Advances in Engineering Software 89 (2015) 3–16

Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

Non-parametric stochastic subset optimization utilizing multivariate boundary kernels and adaptive stochastic sampling

Gaofeng Jia, Alexandros A. Taflanidis*

Department of Civil & Environmental Engineering & Earth Sciences, University of Notre Dame, 156 Fitzpatrick Hall, Notre Dame, IN 46556, USA

ARTICLE INFO

Article history:
Available online 20 July 2015

Keywords:
Non-parametric stochastic subset optimization
Optimization under uncertainty
Kernel density estimation
Adaptive kernel sampling density
Stochastic sampling
Simulation-based optimization

ABSTRACT

The implementation of NP-SSO (non-parametric stochastic subset optimization) to general design under uncertainty problems and its enhancement through various soft computing techniques is discussed. NP-SSO relies on iterative simulation of samples of the design variables from an auxiliary probability density, and approximates the objective function through kernel density estimation (KDE) using these samples. To deal with boundary correction in complex domains, a multivariate boundary KDE based on local linear estimation is adopted in this work. Also, a non-parametric characterization of the search space at each iteration using a framework based on support vector machine is formulated. To further improve computational efficiency, an adaptive kernel sampling density formulation is integrated and an adaptive, iterative selection of the number of samples needed for the KDE implementation is established.

1. Introduction

In any engineering design application, the performance predictions for the system under consideration involve some level of uncertainty, stemming from our incomplete knowledge about the system itself and its environment (representing future excitations). Explicitly accounting for these uncertainties is important for providing optimal configurations that exhibit robust-performance [1-3]. A probabilistic approach provides a rational and consistent framework for performing this task [4], employing probability models to characterize different possible model-parameters values. In this setting, the design objective is typically related to the expected value of a system performance measure, such as failure probability or expected life-cycle cost [2,5-8]. For applications involving complex system models this expected value can rarely be calculated or accurately approximated analytically [9]. For such applications stochastic simulation techniques, which pose no constraints on the complexity of the adopted numerical and probability models, are frequently the only applicable general approach [10,11] for estimating the objective function. This approach, though, entailing a large number of evaluations of the system performance for each objective function estimation, might impose a computational cost that is prohibitive for applications involving computationally expensive models.

To address this challenge, this paper investigates an alternative optimization algorithm for system design optimization under uncertainty, termed non-parametric stochastic subset optimization (NP-SSO). NP-SSO, proposed recently [12] for design problems utilizing the system reliability as the objective function, also relies on stochastic simulation, but rather than calculating the performance objective for specific values of the design variables, establishes a global approximation for it. It is an extension of the SSO algorithm [13], and like SSO it relies on simulation of a sufficient number of samples of the design variables from an auxiliary probability density function, treating them artificially as uncertain. Rather than using the information in these samples to establish an approximation for the average value of the objective function over preselected subsets, as established in SSO, NP-SSO utilizes kernel density estimation (KDE) [14] to approximate the objective function and identify candidate points for the global minimum. An iterative approach is also established within NP-SSO to improve computational efficiency: at each iteration a new compact domain is identified as candidate subset for the optimal design variables and the search is then confined within this domain. A parametric description for that subset, corresponding to a box-bounded domain, was adopted in [12] to support the proposed KDE implementation.

In this paper, based upon [15], NP-SSO is extended to general design under uncertainty problems (not constrained to reliability-optimization), and more importantly it is coupled with various soft-computing techniques to improve its non-parametric characteristics and its numerical efficiency. Specifically, the following advances are introduced here for NP-SSO. A versatile, non-parametric characterization of the subset identified at each iteration is investigated through adoption of KDE with multivariate boundary kernels (KDE-MBK) [16]. To achieve this goal, a support vector machine (SVM) [17] is additionally adopted to facilitate (i) efficient simulation of samples and (ii)

* Corresponding author. Tel.: +1 5746315696.
E-mail address: a.taflanidis@nd.edu (A.A. Taflanidis).

http://dx.doi.org/10.1016/j.advengsoft.2015.06.014
0965-9978/

Nomenclature

AKSD	adaptive kernel sampling density
$\mathbf{x}$	design variable vector
x_i	ith design variable
n_x	dimension of $\mathbf{x}$
X	admissible design space
$\mathbf{x}^*$	optimal design solution
$\boldsymbol{\theta}$	uncertain model parameters
n_θ	dimension of $\boldsymbol{\theta}$
Θ	space of possible values for $\boldsymbol{\theta}$
$p(.)$	probability density function
$h(\mathbf{x}, \boldsymbol{\theta})$	performance measure
$C(\mathbf{x})$	objective function
I	bounded search space
$\hat{C}(\mathbf{x})$	approximation to objective function through stochastic simulation
V_I	volume of set I
$\tilde{C}(\mathbf{x})$	approximation to objective function through NP-SSO
$E_{p(\mathbf{x})p(\boldsymbol{\theta})}[h(\mathbf{x}, \boldsymbol{\theta})]_I$	average value for objective function for $\mathbf{x}$ belonging in I
$\hat{E}_{p(\mathbf{x})p(\boldsymbol{\theta})}[h(\mathbf{x}, \boldsymbol{\theta})]_I$	approximation to $E_{p(\mathbf{x})p(\boldsymbol{\theta})}[h(\mathbf{x}, \boldsymbol{\theta})]_I$ through stochastic simulation
$\pi_I(\mathrm{x}, \boldsymbol{\theta})$	auxiliary density function for $\mathbf{x}$ and $\boldsymbol{\theta}$ with $\mathbf{x}$ belonging in I
$\{\mathbf{x}^j_{..}, \boldsymbol{\theta}^j_{..}\}$	jth sample from $\mathbf{x}_{..}$ or $\boldsymbol{\theta}_{..}$
$\{\mathbf{x}_h, \boldsymbol{\theta}_h\}$	sample set for $\{\mathbf{x}, \boldsymbol{\theta}\}$ from $\pi_I(\mathrm{x}, \boldsymbol{\theta})$
$\{\mathbf{x}_c, \boldsymbol{\theta}_c\}$	sample set for which the system response is evaluated during the stochastic sampling process
$\pi_I(\mathrm{x})$	auxiliary density function for $\mathbf{x}$ with $\mathbf{x}$ belonging in I
$\pi_I(\boldsymbol{\theta})$	auxiliary density function for $\boldsymbol{\theta}$ with $\mathbf{x}$ belonging in I
n_s	number of samples available from $\pi_I(\mathrm{x})$
$\{\mathbf{x}_h\}$	sample set from $\pi_I(\mathrm{x})$
N	number of samples in set $\{\mathbf{x}_c, \boldsymbol{\theta}_c\}$
K_m	product multivariate kernel
$q(\cdot)$	proposal density for $\boldsymbol{\theta}$ for obtaining samples
$S(\mathbf{x})$	support domain for kernel K_m for $\mathbf{x}$
w_i	bandwidth of kernel for x_i
$\underline{K}_m$	multivariate kernel with boundary correction
$K(\cdot)$	univariate kernel for approximation of $\pi_I(\mathrm{x})$
c_0, c_{1i}, d_{kl}	coefficients needed for boundary correction
σ_i	standard deviation of samples $\{\mathbf{x}_h\}$ for ith design variable
I^*	subset of I with $\tilde{C}(\mathbf{x}) \leq c^h$
$\delta(I^*\|I)$	volume ratio between sets I^* and I
$\{\mathbf{x}_u\}$	uniform samples in I
ρ	target volume ratio for $\delta(I^*\|I)$ per iteration of NP-SSO
n_u	number of samples in $\{\mathbf{x}_u\}$
$\{\mathbf{x}_p, \boldsymbol{\theta}_p\}$	samples available for formulation of proposal densities for stochastic sampling in the new search space
n_p	number of samples in $\{\mathbf{x}_p, \boldsymbol{\theta}_p\}$
$H(I^*\|I)$	ratio of average objective function values in I^* and I
$\hat{H}(I^*\|I)$	approximation to $H(I^*\|I)$ using stochastic simulation
$\{\mathbf{x}_o\}$	uniform samples in I^*
c^h	threshold defining subset I^*
n_o	number of samples in $\{\mathbf{x}_o\}$
$\{\mathbf{x}_e, \boldsymbol{\theta}_e\}$	samples for which $h(\mathbf{x}, \boldsymbol{\theta})$ is known for guiding the formulation of proposal densities
n_e	number of samples in $\{\mathbf{x}_e, \boldsymbol{\theta}_e\}$
D_{re}	relative entropy
$K_G(.)$	univariate kernel for AKSD
λ^j	local bandwidth factor for jth sample
$\mathbf{d}$	vector with kernel characteristics
$\mathbf{v}$	subset of $\boldsymbol{\theta}$ targeted for the AKSD
n_v	size of $\mathbf{v}$
t_i	bandwidth for ith component of $\mathbf{v}$
α	sensitivity factor
s_e	percentage reduction for definition of D_{min}
D_{min}	minimum relative entropy considered
c^s	stopping threshold for $H(I^*\|I)$
$.._k$(subscript k)	kth iteration of NP-SSO characteristics
n_{sa}	number of samples per stage for adaptive selection of n_s
n_b	stage count for adaptive selection of n_s
KDE	kernel density estimation
M	number of clusters for $\{\mathbf{x}_o\}$
$..^m$(superscript m)	mth cluster of $\{\mathbf{x}_o\}$ characteristics
ρ_δ	accuracy threshold for identification of subset I^*
$n_{s,max}$	maximum allowable number of samples considered for n_s

estimation of boundary correction terms within the proposed iterative scheme. Furthermore, an adaptive kernel sampling density (AKSD) is adopted to improve the efficiency of the stochastic sampling stage that is required within NP-SSO to generate the necessary samples from the design variables. The characteristics for the AKSD are chosen here to explicitly optimize the anticipated sampling efficiency, utilizing readily available information to perform this optimization. Finally, an adaptive selection of the number of samples needed for the KDE approximation is proposed. This is established through a multi-stage process; a new set of samples is obtained at each stage and then a new KDE approximation is established and a new domain identified (within the same always iteration of NP-SSO). If this domain does not differ significantly from the previous one (obtained utilizing a smaller number of samples), the KDE approximation has sufficient accuracy and the multi-stage sampling stops. The advances proposed here, especially the AKSD and the adaptive selection of the number of samples for the KDE approximation, contribute greatly to the efficiency of NP-SSO.

In the next section the general design under uncertainty problem is reviewed and in Section 3 the NP-SSO framework is presented for such problems. The proposed advances for facilitating complex subset selection are developed in Section 4 whereas in Section 5 an adaptive stochastic sampling implementation is seamlessly integrated in the framework. In Section 6 the adaptive selection of the number of samples for KDE-MBK is presented and the overall adaptive NP-SSO algorithm is reviewed. Finally, Section 7 presents an illustrative example.

2. Design under uncertainty optimization

Consider a system that involves some controllable parameters that define its design, referred to also as design variables and let $\mathbf{x} = [x_1 x_2 \ \ldots \ x_{n_x}] \in X \subset \Re^{n_x}$ be the design vector where X denotes the bounded admissible design space. Let $\boldsymbol{\theta} = [\theta_1 \theta_2 \ldots \theta_{n_\theta}]$ lying in $\Theta \subset \Re^{n_\theta}$ be the vector of uncertain model parameters for the system, where Θ denotes the set of their possible values. A PDF