10th World Congress on Structural and Multidisciplinary Optimization
May 19 -24, 2013, Orlando, Florida, USA

Worst-case design of structures using stopping rules in k-adaptive random sampling approach

Makoto Yamakawa[1], Makoto Ohsaki[2]

[1] Tokyo Denki University, Tokyo, Japan, myamakawa@mail.dendai.ac.jp
[2] Hiroshima University, Hiroshima, Japan, ohsaki@hiroshima-u.ac.jp

Abstract
In the practical design process, uncertainty in the parameters should be appropriately taken into account. In particular, our interest focuses on design problems with dynamic analysis under uncertainty. One of the authors presented a worst-case design of structures based on a random sampling approach, in which constraints are assigned on the worst values of the structural responses. In this case, the optimization problem turns out to be a two-level problem of optimization and anti-optimization. Key concept of the method is estimation of the worst value by Random Search (RS) with order of the function values. The RS is the simplest and most obvious search method; however, it has been also pointed out that the RS is impractical or less efficient. One of the reasons is the difficulty of predicting the exact extremes from only the small samples. Indeed, the theoretical average waiting time of a new record of extremes is infinite and by discarding other record values we lose an enormous amount of information. We may not need such exact extremes in many practical situations. Aiming to improve the efficiency, we relax from the worst value to the k th value. Then we can predict and control more accurately the behavior of RS by using the k th order statistics. In a numerical example, we verify the validity of the estimation of the k th worst value for adequate values of k and number of samples, which is referred to as the stopping rule of RS. The result indicates use of relatively small samples is enough to predict the large number of future samples even for two-level problem with the noisy and non-smooth dynamic analysis under uncertain parameters.

Keywords: uncertainty, anti-optimization, random search, order statistics

1. Introduction

In the practical design process, uncertainty in the parameters should be appropriately taken into account. In particular, our interest focuses on design problems with dynamic analysis under uncertainties. For example, earthquakes are essentially uncertain phenomena. Lack of taking uncertainties of the seismic motions into consideration may cause severe damage in building structures. It is known that transient dynamic analysis is not suitable for conventional design optimization owing to the high computational cost, the difficulty of sensitivity analysis and non-smoothness of the response functions (e.g., [1]). Furthermore, the uncertainty of the inputs makes matters worse. It is highly desired to establish a design method for robust optimization which is applicable to problems with time-consuming dynamic analysis under uncertain parameters.

For such purpose, one of the authors presented a worst-case design of structures based on a random sampling approach, in which constraints are assigned on the worst values of the structural responses [2]. In this case, the optimization problem turns out to be a two-level problem of optimization and anti-optimization. It was shown in the numerical examples that a good approximate optimal solution is found by the method with relatively small number of analyses when compared with Genetic Algorithm (GA) and Tabu Search (TS) [3]. Key concept of the method is estimation of the worst value by Random Search (RS) with order of the function values.

The RS is the simplest and most obvious search method; however, there are still several important issues to deal with. One of them is how we choose the stopping rule. An approach is to estimate the closeness of the current record value of the objective function to its worst value, where the maximum value among the observations is referred to as a record value. It has been also pointed out that RS is impractical or less efficient. Indeed, the theoretical average waiting time of a better record is infinite and by discarding other record values we lose an enormous amount of information. One of the reasons is the difficulty of predicting the exact extremes from only the small samples. We may not need such exact extremes and prefer to know approximately good solution quickly in many practical situations. In the previous study, the author used a stopping rule based on the Chernoff bound under unknown probability distributions. A more sophisticated approach is to estimate the confidence intervals for it by statistical procedures. We attempt to clarify the method proposed in [3] by using the k th order statistics and provide its rigorous theoretical background.

2. Two-level optimization problem

We first define a target optimization problem. The problem is a two-level problem of optimization and anti-optimization, which is quite similar to the one in [3], but we treat continuous uncertainties in lower-level anti-optimization problem. That seems intuitively to be a natural choice. Consider a problem of optimizing some structural and/or material parameters of building structures, e.g., the cross-sections, the story shear strength, the additional damping and so on, which may be selected from the pre-assigned list of standard specifications. The properties are classified into m groups, each of which has the same values. The design variable vector is denoted by $\boldsymbol{J} = (J_1,\cdots,J_m)$, which has integer values. We incorporate uncertainties into the problem. The vector consisting of uncertain parameters is denoted by $\Theta = (\theta_1,\cdots,\theta_r)$, where r is the number of the uncertain parameters. In particular, we assume that $\Theta \in \Omega \subset \mathbb{R}^r$ and $0 < \mathrm{vol}(\Omega) < \infty$, i.e., the uncertain parameters are continuous and bounded. Our formulation is different on this point since the uncertain models are assumed to be discrete variables in [3].

For simplicity, the uncertainties are incorporated into only one constraint function $g(\boldsymbol{J},\Theta)$, which is chosen to be a structural response function. The other constraint functions are denoted by $h_i(\boldsymbol{J})$ $(i = 1,\cdots,l)$, where l is the number of constraints without uncertainties. The objective function is denoted by $\Phi(\boldsymbol{J})$. The problem we are interested in, may be described as

$$\left.\begin{array}{lll} \underset{\boldsymbol{J}\in\mathbb{Z}^m}{\text{Minimize}} & \Phi(\boldsymbol{J}) & \\ \text{subject to} & g(\boldsymbol{J},\Theta) \le \overline{g} & \text{for all } \Theta \in \Omega \\ & h_i(\boldsymbol{J}) \le \overline{h}_i & (i = 1,\cdots,l) \end{array}\right\} \tag{1}$$

where $\overline{g}$ and $\overline{h}_i$ $(i = 1,\cdots,l)$ are given constants. Following the terminology in [2], we translate a two-level problem of optimization and anti-optimization from the original problem (1). In this approach, the lower-level anti-optimization problem is formulated for finding the worst value of the structural response. The worst value is obtained by solving the following anti-optimization problem:

$$\text{Find} \quad g_{\max}(\boldsymbol{J}) = \max_{\Theta\in\Omega} g(\boldsymbol{J},\Theta). \tag{2}$$

By using (2), we can formulate the upper-level optimization problem as

$$\left.\begin{array}{lll} \underset{\boldsymbol{J}\in\mathbb{Z}^m}{\text{Minimize}} & \Phi(\boldsymbol{J}) & \\ \text{subject to} & g_{\max}(\boldsymbol{J}) \le \overline{g} & \\ & h_i(\boldsymbol{J}) \le \overline{h}_i & (i = 1,\cdots,l) \end{array}\right\} \tag{3}$$

This type of formulations is well known in robust optimization and referred to as worst-case approach, bilevel optimization [5], or robust counterpart approach [6] and so on. In the approaches, we work with so called "uncertain-but-bounded" data model [7]. The model has significant methodological advantage but it has been also pointed out that the worst-case-oriented approach decisions can be too conservative and thus impractical.

Ohsaki [3] presented that a random sampling approach can be successfully applied to obtain optimal and anti-optimal solutions within the prescribed accuracy. They mentioned that the solution was obtained by using "approximate worst solution" which is the k th value of the samples. The approach utilizes the order of objective value. This framework is not the worst-case approach in a narrow sense and also not so called "chance constraints" [7] (which is a typical probabilistic model). We focus on the lower-level anti-optimization problem and in particular more clarification of the relation between the accuracy and the stopping rules proposed in [3].

3. Random sampling and probabilistic constraints

To make the short description, we omit the dependence of a given design variable $\boldsymbol{J}$ on the constraint function g because we focus on the anti-optimization problem. Then the constraint corresponding to the anti-optimization problem is given as

$$g(\Theta) \le \overline{g} \quad \text{for } \Theta \in \Omega. \tag{4}$$

Constraint (4) is referred to as "worst-case constraint" and can be rewritten by

$$g_{\max} \le \overline{g}, \tag{5}$$

where

$$g_{\max} = \max_{\Theta \in \Omega} g(\Theta) . \tag{6}$$

We apply RS to solve Problem (6) for obtaining worst solution in some senses. In an algorithm of RS, a sequence of random points $\Theta_1, \Theta_2, \cdots, \Theta_n$ is generated where for each j, $1 \le j \le n$, the point Θ_j has some probability distribution P_j. We will refer to this general scheme as Algorithm 1 (see, e.g., [4]).

Algorithm 1 (general random sampling algorithm)

1. Generate a random point Θ_1 according to a probability distribution P_1 on Ω; evaluate the objective function at this point; set iteration counter $j = 1$.
2. Using the points $\Theta_1, \cdots, \Theta_j$ and the results of the objective function evaluation at these points, check whether $j = n$; that is, check an appropriate stopping condition. If this condition holds, terminate the algorithm.
3. Alternatively, generate a random point Θ_{j+1} according to some probability distribution P_{j+1} and evaluate the objective function at Θ_{j+1}.
4. Substitute $j+1$ for j and return to step 2.

In the algorithm which is often called "Pure Random Search (PRS)" all the distribution P_j are the same (that is, $P_j = P$ for all j) and the points Θ_j are independent; otherwise, the algorithm may be called "Adaptive Random Search" [8]. As a result of the application of PRS we obtain an independent sample $\{\Theta_1, \cdots, \Theta_n\}$ from a distribution P on Ω. Additionally, we obtain an independent sample $\{Y_1 = g(\Theta_1), \cdots, Y_n = g(\Theta_n)\}$ of the objective function values at these points. The samples Y_j are independent identically distributed random variables (iidrv) with the continuous cumulative distribution function (cdf) of a random variable Y which is given by

$$F(t) = \Pr\left\{\Theta \in \Omega : g(\Theta) \le t\right\} = \int_{g(\Theta) \le t} P(d\Theta) = \int_{-\infty}^{t} f(y) dy , \tag{7}$$

where the function f is formally defined as the probability density function (pdf). The iidrv $Y_1, \cdots, Y_n$ are arranged in increasing order of magnitude and the k th value is denoted by $Y_{k,n}$ such that

$$Y_{1,n} \le Y_{2,n} \le \cdots \le Y_{n,n} .$$

For anti-optimization problem, Ohsaki [3] proposed use of the following constraint:

$$Y_{k,n} \le \overline{g} . \tag{8}$$

If we set $k = n$ and $n \to \infty$, Eq.(8) is the same as worst-case constraint (5) with probability one. In next section we will present that Eq.(8) has a close relation with

$$\Pr\left\{F(\overline{g}) \ge \gamma\right\} \ge \beta , \tag{9}$$

where both β and γ are preassigned constants $(0 \le \beta, \gamma \le 1)$. Implications of Eq.(9) are clearer than Eq.(8). Before taking up the main subject, we point out that Eq.(9) is generalization of the chance constraint and the worst-case constraint. When we set $\gamma = 1$, Eq.(9) corresponds to the chance constraint:

$$\Pr\left\{Y \le \overline{g}\right\} \ge \beta . \tag{10}$$

When we set $\beta = \gamma = 1$, Eq.(9) corresponds to the worst-case constraint with probability one:

$$\Pr\left\{Y \le \overline{g}\right\} = 1 . \tag{11}$$

Note that Constraint (11) is not entirely the same as worst-case constraint (5); however, both are the same in the sense of probability. In next section we will show that the number of samples n and the k th value in the samples are closely related to the parameters β and γ. This means that the accuracy of the solution is given as parameters β and γ, by which the stopping rule of Algorithm 1 is decided.

4. Stopping rules in random sampling

We first summarize the basic theory of the k th order statistics. Form the theory, distribution-free tolerance intervals are derived. Then, we present a stopping rules of RS based on the distribution-free interval.

4.1. Distribution-free tolerance interval

The iidrv $Y_1, \cdots, Y_n$ with the common cdf F are arranged in order of magnitude and then written as

$$Y_{1,n} \le Y_{2,n} \le \cdots \le Y_{n,n}, \tag{12}$$

we call $Y_{k,n}$ the k th order statistics $(k = 1, 2, \cdots, n)$, see, e.g., [9][10]. The cdf F is assumed to be continuous but unknown. Let $F_{k,n}(y)$, $(k = 1, \cdots, n)$ denote the cdf of the k th order statistics $Y_{k,n}$. Then we obtain

$$F_{k,n}(y) = \Pr\{Y_{k,n} \le y\} = I_{F(y)}\left(k, n-k+1\right), \tag{13}$$

where $I_p(a,b)$ is incomplete beta function which is given by

$$I_p(a,b) = \frac{1}{B(a,b)} \int_0^p t^{a-1}(1-t)^{b-1} dt, \tag{14}$$

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt. \tag{15}$$

For $0 < \gamma < 1$, define the γ th quantile of Y with the cdf F to be

$$\xi_\gamma = F^{-1}(\gamma), \tag{16}$$

where we define as

$$F^{-1}(\gamma) = \inf\left\{y : F(y) \ge \gamma\right\}. \tag{17}$$

If $F(y)$ is strictly increasing, $F(\xi_\gamma) = \gamma$ holds; otherwise, it holds on some interval. Since $Y_{k,n} < \xi_\gamma$ implies $Y_{k,n} \le \xi_\gamma$, we have

$$\begin{aligned} \Pr\left\{F(Y_{k,n}) \ge \gamma\right\} &= \Pr\left\{Y_{k,n} \ge F^{-1}(\gamma)\right\} \\ &= 1 - \Pr\left\{Y_{k,n} \le \xi_\gamma\right\} \\ &= 1 - I_\gamma\left(k, n-k+1\right). \end{aligned} \tag{18}$$

If we choose the values of n and k such that

$$1 - I_\gamma\left(k, n-k+1\right) \ge \beta, \tag{19}$$

we have

$$\Pr\left\{F(Y_{k,n}) \ge \gamma\right\} \ge \beta. \tag{20}$$

From Eq.(20), $(-\infty, Y_{k,n})$ is known as one-sided "tolerance intervals" (see, e.g., [9]). Clearly, this interval is not dependent on the cdf F and hence the interval is called "distribution-free intervals". Eq.(20) implies the probability is β or more that the interval $(-\infty, Y_{k,n})$ contains at least a proportion γ of the population if Eq.(19) is satisfied. In other words, if the observed value of $Y_{k,n}$ is $y_{k,n}$, one feels at least $100\beta\%$ confident that the true value of the γ th quantile ξ_γ line in the interval $(-\infty, y_{k,n})$, that is,

$$\Pr\left\{\xi_\gamma \in (-\infty, y_{k,n})\right\} \ge \beta. \tag{21}$$

If it holds that $\overline{g} \ge y_{k,n}$ and the condition (19) is satisfied, we have

$$\Pr\left\{\xi_\gamma \in (-\infty, \overline{g})\right\} \ge \Pr\left\{\xi_\gamma \in (-\infty, y_{k,n})\right\} \ge \beta, \tag{22}$$

then we can say that at least $100\beta\%$ confident that at least a proportion γ of the population is less than $\overline{g}$. This rule is not dependent on the cdf F. We do not need to know the cdf F, but which must be continuous.

4.2. Stopping rules of random sampling approach

Stopping rules of Algorithm 1 can be said to be the selection of k and n. We can choose k and n such that the condtion (19) is satisfied a little in excess of β as possible, which is easily found by enumeration method. It is known that the extremes in finite small samples are very sensitive to outlying points. So care must be taken in placing these. For avoiding the difficulties, we find $k, n \in \mathbb{N}$ by solving the following problem:

$$\left.\begin{array}{ll} \underset{k,n}{\text{Minimize}} & n \\ \text{subject to} & 1 - I_\gamma\left(k, n-k+1\right) \geq \beta \\ & 1 \leq k \leq n - \epsilon \end{array}\right\} \tag{23}$$

where $\epsilon \in \mathbb{N}$ is the margin for avoiding extreme order statistics. Intuitively, it seems to be reasonable to use

$$k(n) = \left[(n+1)\gamma\right], \tag{24}$$

where $[x]$ is the greatest integer less than or equal to x. Indeed, the 100γ th sample percentile $Y_{k(n),n}$ becomes a point estimator of ξ_γ for sufficiently large n; however, in finite small samples Eq.(24) does not satisfy the condition (19). Hence, it is plausible to use Eq.(24) as an initial value for solving Problem (23).

5. Numerical example

We show a numerical example, consisting of a seismic design problem using a 10-story shear model and 2 input earthquake motions. The problem is defined as:

$$\left.\begin{array}{lll} \underset{\boldsymbol{x}}{\text{Minimize}} & \Phi(\boldsymbol{x}) = \sum_{i=1}^{10} x_i & \\ \text{subject to} & g(\boldsymbol{x}, \Theta) = \delta_{\max}\left(\boldsymbol{x} + \Theta\right) \leq \overline{g} & \\ & 0.05 \leq x_i \leq 0.4 & (i = 1, \cdots, 10) \\ & \Theta \in \Omega & \end{array}\right\} \tag{25}$$

where design variable vector $\boldsymbol{x} = (x_1, \cdots, x_{10})$ represent the story shear coefficients, $\delta_{\max}(\cdot)$ denotes the maximum story drift angle and $\overline{g}$ is given as 0.01 radian. The configuration of this 10-story shear model is shown in Fig.1 and the relation between i th story drift angle and i th story shear force is given as shown in Fig.2. The maximum story drift angle $\delta_{\max}(\cdot)$ is obtained by nonlinear transient dynamic analysis, in which input earthquake motions are shown in Fig.3. The input accelerations are highly noisy and non-smooth.

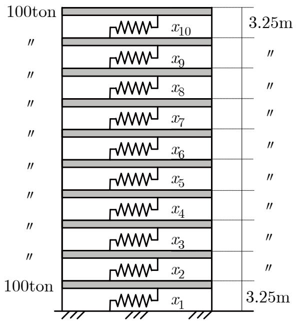

Figure 1: 10-story shear model

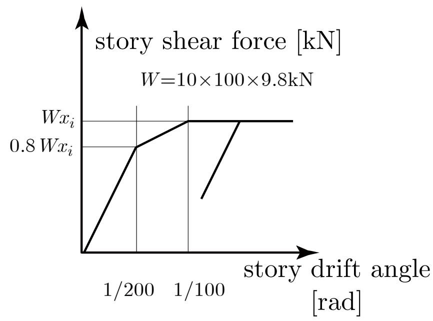

Figure 2: Force-deformation relation

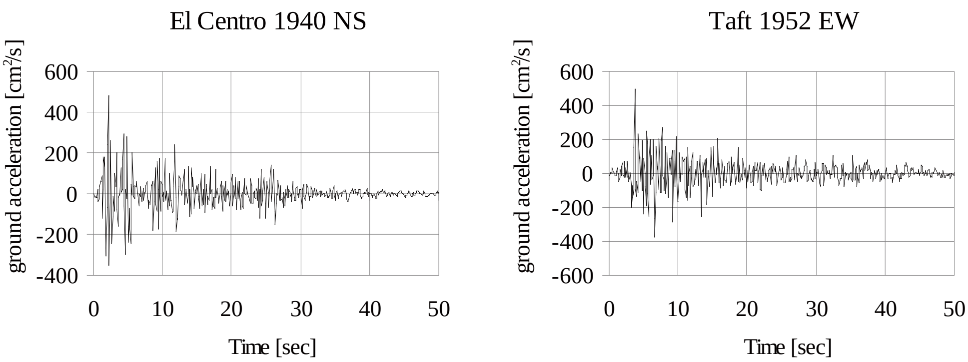

Figure 3: Input earthquake motions (ground acceleration)

Since our interest focuses on the lower-level anti-optimization problem, we only consider following problem:

$$\left.\begin{array}{ll}\underset{\Theta}{\text{Maximize}} & g(\Theta) = \delta_{\max}\left(\boldsymbol{x} + \Theta\right) \\ \text{subject to} & -0.05\boldsymbol{x} \le \Theta \le 0.05\boldsymbol{x}\end{array}\right\} \tag{26}$$

where $\boldsymbol{x}$ is a vector of fixed parameters given by

$$\boldsymbol{x} = \left(0.44, 0.42, 0.40, 0.36, 0.33, 0.30, 0.26, 0.22, 0.17, 0.10\right), \tag{27}$$

which is obtained as a solution of the nominal design optimization without uncertainties as shown in [1]. When we apply Algorithm 1 to Problem (26) and set $\beta = \gamma = 0.9$ and $\epsilon = 10$, the corresponding parameters are decided by solving Problem (23), which is found as

$$(k, n) = (142, 152).$$

Thus we need at least 152 samples. So we stop Algorithm 1 when the number of the observed samples reaches 152. The corresponding upper bound of the tolerance interval is the 142 th value of the samples. The k th value of the total n samples is denoted by $y_{k,n}$. Then in above case the one-sided tolerance interval is denoted by $(-\infty, y_{142,152})$. The value and the extreme values are obtained as

$$y_{1,152} = 0.0092, \quad y_{142,152} = 0.0114, \quad y_{152,152} = 0.0121,$$

where the unit of the values is radian. To summarize, we can say that

$$\Pr\left\{\text{More than 90\% of the future samples is lsss than } 0.0114 \text{ in values}\right\} \ge 0.9. \tag{28}$$

This mean that the solution (27) is surely infeasible in Problem (25). For testing validity of Eq.(28), we further apply Algorithm 1 to the problem until the total number of the observed samples reaches $10,000$. The numbers of population within the interval and proportions to the total samples are obtained as

$$\#\left\{y_i \le 0.0114, i = 1,\cdots,10,000\right\} = 9030,$$

$$\frac{\#\left\{y_i \le 0.0114, i = 1,\cdots,10,000\right\}}{10,000} = 0.903,$$

where y_i denotes the i th observed sample value, which is not order statistics. Thus we could obtain very accurate result, i.e., we could predict the result of large number of future samples from use of relatively small samples. For reference, the extreme values are shown as follows:

$$y_{1,10000} = 0.0090, \quad y_{10000,10000} = 0.0131.$$

Note that we focus on not the true worst value but the proportion within the interval. Hence there is difference between the values of $y_{142,152}$ and $y_{10000,10000}$. We can only say that 90% of the samples are surely within the interval $(-\infty, y_{142,152})$ for large number of future samples. In general, prediction of the extreme values from small samples is not easy and we may not need such exact extremes in many practical situations. The presented method is useful in such situations.

6. Conclusion

We presented a clear definition of the approximation of the worst values based on the k th order statistics and proposed a framework of anti-optimization with new stopping rules of random sampling approach. We investigated the behavior of the stopping rules through a numerical example. The results are as follows:

1. The distribution-free tolerance interval can be used as a stopping rule of random sampling approach for solving anti-optimization problems. It is not needed to know the distribution of the samples to apply the method.
2. We investigated the applicability of the presented method to a seismic design problem with transient dynamic analysis, which is not suitable for conventional design optimization. For such problem, RS is useful.
3. The numerical result indicates use of relatively small samples is enough to predict the large number of future samples even for two-level problem with the complicated dynamic analysis under uncertain parameters.

Acknowledgements
This work was supported by a grant from the Obayashi Foundation. We gratefully acknowledge the support.

References

[1] M. Yamakawa and F. van Keulen, *Derivative-Free Optimization Reflecting the Geometry of Constraints for Nonsmooth Dynamic Responses*, 9th World Congress on Structural and Multidisciplinary Optimization, ISSMO, Shizuoka, Japan, 2011.
[2] I. Elishakoff and M. Ohsaki, *Optimization and anti-optimization of structures under uncertainty*, World Scientific, London, 2010.
[3] M. Ohsaki and M. Katsura, A random sampling approach to worst-case design of structures, *Structural and Multidisciplinary Optimization*, 46 (1), 27-39, 2012.
[4] A. Zhigljavsky and A. Zilinskas, *Stochastic global optimization*. Springer, New York: 2010.
[5] B. Colson, P. Marcotte, and G. Savard, *An overview of bilevel optimization*, *Annals of operations research*, 153(1), 235-256, 2007
[6] H.G. Beyer and S. Bernhard, Robust optimization–a comprehensive survey. *Computer methods in applied mechanics and engineering* 196(33), 3190-3218, 2007.
[7] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton University Press, New Jersey, 2009.
[8] Z.B. Zabinsky, *Stochastic adaptive search for global optimization*. Kluwer Academic Pub, Massachusetts, 2003.
[9] H.A. David and N.N. Haikady, *Order statistics*, 3rd Ed, John Wiley & Sons, Inc., New Jersey, 2003.
[10] P.G. Hoel, *Introduction to mathematical statistics*, 7th Ed, Pearson Education, Inc.,Boston, 2013.