A stochastic gradient descent approach for risk optimization using the Chernoff bound

André G. Carlon[1], Henrique M. Kroetz[2], André J. Torii[3], Rafael H. Lopez[4], Leandro F. F. Miguel[4]

[1]*Computer, Electrical and Mathematical Sciences & Engineering Division, King Abdullah University of Science & Technology (KAUST)*
4700 King Abdullah University of Science and Technology, Thuwal, 3955-6900, Makkah Province, Saudi Arabia
agcarlon@gmail.com
[2]*Center for Marine Studies, Federal University of Paraná (UFPR)*
Pontal do Paraná, 83255-000, Paraná, Brazil
henrique.kroetz@ufpr.br
[3]*Latin American Institute for Technology, Infrastructure and Territory, Federal University for Latin American Integration (UNILA)*
Edifício Comercial Lorivo - Av. Silvio Américo Sasdelli, 1842 - Vila A, Foz do Iguaçu, 85866-000, Paraná, Brazil
ajtorii@hotmail.com
[4]*Center for Optimization and Reliability in Engineering (CORE), Department of Civil Engineering, Federal University of Santa Catarina (UFSC)*
R. Eng. Agronômico Andrei Cristian Ferreira, s/n - Trindade, Florianópolis, 88040-900, Santa Catarina, Brazil
rafael.holdorf@ufsc.br, leandro.miguel@ufsc.br

Abstract. We propose a method for solving Risk Optimization (RO) problems based on the Stochastic Gradient Descent (SGD) methods. SGD is used to minimize the expectation of functions. We approximate each limit state function in the RO problem using the Chernoff bound, thus recasting the original RO problem as an expectation minimization problem. The Chernoff bound approximation requires the evaluation of Monte Carlo sampling, which could be expensive. However, once the Chernoff bound parameters are set, they can be used to cheaply approximate the probabilities of failure of each state limit for several iterations. We propose a heuristic approach to tune the Chernoff bound parameters after a distance from the last update. Moreover, we decay the update distance each iteration, thus guaranteeing that the probabilities of failure approximations are accurate as SGD converges to the optimum solution. We present numerical results supporting the efficiency of our approach to different RO problems with applications in structural engineering. Comparisons of SGD equipped with our Chernoff bound approximation against particle swarm optimization using sample average approximation validate the efficiency of the proposed approach.

Keywords: Stochastic Gradient Descent, Risk Optimization, Chernoff Bound.

1 Introduction

When designing engineering processes, it is often important to consider uncertainties. The two main approaches to optimizing engineering processes considering uncertainties are Reliability-Based Design Optimization (RBDO) and Risk Optimization (RO) [1–3]. In RBDO, one sets the desired reliability level and treats it as a constraint. The RO approach, on the other hand, explicitly incorporates the probabilities of failure on the objective function. Here, we focus on RO, given its generality.

Computing the probability of failure in engineering processes is usually impractical due to the high cost of performing a large number of simulations. The most popular approach is to use sample average approximation by approximating the original problem using Monte Carlo Sampling (thus turning the problem deterministic). Then, solve the problem using standard deterministic optimization methods like Genetic Algorithm, Particle Swarm Optimization, among others [4–6]. This approach has the drawback of requiring a large Monte Carlo sample to be precise, dramatically increasing the cost of each optimization iteration.

Instead of splitting the original problem in a uncertainty quantification and a (deterministic) optimization problem, we focus on solving both problems simultaneously by employing the stochastic gradient descent (SGD) method. The SGD works by sampling the gradient independently each iteration and then controlling the statistical error by decreasing the step-size. The main advantages of the SGD are its low cost per iteration and its ability to converge in high dimensional settings in the random parameters and design spaces. Recently, SGD has been applied to engineering [7–9] and optimal experimental design [10, 11] with success. Moreover, some variations of SGD like Adam [12] are known to converge even in the non-convex multimodal case.

Applying SGD to RO problems is not straightforward since SGD methods require an unbiased gradient estimator to converge. Even though the objective function of RO problems can be written as an expectation by rewriting the probabilities of failure as expectations of indicator functions, this approach results in the loss of continuity of the objective function and is not of practical use. Here, we advance on a previous work by the authors [13] and use the Chernoff bound to approximate the probabilities of failure of each failure mode. Since the Chernoff bound provides an upper bound of the probabilities of failure, we estimate not only the upper bound, defined by a constant, as the slack of the inequality, providing an approximation that is as good as the Monte Carlo Approximation used to tune the Chernoff bound parameters. Moreover, we provide a heuristic approach to update the Chernoff parameters during optimization in such a way as to not overburden the optimization approach with the extra cost.

While on [13] the authors solve some numerical problems using the same method presented here, these do not fully take advantage of the SGD main strength, which is its ability to solve large dimensional problems. Here, we apply Adam to RO problems using the Chernoff bound to a problem with 25 design variables and 12 random parameters. The numerical results support the validity of our approach to solving larger problems.

2 Risk Optimization (RO)

In RO, one is interested in the design parameters $\boldsymbol{d}$ in the search space $\mathcal{D}$ that minimize the objective function f, i.e.,

$$\text{Find } \boldsymbol{d}^* = \underset{\boldsymbol{d}\in\mathcal{D}}{\arg\min}\, f(\boldsymbol{d}), \tag{1}$$

where

$$f(\boldsymbol{d}) \stackrel{\text{def}}{=} C_0(\boldsymbol{d}) + \sum_{i=1}^{m} C_i(\boldsymbol{d}) P_{f,i}(\boldsymbol{d}), \tag{2}$$

The uncertainties are gathered in a random vector $\boldsymbol{X} \in \mathcal{X}$ with probability distribution $\mu_{\boldsymbol{X}}$. For each limit state function g_i, we define the probability of failure as

$$P_{f,i}(\boldsymbol{d}) \stackrel{\text{def}}{=} P(g_i(\boldsymbol{d},\boldsymbol{X}) > 0) \tag{3}$$

$$= \int_{g_i(\boldsymbol{d},\boldsymbol{X})>0} \mu_{\boldsymbol{X}}(\boldsymbol{d},\boldsymbol{X}) d\boldsymbol{X}. \tag{4}$$

For a limit state function $g_i \stackrel{\text{def}}{=} S_i/R_i$, analogous to the one used in [14, Equation 1.15c], we have

$$P_{f,i}(\boldsymbol{d}) = P\left(S_i(\boldsymbol{d},\boldsymbol{X}) > R_i(\boldsymbol{d},\boldsymbol{X})\right) \tag{5}$$

$$= P\left(\frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})} > 1\right). \tag{6}$$

2.1 Chernoff bound on the probabilities of failure

The Chernoff bound is a direct application of the Markov inequality over the exponential of a random variable [15]. Assuming a random variable Y, Chernoff's bound can be written as

$$P(Y \geq a) \leq \frac{\mathbb{E}\left(e^{tY}\right)}{e^{ta}},\ t > 0. \tag{7}$$

We can use (7) to bound the probability of failure. Thus, from (6), we can write

$$P_{f,i}(\boldsymbol{d}) \leq \bar{P}_{f,i}(\boldsymbol{d}) \stackrel{\text{def}}{=} \min_{t_i>0} \frac{\mathbb{E}\left[e^{t_i \frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})}}\right]}{e^{t_i}}. \tag{8}$$

Since the upper-bound $\bar{P}_{f,i}$ might not be tight, we define a correction factor γ_i such that

$$\hat{P}_{f,i}(\boldsymbol{d}) \stackrel{\text{def}}{=} \gamma_i \bar{P}_{f,i}(\boldsymbol{d}) \approx P_{f,i}(\boldsymbol{d}). \tag{9}$$

To avoid direct evaluation of the probability of failure, we use the bound $\hat{P}_{f,i}$ during the optimization procedure. Then, instead of directly minimizing f as presented in (2), we minimize $\hat{f}$ given by

$$\hat{f}(\boldsymbol{d}) \stackrel{\text{def}}{=} C_0(\boldsymbol{d}) + \sum_{i=1}^{m} C_i(\boldsymbol{d})\hat{P}_{f,i}(\boldsymbol{d}). \tag{10}$$

By solving (10) instead of (2), we avoid evaluation of the probabilities of failure and its sensitivities. The next section shows how to efficiently minimize (10) with SGD.

3 Stochastic Gradient Descent (SGD)

The SGD is an optimization method (or rather a family of methods) used to solve stochastic optimization problems, where the objective function to be minimized is an expectation. The main advantage of SGD is that it does not use the true gradient of the objective function but an unbiased estimate of the true gradient instead. Here, to minimize the function $\hat{f}$ of RO, we define an auxiliary function F satisfying

$$\mathbb{E}[F(\boldsymbol{d}, \boldsymbol{X})] = \hat{f}(\boldsymbol{d}). \tag{11}$$

We thus take F as

$$F(\boldsymbol{d}, \boldsymbol{X}) \stackrel{\text{def}}{=} C_0(\boldsymbol{d}) + \sum_{i=1}^{m} \gamma_i C_i(\boldsymbol{d}) \, \frac{e^{t_i \frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})}}}{e^{t_i}}. \tag{12}$$

From (11), we observe that (cf. Appendix B in [9])

$$\begin{aligned} \nabla_{\boldsymbol{d}} \hat{f}(\boldsymbol{d}) &= \nabla_{\boldsymbol{d}} \mathbb{E}[F(\boldsymbol{d}, \boldsymbol{X})] & (13)\\ &= \mathbb{E}[\nabla_{\boldsymbol{d}} F(\boldsymbol{d}, \boldsymbol{X}) + F(\boldsymbol{d}, \boldsymbol{X}) \nabla_{\boldsymbol{d}} \log(\mu_{\boldsymbol{X}}(\boldsymbol{d}, \boldsymbol{X}))]. & (14) \end{aligned}$$

By defining the gradient estimator as

$$\nabla_{\boldsymbol{d}} \mathcal{F}(\boldsymbol{d}, \boldsymbol{X}) \stackrel{\text{def}}{=} \nabla_{\boldsymbol{d}} F(\boldsymbol{d}, \boldsymbol{X}) + F(\boldsymbol{d}, \boldsymbol{X}) \nabla_{\boldsymbol{d}} \log(\mu_{\boldsymbol{X}}(\boldsymbol{d}, \boldsymbol{X})), \tag{15}$$

we see that $\nabla_{\boldsymbol{d}} \mathcal{F}$ is an unbiased estimate for the gradient of $\hat{f}$. For this reason, $\nabla_{\boldsymbol{d}} \mathcal{F}$ can be taken as a search direction for SGD.

The SGD basic update rule for $F : \mathcal{D} \times \mathcal{X} \mapsto \mathbb{R}$ that satisfies (11) is then given by

$$\begin{aligned} &\text{Sample } \boldsymbol{X} \sim \mu_{\boldsymbol{X}} & (16)\\ &\boldsymbol{d}_{k+1} = \boldsymbol{d}_k - \alpha_k \nabla_{\boldsymbol{d}} \mathcal{F}(\boldsymbol{d}, \boldsymbol{X}), & (17) \end{aligned}$$

where α_k is a sequence of decreasing step-sizes [9]. From (12) we have

$$\begin{aligned} \nabla_{\boldsymbol{d}} F(\boldsymbol{d}, \boldsymbol{X}) &= \nabla_{\boldsymbol{d}} C_0(\boldsymbol{d}) + \nabla_{\boldsymbol{d}} \left(\sum_{i=1}^{m} \gamma_i C_i(\boldsymbol{d}) \, \frac{e^{t_i \frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})}}}{e^{t_i}} \right) & (18)\\ &= \nabla_{\boldsymbol{d}} C_0(\boldsymbol{d}) + \sum_{i=1}^{m} \left(\gamma_i \nabla_{\boldsymbol{d}} \left(C_i(\boldsymbol{d}) \right) \, \frac{e^{t_i \frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})}}}{e^{t_i}} + \gamma_i C_i(\boldsymbol{d}) \, \frac{\nabla_{\boldsymbol{d}} e^{t_i \frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})}}}{e^{t_i}} \right), & (19) \end{aligned}$$

with

$$\nabla_{\boldsymbol{d}} e^{t_i \frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})}} = t_i e^{t_i \frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})}} \, \frac{\nabla_{\boldsymbol{d}} S_i(\boldsymbol{d}, \boldsymbol{X}) R_i(\boldsymbol{d}, \boldsymbol{X}) - S_i(\boldsymbol{d}, \boldsymbol{X}) \nabla_{\boldsymbol{d}} R_i(\boldsymbol{d}, \boldsymbol{X})}{R_i(\boldsymbol{d}, \boldsymbol{X})^2}. \tag{20}$$

3.1 Adam algorithm

The Adam algorithm [12] is employed here for the minimization of (10) since it has been observed that it is more robust than the original SGD update rule in the presence of noisy gradients [12]. Adam incorporates small modifications to SGD basic update rule as follows:

$$\text{Sample } \boldsymbol{X} \sim \mu_{\boldsymbol{X}} \tag{21}$$
$$\boldsymbol{m}^{(k+1)} = \beta_1 \boldsymbol{m}^{(k)} + (1-\beta_1)\nabla_{\boldsymbol{d}}\mathcal{F}(\boldsymbol{d}^{(k)}, \boldsymbol{X}) \tag{22}$$
$$\boldsymbol{v}^{(k+1)} = \beta_2 \boldsymbol{v}^{(k)} + (1-\beta_2)(\nabla_{\boldsymbol{d}}\mathcal{F}(\boldsymbol{d}^{(k)}, \boldsymbol{X}))^2 \tag{23}$$
$$\hat{\boldsymbol{m}} = \frac{\boldsymbol{m}^{(k+1)}}{1-\beta_1^{k+1}} \tag{24}$$
$$\hat{\boldsymbol{v}} = \frac{\boldsymbol{v}^{(k+1)}}{1-\beta_2^{k+1}} \tag{25}$$
$$\boldsymbol{d}^{(k+1)} = \boldsymbol{d}^{(k)} - \frac{\alpha_0}{\sqrt{k}}\frac{\hat{\boldsymbol{m}}}{\sqrt{\hat{\boldsymbol{v}}}+\epsilon}, \tag{26}$$

where $0 < \beta_1 < \beta_2 < 1$ and ϵ are parameters to be defined by the user.

3.2 Setting the parameters t_i and γ_i of the Chernoff bound

Note that the tightness of the Chernoff bound in (8) depends on the parameter $t > 0$. In order to get an optimal Chernoff bound, the parameter that minimizes the bound should be employed. Thus, for each limit state, the parameter t_i at iteration (k) is obtained from the minimization problem

$$\text{Find } t_i^{(k)} = \underset{t_i>0}{\arg\min}\ \frac{\mathbb{E}\left[e^{t_i \frac{S_i(\boldsymbol{d},\boldsymbol{X})}{R_i(\boldsymbol{d},\boldsymbol{X})}}\right]}{e^{t_i}}. \tag{27}$$

Once t_i is found, the correction coefficient γ_i for each failure mode can be estimated from

$$\gamma_i = \frac{P_{f,i}(\boldsymbol{d}, \boldsymbol{X})}{\bar{P}_{f,i}(\boldsymbol{d}, \boldsymbol{X})}, \tag{28}$$

Calibrating t and γ requires sampling $\boldsymbol{X}$ and evaluating S and R, which can be costly. Thus, being $\overline{k}$ the iteration when they were last tuned, we update the parameters only if the distance of $\boldsymbol{d}^{(k)}$ to $\boldsymbol{d}^{(\overline{k})}$ exceeds a decaying value,

$$\text{Find } t^{(k)}, \gamma_i^{(k)} \text{ if } \left\|\boldsymbol{d}^{(k)} - \boldsymbol{d}^{(\overline{k})}\right\| \geq \frac{\epsilon_t}{\sqrt{k}}, \tag{29}$$

with ϵ_t being a parameter to be set.

To mitigate the noise in the parameters, we use the moving averages

$$\bar{t}_i^{(k)} \stackrel{\text{def}}{=} \left\lceil \frac{2}{k} \right\rceil \sum_{j=\lfloor \frac{k}{2} \rfloor}^{k} t_i^{(j)}, \quad \bar{\gamma}_i^{(k)} \stackrel{\text{def}}{=} \left\lceil \frac{2}{k} \right\rceil \sum_{j=\lfloor \frac{k}{2} \rfloor}^{k} \gamma_i^{(j)}. \tag{30}$$

4 Size optimization of a truss with Gaussian loads

This example consists in finding the cross-section areas of the 25 elements of a truss structure in order to minimize its volume subject to a penalization on the probabilities of its two failure modes: by compression and by traction. The loads are applied in four nodes and each of the three components of each load are modelled as independent Gaussian variables. The RO objective function to be minimized here is

$$\text{Find } \boldsymbol{d}^* = \underset{\boldsymbol{d}\in\mathcal{D}}{\arg\min} \sum_{e=1}^{25} L_e A_e(\boldsymbol{d}) + 10^5 (P_{f,t}(\boldsymbol{d}) + P_{f,c}(\boldsymbol{d})), \tag{31}$$

where L_e and A_e are, respectively, the length and cross-section area of the e-th element; $P_{f,t}$ is the probability of failure due to traction; and $P_{f,c}$ is the probability of failure due to compression. Failure is considered if the stress on any element exceeds a threshold stress,

$$P_{f,t} = P(\max(\sigma) > f_t), \quad P_{f,c} = P(\min(\sigma) < f_c), \tag{32}$$

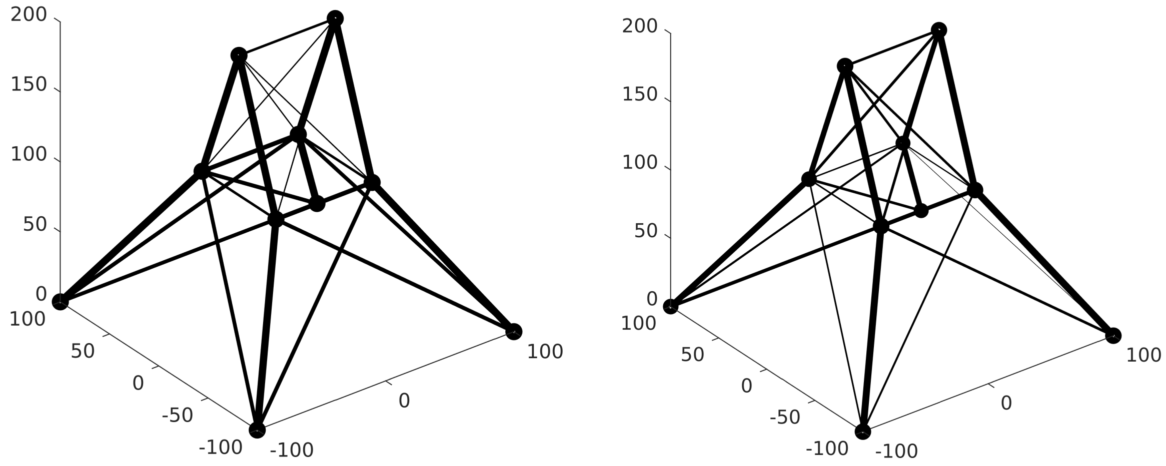
Table 1. For the truss example in Section 4, the positions of the nodes with Gaussian loads, their mean, and standard deviations.

Position	Mean (kN)	Standard deviation (kN)
$[-37.5, 0, 200]$	$[1, -10, -10]$	$[1, 1, 1]$
$[37.5, 0, 200]$	$[0 - 10, -10]$	$[1, 1, 1]$
$[-37.5, 37.5, 100]$	$[0.5, 0, 0]$	$[1, 1, 1]$
$[37.5, 37.5, 100]$	$[0.6, 0, 0]$	$[1, 1, 1]$

where $f_t = 3 \times 10^4$ N/cm^2 and $f_c = -1.2 \times 10^4$ N/cm^2, noting that we use a convention of positive values denoting traction and negative values denoting compression. The Gaussian-distributed loads are applied to four nodes as presented in Table 1.

The cross section areas of the elements are bounded between 0.1 and 100 cm^2, the elasticity modulus of the material is $E = 10^7$ N / cm^2, and the four bottom nodes are considered fixed to the ground. The initial and optimized trusses are presented in Figure 1, where the thickness of the lines represent the cross-section areas of the elements. The step size for Adam used is 0.01, and the other parameters are $\beta_1 = 0.99$, $\beta_2 = 0.9999$, and $\epsilon = 10^{-8}$. We use a Monte Carlo approximation of size 1000 each time we tune the Chernoff bound parameters, which we do according to (29) with $\epsilon_t = 100$. We ran Adam for this problem for 10 iterations, using one gradient evaluation of F per iteration. Moreover, 3000 extra evaluations of F were needed to tune the Chernoff parameters.

In Table 2, we present the truss volume, probabilities of failure due to traction and compression, objective value function, and relative optimality for the starting point and the solution found using Adam. The probabilities of failure are approximated using Monte Carlo Sampling with sample size of 10^6. The reference value was found numerically by running Adam with an independent sample for 10^6 iterations. The convergence of the distance to the optimum of the sequence generated by Adam is presented in Figure 2. It is clear that Adam succeeds in converging to the optimum even with just one evaluation of $\nabla_{\boldsymbol{d}} F$ per iteration. The total number of 10000 $\nabla_{\boldsymbol{d}} F$ evaluations and 3000 extra F evaluations is comparatively small considering the probability of failure at the optimum. A sample average approximation approach would require a large Monte Carlo sample and would thus not be able to perform many iterations with this budget.



(a) Initial structure

(b) Optimized structure

Figure 1. Truss structure for Example in Section 4. The widths of the lines represent the areas of the cross-sections of the elements.

5 Conclusions

Solving Risk Optimization (RO) problems is often an expensive task. In this work, we employ the Chernoff bound of the probabilities of failure to recast the original RO problem as an expectation minimization. An approx-

Table 2. Results for the example in Section 4.

	Volume (cm^3)	$P_{f,t}$	$P_{f,c}$	$f(\boldsymbol{d})$	$\left\lvert \frac{f(\mathbf{d}^*)-f(\mathbf{d})}{f(\mathbf{d}^*)} \right\rvert$
Starting point	4981.5	6.60×10^{-5}	3.55×10^{-1}	40498.37	1034.72%
Adam solution	3814.4	0	2.29×10^{-4}	3837.31	7.52%
Reference solution	3833.1	0	1.02×10^{-3}	3569.03	0%

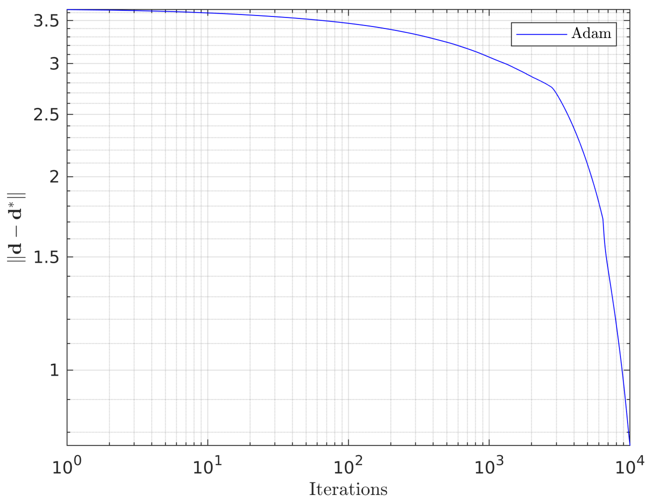


Figure 2. Convergence of the distance to the optimum for the truss example in Section 4.

imation of the original RO problem can be obtained using Monte Carlo sampling to get two parameters per failure mode, the Chernoff parameter that gets the tightest upper bound and a second parameter that corrects the bound slackness. Then, we use Adam, a robust Stochastic Gradient Descent method, to minimize the approximated RO problem. Moreover, we provide an heuristic approach to tune the Chernoff parameters of each failure mode during optimization in order not to render the optimization too expensive. The methodology is validated by optimizing the design of a truss structure subject to random loads. The low cost of the procedure and the good convergence achieved show that this approach is suited for RO, especially in the case when the dimension of the design and random parameter spaces do not allow the use of surrogate models.

Acknowledgements. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ - grant number 307133/2020-6.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] R. H. Lopez and A. T. Beck. Rbdo methods based on form: a review. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 34, n. 4, pp. 506–514, 2012.
[2] A. T. Beck and W. J. Gomes. A comparison of deterministic, reliability-based and risk-based structural optimization under uncertainty. *Probabilist Eng Mech*, vol. 28, pp. 18 – 29, 2012.
[3] A. T. Beck, R. K. Tessari, and H. M. Kroetz. System reliability-based design optimization and risk-based optimization: a benchmark example considering progressive collapse. *Engineering Optimization*, vol. 51, n. 6, pp. 1000–1012, 2019a.
[4] H. Kroetz, M. Moustapha, A. Beck, and B. Sudret. A two-level kriging-based approach with active learning for solving time-variant risk optimization problems. *Reliability Engineering & System Safety*, vol. 203, pp. 107033,

2020.
[5] A. T. Beck, R. K. Tessari, and H. M. Kroetz. System reliability-based design optimization and risk-based optimization: a benchmark example considering progressive collapse. *Engineering Optimization*, vol. 51, n. 6, pp. 1000–1012, 2019b.
[6] A. T. Beck and de W. J. Santana Gomes. A comparison of deterministic, reliability-based and risk-based structural optimization under uncertainty. *Probabilistic Engineering Mechanics*, vol. 28, pp. 18–29. Computational Stochastic Mechanics — CSM6, 2012.
[7] S. De, K. Maute, and A. Doostan. Reliability-based topology optimization using stochastic gradients. *Structural and Multidisciplinary Optimization*, vol. 64, n. 5, pp. 3089–3108, 2021.
[8] S. De, J. Hampton, K. Maute, and A. Doostan. Topology optimization under uncertainty using a stochastic gradient-based approach. *Structural and Multidisciplinary Optimization*, vol. 62, n. 5, pp. 2255–2278, 2020.
[9] A. G. Carlon, R. H. Lopez, L. F. d. R. Espath, L. F. F. Miguel, and A. T. Beck. A stochastic gradient approach for the reliability maximization of passively controlled structures. *Engineering Structures*, vol. 186, pp. 1–12, 2019.
[10] X. Huan and Y. Marzouk. Gradient-based stochastic optimization methods in bayesian experimental design. *International Journal for Uncertainty Quantification*, vol. 4, n. 6, 2014.
[11] A. G. Carlon, B. M. Dia, L. Espath, R. H. Lopez, and R. Tempone. Nesterov-aided stochastic gradient methods using laplace approximation for bayesian design optimization. *Computer Methods in Applied Mechanics and Engineering*, vol. 363, pp. 112909, 2020.
[12] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, vol. , 2014.
[13] A. G. Carlon, H. M. Kroetz, A. J. Torii, R. H. Lopez, and L. F. F. Miguel. Risk optimization using the chernoff bound and stochastic gradient descent. *Reliability Engineering & System Safety*, vol. 223, pp. 108512, 2022.
[14] R. E. Melchers and A. T. Beck. *Structural reliability analysis and prediction*. John Wiley & Sons, 2018.
[15] E. Lehman, T. Leighton, and A. R. Meyer. Mathematics for computer science. Technical report, Technical report, 2006. Lecture notes, 2010.