2004-07

Implementable Algorithm for Stochastic Optimization Using Sample Average Approximations

Royset, J.O.

Plenum Publishing Corporation

þÿJ.O. Royset and E. Polak, 2004, Implementable Algorithm for Stochastic
þÿSample Average Approximations, Journal of Optimization Theory and Ap

JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 122, No. 1, pp. 157–184, July 2004 (© 2004)

Implementable Algorithm for Stochastic Optimization Using Sample Average Approximations[1]

J. O. Royset[2] and E. Polak[3]

Abstract. We develop an implementable algorithm for stochastic optimization problems involving probability functions. Such problems arise in the design of structural and mechanical systems. The algorithm consists of a nonlinear optimization algorithm applied to sample average approximations and a precision-adjustment rule. The sample average approximations are constructed using Monte Carlo simulations or importance sampling techniques. We prove that the algorithm converges to a solution with probability one and illustrate its use by an example involving a reliability-based optimal design.

Key Words. Stochastic optimization, sample average approximations, Monte Carlo simulations, reliability-based optimal designs.

1. Introduction

A wide range of engineering decisions is subject to uncertainties caused by insufficient information about the system properties, the random occurrence of events such as earthquakes or explosions, the inaccurate representation of the real-world system by mathematical models, etc. In such situations, reliability-based optimization can be used as a decision-support tool.

[1]This work was sponsored by the Research Associateship Program at the National Research Council, the Taisei Chair in Civil Engineering at UC Berkeley, and the National Science Foundation under Grant ECS-9900985. Valuable comments by Professor Kevin Wood, US Naval Postgraduate School, Monterey, California, are acknowledged.

[2]Research Associate, Department of Operations Research, US Naval Postgraduate School, Monterey, California.

[3]Professor, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California.

The reliability-based optimal design problem takes the form

$$(P) \quad \min_{x \in X} \left\{ c_0(x) + \sum_{k \in K} c_k(x) p_k(x) | p_k(x) \leq \hat{p}_k, k \in K \right\}, \tag{1a}$$

where the design variables are denoted by $x \in X \subset \mathbb{R}^n$ and the failure probabilities $p_k(x), k \in K = \{1, 2, \ldots, K_g\}$, are given by the m-dimensional integrals

$$p_k(x) = \int_{\mathcal{F}_k(x)} \varphi_m(u) du, \quad k \in K, \tag{1b}$$

with $\varphi_m : \mathbb{R}^m \to \mathbb{R}$ being the m-dimensional standard normal probability density function and $\mathcal{F}_k(x) \subset \mathbb{R}^m$ being a failure domain. Furthermore, $c_k : \mathbb{R}^n \to \mathbb{R}, k \in \{0, 1, \ldots, K_g\}$, are continuously differentiable functions, with $c_0(\cdot)$ describing the initial cost and $c_k(\cdot), k > 1$, the cost associated with the kth failure mode. The values $\hat{p}_k$ are predefined bounds on the failure probabilities.

Problems of the form P are difficult to solve for at least two reasons. First, rarely $p_k(\cdot)$ can be evaluated exactly; hence, approximations are usually needed. Second, $p_k(\cdot), k \in K$, and their approximations can be nonsmooth.

In the engineering literature, we find several approaches for solving problem P. These approaches include the use of response surface techniques (Ref. 1), surrogate functions (Ref. 2), and first-order approximations to the failure probability (Refs. 3–5). These approaches work satisfactorily under certain conditions, but are not proved to converge to a solution of problem P.

Problems of the form P are special cases of stochastic optimization problems (see e.g. Refs. 6–7). The two main techniques for solving stochastic optimization problems are stochastic quasigradient methods (Refs. 8–12) and sample average approximations (Refs. 13–19).

Stochastic quasigradient methods are not applicable to problems involving failure probability constraints. In principle, such constraints can be removed by including penalty terms or barrier terms in the objective function. However, the details of such an approach do not appear to have been worked out. Additionally, Ref. 19 reports that stochastic quasigradient methods are less robust numerically than sample average approximations due to the difficulty of selecting an efficient stepsize. Thus, stochastic quasigradient methods do not appear to be a good choice for solving reliability-based optimal design problems of the form P.

The results available for sample average approximations include the fact that the minimizers and minimum values of sample average approximations converge to a minimizer and minimum value of the original problem, respectively, as the number of samples goes to infinity. Techniques for checking whether a given design is sufficiently close to stationarity or optimality can be found in Refs. 17 and 19 and, for the case of deterministic constraints, in Ref. 18. These results provide guidance for the selection of one or more approximating problems to be solved using some optimization algorithm.

In this paper, we develop a new implementable algorithm for the solution of problem P based on sample average approximations. Under certain assumptions, we show that both the failure probability and the corresponding sample average approximations are smooth. The sample average approximations give rise to a sequence of smooth approximating problems corresponding to gradually larger sample sizes. Rather than picking a particularly accurate approximation problem to solve, we use a much more efficient diagonalization technique, which consists of starting out with a coarse approximation and then proceeding recursively as follows. One applies a nonlinear programming algorithm to the current approximation until a precision-adjustment test indicates that it is time to move on to a higher-precision approximation whose solution is initiated with the last iterate of the preceding nonlinear programming calculation, i.e., a warm start. Hence, a reasonably good design is obtained before numerically costly approximations are required. We show that the algorithm converges to a local solution with probability 1 and illustrate its behavior with a numerical example.

2. Definition of the Failure Domain

In accordance with common practice in structural and mechanical engineering (see e.g. Ref. 20), we express the uncertainties in engineering design by means of a time-invariant probabilistic model defined in terms of an m-dimensional vector V of random variables. Failures of a structure are defined in terms of limit-state functions $G_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, k \in K$. It is theoretically and computationally convenient to introduce a bijective transformation of realizations v of the random vector V into realizations u of a standard normal random vector U. Such transformations can be defined under weak assumptions when the probability distribution of V is continuous. For a given design vector x, let $T_x : \mathbb{R}^m \to \mathbb{R}^m$ denote this transformation. Replacing v by $T_x^{-1}(u)$, gives the equivalent limit-state functions $g_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, k \in K$, defined by

$$g_k(x,u) = G_k(x, T_x^{-1}(u)).$$

The failure domain $\mathcal{F}_k(x)$ associated with the kth limit-state function is then defined by

$$\mathcal{F}_k(x) = \{u \in \mathbb{R}^m | g_k(x,u) \leq 0\}. \tag{2}$$

Before we proceed, we adopt the following notation. For any set $A \subset \mathbb{R}^m$, we define the indicator function

$$I_A(u) = 1, \quad \text{whenever } u \in A,$$
$$I_A(u) = 0, \quad \text{otherwise.}$$

The components of vectors are given by superscripts; i.e.,

$$a = (a^1, a^2, \ldots, a^q) \in \mathbb{R}^q.$$

Note that, in this notation,

$$p_k(x) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} I_{\mathcal{F}_k(x)}(u) \varphi_m(u) du^1 \ldots du^m, \quad k \in K. \tag{3}$$

3. Properties of the Failure Probability

A sufficient condition for the failure probability to be continuous is found in Ref. 21.

Assumption A1. We assume that, for each $k \in K$,

(i) the limit-state function $g_k(\cdot,\cdot)$ is continuous,
(ii) $M(\{u \in \mathbb{R}^m | g_k(x,u) = 0\}) = 0$, for all $x \in \mathbb{R}^n$, where $M(S) = \int_S \varphi_m(u) du$, for any set $S \subset \mathbb{R}^m$.

Theorem 3.1. See Ref. 21. If Assumption A1 is satisfied, then the failure probabilities $p_k(\cdot), k \in K$, defined in (1b) and (2), are continuous.

A sufficient condition for the failure probability to be continuously differentiable becomes apparent when the equation $g_k(x,u) = 0$ can be solved uniquely in terms of one of the components of u. In Assumption A2 below, this solution is denoted by h_k and is clearly a function of x and the other components of u. Before we proceed with a precise statement, we adopt the following notation. For any $u \in \mathbb{R}^m$, let

$$(u^1, \bar{u}) = u, \quad \text{with } u^1 \in \mathbb{R} \text{ and } \bar{u} = (u^2, u^3, \ldots, u^m) \in \mathbb{R}^{m-1}.$$

Assumption A2. We assume that, for each $k \in K$,

(i) there exists a function $h_k : \mathbb{R}^n \times \mathbb{R}^{m-1} \to \mathbb{R}$ such that, for all $x \in \mathbb{R}^n$ and $\bar{u} \in \mathbb{R}^{m-1}$, $g_k(x, (h_k(x, \bar{u}), \bar{u})) = 0$,
(ii) for all $x \in \mathbb{R}^n$ and $\bar{u} \in \mathbb{R}^{m-1}$, $\partial g_k(x, (h_k(x, \bar{u}), \bar{u}))/\partial u^1 \neq 0$,
(iii) the limit-state function $g_k(\cdot, \cdot)$ is continuously differentiable.

Assumption A2 is satisfied in many practical applications. For example, Assumption A2(i) holds when

$$g_k(x, u) = \tilde{g}_k(x, \bar{u}) + f(x)u^1,$$

with $\tilde{g}_k(\cdot, \cdot)$ some function and $f(x) > 0$ for all $x \in \mathbb{R}^n$.

In Assumptions A1 and A2, the statements are required to hold for all $x \in \mathbb{R}^n$. However, it is sufficient that these statements hold on a sufficiently large subset of the design space containing all relevant designs. Since the characterization of such a subset is application dependent, for generality, we have adopted stronger assumptions than needed typically in practice.

In the following, we denote the standard normal cumulative distribution function by $\Phi(\cdot)$, i.e.,

$$\Phi(\alpha) = \int_{-\infty}^{\alpha} \varphi_1(a)da. \tag{4}$$

Theorem 3.2. Suppose that Assumption A2 holds.

(i) If $\alpha < h_k(x, \bar{u})$ implies $g_k(x, (\alpha, \bar{u})) < 0$, then $p_k(\cdot)$ is continuously differentiable and

$$p_k(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi(h_k(x, \bar{u}))\varphi_{m-1}(\bar{u})du^2 \ldots du^m, \tag{5a}$$

$$\nabla p_k(x) = -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_1(h_k(x, \bar{u})) \frac{\nabla_x g_k(x, (h_k(x, \bar{u}), \bar{u}))}{\partial g_k(x, (h_k(x, \bar{u}), \bar{u}))/\partial u^1} \times \varphi_{m-1}(\bar{u})du^2 \ldots du^m. \tag{5b}$$

(ii) If $\alpha > h_k(x, \bar{u})$ implies $g_k(x, (\alpha, \bar{u})) < 0$, then $p_k(\cdot)$ is continuously differentiable and

$$p_k(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi(-h_k(x, \bar{u}))\varphi_{m-1}(\bar{u})du^2 \ldots du^m, \tag{5c}$$

$$\nabla p_k(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_1(h_k(x, \bar{u})) \frac{\nabla_x g_k(x, (h_k(x, \bar{u}), \bar{u}))}{\partial g_k(x, (h_k(x, \bar{u}), \bar{u}))/\partial u^1} \times \varphi_{m-1}(\bar{u})du^2 \ldots du^m. \tag{5d}$$

Proof. First, consider (i). It follows from (3), Assumption A2(i), and (4) that

$$\begin{aligned} p_k(x) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{\mathcal{F}_k(x)}(u)\varphi_m(u)du^1du^2\ldots du^m \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{h_k(x,\bar{u})} \varphi_m(\bar{u})du^2\ldots du^m \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi(h_k(x,\bar{u}))\varphi_{m-1}(\bar{u})du^2\ldots du^m. \end{aligned} \tag{6a}$$

Since $g(\cdot,\cdot)$ is continuously differentiable and $\partial g_k(x,(h_k(x,\bar{u}),\bar{u}))/\partial u^1 \neq 0$ by Assumption A2, $h_k(\cdot,\cdot)$ is also continuously differentiable with

$$\nabla_x h_k(x,\bar{u}) = -\frac{\nabla_x g_k(x,(h_k(x,\bar{u}),\bar{u}))}{\partial g_k(x,(h_k(x,\bar{u}),\bar{u}))/\partial u^1}. \tag{6b}$$

Hence, from (6a) and (6b), and from the fact that the differentiation and integration operators can be interchanged, we obtain (5b). Item (ii) follows by the same arguments as above. □

In Ref. 12, we find a result similar to Theorem 3.2. It is shown in Ref. 12 that the failure probability is continuously differentiable when the random variables V are bounded. The result in Ref. 12 holds also when $g_k(\cdot,\cdot)$ is a nonsmooth function of the form

$$g_k(x,u) = \max_{j\in J_k} \phi_{k,j}(x,u),$$

with $\phi_{k,j}(\cdot,\cdot)$ being smooth and J_k being a set of finite cardinality

In the following, we assume without loss of generality that (5a) and (5b) hold, but not (5c) and 5(d). If the assumption of item (ii) in Theorem 3.2 holds, we can transform always the problem into a case satisfying the assumption of item (i) in Theorem 3.2 by replacing u^1 by $-u^1$ in the definition of $g_k(\cdot,\cdot)$.

4. Algorithm

We derive an algorithm for solving problem P in two steps. In the first step, we use sampling techniques to estimate the failure probability. This gives rise to a family of approximating problems, which is increasingly accurate as the number of sample points increases. In the second step, we develop an adaptive precision adjustment rule and incorporate it in a diagonalization type algorithm, which we show to converge to local solutions of problem P in (1a).

4.1. Approximation Result. The Monte Carlo simulation estimates of (5a) and (5b) are

$$p_{k,N}(x) = (1/N)\sum_{j=1}^{N} \Phi(h_k(x, \bar{u}_j)), \tag{7a}$$

$$\nabla p_{k,N}(x) = -(1/N)\sum_{j=1}^{N} \varphi_1(h_k(x, \bar{u}_j)) \frac{\nabla_x g_k(x, (h_k(x, \bar{u}_j), \bar{u}_j))}{\partial g_k(x, (h_k(x, \bar{u}_j), \bar{u}_j))/\partial u^1}, \tag{7b}$$

where $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_N$ are realizations of a collection of independent standard normal *(m-1)*-dimensional random vectors $\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_N$.

Instead of generating sample points according to a standard normal distribution, as in (7a) and (7b), we can sample according to other probability distributions. This approach leads to importance sampling, which tends to improve the failure probability estimates by concentrating the samples in the most relevant region. A typical selection of a sampling distribution is a (non-standard) normal distribution. In this paper, we use a normal random vector with mean vector $\mu \in \mathbb{R}^{m-1}$ and variance-covariance matrix $\sigma \mathrm{I}, \sigma > 0$, where I is the $(m-1) \times (m-1)$ identity matrix. The importance sampling estimates of (5a) and (5b) are

$$p_{k,\mathrm{N}}(x; \mu, \sigma) = (1/\mathrm{N})\sum_{j=1}^{\mathrm{N}} \Phi(h_k(x, \sigma \bar{u}_j + \mu)) \frac{\varphi_{m-1}(\sigma \bar{u}_j + \mu)}{\varphi_{m-1}(\bar{u}_j)/\sigma^{m-1}}, \tag{8a}$$

$$\begin{aligned} \nabla p_{k,N}(x; \mu, \sigma) = &-(1/N)\sum_{j=1}^{N} \varphi_1(h_k(x, \sigma \bar{u}_j + \mu)) \\ &\times \frac{\nabla_x g_k(x, (h_k(x, \sigma \bar{u}_j + \mu), \sigma \bar{u}_j + \mu))}{\partial g_k(x, (h_k(x, \sigma \bar{u}_j + \mu), \sigma \bar{u}_j + \mu))/\partial u^1} \\ &\times \frac{\varphi_{m-1}(\sigma \bar{u}_j + \mu)}{\varphi_{m-1}(\bar{u}_j)/\sigma^{m-1}}, \end{aligned} \tag{8b}$$

where we have used the fact that $\sigma \bar{u} + \mu$ is a normal random vector with mean vector μ and variance-covariance matrix $\sigma \mathrm{I}$. Note that, for $\sigma = 1$ and $\mu = 0$, (8a) and (8b) simplify to (7a) and (7b). For $\mu \neq 0$, the samples are not centered at the origin, but hopefully in a more relevant region.

The statistical estimators corresponding to the estimates in (8a) and (8b) are given by

$$p^*_{k,\mathrm{N}}(x; \mu, \sigma) = (1/N)\sum_{j=1}^{\mathrm{N}} \varphi_1(h_k(x, \sigma \bar{\mathrm{U}}_j + \mu)) \frac{\varphi_{m-1}(\sigma \bar{\mathrm{U}}_j + \mu)}{\varphi_{m-1}(\bar{U}_j)/\sigma^{m-1}} \tag{9a}$$

$$\nabla p^*_{k,N}(x;\mu,\sigma) = -(1/N)\sum_{j=1}^{N}\varphi_1(h_k(x,\sigma\bar{U}_j+\mu)) \times \frac{\nabla_x g_k(x,(h_k(x,\sigma\bar{U}_j+\mu),\sigma\bar{U}_j+\mu))}{\partial g_k(x,(h_k(x,\sigma\bar{U}_j+\mu),\sigma\bar{U}_j+\mu))/\partial u^1} \times \frac{\varphi_{m-1}(\sigma\bar{U}_j+\mu)}{\varphi_{m-1}(\bar{U}_j)/\sigma^{m-1}}. \tag{9b}$$

In the following, we use the abbreviation w.p.l for the statement "with probability one".

Theorem 4.1. Let Assumptions A1, A2(i) hold and let the sampling parameters in (9a) be either $\mu=0$ and $\sigma=1$ (Monte Carlo) or $\mu\in\mathbb{R}^{m-1}$ and $\sigma>1$ (importance sampling). For every $\kappa>0$ and bounded set $S\subset\mathbb{R}^n$, there exists a constant $C_1\in(0,\infty)$ such that, for all $x\in S$, $N\in\mathbb{N}=\{1,2,3,\ldots\}$, and $k\in K$,

$$|\,p^*_{k,N}(x;\mu,\sigma)-p_k(x)\,|\le C_1(\log N)^{\kappa+1/2}/\sqrt{N}, \qquad \text{w.p.1.} \tag{10}$$

Proof. According to Theorem 37 in Ref. 22, we need to show only that

$$\Phi(h_k(x,\sigma\bar{U}+\mu))\varphi_{m-1}(\sigma\bar{U}+\mu)/(\varphi_{m-1}(\bar{U})/\sigma^{m-1})-p_k(x)$$

has zero mean and bounded variance. The expectation of $p^*_{k,N}(x;\mu,\sigma)$ is

$$E[p^*_{k,N}(x;\mu,\sigma)]=E\left[\Phi(h_k(x,\sigma\bar{U}+\mu))\frac{\varphi_{m-1}(\sigma\bar{U}+\mu)}{\varphi_{m-1}(\bar{U})/\sigma^{m-1}}\right], \tag{11a}$$

where $\bar{U}$ is a standard normal $(m-1)$- dimensional random vector. By the use of conditioning and a change of variables, we obtain from (11a) that

$$\begin{aligned}
&E[p^*_{k,\mathrm{N}}(x;\mu,\sigma)]\\
&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}E\left[\Phi(h_k(x,\sigma\bar{U}+\mu))\frac{\varphi_{m-1}(\sigma\bar{U}+\mu)}{\varphi_{m-1}(\bar{U})/\sigma^{m-1}}\,|\,\bar{U}=\bar{u}\right]\\
&\quad\times\varphi_{m-1}(\bar{u})du^2\ldots du^m\\
&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Phi(h_k(x,\sigma\bar{u}+\mu))\frac{\varphi_{m-1}(\sigma\bar{u}+\mu)}{\varphi_{m-1}(\bar{u})/\sigma^{m-1}}\\
&\quad\times\varphi_{m-1}(\bar{u})du^2\ldots du^m\\
&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Phi(h_k(x,(\bar{u}))\varphi_{m-1}(\bar{u})du^2\ldots du^m\\
&=p_k(x).
\end{aligned} \tag{11b}$$

Clearly,

$$E[\Phi(h_k(x,\sigma\bar{U}+\mu))\varphi_{m-1}(\sigma\bar{U}+\mu)/(\varphi_{m-1}(\bar{U})/\sigma^{m-1})-p_k(x)]=0. \tag{11c}$$

Next, we consider the variance. If $\mu=0$ and $\sigma=1$, then

$$\varphi_{m-1}(\sigma\bar{u}+\mu)/(\varphi_{m-1}(\bar{u})/\sigma^{m-1})=1$$

If $\mu\in\mathbb{R}^{m-1}$ and $\sigma>1$, then there exists a constant $C\in[1,\infty)$ such that

$$\varphi_{m-1}(\sigma\bar{u}+\mu)/(\varphi_{m-1}(\bar{u})/\sigma^{m-1})\le C, \quad \text{for all } \bar{u}\in\mathbb{R}^{m-1}.$$

Hence,

$$\begin{aligned}&\mathrm{Var}\left[\Phi(h_k(x,\sigma\bar{U}+\mu))\frac{\varphi_{m-1}(\sigma\bar{U}+\mu)}{\varphi_{m-1}(\bar{U})/\sigma^{m-1}}-p_k(x)\right]\\&=E\left[\left(\Phi(h_k(x,\sigma\bar{U}+\mu))\frac{\varphi_{m-1}(\sigma\bar{U}+\mu)}{\varphi_{m-1}(\bar{U})/\sigma^{m-1}}-p_k(x)\right)\right]\\&\le C^2+1,\end{aligned} \tag{11d}$$

for all $x\in S$ and $k\in K$. Now, the result follows by Theorem 37 in Ref. 22. □

We can define a class of approximations to problem P in terms of the sampling estimator in (9a). For any $N\in\mathbb{N}$, $\mu\in\mathbb{R}^{m-1}$, and $\sigma>0$, let the approximating problems P_N^* be

$$\begin{aligned}(\mathrm{P}_N^*)\ \min_{x\in\mathbb{R}^n}\Big\{&c_0(x)+\sum_{k=1}^{k_g}c_k(x)\mathrm{p}_N^*(x;\mu,\sigma)\mid\\&\mathrm{P}_N^*(x;\mu,\sigma)\le\hat{p}_k,\,k\in K,\,x\in X\Big\}.\end{aligned} \tag{12}$$

The deterministic problem corresponding to P_N^* is denoted P_N and is defined by (12) with $\mathrm{P}_{k,N}^*(x;\mu,\sigma)$ replaced by $\mathrm{p}_{k,N}(x;\mu,\sigma)$. For the remainder of the paper, we assume that the subset $X\subset\mathbb{R}^n$ [see problem P in (1a)] are given by

$$X=\{x\in\mathbb{R}^n\mid f_j(x)\le 0,\,j\in J\}, \tag{13}$$

where

$$f_j:\mathbb{R}^n\to\mathbb{R},\qquad j\in J=\{1,2,\ldots,J_f\}, \tag{14}$$

are deterministic, continuously differentiable, constraint functions. Also, let $\mathcal{N}\subset\mathbb{N}$ be an infinite sequence of strictly increasing integers. Under a constraint qualification, we state our asymptotic approximation result.

Assumption A3. For each $x \in \Gamma$, there exist a direction $d \in \mathbb{R}^n$ and an $\epsilon_0 > 0$ such that, for all $\epsilon \in (0, \epsilon_0]$, $x_\epsilon = x + \epsilon d$ satisfies $f_j(x_\epsilon) < 0$ for all $j \in J$ and $p_k(x_\epsilon) < \hat{p}_k$ for all $k \in K$.

Theorem 4.2. Let Assumptions A1, A2(i), and A3 hold. If $\{\hat{x}_N^*\}_{N \epsilon \mathcal{N}}$ is a sequence of global solutions of $\{P_N^*\}_{N \epsilon \mathcal{N}}$, then every accumulation point of $\{\hat{x}_N^*\}_{N \epsilon \mathcal{N}}$ is a global solution of problem P with probability 1.

Proof. See the Appendix. □

Note that results similar to the one in Theorem 4.2 can be found in Refs. 13, 15. 19.

4.2. Implementable Algorithm Under Assumptions A1 and A2, problem P is a smooth nonlinear program with stationary points defined by the Fritz John conditions (see e.g. Ref.23). We find it convenient to express those conditions by means of a nonpositive, continuous, optimality function $\theta : \mathbb{R}^n \to \mathbb{R}$, defined by

$$\begin{aligned}\theta(x) = -\min_{\nu \in \Sigma} \Bigg\{ & \sum_{k \in K} \nu^k [\psi(x)_+ - p_k(x) + \hat{p}_k] \\ & + \sum_{j \in J} \nu^{j+K_g} \left[\psi(x)_+ - f_j(x)\right] \\ & + (1/2\delta) \Bigg\| \sum_{k \in K} \nu^k \nabla p_k(x) + \sum_{j \in J} \nu^{j+K_g} \nabla f_j(x) \\ & + \nu^{K_g + J_f + 1} \nabla f_0(x) \Bigg\|^2 + \nu^{K_g + J_f + 1} \gamma \psi(x)_+ \Bigg\} \end{aligned} \tag{15}$$

with the parameters $\gamma, \delta > 0$ and

$$f_0(x) = c_0(x) + \sum_{k \in K} c_k(x) p_k(x), \tag{16}$$

$$\psi(x) = \max \left\{ \max_{k \in K} \left\{ \hat{p}_k(x) - \hat{p}_k \right\}, \max_{j \in J} f_j(x) \right\}, \tag{17}$$

$$\begin{aligned}\Sigma = \Bigg\{ \nu \in \mathbb{R}^{K_g + J_f + 1} \Bigg| & \sum_{i=1}^{K_g + J_f + 1} \nu^i = 1, \\ & \nu^i \geq 0, \forall i \in \{1, 2, \ldots, K_g + J_f + 1\} \Bigg\}, \end{aligned} \tag{18}$$

$$\psi(x)_+ = \max\{0, \psi(x)\}. \tag{19}$$

Theorem 4.3. (See Ref. 23). Let Assumption A2 hold. If $\hat{x} \in \mathbb{R}^n$ is a local minimizer for problem P, then $\hat{x}$ satisfies the Fritz. John conditions and $\theta(\hat{x}) = 0$. □

A sufficient condition for uniform convergence of (8b) to (5b) follows.

Assumption A4. For every bounded set $S \subset \mathbb{R}^n$, there exists constants $C_2, C_3 \in (0, \infty)$ such that

$$|\partial g_k(x, (h_k(x, \bar{u}), \bar{u}))/\partial x^i| \leq C_2 \text{ and } |\partial g_k(x, (h_k(x, \bar{u}), \bar{u}))/\partial u^1| \geq C_3,$$

for all $x \in S, \bar{u} \in \mathbb{R}^{m-1}, k \in K,$ and $i \in \{1, 2, \ldots, n\}$.

Assumption A4 can be difficult to check in practice. However, cases such as

$$g(x, u) = \tilde{g}(x, \bar{u}) + f(x)u^1,$$

with $f(x) > 0$ for $x \in \mathbb{R}^n$ and bounded random variable V, satisfy Assumption A4.

Theorem 4.4. Let Assumptions A2, A4 hold and let the sampling parameters in (9b) be either $\mu = 0$ and $\sigma = 1$ (Monte Carlo) or $\mu \in \mathbb{R}^{m-1}$ and $\sigma > 1$ (importance sampling). For every $\kappa > 0$ and bounded set $S \subset \mathbb{R}^n$, there exists a constant $C_4 \in (0, \infty)$ such that, for all $x \in S$, $N \in \mathbb{N}$, and $k \in K$,

$$\|\nabla p^*_{k,N}(x; \mu, \sigma) - \nabla p_k(x)\| \leq C_4 (\log N)^{\kappa+1/2}/\sqrt{N}, \quad \text{w.p.1.} \tag{20}$$

Proof. The result follows by the same arguments as in the proof of Theorem 4.1. □

We formulate an algorithm model for solving problem P, which makes use of an algorithm map $A_N : \mathbb{R}^n \to \mathbb{R}^n$. For a given number of sample points N, recursive use of the algorithm map A_N will compute a stationary point of P_N. Any algorithm map satisfying an assumption stated below can be combined with the algorithm model to generate an implementable algorithm.

The algorithm model considers the sequence of approximating problems $\{P_N\}_{N \in \mathcal{N}}$. At a given number of sample points N', the algorithm model computes iterates that approach a stationary point of the

approximating problem $P_{N'}$. When the current iterate is sufficiently close to a stationary point for $P_{N'}$, as determined by a test described below, the number of sample points is increased to $N'' > N'$. The algorithm model then continues by computing iterates that approach a stationary point of $P_{N''}$ until the test again determines that the number of sample points must be increased. The last iterate of the previous approximation level is used as a warm start for the next approximation level. The iterates generated by the algorithm model gradually get closer and closer to a stationary point of the current approximating problem before the number of sample points is increased. In essence, the algorithm model computes approximating solutions to a sequence of approximating problems $\{P_N\}_{N\in\mathcal{N}}$ with higher and higher precision as the number of iterations increases. Such a diagonalization scheme is substantially more efficient than choosing a single high-precision approximation and solving that directly.

Before we describe the algorithm model, we must establish additional notation and state the assumption about the algorithm map. Let

$$F_N(x', x''; \mu, \sigma) = \max\{f_{0,N}(x''; \mu, \sigma) - f_{0,N}(x'; \mu, \sigma) - \gamma\psi_N(x'; \mu, \sigma)_+, \psi_N(x''; \mu, \sigma) - \psi_N(x'; \mu, \sigma)_+\}, \tag{21a}$$

$$f_{0,N}(x'; \mu, \sigma) = c_0(x) + \sum_{k\in K} c_k(x) p_{k,N}(x; \mu, \sigma), \tag{21b}$$

$$\psi_N(x; \mu, \sigma) = \max\left\{\max_{k\in K}\{p_{k,N}(x; \mu, \sigma) - \hat{p}_k\}, \max_{j\in J} f_j(x)\right\}, \tag{21c}$$

$$\psi_N(x; \mu, \sigma)_+ = \max\{0, \psi_N(x; \mu, \sigma)\}, \tag{21d}$$

with $\gamma > 0$ being a constant.

Assumption A5. Let $\mu \in \mathbb{R}^{m-1}$ and $\sigma > 0$ be the sampling parameters in (8a) and (8b). For any $N \geq N_0$, N_0 some integer, let $A_N : \mathbb{R}^n \to \mathbb{R}^n$ be an algorithm map for P_N with the property that, for every $x \in \mathbb{R}^n$ such that $\theta(x) < 0$, there exist $\rho_x > 0$, $N_x \in \mathbb{N}$, and $\delta_x > 0$ such that

$$F_N(x', x''; \mu, \sigma) \leq -\delta_x, \tag{22}$$

for all $N \geq N_x$, $x'' \in A_N(x')$, and x' satisfying $\|x - x'\| \leq \rho_x$.

Algorithm 4.1. Algorithm Model for Solving Problem P.

Parameters. $k > 0, \eta > 0, \gamma > 0$, and either $(\mu = 0, \sigma = 1)$ or $(\mu \in \mathbb{R}^{m-1}, \sigma > 1)$.

Data. $N_0 \in \mathbb{N}$, $x_1 \in \mathbb{R}^n$, an infinite sequence $\bar{u}_1, \bar{u}_2, \bar{u}_3, \ldots$ of generated realizations of independent standard normal $(m-1)$-dimensional random vectors $\bar{U}_1, \bar{U}_2, \bar{U}_3, \ldots$, and an infinite set $\mathcal{N}$ of strictly increasing positive integers.

Step 0. Set $i = 1$.

Step 1. Compute the smallest $N_i \in \mathcal{N}$ and the corresponding x_{i+1} such that $N_i \geq N_{i-1}, x_{i+1} = A_{N_i}(x_i)$, and

$$F_{N_i}(x_i, x_{i+1}; \mu\sigma) \leq -\eta(\log N_i)^{\kappa+1/2}/\sqrt{N_i}. \tag{23}$$

Step 2. Replace i by $i+1$, and go to Step 1.

Proposition 4.1. Let Assumptions A2 and A5 hold. Consider Algorithm 4.1.

(i) If $x_i \in \mathbb{R}^n$ is not stationary , i.e., $\theta(x_i) < 0$, then there exists a finite $N_i \in \mathcal{N}$ such that (23) is satisfied.

(ii) Whenever Algorithm 4.1 constructs an infinite sequence $\{x_i\}_{i=1}^{\infty}$ that has an accumulation point $\hat{x}$, the accompanying sequence $\{N_i\}_{i=1}^{\infty}$ diverges to infinity, i.e., $N_i \to \infty$, as $i \to \infty$.

Proof. See the Appendix. □

Theorem 4.5. Let Assumptions A2, A5 hold and assume that Algorithm 4.1 has constructed a bounded sequence $\{x_i\}_{i=1}^{\infty}$ with accumulation point $\hat{x}$. Then, $\hat{x}$ is a stationary point for problem P, i.e., $\theta(\hat{x}) = 0$.

Proof. See the Appendix. □

In particular, the Polak-He algorithm map (see Section 2.6 of Ref. 23), given by

$$A_N(x) = x + \lambda_N(x; \mu, \sigma) h_N(x; \mu, \sigma), \tag{24a}$$

can be used in Algorithm 4.1, where $\lambda_N(x; \mu, \sigma)$ is the Armijo stepsize,

$$\begin{aligned} \lambda_N(x; \mu, \sigma) = \max_{k \in \mathbb{N}} \Big\{ & \beta^k \mid F_N(x, x + \beta^k h_N(x; \mu, \sigma); \mu, \sigma) \\ & \leq \beta^k \alpha \theta_N(x; \mu, \sigma) \Big\}, \end{aligned} \tag{24b}$$

with parameters $\alpha \in (0, 1]$, $\beta \in (0, 1)$, and for some parameters $\gamma, \delta > 0$,

$$\theta_N(x; \mu, \sigma) = -\min_{v \in \Sigma} \left\{ \sum_{k \in K} v^k \left[\psi_N(x; \mu, \sigma)_+ - p_{k,N}(x; \mu, \sigma) + \hat{p}_k \right] \right.$$

$$+\sum_{j\in J}\nu^{j+K_g}\left[\psi_N(x;\mu,\sigma)_+ - f_j(x)\right]$$

$$+(1/2\delta)\left\|\sum_{k\in K}\nu^k\nabla p_{k,N}(x;\mu,\sigma)\right.$$

$$\left.+\sum_{j\in J}\nu^{j+K_g}\nabla f_j(x)+\nu^{K_g+J_f+1}\nabla f_{0,N}(x;\mu,\sigma)\right\|^2$$

$$+\nu^{K_g+J_f+1}\gamma\psi N(x;\mu,\sigma)_+\Big\}, \tag{24c}$$

and where the search direction is

$$h_N(x;\mu,\sigma)=-(1/\delta)$$

$$\times\left(\sum_{k\in K}\hat{\nu}^k\nabla p_{k,N}(x;\mu,\sigma)\right.$$

$$\left.+\sum_{j\in J}\hat{\nu}^{j+k_g}\nabla f_j(x)+\hat{\nu}^{k_g+J_f+1}\nabla f_{0,N}(x;\mu,\sigma)\right), \tag{24d}$$

with $\hat{\nu}=(\hat{\nu}^1,\ldots,\hat{\nu}^{K_g+J_f+1})$ being any solution of (24c). The parameter γ in (24c) must be set equal to the value of the algorithm parameter γ in Algorithm 4.1. The Polak-He algorithm map satisfies Assumption A5. □

Proposition 4.2. Let Assumptions A2 and A4 hold. For any $N\in\mathbb{N}$, let the algorithm map $A_N(\cdot)$ be defined by one iteration of the Polak-He algorithm, i.e., (24a), with (24b)-(24d), applied to P_N, with the same values of the parameters $\alpha,\beta,\sigma,\gamma$ for all $N\in\mathbb{N}$. Then, $A_N(\cdot)$ satisfies Assumption A5; i.e., for every $x\in\mathbb{R}^n$ such that $\theta(x)<0$, there exist $\rho_x>0$, $N_x\in\mathbb{N}$, and $\delta_x>0$ such that

$$F_N(x',x'';\mu,\sigma)\le-\delta_x, \tag{25}$$

for all $x''\in A_N(x')$, $N\ge N_x$, and x' satisfying $\|x-x'\|\le\rho_x$.

Proof. See the Appendix. □

Note that the one-dimensional root-finding problem in the evaluation of $p_{k,N}(\cdot)$ and $\nabla p_{k,N}(\cdot)$ usually cannot be solved exactly in finite computing time. Hence, Step 1 of Algorithm 4.1 may be considered as a conceptual step. We could have introduced an additional precision parameter that ensures gradually better accuracy in the root finding as Algorithm 4.1 progresses, or we would have prescribed a rule saying that the root

finding algorithm (e.g., the secant method) should terminate after kN_i iterations for some $k > 0$. This would have lead to an implementable algorithm with the same behavior as Algorithm 4.1. For simplicity of the presentation above, we have not included this issue in the discussion. In fact, the issue of root finding is not problematic in practice. One-dimensional root-finding problems can be solved in a few iterations with high accuracy using standard algorithms. Hence, the root-finding problem can be solved with a fixed precision for all iterations in Algorithm 4.1 giving a negligible error.

5. Numerical Example

Consider a short structural column with a rectangular cross section of dimensions b and h and material yield strength Y, which is subjected to biaxial bending moments M_1, M_2 and axial force P_a. Assuming an elastic perfectly plastic material, the reliability of the column is defined by the limit-state function

$$G(x, v) = 1 - \frac{4m_1}{bh^2 y} - \frac{4m_2}{b^2 hy} - \left(\frac{pa}{bhy}\right)^2, \tag{26}$$

where $v = (m_1, m_2, p_a, y) \in \mathbb{R}^4$ denotes a realization of the random vector $V = (M_1, M_2, P_a, Y)$ and $x = (b, h) \in \mathbb{R}^2$ denotes the vector of design variables. Since $K_g = 1$, we have simplified the notation by dropping the subscript k. We assume that M_1, M_2, P_a, Y are statistically independent lognormal random variables with means 250 kNm,125 kNm, 2500 kN, 40 MPa respectively and coefficients of variation 0.3, 0.3, 0.2, 0.1 respectively.

Suppose that the column is to be designed for minimum cross-sectional area $A = bh$, subject to the failure probability constraint

$$p(x) \leq 0.00134990$$

and deterministic constraints

$$b, h \geq 0 \quad \text{and} \quad 0.5 \leq b/h \leq 2.$$

Hence,

$$c_0(x) = bh, \;\; c_k(x) = 0, \;\; k \in K,$$
$$f_1(x) = -b, \;\; f_2(x) = -h, \;\; f_3(x) = b/h - 2, \;\; f_4(x) = 1/2 - b/h$$

in (1a).

As required by Assumption A2(i), we select to solve the equations $g(x,u)=0$ for the standardized random variable corresponding to m_1. In Algorithm 4.1, we used the algorithm parameters

$$\kappa=0.0001,\ \eta=0.0001,\ \gamma=2.$$

Note that smaller η implies that the test in (23) becomes easier to pass. Hence, smaller η results in an initially slower increase in the number of sample points. Additionally, we select the sampling parameters to be

$$\mu=(2,2,-1),\quad \sigma=1.01,$$

where the sampling density has been shifted into the negative range for the resistance variable and the positive range for the load variables. The parameters in the Polak-He algorithm map are selected to be

$$\alpha=0.5,\quad \beta=0.8,\quad \delta=1.$$

The initial number of sample points is $N_0=1000$ and the consecutive number of sampling points are given by the set

$$\mathcal{N}=\left\{10^3, 5\cdot 10^3, 2.5\cdot 10^4, 1.25\cdot 10^5, 6.25\cdot 10^5, 3.125\cdot 10^6, \dots\right\}.$$

We do not specify $\mathcal{N}$ beyond $3.125\cdot 10^6$, because we plan to terminate the calculations when the number of sample points is about to be increased beyond $3.125\cdot 10^6$.

Algorithm 4.1 is implemented using Matlab (Ref. 24) with the QP-solver quadprog and the example is run on a 1.7 GHz laptop computer. The calculations are terminated after 108 iterations of Algorithm 4.1. Results after various number of iterations are summarized in Table 1, where estimates of the failure probabilities using both N_i and $N^*=3.125\cdot 10^6$ number of samples are given. The estimates after $3.125\cdot 10^6$ samples have a coefficient of variations of less than 0.5%. The convergence of the iterates x_i and of the objective function $c_0(x_i)$ and the increase of the number of sample points N_i are shown in Figures 1–3 respectively.

We see from Table 1 and Figures 1–3 that the number of sample points N_i is initially small but increases as less progress is made towards the solution of the approximating problems. In fact, most iterations are performed on problems involving 5000 or fewer samples ($i=68$ and $i=95$ correspond to the last iterates with $N_i=1000$ and $N_i=5000$ respectively). This illustrates a significant advantage of Algorithm 4.1: Coarse estimates

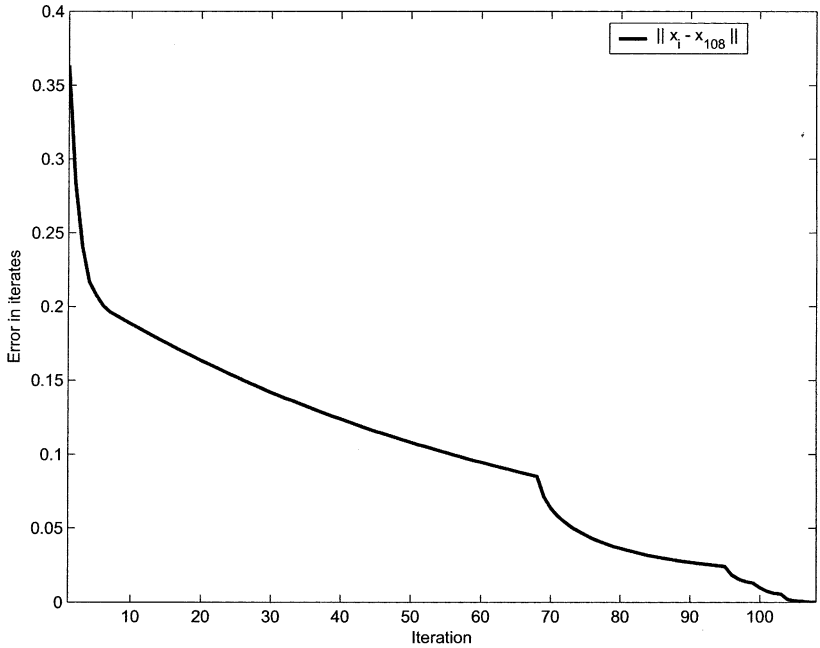

Fig. 1. Convergence in the iterates x_i.

of the failure probabilities are used until a reasonably good design is obtained. Using this design as a warm start, it is necessary only to perform a few iterations with high-precision, computationally expensive, estimates of the failure probabilities to obtain an optimal design. Because of the long computing time needed for iterations with $N_i = 3.125 \cdot 10^6$, it is practically impossible to run this example with a constant sample size of $N_i = 3.125 \cdot 10^6$ for all $i \in \mathbb{N}$.

It is also seen that the value of the objective function may increase slightly when the number of sample points is increased. This occurs because the failure probability is reestimated using a large number of sample points.

6. Conclusions

We have developed an implementable algorithm for the solution of stochastic optimization problems based on sample average approximations. The algorithm is illustrated by an example from the area of structural design. Our algorithm is shown to converge to a solution of the problem

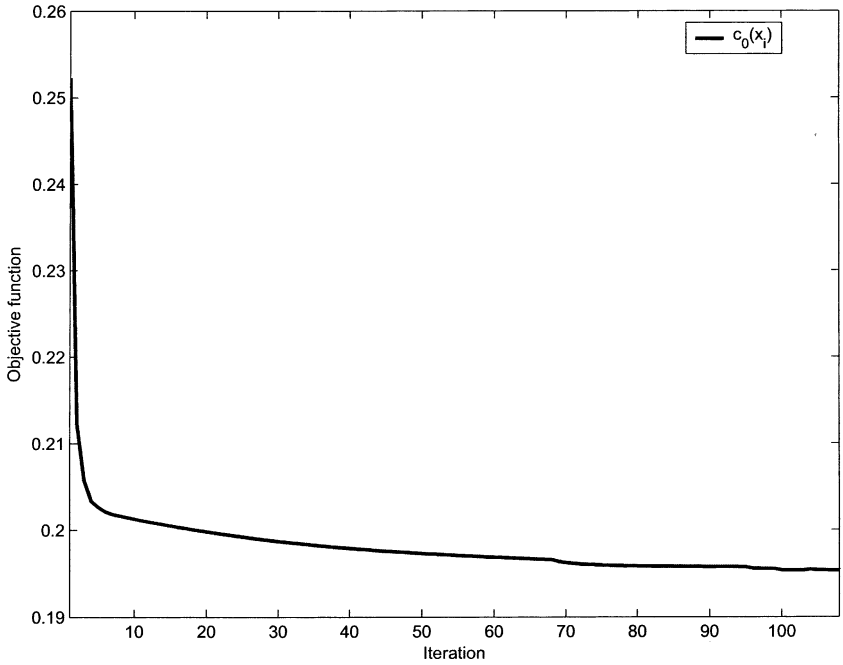

Fig. 2. Convergence in the objective $C_0(x_i)$.

with probability one under fairly general conditions. The algorithm uses an adaptive scheme to control the precision of the failure probability estimates, which reduces the computing time significantly. Initially, only a small number of sample points is used to estimate the failure probability by means of simulation techniques. As the algorithm progresses toward a solution, the number of sample points is increased to obtain a high-quality solution.

The algorithm that we presented may require a large number of evaluations of the limit-state functions and their gradients. In particular, this is the case for problems involving small failure probabilities. Hence, the algorithm is not applicable to problems with limit-state functions that computationally are extremely costly to evaluate. In such cases, different types of approximation must be introduced (see e.g. Ref.2). For computationally inexpensive limit-state functions, the algorithm appears to be efficient, particularly when the importance sampling option is utilized. In view of the increasing speed of computers and the possibility for parallel processing, we except the algorithm to be applicable in the near future to problems with moderately costly limit-state functions.

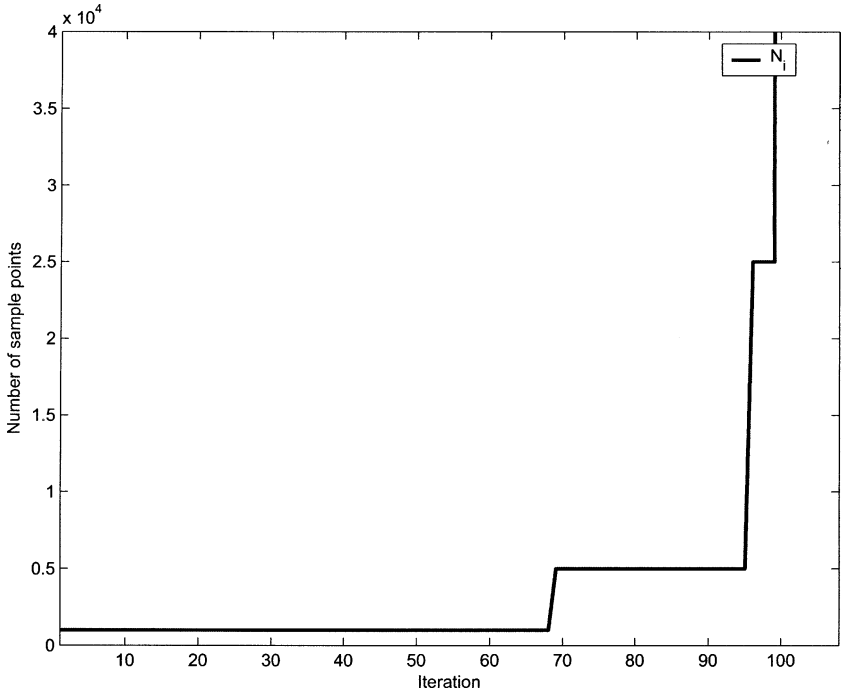

Fig. 3. Increase in the number of sample points N_i.

Table 1. Results for design of rectangulr column.

i	x_i	$c_0(x_i)$	$p_{N_i}(x_i)$	N_i	p_{N_i}
1	(1.00000, 1.00000)	0.25232	0.00003293	1000	0.00000000
68	(0.35658, 0.55134)	0.19660	0.00134986	1000	0.00134639
95	(0.32461, 0.60303)	0.19575	0.00135009	5000	0.00129693
99	(0.31920, 0.61259)	0.19554	0.00134986	$2.5 \cdot 10^4$	0.00129693
103	(0.31549, 0.61919)	0.19535	0.00134988	$1.25 \cdot 10^5$	0.00135338
106	(0.31331, 0.62359)	0.19538	0.00134985	$6.25 \cdot 10^5$	0.00134129
108	(0.31293, 0.62423)	0.19534	0.00134987	$3.125 \cdot 10^6$	0.00134987

7. Appendix: Proofs

Proof of Theorem 4.2. Let $\bar{u}_1, \bar{u}_2, \bar{u}_3, \ldots$ be an infinite sequence of generated realizations of the independent standard normal $(m-1)$-dimensional random vectors $\bar{U}_1, \bar{U}_2, \bar{U}_3, \ldots$. By Theorem 4.1, for every bounded set $S \subset \mathbb{R}^n$ and $\kappa > 0$, there exists a constant $C_S < \infty$ such that, for all

$N \in \mathbb{N}$ and $x \in S$,

$$|p_{k,N}(x;\mu,\sigma) - p_k(x)| \le C_S(\log N)^{\kappa+1/2}/\sqrt{N}. \tag{27a}$$

Hence we proceed now in a deterministic setting. For any $N \in \mathbb{N}$ and $\mu \in \mathbb{R}^{m-1}$, $\sigma > 0$, let

$$\Gamma_N = \left\{x \in X \,|\, p_{k,N}(x;\mu\sigma) \le \hat{p}_k, k \in K\right\}. \tag{27b}$$

In view of Theorems 3.3.2 and 3.3.3 in Ref.23, we need only to show that two conditions hold:

(i) For every $x \in \Gamma = \{x \in X | p_k(x) \le \hat{p}_k \in K\}$, there exists a sequence $\{x_N\}_{N\in\mathcal{N}}$, with $x_N \in \Gamma_N$, such that $x_N \to^{\mathcal{N}} x$ as $N \to \infty$ and $\lim \sup_{N\to\infty} p_{k,N}(x_N;\mu,\sigma) \le p_k(x)$ for all $k \in K$.

(ii) For every infinite sequence $\{x_N\}_{N\in L}$, such that $x_N \in \Gamma_N$ for all $N \in L$ and $x_N \to^L x$ as $N \to \infty$ we have that $x \in \Gamma$ and $\lim \inf_{N\to\infty} p_{k,N}(x_N;\mu,\sigma) \ge p_k(x)$ for all $k \in K$.

First consider (i). Let $x \in \Gamma$ be arbitrary. By Assumption A3 and (27a), $\arg\min_{x'\in\Gamma_N} \|x' - x\| \ne 0$ for sufficiently large N. Hence, we can define $x_N \in \arg\min_{x'\in\Gamma_N} \|x' - x\|$ for all $N \in \mathcal{N}$ without loss of generality. For the sake of a contradiction, suppose that $x_N \not\to^{\mathcal{N}} x$ as $N \to \infty$. Then, there exists a $\delta > 0$ such that $||x_N - x|| \ge \delta$ for all $N \in \mathcal{N}$. By Assumption A3, there exists $\hat{x} \in \mathbb{R}^n$ such that $||\hat{x} - x|| < \delta$ and $f_j(\hat{x}) < 0$ for all $j \in J$ and $p_k(\hat{x}) < \hat{p}_k$ for all $k \in K$. By (27a), there exists an $\hat{N} \in \mathcal{N}$ such that, for all $N > \hat{N}$, $N \in \mathcal{N}$, we have that $\hat{x} \in \Gamma_N$, which is a contradiction. Hence, $x_N \to^{\mathcal{N}} x$ as $N \to \infty$. By Theorem 3.1 $p_k(.)$ is continuous. Consequently, it follows from (27a) that

$$\begin{aligned} |\, p_{k,N}(x_N;\mu,\sigma) - p_k(x)\,| \;\le\; & |\, p_{k,N}(x_N;\mu,\sigma) - p_k(x_N)\,| \\ & + |\, p_k(x_N) - p_k(x)\,| \to^{\mathcal{N}} 0, \end{aligned} \tag{28a}$$

as $N \to \infty$. Hence, the first part of the proof is complete. Next consider (ii). Let $\{x_N\}_{N\in L}$ be an infinite sequence, with $N \subset \mathcal{N}$ and the properties that $x_N \in \Gamma_N$ for all $N \in L$ and $x_N \to^L x$ as $N \to \infty$. Hence, using (28a), we have that

$$p_k(x) = \lim p_{k,N}(x_N;\mu,\sigma) \le 0. \tag{28b}$$

It follows that $x \in \Gamma$. This completes the proof. □

Proof of Proposition 4.1. First, consider (i). By Assumption A5, there exist $N_{x_i}, \in \mathbb{N}$ and $\delta_{x_i} > 0$ such that

$$F_N(x_i, x''; \mu, \sigma) \leq -\delta_{x_i}, \tag{29a}$$

for all $x'' \in A_N(x_i)$ and $N \geq N_{x_i}, N \in \mathcal{N}$. Since $(\log N)^{\kappa+1/2}/\sqrt{N} \to 0$ as $N \to \infty$, there exists $N' \geq N_{x_i}$, such that, for all $N \geq N', N \in \mathcal{N}$, and $x_{i+1} \in A_N(x_i)$,

$$F_N(x_i, x_{i+1}; \mu, \sigma) \leq -\delta_{x_i} \leq -\eta(\log N)^{\kappa+1/2}/\sqrt{N}. \tag{29b}$$

Hence, we see that (23) is satisfied with $N \geq N'$.

Next, consider (ii). For the sake of a contradiction, suppose that the monotone increasing sequence $\{N_i\}_{i=1}^{\infty}$ is bounded. Then, there exists an $i_0 \in \mathbb{N}$ such that $N_i = N_{i_0} < \infty$ for all $i \geq i_0$. To simplify the notation, let $N^* = N_{i_0}$. Then, by the test in (23),

$$F_{N^*}(x_i, x_{i+1}; \mu, \sigma) \leq -\eta(\log N^*)^{\kappa+1/2}/\sqrt{N^*}, \tag{29c}$$

for all $i \geq i_0$. Since it follows from (21a) that

$$\begin{aligned} &\psi_{N^*}(x_{i+1}; \mu, \sigma) - \psi_{N^*}(x_i; \mu, \sigma)_+ \\ &\leq F_{N^*}(x_i, x_{i+1}; \mu, \sigma) \leq -\eta(\log N^*)^{\kappa+1/2}/\sqrt{N^*}, \end{aligned} \tag{29d}$$

for all $i \geq i_0$, we conclude from (29d) that there must exist an $i_1 \geq i_0$ such that $\psi_{N^*}(x_i, \mu, \sigma) \leq 0$ for all $i \geq i_1$. Hence for all $i \geq i_1, \psi_{N^*}(x_i, \mu, \sigma)_+ = 0$; therefore, in view of (21a) and (29c),

$$\begin{aligned} &f_{0,N^*}(x_{i+1}; \mu, \sigma) - f_{0,N^*}(x_i; \mu, \sigma) \\ &\leq F_{N^*}(x_i, x_{i+1}; \mu, \sigma) \leq -\eta(\log N^*)^{\kappa+1/2}/\sqrt{N^*}, \end{aligned} \tag{29e}$$

for all $i \geq i_1$. Hence, we conclude that $f_{0,N^*}(x_i; \mu, \sigma) \to -\infty$ as $i \to \infty$. Since by continuity $f_{0,N^*}(x_i, \mu, \sigma) \to^L f_{0,N^*}(\hat{x}; \mu, \sigma)$ as $i \to \infty$, where $L \subset \mathbb{N}$ is such that $x_i \to^L \hat{x}$ as $i \to \infty$, we have a contradiction. Hence, we must have that $N_i \to \infty$ as $i \to \infty$. □

Proof of Theorem 4.5. Let $\{x_i\}_{i=1}^{\infty}$ be an infinite sequence generated by Algorithm 4.1 with accumulation point $\hat{x}$. Suppose that $\{x_i\}_{i=1}^{\infty} \subset S$, where $S \subset \mathbb{R}^n$ is bounded. Let $C_1 < \infty$ be a constant such that

$$| p_{k,N}(x; \mu, \sigma) - p_k(x) | \leq C_1(\log N)^{k+1/2}/\sqrt{N}, \tag{29f}$$

for all $x \in S$. Similarly to (21a), (21b), (21c), we define

$$F(x', x'') = \max\left\{ f_0(x'') - f_0(x') - \gamma\psi(x')_+, \psi(x'') - \psi(x')_+ \right\}, \tag{29g}$$

where $f_0(\cdot)$ and $\psi(\cdot)$ are defined in (17) and (18) respectively. Hence, there exists a constant $C_S < \infty$ such that, for all $N \in \mathbb{N}, i \in \mathbb{N}$, and $\kappa > 0$,

$$F(x_i, x_{i+1}) \leq F_N, (x_i, x_{i+1}; \mu, \sigma) + C_S(\log N_i)^{\kappa+1/2}/\sqrt{N_i}. \tag{30a}$$

Let $\kappa^* > 0$ be the value of the parameter κ in Algorithm 4.1, Step 1. Hence, because of the imposed condition (23), with κ replaced by κ^*, and because of (30a),

$$\begin{aligned} F(x_i, x_{i+1}) &\leq -\eta(\log\ N_i)^{\kappa^*+1/2}/\sqrt{N_i} + C_S(\log N_i)^{\kappa+1/2}/\sqrt{N_i} \\ &= -(\log N_i)^{\kappa^*+1/2}\left(\eta - C_S(\log N_i)^{\kappa-\kappa^*}\right)/\sqrt{N_i}. \end{aligned} \tag{30b}$$

By Proposition 4.1 and (30b), it follows that, for each $\kappa \in (0, \kappa^*)$, there exists an $i_\kappa \in \mathbb{N}$ such that, for all $i \geq i_\kappa$,

$$F(x_i, x_{i+1}) \leq 0. \tag{30c}$$

Consequently, if $\psi(x_i) > 0$ for all $i \geq i_\kappa$, then it follows from the definition of $F(x_i, x_{i+1})$ that $\{\psi(x_i)\}_{i=i_\kappa}^{\infty}$ is a monotone decreasing sequence with an accumulation point $\psi(\hat{x})$. Therefore, it follows that $\psi(x_i) \to \psi(\hat{x})$ as $i \to \infty$. Alternatively, if there exists $i_1 \geq i_\kappa$ such that $\psi(x) \leq 0$, then because of (30c) and the definition of $F(x_i, x_{i+1})$, $\psi(x_i) \leq 0$ for all $i \geq i_1$ and $\{f_0(x_i)\}_{i=i_1}^{\infty}$ is a monotone decreasing sequence with an accumulation point $f_0(\hat{x})$. Consequently, $f_0(x_i) \to f_0(\hat{x})$ as $i \to \infty$.

Now, for the sake of a contradiction, suppose that $\theta(\hat{x}) < 0$ and that $L \subset \mathbb{N}$ is such that $x_i \to^L \hat{x}$ as $i \to \infty$ Then, by Assumption A5, there exist an $i_2 \in \mathcal{N}$ and a $\delta_{\hat{x}} > 0$ such that, for all $i \in L, i \geq i_2$,

$$F_{N_i}(x_i, x_{i+1}; \mu, \sigma) \leq -\delta_{\hat{x}} < 0; \tag{30d}$$

hence, because of (30a),

$$F(x_i, x_{i+1}) \leq C_S(\log N_i)^{\kappa+1/2}/\sqrt{N_i} - \delta_{\hat{x}}. \tag{30e}$$

Since, by Proposition 4.1,

$$(\log N_i)^{\kappa+1/2}/\sqrt{N_i} \to 0, \quad \text{as } i \to \infty,$$

if follows from (30e) that there exists $i_3 \geq i_2$ such that $F(x_i, x_{i+1}) \leq -\delta_{\hat{x}}/2$ for all $i \geq i_3, i \in L$. But this contradicts the fact that, for $\kappa \in (0, \kappa^*)$, either $\psi(x_i) \to \psi(\hat{x})$ as $i \to \infty$ or $f_0(x_i) \to f_0(\hat{x})$ as $i \to \infty$. Hence, $\theta(\hat{x}) = 0$. □

Lemma 7.1. Let Assumptions A2 and A4 hold. Let $\mu \in \mathbb{R}^{m-1}$ and $\sigma > 0$ be the sampling parameters in (8a) and (8b). For every bounded set $S \subset \mathbb{R}^n$, there exists a constant $C_S < \infty$ such that, for $x \in S$ and $\kappa > 0$,

$$|\theta_N^*(x;\mu,\sigma) - \theta(x)| \leq C_S(\log N)^{\kappa+1/2}/\sqrt{N}, \quad \text{w.p.1}, \tag{31a}$$

where

$$\begin{aligned}\theta_N^*(x;\mu,\sigma) = -\min_{\nu\in\sum} \Bigg\{ & \sum_{k\in K} \nu^k [\psi_N^*(x;\mu,\sigma)_+ - p_{k,N}^*(x;\mu,\sigma) + \hat{p}_k] \\ & + \sum_{j\in J} \nu^{j+K_g} [\psi_N^*(x;\mu,\sigma)_+ - f_j(x)] \\ & + (1/2\delta) \| \sum_{k\in K} \nu^k \nabla p_{k,N}^*(x;\mu,\sigma) \\ & + \sum_{j\in J} \nu^{j+K_g} \nabla f_j(x) + \nu^{K_g+J_f+1} \nabla f_{0,N}^*(x;\mu,\sigma) \|^2 \\ & + \nu^{K_g+J_f+1} \gamma \psi_N^*(x;\mu,\sigma)_+ \Bigg\},\end{aligned} \tag{31b}$$

with

$$f_{0,N}^*(x;\mu,\sigma) = c_0(x) + \sum_{k\in K} c_k(x) p_{k,N}^*(x;\mu,\sigma), \tag{31c}$$

$$\psi_N^*(x;\mu,\sigma) = \max\left\{ \max_{k\in K} \left\{ p_{k,N}^*(x;\mu,\sigma) - \hat{p}_k \right\}, \max_{j\in J} f_j(x) \right\}, \tag{31d}$$

$$\psi_N^*(x;\mu,\sigma)_+ = \max\left\{0, \psi_N^*(x;\mu,\sigma)\right\}. \tag{31e}$$

Proof. Let $\bar{u}_1, \bar{u}_2, \bar{u}_3, \ldots$ be an infinite sequence of generated realizations of the independent standard normal $(m-1)$-dimensional random vectors $\bar{U}_1, \bar{U}_2, \bar{U}_3, \ldots$. By Theorems 4.1 and 4.4, for every bounded set $S \subset \mathbb{R}^n$ and $\kappa > 0$, there exist constants $C_1, C_4 < \infty$ such that, for all $N \in \mathcal{N}$ and $x \in S$, (29f) and

$$\| \nabla p_{\kappa,N}(x;\mu,\sigma) - \nabla p_\kappa(x) \| \leq C_4(\log N)^{\kappa+1/2}/\sqrt{N} \tag{31f}$$

hold. Hence, we can proceed now in a deterministic setting.

For every $h \in \mathbb{R}^n$, let

$$\begin{aligned}&\tilde{\psi}_N(x, x+h;\mu,\sigma) \\ &= \max\left\{ -\gamma\psi_N(x;\mu,\sigma)_+ + \langle \nabla f_{0,N}(x;\mu,\sigma), h\rangle, \right. \\ &\quad \left. \max_{k\in K} \left\{ p_{k,N}(x;\mu,\sigma) - \hat{p}_k - \psi_N(x;\mu,\sigma)_+ \right.\right.\end{aligned}$$

$$+\langle \nabla p_{k,N}(x;\mu,\sigma), h\rangle\big\},$$
$$\max_{j\in J}\{f_j(x)-\psi_N(x;\mu,\sigma)_+ + \langle \nabla f_j(x), h\rangle\}\}$$
$$+\delta\,||\,h\,||^2/2, \tag{31g}$$

$$\tilde{\psi}(x,x+h) = \max\{-\gamma\psi(x)_+ + \langle \nabla f_0(x), h\rangle,$$
$$\max_{k\in K}\{p_k(x)-\hat{p}_k-\psi(x)_+ + \langle \nabla p_k(x), h\rangle\},$$
$$\max_{j\in J}\{f_j(x)-\psi(x)_+ + \langle \nabla f_j(x), h\rangle\}\}$$
$$+\delta\,||\,h\,||^2/2. \tag{31h}$$

Now, using (29f) and (31f), we obtain that there exists a constant $C_S<\infty$ such that, for all $x\in S$ and $N\in\mathbb{N}$,

$$|\max_{k\in K}\{p_{k,N}(x;\mu,\sigma)-\hat{p}_k-\psi_N(x;\mu,\sigma)_+ + \langle \nabla p_{k,N}(x;\mu,\sigma), h\rangle\}$$
$$-\max_{k\in K}\{p_k(x)-\hat{p}_k-\psi(x)_+ + \langle \nabla p_k(x), h\rangle\}\,|$$
$$\le 2C_1(\log N)^{\kappa+1/2}/\sqrt{N} + C_4(\log N)^{\kappa+1/2}\,||\,h\,||/\sqrt{N}, \tag{31i}$$

$$|\,(-\gamma\psi_N(x;\mu,\sigma)_+ + \langle \nabla f_{0,N}(x;\mu,\sigma), h\rangle)$$
$$-(-\gamma\psi(x)_+ + \langle \nabla f_0(x), h\rangle)\,|$$
$$\le \gamma C_1(\log N)^{\kappa+1/2}/\sqrt{N} + C_S(\log N)^{\kappa+1/2}\,||\,h\,||/\sqrt{N}. \tag{31j}$$

Hence, for all $h\in\mathbb{R}^n$, $x\in S$, and $N\in\mathbb{N}$,

$$|\,\tilde{\psi}_N(x,x+h;\mu,\sigma)-\tilde{\psi}(x,x+h)\,|$$
$$\le \max\{\gamma,2\}C_1(\log N)^{\kappa+1/2}/\sqrt{N}$$
$$+\max\{C_4,C_S\}(\log N)^{\kappa+1/2}\,||\,h\,||/\sqrt{N}. \tag{31k}$$

Next, let $h_N(x;\mu,\sigma)$ be given by (24d) and let $h(x)$ be given by

$$h(x) = -(1/\delta)\left(\sum_{k\in K}\hat{v}^k\nabla p_k(x) + \sum_{j\in J}\hat{v}^{j+K_g}\nabla f_j(x) + \hat{v}^{K_g+J_f+1}\nabla f_0(x)\right), \tag{31l}$$

where $\hat{v}=(\hat{v}^1,\ldots,\hat{v}^{K_g+J_f+1})$ is any solution of (16). Then, $h(x)$ is bounded for all $x\in S$, because it is defined as a linear combination of bounded vector-valued functions. By the same argument and the fact that (31f) holds, $h_N(x;\mu,\sigma)$ is bounded for all $x\in S$ and $N\in\mathcal{N}$. From Theorem 2.2.8 of Ref. 23, we have that

$$\theta_N(x;\mu,\sigma) = \min_{h\in\mathbb{R}^n}\tilde{\psi}_N(x,x+h;\mu,\sigma), \tag{31m}$$

$$\theta(x) = \min_{h \in \mathbb{R}^n} \tilde{\psi}(x, x+h). \tag{31n}$$

Hence, there exists $C_S^* < \infty$ such that, for all $x \in \mathbb{R}^n$,

$$\begin{aligned}
\theta(x) &\leq \tilde{\psi}(x, x + h_N(x;\mu,\sigma)) \\
&\leq \tilde{\psi}_N(x, x + h_N(x;\mu,\sigma);\mu,\sigma) \\
&\quad + (\max\{\gamma, 2\}C_1 + \max\{C_4, C_S\}\,||\,h_N(x;\mu,\sigma)\,||) \\
&\quad \times (\log N)^{\kappa+1/2}/\sqrt{N} \\
&\leq \theta_N(x;\mu,\sigma) + C_S^*(\log N)^{\kappa+1/2}/\sqrt{N}, \qquad (31o) \\
\theta(x) &= \tilde{\psi}(x, x + h(x)) \\
&\geq \tilde{\psi}_N(x, x + h(x);\mu,\sigma) \\
&\quad - (\max\{\gamma, 2\}C_1 + \max\{C_4, C_S\}\,||\,h(x)\,||)(\log N)^{\kappa+1/2}\sqrt{N} \\
&\geq \theta_N(x;\mu,\sigma) - C_S^*(\log N)^{K+1/2}\sqrt{N}. \qquad (31p)
\end{aligned}$$

This completes the proof. □

Proof of Preposition 4.2. Let $\bar{u}_1, \bar{u}_2, \bar{u}_3, \ldots$ be an infinite sequence of generated realizations of the independent standard normal $(m-1)$-dimensional random vectors $\bar{U}_1, \bar{U}_2, \bar{U}_3, \ldots$. Let S be any bounded set in $\mathbb{R}^n$ with corresponding constants $C_1, C_4 < \infty$ such that (29f) and (31f) hold for all $x \in S$ and $N \in \mathbb{N}$. Hence, we can proceed now in a deterministic setting.

The search direction $h_N(x;\mu,\sigma)$ is bounded for all $x \in S$, because it is defined as a linear combination of bounded vector-valued functions [see (24d) and (31f)]. The bound is independent of N due to (31f). By the mean-value theorem, we obtain that

$$\begin{aligned}
&F_N(x, x + \lambda h_N(x;\mu,\sigma);\mu,\sigma) \\
&\quad = \max\big\{-\gamma\psi_N(x;\mu,\sigma)_+ + \lambda\langle\nabla f_{0,N}(x;\mu,\sigma), h_N(x;\mu,\sigma)\rangle \\
&\quad + \lambda\int_0^1 \langle\nabla f_{0,N}(x + s\lambda h_N(x;\mu,\sigma);\mu,\sigma) \\
&\qquad - \nabla f_{0,N}(x;\mu,\sigma), h_N(x;\mu,\sigma)\rangle ds\big\}, \\
&\max_{k\in K}\{p_{k,N}(x;\mu,\sigma) - \hat{p}_k - \psi_N(x;\mu,\sigma)_+ \\
&\quad + \lambda\langle\nabla p_{k,N}(x;\mu,\sigma), h_N(x;\mu,\sigma)\rangle \\
&\quad + \lambda\int_0^1 \langle\nabla p_{k,N}(x + s\lambda h_N(x;\mu,\sigma);\mu,\sigma) \\
&\qquad - \nabla p_{k,N}(x;\mu,\sigma), h_N(x;\mu,\sigma)\rangle ds\big\}, \\
&\max_{j\in J}\{f_j(x) - \psi_N(x;\mu,\sigma)_+ + \lambda\langle\nabla f_j(x), h_N(x;\mu,\sigma)\rangle
\end{aligned}$$

$$+\lambda \int_0^1 \langle \nabla f_j(x + s\lambda h_N(x;\mu,\sigma)) - \nabla f_j(x), h_N(x;\mu,\sigma)\rangle ds\}\Big\}. \tag{32a}$$

By (31f) and the fact that $\nabla_{p_{k,N}}(\cdot;\mu,\sigma)$ and $\nabla p_\kappa(\cdot)$ are uniformly continuous on bounded sets, we have from (32a) that, for any $\epsilon > 0$, there exists a $\lambda_0 \in (0, 1]$ such that, for all $\lambda \in (0, \lambda_0]$, $x \in S$, and $N \in \mathbb{N}$,

$$\begin{aligned}
&F_N(x, x + \lambda h_N(x;\mu,\sigma);\mu,\sigma)\\
&\le \max\Big\{-\gamma\psi_N(x;\mu,\sigma)_+ + \langle \nabla f_{0,N}(x;\mu,\sigma), h_N(x;\mu,\sigma)\rangle,\\
&\max_{k\in K}\{p_{k_N}(x;\mu,\sigma) - \hat{p}_k - \psi_N(x;\mu,\sigma)_+\\
&+\langle \nabla p_{k,N}(x;\mu,\sigma), h_N(x;\mu,\sigma)\rangle,\\
&\max_{j\in J}\{f_j(x) - \psi_N(x;\mu,\sigma)_+ + \langle \nabla f_j(x), h_N(x;\mu,\sigma)\rangle\Big\}\\
&+\delta \| h_N(x;\mu,\sigma)\|^2/2 + \lambda(\epsilon - \delta\|h_N(x;\mu,\sigma)\|^2/(2\lambda)).
\end{aligned} \tag{32b}$$

Since $h_N(x;\mu,\sigma)$ is bounded for all $x \in S$ and $N \in \mathbb{N}$, there exists a $\lambda_1 \in (0, \lambda_0]$ such that

$$\epsilon - \delta\|h_N(x;\mu,\sigma)\|^2/(2\lambda) \le 0, \tag{32c}$$

for all $\lambda \in (0, \lambda_1]$, $x \in S$, and $N \in \mathbb{N}$. Hence, by using (31f) and (31l), it follows that

$$\begin{aligned}
&F_N(x, x + \lambda h_N(x;\mu,\sigma)) - \alpha\lambda\theta_N(x;\mu,\sigma)\\
&\quad \le (1-\alpha)\lambda\theta_N(x;\mu,\sigma) \le 0,
\end{aligned} \tag{32d}$$

for all $\lambda \in (0, \lambda_1]$, $x \in S$, $N \in \mathbb{N}$, and algorithm parameter $\alpha \in (0, 1]$. Consequently, for any $x \in S$, the Polak-He algorithm computes a stepsize $\lambda_N(x;\mu,\sigma) \ge \beta\lambda_1$. Hence, for any $x' \in S$ and $x'' \in A_N(x')$,

$$F_N(x', x'') \le \alpha\lambda_N(x';\mu,\sigma)\theta_N(x';\mu,\sigma) \le \alpha\beta\lambda_1\theta_N(x';\mu,\sigma). \tag{32e}$$

Suppose that $\hat{x} \in \mathbb{R}^n$ is that $\theta(\hat{x}) < 0$. Since $\theta(\cdot)$ is continuous (Theorem 2.2.8 in Ref. 23), there exists a $\rho > 0$ such that

$$\theta(x) \le \theta(\hat{x})/2, \text{ for all } x \in \mathbb{B}(\hat{x}, \rho) = \{x \in \mathbb{R}^n \mid \|x - \hat{x}\| \le \rho\}.$$

By Lemma 7.1, there exists a $C < \infty$ such that

$$|\theta_N(x;\mu,\sigma) - \theta(x)| \le C(\log N)^{\kappa+1/2}/\sqrt{N}, \tag{32f}$$

for all $x \in \mathbb{B}(\hat{x}, \rho)$. Hence, for all $x \in \mathbb{B}(\hat{x}, p)$,

$$\theta_N(x;\mu,\sigma) \le \theta(x) + C(\log N)^{k+1/2}/\sqrt{N}. \tag{32g}$$

Since $\theta(\cdot)$ is continuous and the right-hand side of (32f) vanishes as $N \to \infty$, it follows from (32g) that there exists a $\hat{\rho} \in (0, \rho]$ and an $\hat{N} \in \mathbb{N}$ such that

$$\theta_N(x; \mu, \sigma) \leq \theta(\hat{x}), \quad \text{for all } x \in \mathbb{B}(\hat{x}, \hat{\rho}) \text{ and } N \geq \hat{N}.$$

Let the bounded set $S = \mathbb{B}(\hat{x}, \hat{\rho})$. Then, we conclude from (32e) that, for all $x' \in \mathbb{B}(\hat{x}, \hat{\rho})$, $N \geq \hat{N}$, and $x'' \in A_N(x')$,

$$F_N(x', x'') \leq \alpha\beta\lambda_1\theta_N(x'; \mu, \sigma) \leq \alpha\beta\lambda_1\theta(\hat{x})/4 = -\delta_{\hat{x}} < 0. \tag{32h}$$

This completes the proof. □

References

1. Gasser, M., and Schueller, G.I., *Some Basic Principles in Reliability-Based Optimization (RBO) of Structures and Mechanical Components*, Stochastic Programming Methods and Technical Applications, Edited by K. Marti and P. Kall, Springer, Berlin, Germany, 1998.
2. Torczon, V., and Trosset, M.W., *Using Approximations to Accelerate Engineering Design Optimization*, AIAA Paper 98–4800, Proceedings of the 7th AIAA/ USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, St. Louis, Missouri, 1998.
3. Enevoldsen, I., and Sorensen, J.D., *Reliability- Based Optimization In Structural Engineering*, Structural Safety, Vol. 15, pp. 169–196,1994.
4. Kuschel, N., and Rackwitz, R., *Optimal Design under Time-Variant Reliability Constraints*, Structural Safety, Vol. 22, pp. 113–127, 2000.
5. Royset, J.O., Der Kiureghian, A., and Polak, E., *Reliability-Based Optimal Structural Design by the Decoupling Approach*, Reliability Engineering and System Safety, Vol. 73, pp. 213–221, 2001.
6. Prekopa, A., *Stochastic Programming*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1995.
7. Birge, J.R., and Louveaux, F., *Introduction to Stochastic Programming*, Springer, New York, NY, 1997.
8. Ermoliev, Yu., *Stochastic Quasigradient Methods*, Numerical Techniques for Stochastic Optimization, Edited by Yu. Ermoliev and R.J.B. Wets, Springer, New York, NY, pp. 141–186,1988.
9. Gaivoronski, A., *Implementation of Stochastic Quasigradient Methods*, Numerical Techniques for Stochastic Optimization, Edited by Yu. Ermoliev and R.J.B. Wets, Springer, New York, NY, pp. 313–352, 1998.
10. Benveniste, A., Metivier, M., and Priouret, P., *Adaptive Algorithms and Stochastic Approximations*, Springer, New York, NY, 1990.
11. Marti, K., *Solving Stochastic Structural Optimization Problems By RSM-Based Stochastic Approximation Methods: Gradient Estimation In Case of*

Intermediate Variables, Mathematical Methods of Operations Research, Vol. 46, pp. 409–434, 1997.

12. Tretiakov, G., *Stochastic Quasigradient Algorithms for Maximization of the Probability Function: A New Formula for the gradient of the Probability Function*, Stochastic Optimization Techniques, Edited by K. Marti, Springer, New York, NY, 2002.
13. King, A.J., and Wets, R.J.B., *Epiconvergence of Convex Stochastic Programs*, Stochastics, Vol. 34, pp. 83–91, 1991.
14. Higle, J.L., and Sen, S., *Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse*, Mathematics of Operations Research, Vol. 16, pp. 650–669,1991.
15. King, A.J., and Rockafellar, R.T., *Asymptotic Theory for Solutions in Statistical Estimation and Stochastic Programming*, Mathematics of Operations Research, Vol. 18, pp. 148–162, 1993.
16. Infanger, G., *Planning under Uncertainty, Solving Large-Scale Stochastic Linear Programs*, Boyd and Fraser Publishers, Boston, Massachusetts,1994.
17. Shapiro, A., and Homem-De-Mello, T., *A Simulation- Based Approach to Two-Stage Stochastic Programming with Recourse*, Mathematical Programming, Vol. 81, pp. 301–325, 1998.
18. Mak, W.K., Morton, D.P., and Wood, R.K. *Monte Carlo Bounding Techniques for Determining Solution Quality in Stochastic Programs*, Operations Research Letters, Vol. 24, pp. 47–56, 1999.
19. Shapiro, A., *Stochastic Programming by Monte Carlo Simulation Methods,* Stochastic Programming E-Print Series, 2000; see http://dochost.rz.huberlin.de/speps/.
20. Ditlevsen, O., and Madsen, H.O., *Structural Reliability Methods*, Wiley, New York, NY, 1996.
21. Polak, E., Wets, R.J.B., Der Kiureghian, A., *On the Approach to Optimization of Problems with Probabilistic Cost and Constraints*, Nonlinear Optimization and Related Topics, Edited by G. Di Pillo and F. Giannessi, Publisher, Town, State, pp. 299–316, 2000.
22. Pollard, D., *Convergence of Stochastic Processes*, Springer, New York, NY, 1984.
23. Polak, E., *Optimization: Algorithms and Consistent Approximations*, Springer, New York, NY, 1997.
24. Matlab, MathWorks, Natick, Massachusetts, Matlab Version 6.5, Optimization Toolbox 2.2, 2002; see http:/www.mathworks.com.