymbol{\varphi},\boldsymbol{\theta})\, d\boldsymbol{\theta}\, d\boldsymbol{\varphi} \tag{8}$$

and corresponds to the expected value in the augmented uncertain space. This expected value is not explicitly needed, but it can be obtained though stochastic simulation, which leads to an expression similar to (4) but with the pair $[\boldsymbol{\varphi}, \boldsymbol{\theta}]$ defining the uncertain parameters. The transformation of the performance measure in (6) may be necessary to ensure that $\pi(\boldsymbol{\varphi},\boldsymbol{\theta}) \geq 0$. For many stochastic design problems $h(\boldsymbol{\varphi},\boldsymbol{\theta}) \geq 0$ and the transformation in (6) is unnecessary.

In terms of the auxiliary PDF, the objective function $E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]$ can be expressed as:

$$E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})] = \frac{\pi(\boldsymbol{\varphi})}{p(\boldsymbol{\varphi})} E_{\boldsymbol{\varphi},\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})] \tag{9}$$

where the marginal PDF $\pi(\boldsymbol{\varphi})$ is equal to:

$$\pi(\boldsymbol{\varphi}) = \int_{\Theta} \pi(\boldsymbol{\varphi},\boldsymbol{\theta})\, d\boldsymbol{\theta} \tag{10}$$

Define, now:

$$J(\boldsymbol{\varphi}) = \frac{E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]}{E_{\boldsymbol{\varphi},\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]} = \frac{\pi(\boldsymbol{\varphi})}{p(\boldsymbol{\varphi})} \tag{11}$$

Since $E_{\boldsymbol{\varphi},\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]$ can be viewed simply as a normalizing constant, minimization of $E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]$ is equivalent to the minimization of $J(\boldsymbol{\varphi})$. For this purpose the marginal PDF $\pi(\boldsymbol{\varphi})$ in the numerator of $J(\boldsymbol{\varphi})$ must be evaluated. Samples of this PDF can be obtained through stochastic simulation techniques [11]; for example, by using direct Monte Carlo simulation (MCS) or Markov Chain Monte Carlo (MCMC) sampling. Appendix A provides a brief description of two such algorithms. These algorithms will give sample pairs $[\boldsymbol{\varphi},\boldsymbol{\theta}]$ that are distributed according to the joint distribution $\pi(\boldsymbol{\varphi},\boldsymbol{\theta})$. Their $\boldsymbol{\varphi}$ component corresponds to samples from the marginal distribution $\pi(\boldsymbol{\varphi})$. Analytical approximations of $\pi(\boldsymbol{\varphi})$ based on these samples, using for example kernel density methods or the maximum entropy method, can be problematic for complicated problems, such as when n_φ is large or the sensitivity for some design variables is complex [7]. In the SSO framework, such approximation of $\pi(\boldsymbol{\varphi})$ is avoided. The sensitivity analysis is performed by looking at the average value (or equivalently volume density) of $J(\boldsymbol{\varphi})$ over any subset of the design space $I \subset \Phi$, denoted by $H(I)$:

$$H(I) \triangleq \frac{\int_I J(\boldsymbol{\varphi}) d\boldsymbol{\varphi}}{V_I} \propto \frac{\int_I E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})] d\boldsymbol{\varphi}}{V_I} \tag{12}$$

This term is also proportional, ignoring normalization constants, to the average value of the objective function in set I (which can be also considered [8] as the "normalized average set content"). To simplify the evaluation of $H(I)$, a uniform distribution is chosen for $p(\boldsymbol{\varphi})$. Note that $p(\boldsymbol{\varphi})$ does not reflect the uncertainty in $\boldsymbol{\varphi}$ but is simply a device for formulating the augmented problem and thus can be selected according to convenience. Finally, $H(I)$ and an estimate of it based on the samples from $\pi(\boldsymbol{\varphi})$ obtained as described previously, are given, respectively, by:

$$H(I) = \frac{V_\Phi}{V_I} \int_I \pi(\boldsymbol{\varphi}) d\boldsymbol{\varphi} \tag{13}$$

$$\hat{H}(I) = \frac{N_I / V_I}{N_\Phi / V_\Phi} \tag{14}$$

where N_I and N_Φ denote the number of samples from $\pi(\boldsymbol{\varphi})$ belonging to the sets I and Φ, respectively, and V_Φ is the volume of the design space Φ. The estimate for $H(I)$ is equal to the ratio of the volume density of samples from $\pi(\boldsymbol{\varphi})$ in sets I and Φ. The coefficient of variation (c.o.v.) for this estimate depends on the simulation technique used for obtaining the samples from $\pi(\boldsymbol{\varphi})$. For a broad class of sampling algorithms this c.o.v. may be expressed as:

$$\text{c.o.v.}\, \hat{H}(I) = \sqrt{\frac{1 - P(\boldsymbol{\varphi} \in I)}{N \cdot P(\boldsymbol{\varphi} \in I)}} \approx \sqrt{\frac{1 - N_I / N_\Phi}{N \cdot N_I / N_\Phi}}\,, \qquad P(\boldsymbol{\varphi} \in I) \triangleq \int_I \pi(\boldsymbol{\varphi}) d\boldsymbol{\varphi} \approx N_I / N_\Phi \tag{15}$$

where $N = N_\Phi/(1+\gamma)$, $\gamma \geq 0$, is the effective number of independent samples. If direct Monte Carlo techniques are used then $\gamma = 0$, but if Markov Chain Monte Carlo (MCMC) sampling is selected then $\gamma > 0$ because of the

correlation of the generated samples. Ultimately, the value of γ depends on the characteristics of the algorithm used.

For the uniform PDF for $p(\boldsymbol{\varphi})$, note that $H(\Phi)=1$ and $H(I)$ is equal to the ratio:

$$H(I)=\frac{\int_I E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]d\boldsymbol{\varphi}/V_I}{\int_\Phi E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]d\boldsymbol{\varphi}/V_\Phi} \tag{16}$$

where the integrals in the numerator and denominator correspond to the average value of the objective function in sets I and Φ, respectively (average set contents). Thus $H(I)$ expresses the average relative sensitivity of $E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]$ to $\boldsymbol{\varphi}$ within the set $I\subset\Phi$: greater sensitivity, i.e. bigger contrast in the average value (i.e. volume density) of the objective function, corresponds to smaller values for $H(I)$.

2.2 Subset optimization

Consider a set of admissible subsets A in Φ that have some predetermined shape and some size constraint, for example related to the set volume, and define the *deterministic subset optimization*:

$$I^*=\arg\min_{I\in A} H(I) \tag{17}$$

Based on the estimate in (13), this optimization is approximately equal to the following *stochastic subset optimization*:

$$\hat{I}=\arg\min_{I\in A}\hat{H}(I)=\arg\min_{I\in A}\frac{N_I}{V_I} \tag{18}$$

i.e., to identification of the set $I\in A$ that contains the smallest volume density N_I/V_I of samples.

Optimization (17) identifies the set that gives the smallest average value of $J(\boldsymbol{\varphi})$ (or equivalently $E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]$) within the class of admissible subsets A. If set A is properly chosen, for example if its shape is "close" to the contours of $E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]$ in the vicinity of $\boldsymbol{\varphi}^*$, then $\boldsymbol{\varphi}^*\in I^*$ for the optimization in (17); this is true even if the constraints of the design problem are active so that $\boldsymbol{\varphi}^*$ is an active optimum rather than an interior optimum in Φ. This whole argument is not necessarily true for the optimization in (18) because only estimates of $H(I)$ are used. $\hat{I}$ is simply the set, within the admissible subsets A, that has the largest likelihood, in terms of the information available through the obtained samples, of including $\boldsymbol{\varphi}^*$. This likelihood defines the quality of the identification and ultimately depends [8] on $H(I)$; taking into account the fact that the average value of $E_{\boldsymbol{\theta}}[h_s(\boldsymbol{\varphi},\boldsymbol{\theta})]$ in the neighborhood of the optimal solution is the smallest in Φ, it is evident (see equation (16)) that smaller values of $H(\hat{I})$ correspond to greater plausibility for the set $\hat{I}$ to include $\boldsymbol{\varphi}^*$. Since only estimates of $H(\hat{I})$ are available in the stochastic identification, the quality depends, ultimately, on both: (a) the estimate $\hat{H}(\hat{I})$ and (b) its coefficient of variation (defining the accuracy of that estimate). Smaller values for these parameters correspond to better quality of identification. Both of them should be used as a measure of this quality.

Note that the computational cost for obtaining the samples needed for optimization (18) is comparable to the cost required for a single evaluation of the objective function in (4), depending on how many samples are simulated and the details of the algorithm used.

2.3 Iterative approach

The relative size of the admissible subsets I define (a) the resolution of $\boldsymbol{\varphi}^*$ and (b) the accuracy information about $\hat{H}(I)$ that is extracted from the samples from $\pi(\boldsymbol{\varphi})$. Selecting smaller size for the admissible sets leads to better resolution for $\boldsymbol{\varphi}^*$. At the same time, though, this selection leads to smaller values for the ratio N_I/N_Φ (since smaller number of samples are included in smaller sets) and thus it increases the c.o.v. (reduces accuracy) of the estimation, as seen from (15). In order to maintain the same quality for the estimate, the effective number of independent samples must be increased, which means that more simulations must be performed. Since we are interested in subsets in Φ with small average value, the required number of simulations to gather accurate information for subsets with small size is large, i.e. small percentage of the simulated samples fall in these subsets. To account for this characteristic and to increase the efficiency of the identification process, an iterative approach can be adopted. At iteration k, additional samples in $\hat{I}_{k-1}$ (where $\hat{I}_0=\Phi$) that are distributed according to $\pi(\boldsymbol{\varphi})$ are obtained. A region $\hat{I}_k\subset\hat{I}_{k-1}$ for the optimal design parameters is then identified as above. The quality of the identification is improved by applying such an iterative scheme, since the ratio of the samples in $\hat{I}_{k-1}$ to the one in $\hat{I}_k$ is larger than the equivalent ratio when comparing $\hat{I}_k$ and the original design space Φ. The samples $[\boldsymbol{\varphi},\boldsymbol{\theta}]$ available from the previous iteration, whose $\boldsymbol{\varphi}$ component lies inside set $\hat{I}_{k-1}$, can be exploited for improving the efficiency of the sampling process. In terms of the algorithms described in Appendix A this may be established for example by (a) forming better proposal PDFs or (b) using the samples already available as seeds for MCMC simulation (since these samples already follow the target distribution).

This iterative approach leads to a favorable feature for the computational cost of SSO with respect to the dimension of the search space (number of design variables, n_φ). For a specific reduction $\delta_k=V_{I_k}/V_{I_{k-1}}$ of the size (volume) of the search space in some step of the set identification, the corresponding average size reduction per design variable is $\sqrt[n_\varphi]{\delta_k}$. This means that if the identification was performed in one step, the logarithmic average size reduction per variable would be inversely proportional to the dimension of the search space n_φ

(assuming δ_k remains the same). In the suggested iterative approach though, in n_{iter} iterations the average size reduction per design variables is:

$$\sqrt[n_\varphi]{\prod_{k=1}^{n_{iter}} \delta_k} = \left(\sqrt[n_\varphi]{\delta_{mean}{}^{n_{iter}}} \right) = \left(\delta_{mean} \right)^{n_{iter}/n_\varphi} \tag{19}$$

where

$$\delta_{mean} = \sqrt[n_{iter}]{\prod_{k=1}^{n_{iter}} \delta_k} \tag{20}$$

is the geometric mean of the volume reductions over all of the iterations (note that if $\delta_k=\delta$, then $\delta_{mean}=\delta$). Thus, for the same average total size reduction over all design variables (left hand side of equation (19)), the number of required iterations is proportional to the dimension of the design space (look at the exponent in right hand side of equation (19)), assuming that the mean reduction of the volume over all iterations, δ_{mean}, is comparable. This argument shows that the efficiency of SSO decreases linearly with the dimension of the design space, so SSO should be considered appropriate for problems that involve a large number of design variables.

2.4 Selection of admissible subsets

Proper selection of the geometrical shape and size of the admissible sets is important for the efficiency of SSO. The geometrical shape should be chosen so that the challenging, non-smooth optimization (18) can be efficiently solved while the sensitivity of the objective function to each design variable is fully explored. For example, a hyper-rectangle or a hyper-ellipse can be a appropriate choices for shape of admissible subsets, depending also on the shape of the design space Φ. For problems involving complex design spaces, special attention is needed for the proper selection of the geometrical shape for I. In such cases, re-definition of the admissible design space in (3) might be beneficial; this can be established by including some of the complex constraints of (2) as part of the objective function (as penalty terms). Note, additionally, that the difference in size (volume) between Φ and the largest possible subset $I \subset \Phi$ should not be large. If this property does not hold, then the size reduction in the first iteration of SSO will necessarily have to be at least as big as this volume difference. This feature might reduce the quality of the SSO identification in the first iteration. This problem may be circumvented by appropriate adjustment of the design space, Φ, based on the desired shapes for the class of admissible subsets. For example, a superset Φ_{sup} that has shape similar to the one of the admissible subsets, and circumscribes the desired design set Φ can be selected as initial search space for the optimal system configuration.

The size of admissible subsets, now, is related to the quality of identification as discussed earlier. Selection of this size can be determined, for example, by incorporating a constraint for either (i) the volume ratio $\delta=V_I/V_\Phi$ or (ii) the number of samples ratio $\rho=N_I/N_\Phi$. The first choice cannot be directly related to any of the measures of quality of identification; thus proper selection of δ is not straightforward, though our experience indicates that δ close to 0.25 is, in general, an appropriate option. The second choice allows for directly controlling the coefficient of variation (see (15)) and thus one of the parameters influencing the quality of identification. This is the selection that is discussed in more detail here, that is, $A_\rho=\{I \subset \Phi: \rho=N_I/N_\Phi\}$.

The volume (size) of the admissible subsets, δ_k, in this scheme is adaptively chosen so that the ratio of samples in the identified set is equals to ρ. The choice of the value for ρ affects the efficiency of the identification. If ρ is large, fewer number of samples is required for the same level of accuracy (c.o.v. in (15)). However, a large value of ρ means that the size of the identified subsets will decreases slowly (larger size sets are identified), requiring more steps to converge to the optimal solution. The choice of the constraint ρ is a trade-off between the number of samples required in each step and the number of steps required to converge to the optimal design choice. In the applications we have investigated so far it was found that choosing ρ=0.1-0.2 yields good efficiency. An adaptive scheme can also be applied: smaller values of ρ may be selected in the first iterations of SSO when the sensitivity of the design problem is large and so the values of $\hat{H}(\hat{I})$ small (see discussion later on too). As the algorithm converges to the optimal design configuration, $\hat{H}(\hat{I})$ increases and larger values of ρ can be chosen to decrease the c.o.v. of $\hat{H}(\hat{I})$ and thus improve the identification quality.

For problems with multiple local minima of the objective function, the subset identified by SSO does not necessarily include the global optimum, even in the context of the deterministic subset optimization (17). That subset will have the smallest estimated average value for the objective function within the selected class of admissible subsets; this does not guarantee automatically that it will include the global optimum. The selection of the geometrical shape of the admissible subsets and, more importantly, for ρ may determine whether that subset includes the global optimal solution or only a local one. This topic deserves a more detailed investigation which is left to future work.

2.5 Identification of optimal subsets

Another important issue for the efficiency of SSO is the identification of the optimal sets within the class of admissible subsets selected, i.e., optimization (18). A fundamental remark regarding this optimization is that the position in the search space of a set $I \in A$ and the number of sample points in it is non-continuous. Thus, only methods appropriate for non-smooth optimization problems, such as genetic algorithms or direct search

methods (see [12] for more details), should be applied for this optimization. The choice of mathematical description of the subsets I is important for the efficiency of these algorithms. Reference [13] provides a detailed discussion of normalization processes that can increase this efficiency for admissible subsets that are hyper-ellipses or hyper-rectangles. Note, additionally, that the evaluation of the objective function and the constraint for this optimization involve small computational effort; they simply consist of counting the number of samples within the subset, calculating the volume of the set, and checking that $I \subset \Phi$. Thus the optimizer can explore many candidate subsets during the solution of the problem (18) at a small computational cost. In our applications, exploration of 10^5 candidate solutions takes on the average 10-20 sec. This cost is orders of magnitude smaller than the cost of the stochastic simulation needed for generating the samples from $\pi(\boldsymbol{\varphi})$ for problems involving complex system models. Thus the computational efficiency of SSO is primarily determined by the efficiency of the stochastic simulation stage, not optimization (18).

The optimization (18) can therefore be efficiently solved if an appropriate algorithm is chosen and, additionally, the admissible subsets are appropriately mathematically described. Because of the significance of the optimization in the accuracy of SSO, special attention is warranted to guarantee that the identification of the optimal subsets is robust. A simpler geometrical characterization of the admissible subsets should be preferred when there is doubt about the reliability of the optimization process if more complex ones were chosen.

2.6 SSO algorithm

The SSO algorithm is summarized as follows (Figure 2 illustrates some important steps):

Initialization: Define the bounded design space Φ, and the desired geometrical shape for the subsets I. Decide on the desired number of samples N and on the value for the constraint ρ.

Step k: Use some sampling procedure, such as MC simulation for the 1st step and MCMC simulation for subsequent steps, in order to simulate N samples (or effective samples) from $\pi(\boldsymbol{\varphi})$ inside the subset $\hat{I}_{k-1}$. Identify subset $\hat{I}_k$ as:

$$\hat{I}_k = \arg\min_{I \in A_{\rho,k}} N_I / V_I$$
$$A_{\rho,k} = \left\{ I \subset \hat{I}_{k-1} : \rho = N_I / N \right\} \tag{21}$$

Keep only the $N_{\hat{I}_k}$ samples that belong to the subset $\hat{I}_k$ (exploit these samples in the next step).

Stopping criteria: At each step, estimate ratio:

$$\hat{H}(\hat{I}_k) = \frac{N_{\hat{I}_k}}{N_{\hat{I}_{k-1}}} \frac{V_{\hat{I}_{k-1}}}{V_{\hat{I}_k}} \tag{22}$$

and its coefficient of variation according to the simulation algorithm used. Based on these two quantities and the desired quality of the identification (see Section 4.1), decide on whether to (a) stop (or even increase N to obtain better accuracy information about $\hat{H}(\hat{I}_k)$) or (b) proceed to step k+1.

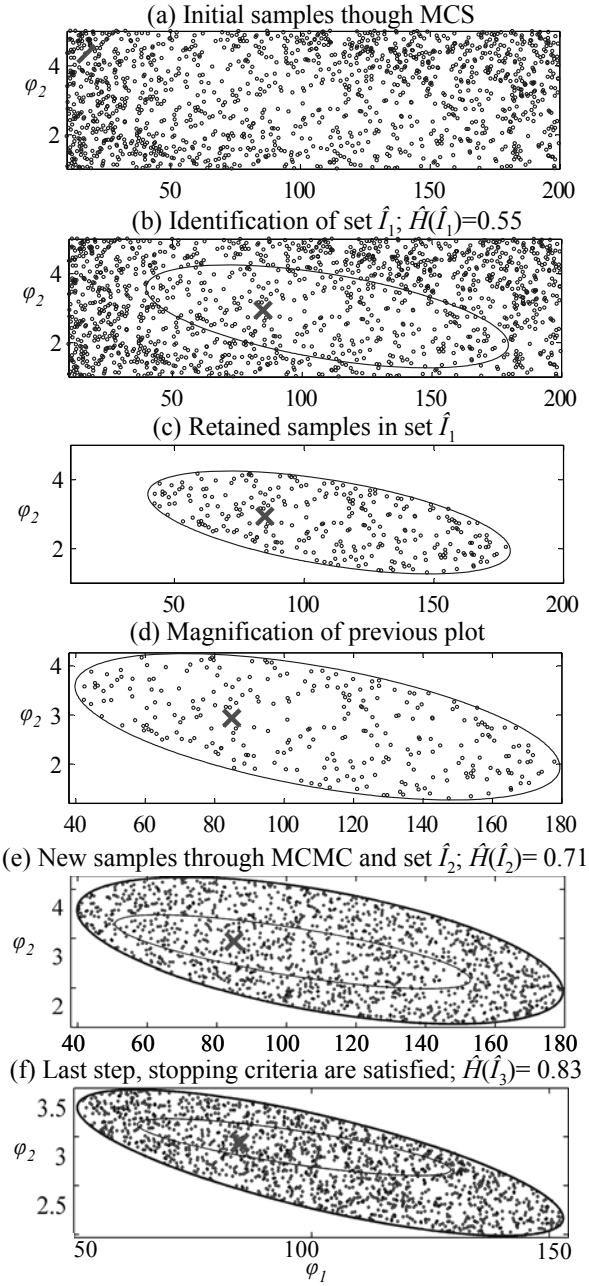

Figure 2: Illustration of SSO procedure (the X in these plots corresponds to actual optimal solution).

Figure 2 also demonstrates the dependence of the quality of the identification on $\hat{H}(\hat{I}_k)$ for a two-dimensional example. This ratio expresses the difference

in volume density of the samples inside and outside the identified set $\hat{I}_k$. In the first iteration, this difference is clearly visible. As SSO evolves and converges to subsets with smaller sensitivity to the objective function, the difference becomes smaller and by the last iteration (Figure 2(f)), it is difficult to visually discriminate which region in the set has smaller volume density of failed samples. This corresponds to a decrease in the quality of the identification.

The SSO algorithm described in this section, will adaptively converge to a relatively small sub-region for the optimal design variables $\boldsymbol{\varphi}^*$ within the original design space. This is true even if the constraints of problem (2) are active for the optimal design point (so it is not an interior point of the feasible region); in this case the sub-region will include the active optimum $\boldsymbol{\varphi}^*$, a characteristic that has been verified in the example considered in [9]. The size, now, of the sub-region identified by SSO depends on the sensitivity of the objective function around the optimal point. If that sensitivity is large then SSO will ultimately converge to a "small" set $\hat{I}_{sso}$, satisfying at the same time the accuracy requirements that make it highly likely that $\boldsymbol{\varphi}^*$ is in $\hat{I}_{sso}$. The center point of this set, denoted herein as $\boldsymbol{\varphi}_{SSO}$, gives the estimate for the optimal design variables. Additionally SSO gives information about the local behavior of the objective function. As long as the shape of the admissible subsets is close to the contours of the objective function near the optimal design point, the subset identified in the last stage of SSO provides a good estimate for these contours. For improving the accuracy of this estimate, it might be desirable to increase the number of samples N, in the last iteration of SSO, in order to obtain more information for $\pi(\boldsymbol{\varphi})$. Also, selection of the shape of admissible subsets as hyper-ellipses should be more appropriate for this purpose since the contours of the objective function are expected to fit better to hyper-elliptical shapes near the optimal design point.

In cases, though, that the sensitivity of the objective function around the optimal point is not large enough, convergence to a small subset might be problematic and will require increasing the number of samples in order to satisfy the requirement for the quality of identification. Another important issue related to the identification in such cases is that there is no measure of the quality of the identified solution, i.e. how close $\boldsymbol{\varphi}_{SSO}$ is to $\boldsymbol{\varphi}^*$, that can be directly established through the SSO results. If the identification is performed multiple times, the c.o.v. of $\{\hat{E}_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi}_{SSO,i},\boldsymbol{\theta})]\}$ could be considered a good candidate for characterizing this quality. This might not be always a good measure though. For example, if the choice for admissible subsets is inappropriate for the problem considered, it could be the case that consistent results are obtained for $\boldsymbol{\varphi}_{SSO}$ (small c.o.v.) that are far from the optimal design choice $\boldsymbol{\varphi}^*$. Also, this approach involves higher computational cost because of the need to perform the identification multiple times. For such cases, it could be more computationally efficient (instead of increasing N in SSO and performing the identification multiple times) and more accurate (in terms of identifying the true optimum), to combine SSO with some other optimization algorithm for pinpointing $\boldsymbol{\varphi}^*$. A discussion of topics related to such algorithms is presented next.

3. Stochastic optimization

We go back to the original formulation of the objective function in (1). In principle, though, the techniques discussed here are applicable to the case that the performance measure $h(\boldsymbol{\theta},\boldsymbol{\varphi})$ is replaced by the positive function $h_s(\boldsymbol{\theta},\boldsymbol{\varphi})$ used in the SSO setting (given by (6)).

3.1 Common random numbers and exterior sampling

The efficiency of stochastic optimizations such as (5) can be enhanced by the reduction of the absolute and/or relative importance of the estimation error $e_N(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)$. The absolute importance may be reduced by obtaining more accurate estimates of the objective function, i.e. by reducing the error $e_N(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)$ and thus the variance of the estimates. This can be established in various ways, for example by using importance sampling (see Section 4.2) or by selecting a larger sample size N in (4), but these typically involve extra computational effort. It is, thus, more efficient to seek a reduction in the relative importance of the estimation error. This means reducing the variance of the difference of estimates $\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^1,\boldsymbol{\Omega}_N^1)]$ and $\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^2,\boldsymbol{\Omega}_N^2)]$ that correspond to two different design choices $\boldsymbol{\varphi}^1$ and $\boldsymbol{\varphi}^2$. This variance can be decomposed as:

$$\begin{aligned} var\left(\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^1,\boldsymbol{\Omega}_N^1)]-\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^2,\boldsymbol{\Omega}_N^2)]\right)= \\ var\left(\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^1,\boldsymbol{\Omega}_N^1)]\right)+\mathrm{var}\left(\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^2,\boldsymbol{\Omega}_N^2)]\right) \\ -2\,\mathrm{cov}\left(\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^1,\boldsymbol{\Omega}_N^1)],\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^2,\boldsymbol{\Omega}_N^2)]\right) \end{aligned} \quad (23)$$

Deliberately introducing dependence in the evaluation of $\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^1,\boldsymbol{\Omega}_N^1)]$ and $\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi}^2,\boldsymbol{\Omega}_N^2)]$, increases the covariance (i.e. increases their correlation) and thus decreases their variability (the variance on the left). This decrease in the variance improves the efficiency of the comparison of the two estimates; it may be equivalently considered as creating a consistent estimation error. This task can be achieved by adopting common random numbers (CRN), i.e. $\boldsymbol{\Omega}_N^2=\boldsymbol{\Omega}_N^1$, in the simulations generating the two different estimates. Figure 3 shows the influence of such a selection: the curves that correspond to CRN are characterized by consistent estimation error and are smoother. Continuity and monotonicity of the output with respect to the random number input are key issues for improving the efficiency of stochastic comparisons when using common random

numbers [14]. If $h(\boldsymbol{\varphi},\boldsymbol{\theta})$ is sufficiently smooth then these two aforementioned requirements can be typically guaranteed, as long as the design choices compared are not too far apart in the design variable space, i.e. the systems compared are not too different so that the model parameter values have a similar effect on the different system outputs. In such cases it is expected that use of CRN will be advantageous. Note that if the systems compared are significantly different, i.e. correspond to $\boldsymbol{\varphi}$ that are not close, then CRN do not necessarily guarantee efficiency. This might occur, for examples, when the regions of Θ that contribute most to the integral of the expected value for the two systems are drastically different and the sample sets $\boldsymbol{\Omega}_N^2 = \boldsymbol{\Omega}_N^1$ selected do not efficiently represent both of these regions. This feature is also indicated in curve (iv) in Figure 3; the estimation error is not consistent along the whole range of $\boldsymbol{\varphi}$ for the CRN curves (but for local comparisons it appears to be consistent).

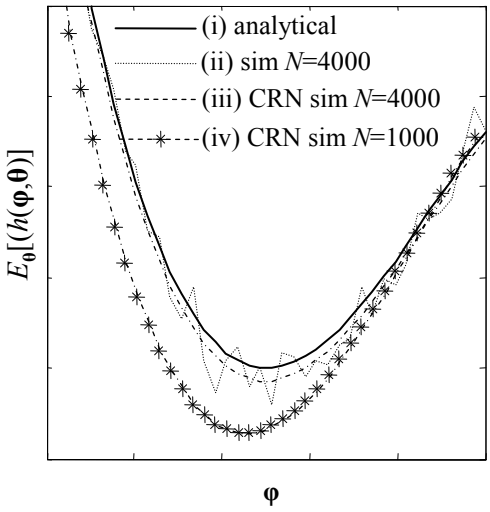

Figure 3: Illustration of some points in CRN-based evaluation of an objective function

A notion closely related to CRN is the exterior sampling approximation (ESA) for stochastic design optimizations. ESA adopts the same stream of random numbers throughout all iterations in the optimization process, thus transforming problem (5) into a deterministic system design problem, which can be solved by any appropriate routine. Usually ESA is implemented by selecting N "large enough" to give good quality estimates for the objective function and thus more accurate solutions to the optimization problem (see [5] for more details). The quality of the solution obtained through ESA is commonly assessed by solving the optimization problem multiple times, for different independent random sample streams. Even though the computational cost for the ESA deterministic optimization is typically smaller than that of the original stochastic search problem, the overall efficiency may be worse because of the requirement to perform the optimization multiple times.

3.2 Appropriate optimization algorithms

Both gradient-based and gradient-free algorithms can be used in conjunction with CRN or ESA and can be appropriate for stochastic optimizations.

Gradient-based algorithms: these use derivative information to iterate in the direction of steepest descent for the objective function. Only local designs are compared in each iteration, which makes the implementation of CRN efficient and allows for use of stochastic approximations [15] which can significantly improve the computational efficiency of stochastic search applications [4]. The latter approximation is performed by establishing (through proper recursive formulas) an equivalent averaging across the iterations of the optimization algorithm, instead of getting higher accuracy estimates for the objective function at each iteration, that is averaging over one single iteration. In simple design problems the performance measure $h(\boldsymbol{\varphi},\boldsymbol{\theta})$ may be such that the gradient of the objective function with respect to $\boldsymbol{\varphi}$ can be obtained through a single stochastic simulation analysis [4, 16]. In many cases though, the models used are generally complex, and it is difficult, or impractical, to develop an analytical relationship between the design variables and the objective function. Finite difference numerical differentiation is often the only possibility for obtaining information about the gradient vector but this involves computational cost which increases linearly with the dimension of the design parameters. Simultaneous-perturbation stochastic approximation (SPSA) [4, 17] is an efficient alternative search method. It is based on the observation that one properly chosen simultaneous random perturbation in all components of $\boldsymbol{\varphi}$ provides as much information for optimization purposes in the long run as a full set of one at a time changes of each component. Thus, it uses only two evaluations of the objective function, in a direction randomly chosen at each iteration, to form an approximation to the gradient vector.

Gradient-free algorithms: these include, for example, evolutionary algorithms, direct search and objective function approximation methods and are based on comparisons of design choices that are distributed in large regions of the design space. They require information only for the objective function which makes them highly appropriate for stochastic optimizations [18, 19] because they avoid the difficulty of obtaining derivative information. They involve, though, significant computational effort if the dimension of the design variable space is high. Use of CRN in these algorithms may only improve the efficiency of the comparisons in special cases; for example, if the size (volume) of the design space is "relatively small" and thus the design variables being compared are always close to each other.

More detailed discussion of algorithms for stochastic optimization can be found in [4] and [5]. Only SPSA using CRN is briefly summarized here.

3.3 Simultaneous perturbation stochastic approximation using common random numbers

The implementation of SPSA takes the iterative form:

$$\boldsymbol{\varphi}_{k+1} = \boldsymbol{\varphi}_k - \alpha_k \mathbf{g}_k\left(\boldsymbol{\varphi}_k, \boldsymbol{\Omega}_N\right) \tag{24}$$

where $\boldsymbol{\varphi}_1 \in \Phi$ is the chosen point to initiate the algorithm and the j^{th} component for the CRN simultaneous perturbation approximation to the gradient vector in the k^{th} iteration, $\mathbf{g}_k(\boldsymbol{\Omega}_{N,k},\boldsymbol{\varphi})$, is given by:

$$g_{k,j} = \frac{\hat{E}_{\boldsymbol{\theta},N}\left(\boldsymbol{\varphi}_k + c_k\boldsymbol{\Delta}_k, \boldsymbol{\Omega}_{N,k}\right) - \hat{E}_{\boldsymbol{\theta},N}\left(\boldsymbol{\varphi}_k - c_k\boldsymbol{\Delta}_k, \boldsymbol{\Omega}_{N,k}\right)}{2c_k\Delta_{\kappa,j}} \tag{25}$$

where $\boldsymbol{\Delta}_\kappa \in \mathbb{R}^{n_\varphi}$ is a vector of mutually independent random variables that defines the random direction of simultaneous perturbation for $\boldsymbol{\varphi}_k$ and that satisfies the statistical properties given in [4]. A symmetric Bernoulli ± 1 distribution is typically chosen for the components of $\boldsymbol{\Delta}_\kappa$. The selection for the sequences $\{c_k\}$ and $\{\alpha_k\}$ is discussed in detail in [17]. A choice that guarantees asymptotic convergence to $\boldsymbol{\varphi}^*$ is $\alpha_k=\alpha/(k+w)^\beta$ and $c_k=c_1/k^\zeta$, where $4\zeta-\beta>0$, $2\beta-2\zeta>1$, with $w,\zeta>0$ and $0<\beta\leq 1$. This selection leads to a rate of convergence that asymptotically approaches $k^{-\beta/2}$ when CRN is used [17]. The asymptotically optimal choice for β is, thus, 1. In applications where efficiency using a small number of iterations is sought after, use of smaller values for β are suggested in [4]. For complex structural design optimizations, where the computational cost for each iteration of the algorithm is high, the latter suggestion should be adopted.

Regarding the rest of the parameters for the sequences $\{c_k\}$ and $\{\alpha_k\}$: w is typically set to 10% of the number of iterations selected for the algorithm and the initial step c_1 is chosen "close" to the standard deviation of the measurement error $e_N(\boldsymbol{\Omega}_N,\boldsymbol{\varphi}_1)$. The value of α can be determined based on the estimate of $\mathbf{g}_1$ and the desired step size for the first iteration. Some initial trials are generally needed in order to make a good selection for α, especially when little prior insight is available for the sensitivity of the objective function to each of the design variables. Convergence of the iterative process is judged based on the value $\left\|\boldsymbol{\varphi}_{k+1}-\boldsymbol{\varphi}_k\right\|$ in the last few steps, for an appropriate selected vector norm. Blocking rules can also be applied in order to avoid potential divergence of the algorithm, especially in the first iterations (see [4] for more details).

4. An efficient framework for stochastic design using stochastic simulation

4.1 Framework Outline

As already mentioned, a two-stage framework for stochastic system design may be established by combining the algorithms presented in the previous two sections. In the first stage, SSO is implemented in order to explore the sensitivity of the objective function and adaptively identify a subset $I_{SSO}\subset\Phi$ containing the optimal design variables. In the second stage, any appropriate stochastic optimization algorithm is implemented in order to pinpoint the optimal solution within I_{SSO}. The specific algorithm selected for the second stage determines the level of quality that should be established in the SSO identification. If a method is chosen that is restricted to search only within I_{SSO} (typically characteristic of gradient-free methods), then better quality is needed. The iterations of SSO should stop at a larger size set, and establish greater plausibility that the identified set includes the optimal design point. If, on the other hand, a method is selected that allows for convergence to points outside the identified set, lower quality may be acceptable in the identification. Our experience indicates that a value around 0.8 for $\hat{H}(\hat{I}_k)$ with a c.o.v. of 4% for that estimate, indicates high certainty that $\hat{I}_k$ includes the optimal solution. Of course, this depends on the characteristics of the problem too and particularly on the selection of the shape of admissible subsets, but this guideline has proved to be robust in the applications we have considered so far.

The efficiency of the stochastic optimization considered in the second stage is influenced by (a) the size of the design space Φ defined by its volume V_Φ, and, depending on the characteristics of the algorithm chosen, by (b) the initial point $\boldsymbol{\varphi}_1$ at which the algorithm is initiated, and (c) the knowledge about the local behavior of the objective function in Φ. The SSO stage gives valuable insight for all these topics and can, therefore, contribute to increasing the efficiency of convergence to the optimal solution $\boldsymbol{\varphi}^*$ (as has been illustrated in [7, 8]). The set I_{SSO} has smaller size (volume) than the original design space Φ. Also, it is established that the sensitivity of the objective function with respect to all components of $\boldsymbol{\varphi}$ is small and, depending on the selection of the shape of admissible subsets, the correlation between the design variables is identified (see also discussion in the example later). This allows for efficient normalization of the design space (in selecting step sizes) or choice of approximating functions (for example, for objective function approximation methods).

In particular for tuning the SPSA algorithm presented in Section 3.2 using information from SSO, the following guidelines can be applied: $\boldsymbol{\varphi}_1$ should be selected as the center of the set I_{SSO} and parameter a chosen so that the

initial step for each component of $\boldsymbol{\varphi}$ is smaller than a certain fraction (chosen as 1/10) of the respective size of I_{SSO}, based on the estimate for $\mathbf{g}_1$. This estimate should be averaged over n_g (chosen as 6) evaluations because of its importance in the efficiency of the algorithm. Also, no movement in any direction should be allowed that is greater than a quarter of the size of the respective dimension of I_{SSO} (blocking rule).

The information from the SSO stage can also be exploited in order to reduce the variance of the estimate $\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)]$ by using importance sampling. This choice is discussed next.

4.2 Importance sampling

Importance sampling (IS) [11] is an efficient variance reduction technique. It is based on choosing an importance sampling density $p_{is}(\boldsymbol{\theta}|\boldsymbol{\varphi})$, to generate samples in those regions of Θ that contribute more to the integral of $E_{\boldsymbol{\theta}}[h(\boldsymbol{\varphi},\boldsymbol{\theta})]$. An estimate for the integral in (1) is given in this case by:

$$\hat{E}_{\boldsymbol{\theta},N}[h(\boldsymbol{\varphi},\boldsymbol{\Omega}_N)] = \frac{1}{N}\sum_{i=1}^{N} h(\boldsymbol{\varphi},\boldsymbol{\theta}_i) R(\boldsymbol{\theta}_i \mid \boldsymbol{\varphi}) \tag{26}$$

where the samples $\boldsymbol{\theta}_i$ are simulated according to $p_{is}(\boldsymbol{\theta}|\boldsymbol{\varphi})$ and $R(\boldsymbol{\theta}_i|\boldsymbol{\varphi})=p(\boldsymbol{\theta}_i|\boldsymbol{\varphi})/p_{is}(\boldsymbol{\theta}_i|\boldsymbol{\varphi})$ is the importance sampling quotient. The main problem is how to choose a good IS density, i.e. a density that reduces the variability of the estimate in (26). The optimal density is simply the PDF that is proportional to the absolute value of the integrand of (1) [11]:

$$p_{is,opt}(\boldsymbol{\theta} \mid \boldsymbol{\varphi}) = \frac{|h(\boldsymbol{\varphi},\boldsymbol{\theta})|\, p(\boldsymbol{\theta} \mid \boldsymbol{\varphi})}{E_{\boldsymbol{\theta}}[|h(\boldsymbol{\varphi},\boldsymbol{\theta})|]} \propto |h(\boldsymbol{\varphi},\boldsymbol{\theta})|\, p(\boldsymbol{\theta} \mid \boldsymbol{\varphi}) \tag{27}$$

Samples for $\boldsymbol{\theta}$ that are distributed proportional to $h_s(\boldsymbol{\varphi},\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\varphi})$ when $\boldsymbol{\varphi} \in I_{SSO}$ are available from the last iteration of the SSO stage. They simply correspond to the $\boldsymbol{\theta}$ component of the available sample pairs $[\boldsymbol{\varphi},\boldsymbol{\theta}]$. Re-sampling can be performed within these samples, using weighting factors $|h(\boldsymbol{\varphi}_i,\boldsymbol{\theta}_i)|/h_s(\boldsymbol{\varphi}_i,\boldsymbol{\theta}_i)$ for each sample, in order to simulate samples from $|h(\boldsymbol{\varphi}_i,\boldsymbol{\theta}_i)|p(\boldsymbol{\theta}|\boldsymbol{\varphi})$ when $\boldsymbol{\varphi} \in I_{SSO}$. The efficiency of this re-sampling procedure depends on how different $h_s(\boldsymbol{\varphi}_i,\boldsymbol{\theta}_i)$ and $h(\boldsymbol{\varphi}_i,\boldsymbol{\theta}_i)$ are. In most cases the difference will not be significant, or even zero, and good efficiency can be established. Alternatively, $h_s(\boldsymbol{\varphi},\boldsymbol{\theta})$ can be used as performance measure in the second stage of the optimization; this choice would be inappropriate if s was negative because it makes the loss function less sensitive to the uncertain parameters $\boldsymbol{\theta}$, thus possible reducing the efficiency of IS.

The samples simulated from $h(\boldsymbol{\varphi},\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\varphi})$ can be finally used to create an importance sampling density $p_{is}(\boldsymbol{\theta})$ to use for $p_{is}(\boldsymbol{\theta}|\boldsymbol{\varphi})$, since the set I_{SSO} is relatively small. This last feature is important; it means that the different system configurations compared are not too different and thus the suggested IS densities will be efficient for all of them, i.e. for all $\boldsymbol{\varphi}$ in I_{SSO}. Reference [20] provides a method for creating IS densities using samples of the model parameters.

For problems with a high-dimensional vector $\boldsymbol{\theta}$, efficiency of IS can be guaranteed only under strict conditions [21]. An alternative approach can be applied for such cases: the uncertain parameter vector is partitioned into two sets, Θ_1 and Θ_2. Θ_1 is comprised of parameters that individually do not significantly influence the loss function (they have significant influence only when viewed as a group), while Θ_2 is comprised of parameters that have individually a strong influence on $h(\boldsymbol{\varphi},\boldsymbol{\theta})$. The latter set typically corresponds to a low-dimensional vector. IS is applied for the elements of Θ_2 only. This approach is similar to the one discussed in [22] and circumvents the problems that may appear when applying IS to design problems involving a large number of uncertain parameters.

5. Illustrative example: reliability based design for a base-isolated structure

The proposed framework is illustrated in an example considering the reliability-based optimization of a base-isolation system for a three-story structure. Initially some brief comments are presented for reliability optimization problems (ROP) and the special characteristics they involve. We use the terminology ROP to distinguish the approach taken here, which is the direct optimization of the reliability of the system, from common Reliability-Based Design Optimization problems, where the reliability requirements are forced by restrictions put on the admissible design space.

5.1 Reliability-based optimal design

In a reliability context, the overall performance of a system is quantified in terms of the plausibility of the occurrence of unacceptable performance, i.e."failure", based on all available information. This probability is expressed as:

$$P(F \mid \boldsymbol{\varphi}) = E_{\boldsymbol{\theta}}[I_F(\boldsymbol{\varphi},\boldsymbol{\theta})] = \int_{\Theta} I_F(\boldsymbol{\varphi},\boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \boldsymbol{\varphi})\, d\boldsymbol{\theta} \tag{28}$$

where $I_F(\boldsymbol{\varphi},\boldsymbol{\theta})$ is the indicator function of failure, which is 1 if the system that corresponds to $(\boldsymbol{\varphi},\boldsymbol{\theta})$ fails, i.e. its response departs from the acceptable performance set, and 0 if it does not. A limit state function $\tilde{g}(\boldsymbol{\varphi},\boldsymbol{\theta})$ is typically used to evaluate the model performance. If $\tilde{g}(\boldsymbol{\varphi},\boldsymbol{\theta})>0$ then the model is considered to perform unacceptably, i.e. fail. Reference [23] provides useful information for estimation of the integral in (28) using stochastic simulation. It discusses a highly efficient algorithm, called Subset Simulation, for estimation of small failure probabilities. Also it provides guidelines for efficient MCMC sampling in reliability problems, that can be useful in the SSO setting for obtaining samples

from $\pi(\boldsymbol{\varphi},\boldsymbol{\theta})$. Note that for reliability problems, $\pi(\boldsymbol{\varphi},\boldsymbol{\theta})$ simply corresponds to the conditional on system failure pdf $p(\boldsymbol{\varphi},\boldsymbol{\theta}|F)$ and the sample pairs for $[\boldsymbol{\varphi},\boldsymbol{\theta}]$ that are distributed according to $\pi(\boldsymbol{\varphi},\boldsymbol{\theta})$ can be simply considered as failed samples.

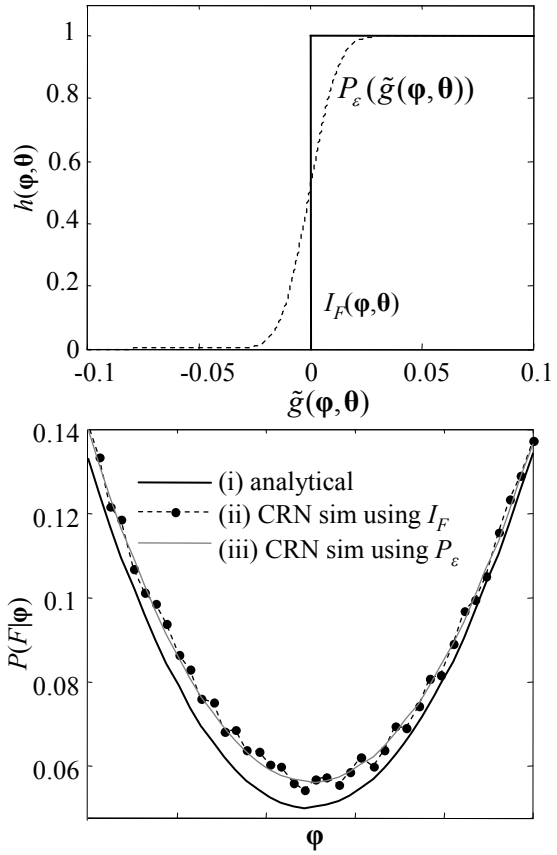

Figure 4. Comparison between (a) the two candidate performance measures for reliability objective problems and (b) the objective function estimates obtained through use of CRN for these two candidates.

An important concept for reliability-based design is the existence of a model prediction error ε, i.e. an error between the model-predicted response and the actual system response. This error can be model probabilistically as a random variable [24] and added to $\boldsymbol{\theta}$ to form an augmented uncertain parameter vector, comprised of both the model parameters and the model prediction error. It is possible in some cases to analytically evaluate the part of the integral (28) that corresponds to this prediction error: let $g(\boldsymbol{\varphi})>0$ and $\tilde{g}(\boldsymbol{\varphi},\boldsymbol{\theta})>0$ be the limit state quantities defining the system's and model's failure respectively, and define ε so that the relationship $\tilde{g}(\boldsymbol{\varphi},\boldsymbol{\theta})=g(\boldsymbol{\varphi})+\varepsilon(\boldsymbol{\varphi},\boldsymbol{\theta})$ holds. Then, if $P_\varepsilon(.)$ is the cumulative distribution function for the model prediction error $\varepsilon(\boldsymbol{\varphi},\boldsymbol{\theta})$, the failure probability can be equivalently expressed as [7]:

$$P(F\,|\,\boldsymbol{\varphi}) = E_{\boldsymbol{\theta}}[P_\varepsilon(\tilde{g}(\boldsymbol{\varphi},\boldsymbol{\theta}))] = \int_\Theta P_\varepsilon(\tilde{g}(\boldsymbol{\varphi},\boldsymbol{\theta}))\,p(\boldsymbol{\theta}\,|\,\boldsymbol{\varphi})\,d\boldsymbol{\theta} \qquad (29)$$

where in this case the vector $\boldsymbol{\theta}$ corresponds solely to the uncertain parameters for the system and excitation model, i.e. excluding the prediction error. Thus, the performance measure in optimal reliability problems corresponds either to (a) $h(\boldsymbol{\varphi},\boldsymbol{\theta})=I_F(\boldsymbol{\varphi},\boldsymbol{\theta})$ or (b) $h(\boldsymbol{\varphi},\boldsymbol{\theta})=P_\varepsilon(\tilde{g}(\boldsymbol{\varphi},\boldsymbol{\theta}))$, depending on which formulation is adopted, (28) or (29). In Figure 4(a) these two loss functions are compared when ε is modeled as a Gaussian variable with mean 0 and standard deviation 0.01.

Note that since the indicator function $I_F(\boldsymbol{\varphi},\boldsymbol{\theta})$ is discontinuous, the requirements that were discussed in Section 3.1 for establishing efficiency for usage of CRN cannot be guaranteed when formulation (28) is adopted. It is thus beneficial to use the formulation (29) for the probability of failure in CRN-based optimizations. For design problems where no prediction error in the model response is actually assumed, a small fictitious error should be chosen so that the optimization problems with and without the model prediction error are practically equivalent, i.e. correspond to the same optimum. Figure4(b) illustrates this concept; the influence on $P(F|\boldsymbol{\varphi})$ of the two different performance measures and the advantage of selecting $P_\varepsilon(\tilde{g}(\boldsymbol{\varphi},\boldsymbol{\theta}_s))$ is clearly demonstrated.

5.2 Structural and excitation models

A three-story symmetric building is considered and it is modeled (Figure 5) as a planar shear building with uncertain inter-story stiffness and uncertain classical modal damping. The lumped mass of the top story is 636 ton while it is 817 ton for the bottom two. The inter-story stiffnesses k_i of all the stories are parameterized by $k_i=\theta_{k,i}\hat{k}_i$, i=1,2,3, where $[\hat{k}_i]$=[633.9, 443.7, 253.6] MN/m are the most probable values for the inter-story stiffness, and $\theta_{k,i}$ are nondimensional uncertain parameters, assumed to be correlated Gaussian variables with mean value one and covariance matrix with elements:

$$\mathbf{K}_{ij} = (0.1)^2 \exp[-(i-j)^2/2^2] \qquad (30)$$

that roughly imply significant correlation between inter-story stiffness's within two stories apart and c.o.v. of 10%. The damping ratios for the modes are modeled as independent Gaussian variables with mean value 5% and coefficient of variation 10%.

A base-isolation system, with lead-rubber bilinear isolators and supplemental viscous dampers at the base level, is designed for the protection of the superstructure. The mass of the base is 999 ton. The pre-yield, K_{pr}, and post-yield, K_p, stiffness and the yield force, F_y, are the design variables $\boldsymbol{\varphi}$ for the isolation system along with the damping coefficient, c_d, for the dampers. A simplified problem with only two design variables is also formulated

by setting the post-yield stiffness equal to 15% of the pre-yield stiffness and the viscous dampers to 5% critical damping assuming a nominal period of 2.25 sec for the isolated structure (thus only K_{pr} and F_y are design variables in this case). The simplified problem is denoted by D_1 and the full one by D_2.

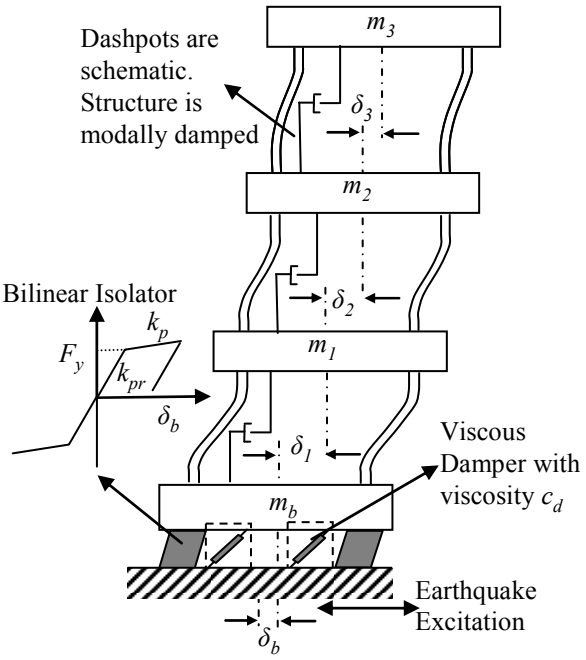

Figure 5: Structural model assumed in the study

In order to estimate the structural system reliability, probability models are established for the seismic hazard at the structural site, which corresponds to potentially damaging future near-fault ground motions (since we discuss base-isolated structures). The uncertainty in moment magnitude for seismic events, M, is modeled by the Gutenberg-Richter relationship [25] truncated on the interval $[M_{min}, M_{max}]=[6, 8]$, leading to a PDF :

$$p(M) = \frac{b\exp(-bM)}{\exp(-bM_{\min}) - \exp(-bM_{\max})} \qquad (31)$$

with regional seismicity factors selected as $b=0.7\log_e(10)$. For the uncertainty in the event location, the logarithm of the epicentral distance, r, for the earthquake events is assumed to follow a Gaussian distribution with mean log(20) km and standard deviation 0.5. Figure 7(a) illustrates the PDFs for M and r.

For the ground motion, the probabilistic model described in detail in [26] is adopted: the high-frequency and low-frequency (long-period) components of the earthquake excitation are separately modeled and then combined to form the acceleration input. The high-frequency component is modeled by the stochastic method (see [27] for more details) which involves modifying a white-noise sequence $\mathbf{Z}_w$ by (i) a time-domain envelope function and (ii) a frequency-domain filter, that are both expressed as nonlinear functions of the moment magnitude and the epicentral distance of the seismic event. The specific characteristics for these two steps are the ones described in [26]. The long period (pulse-like) component of near-fault ground motions is modeled by the pulse model presented in [28], which involves a simple analytical expression for the time history of the pulse. The magnitude and the frequency of the pulse are chosen according to the probabilistic models for near-field pulses in rock sites given in [29] (the logarithms of both are modeled as Gaussian distributions). The number of half-cycles, γ, and phase, ν, for the near-fault pulse are additional model parameters. The probability models for these parameters are chosen, respectively, as Gaussian with mean 1.8 and standard deviation 0.3 and uniform in the range $[-\pi/2, \pi/2]$. Figure 6 illustrates a sample excitation generated according to this ground motion model. The existence of the near-field pulse is evident in the velocity time history.

The uncertain parameter vector in this design problem consists of the structural model parameters, $\boldsymbol{\theta}_s$, the seismological parameter $\boldsymbol{\theta}_g=[M, r]$, the additional parameters for the near-fault pulse $\boldsymbol{\theta}_p$ and the white-noise sequence, $\mathbf{Z}_w$, so $\boldsymbol{\theta}=[\boldsymbol{\theta}_s, \boldsymbol{\theta}_g, \boldsymbol{\theta}_p, \mathbf{Z}_w]$. Note that $\mathbf{Z}_w$ corresponds to a 5001-dimensional vector in this case (partitioning of a 50 sec time window into intervals of 0.01sec).

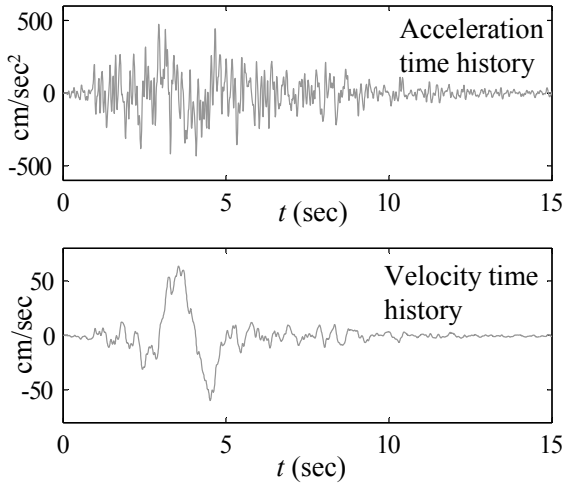

Figure 6. Sample near-field earthquake excitation

Failure for the system is defined to be that any of the inter-story drifts, base displacement or shear force at the first story exceeds the thresholds 0.0105m, 0.25 m and 0.24g of the superstructure mass, respectively. Function $\tilde{g}(\boldsymbol{\varphi}, \boldsymbol{\theta})$ is defined as the logarithm of the maximum over the excitation duration of these performance variables (normalized by their respective threshold). A small model prediction-error is assumed that is Gaussian with mean 0 and standard deviation 0.05. The initial design interval for each design variable is defined as $K_{pr} \in [10,240]$ MN/m,

$F_y \in [1,5]$ MN, $K_p \in [1,50]$ MN/m and $c_d \in$ [0.1,10] MNsec/m.

5.3 Optimization algorithm characteristics

The two-stage framework discussed in section 4.1 is used. For the SSO algorithm (Section 2.5), the parameter selections are: ρ=0.2, N=3000. The shape for the admissible sets I is selected as a hyper-ellipse and the adaptive identification is stopped when $\hat{H}(\hat{I}_k)$ becomes larger than 0.8. For the non-smooth optimization of (21) an algorithm based on direct-search is chosen [30].

SPSA with CRN (Section 3.3) is adopted for the second stage of the optimization, with parameter selection: β=0.71, ζ=0.25, N=1500. Convergence is judged by looking at the norm $\|\boldsymbol{\varphi}_{k+1} - \boldsymbol{\varphi}_k\|_\infty$ for each of the five last iterations. If that norm is less than 0.2% (normalized by the dimension of the initial design space), then we assume that convergence to the optimal solution has been established. Formulation (29) is used in the SPSA optimization in order to improve the efficiency of CRN comparisons, as discussed earlier. The guidelines discussed in section 4.1 are adopted for selection of step sizes and blocking rules for SPSA. To implement these guidelines, normalization of the search space is performed so that the hyper-ellipse I_{SSO} is transformed into a unit radius hyper-sphere.

For the second optimization stage (SPSA), following the discussion in Section 5.2, importance sampling densities are established for the structural, near-field pulse model parameters and the seismological parameters, but not for the high-dimensional white-noise sequence $\mathbf{Z}_w$. Apart from the phase of the near-field pulse, v, for which the samples from $\pi(v)$ were found to be distributed similar to $p(v)$ (indicating that this model parameter has small influence on the response of the model), for the rest of the parameters, the IS PDFs were approximated by Gaussian distributions, like the prior distributions $p(\boldsymbol{\theta}_i)$, but with a shifted mean value and variance, selected as the mean and variance of the samples from the SSO stage that are distributed according to $\pi(\boldsymbol{\varphi})$. Figure 7(b) illustrates this concept for M and r for problem D_1. Note that the IS PDF for M and r are significantly different from their initial distribution; since these seismological parameters are expected to have a strong influence on the model response, the difference between the distributions is expected to have a big effect on the accuracy of the estimation. For problem D_1 the c.ov. for $\hat{P}(F|\boldsymbol{\varphi})$ for a sample size N=1500 is 12.5% without using IS and 4.4% when IS is used. This c.o.v. is of the same level for all values of $\boldsymbol{\varphi} \in I_{SSO}$ (since the I_{SSO} set is relatively small good efficiency is achieved). Note that the c.o.v. varies according to $1/\sqrt{N}$ [11]; thus, the sample size for direct estimation (i.e. without use of IS) of $\hat{P}(F|\boldsymbol{\varphi})$ with the same level of accuracy as in the case when IS is applied would be approximately 8 times larger. This illustrates the efficiency increase that can be established by the IS scheme discussed earlier. Similarly, for problem D_2 the c.o.v. is 4.9% when IS was used and 16.2% when not. In this case a sample size that is 11 times larger is needed for direct estimation of $\hat{P}(F|\boldsymbol{\varphi})$ for establishing the same accuracy as with the IS scheme.

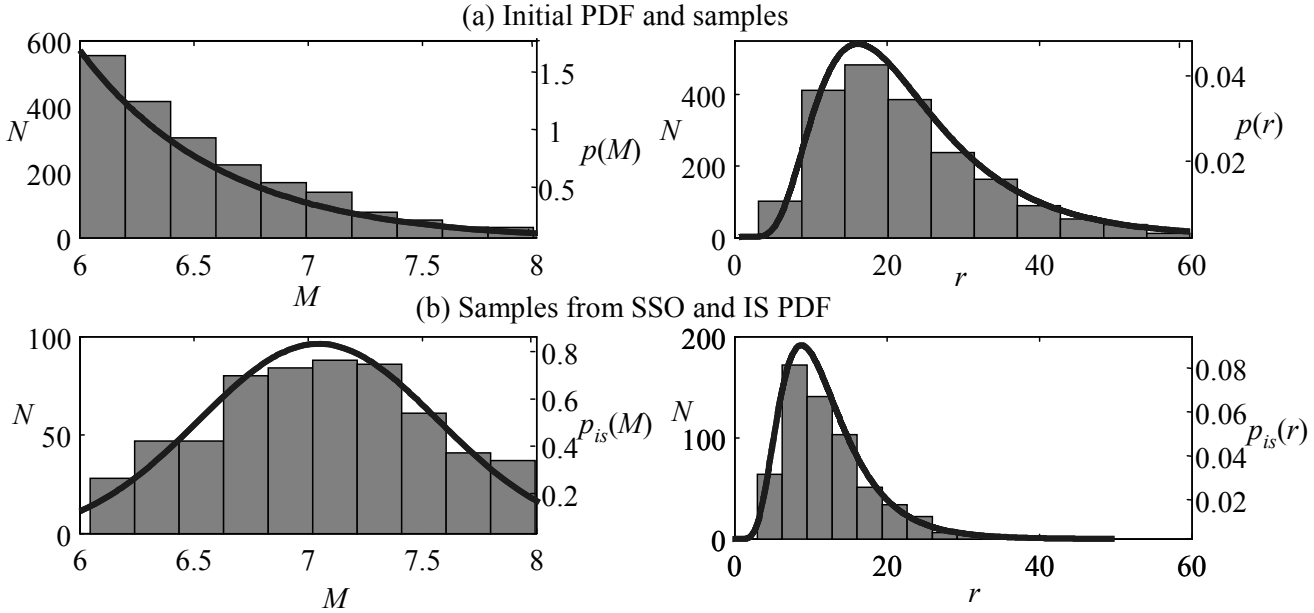

Figure 7. Details about importance sampling densities formulation for M and r (a) Initial PDF and samples and (b) samples from SSO stage and chosen IS densities for D1.

Table 1: Results from a sample run of the SSO algorithm for two design problems

Problem D_1 $n_\varphi=2$, $\hat{I}_{SSO}=\hat{I}_3$	Iteration of SSO 1	2	3	Problem D_2 $n_\varphi=4$, $\hat{I}_{SSO}=\hat{I}_6$	Iteration of SSO 1	2	3	4	5	6
$V_{\hat{I}_k} / V_{\hat{I}_{k-1}}$	0.381	0.265	0.241	$V_{\hat{I}_k} / V_{\hat{I}_{k-1}}$	0.286	0.345	0.305	0.271	0.259	0.230
$\sqrt[n_\varphi]{V_{\hat{I}_k} / V_{\hat{I}_{k-1}}}$	0.617	0.514	0.491	$\sqrt[n_\varphi]{V_{\hat{I}_k} / V_{\hat{I}_{k-1}}}$	0.731	0.7676	0.743	0.722	0.713	0.693
$\hat{H}(\hat{I}_k)$	0.525	0.754	0.830	$\hat{H}(\hat{I}_k)$	0.698	0.580	0.657	0.738	0.772	0.865
$V_{\hat{I}_{SSO}} / V_{\Phi}$	0.0243			$V_{\hat{I}_{SSO}} / V_{\Phi}$	$4.85\ 10^{-4}$					
$\sqrt[n_\varphi]{V_{\hat{I}_{SSO}} / V_{\Phi}}$	0.156			$\sqrt[n_\varphi]{V_{\hat{I}_{SSO}} / V_{\Phi}}$	0.149					

5.4 Results and discussion

Results for a sample run are presented in Table 1 for the SSO and in Table 2 and Figures 8 and 9 for the combined optimization framework. The SSO algorithm converged in 3 iterations for problem D_1 and in 6 iterations for problem D_2 to sets that are characterized by small sensitivity to the objective function (I_{sso} in Figures 8 and 9 with center $\boldsymbol{\varphi}_{SSO}$). SPSA was then used to pinpoint the optimal solution, $\boldsymbol{\varphi}^*$, within these sets (point X in the aforementioned figures). The results of this sample optimization run are discussed in more detail next, focusing on the aspects related to the novel SSO algorithm.

Table 2: Cumulative results for a sample run of the stochastic optimization framework

		$\boldsymbol{\varphi}_{SSO}$	$\hat{P}(F \mid \boldsymbol{\varphi}_{SSO})$	$\boldsymbol{\varphi}^*$	$\hat{P}(F \mid \boldsymbol{\varphi}^*)$
D_1	K_{pr}	98.05	0.0391	92.75	0.0386
	F_y	2.35		2.44	
D_2	K_{pr}	68.2	0.0240	53.2	0.0229
	F_y	1.76		1.92	
	K_p	15.56		13.93	
	c_d	4.21		4.24	

SSO leads to a significant reduction of the size (volume) of the search space for both design problem; this is evident in both Table 1 (last two rows) as well as Figures 8 and 9. These two figures and Table 2 also illustrate that SSO efficiently identifies a subset for the optimal design variables; the converged optimal solution in the second stage, $\boldsymbol{\varphi}^*$, is close to the center of the set that is identified by SSO, $\boldsymbol{\varphi}_{SSO}$; also the objective function at that center point $\hat{P}(F \mid \boldsymbol{\varphi}_{SSO})$ is not significantly different than the optimal value $\hat{P}(F \mid \boldsymbol{\varphi}^*)$ (Table 2). Thus, selection of $\boldsymbol{\varphi}_{SSO}$ as the design choice leads to a sub-optimal design but close to the true optimum in terms of both the design vector selection and its corresponding performance. All these characteristics illustrate the effectiveness and quality of the set identification in SSO. Note that as the algorithm evolves, this quality, expressed by $\hat{H}(\hat{I}_k)$, decreases (Table 1). Within I_{SSO}, the small sensitivity of the objective function to $\boldsymbol{\varphi}$ cannot be easily captured by samples obtained by stochastic simulation, unless a large number of them are available. Instead, SPSA is chosen here for pinpointing the optimal design variables.

For the design problem D_2 the difference of the shapes of the initial design space and the admissible subsets is considerable (difference of volume of four-dimensional hyper-rectangles and the inscribed hyperellipses). As discussed earlier, this leads to some small loss of quality for the first stage of the identification; $\hat{H}(\hat{I}_k)$ in the first iteration is larger than the second one, though typically it is expected to be a decreasing function of the iteration number [7]. This feature does not influence, though, the overall quality of the SSO identification as evident by the rest of the results.

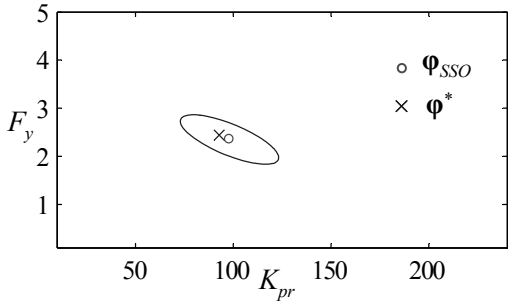

Figure 8. Sets I_{SSO} (ellipse) and Φ (rectangle) for problem D_1

The ability of SSO to estimate contours of the objective function and capture the correlation between the design variables is illustrated in Figures 8 and 9. This correlation can be demonstrated also by looking at the

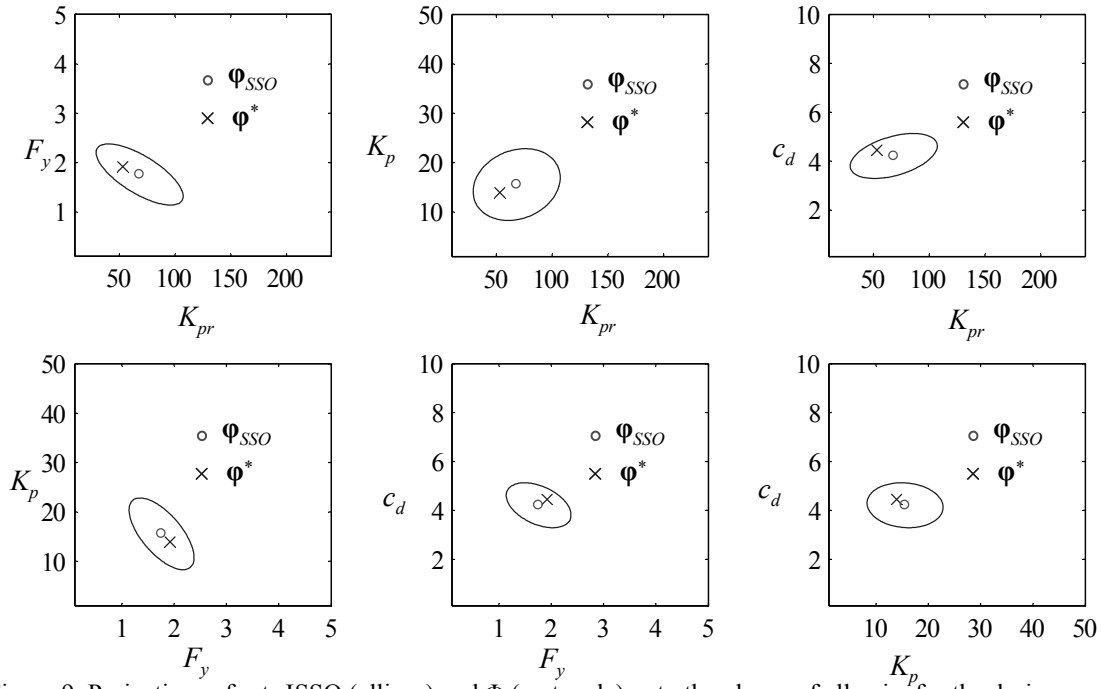

Figure 9. Projections of sets ISSO (ellipse) and Φ (rectangle) onto the planes of all pairs for the design parameters for problem D2

normalized positive definite matrix, **M**, that defines the ellipses:

$$\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_{SSO}\right)^{T}\mathbf{D}^{1/2}\mathbf{M}^{-1}\mathbf{D}^{1/2}\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_{SSO}\right)=1 \tag{32}$$

where normalization is used in order to make all diagonal elements of **M** be unity. The off-diagonal elements of **M** show the correlation between the different design variables (similar to the concept of correlation for a Gaussian PDF); the larger the absolute values of these elements the more important the correlation between the design variables. For the sample run of SSO discussed in this section, these matrices are respectively:

$$\mathrm{D}_1:\mathbf{M}=\begin{array}{c} \begin{matrix} K_{pr} & F_y \end{matrix} \\ \begin{bmatrix} 1 & -0.74 \\ -0.74 & 1 \end{bmatrix} \begin{matrix} K_{pr} \\ F_y \end{matrix} \end{array} \tag{33}$$

$$\mathrm{D}_2:\mathbf{M}=\begin{array}{c} \begin{matrix} K_{pr} & F_y & K_p & c_d \end{matrix} \\ \begin{bmatrix} 1 & -0.71 & 0.19 & 0.45 \\ -0.71 & 1 & -0.66 & -0.43 \\ 0.19 & -0.66 & 1 & -0.09 \\ 0.45 & -0.43 & -0.09 & 1 \end{bmatrix} \begin{matrix} K_{pr} \\ F_y \\ K_p \\ c_d \end{matrix} \end{array} \tag{34}$$

The correlation between pairs of design variables indicated by these matrices can be visually verified in Figures 8 and 9; the highly-correlated are variables K_{pr} and F_y in design problem D_1 and pairs of variables K_{pr}-F_y, K_{pr}-c_d, F_y-K_p and F_y-c_d in problem D_2. These higher correlations mean that these pairs of design variables can be traded off without significantly changing the objective function. Note, also, that the correlation relationship between K_{pr} and F_y is the same in both design problems.

The influence of the number of the design variables in the efficiency of SSO, now, is evident when comparing the results in Table 1 between the two design cases. For D_2 the average reduction of the size for each design variable (second row of Table 1) is much smaller which leads to more iterations until a set with small sensitivity to the objective function is identified. The proportionality dependence with regard to the computational cost of SSO argued in Section 2.3 is also verified in the context of the example discussed here: the mean total length reduction over all iteration for D_1 and D_2, corresponding to the quantity in equation (20), is similar (look at last row in Table 1) but the number of required iterations for convergence to the I_{sso} set for D_2 (which has twice as many design variables) is only double.

A comment with respect to the effectiveness of the base isolation protection system in this example is, finally, warranted. Design D_2 involves much larger versatility in the selection of the characteristics of the isolation system, and thus leads to smaller failure probability, as expected. The period of the base isolated structures for a displacement of base displacement of 25cm is 2.4 sec for D_1 and 2.47 sec for D_2. These selections seem reasonable based on common design practice for base isolation systems and verify the validity of the stochastic framework chosen (system and excitation models, modeling uncertainty and objective function formulation).

5.5 Efficiency of optimization framework

The efficiency of the suggested optimization framework is judged by performing the same optimization by applying only SPSA, without using any information from SSO; the optimization in this case is performed with respect to the initial design space, Φ, by randomly selecting the starting point for the algorithm. In this case IS is not implemented; since search inside the whole design space Φ is considered, it is unclear how samples of $\boldsymbol{\theta}$ can be obtained to form the IS densities and separately establishing an IS density for each design choice $\boldsymbol{\varphi}$ is too computationally expensive. Larger values for the sample size, N, are chosen in order to establish similar c.o.v. near the optimal solution as in the case that IS is implemented; according to previous discussion the selection for N is 8 times larger than in the two-stage approach for problem D_1 and 11 for D_2 . In 10 runs, the average number of iterations (and its standard deviation) for convergence to the optimal solution was (a) for D_1: with SSO 20.42 (4.5) and without 59.5 (21.4) and (b) for D_2: with SSO 30.5 (9.4) and without 98.4 (36.9). In order to evaluate the effectiveness of the two-stage framework discussed, note that the computational cost of one iteration of SSO, in terms of system simulations needed, is equal to the computational cost of one iteration of SPSA; thus, one can simply add three and six additional iterations to the means for problems D_1 and D_2 in (a) and (b), respectively, when SSO is used. This comparison illustrates the effectiveness of the proposed two-stage optimization framework to enhance the efficiency of stochastic optimization. It should also be noted that use of SSO leads to more consistency in the optimization efficiency (smaller variation). The better starting point of the algorithm, as well as the smaller size of the search space which allows for better normalization, that SSO can provide, are the features that contribute to this improvement in efficiency. If we consider the added efficiency because of the use of IS when the combined framework is chosen, then the computational advantages from using this framework are even higher: in this example the computational cost of each iteration of SPSA is 8 and 11 times smaller for problems D_1 and D_2, respectively, in the setting of the combined framework.

5. Conclusions

In this study, we discussed robust stochastic optimization problems that entail as objective function the expected value of any system performance measure. Stochastic simulation was considered for evaluation of the system performance. This simulation-based approach allows for explicit consideration of (a) nonlinearities in the models assumed for the system and its future excitation and (b) complex system performance measures (or complex decision-theoretic utility functions). A two-stage framework for the associated optimization problem was discussed. The first stage implements an innovative approach, called Stochastic Subset Optimization (SSO), which is based on formulating an augmented design problem that artificially considers the design variables as uncertain with uniform probability distribution over the admissible design space. Then the original objective function is proportional to an ancillary PDF, samples of which can be obtained through advanced stochastic simulation techniques. Using the information contained in samples SSO efficient explores the sensitivity of the objective function to the design variables and adaptively identifies sub-regions inside the original design space that (a) have the highest likelihood of including the optimal design variables within the class of admissible subsets considered and (b) are characterized by small sensitivity with respect to each design variable (so these subsets consist of near-optimal solutions). For efficiently establishing higher accuracy for the optimal design variables, SSO can be followed by a second stage involving some other stochastic search algorithm. The information for the sensitivity of the objective function to the design variables available from the SSO stage can be used to readily tune the characteristics of the algorithms used in the second stage. This two-stage approach can provide overall enhanced efficiency and accuracy of the optimization process. Simultaneous perturbation stochastic approximation was suggested for the second stage in this study and suggestions were provided for improved efficiency of the overall framework.

With respect to SSO, guidelines were suggested for establishing good quality in the identification, for selecting the appropriate shape and size for the class of admissible subsets, and for the stopping criteria for the iterative process. An important step for the accuracy of SSO is the identification of the optimal subsets within the considered class of admissible subsets. This optimization sub-problem (within the SSO framework) is quite challenging, because it has non-smooth characteristics, and it corresponds to a potential difficulty associated with the SSO implementation. Attention needs to be given to the appropriate selection of algorithms and parameterization of the admissible subsets to ensure a reliable solution to this sub-problem. For the second stage of the two-stage optimization framework, the use of common random numbers was extensively discussed. Implementation of importance sampling for this stage was also considered by using the information available in the last iteration of SSO.

An example was presented that shows the efficiency of the proposed methodology; it considers the optimization of a base-isolation system for a three-story structure. The optimization of the reliability of the system considering future near-fault ground motions was adopted as the design objective. A realistic probability model was chosen for representing these ground motions. The

structural performance was evaluated by nonlinear simulation that incorporates all important characteristics of the behavior of the structural system and all available information about the structural model and expected future earthquakes. It was demonstrated in the context of this example that SSO can efficiently identify a set that contains the optimal design variables and is characterized by small sensitivity with respect to the objective function, and that it also can improve the efficiency of SPSA when combined as in the suggested optimization framework. Additionally, SSO was demonstrated to be able to describe the correlation between the design variables in terms of the contours for the objective function. The implementation of importance sampling using the information for SSO was also shown enhance the overall efficiency.

Appendix A: Sampling techniques

Two algorithms that can be used for simulating samples from $\pi(\boldsymbol{\varphi},\boldsymbol{\theta})$ are discussed here. The first one is the accept-reject method which may be considered to belong to direct Monte Carlo methods. First, choose an appropriate proposal PDF $f(\boldsymbol{\varphi},\boldsymbol{\theta})$ and then generate independent samples as follows:

(1) Randomly simulate candidate sample $[\boldsymbol{\varphi}_c,\ \boldsymbol{\theta}_c]$ from $f(\boldsymbol{\varphi},\boldsymbol{\theta})$ and u from uniform (0,1).

(2) Accept $[\boldsymbol{\varphi},\boldsymbol{\theta}]=[\boldsymbol{\varphi}_c,\boldsymbol{\theta}_c]$ if

$$h_s(\boldsymbol{\varphi}_c,\boldsymbol{\theta}_c)\frac{p(\boldsymbol{\varphi}_c,\boldsymbol{\theta}_c)}{Mf(\boldsymbol{\varphi}_c,\boldsymbol{\theta}_c)} > u \qquad \text{where } M > \max_{\boldsymbol{\varphi},\boldsymbol{\theta}} h_s(\boldsymbol{\varphi},\boldsymbol{\theta})\frac{p(\boldsymbol{\varphi},\boldsymbol{\theta})}{f(\boldsymbol{\varphi},\boldsymbol{\theta})} \tag{35}$$

(3) Return to (1) otherwise

The second one is the Metropolis-Hastings algorithm, which is a Markov Chain Monte Carlo method and is expressed through the iterative form:

(1) Randomly simulate a candidate sample $[\tilde{\boldsymbol{\varphi}}_{k+1},\tilde{\boldsymbol{\theta}}_{k+1}]$ from a proposal PDF $q(\tilde{\boldsymbol{\varphi}}_{k+1},\tilde{\boldsymbol{\theta}}_{k+1}|\boldsymbol{\varphi}_k,\boldsymbol{\theta}_k)$.

(2) Compute acceptance ratio:

$$r_{k+1}=\frac{h_s(\tilde{\boldsymbol{\varphi}}_{k+1},\tilde{\boldsymbol{\theta}}_{k+1},)p(\tilde{\boldsymbol{\varphi}}_{k+1},\tilde{\boldsymbol{\theta}}_{k+1},)q(\boldsymbol{\varphi}_k,\boldsymbol{\theta}_k\,|\,\tilde{\boldsymbol{\varphi}}_{k+1},\tilde{\boldsymbol{\theta}}_{k+1})}{h_s(\boldsymbol{\varphi}_{k+1},\boldsymbol{\theta}_{k+1},)p(\boldsymbol{\varphi}_{k+1},\boldsymbol{\theta}_{k+1},)q(\tilde{\boldsymbol{\varphi}}_{k+1},\tilde{\boldsymbol{\theta}}_{k+1}\,|\,\boldsymbol{\varphi}_k,\boldsymbol{\theta}_k)} \tag{36}$$

(3) Simulate u from uniform (0,1) and set

$$[\boldsymbol{\varphi}_{k+1},\boldsymbol{\theta}_{k+1}]=\begin{cases}[\tilde{\boldsymbol{\varphi}}_{k+1},\tilde{\boldsymbol{\theta}}_{k+1}] \text{ if } r_{k+1}\ge u\\ [\boldsymbol{\varphi}_k,\boldsymbol{\theta}_k] \text{ otherwise}\end{cases} \tag{37}$$

In this case the samples are correlated (the next sample depends on the previous one) but follow the target distribution after a burn-in period, i.e. after the Markov chain reaches stationarity. The algorithm is particularly efficient when samples that follow the target distribution are already available. Since the initial samples are distributed according to $\pi(\boldsymbol{\theta},\boldsymbol{\varphi})$, the Markov Chain generated in this way is always in its stationary state and all simulated samples follow the target distribution. The efficiency of both these sampling algorithms depends on the proposal PDFs $f(\boldsymbol{\varphi},\boldsymbol{\theta})$ and $q(\tilde{\boldsymbol{\varphi}},\tilde{\boldsymbol{\theta}}|\boldsymbol{\varphi},\boldsymbol{\theta})$. These PDFs should be chosen to closely resemble $h(\boldsymbol{\varphi},\boldsymbol{\theta})p(\boldsymbol{\varphi},\boldsymbol{\theta})$ and still be easy to sample from. If the first feature is established then most samples are accepted and the efficiency of the algorithm is high (see [11, 23] for more details).

References

1. E.T. Jaynes, Probability theory: the logic of science., Cambridge University Press, Cambridge, UK, 1996.
2. I. Enevoldsen and J.D. Sorensen, Reliability-based optimization in structural engineering. Structural Safety. 15 (1994) 169-196.
3. M. Gasser and G.I. Schueller, Reliability-based optimization of structural systems. Mathematical Methods of Operations Research. 46 (1997) 287-307.
4. J.C. Spall, Introduction to stochastic search and optimization, Wiley-Interscience, New York, 2003.
5. A. Ruszczynski and A. Shapiro, Stochastic programming, Elsevier, New York, 2003.
6. H.A. Jensen, Structural optimization of linear dynamical systems under stochastic excitation: a moving reliability database approach. Computer Methods in Applied Mechanics and Engineering. 194 (2005) 1757-1778.
7. A.A. Taflanidis and J.L. Beck, Stochastic subset optimization for optimal reliability problems. Probabilistic Engineering Mechanics. 24 (2008) 324-338.
8. A.A. Taflanidis and J.L. Beck, Stochastic subset optimization for stochastic design. in: Proc. *ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering*, (Rethymno, Greece, 2007).
9. A.A. Taflanidis and J.L. Beck, Reliability-based optimal design by efficient stochastic simulation. in: Proc. *5th International Conference on Computational Stochastic Mechanics*, (Rhodes, Greece, 2006).
10. S.K. Au, Reliability-based design sensitivity by efficient simulation. Computers and Structures. 83 (2005) 1048-1061.
11. C.P. Robert and G. Casella, Monte Carlo statistical methods, Springer, New York ,NY, 2004.
12. P.M. Pardalos and M.G.C. Resende, Handbook of applied optimization, Oxford University Press, 2002.
13. A.A. Taflanidis, Stochastic system design and applications to stochastically robust structural control, EERL Report 2007-5, California Institute of Technology, Pasadena, 2007.
14. P. Glasserman and D.D. Yao, Some guidelines and guarantees for common random numbers. Management Science. 38 (1992) 884-908.
15. H.J. Kushner and G.G. Yin, Stochastic approximation and recursice algorithms and applications, Springer Verlag, New York, 2003.
16. J.O. Royset and E. Polak, Reliability-based optimal design using sample average approximations. Probabilistic Engineering Mechanics. 19 (2004) 331-343.

17. N.L. Kleinmann, J.C. Spall, and D.C. Naiman, Simulation-based optimization with stochastic approximation using common random numbers. Management Science. 45 (1999) 1570-1578.
18. J.L. Beck, E. Chan, A. Irfanoglu, and C. Papadimitriou, Multi-criteria optimal structural design under uncertainty. Earthquake Engineering and Structural Dynamics. 28 (1999) 741-761.
19. N.D. Lagaros, M. Papadrakakis, and G. Kokossalakis, Structural optimization using evolutionary algorithms. Computers and Structures. 80 (2002) 571-589.
20. S.K. Au and J.L. Beck, A new adaptive importance sampling scheme. Structural Safety. 21 (1999) 135-158.
21. S.K. Au and J.L. Beck, Importance sampling in high dimensions. Structural Safety. 25 (2001) 139-163.
22. H.J. Pradlwater, G.I. Schueller, P.S. Koutsourelakis, and D.C. Champris, Application of line sampling simulation method to reliability benchmark problems. Structural Safety. (2007).
23. S.K. Au and J.L. Beck, Subset simulation and its applications to seismic risk based on dynamic analysis. Journal of Engineering Mechanics, ASCE. 129 (2003) 901-917.
24. J.L. Beck and L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework. Journal of Engineering Mechanics. 124 (1998) 455-461.
25. S.L. Kramer, Geotechnical earthquake engineering, Prentince Hall, 2003.
26. A.A. Taflanidis, J.T Scruggs and J.L. Beck, Smart base isolation design including model uncertainty in ground motion characterization. in: Proc. *4th International Conference on Earthquake Geotechnical Engineering*, (Thessaloniki, Greece, 2007).
27. D.M. Boore, Simulation of ground motion using the stochastic method. Pure and Applied Geophysics. 160 (2003) 635-676.
28. G.P. Mavroeidis and A.P. Papageorgiou, A mathematical representation of near-fault ground motions. Bulletin of the Seismological Society of America. 93 (2003) 1099-1131.
29. J.D. Bray and A. Rodriguez-Marek, Characterization of forward-directivity ground motions in the near-fault region. Soil Dynamics and Earthquake Engineering. 24 (2004) 815-828.
30. K. Holmstrom, A.O. Goran, and M.M. Edvall, User's guide for TOMLAB 5.8, Tomlab Optimization Inc. www.TOMLAB.biz, San Diego, CA, 2007.