

Global optimization for performance-based design using the Asymptotically Independent Markov Sampling Method

K. M. Zuev

University of Southern California, 3620 S. Vermont Ave, KAP 108, Los Angeles, CA 90089-2532, USA.

J. L. Beck

California Institute of Technology, MC 9-94, Pasadena, CA 91125, USA.

ABSTRACT: In this paper, we introduce a new efficient stochastic simulation method, AIMS-OPT, for approximating the set of globally optimal solutions when solving optimization problems such as optimal performance-based design problems. This method is based on Asymptotically Independent Markov Sampling (AIMS), a recently developed advanced simulation scheme originally proposed for Bayesian inference. Instead of a single approximation of the optimal solution, AIMS-OPT produces a set of nearly optimal solutions where the accuracy of the near-optimality is controlled by the user. Having a set of nearly optimal system designs can be advantageous in many practical cases such as when there exists a whole set of optimal designs or in multi-objective optimization where there is a Pareto optimal set. AIMS-OPT is also useful for efficient exploration of the global sensitivity of the objective function to the design parameters. The efficiency of AIMS-OPT is demonstrated with several examples which have different topologies of the optimal solution sets. Comparison is made with the results of applying Simulated Annealing, a well-known stochastic optimization algorithm.

1 INTRODUCTION

Consider the general problem of optimal system design under uncertainty. Suppose that various possibilities for the design of a system are defined by some controllable *design parameters* $\varphi \in \Phi \subset \mathbb{R}^{N_\varphi}$, where Φ denotes the bounded space of possible designs. Assume that a single model class ((Beck 2010)) is chosen to represent the uncertain system behavior and its uncertain future excitation, where each probability model in the class is specified by the *model parameters* $\theta \in \Theta \subset \mathbb{R}^{N_\theta}$. Since there is uncertainty in which model best describes the real system behavior, a probability density function $\pi(\theta|\varphi)$, which incorporates available knowledge about the system, is assigned to the model parameters.

Let $h : \mathbb{R}^{N_\varphi} \times \mathbb{R}^{N_\theta} \rightarrow \mathbb{R}$ denote the performance measure of the system, e.g. a utility or loss function. In *performance-based design optimization* (PBDO), the goal is then to find the optimal design φ^* that minimizes (maximizes) the expected loss (utility) function

$$\mathcal{H}(\varphi) = \mathbb{E}_\pi[h(\varphi, \theta)] = \int_\Theta h(\varphi, \theta)\pi(\theta|\varphi)d\theta, \quad (1)$$

where $\mathbb{E}_\pi[\cdot]$ denotes the expectation with respect to the distribution $\pi(\theta|\varphi)$ for θ . For example,

in reliability-based design optimization (RBDO) (Gasser & Schuëller 1997, Taflanidis & Beck 2008, Jensen, Valdebenito, & Schuëller 2008), the performance measure is $h(\varphi, \theta) = I_F(\varphi, \theta)$, where $I_F(\varphi, \theta)$ is the indicator function of the failure domain $F \subset \mathbb{R}^{N_\varphi} \times \mathbb{R}^{N_\theta}$: $I_F(\varphi, \theta) = 1$ if the system model corresponding to θ and φ fails (i.e. the model output is not acceptable according to the performance criteria) and $I_F(\varphi, \theta) = 0$ otherwise. In this case, the objective function for minimization is the probability of failure, i.e. $\mathcal{H}(\varphi) = \mathbb{P}(F|\varphi)$.

Assume for definiteness that the performance measure is interpreted as a loss function, i.e. lower values of $h(\varphi, \theta)$ correspond to better performance. The *performance-based design optimization problem* takes then the following form for the *set* of optimal designs:

$$\Phi^* = \{\varphi^*\} = \arg \min_{\varphi \in \Phi} \mathbb{E}_\pi[h(\varphi, \theta)]. \quad (2)$$

Solving the optimization problem (2) involves finding the global minimum of $\mathcal{H}(\varphi)$ which is well-known to be very challenging, especially when there may be multiple optimal solutions. Various global optimization algorithms have been devised, including Genetic Algorithm (Holland 1975), Simulated Annealing (Kirkpatrick, Gelatt, & Vecchi 1983),

Particle Swarm Optimization (Kennedy & Eberhart 1995), Generalized Trajectory Methods (Yang & Beck 1998), etc. A survey of computational methods in optimization under uncertainties is given in (Schuëller & Jensen 2008).

In this paper, we introduce a new efficient stochastic simulation method for solving the optimization problem (2) that can handle multiple optimal solutions and that was motivated by optimal performance-based design under uncertainty, although it can handle general objective functions $\mathcal{H}(\varphi)$, including those that are specified deterministically. This method is based on Asymptotically Independent Markov Sampling (AIMS), a recently developed advanced simulation scheme originally proposed for Bayesian inference (Beck & Zuev 2013). Instead of a single approximation $\hat{\varphi} \approx \varphi^*$, the new method, denoted AIMS-OPT, produces a set $\{\hat{\varphi}_1, \dots, \hat{\varphi}_n\}$ of nearly optimal solutions. Having a set of nearly optimal solutions can be advantageous in many practical situations. First, the solution of problem (2) may not be unique, i.e. there may exist a finite or infinite set $\Phi^* \subset \Phi$ of solutions, then optimization methods that produce only a single point estimate $\hat{\varphi} \approx \varphi^* \in \Phi^*$ would not provide a complete picture of the optimal solution set Φ^* . Furthermore, there may be a set of nearly optimal solutions in Φ whose objective function values differ by an inconsequential amount from the minimum value $\mathcal{H}(\varphi^*)$, $\varphi^* \in \Phi^*$, that may be of interest for reasons not quantified by the objective function $\mathcal{H}(\varphi)$.

AIMS-OPT generates a set of nearly optimal solutions $\{\hat{\varphi}_1, \dots, \hat{\varphi}_n\} \subset \Phi_T^*$ approximately uniformly distributed over a neighborhood Φ_T^* of the optimal solution set, $\Phi^* \subset \Phi_T^*$, where T is a user-specified parameter controlling the size of Φ_T^* ($\lim_{T \rightarrow 0} \Phi_T^* = \Phi^*$), and therefore defines the meaning of “nearly optimal”. AIMS-OPT is also useful for efficient exploration of the global sensitivity of the objective function to the parameters φ : at each stage k , the algorithm generates a set $\{\varphi_1^{(k)}, \dots, \varphi_n^{(k)}\} \subset \Phi_{T_k}^*$ that converges to the set of optimal solutions as $k \rightarrow \infty$: $\Phi \equiv \Phi_{T_0}^* \supset \Phi_{T_1}^* \supset \dots \supset \Phi_{T_k}^* \supset \dots \supset \Phi^*$.

The rest of the paper is organized as follows. In Section 2, the two special cases of the Metropolis-Hastings algorithm that lie at the heart of the proposed algorithm, are briefly reviewed. In Section 3, the AIMS-OPT method is introduced. The efficiency of AIMS-OPT is illustrated in Section 4 with several examples. Concluding remarks are made in Section 5.

2 MCMC SAMPLING

Markov chain Monte Carlo (MCMC), a family of stochastic simulation algorithms for sampling from arbitrary probability distributions, lies at the heart of AIMS-OPT. These algorithms are based on constructing a Markov chain whose state probability distribution converges to any desired target distribution as its

stationary distribution.

The Metropolis-Hastings algorithm (Hastings 1970), the most popular MCMC technique, works as follows. Suppose we want to generate samples from a probability distribution $p(\varphi)$ on Φ . Let $q(\xi|\varphi)$ be a distribution for $\xi \in \Phi$, which may or may not depend on $\varphi \in \Phi$. Assume that $q(\xi|\varphi)$ is easy to sample from and it is either computable (up to a multiplicative constant) or symmetric, i.e. $q(\xi|\varphi) = q(\varphi|\xi)$. The sampling distribution $q(\xi|\varphi)$ is called the *proposal distribution*. Starting from essentially any $\varphi_1 \in \Phi$, the Metropolis-Hastings algorithm proceeds by iterating the following two steps.

1. Generate a *candidate* state ξ from the proposal density $q(\xi|\varphi_j)$.
2. Either accept ξ as the next state of the Markov chain, $\varphi_{j+1} = \xi$, with probability

$$\alpha(\xi|\varphi_j) = \min \left\{ 1, \frac{p(\xi)q(\varphi_j|\xi)}{p(\varphi_j)q(\xi|\varphi_j)} \right\}, \quad (3)$$

or reject ξ and set $\varphi_{j+1} = \varphi_j$ with the remaining probability $1 - \alpha(\xi|\varphi_j)$.

It can be shown (see, for example, (Robert & Casella 2004)), that under fairly weak conditions, $p(\varphi)$ is the stationary distribution of the Markov chain $\varphi_1, \varphi_2, \dots$, i.e. the distribution of φ_j converges to $p(\varphi)$ as $j \rightarrow \infty$.

The two main special cases of the Metropolis-Hastings algorithm are Independent Metropolis-Hastings (IMH), where the proposal distribution $q(\xi|\varphi) = q_g(\xi)$ is independent of φ (so q_g is a *global proposal*), and Random Walk Metropolis-Hastings (RWMH), where the proposal distribution is of the form $q(\xi|\varphi) = q_l(\xi - \varphi)$, i.e. a candidate state is proposed as $\xi = \varphi_j + \varepsilon_j$, where $\varepsilon_j \sim q_l$ is a random perturbation (so q_l is a *local proposal*). In both cases, the choice of the proposal distribution strongly affects the efficiency of the algorithms. For IMH to work well, as with importance sampling, the proposal distribution must be a good approximation of the *target distribution* $p(\varphi)$, otherwise a large fraction of the candidate samples will be rejected and the Markov chain will be too slow in covering the important regions of $p(\varphi)$. When, however, it is possible to find a proposal q_g such that $q_g \approx p$, IMH should always be preferred to RWMH because of better efficiency. Unfortunately, such a proposal is usually difficult to construct when the target distribution $p(\varphi)$ is complex. This limits the applicability of IMH.

Since the random walk proposal q_l is local, it is less sensitive to the target distribution. That is why, in practice, RWMH is more robust and used more frequently than IMH. Nonetheless, there are settings where RWMH also does not work well because of the complexity of the target distribution. Although convergence of the Markov chain $\varphi_1, \varphi_2, \dots$ to its stationary distribution $p(\varphi)$ is true in theory, in practice, it is often very difficult to check whether the chain has reached its steady-state or not.

3 ASYMPTOTICALLY INDEPENDENT MARKOV SAMPLING FOR OPTIMIZATION

Recently a new advanced stochastic simulation scheme, called Asymptotically Independent Markov Sampling (AIMS), was developed for computational Bayesian inference (Beck & Zuev 2013). This scheme efficiently combines importance sampling, Markov chain Monte Carlo, and annealing for sampling from any target distribution. In this paper, we extend the applications of AIMS to global optimization problems by introducing AIMS-OPT for solving the optimization problem (2).

The starting point of AIMS-OPT is the concept of *annealing* (or *tempering*), which is based on the following simple but important observation: finding the global minimum of the objective function $\mathcal{H}(\varphi)$ is equivalent to finding the global maximum of $\exp(-\mathcal{H}(\varphi)/T)$ for any given “temperature” $T > 0$ (Kirkpatrick, Gelatt, & Vecchi 1983). Let us define a “tempered” distribution on the bounded admissible parameter space Φ as follows:

$$p_T(\varphi) \propto \exp(-\mathcal{H}(\varphi)/T)I_\Phi(\varphi), \quad (4)$$

where $I_\Phi(\varphi)$ is the indicator function of Φ . Note that the tempered PDF (4) becomes flatter as the temperature T increases, i.e. as $p_T(\varphi)$ gets “hotter”; and it becomes spikier as T decreases toward zero, i.e. as $p_T(\varphi)$ gets “cooler”. More precisely,

$$\lim_{T \rightarrow \infty} p_T(\varphi) = U_\Phi(\varphi) \quad \text{and} \quad \lim_{T \rightarrow 0} p_T(\varphi) = U_{\Phi^*}(\varphi), \quad (5)$$

where $U_\Phi(\varphi)$ and $U_{\Phi^*}(\varphi)$ are the uniform distributions on the parameter space Φ and the optimal solution set Φ^* , respectively. $U_{\Phi^*}(\varphi)$ may be a discrete or continuous distribution. In particular, if the optimization problem (2) has a unique solution, i.e. the optimal solution set Φ^* consists only of a single point, $\Phi^* = \{\varphi^*\}$, then $\lim_{T \rightarrow 0} p_T(\varphi) = \delta_{\varphi^*}(\varphi)$, where $\delta_{\varphi^*}(\varphi)$ is the Dirac mass at φ^* .

The key idea behind annealing is the following: as the temperature T decreases, the tempered distribution $p_T(\varphi)$ puts more and more of its probability mass (converging to one) into the set of optimal solutions Φ^* . Therefore, when T is close to zero, a sample drawn from $p_T(\varphi)$ will be in a neighborhood Φ_T^* of Φ^* with a very high probability. Here, Φ_T^* denotes the so-called “practical support” of $p_T(\varphi)$, i.e. the region that contains almost all probability mass of $p_T(\varphi)$ with the property that $\lim_{T \rightarrow 0} \Phi_T^* = \Phi^*$.

Let $\infty = T_0 > T_1 > \dots > T_k > \dots$ be a sequence of monotonically decreasing temperatures with $\lim_{k \rightarrow \infty} T_k = 0$, and

$$p_0(\varphi) = U_\Phi(\varphi), \quad (6)$$

$$p_k(\varphi) \propto \exp(-\mathcal{H}(\varphi)/T_k)I_\Phi(\varphi), \quad k = 1, 2, \dots$$

be the corresponding sequence of tempered PDFs on Φ . In AIMS-OPT, we sequentially generate samples from the tempered distributions in (6) in the following way. Importance sampling with $p_{k-1}(\varphi)$ as the ISD is used for a construction of an approximation $\hat{p}_{k,n}(\varphi)$ of $p_k(\varphi)$, which is based on samples $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)} \sim p_{k-1}(\varphi)$. This approximation is then employed as the global proposal distribution for sampling from $p_k(\varphi)$ by the IMH algorithm. The tempered distributions in (6) are constructed adaptively, using the essential sample size (ESS) to measure how much $p_{k-1}(\varphi)$ differs from $p_k(\varphi)$ (Beck & Zuev 2013). When the number of samples $n \rightarrow \infty$, the approximation $\hat{p}_{k,n}(\varphi)$ converges to $p_k(\varphi)$, providing the optimal proposal distribution. In other words, when $n \rightarrow \infty$, the corresponding MCMC sampler produces independent samples, which is the reason for “Asymptotically Independent Markov Sampling” in naming the algorithm.

We will refer to k and T_k as the *annealing level* and the *annealing temperature* at level k , respectively. In the next subsection, we assume that T_k is given and therefore the tempered distribution $p_k(\varphi)$ is also known (up to a normalizing constant). In Subsection 3.2, we describe how to choose the annealing temperatures adaptively.

3.1 AIMS-OPT at annealing level k

First, we describe how AIMS-OPT generates samples $\varphi_1^{(k)}, \dots, \varphi_n^{(k)}$ from $p_k(\varphi)$ based on the samples $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)} \sim p_{k-1}(\varphi)$ obtained at the previous annealing level.

Let \mathcal{P}_k be any Markov transition kernel such that $p_k(\varphi)$ is a stationary distribution with respect to \mathcal{P}_k . By definition, this means that

$$p_k(\varphi)d\varphi = \int_{\Phi} \mathcal{P}_k(d\varphi|\xi)p_k(\xi)d\xi \quad (7)$$

Applying importance sampling with the ISD $p_{k-1}(\varphi)$ to integral (7), we obtain:

$$p_k(\varphi)d\varphi = \int_{\Phi} \mathcal{P}_k(d\varphi|\xi) \frac{p_k(\xi)}{p_{k-1}(\xi)} p_{k-1}(\xi) d\xi$$

$$\approx \sum_{j=1}^n \mathcal{P}_k(d\varphi|\varphi_j^{(k-1)}) \bar{\omega}_j^{(k-1)} \stackrel{\text{def}}{=} \hat{p}_{k,n}(d\varphi), \quad (8)$$

where $\hat{p}_{k,n}(d\varphi)$ will be used as the *global* proposal distribution in the IMH algorithm for sampling from $p_k(\varphi)$, and

$$\omega_j^{(k-1)} = \frac{p_k(\varphi_j^{(k-1)})}{p_{k-1}(\varphi_j^{(k-1)})}, \quad \bar{\omega}_j^{(k-1)} = \frac{\omega_j^{(k-1)}}{\sum_{j=1}^n \omega_j^{(k-1)}}, \quad (9)$$

are the importance weights and normalized importance weights, respectively. Note that to compute

$\bar{\omega}_j^{(k-1)}$, we do not need to know the normalizing constants of $p_{k-1}(\varphi)$ and $p_k(\varphi)$. If adjacent tempered distributions $p_{k-1}(\varphi)$ and $p_k(\varphi)$ are sufficiently close (in other words, if the temperature change $\Delta T_k = T_k - T_{k-1}$ is small enough), then the variability of the importance weights (9) will be mild, and, therefore, we can expect that, for reasonably large n , approximation (8) is accurate. A simple illustrative example of approximation (8) is given in (Beck & Zuev 2013).

From now on, we consider a special case where \mathcal{P}_k is the RWMH transition kernel. In this case, it can be written as follows:

$$\mathcal{P}_k(d\varphi|\xi) = q_k(\varphi|\xi) \min \left\{ 1, \frac{p_k(\varphi)}{p_k(\xi)} \right\} d\varphi + (1 - a_k(\xi)) \delta_\xi(d\varphi), \quad (10)$$

where $q_k(\varphi|\xi)$ is a symmetric *local* proposal density, and $a_k(\xi)$ is the probability of having a proper transition from ξ to $\Phi \setminus \{\xi\}$:

$$a_k(\xi) = \int_{\Phi} q_k(\varphi|\xi) \min \left\{ 1, \frac{p_k(\varphi)}{p_k(\xi)} \right\} d\varphi \quad (11)$$

For sampling from $p_k(\varphi)$, we will use the Metropolis-Hastings algorithm with the *global* proposal distribution $\hat{p}_{k,n}(d\varphi)$. To accomplish this, we have to be able to compute the ratio $\hat{p}_{k,n}(\varphi)/\hat{p}_{k,n}(\xi)$ for any $\varphi, \xi \in \Phi$ as a part of the expression for the acceptance probability (3), which, in our case, reduces to

$$\alpha_k(\xi|\varphi) = \min \left\{ 1, \frac{p_k(\xi)\hat{p}_{k,n}(\varphi)}{p_k(\varphi)\hat{p}_{k,n}(\xi)} \right\} \quad (12)$$

However, the distribution $\hat{p}_{k,n}(d\varphi)$ does not have a density since it has both continuous and discrete components and, therefore, the ratio $\hat{p}_{k,n}(\varphi)/\hat{p}_{k,n}(\xi)$ might not make sense. Nevertheless, this technical ‘‘lack-of-continuity problem’’ can be overcome by replacing the sample space Φ with $\Phi \setminus \{\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)}\}$ (see details in (Beck & Zuev 2013)), and this leads to the following algorithm for sampling from the tempered distribution $p_k(\varphi)$.

AIMS-OPT at annealing level k

Input:

▷ $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)} \sim p_{k-1}(\varphi)$, samples generated at annealing level $k-1$;

▷ $\varphi_1^{(k)} \in \Phi \setminus \{\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)}\}$, initial state of a Markov chain;

▷ $q_k(\varphi|\xi)$, symmetric proposal density associated with the RWMH kernel.

Algorithm:

for $i = 1, \dots, n-1$ **do**

1) Generate a candidate state

$$\xi \sim \sum_{j=1}^n \bar{\omega}_j^{(k-1)} q_k(\xi|\varphi_j^{(k-1)}) \quad (13)$$

a. Select j from $\{1, \dots, n\}$ with respective probabilities $\bar{\omega}_1^{(k-1)}, \dots, \bar{\omega}_n^{(k-1)}$ given by (9).

b. Generate $\xi \sim q_k(\xi|\varphi_j^{(k-1)})$.

2) Update $\varphi_i^{(k)} \rightarrow \varphi_{i+1}^{(k)}$ by accepting or rejecting ξ as follows:

Set

$$\varphi_{i+1}^{(k)} = \begin{cases} \xi, & \text{with prob. } \mathcal{A}(\varphi_i^{(k)}, \varphi_j^{(k-1)}, \xi), \\ \varphi_i^{(k)}, & \text{with the remaining prob.,} \end{cases} \quad (14)$$

where

$$\begin{aligned} \mathcal{A}(\varphi_i^{(k)}, \varphi_j^{(k-1)}, \xi) &= \min \left\{ 1, \frac{p_k(\xi)}{p_k(\varphi_j^{(k-1)})} \right\} \\ &\times \min \left\{ 1, \frac{p_k(\xi) \sum_{j=1}^n \bar{\omega}_j^{(k-1)} q_k(\varphi_i^{(k)}|\varphi_j^{(k-1)})}{p_k(\varphi_i^{(k)}) \sum_{j=1}^n \bar{\omega}_j^{(k-1)} q_k(\xi|\varphi_j^{(k-1)})} \right\} \\ &\times \frac{\min \left\{ 1, \frac{p_k(\varphi_i^{(k)})}{p_k(\varphi_j^{(k-1)})} \right\}}{\min \left\{ 1, \frac{p_k(\xi)}{p_k(\varphi_j^{(k-1)})} \right\}} \end{aligned} \quad (15)$$

end for

Output:

► $\varphi_1^{(k)}, \dots, \varphi_n^{(k)}$, n states of a Markov chain with a stationary distribution $p_k(\varphi) \propto \exp(-\mathcal{H}(\varphi)/T_k) I_\Phi(\varphi)$.

The proof that $p_k(\varphi)$ is indeed a stationary distribution for the Markov chain generated by AIMS-OPT will be given in the corresponding journal publication.

The choice of the local proposal density $q_k(\varphi|\xi)$ associated with the RWMH kernel determines how efficiently the Markov chain generated by AIMS-OPT at level k explores local neighborhoods of samples $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)}$ generated at the previous level. This makes the choice of $q_k(\varphi|\xi)$ very important. It has been observed by many researchers that the efficiency of Metropolis-Hastings based MCMC methods is not sensitive to the type of the proposal density; however, it strongly depends on its spread. For this reason, we use a truncated Gaussian density as the local proposal:

$$q_k(\varphi|\xi) \propto \mathcal{N}(\varphi|\xi, c_k \mathbb{I}) I_\Phi(\varphi) I_\Phi(\xi), \quad (16)$$

where ξ and $c_k \mathbb{I}$ are the mean and diagonal covariance matrix, respectively. The scaling parameter c_k determines the spread of the local proposal distribution. The optimal values for c_k are, of course, problem dependent. As a general recommendation, c_k should decay with k , since the tempered distributions $p_k(\varphi)$ become more and more concentrated when k increases.

3.2 The full AIMS-OPT method

At the zeroth annealing level, $k = 0$, we generate samples $\varphi_1^{(0)}, \dots, \varphi_n^{(0)}$ uniformly distributed over the admissible parameter space Φ . Then, using the algorithm described in the previous subsection, we generate samples $\varphi_1^{(1)}, \dots, \varphi_n^{(1)}$, which are approximately distributed according to tempered distribution $p_1(\varphi) \propto \exp(-\mathcal{H}(\varphi)/T_1)$ and, therefore, are better solutions if compared with $\varphi_1^{(0)}, \dots, \varphi_n^{(0)}$. We proceed like this until the annealing temperature T_k is small enough so that the corresponding samples $\varphi_1^{(k)}, \dots, \varphi_n^{(k)} \sim p_k(\varphi) \approx U_{\Phi^*}(\varphi)$ are approximately uniformly distributed over the optimal solution set Φ^* . To make the description of AIMS-OPT complete, we have to explain how to choose the annealing temperatures T_k , for $k = 1, 2, \dots$, and provide the stopping criterion.

3.2.1 Adaptive annealing schedule

It is clear that the choice of the annealing temperatures is very important, since, for instance, it affects the accuracy of the importance sampling approximation (8) and, therefore, the efficiency of the whole AIMS-OPT method. At the same time, it is difficult to make a rational choice of the T_k -values in advance, since this requires some prior knowledge about the optimal solution set Φ^* and the global sensitivity of the objective function $\mathcal{H}(\varphi)$ to the parameters φ , which is usually not available. For this reason, we propose an adaptive way of choosing the annealing schedule.

In importance sampling, a useful measure of degeneracy of the method is the *effective sample size* (ESS) n_{eff} introduced in (Kong, Liu, & Wong 1994) and (Liu 1996). The ESS measures how similar the importance sampling distribution $p_{k-1}(\varphi)$ is to the target distribution $p_k(\varphi)$. Suppose n independent samples $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)}$ are generated from $p_{k-1}(\varphi)$, then the ESS of these samples is defined as

$$n_{\text{eff}} = \frac{n}{1 + \text{var}_{p_{k-1}}[\omega(\varphi)]} = \frac{n}{\mathbb{E}_{p_{k-1}}[\omega(\varphi)^2]}, \quad (17)$$

where $\omega(\varphi) = p_k(\varphi)/p_{k-1}(\varphi)$. The ESS can be interpreted as implying that n weighted samples $(\varphi_1^{(k-1)}, \omega_1^{(k-1)}), \dots, (\varphi_n^{(k-1)}, \omega_n^{(k-1)})$ are worth n_{eff} ($\leq n$) identically and independently distributed (i.i.d.) samples drawn from the target distribution $p_k(\varphi)$. One cannot evaluate the ESS exactly but an estimate \hat{n}_{eff} of n_{eff} is given by

$$\hat{n}_{\text{eff}} = \frac{1}{\sum_{j=1}^n (\bar{\omega}_j^{(k-1)})^2}, \quad (18)$$

where $\bar{\omega}_j^{(k-1)}$ is the normalized importance weight of $\varphi_j^{(k-1)}$.

At annealing level k , when the temperature T_{k-1} is already known, the problem is to define T_k . Let $\gamma = \hat{n}_{\text{eff}}/n \in (0, 1)$ be a prescribed threshold that characterizes the “quality” of the weighted sample (the larger γ is, the “better” the weighted sample is). Then we obtain the following equation:

$$\sum_{j=1}^n (\bar{\omega}_j^{(k-1)})^2 = \frac{1}{\gamma n} \quad (19)$$

It is easy to see that this equation can be expressed as an equation on T_k . Solving this equation for T_k gives us the value of the annealing temperature at level k .

The threshold γ affects the speed of annealing. If γ is very small, i.e. close to zero, then AIMS-OPT will have very few tempered distributions, and this will lead to inaccurate results for a moderate number of samples n . On the other hand, if γ is very large, i.e. close to one, then AIMS-OPT will have too many tempered distributions, which will make the algorithm computationally very expensive. As discussed in (Beck & Zuev 2013), in the context of sampling the posterior distribution, $\gamma = 1/2$ is usually a reasonable choice of the threshold.

3.2.2 Stopping criterion

As $k \rightarrow \infty$, the adaptively chosen annealing temperature T_k decreases toward zero, the tempered distribution $p_k(\varphi)$ converges to the uniform distribution $U_{\Phi^*}(\varphi)$ over the optimal solution set Φ^* , and, therefore, the generated samples $\varphi_1^{(k)}, \dots, \varphi_n^{(k)}$ become more and more uniformly distributed over Φ^* . In practice, however, “absolute zero” $T_k = 0$ cannot be reached, and the algorithm should stop using some stopping rule. The proposed stopping criterion is based on the sample coefficient of variation (COV) of the objective function $\mathcal{H}(\varphi)$. Let δ_k denote the sample COV of $\mathcal{H}(\varphi_1^{(k)}), \dots, \mathcal{H}(\varphi_n^{(k)})$, i.e.

$$\delta_k = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n \left(\mathcal{H}(\varphi_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \mathcal{H}(\varphi_j^{(k)}) \right)^2}}{\frac{1}{n} \sum_{j=1}^n \mathcal{H}(\varphi_j^{(k)})} \quad (20)$$

We use δ_k as a measure of sensitivity of the objective function to the parameters φ in the domain $\Phi_{T_k}^*$, the practical support of $p_k(\varphi)$. If samples $\varphi_1, \dots, \varphi_n \sim U_{\Phi^*}(\varphi)$, then their COV is zero, since $\mathcal{H}(\varphi_j) = \min_{\varphi \in \Phi} \mathcal{H}(\varphi)$ for all $j = 1, \dots, n$. Therefore, δ_k converges to zero, when $k \rightarrow \infty$. This suggests the following stopping criterion: run the algorithm until δ_k becomes less than some fraction α of the initial sample COV δ_0 ; in other words, stop when the following condition is fulfilled:

$$\delta_k < \alpha \delta_0 \stackrel{\text{def}}{=} \delta_{\text{target}}, \quad (21)$$

where $\alpha \in (0, 1)$ and δ_{target} is the target sample COV.

Combining the AIMS-OPT algorithm at a given annealing level with the described adaptive annealing schedule and stopping rule gives rise to the following procedure.

The AIMS-OPT method

Input:

- ▷ $\mathcal{H}(\varphi)$, objective function in (2);
- ▷ γ , threshold for the ESS;
- ▷ α , threshold for the stopping rule;
- ▷ n , the number of Markov chain states to be generated at each annealing level;
- ▷ $q_1(\varphi|\xi), q_2(\cdot|\xi), \dots$, where $q_k(\varphi|\xi)$ is the symmetric proposal density associated with the RWMH kernel at annealing level k .

Algorithm:

- Set $k = 0$, current annealing level.
 - Set $T_0 = \infty$, current annealing temperature.
 - Sample $\varphi_1^{(0)}, \dots, \varphi_n^{(0)} \sim U_\Phi(\varphi)$.
 - Calculate δ_0 using (20).
 - Set $\delta_{\text{target}} = \alpha\delta_0$, the target sample COV.
 - while** $\delta_k > \delta_{\text{target}}$ **do**
 - Find T_{k+1} from equation (19).
 - Calculate normalized importance weights $\bar{\omega}_j^{(k)}$, $j = 1, \dots, n$ using (9).
 - Generate a Markov chain $\varphi_1^{(k+1)}, \dots, \varphi_n^{(k+1)}$ with the stationary distribution $p_{k+1}(\varphi)$ using AIMS-OPT at annealing level $k + 1$.
 - Calculate δ_{k+1} using (20).
 - Increment k to $k + 1$.
 - end while**
 - Set $K = k$, the total number of tempered distributions in the annealing schedule.
 - Set $\tau = T_k$, the smallest annealing temperature.
 - Output:
 - ▶ $\varphi_1^{(K)}, \dots, \varphi_n^{(K)} \sim U_{\Phi^*}(\varphi)$, samples that are approximately uniformly distributed over the optimal solution set Φ^* .
-

4 ILLUSTRATIVE EXAMPLES

To illustrate the effectiveness of AIMS-OPT for solving the optimization problem (2), we consider the following three test functions defining objective functions by (1):

$$h_1(\varphi, \theta) = 1 + \left(\varphi_1 - \frac{a}{2}\right)\theta_1 + \left(\varphi_2 - \frac{a}{2}\right)\theta_2,$$

$$h_2(\varphi, \theta) = 1 + \left(\varphi_1 - \frac{a}{2}\right)\left(\varphi_2 - \frac{a}{2}\right)\theta_1\theta_2,$$

$$h_3(\varphi, \theta) = 4a - \theta_1 \text{sign}\left(\varphi_1 - \frac{a}{2}\right) - \theta_2 \text{sign}\left(\varphi_2 - \frac{a}{2}\right),$$

where the admissible parameter space is the square $(\varphi_1, \varphi_2) \in \Phi = [0, a] \times [0, a]$, $a = 10$, and the model parameters $\theta_1 \sim \mathcal{N}\left(\varphi_1 - \frac{a}{2}, 1\right)$ and $\theta_2 \sim \mathcal{N}\left(\varphi_2 - \frac{a}{2}, 1\right)$.

4.1 Exact optimal solution sets

It is easy to evaluate the corresponding objective functions analytically, $\mathcal{H}_i(\varphi) = \mathbb{E}_\pi[h_i(\varphi, \theta)]$, $i = 1, 2, 3$,

$$\mathcal{H}_1(\varphi) = 1 + \left(\varphi_1 - \frac{a}{2}\right)^2 + \left(\varphi_2 - \frac{a}{2}\right)^2,$$

$$\mathcal{H}_2(\varphi) = 1 + \left(\varphi_1 - \frac{a}{2}\right)^2 \left(\varphi_2 - \frac{a}{2}\right)^2, \quad (22)$$

$$\mathcal{H}_3(\varphi) = 4a - \left|\varphi_1 - \frac{a}{2}\right| - \left|\varphi_2 - \frac{a}{2}\right|.$$

The optimal solution sets for these three case are, therefore,

$$\Phi_1^* = \left(\frac{a}{2}, \frac{a}{2}\right),$$

$$\Phi_2^* = \left\{\varphi \in \Phi \mid \varphi_1 = \frac{a}{2} \text{ or } \varphi_2 = \frac{a}{2}\right\}, \quad (23)$$

$$\Phi_3^* = (0, 0) \cup (0, a) \cup (a, 0) \cup (a, a).$$

We will refer to Φ_1^* , Φ_2^* , and Φ_3^* as “center”, “cross”, and “corners”, respectively. Note that while “center” is a relatively simple case — Φ_1^* consists of a single point located at the center of Φ — “cross” and “corners” are quite challenging cases: Φ_2^* has complicated geometry and Φ_3^* consists of four different points situated far from each other.

4.2 Approximation of optimal solution sets using AIMS-OPT

To estimate the objective functions $\mathcal{H}_i(\varphi)$, for each value of φ , $N = 10^3$ samples of the model parameters θ were used in the following Monte Carlo estimator

$$\hat{\mathcal{H}}_i(\varphi) = \frac{1}{N} \sum_{j=1}^N h_i(\varphi, \theta_j) \approx \mathcal{H}_i(\varphi). \quad (24)$$

The left panels of Figures 1, 2, and 3 display the scatterplots of $n = 10^3$ samples of the design parameters obtained from AIMS-OPT for “center”, “cross”, and “corners”, respectively. In these figures, k denotes the annealing level. The parameters of the algorithm were chosen as follows: the threshold for the ESS $\gamma = 1/2$; the threshold for the stopping rule $\alpha = 0.05$ for “center” and “corners” and $\alpha = 0.01$ for “cross”; the local proposal density $q_k(\varphi|\xi) \propto \mathcal{N}(\varphi|\xi, c_k \mathbb{I}) I_\Phi(\varphi) I_\Phi(\xi)$, where $c_0 = 0.1$ and $c_{k+1} = c_k/4$. This implementation of AIMS-OPT leads to a total number of $K = 6, 4,$ and 5 tempered distributions for “center”, “cross”, and “corners”, respectively. Note that the total number of tempered distributions K can be considered as a measure of the global sensitivity of the objective function to the parameters φ : the larger K , the more sensitive the objective function is.

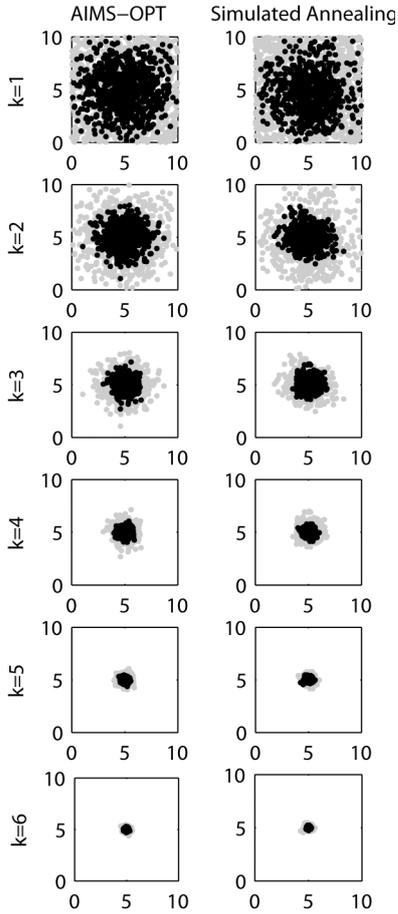


Figure 1: Case “Center”: the left and right panels display the scatterplots of $n = 10^3$ samples obtained from AIMS-OPT and SA, respectively; k denotes the annealing level. Grey and black dots represent $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)}$ and $\varphi_1^{(k)}, \dots, \varphi_n^{(k)}$, respectively.

The theoretical minimum values of the objective functions are $\min_{\varphi \in \Phi} \mathcal{H}_1(\varphi) = 1$, $\min_{\varphi \in \Phi} \mathcal{H}_2(\varphi) = 1$, $\min_{\varphi \in \Phi} \mathcal{H}_3(\varphi) = 3a = 30$. The minimum and the maximum values of the objective functions computed for the samples generated at the last annealing levels ($K = 6, 4$, and 5 for for “center”, “cross”, and “corners”, respectively) are $[\min_{j=1, \dots, n} \mathcal{H}_1(\varphi_j^{(6)}), \max_{j=1, \dots, n} \mathcal{H}_1(\varphi_j^{(6)})] = [1.02, 1.12]$, $[\min_{j=1, \dots, n} \mathcal{H}_2(\varphi_j^{(4)}), \max_{j=1, \dots, n} \mathcal{H}_2(\varphi_j^{(4)})] = [1.01, 1.03]$, $[\min_{j=1, \dots, n} \mathcal{H}_3(\varphi_j^{(5)}), \max_{j=1, \dots, n} \mathcal{H}_3(\varphi_j^{(5)})] = [30.012, 31.10]$.

Let us now compare the performance of AIMS-OPT with Simulated Annealing (SA) (Kirkpatrick, Gelatt, & Vecchi 1983, Robert & Casella 2004). Since we are interested in optimization problems where there may be multiple optimal solutions, we use a population-based SA with adaptive annealing schedule rather than the original algorithm that uses a single thread where the temperature is changed after each Markov chain update by using a pre-set annealing schedule. Thus, the only difference between AIMS-OPT and a population-based SA is in how samples at a given annealing level are generated. In AIMS-OPT, samples $\varphi_1^{(k)}, \dots, \varphi_n^{(k)}$ are obtained using IMH

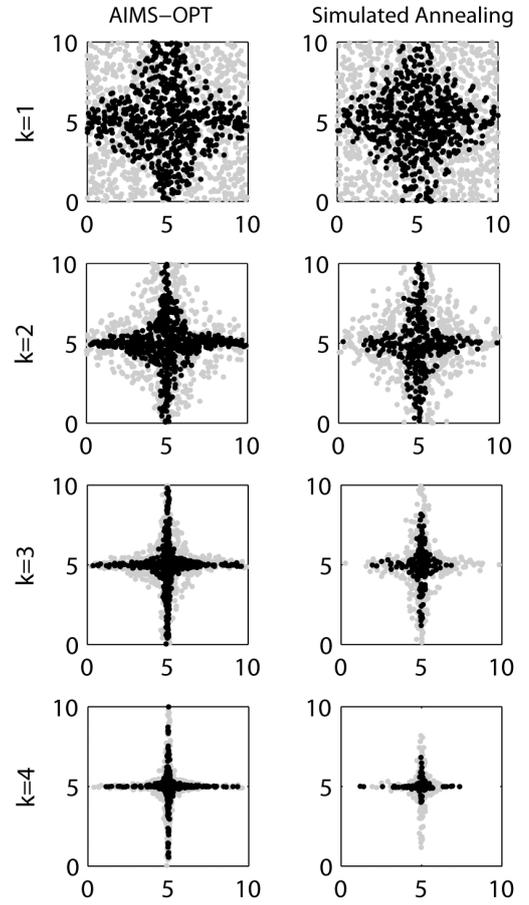


Figure 2: Case “Cross”: the left and right panels display the scatterplots of $n = 10^3$ samples obtained from AIMS-OPT and SA, respectively; k denotes the annealing level. Grey and black dots represent $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)}$ and $\varphi_1^{(k)}, \dots, \varphi_n^{(k)}$, respectively.

with the global proposal distribution $\hat{p}_{k,n}(d\varphi)$ which is constructed based on the samples $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)}$ from the previous annealing level. In SA, RWMH with the local proposal distribution $q_{loc,k}(\varphi|\xi)$ is used instead. To capture the global structure of the target distribution, the local proposal distribution is often chosen to be Gaussian of the following form:

$$q_{loc,k}(\varphi|\xi) \propto \mathcal{N}(\varphi|\xi, c\Sigma_k)I_{\Phi}(\varphi)I_{\Phi}(\xi), \quad (25)$$

where Σ_k is the sample covariance matrix, $\Sigma_k = \sum_{j=1}^n \bar{\omega}_j^{(k-1)} (\varphi_j^{(k-1)} - \bar{\varphi}^{(k)}) (\varphi_j^{(k-1)} - \bar{\varphi}^{(k)})^T$, where $\bar{\varphi}^{(k)} = \sum_{j=1}^n \bar{\omega}_j^{(k-1)} \varphi_j^{(k-1)}$, and $\bar{\omega}_j^{(k-1)}$ are the normalized importance weights from (9), and c is the scaling factor. In all examples, we used an approximately optimal value of the scaling factor, $c = 0.1$.

The right panels of Figures 1, 2, and 3 display the scatterplots of $n = 10^3$ samples obtained from SA for “center”, “cross”, and “corners”, respectively. In the first (simple) case “center”, both AIMS-OPT and SA successfully generate samples in the vicinity of the optimal solution set Φ_1^* . In a more complicated case “cross”, AIMS-OPT approximates Φ_2^* more accurately than SA. Finally, in the most challenging case “corners”, AIMS-OPT clearly outperforms SA: while the AIMS-OPT algorithm successfully finds all four optimal solutions $(0, 0)$, $(0, a)$, $(a, 0)$, and (a, a) ,

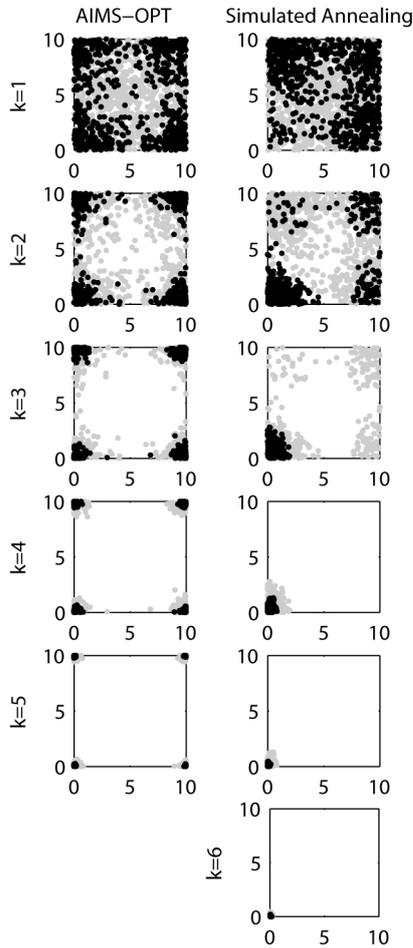


Figure 3: Case “Corners”: the left and right panels display the scatterplots of $n = 10^3$ samples obtained from AIMS-OPT and SA, respectively; k denotes the annealing level. Grey and black dots represent $\varphi_1^{(k-1)}, \dots, \varphi_n^{(k-1)}$ and $\varphi_1^{(k)}, \dots, \varphi_n^{(k)}$, respectively.

SA finds only one, $(0, 0)$. Note also that in this case, the tempered distributions $p_k(\varphi)$ “cool down” slower under SA, and, as a result, SA requires six annealing levels while AIMS-OPT requires only five.

5 CONCLUSIONS

In this paper, a new stochastic simulation method, denoted AIMS-OPT, is introduced for solving global optimization problems such as those that arise in optimal performance-based design. AIMS-OPT is based on Asymptotically Independent Markov Sampling (AIMS), a recently developed advanced simulation scheme originally proposed for Bayesian inference (Beck & Zuev 2013). The main feature of AIMS-OPT is that instead of a single approximation of the optimal solution, the algorithm produces a set of nearly optimal solutions. This can be advantageous in many practical cases, e.g. in multi-objective optimization or when there exists whole set of optimal solutions. Also, AIMS-OPT can be used for exploration of the global sensitivity of the objective function.

The efficiency of AIMS-OPT is demonstrated with several examples, which have different topologies of the optimal solution sets. A comparison is made with

Simulated Annealing (SA). If the optimal solution set has a relatively simple geometry, then the performances of AIMS-OPT and SA are comparable; however, in more complicated cases, AIMS-OPT outperforms SA. In the corresponding journal publication, applications of AIMS-OPT to higher dimensional examples as well as to a reliability-based design optimization problem will be considered.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under award number EAR-0941374 to the California Institute of Technology. This support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect those of the National Science Foundation.

REFERENCES

- Beck, J. (2010). Bayesian system identification based on probability logic. *Struct Control Health Monit* 17, 825–847.
- Beck, J. & K. Zuev (2013). Asymptotically independent markov sampling: a new mcmc scheme for bayesian inference. *accepted by Int J for Uncertainty Quantification*.
- Gasser, M. & G. Schuëller (1997). Reliability-based optimization of structural systems. *Math Meth Oper Res* 46, 287–307.
- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika* 57, 97–109.
- Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, Michigan: University of Michigan Press.
- Jensen, H., M. Valdebenito, & G. Schuëller (2008). An efficient reliability-based optimization scheme for uncertain linear systems subject to general gaussian excitation. *Comp Meth in Applied Mech and Eng* 198, 72–87.
- Kennedy, J. & R. Eberhart (1995). Particle swarm optimization. *Proc of IEEE Int Conf on Neural Networks* 4, 1942–1948.
- Kirkpatrick, S., C. Gelatt, & M. Vecchi (1983). Optimization by simulated annealing. *Science* 220, 671–680.
- Kong, A., J. Liu, & W. Wong (1994). Sequential imputations and bayesian missing data problems. *J American Stat Assoc* 89, 278–288.
- Liu, J. (1996). Metropolized independent sampling with comparison to rejection sampling and importance sampling. *Stat and Comp* 6, 113–119.
- Robert, C. & G. Casella (2004). *Monte Carlo Statistical Methods*. Springer Texts in Statistics.
- Schuëller, G. & H. Jensen (2008). Computational methods in optimization considering uncertainties – an overview. *Comp Meth in Applied Mech and Eng* 198, 2–13.
- Taflanidis, A. & J. Beck (2008). Stochastic subset optimization for optimal reliability problems. *Prob Eng Mech* 23, 324–338.
- Yang, C. & J. Beck (1998). Generalized trajectory methods for finding multiple extrema and roots of functions. *J Optim Theory Appl* 97, 211–227.