
Reliability Optimization Using Progressive Batching L-BFGS

Mohammad Etesam and Gholam Reza Mohtashami Borzadaran*

*Department of Statistics, Ferdowsi University of Mashhad, P. O. Box 1159,
Mashhad 91775, Iran*

E-mail: m.eatesam1@gmail.com; grmohtashami@um.ac.ir

**Corresponding Author*

Received 10 May 2024; Accepted 20 October 2024

Abstract

Reliability optimization can be applied to find parameters that increase reliability and decrease costs, in the presence of uncertainty. Nowadays, with the increasing complexity of systems, it is important to find suitable optimization methods. In this regard, we can refer to gradient-based optimization methods. The power of stochastic gradient-based approaches in optimization under uncertainty resides in efficiency in using sampling information. These methods allow applying a small sample size in updating problem parameters. Using a small sample size also has its disadvantages, and it leads to oscillation around the minimum point when approaching the minimum. One of the ways to solve this problem is to use progressive batching. Here, to increase stability Progressive Batching L-BFGS (PB-LBFGS) and Progressive Batching L-BFGS with momentum (PB-mLBFGS) are used for reliability optimization, and with an example, the effectiveness of these approaches is compared with some other optimization methods.

Keywords: Optimization under uncertainty, progressive batching L-BFGS, reliability optimization; stochastic gradients.

Journal of Reliability and Statistical Studies, Vol. 17, Issue 2 (2024), 321–332.

doi: 10.13052/jrss0974-8024.1723

© 2024 River Publishers

1 Introduction

In some reliability optimization problems, the mean of the cost function under given restrictions is minimized. This type of problem is called Reliability-Based Optimization (RBO). To solve RBO problems efficiently, one way is using a stochastic approximation of gradients [1]. Complex RBO problems can be solved well with stochastic optimization, especially when the size of the system is large. In reliability analysis, attaining full information requires too many simulations. Reducing the number of needed simulations is one of the benefits of using stochastic gradients [1]. Different optimization methods are deployed for RBO problems.

In [2] the efficiency of Stochastic Gradient Descent (SGD) algorithms for structural reliability maximization of buildings is investigated with a focus on computational cost reduction; They used the design of Friction tuned mass damper (FTMD) as an example problem and the Accelerated Stochastic Gradient Descent (ASGD) algorithm for optimization.

In the redundancy allocation problem for systems consisting of multiple k-out-of-n [3], a methodology for maximization of system reliability was developed by transforming the problem and defining new decision variables to yield an equivalent zero-one integer programming problem.

In [4] the application of SGD, Adam, AdaMax, Nadam, AMSGrad, and AdamSE algorithms in solving the mean-variance portfolio optimization problem is studied and it is shown that the AdamSE algorithm is a better approach, especially for solving the mean-variance portfolio optimization problem. [5] solves the reliability optimization problem for the Circular k-out-of-n system by introducing the BIGA method, which combines the genetic algorithm with Birnbaum importance. For cost optimization and reliability parameter extraction of a complex engineering system named Heat Removal System of a nuclear power generation plant safety system, [12] used metaheuristic algorithms, like Cuckoo Search algorithm and Grey Wolf optimizer. [14] conducted a reliability analysis of an excavator from its field failure data and effectively improved its life cycle cost.

The general form of a reliability optimization problems is

$$\begin{aligned}
 & \min_{\theta} E[C(\theta, X)], \theta \in \Omega_{\theta} \\
 & \text{Subject to} \\
 & h_i(\theta) \leq 0, i = 1, \dots, n_C \\
 & p_j(\theta) \leq p_j^{tol}, j = 1, \dots, n_p,
 \end{aligned} \tag{1}$$

Where θ is the design variables that should be optimized over its range Ω_θ ; E is the expectation operator; C is a cost function; X is the vector of uncertain variables; h_i denotes the constraints on the problem; p_j is the probability of occurrence of the j th event and p_j^{tol} is its maximum tolerance. Monte Carlo sampling can be applied for assessing the reliability of a system [7]. In optimization under uncertainty gradients can be approximated with the finite difference method [8]; this method is used extensively for solving RBO problems.

In optimization under uncertainty with gradients, it is common to use coordinate-wise methods such as SGD. Recently, quasi-Newton methods that use vector operators have also been introduced for optimization under uncertainty [11, 15]. In this study, the goal is to compare the performance of Progressive Batching L-BFGS (PB-LBFGS) and Progressive Batching L-BFGS with momentum (PB-mLBFGS) with some coordinate-wise methods in terms of the number of required samplings.

Next, in Section 2, the proposed method for reliability optimization is introduced. In Section 3, using a simulated example, the performance of the proposed optimization method is compared with some other methods. Finally, the conclusion is made in part 4.

2 Proposed Method

Stochastic Gradient Descent is an algorithm for optimization. Compared to Batch methods, there are some motivations for using stochastic methods. It is practical (when the dataset is very big) and theoretical (SG employs information more efficiently) motivations for following a stochastic approach [9]. In reliability optimization, there is a loss function that shall be minimized or maximized over its parameters. for example , we can maximize the reliability of the system; or we can minimize some predefined loss (i.e., expectation of a fiscal loss of the system). Suppose that the loss function is in the form of:

$$\int_X C(\theta, X) f_X(\theta, x) dx, \tag{2}$$

where f is the density function of the failure time and θ is the parameters of the problem. Sometimes the integral has a closed-form solution and we can derive it in terms of θ :

$$G(\theta) = \int_X C(\theta, X) f_X(\theta, x) dx. \tag{3}$$

Then for finding the optimal solution of the problem we need to differentiate respect to d and then solve the equation

$$\nabla g(d) = 0, \quad (4)$$

And then derive the solution \hat{d} . Sometimes $G(d)$ can be obtained, but we cannot solve (4) for d . In this case, a numerical approach (i.e., gradient descent method) can be employed.

Sometimes in equation (2), solving the integral is a complicated task and we cannot obtain $G(d)$. In this case, first, the loss is turned into this form [2]:

$$\begin{aligned} \nabla_{\theta} \bar{C}(\theta) &:= \nabla_{\theta} E_X[C(\theta, X)] \\ &= \nabla_{\theta} \int_X C(\theta, x) f_X(\theta, x) dx \\ &= \int_X \nabla_{\theta} \beta(d, x) f_X(\theta, x) dx \\ &\quad + \int_X C(\theta, x) \nabla_{\theta} (\log f_X(\theta, x)) f_X(\theta, x) dx \\ &= E_X [\nabla_{\theta} C(\theta, X) + C(\theta, x) \nabla_{\theta} (\log f_X(\theta, x))]. \end{aligned} \quad (5)$$

For most problems computing the above expectation directly is not possible; So, the above integral can be approximated with the Monte Carlo sampling method. This procedure is known as Sample average approximation (SAA) [8]:

$$E[C(\theta, x)] \approx \frac{1}{n} \sum_{i=1}^n C(\theta, x_i). \quad (6)$$

In the above formula, normally Monte Carlo simulation with a sampling strategy is used [8]. An accurate approximation needs a very large sampling size. Here, stochastic optimization methods can be deployed that works with few sampling sizes [1]. Sometimes we cannot obtain $\nabla_{\theta} C$. In this case, a finite-difference approximation to the gradients can be used [10]. The j th element of the gradient can be obtained from:

$$[\nabla_{\theta} C(\theta, x_i)]_j \approx \frac{C(\theta + c^j \cdot h, x) - C(\theta, x)}{h}, \quad (7)$$

where c^j is a vector of 0s except its j th element that is 1; and h is a positive number with a little magnitude, and can be obtained from:

$$h = \max\{1, |\theta_j|\} \sqrt{\epsilon_M}. \quad (8)$$

Here, ϵ_M is the machine precision [10]. Finite-difference approximation to the gradients is a simple method and only needs Monte Carlo simulation; so, it can be deployed in complicated models. Also, from formula (2.4), a finite-difference approximation to the gradients can be computed with a parallel procedure and is applicable when the number of components is very large.

A quasi-Newton method based on momentum, called Momentum-based L-BFGS (mLBFGS) is proposed in [15]. Progressive Batching L-BFGS (PB-LBFGS) is a quasi-Newton algorithm for optimization that uses stochastic gradients [11]. The method uses a stochastic backtracking line search [11]. A challenge in the stochastic version of L-BFGS and generally stochastic optimization is sample size selection. A lower bound for sample size is introduced in [11].

The functionality of a reliability system can be depicted with a reliability network [13]. Each network has a start point and a terminal point. An example of a reliability network is depicted in Figure 1.

In each simulation, given the failure time of each component, the failure time of the system can be computed with a function $T(t_1, t_2, \dots, t_n)$ where t_i is the failure time of the i th component. The cost function C is a function of T and model parameters θ .

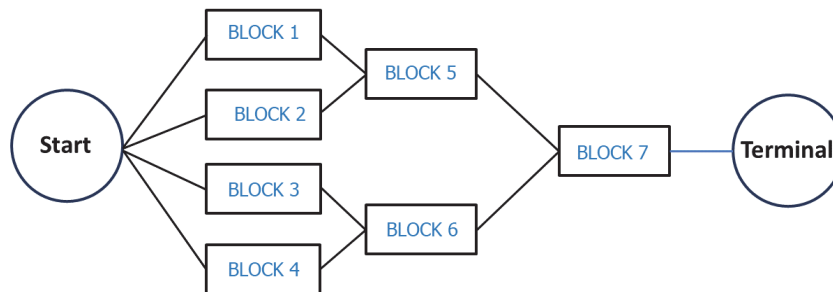


Figure 1 An example of reliability network of a system. The system consists of seven blocks. Each block is a series system with 50 components as shown in Figure 2.

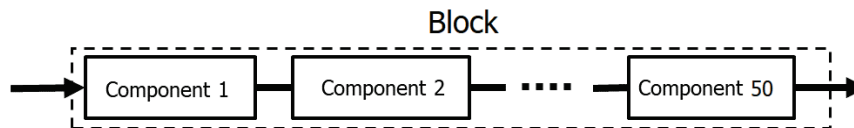


Figure 2 The architecture of the blocks used in Figure 1. Each block is a series system consisting of 50 components with independent exponential distributions.

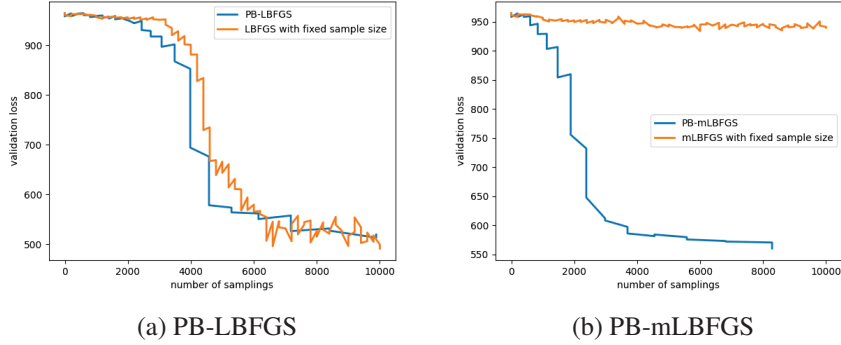


Figure 3 The effect of progressive batching on the RBO problem in Figure 1.

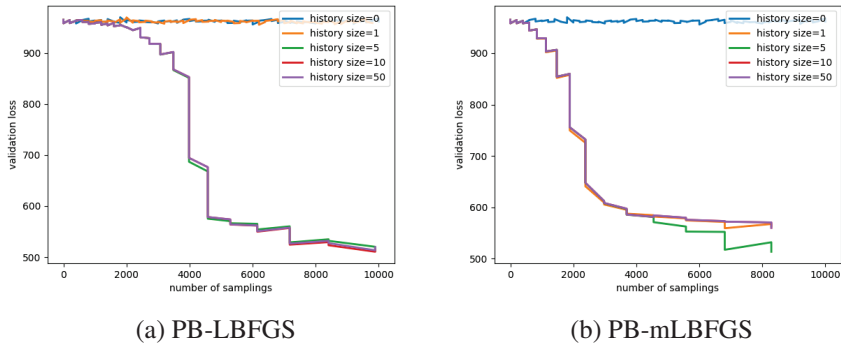


Figure 4 The effect of different history sizes on the PB-LBFGS and PB-mLBFGS.

3 Empirical Results

SGD uses coordinate-wise operations; But, L-BFGS uses vector operations; So, L-BFGS is more susceptible to machine precision and, when the magnitude of gradients is very small, the performance of L-BFGS may decay much more compared to SGD. In addition, when the relationship between problem parameters is not very complicated, L-BFGS is an unnecessarily complex option. In reliability network depicted in Figure 1, C_i consists of a series of 50 components. The network consists of 350 independent components. Time to failure of the i th component has an exponential distribution with parameter θ_i . The cost function for each simulation is $\sum_{i=1}^{350} \theta'_i / 350 + 1000 * \exp(-T(t_1, t_2, \dots, t_n))$ where $\theta' = \theta.W$. T is the failure time and W is a matrix with positive entries that indicates relationship between coordinates of θ and the loss. The initial batch size for optimization methods is 200, and validation is performed on 500,000 simulated data. For PB-mLBFGS

and PB-LBFGS Armijo line search is used, and in each iterations variance bound is checked, and when necessary sample size is multiplied by 1.2. The history size's initial value is 5, and Momentum value for PB-mLBFGS will be 0.2, unless otherwise stated. Optimization is performed with ADAM and SGD with momentum, in addition to PB-LBFGS and PB-mLBFGS.

The effect of using a progressive batching mechanism or not using it in algorithms PB-LBFGS and PB-mLBFGS, in terms of the number of samplings is depicted in Figure 3. With progressive batching PB-LBFGS reaches the solution with a smaller number of samplings, also its final solution is more stable. PB-mLBFGS can reach the solution by using progressive batching, but otherwise, it will not converge. So, progressive batching helps the optimizer to behave more stably near the minimum and to use information more efficiently.

The quasi-Newtonian algorithms obtain the direction of optimization with the help of values stored in their history. The Figure 4 shows the effect of history size on optimization. Algorithms PB-LBFGS and PB-mLBFGS cannot find a descent direction when history size is zero. If history size is one, algorithm PB-mLBFGS converges, which indicates its better performance than PB-LBFGS in stochastic conditions.

The PB-mLBFGS algorithm has a parameter called momentum. If it is one, the algorithm will convert to PB-LBFGS. Figure 5 shows the effect of different momentums in PB-mLBFGS. If momentum is one, the optimizer does not converge, because the progressive batching mechanism recognizes the generated direction as stable, and the batch size remains constant. As a result, the directions produced by PB-mLBFGS will not have the necessary efficiency. Also, these results show that using momentum can be useful in optimization under uncertainty with quasi-Newton optimizer.

Figure 6 shows the sample size of PB-LBFGS and PB-mLBFGS algorithms in each iteration, which can be used in manual batch size increments, to consume less computations..

Figure 7 shows the performance of the algorithms PB-LBFGS and PB-mLBFGS, compared to the coordinate-wise methods of SGD and ADAM. As can be seen, the algorithms PB-LBFGS and PB-mLBFGS reached a solution, with a much smaller number of samples. In addition, PB-mLBFGS can reach the solution with fewer samples than PB-LBFGS. It is worth mentioning, in terms of computation time, PB-LBFGS and PB-mLBFGS are slower because the progressive batching mechanism uses relevantly time-consuming operations; But PB-LBFGS and PB-mLBFGS automatically make the optimizer more stable, and also guide us on how to increment sample size manually.

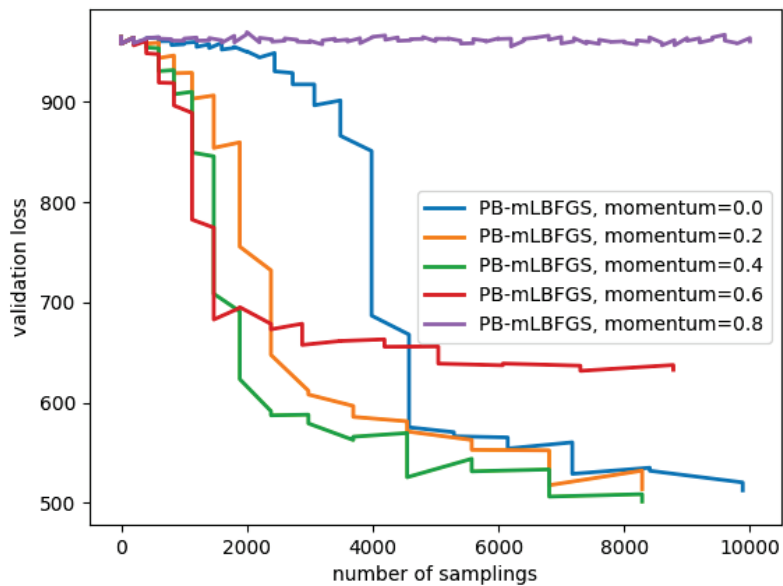


Figure 5 The effect of different momentum values on the PB-mLBFGS.

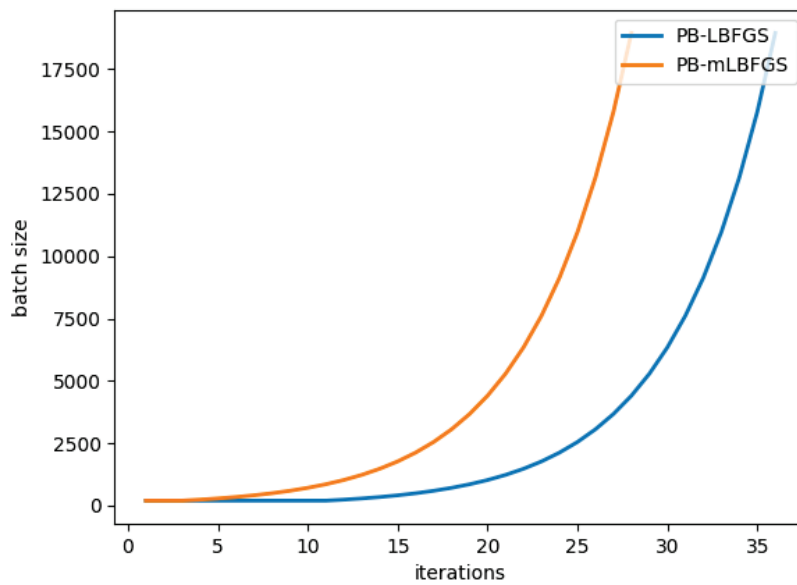


Figure 6 Sample size in terms of number of iterations.

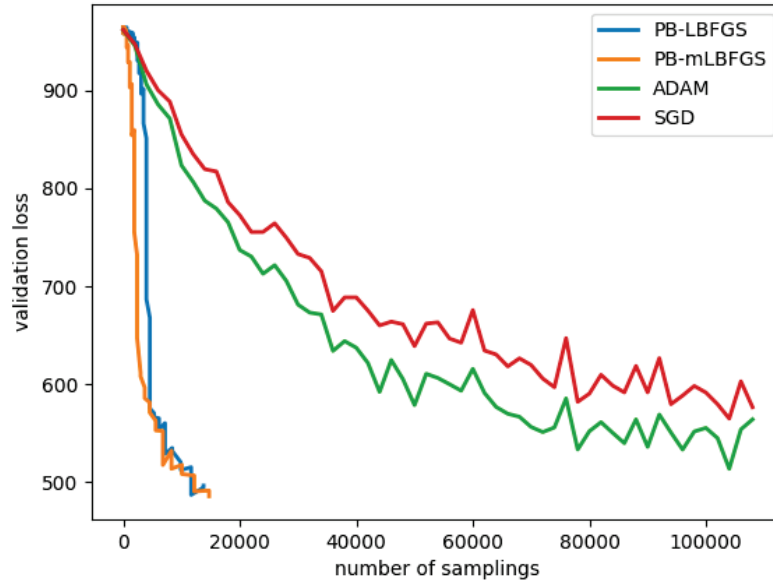


Figure 7 validation loss in terms of number of samplings for the problem depicted in Figure 1.

4 Conclusion

In this work first, the reliability optimization problem is introduced. Then, the procedures of cost approximation and gradient evaluation are explained. The performance of the PB-LBFGS and PB-mLBFGS methods in optimizing a reliability network consisting of 350 components was compared with some other optimization methods like ADAM and SGD. The effect of progressive batching, different history sizes, and different momentum values on the PB-LBFGS and PB-mLBFGS algorithms, in a simulated RBO problem was investigated. PB-LBFGS and PB-mLBFGS use an expression for computing an efficient number of samplings in each iteration. PB-mLBFGS without progressive batching may work worse than PB-LBFGS, and work better than it with using progressive batching. The results in an example showed that these methods applies to RBO problems compared to other optimizers like ADAM and SGD.

For future work, considering that quasi-Newton methods are evolving, the performance of newer stochastic quasi-Newton methods should be investigated; Also, the effect of different line search strategies in such problems.

References

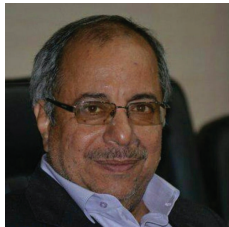
- [1] S. De, K. Maute, and A. Doostan. Reliability-based Topology Optimization using Stochastic Gradients. *Structural and Multidisciplinary Optimization*, 64(5): 3089–3108, 2021.
- [2] A. G. Carlon, R. H. Lopez, L. F. D. R. Espath, L. F. F. Miguel, and A. T. Beck. A stochastic gradient approach for the reliability maximization of passively controlled structures. *Engineering Structures*, 186: 1–12, 2019.
- [3] B. Chandrasekaran and W. Swartout. System reliability optimization with k-out-of-n subsystems. *Quality and Safety Engineering*, 7: 129–142, 2000.
- [4] S. S. Su and S. L. Kek. An Improvement of Stochastic Gradient Descent Approach for Mean-Variance Portfolio Optimization Problem. *Journal of Mathematics*, 2021.
- [5] S. Zhang, J. B. Zhao, H. G. Li, and N. Wang. Reliability optimization and importance analysis of circular-consecutive k-out-of-n system. *Mathematical Problems in Engineering*, (1), 1831537, 2017.
- [6] M.T. Todinov. A survey on approaches for reliability-based optimization. *Structural and Multidisciplinary Optimization*, 42:645–663, 2010.
- [7] W. Broding, F. Diederich, and P. Parker. Structural optimization and design based on a reliability design criterion. *Journal of Spacecraft and Rockets*, 1(1):56–61, 1964.
- [8] L. A. Hannah. Stochastic Optimization. Available online: <http://www.st.columbia.edu/~liam/teaching/compstat-spr14/lauren-notes.pdf>.
- [9] L. Bottou, F. E. Cutis, and J. Nocedal. Optimization methods for large-scale machine learning. *Siam Review*, 60(2):223–311, 2018.
- [10] H. J. M. Shi, Y. Xie, M. Q. Xuan, and J. Nocedal. Adaptive finite-difference interval estimation for noisy derivative-free optimization. *SIAM Journal on Scientific Computing*, 44(4): A2302–A2321, 2022.
- [11] R. Bollapragada, D. Mudigere, J. Nocedal, H. J. M. Shi, and P. T. P. Tang. A progressive batching L-BFGS method for machine learning. *International Conference on Machine Learning*, 620–629, 2018.
- [12] A. Kumar, S. Pant, and M. Ramt. Cost Optimization and Reliability Parameter Extraction of a Complex Engineering System. *Journal of Reliability and Statistical Studies*, 27:99–116, 2023.
- [13] M.T. Todinov. *Reliability and risk models: setting reliability requirements*. Chichester: John Wiley & Sons, 2007.

- [14] S. K. Mishra, N. K. Goyal, and A. Mukherjee. Reliability Analysis and Life Cycle Cost Optimization of Hydraulic Excavator. *Journal of Reliability and Statistical Studies*, 297–328, 2023.
- [15] Y. Niu, Z. Fabian, S. Lee, M. Soltanolkotabi, and S. Avestimehr. MI-bfgs: A momentum-based l-bfgs for distributed large-scale neural network optimization. arXiv preprint arXiv:2307.13744, 2023.

Biographies



Mohammad Etesam received his B.S. degree in computer software engineering and an MSc degree in statistics from Ferdowsi University of Mashhad in Iran. He is pursuing his PhD in artificial intelligence at Ferdowsi University of Mashhad. His research interests include reliability, and learning.



Gholam Reza Mohtashami Borzadaran received his B.S. degree in Statistics from the Faculty of Informatic and Statistics in Iran and MSc degree in Statistics from Shahid Beheshti University in Iran and PhD in Statistical Inference from Sheffield University, Sheffield, England (UK) in 1997. He is a professor in Statistics in the Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Iran. His research is focused on information theory and reliability.

