XL CILAMCE
IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING
NOVEMBER 11-14, 2019
Praiamar Natal Hotel & Convention
Natal, RN-BRAZIL

MAXIMIZATION OF THE RELIABILITY OF FRICTION TUNED MASS DAMPERS USING STOCHASTIC GRADIENT METHODS

André Gustavo Carlon
Rafael Holdorf Lopez
agcarlon@gmail.com
rafael.holdorf@ufsc.br
Federal University of Santa Catarina (UFSC), Department of Civil Engineering
Rua João Pio Duarte da Silva, Florianópolis, 88040-970, SC, Brazil
Luis Espath
espath@gmail.com
King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Science and Engineering Division (CEMSE)
Thuwal, 23955-6900, Jeddah, Saudi Arabia
Leandro Fleck Fadel Miguel
leandro.miguel@ufsc.br
Federal University of Santa Catarina (UFSC), Department of Civil Engineering
Rua João Pio Duarte da Silva, Florianópolis, 88040-970, SC, Brazil
André Teófilo Beck
atbeck@sc.usp.br
University of São Paulo (USP), São Carlos School of Engineering, Structural Engineering Department
São Carlos, 13566-590, SP, Brazil

Abstract. We use an efficient stochastic optimization framework to optimize the design of friction tuned mass dampers (FTMD). To deal with the uncertainties in the model, we use a reliability analysis based on the out-crossing rate approach. We formulate the unbiased gradient estimator of the reliability index with respect to the design parameters, a condition for the use of stochastic gradient descent (SGD) and its variations. We couple SGD with Nesterov's acceleration, Polyak–Ruppert averaging, and a restart technique to evaluate their improvement on SGD efficiency for design optimization. To assess the performance of the proposed stochastic optimization framework, we optimize the parameters of an FTMD in a steel frame building. Comparing the obtained results with the state-of-the-art in the literature, we observe a reduction of up to an order of magnitude in cost to achieve a given accuracy. Given that an unbiased estimator for the true gradient can be evaluated, stochastic gradient methods can efficiently perform local search in design problems on the presence of uncertainties.

Keywords: Tuned mass dampers, Stochastic optimization, Stochastic gradient descent, Reliability

1 Introduction

In many cases it is important to consider the reliability of an engineering structure under seismic loadings, since the effects of seismic excitations can lead to failure of the structure. To reduce the probability of failure, one solution is to use passive dissipation systems [1]. In this paper, we focus on passive control of buildings using friction tuned mass dampers (FTMDs), where friction is used to reduce vibration amplitudes [2–7]. However, FTMDs have some parameters to be tuned, namely, their friction force magnitude, stiffness, and mass. Moreover, the optimal tuning of these parameters is not trivial and depends on the structure and loadings. Since we want to minimize the probability of failure, we have a coupled reliability and optimization problem. These problems are generally expensive to solve due to the high number of reliability analysis required; at least one per optimization iteration.

To efficiently optimize FTMDs, we use stochastic gradient descent (SGD) [8] and some of its variations. These methods are used in stochastic optimization to minimize (or maximize) the expected value of a function, given that an unbiased gradient estimator is available. The advantage of SGD is that it simultaneously solves the uncertainty quantification and the optimization problems, resulting in a cheaper procedure in comparison to combining deterministic local search methods with Monte Carlo integration. Moreover, SGD does not introduce any bias in the optimum estimate, progressively converging to the true optimum as optimization progresses. To further improve SGD, we couple it with state-of-the-art techniques in stochastic optimization, namely, Polyak–Ruppert averaging [9, 10], Nesterov's acceleration [11], and an acceleration restart technique [12], furnishing the accelerated stochastic gradient descent (ASGD). The ASGD method has been employed in stochastic optimization with success by the authors [13], thus, being a promising approach to optimize FTMDs.

To assess the proposed methodology, we optimize an FTMD at the top of a steel frame building subject to a seismic excitation generated using the Kanai-Tajimi filtering [14, 15]. We model both stiffness and mass of each element as random variables, besides building, FTMD, and Kanai-Tajimi filtering parameters, totaling 148 random variables. To compare the performance of the stochastic gradient methods in optimzing the FTMD parameters, we use the Nelder-Mead (NM) [16] and the Sequential Quadratic Programming (SQP) [17] algorithms. Both NM and SQP are used along with sample size averaging [18] to reduce the variance, thus, reducing the optimization problem to a deterministic (and biased) one, at the cost of evaluating at least one Monte Carlo Integration each optimization iteration.

2 Optimal design of passive control devices

To improve the performance of the FTMDs, we seek the design parameters vector that maximize the reliability of the building when subject to transient loadings. We model the optimization problem as

$$\begin{aligned} \text{Find } \mathbf{d}^* = \underset{\mathbf{d} \in \mathbb{R}^{n_d}}{\arg\max}\ \bar{\beta}(\mathbf{d}) \\ \text{Subject to } \mathbf{d}_{\text{min}} \leq \mathbf{d} \leq \mathbf{d}_{\text{max}}, \end{aligned} \tag{1}$$

where $\mathbf{d}$ is the design parameters vector, $\mathbf{d}_{\text{min}}$ and $\mathbf{d}_{\text{max}}$ are the bound constraints, $\bar{\beta}$ is the reliability index and n_d is the number of design variables. In the case of FTMDs, we model the friction force magnitude (f_F) and stiffness (k_F) of the FTMDs as design variables.

The evaluation of $\bar{\beta}$ for the optimal design of FTMD is a time dependent reliability problem [19, 20]. To solve this problem, we utilize the up-crossing rate approach, which consists in evaluating the probability of a quantity of interest exceeding a barrier value b. In this work, we use statistics about the absolute displacement (z) and velocity ($\dot{z}$) of the top floor of the building to define the up-crossing rate. Gathering the uncertainties in a vector $\mathbf{x} \in \mathbf{X}$, we define the up-crossing rate for a barrier b as

$$v_z^+(\mathbf{d}, \mathbf{x}) = \frac{\sigma_{\dot{z}}(\mathbf{d}, \mathbf{x})}{\sigma_z(\mathbf{d}, \mathbf{x})} \frac{1}{2\pi} \exp\left(-\frac{b^2}{2\left(\sigma_z(\mathbf{d}, \mathbf{x})\right)^2} \right), \tag{2}$$

where σ_z and $\sigma_{\dot{z}}$ are, respectively, the standard deviation of displacement and velocity responses [21]. Defining the probability of the up-crossing rate being violated as P_f, we define β as

$$\beta(\mathbf{d}, \mathbf{x}) \stackrel{\text{def}}{=} -\Phi^{-1}(P_f(\mathbf{d}, \mathbf{x})), \tag{3}$$

where Φ is the cumulative probability density of the standard Gaussian. The reliability index is calculated as

$$\bar{\beta}(\mathbf{d}) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{x}}[\beta(\mathbf{d}, \mathbf{x})] \tag{4}$$

$$= \int_{\mathbf{X}} \beta(\mathbf{d}, \mathbf{x}) f_{\mathbf{x}}(\mathbf{d}, \mathbf{x}) \mathrm{d}\mathbf{x}, \tag{5}$$

where $f_{\mathbf{x}}$ is the pdf of $\mathbf{x}$. For a further discussion on the evaluation of the probability of failure and reliability index in the up-crossing rate approach, as well as a discussion on the statistical linearization used to deal with FTMDs nonlinearities, the authors refer to their previous work [21].

Given that the reliability index requires the evaluation of an expected value, the optimization problem in (1) is one of maximization of an expected value, thus, an stochastic optimization problem.

3 Stochastic optimization

Rewriting (1) as a stochastic optimization problem furnishes

$$\begin{aligned} \text{Find } \mathbf{d}^* &= \underset{\mathbf{d} \in \mathbb{R}^{n_d}}{\arg\max}\, \mathbb{E}_{\mathbf{x}}[\beta(\mathbf{d}, \mathbf{x})] \\ \text{Subject to } \mathbf{d}_{\text{min}} &\leq \mathbf{d} \leq \mathbf{d}_{\text{max}}. \end{aligned} \tag{6}$$

One way of solving (6) is to use the steepest descent method, also known in the stochastic case as full-gradient descent (FGD). Algorithm 1 presents the iterative procedure for solving the problem in (6) using FGD. The FGD evaluates the true gradient $\nabla_{\mathbf{d}} \mathbb{E}_{\mathbf{x}}[\beta]$ every iteration and performs a step with size α in

Algorithm 1 Pseudocode for FGD.

1: **procedure** FGD($\mathbf{d}_0, \alpha$)
2: **for** $k = 1, 2, ...$ **do**
3: $\mathbf{d}_k = \mathbf{d}_{k-1} + \alpha \nabla_{\mathbf{d}} \mathbb{E}_{\mathbf{x}} \left[\beta(\mathbf{d}_{k-1}, \mathbf{x})\right]$
4: **end for**
5: $\hat{\mathbf{d}} = \mathbf{d}_k$
6: **end procedure**

its direction. When the iterative process finishes, the last $\mathbf{d}$ is taken as $\hat{\mathbf{d}}$, the estimate of the optimum $\mathbf{d}^*$. The FGD method has a guaranteed convergence for convex problems, however, the evaluation of the true gradient every iteration can be expensive. If the expectation of the gradient needs to be approximated by a Monte Carlo integration (MCI), a MCI must be evaluated every iteration.

3.1 Stochastic gradient descent

One alternative to using FGD is to use SGD [8]. The SGD method consists in using an imprecise estimate of the gradient as a descent (or ascent) direction. Thus, the expensive estimation of the true gradient by MCI is substituted by a cheaper estimate, reducing the total cost of optimization. The SGD algorithm converges almost surely to the optimum if the step-size decreases every iteration as $\alpha_k = \alpha_0/k$.

To use SGD, an unbiased estimator of the true gradient $\nabla_{\mathbf{d}}\mathbb{E}[\beta(\mathbf{d},\mathbf{x})]$ is needed. Developing the true gradient for our problem in (6) results in

$$\nabla_{\mathbf{d}}\mathbb{E}[\beta(\mathbf{d},\mathbf{x})] = \nabla_{\mathbf{d}}\int_{\mathbf{X}}\beta(\mathbf{d},\mathbf{x})f_{\mathbf{x}}(\mathbf{d},\mathbf{x})\mathrm{d}\mathbf{x} \tag{7}$$

$$= \int_{\mathbf{X}}\nabla_{\mathbf{d}}\beta(\mathbf{d},\mathbf{x})f_{\mathbf{x}}(\mathbf{d},\mathbf{x})\mathrm{d}\mathbf{x} + \int_{\mathbf{X}}\beta(\mathbf{d},\mathbf{x})\nabla_{\mathbf{d}}f_{\mathbf{x}}(\mathbf{d},\mathbf{x})\mathrm{d}\mathbf{x} \tag{8}$$

$$= \int_{\mathbf{X}}\nabla_{\mathbf{d}}\beta(\mathbf{d},\mathbf{x})f_{\mathbf{x}}(\mathbf{d},\mathbf{x})\mathrm{d}\mathbf{x} + \int_{\mathbf{X}}\beta(\mathbf{d},\mathbf{x})\nabla_{\mathbf{d}}\log(f_{\mathbf{x}}(\mathbf{d},\mathbf{x}))f_{\mathbf{x}}(\mathbf{d},\mathbf{x})\mathrm{d}\mathbf{x} \tag{9}$$

$$= \mathbb{E}[\nabla_{\mathbf{d}}\beta(\mathbf{d},\mathbf{x}) + \beta(\mathbf{d},\mathbf{x})\nabla_{\mathbf{d}}\log(f_{\mathbf{x}}(\mathbf{d},\mathbf{x}))]. \tag{10}$$

We define the gradient estimator for stochastic gradient methods as $\mathcal{G}$, with

$$\mathcal{G}(\mathbf{d},\mathbf{x}) \stackrel{\text{def}}{=} \nabla_{\mathbf{d}}\beta(\mathbf{d},\mathbf{x}) + \beta(\mathbf{d},\mathbf{x})\nabla_{\mathbf{d}}\log(f_{\mathbf{x}}(\mathbf{d},\mathbf{x})), \tag{11}$$

for our case. Since $\mathbb{E}[\mathcal{G}] = \nabla_{\mathbf{d}}\mathbb{E}[\beta]$, $\mathcal{G}$ is an unbiased estimator of the true gradient and, thus, can be used in stochastic gradient methods. The second term in the right hand side of 11 results from the dependency of $f_{\mathbf{x}}$ on $\mathbf{d}$ (cf. Appendix B in [21]).

To evaluate $\mathcal{G}$ in stochastic gradient methods, one or more realizations of $\mathbf{x}$ are independently sampled each iteration. The pseudocode for the solution of (6) using SGD is presented in Algorithm 2. In this paper, we refer to simple SGD as *vanilla* SGD, since other methods will later be built upon this idea.

Algorithm 2 Pseudocode for *vanilla* SGD.

procedure SGD($\mathbf{d}_0, \alpha_0$)
  for $k = 1, 2, ...$ **do**
    Sample random variables $\mathbf{x}_k$
    $\alpha_k = \frac{\alpha_0}{k}$
    $\mathbf{d}_k = \mathbf{d}_{k-1} + \alpha_k\mathcal{G}(\mathbf{d}_{k-1}, \mathbf{x}_k)$
  end for
  $\hat{\mathbf{d}} = \mathbf{d}_k$
end procedure

The optimization path of FGD and SGD over the contour of a benchmark function is presented in Figure 1 to illustrate their behavior. It can be seen that SGD moves is direction of the optimum, even with noisy gradient estimates.

The *vanilla* SGD main disadvantages are that the step-size rapidly decreases and that SGD convergence is very sensible to the initial step-size. To improve on SGD, we combine it with state-of-the-art techniques in stochastic optimization, namely Polyak–Ruppert averaging [9, 10], Nesterov's acceleration [11], and a restart technique [12].

3.2 Polyak–Ruppert averaging

Polyak and Juditsky [10] and Ruppert [9] independently developed an averaging method that allows the step-size in SGD to decrease slowly. The idea behind Polyak–Ruppert averaging is to use the average of the optimization path as an optimum estimate. Let the average at the k-th iteration be

$$\bar{\mathbf{d}}_k \stackrel{\text{def}}{=} \frac{1}{k}\sum_{i=0}^{k}\mathbf{d}_i. \tag{12}$$

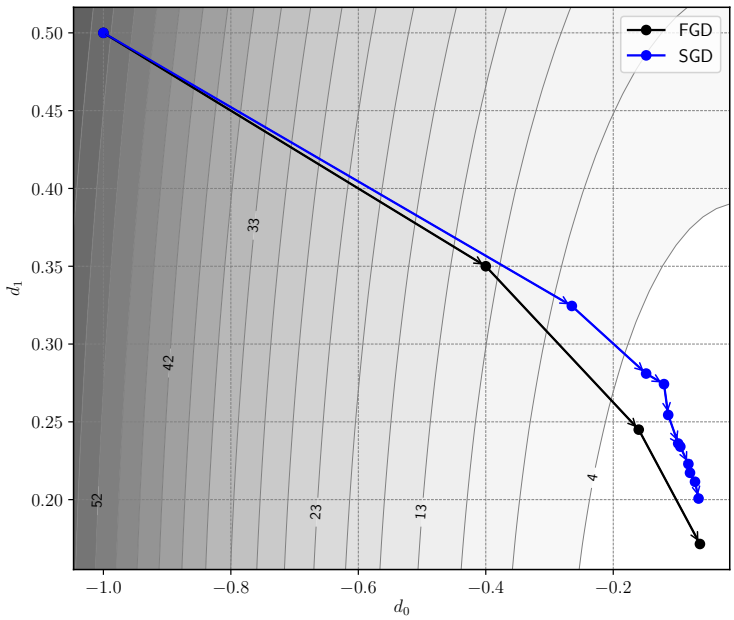

Figure 1. Path of full-gradient descent and stochastic gradient descent.

With step-size $\alpha_k = \frac{\alpha_0}{\sqrt{k}}$, $\mathbf{d}_k$ does not converge to the optimum, however, $\bar{\mathbf{d}}_k$ does [22]. Nemirovski uses a weighted averaging with the step-size as a weight,

$$\bar{\mathbf{d}}_k \overset{\text{def}}{=} \left(\sum_{\frac{k}{2} \leq i \leq k} \alpha_i \right)^{-1} \sum_{\frac{k}{2} \leq i \leq k} \alpha_i \mathbf{d}_i, \tag{13}$$

which is the averaging we opt to use in this paper.

We present the optimization path over the contour of a benchmark function for both $\mathbf{d}$ and $\bar{\mathbf{d}}$ in Figure 2. The distances to the optimum of both $\mathbf{d}$ and $\bar{\mathbf{d}}$ are presented in Figure 3. It can be observed in Figures 2 and 3 that the Polyak–Ruppert averaging consistently moves towards the optimum even though $\mathbf{d}$ does not.

We refer to SGD with Polyak–Ruppert simply as SGD, since *vanilla* SGD is not used in this paper due to its poor performance. We use a tolerance to the change in $\bar{\mathbf{d}}$ in consecutive iterations as a stop criterion.

The pseudocode of SGD with Polyak–Ruppert averaging is presented in Algorithm 3, with the lines related to the averaging colored in green.

Algorithm 3 Pseudocode for SGD with Polyak–Ruppert averaging.

procedure SGD($\mathbf{d}_0, \alpha_0$)
  $\bar{\mathbf{d}}_0 = \mathbf{d}_0$
  for $k = 1, 2, \ldots$ **do**
    Sample random variables $\mathbf{x}_k$
    $\alpha_k = \frac{\alpha_0}{\sqrt{k}}$
    $\mathbf{d}_k = \mathbf{d}_{k-1} + \alpha_k \mathcal{G}(\mathbf{d}_{k-1}, \mathbf{x}_k)$
    $\bar{\mathbf{d}}_k = \left(\sum_{\frac{k}{2} \leq i \leq k} \alpha_i \right)^{-1} \sum_{\frac{k}{2} \leq i \leq k} \alpha_i \mathbf{d}_i$ ▷ Polyak–Ruppert averaging
  end for
  $\hat{\mathbf{d}} = \bar{\mathbf{d}}_k$
end procedure

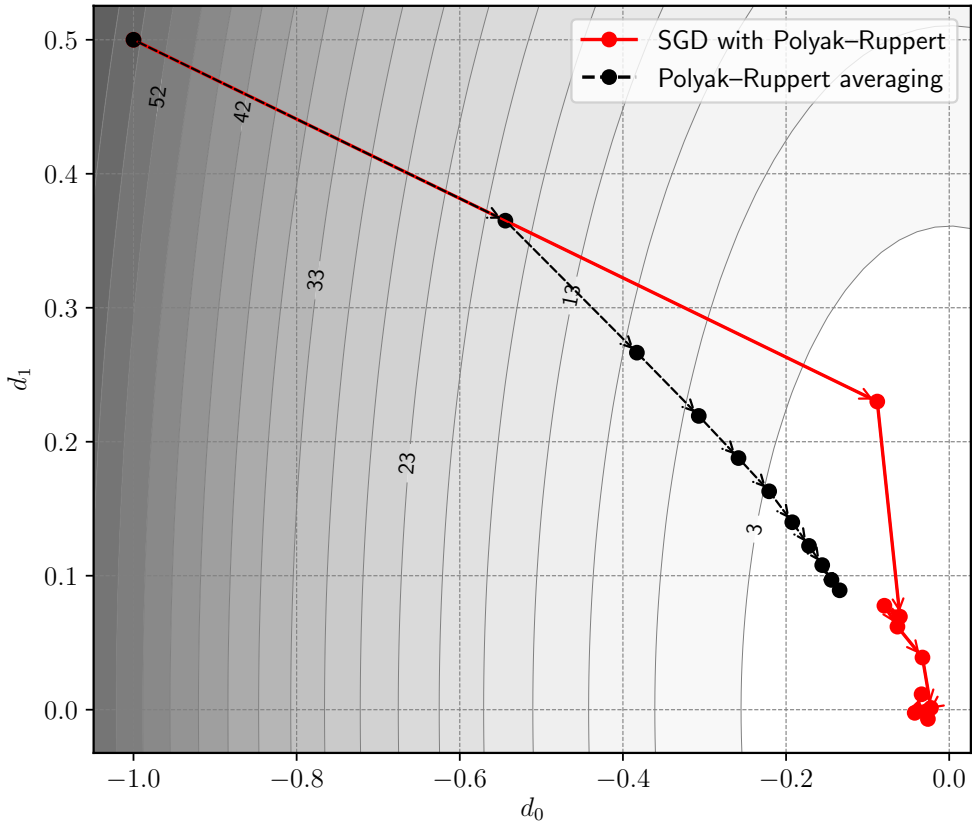

Figure 2. Example of SGD with Polyak–Ruppert averaging.

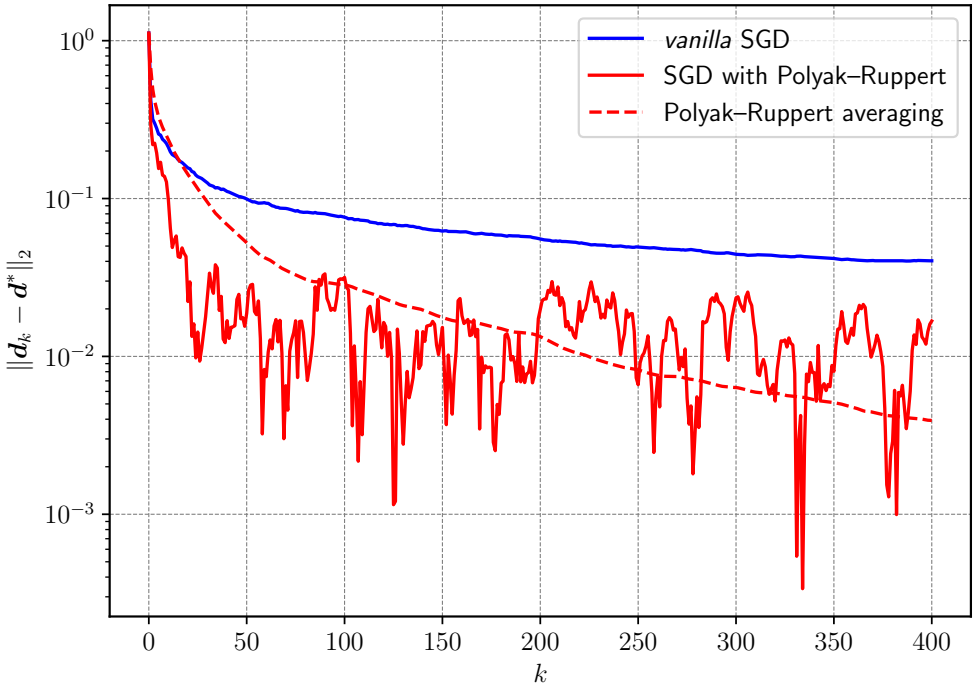

Figure 3. Convergence of vanilla SGD, SGD with Polyak–Ruppert and its average.

3.3 Nesterov's acceleration in SGD

To improve on SGD, we use Nesterov's acceleration [11]. Nesterov's acceleration is a momentum method that was developed for deterministic optimization. Momentum methods have been gaining attention from the stochastic optimization community due to their remarkable performance in machine learning training [23]. Nesterov's acceleration uses information from previous iterations to perform an extra step in comparison to steepest descent. We write the Nesterov's accelerated stochastic gradient descent (ASGD) method for solving (6) as

$$\begin{cases} \mathbf{z}_k = \mathbf{d}_{k-1} + \alpha_k \mathcal{G}(\mathbf{d}_{k-1}, \mathbf{x}) \\ \\ \mathbf{d}_k = \mathbf{z}_k + \gamma_k(\mathbf{z}_k - \mathbf{z}_{k-1}), \end{cases} \tag{14}$$

where the sequence $(\gamma_k)_{k\geq 0}$ is given by

$$\gamma_k = \frac{\lambda_{k-1}(1-\lambda_{k-1})}{\lambda_{k-1}^2 + \lambda_k}, \tag{15}$$

and the sequence $(\lambda_k)_{k\geq 0}$ solves

$$\lambda_k^2 = (1-\lambda_k)\lambda_{k-1}^2 + q\lambda_k, \quad \lambda_0 = 1. \tag{16}$$

The variable q is a positive real number between 0 and 1 that defines how much acceleration is imposed; 0 is maximum acceleration and 1 furnishes the steepest descent algorithm. According to Nesterov [24], if the objective function is μ strongly-convex and has L-Lipschitz gradient, the optimal value for q is $q^* = \mu/L$.

Figure 4 illustrates Nesterov's accelerated steps over the contour of a benchmark function, where the steepest descent step ($\mathbf{z}_k$) is in red and the accelerated step ($\gamma_k(\mathbf{z}_k - \mathbf{z}_{k-1})$) is in blue. Figure 5 shows

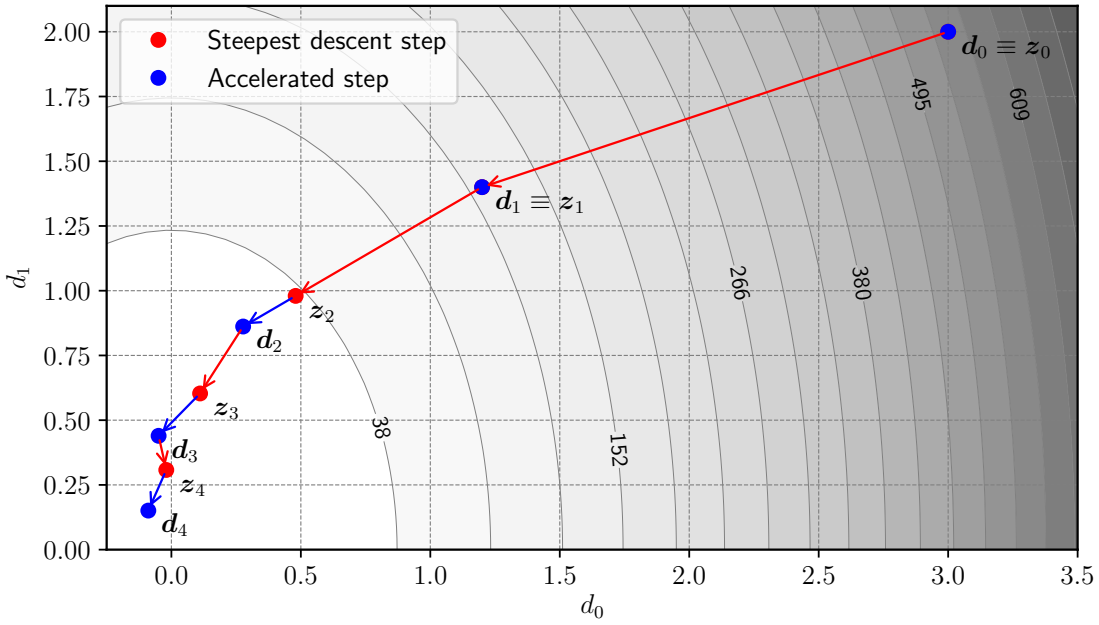

Figure 4. Example of Nesterov's acceleration.

how the Nesterov's accelerated step, presented in Eq. 14, is performed.

The pseudocode of ASGD with Polyak–Ruppert averaging is presented in Algorithm 4.

Nesterov's acceleration with q^* has optimal convergence rate for convex deterministic problems, however, estimating q^* can be expensive [12].

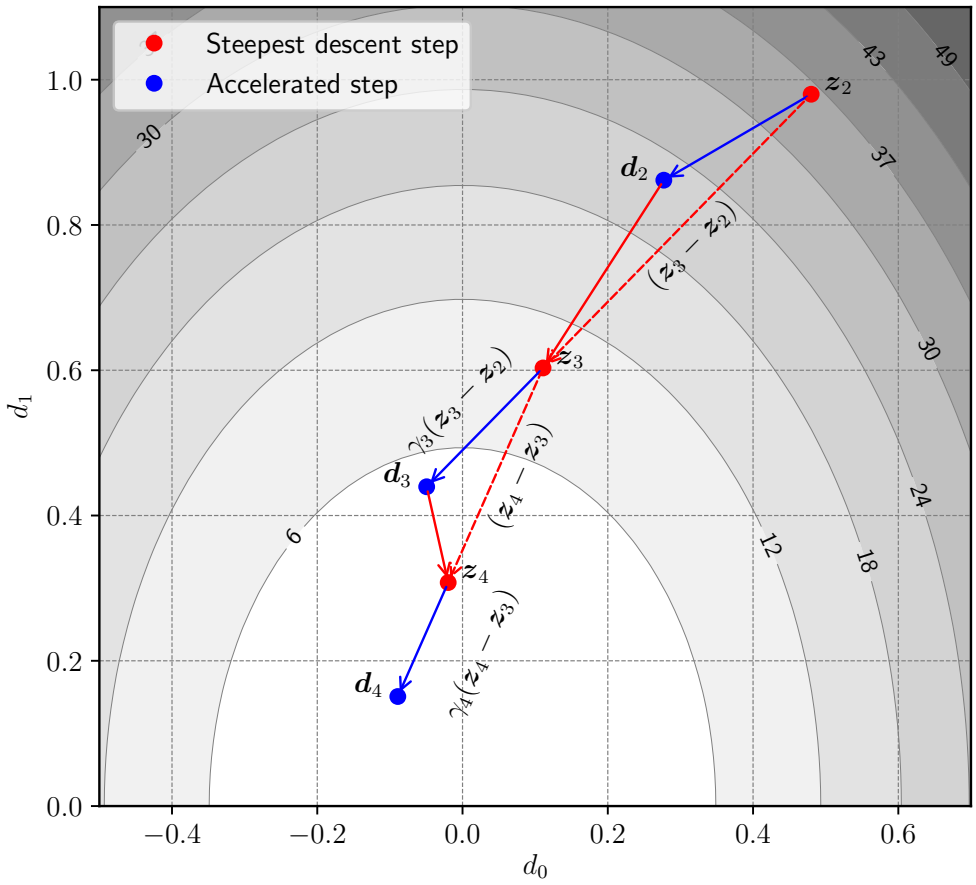

Figure 5. Representation of how accelerated steps are taken.

Acceleration restart

O'Donoghue and Candès [12] proposed a restart technique to avoid the need of evaluating q^*. If, for example, we are looking for the minimizer of a function f, the technique consists in restarting the acceleration whenever the optimization step $(\mathbf{d}_k - \mathbf{d}_{k-1})$ has negative inner product with respect to steepest descent direction. $(-\nabla_{\mathbf{d}} f)$, i.e.,

$$\langle -\nabla_{\mathbf{d}} f(\mathbf{d}_k), \mathbf{d}_k - \mathbf{d}_{k-1} \rangle < 0. \tag{17}$$

The restart technique is illustrated in Figure 6, where the candidate $\mathbf{d}_7$ is rejected for being an upward move, thus, acceleration is restarted and $\mathbf{d}_7$ is set as $\mathbf{z}_7$.

For the deterministic case, using the restart technique results in the same optimal linear convergence as using q^*, as can be observed in Figure 7. For the ASGD case, we substitute the true gradient by its estimate, thus, for the optimization problem in (6), the acceleration must be restarted whenever

$$\langle \mathcal{G}(\mathbf{d}_{k-1}, \mathbf{x}_k), \mathbf{d}_k - \mathbf{d}_{k-1} \rangle < 0. \tag{18}$$

We present the pseudocode of ASGD with restart in Algorithm 5.

In the next section, we refer to SGD with Polyak–Ruppert simply as SGD, and ASGD with both Polyak–Ruppert averaging and the restart technique as ASGD.

Algorithm 4 Pseudocode for *vanilla* ASGD.

procedure ASGD($\mathbf{d}_0, \alpha_0, q$)
  $\bar{\mathbf{d}}_0 = \mathbf{d}_0, \mathbf{z}_0 = \mathbf{d}_0, \lambda_0 = 1$
  for $k = 1, 2, \ldots$ **do**
    Sample random variables $\mathbf{x}_k$
    Solve λ_k for $\lambda_k^2 = (1 - \lambda_k)\lambda_{k-1}^2 + q\lambda_k$
    $\gamma_k = \frac{\lambda_{k-1}(1-\lambda_{k-1})}{\lambda_{k-1}^2+\lambda_k}$
    $\alpha_k = \frac{\alpha_0}{\sqrt{k}}$ ▷ Decreasing step scheme
    $\mathbf{z}_k = \mathbf{d}_{k-1} + \alpha_k \nabla_{\mathbf{d}} \beta(\mathbf{d}_{k-1}, \mathbf{x}_k)$
    $\mathbf{d}_k = \mathbf{z}_k + \gamma_k(\mathbf{z}_k - \mathbf{z}_{k-1})$ ▷ Nesterov's accelerated step
    $\bar{\mathbf{d}}_k = \left(\sum_{\frac{k}{2} \le i \le k} \alpha_i\right)^{-1} \sum_{\frac{k}{2} \le i \le k} \alpha_i \mathbf{d}_i$ ▷ Polyak–Ruppert averaging
  end for
  $\hat{\mathbf{d}} = \bar{\mathbf{d}}_k$
end procedure

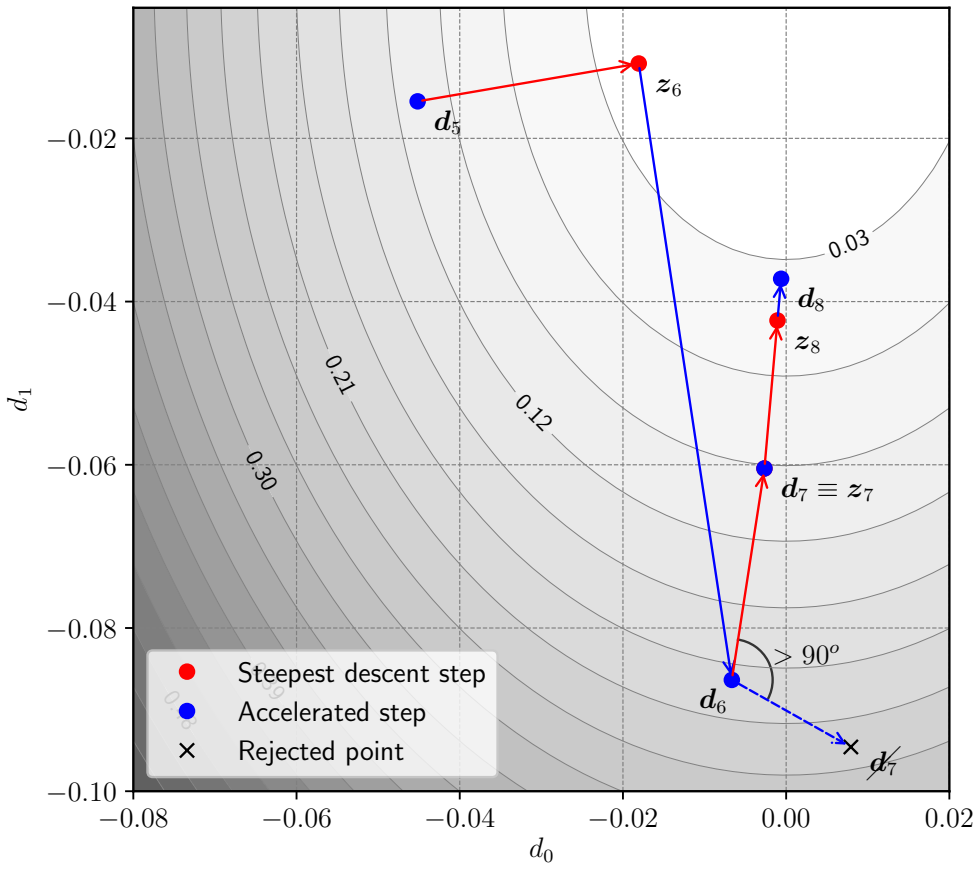

Figure 6. Illustration of restart technique for Nesterov's acceleration.

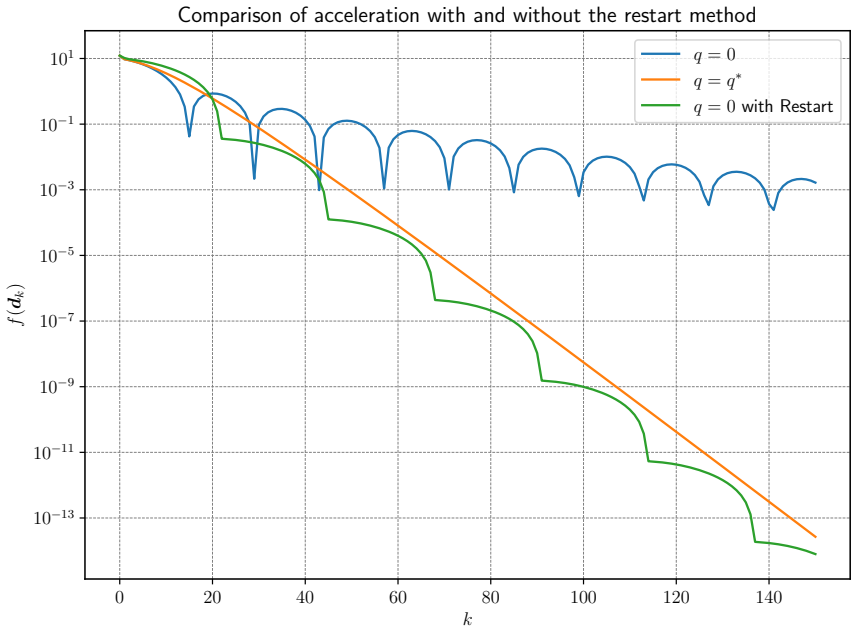

Figure 7. Convergence of the Nesterov acceleration with the restart technique in deterministic convex benchmark example.

Algorithm 5 Pseudocode for ASGD with restart.

procedure ASGD($\mathbf{d}_0, \alpha_0, q$)
  $\bar{\mathbf{d}}_0 = \mathbf{d}_0, \mathbf{z}_0 = \mathbf{d}_0, \lambda_0 = 1$
  for $k = 1, 2, \ldots$ **do**
    Sample random variables $\mathbf{x}_k$
    Solve λ_k for $\lambda_k^2 = (1 - \lambda_k)\lambda_{k-1}^2 + q\lambda_k$
    $\gamma_k = \frac{\lambda_{k-1}(1-\lambda_{k-1})}{\lambda_{k-1}^2+\lambda_k}$
    $\alpha_k = \frac{\alpha_0}{\sqrt{k}}$ ▷ Decreasing step scheme
    $\mathbf{z}_k = \mathbf{d}_{k-1} + \alpha_k \mathcal{G}(\mathbf{d}_{k-1}, \mathbf{x}_k)$
    $\mathbf{d}_k = \mathbf{z}_k + \gamma_k(\mathbf{z}_k - \mathbf{z}_{k-1})$ ▷ Nesterov's accelerated step
    if $\langle \nabla_{\mathbf{d}} \beta(\mathbf{d}_{k-1}, \mathbf{x}_k), \mathbf{d}_k - \mathbf{d}_{k-1} \rangle < 0$ **then** ▷ Restart technique
      $\lambda_{k+1} = 1$
    end if
    $\bar{\mathbf{d}}_k = \left(\sum_{\frac{k}{2} \le i \le k} \alpha_i\right)^{-1} \sum_{\frac{k}{2} \le i \le k} \alpha_i \mathbf{d}_i$ ▷ Polyak–Ruppert averaging
  end for
  $\hat{\mathbf{d}} = \bar{\mathbf{d}}_k$
end procedure

4 Numerical example

In this example, we optimize the friction force magnitude and stiffness of a single FTMD in the top of a steel frame building subject to seismic excitation in order to maximize its reliability index. The structure is modeled as a planar steel building frame, with three spans (23.77 m wide) and ten stories (37.42 m high), obtained from [25]. The linear elastic finite element model is discretized in 70 elements and 44 nodes, resulting in 132 degrees of freedom. Regarding the mass and damping, we adopted a consistent mass matrix and the classical Rayleigh proportional damping. The damping ratio is fixed as 5% for the first and second vibration modes. We obtained the geometric properties of the steel profiles from [26]. An additional mass of 44 t is considered per story, due to slabs and some eventual overload. The geometry of the structure of the building and the cross section of its elements are presented in Figure 8.

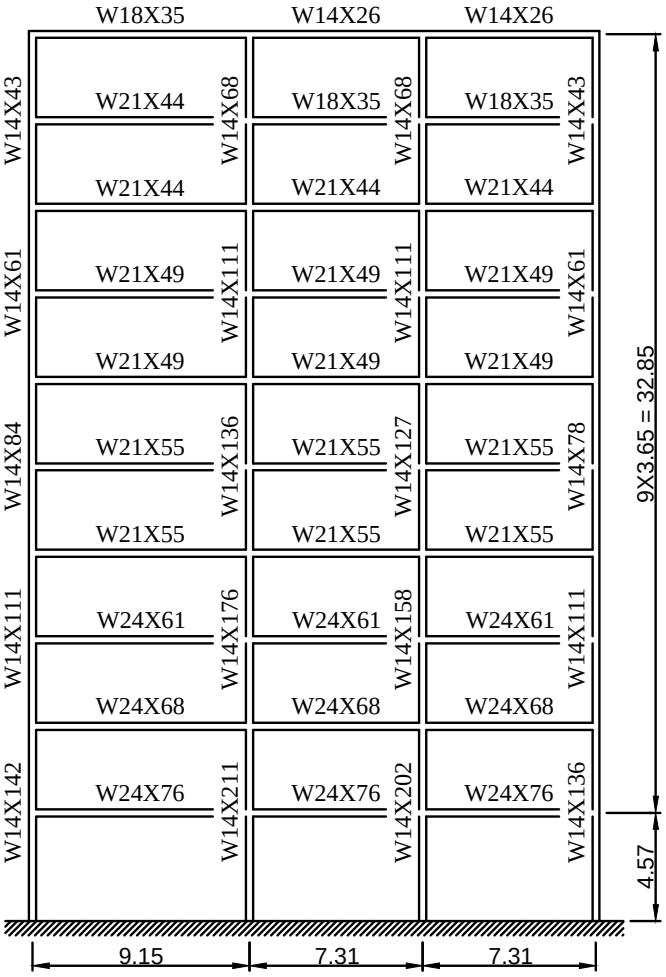

Figure 8. Planar steel frame (dimensions in m): Adapted from [25]

The seismic excitation is simulated by a zero mean Gaussian stationary stochastic process, $\ddot{\mathbf{z}}_{\mathbf{0}}(t)$, filtered through the Kanai-Tajimi [14, 15] model, with power spectral density

$$S(\omega) = S_0 \left[\frac{{\omega_g}^4 + 4{\omega_g}^2{\xi_g}^2\omega^2}{(\omega^2 - {\omega_g}^2)^2 + 4{\omega_g}^2{\xi_g}^2\omega^2} \right], \tag{19}$$

where S_0 is the constant spectral density, ω is the angular frequency and, ξ_g and ω_g are, respectively, the damping ratio and the natural frequency of the filter, which depend on the local soil.

We individually model the elasticity modulus and specific mass of each element as Gamma-distributed. Moreover, the additional mass per story and the damping ratio of the whole structure are modeled as Gamma-distributed. The parameters of the FTMD, its mass, stiffness and friction force magnitude are modeled as Gamma-distributed, with the mean value of the stiffness and friction force magnitude being our design parameters. Finally, we consider the uncertainties related to a seismic excitation, the natural frequency of the Kanai-Tajimi filter (ω_g) and its damping ratio (ξ_g) are modeled with truncated-Normal distribution, whereas the peak ground acceleration (PGA) is modeled as log-normal. The parameters of the distributions are presented in Table 1.

Table 1. Statistical information about the random variables

Random variable	Probability distribution	Mean value	Coeff. of var. (%)	NRV[1]
Building frame				
Elasticity modulus (steel)	Gamma	200 GPa	5.0	70
Specific mass (steel)	Gamma	7500 kg/m^3	5.0	70
Additional mass per story	Gamma	44 t	5.0	1
Damping ratio	Gamma	0.05	10.0	1
FTMD				
Mass	Gamma	1.94 t	5.0	1
Stiffness	Gamma	$\mu_{k_{F_1}}(\mathbf{d})$	10.0	1
Friction force magnitude	Gamma	$\mu_{f_{F_1}}(\mathbf{d})$	10.0	1
Seismic excitation				
Natural frequency of the filter	Truncated-Normal	37.3 rad/s	20.0	1
Damping ratio of the filter	Truncated-Normal	0.30	20.0	1
PGA	Log-normal	0.500g	20.0	1

Since the probability distribution of the stiffness and friction force magnitude of the FTMD are dependent on design variables, the gradient $\nabla_{\mathbf{d}} f_{\mathbf{X}}$ in (11) does not vanish. The partial derivative of the logarithm of the pdf of a Gamma-distributed variable with respect to its mean is

$$\frac{\partial}{\partial \mu_i}\left(\log f_{\mathbf{x}}(\mu_i, \mathbf{x})\right) = \frac{-\mu_i + \mathbf{x}}{\varpi^2 \mu_i^2}. \tag{20}$$

Thus, for this case, the i-th element of $\mathcal{G}$ is

$$\mathcal{G}_i(\mathbf{d}, \mathbf{x}) = \frac{\partial}{\partial d_i}\beta(\mathbf{d}, \mathbf{x}) + \beta(\mathbf{d}, \mathbf{x})\frac{-\mu_i + \mathbf{x}}{\varpi_i^2 \mu_i^2}, \tag{21}$$

where μ_i and ϖ_i are, respectively, the mean and coefficient of variation of the i-th element of $\mathbf{d}$ (cf. Appendix B at [21])

To evaluate the performance of SGD and ASGD in the solution of the present numerical example, we compare their efficiency with NM and SQP methods coupled with sample average approximation. We randomly sample 10 starting points and perform optimization from these points using the four optimization methods (SGD, ASGD, NM, and SQP), using the same random number generator seed for each method. We run the algorithms until the stop criterion: $|\mathbf{d}_k - \mathbf{d}_{k-1}|_2 < 10^{-5}$. For SGD and ASGD, Polyak–Ruppert averaging $\bar{\mathbf{d}}$ is used instead of $\mathbf{d}$ for the stop criterion. Moreover, the barrier b on the horizontal displacement, as in (2), is set as 0.467; value obtained by Curadelli and Amani [1] from a static nonlinear analysis, and also used by Mantovani et al. [7].

[1]Number of random variables

Results

We present the mean value and standard deviation of the objective function and computational cost (OFE) reached by each algorithm over the 10 runs in Table 2.

Table 2. Single FTMD: Results for different optimization methods.

Optimization method	Objective function $(\bar{\beta})$		Cost ($10^3\times$ OFE)	
	mean	std	mean	std
NM	4.124	8.8×10^{-3}	63.6	6.72
SQP (fmincon)	4.128	2.8×10^{-3}	28.4	15.9
SGD	4.187	8.3×10^{-3}	6.05	3.62
ASGD	4.191	3.0×10^{-3}	1.14	0.45

A histogram of the cost required by each of the tested algoritms (Nelder-Mead, SQP, SGD, and ASGD) to solve this numerical example in ten independent runs is presented in Figure 9.

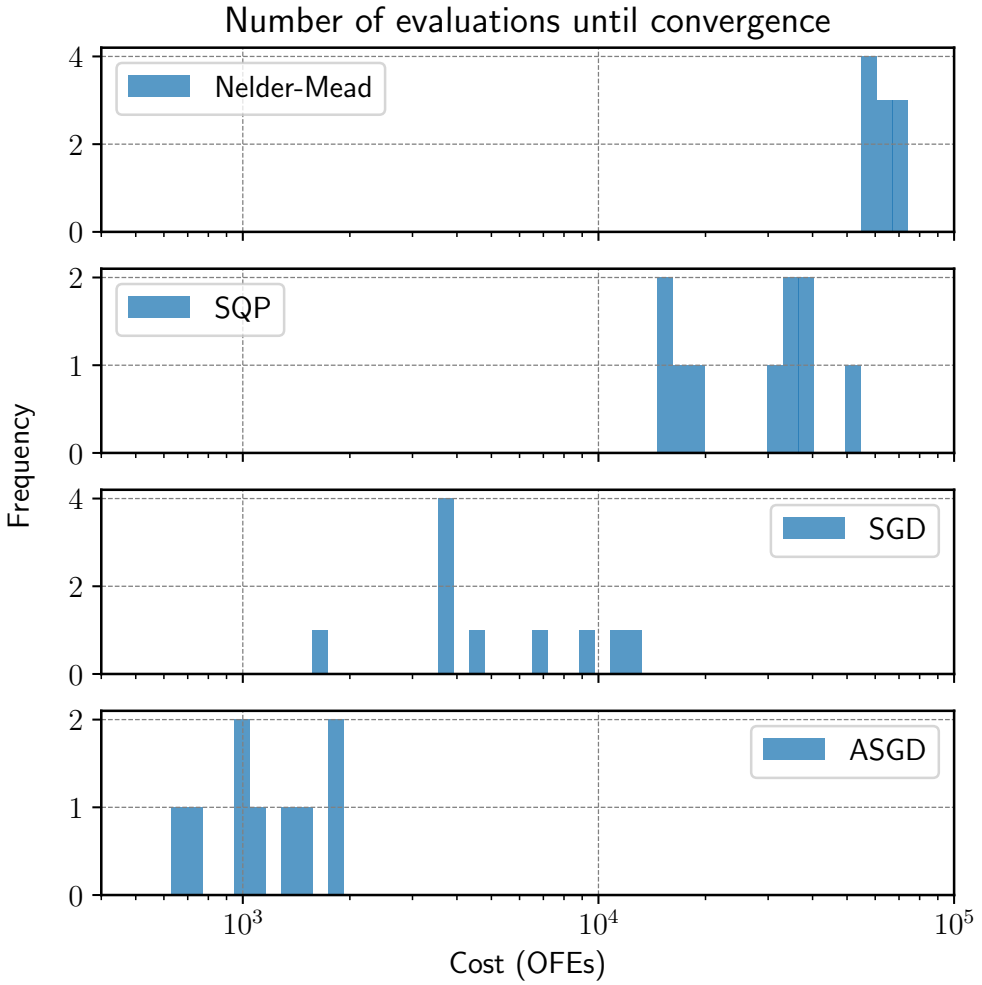

Figure 9. Single FTMD: Histogram of optimization costs for the different methods.

For the sake of comparison, the ten initial points are presented in Table 3 along with their respective reliability indexes. It can be observed, comparing with the results in Table 2, that ASGD is able to improve the reliability index. Averaging the reliability indexes at the start results in a probability of

Table 3. Starting points of the independent runs and their reliability indexes.

$\mathbf{d}_0$	k_1 (kN/m)	f_1 (kN)	$\hat{\beta}$
1	37.91	0.52	3.41
2	48.94	2.31	3.43
3	61.08	2.31	3.36
4	44.63	0.36	3.29
5	45.64	2.03	3.47
6	42.40	0.97	3.43
7	40.91	0.14	3.26
8	3.49	0.82	3.41
9	51.87	1.42	3.37
10	33.53	2.09	3.77

failure of 3.13×10^{-4}, whereas, considering the optimized design parameters, this probability reduces to 1.89×10^{-5}. Hence, we can conclude that the tested methods successfully improve the efficiency of the FTMD without introducing more mass, reducing the probability of failure up to one order of magnitude in comparison to randomly generated values. Moreover, ASGD is able to perform optimization with a fraction of the cost of the other methods, as can be observed in Figure 9.

5 Conclusion

In the present paper, we studied the performance of stochastic gradient methods in the maximization of the reliability index of a building with respect to the parameters of a friction tuned mass damper (FTMD). This optimization problem is a coupled reliability and optimization problem, where uncertainties in the building structure, loadings and the FTMDs must be taken into consideration. Moreover, the FTMDs have a non-linear model that increases even more the cost of the optimization process. Thus, we are interested in investigating if stochastic gradient methods are able to reduce the number of model evaluations required to find the optimum.

We used two stochastic gradient methods in this paper, the stochastic gradient descent (SGD) and the accelerated stochastic gradient descent (ASGD), where, in both cases, Polyak–Ruppert was used, and, for ASGD, an acceleration restart technique was also employed. We presented the pseudocodes for the algorithms used, namely, *vanilla* SGD, SGD with Polyak–Ruppert averaging, *vanilla* ASGD, and ASGD with restart technique. The pseudocodes allow other authors to reproduce the results obtained and to adapt the methods to their needs. Moreover, we deduced the unbiased gradient estimator of the reliability index, which is required for stochastic gradient optimization. This deduction is not trivial, given the dependency of the statistics of the random parameters on the design variables.

To test the optimization methods, we optimized the FTMD at the top of a building using SGD, ASGD, Nelder-Mead (NM), and sequential quadratic programming (SQP). We modeled uncertainties in the example as Mantovani et al [7], resulting in 148 random variables. SGD was able to perform better than NN and SQP, however, ASGD performed even better, finding better optima with less model evaluations, in average. ASGD found the optima with an average of 1140 model evaluations, whereas SQP needed 28400 model evaluations and NM required 63600 model evaluations.

The stochastic gradient methods tested were able to efficiently solve the passive control problem, providing accurate results with a fraction of the cost of classical methods. Moreover, we conclude that

Nesterov acceleration and the restart technique were able to further improve the performance of SGD. Hence, given that an unbiased gradient estimator is available, stochastic gradient methods are a viable alternative for local search on the presence of uncertainties.

Acknowledgments

The authors gratefully acknowledge the financial support of CNPq (National Counsel of Technological and Scientific Development) and CAPES (Coordination of Superior Level Staff Improvement).

References

[1] Curadelli, R. O. & Amani, M., 2014. Integrated structure-passive control design of linear structures under seismic excitations. *Engineering Structures*, vol. 81, pp. 256–264.

[2] Inaudi, J. A. & Kelly, J. M., 1995. Optimum absorber parameters for minimizing vibration response. *Journal of Engineering Mechanics*, vol. 121, n. 1, pp. 142–149.

[3] Ricciardelli, F. & Vickery, B. J., 1999. Tuned vibration absorbers with dry friction damping. *Earthquake Engineering & Structural Dynamics*, vol. 28, n. 7, pp. 707–723.

[4] Gewei, Z. & Basu, B., 2010. A study on friction-tuned mass damper: harmonic solution and statistical linearization. *Journal of Vibration and Control*, vol. 17, n. 5, pp. 721–731.

[5] Barbosa, A. R. & Ramadhan, G., 2014. Seismic performance of a tall diagrid steel building with tuned mass dampers. *International Journal of Innovations in Materials Science and Engineering*, vol. 1, n. 2, pp. 90–102.

[6] Pisal, A. Y. & Jangid, R. S., 2014. Seismic response of multi-story structure with multiple tuned mass friction dampers. *International Journal of Advanced Structural Engineering*, vol. 6, n. 1, pp. 1–13.

[7] Mantovani, G. Z., Fadel Miguel, L. F., Lopez, R. H., Miguel, L. F. F., & Torii, A. J., 2017. Optimum design of multiple friction tuned mass dampers under seismic excitations. *Proceedings of the 6th international symposium on solid mechanics, Joinville*, vol. , pp. 521–535.

[8] Robbins, H. & Monro, S., 1951. A stochastic approximation method. *The annals of mathematical statistics*, vol. 1, pp. 400–407.

[9] Ruppert, D., 1988. Efficient estimations from a slowly convergent Robbins-Monro process. Technical report, Cornell University Operations Research and Industrial Engineering.

[10] Polyak, B. T. & Juditsky, A. B., 1992. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, vol. 30, n. 4, pp. 838–855.

[11] Nesterov, Y., 1983. A method of solving a convex programming problem with convergence rate o (1/k2). In *Soviet Mathematics Doklady*, volume 27, pp. 372–376.

[12] O'Donoghue, B. & Candès, E., 2015. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, vol. 15, n. 3, pp. 715–732.

[13] Carlon, A. G., Dia, B. M., Espath, L. F., Lopez, R. H., & Tempone, R., 2018. Nesterov-aided stochastic gradient methods using Laplace approximation for Bayesian design optimization. *arXiv preprint arXiv:1807.00653*, vol. 1.

[14] Tajimi, H., 1960. A statistical method of determing the maximum response of a building structure during an earthquake. In *Proc. 2nd World Conf. Earthq. Eng.*, pp. 781–797.

[15] Kanai, K., 1961. An empirical formula for the spectrum of strong earthquake motions. *Earthquake research report*, vol. 39, pp. 85–95.

[16] Nelder, J. A. & Mead, R., 1965. A simplex method for function minimization. *The computer journal*, vol. 7, n. 4, pp. 308–313.

[17] Boggs, P. T. & Tolle, J. W., 1995. Sequential quadratic programming. *Acta numerica*, vol. 4, pp. 1–51.

[18] Kleywegt, A. J., Shapiro, A., & Homem-de Mello, T., 2002. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, vol. 12, n. 2, pp. 479–502.

[19] Melchers, R. E. & Beck, A. T., 2018. *Structural Reliability Analysis and Prediction*. John Wiley & Sons.

[20] Lopez, R. H., Fadel Miguel, L. F., & Beck, A. T., 2014. Tuned mass dampers for passive control of structures under earthquake excitations. *Encyclopedia of Earthquake Engineering*, vol. , pp. 1–12, Springer Berlin Heidelberg.

[21] Carlon, A., Lopez, R., Espath, L., Miguel, L., & Beck, A., 2019. A stochastic gradient approach for the reliability maximization of passively controlled structures. *Engineering Structures*, vol. 186, pp. 1–12.

[22] Nemirovski, A., 2005. *Efficient methods in convex programming*. TECHNION.

[23] Qian, N., 1999. On the momentum term in gradient descent learning algorithms. *Neural networks*, vol. 12, n. 1, pp. 145–151.

[24] Nesterov, Y., 2013. Introductory lectures on convex optimization: A basic course.

[25] Bertero, V. V. & Kamil, H., 1975. Nonlinear seismic design of multistory frames. *Canadian Journal of Civil Engineering*, vol. 2, n. 4, pp. 494–516.

[26] 1973. *Manual of Steel Construction*, American Institute of Steel Construction, Seventh edition.