AGGREGATION AND REGULARIZATION SCHEMES: A PROBABILISTIC POINT OF VIEW

A. J. TORII, J. R. DE FARIA AND A. A. NOVOTNY

Abstract. The main issue with existing constraint aggregation and regularization approaches (e.g. p-norm, p-mean and KS approaches) is that very high exponents are required to reduce the bias when the number of active constraints is large. In practice, the exponent employed for such approaches is limited due to numerical reasons. Thus, the results often present significant bias that may affect the quality of the designs obtained. In this work we propose a novel probabilistic constraint aggregation/regularization approach that does not suffer from such issues, allowing for very sharp regularization of extreme values (maximum or minimum) even in the presence of a large number of closely spaced values (e.g. several active constraints). Bias estimates for the p-norm and the proposed approach are also derived, that can be employed for choice of appropriate regularization parameters. The proposed approach is compared with p-norm regularization in the context of structural topology optimization considering maximization of the first natural vibration frequency. The topology optimization problem is solved with the help of an efficient algorithm based on the topological derivative method combined with a level-set representation of the design domain. In the numerical examples we demonstrate that very small aggregation bias can be obtained with the proposed approach. In contrast, the same is not possible with the p-norm approach, because the required exponent becomes too large and blocks the algorithm from running. These results demonstrate that the proposed probabilistic approach is more appropriate for regularization and aggregation in the presence of closely spaced values.

1. Introduction

An important topic for structural and multidisciplinary optimization is that of regularization and constraint aggregation approaches. By constraint aggregation we mean: to group several smooth constraints into a single (or small number of) smooth constraint. By regularization, on the other hand, we mean: to make smooth approximations to otherwise non-smooth quantities, such as extreme values (i.e., minimum/maximum eigenvalues). Note that constraint aggregation and regularization are not the same thing strictly speaking. However, these two subjects are closely related because most existing approaches for aggregation and regularization are based on the p-norm, the p-mean or the KS function (Kreisselmeier and Steinhauser, 1979). For this reason, we do not distinguish between the two for the rest of this paper.

The literature concerning aggregation approaches for structural optimization is so vast that it is difficult to present a fair perspective of the subject. A recent and detailed literature review can be found in the works by Fernández et al. (2019) and da Silva et al. (2021). The works by Yang and Chen (1996); Duysinx and Sigmund (1998); Poon and Martins (2007); París et al. (2009); Le et al. (2010); París et al. (2010); Kennedy and Hicken (2015); Gao et al. (2015); Verbart et al. (2017); Wang and Qian (2018) also present a general overview of the subject. See also the works by Torii et al. (2015); Torii and de Faria (2017); Ferrari and Sigmund (2019); Quinteros et al. (2021); Dalklint et al. (2021)

Key words and phrases. aggregation, regularization, eigenvalues, constraint, sensitivity, optimization.

for an overview of regularization schemes in the context of eigenvalues-based structural optimization problems.

The main idea behind regularization approaches is to make a smooth approximation to an otherwise non-smooth (or singular) functional, such as maximum or minimum values in a set. This occurs, for example, in problems involving buckling or vibration frequencies, where one is frequently interested in smallest magnitude eigenvalues. Since the minimum/maximum operators may become non-smooth, when minimum/maximum values are not unique, the resulting quantity of interest (e.g. smallest magnitude eigenvalue) may also become non-smooth. It is then necessary to employ some smooth approximation to the quantity of interest, in order to be able to employ widely available optimization algorithms that require the functions involved to be smooth. In the case of aggregation approaches, the idea is to transform a set of constraints into a single one. In this way it is possible to ensure that all constraints are satisfied by controlling only the aggregate constraint. This can be accomplished, for example, by taking the maximum value among the constraints (or the minimum, depending on the context). However, this direct aggregation approach leads to non-smooth aggregate constraints. For this reason, constraint aggregation generally requires an additional regularization step, where the otherwise non-smooth maximum/minimum operator is now replaced by some smooth approximation. In both cases, the problem is that the maximum and minimum operators are not always differentiable (i.e., not always smooth). This can lead to numerical issues during the optimization procedure, such as slow convergence rate and poor solutions, since most optimization algorithms widely available were developed for smooth problems.

A largely employed approach to overcome non-smooth maximum/minimum operators is to replace them by p-norm, p-mean and KS functions. Unfortunately, p-norm, p-mean and KS maximum aggregation approaches share a common issue: when we have several closely spaced extreme values, very high exponents are required in order to produce sharp approximations[1]. This is a serious issue in stress-constrained topology optimization problems, for example, where the number of constraints is usually very large and the structure is generally almost fully stressed. In these situations, very large exponents are usually required (e.g. 50, 100, 200), leading to serious numerical difficulties related to rounding errors and possibly blocking the computational routines from running. The issue is so serious that some authors proposed cluster aggregation approaches, where local constraints are not aggregate into a single constraint, but rather aggregated into a set of cluster constraints, each containing a limited number of local constraints (París et al., 2010). By limiting the number of local constraints per cluster it is then possible to employ lower exponents in order to get sharp approximations.

In this context, the goal of this work is threefold. First, we demonstrate why it is infeasible to obtain sharp approximations using p-norm and KS based approaches when the number of closely spaced values is large. We then propose a novel probabilistic aggregation/regularization approach that avoids employment of high order exponents. This allows sharp approximations even in the presence of a large number of closely spaced extreme values. Finally, we present relative bias estimates for the p-norm approach and the proposed approach. These estimates should be useful for parameter tuning and can be employed for adaptive schemes in the future.

The rest of this paper is organized as follows. In Section 2 we discuss some properties of the p-norm and KS approaches. Relative bias estimates for these approaches are also

[1]By sharp we mean an approximation that is close to the exact value. In this case, a sharp aggregation/regularization approach is one that gives a value close to the maximum/minimum.

presented. We then propose a novel probabilistic approach in Section 3, that avoids high order exponents. Bias estimates for this approach are also derived and compared with the p-norm approach. In Section 4 we describe the optimization problem and the algorithm employed to test the proposed approach. Numerical examples are presented in Section 5. The conclusions of this work are summarized in Section 6.

2. p-NORM REGULARIZATION

Here we present a brief review of the p-norm approach. The p-mean approach is very similar to the p-norm approach, apart from the fact that it produces a lower bound for the maximum value. For this reason, the p-mean approach is not addressed here. Consider the maximum value

$$\lambda_{\max} = \max_{i \in \mathcal{N}_n} |\lambda_i|, \qquad \text{with} \qquad \mathcal{N}_n = \{1, 2, \cdots, n\}. \tag{2.1}$$

The p-norm regularization scheme follows from the property

$$\lambda_{\max} = \lim_{p \to \infty} \|\lambda\|_p, \tag{2.2}$$

where

$$\overline{\lambda}_p = \|\lambda\|_p = \left(\sum_{i=1}^{n} \lambda_i^p \right)^{1/p}, \tag{2.3}$$

is the p-norm of the vector λ. This regularization scheme only makes sense if λ_i is always non-negative, because the p-norm converges to the maximum component in absolute value.

In practice we take p large enough and write the approximation

$$\overline{\lambda}_p \approx \lambda_{\max}. \tag{2.4}$$

For finite p this approximation is smooth (i.e., is continuously differentiable) as long as λ_i are continuously differentiable. It is also known that

$$\lambda_{\max} \leq \overline{\lambda}_p, \tag{2.5}$$

i.e., the regularization scheme produces an upper bound. A p-norm approach for regularization of the minimum was presented by Torii and de Faria (2017) and employed in the works by Quinteros et al. (2021) and Dalklint et al. (2021).

2.1. KS regularization: a p-norm point of view.

The KS (Kreisselmeier and Steinhauser, 1979) regularization scheme is given by

$$\overline{\lambda}_{\text{KS}} = \frac{1}{p_{KS}} \ln \left(\sum_{i=1}^{n} \exp\left(p_{KS} \lambda_i\right) \right), \tag{2.6}$$

where p_{KS} is the KS regularization parameter. Note that we can write

$$\begin{aligned} \overline{\lambda}_{\text{KS}} &= \ln \left(\sum_{i=1}^{n} \left(\exp\left(\lambda_i\right)\right)^{p_{KS}} \right)^{1/p_{KS}} \\ &= \ln \| \exp(\lambda) \|_{p_{KS}}, \end{aligned} \tag{2.7}$$

where $\exp(\lambda)$ represents the vector with components

$$\exp(\lambda) = \left\{e^{\lambda_1}, e^{\lambda_2}, \cdots, e^{\lambda_n}\right\}. \tag{2.8}$$

This puts in evidence that the KS regularization approach is a kind of p-norm approach involving exponential. Employment of exponential can be viewed as a very smart way of accounting for the possibility of negative λ_i. From the properties of the p-norm we conclude that

$$\lim_{p_{KS}\to\infty} \ln \|\exp(\lambda)\|_{p_{KS}} = \lambda_{\max}, \tag{2.9}$$

and

$$\ln \|\exp(\lambda)\|_{p_{KS}} \geq \lambda_{\max}. \tag{2.10}$$

These results also indicate a more general family of aggregation schemes, given by

$$\overline{\lambda} = g^{-1}\left(\|g(\lambda)\|_p\right), \tag{2.11}$$

where g is an increasing homeomorphism and g^{-1} its inverse. Appropriate choices of g can then be employed to obtain aggregation schemes with desirable properties. In particular, for p-norm and KS approaches we take $g := id(\cdot)$ (the identity function) and $g := \exp(\cdot)$, respectively. Here we do not study this subject in more details, even though it clearly deserves further investigation in the future.

2.2. **Sharpness of p-norm and KS approaches.** In this section we demonstrate why very high powers p must be employed to get $\overline{\lambda}$ close to $\lambda_{\max}$ when we have closely spaced extreme values. Here we define this property as sharpness of the regularization scheme, i.e., the difference between the true maximum/minimum value and the corresponding regularization scheme.

Suppose the values are sorted as to satisfy

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0, \tag{2.12}$$

and we have m repeated maximum values

$$\lambda_{\max} = \lambda_1 = \lambda_2 = \cdots = \lambda_m \neq 0,\ m \leq n. \tag{2.13}$$

In the analysis below, the particular cases where $\lambda_{\max} = 0$ or $m = n$ are trivially satisfied and, therefore, we will assume without loss of generality that $\lambda_{\max} > 0$ and $m < n$. Taking into account Eq. (2.11) and assuming that g is a function, the p-norm then results

$$\begin{aligned} \overline{\lambda} &= g^{-1}\left(\left[\sum_{i=1}^{n}\left(g(\lambda_i)\right)^p\right]^{1/p}\right) \\ &= g^{-1}\left(\left[m g(\lambda_i)^p + \sum_{i=m+1}^{n}\left(g(\lambda_i)\right)^p\right]^{1/p}\right), \end{aligned} \tag{2.14}$$

and for $p \to \infty$ we have

$$\begin{aligned}\lim_{p\to\infty}\overline{\lambda} &= \lim_{p\to\infty} g^{-1}\left(m^{1/p}g(\lambda_{\max})\left[1+\frac{1}{m}\sum_{i=m+1}^{n}\left(\frac{g(\lambda_i)}{g(\lambda_{\max})}\right)^{p}\right]^{1/p}\right)\\ &= g^{-1}\left(\lim_{p\to\infty}[m^{1/p}g(\lambda_{\max})]\right) = \lambda_{\max}.\end{aligned} \tag{2.15}$$

From the above result we conclude that when we have m repeated maximums, for large $p\to\infty$ we have approximately

$$\boxed{\overline{\lambda} = g^{-1}(m^{1/p}g(\lambda_{\max}))}. \tag{2.16}$$

Note that for the p-norm approach we have[2]

$$\overline{\lambda}_p = m^{1/p}\lambda_{\max}. \tag{2.17}$$

The error of the p-norm approach is given by

$$e_p = \overline{\lambda}_p - \lambda_{\max} = \left(m^{1/p}-1\right)\lambda_{\max}, \tag{2.18}$$

and we get the inconvenient factor $(m^{1/p}-1)$, that hinders convergence to the true maximum. This is an issue mainly when we get several closely spaced extreme values (i.e., when m is large). That is why very large values of p must be employed in stress constraint aggregation in order to get sharp estimates, for example.

With KS function things are not much different. In this case

$$\overline{\lambda}_{\text{KS}} = \frac{1}{p_{KS}}\ln(m) + \lambda_{\max}, \tag{2.19}$$

The error of the KS function is given by

$$e_{\text{KS}} = \overline{\lambda}_{\text{KS}} - \lambda_{\max} = \frac{1}{p_{KS}}\ln(m), \tag{2.20}$$

and we get another inconvenient term $1/p_{KS}\ln(m)$ hindering convergence again.

From Eqs. (2.3) and (2.6) we observe that the p-norm and the KS approach will have the same effective exponent when we take $p_{KS} = p/\lambda_{\max}$. In this case we have

$$\begin{aligned}e_{\text{KS}} - e_p &= \frac{\lambda_{\max}}{p}\ln(m) - \left(m^{1/p}-1\right)\lambda_{\max}\\ &= \left(\frac{\ln(m)}{p} - \left(m^{1/p}-1\right)\right)\lambda_{\max}.\end{aligned} \tag{2.21}$$

For $p\to\infty$ both terms in parenthesis converge to zero with similar rate and we thus conclude that

$$\lim_{p\to\infty}(e_{\text{KS}} - e_p) = 0 \tag{2.22}$$

when $p_{KS} = p/\lambda_{\max}$. This means that the two approaches are asymptotically equivalent under the above circumstances. These results indicate that the two approaches should have similar behavior in practice, as long as the same effective exponent is employed.

[2]For the p-mean approach we would get $\|\lambda\|_p = (m/n)^{1/p}\lambda_{\max}$.

This occurs because both are based on the p-norm. For this reason, we do not distinguish between the KS and p-norm approaches further in this work.

Finally, note that the relative bias of the p-norm approach results

$$\begin{aligned} \epsilon_p &= \frac{e_p}{\lambda_{\max}} \\ &= m^{1/p} - 1. \end{aligned} \tag{2.23}$$

Solving Eq. (2.23) for p gives

$$p = \frac{1}{\log_m (1 + \epsilon_p)}, \tag{2.24}$$

that can be employed to set p in order to get relative bias ϵ_p for a given m.

For $m = 10$ and $\epsilon_p = 1/100$, for example, Eq. (2.24) gives $p = 231.41$. This puts in evidence that p-norm approaches are unfeasible in practice when m is large (i.e., when we get several closely spaced extreme values), since too large values for p are required in this case. If we consider only two closely spaced extreme values (i.e., $m = 2$) and $\epsilon_p = 1/100$, Eq. (2.24) now gives $p = 69.66$. We observe that even in this case a high power p is required, that may lead to computational issues in some situations. Note that the p-mean and KS approaches will also lead to similar limitations because of their similarity to the p-norm approach.

3. Proposed probabilistic regularization approach

In order to overcome the issues point-out in the previous section we propose a novel probabilistic regularization scheme. We consider first regularization of the maximum value. Regularization of the minimum is very similar, as demonstrated latter.

Consider then the maximum value

$$\lambda_{\max} = \max_{i \in \mathcal{N}_n} \lambda_i, \qquad \text{with} \qquad \mathcal{N}_n = \{1, 2, \cdots, n\}. \tag{3.1}$$

The difficulty with this direct approach is that the maximum value may be non-differentiable. Suppose the values are sorted as to satisfy Eq. (2.12). Then the maximum value is non-differentiable when it is not unique, i.e., when the situation from Eq. (2.13) occurs.

In this work we avoid non-differentiability of $\lambda_{\max}$ by applying random perturbations to λ_i. Consider the perturbations[3]

$$\alpha_i = Y_i \lambda_i, \quad i \in \mathcal{N}_n, \tag{3.2}$$

where Y_i are independent random variables. Here we assume that Y_i have continuous uniform distribution on the interval $[1 - h/2, 1 + h/2]$, i.e., with the probability density given by (Rubinstein and Kroese, 2008)

$$f(y) = \begin{cases} \frac{1}{h} & \text{if } 1 - h/2 \leq y \leq 1 + h/2 \\ 0 & \text{otherwise} \end{cases}. \tag{3.3}$$

In this case, the expected value $\mathbb{E}[Y_i]$ and the variance $\mathbb{V}[Y_i]$ result (Rubinstein and Kroese, 2008)

[3]Note that shift perturbations $\alpha_i = Y_i + \lambda_i$ could also have been employed, by taking $\mathbb{E}[Y_i] = 0$. Besides, other distributions can also be employed for Y_i.

$$\mathbb{E}\left[Y_i\right] = 1, \tag{3.4}$$

$$\mathbb{V}\left[Y_i\right] = \frac{h^2}{12}. \tag{3.5}$$

We then take the regularization scheme

$$\boxed{\overline{\lambda}_h = \mathbb{E}\left[\alpha_{\max}\left(Y\right)\right]}, \tag{3.6}$$

where $Y = \{Y_1, Y_2, \cdots, Y_n\}$ and

$$\begin{aligned} \alpha_{\max}\left(Y\right) &= \max_{i \in \mathcal{N}_n} \alpha_i \\ &= \max_{i \in \mathcal{N}_n} (Y_i \lambda_i). \end{aligned} \tag{3.7}$$

Note that the regularization parameter h actually defines the range of perturbations applied to λ_i.

The quantity $\alpha_{\max}\left(Y\right)$ is non-differentiable for

$$y_1\lambda_1 = y_2\lambda_2 = \cdots = y_n\lambda_n, \tag{3.8}$$

where y_i are realizations of the random variables. However, the above condition defines a zero measure set with respect to the probability density of the random vector Y (Loève, 1977; Shriryaev, 1995), unless $\lambda_1 = \lambda_2 = ... = \lambda_n = 0$. In other words, the probability of occurrence of the above event is null if some $\lambda_i \neq 0$. In this case the expected value $\overline{\lambda}_h$ is actually differentiable for $h > 0$. Also note that

$$\lim_{h \to 0} \mathbb{E}\left[\alpha_{\max}\left(Y\right)\right] = \lambda_{\max}, \tag{3.9}$$

since in this case the perturbations are removed. This demonstrates that the regularization converges to the exact value for $h \to 0$.

Invoking the Dominated Convergence Theorem (Loève, 1977; Shriryaev, 1995), sensitivity with respect to a design parameter τ results

$$\begin{aligned} \frac{d\overline{\lambda}_h}{d\tau} &= \frac{d}{d\tau}\mathbb{E}\left[\alpha_{\max}\left(Y\right)\right] \\ &= \mathbb{E}\left[\frac{d}{d\tau}\alpha_{\max}\left(Y\right)\right], \end{aligned} \tag{3.10}$$

that can be estimated with Monte Carlo Simulation (MCS) by (Rubinstein and Kroese, 2008)

$$\overline{\lambda}_h \approx \frac{1}{N}\sum_{y \in S} \alpha_{\max}(y), \tag{3.11}$$

$$\frac{d\overline{\lambda}_h}{d\tau} \approx \frac{1}{N}\sum_{y \in S} \frac{d\alpha_{\max}(y)}{d\tau}, \tag{3.12}$$

with

$$\alpha_{\max}(y) = \max_{i \in \mathcal{N}_n}(y_i \lambda_i), \tag{3.13}$$

where S is a sample of size N for the random vector Y, y are realizations of the random vector and y_i are components of y. In order to avoid oscillations due to random sampling we recommend taking a fixed sample during the entire optimization procedure. This is known as Common Random Variable Approach (Rubinstein and Kroese, 2008).

Note that evaluation of Eqs. (3.11) and (3.12) does not require additional evaluations of λ_i and its sensitivity, because sampling only requires random scaling of λ_i and identification of the resulting maximum for each realization of Y. For this reason, the computational cost of evaluating Eqs. (3.11) and (3.12) can be neglected in comparison to the computational cost for evaluation of λ_i and its sensitivities.

Regularization of the minimum can be written as

$$\boxed{\overline{\lambda}_h = \mathbb{E}\left[\alpha_{\min}\left(Y\right)\right]}, \tag{3.14}$$

with

$$\alpha_{\min}\left(Y\right) = \min_{i \in \mathcal{N}_n} \alpha_i. \tag{3.15}$$

This is another advantage of the proposed approach: regularization of the maximum and regularization of the minimum are very similar in practice. Note that this is not true for the p-norm and the KS approaches, that were originally conceived for regularization of the maximum and require some adaptation in order to be applied for regularization of the minimum (Torii and de Faria, 2017). Finally, it should be stressed that the proposed probabilistic approach does not require the values λ_i to be non-negative.

3.1. **Sharpness of the proposed approach.** We now evaluate the sharpness of the proposed aggregation approach. For m repeated maximum values and small $h \to 0$ we have

$$\begin{aligned} \alpha_{\max}\left(Y\right) &= \max_{i \in \mathcal{N}_m} \alpha_i \\ &= \max_{i \in \mathcal{N}_m} (Y_i \lambda_i) \\ &= \lambda_{\max} \max_{i \in \mathcal{N}_m} Y_i, \end{aligned} \tag{3.16}$$

and thus

$$\overline{\lambda}_h = \lambda_{\max} \mathbb{E}\left[\max_{i \in \mathcal{N}_m} Y_i\right]. \tag{3.17}$$

From Order Statistics we know that (David and Nagaraja, 2003)

$$\mathbb{E}\left[\max_{i \in \mathcal{N}_m} Y_i\right] = \mathbb{E}[Y_i] + k_m \sqrt{\mathbb{V}[Y_i]}, \tag{3.18}$$

where k_m is a constant that depends on m, with $k_2 = 0.5744$, $k_{10} = 1.4171$, $k_{100} = 1.6978$ and $k_{1000} = 1.7286$ for the continuous uniform distribution[4]. Also note that for the uniform distribution k_m is bounded by

[4]Note that for Y_i with Normal distribution we have $k_2 = 0.5642$, $k_{10} = 1.5388$, $k_{100} = 2.5076$ and $k_{1000} = 3.2414$ (David and Nagaraja, 2003). These values are, in general, higher than those obtained for Y_i with uniform distribution. Consequently, employment of Y_i with Normal distribution would reduce the sharpness of the estimates for large m. That is the reason why we choose Y_i with uniform distribution in this work. However, for small m (e.g. $m \leq 10$) employment of Y_i with Normal distribution would likely lead to very similar results.

$$k_m \leq \sqrt{3}, \tag{3.19}$$

because $\mathbb{E}\left[\max_{i\in\mathcal{N}_m} Y_i\right]$ must lie in the range of the distribution. Thus, for m repeated values we get, from Eq. (3.5),

$$\boxed{\overline{\lambda}_h = \left(1 + \frac{\sqrt{3}}{6} k_m h\right) \lambda_{\max} \leq \left(1 + \frac{1}{2} h\right) \lambda_{\max}}, \tag{3.20}$$

for $h \to 0$. We again get an inconvenient term of the form $k_m h\sqrt{3}/6$ that hinders convergence when we have closely spaced extreme values, as occurs for the p-norm and the KS approaches.

However, there is no computational issues arising from small h in this case, since there are no high order exponents involved. For this reason, we are able to get sharp estimates by taking h very small. This is the main advantage of the proposed approach. The regularization parameter can be reduced indefinitely without breaking-down the algorithm. This allows very sharp regularization.

From Eq. (3.20) we observe that the bias of the proposed approach results

$$\begin{aligned} e_h &= \overline{\lambda}_h - \lambda_{\max} \\ &= \frac{\sqrt{3}}{6} k_m h \lambda_{\max}, \end{aligned} \tag{3.21}$$

and the relative bias results

$$\epsilon_h = \frac{e_h}{\lambda_{\max}} = \frac{\sqrt{3}}{6} k_m h, \tag{3.22}$$

for $h \to 0$. Thus, by setting h as

$$h = \frac{2\sqrt{3}}{k_m} \epsilon_h, \tag{3.23}$$

we should get a relative bias close to ϵ_h.

Besides, from Eq. (3.19) we also get

$$h \geq 2\epsilon_h. \tag{3.24}$$

Thus, by setting h as

$$h = 2\epsilon_h, \tag{3.25}$$

we should get a relative bias smaller than ϵ_h no matter how many closely spaced extreme values we have (i.e., for any value of m). For $\epsilon_h = 1/100$ we get $h = 2/100$, for example. Again, note that these results hold for small h (i.e., $h \to 0$).

3.2. **Comparison to the p-norm approach.** In Table 1 we present suggested values for h when the Probabilistic Regularization Approach is employed. The values were obtained with Eq. (3.23). The results are not exact in practice because the expressions hold for $h \to 0$.

Suggested values for the parameter p of the p-norm approach are presented in Table 2. The values were obtained with Eq. (2.24) and were rounded up to the next integer. The results are also not exact in practice because the expressions hold for $p \to \infty$. These results can also be easily adapted for the p-mean and the KS approaches.

TABLE 1. Suggested values for h in Probabilistic Regularization Approach

ϵ_h	$m = 2$	$m = 10$	$m = 100$	$m \geq 1000$
10.0%	0.6031	0.2445	0.2040	0.2004
5.0%	0.3015	0.1222	0.1020	0.1002
1.0%	0.0603	0.0244	0.0204	0.0200
0.5%	0.0302	0.0122	0.0102	0.0100
0.1%	0.0060	0.0024	0.0020	0.0020

TABLE 2. Suggested values for p in p-norm Approach

ϵ_p	$m = 2$	$m = 10$	$m = 100$	$m = 1000$
10.0%	8	25	49	73
5.0%	15	48	95	142
1.0%	70	232	463	695
0.5%	139	462	924	1386
0.1%	694	2304	4608	6912

From Tables 1 and 2 we observe that the approaches can be tuned to have equivalent sharpness. In order to get a relative bias close to 5% in the presence of two closely spaced extreme values ($m = 2$), for example, we can take $h = 0.3015$ and $p = 15$.

However, from Table 2 we observe that very high values must be taken for p in the presence of several closely spaced extreme values (e.g. $m \geq 10$) if we require a relative bias smaller than 1%. For $m = 10$, for example, if we require a relative bias equal to 1% we should take p close to 232. This would likely lead to numerical issues. Even for $m = 2$ and a relative bias of 1% we already need a large value for p ($p = 70$). This puts in evidence that the p-norm approach (and the related p-mean and KS approaches) quickly become infeasible from the computational point of view if m is large and a small relative bias is required.

From Table 1 we observe that small values for h are required in the presence of several closely spaced extreme values (e.g. $m \geq 100$) if small relative bias is demanded. However, employment of very small values for h do not lead to numerical issues, as demonstrated in the numerical examples. For this reason the proposed approach is able to obtain sharp approximations even in the presence of several closely spaced extreme values.

4. Application problem

In order to illustrate the effectiveness of the proposed regularization approach, we apply the concepts previously developed to the problem of structural topology optimization for maximization of the first natural vibration frequency.

4.1. **Problem setting.**

We assume that the natural vibration frequencies of the structure are ordered as

$$\omega_1 \leq \omega_2 \leq \omega_3 \leq \cdots , \tag{4.1}$$

where ω_i are the vibration frequencies of the structure. In this case

$$\omega_1 = \min_{i \in \mathcal{N}} \omega_i, \ \mathcal{N} = \{1, 2, \cdots\} \tag{4.2}$$

is the fundamental vibration frequency of the structure. Here we consider the optimization problem: Find the design parameter τ, such that

$$\min \left[F(\tau) = \mu_1 \frac{1}{\overline{\omega}(\tau)} + \mu_2 V(\tau) \right], \tag{4.3}$$

where $V(\tau)$ is the volume of material, $\overline{\omega}(\tau)$ is the regularized minimum obtained with the p-norm approach or the proposed probabilistic approach and $\mu_1, \mu_2 > 0$ are given weights. This is a multi-objective optimization problem where we wish to maximize $\overline{\omega}(\tau)$ and minimize the volume of material employed. Although the formulation is simple, it is appropriate for the comparisons made in this work.

The natural vibration frequencies are given by

$$\omega_i = \sqrt{\lambda_i}, \tag{4.4}$$

where λ_i are the eigenvalues, solutions to the following spectral problems: Find (λ_i, u_i), such that

$$\left\{ \begin{array}{rcll} -\text{div}\sigma(u_i) & = & \lambda_i \rho u_i & \text{in } \mathcal{D}, \\ u_i & = & 0 & \text{on } \Gamma_D, \\ \sigma(u_i)n & = & 0 & \text{on } \Gamma_N, \end{array} \right. \tag{4.5}$$

where u_i are the associated eigenfunctions. The hold-all domain is denoted as $\mathcal{D}$, with boundary $\partial\mathcal{D} = \Gamma_D \dot{\cup} \Gamma_N$. The stress tensor $\sigma(u_i)$ and the parameter ρ are given by:

$$\sigma(u_i) = \tau \mathbb{C} \varepsilon(u_i) \quad \text{and} \quad \rho = \tau \rho_0 t, \tag{4.6}$$

with the thickness t and density ρ_0 assumed to be constant everywhere. The linearized Green tensor is defined as follows

$$\varepsilon(u_i) = \frac{1}{2}(\nabla u_i + \nabla u_i^\top). \tag{4.7}$$

The elasticity tensor is written as

$$\mathbb{C} = \frac{tE}{1-\nu^2}((1-\nu)\mathbb{I} + \nu(\mathrm{I} \otimes \mathrm{I})), \tag{4.8}$$

in which E is the Young modulus and ν the Poisson ratio, both considered constants everywhere. In addition, $\mathbb{I}$ and I are the fourth and the second order identity tensors, respectively. Finally, the statement of the problem is complemented with the definition of a piecewise constant function τ, such that:

$$\tau(x) := \left\{ \begin{array}{ll} 1, & \text{if } x \in \Omega, \\ \gamma_0, & \text{if } x \in \mathcal{D} \setminus \overline{\Omega}, \end{array} \right. \tag{4.9}$$

where $\overline{\Omega}$ is the closure of Ω and $0 < \gamma_0 \ll 1$ is used to mimic voids.

For the probabilistic approach we employ the regularization scheme from Eq. (3.14) directly. We again emphasize that the sample employed for MCS is kept fixed during the optimization procedure in order to avoid numerical oscillations arising from MCS variability.

Since the p-norm approaches are tailored to address regularization of the maximum, a transformation scheme must be used in order to allow regularization of the minimum. The scheme employed here for p-norm regularization of the minimum is given by (Torii and de Faria, 2017)

$$\overline{\omega} = \frac{1}{\| q \|_p}, \tag{4.10}$$

with

$$q = \left\{ \frac{1}{\omega_1}, \frac{1}{\omega_2}, \cdots, \frac{1}{\omega_n} \right\}. \tag{4.11}$$

4.2. **Topological derivative method.** The optimization problem (4.3) is solved with the help of a level-set domain representation method combined with the topological derivative concept (Amstutz and Andrä, 2006). The associated topological derivative is given by:

Theorem 1. *Let us consider that the λ_i-th eigenvalue of Eq. (4.5) is simple. Then, its topological derivative, with respect to the nucleation of a small inclusion endowed with different material from the background, is given by*

$$D_{\mathcal{T}}\lambda_i = -\frac{\mathbb{P}_\gamma \sigma(u_i) \cdot \varepsilon(u_i) - (1-\gamma)\lambda_i \rho \|u_i\|^2}{\displaystyle\int_{\mathcal{D}} \rho \|u_i\|^2}, \tag{4.12}$$

where u_i is the eigenvector associated with the λ_i-th eigenvalue of Eq. (4.5) and $\gamma(x)$ is the contrast on the material properties defined as

$$\gamma(x) = \begin{cases} \gamma_0, & x \in \Omega, \\ \gamma_0^{-1}, & x \in \mathcal{D} \setminus \Omega, \end{cases}$$

where $0 < \gamma_0 \ll 1$. We are interested into two particular cases, which are:
Case 1. *Let us consider $x \in \Omega$. In this case, $\tau = 1$ and a small portion of material is removed. Then the topological derivative reads*

$$D_{\mathcal{T}}\lambda_i = -\frac{\mathbb{P}_0 \sigma(u_i) \cdot \varepsilon(u_i) - \lambda_i \rho_0 t \|u_i\|^2}{\displaystyle\int_{\mathcal{D}} \rho \|u_i\|^2}. \tag{4.13}$$

where the polarization tensor $\mathbb{P}_0$ is written as

$$\mathbb{P}_0 = \frac{4}{1+\nu}\mathbb{I} - \frac{1-3\nu}{1-\nu^2}(\mathrm{I} \otimes \mathrm{I}). \tag{4.14}$$

Finally, the topological derivative of the volume $V(\tau)$ is trivially given by

$$D_{\mathcal{T}}V = -1. \tag{4.15}$$

Case 2. *Now, let us consider $x \in \mathcal{D} \setminus \overline{\Omega}$. In this case $\tau = \gamma_0 \ll 1$ and a small portion of material is added. Then the topological derivative can be written as*

$$D_{\mathcal{T}}\lambda_i = -\frac{\mathbb{P}_\infty \sigma(u_i) \cdot \varepsilon(u_i) + \lambda_i \rho_0 t \|u_i\|^2}{\displaystyle\int_{\mathcal{D}} \rho \|u_i\|^2}. \tag{4.16}$$

with the polarization tensor $\mathbb{P}_\infty$ given by

$$\mathbb{P}_\infty = -\frac{4}{3-\nu}\mathbb{I} - \frac{1-3\nu}{(1+\nu)(3-\nu)}(\mathrm{I} \otimes \mathrm{I}). \tag{4.17}$$

Finally, the topological derivative of the volume $V(\tau)$ is trivially given by

$$D_{\mathcal{T}}V = +1. \tag{4.18}$$

Proof. See, for instance, the paper by Amstutz (2011). □

The topological derivative of the shape functional $F(\tau)$ from Eq. (4.3) is given by

$$D_{\mathcal{T}}F = -\frac{\mu_1}{\overline{\omega}^2} D_{\mathcal{T}}\overline{\omega} + \mu_2 D_{\mathcal{T}}V, \tag{4.19}$$

where $D_{\mathcal{T}}\overline{\omega}$ has to be derived according to the regularization procedure we are dealing with. From Theorem 1, the topological derivatives of the regularized fundamental vibration frequency of the structure are given by

$$D_{\mathcal{T}}\overline{\omega} = \sum_{i=1}^{n} \left(\frac{q_i}{\|q\|_p} \right)^{p+1} D_{\mathcal{T}}\omega_i, \tag{4.20}$$

in the case of p-norm regularization (Eq. 2.3) and by

$$D_{\mathcal{T}}\overline{\omega} \approx \frac{1}{N} \sum_{y \in S} D_{\mathcal{T}}\alpha_{\max}(y), \tag{4.21}$$

in the case of the probabilistic approach regularization (Eq. 3.11), with

$$\alpha_{\max}(y) = \max_{i \in \mathcal{N}_n} (y_i \omega_i). \tag{4.22}$$

Finally, $D_{\mathcal{T}}\omega_i$ is simply obtained as follows

$$D_{\mathcal{T}}\omega_i = \frac{D_{\mathcal{T}}\lambda_i}{2\sqrt{\lambda_i}}, \tag{4.23}$$

with $D_{\mathcal{T}}\lambda_i$ and $D_{\mathcal{T}}V$ given by Theorem 1.

4.3. **Topology optimization algorithm.** Now, we have all elements to explain the topology design algorithm proposed by Amstutz and Andrä (2006). It consists basically in achieving a local optimality condition for the minimization problem (4.3), given in terms of the topological derivative and a level-set function. In particular, the domain $\Omega \subset \mathcal{D}$ and the complement $\mathcal{D} \setminus \Omega$ are characterized by a level-set function Ψ of the form:

$$\Omega = \{x \in \mathcal{D} : \Psi(x) < 0\} \quad \text{and} \quad \mathcal{D} \setminus \Omega = \{x \in \mathcal{D} : \Psi(x) > 0\}, \tag{4.24}$$

where Ψ vanishes on the interface between Ω and $\mathcal{D} \setminus \Omega$. A local sufficient optimality condition for problem (4.3), under a class of domain perturbations given by ball-shaped inclusions, can be stated as

$$D_{\mathcal{T}}F(x) > 0 \quad \forall x \in \mathcal{D}, \tag{4.25}$$

where $D_{\mathcal{T}}F(x)$ is the topological derivative of the shape functional $F(\tau)$ at $x \in \mathcal{D}$. Therefore, let us define the quantity

$$G(x) := \begin{cases} -D_{\mathcal{T}}F(x) & \text{if } \Psi(x) < 0, \\ +D_{\mathcal{T}}F(x) & \text{if } \Psi(x) > 0, \end{cases} \tag{4.26}$$

which allows to rewrite the condition from Eq. (4.25) in the following equivalent form:

$$\begin{cases} G(x) < 0 & \text{if } \Psi(x) < 0, \\ G(x) > 0 & \text{if } \Psi(x) > 0. \end{cases} \tag{4.27}$$

We observe that Eq. (4.27) is satisfied, where the quantity G coincides with the level-set function Ψ up to a strictly positive factor, namely $\exists\, \varphi > 0 : G = \varphi\Psi$. In order to fulfil this condition numerically, we start by choosing an initial level-set function Ψ_0. In a generic iteration k, we compute the function G_k associated with the level-set function Ψ_k. Thus,

the new level-set function Ψ_{k+1} is updated according to the following linear combination between the functions G_k and Ψ_k:

$$\begin{aligned} &\Psi_0 : \|\Psi_0\|_{L^2(\mathcal{D})} = 1, \\ &\Psi_{k+1} = (1-w)\Psi_k + w\frac{G_k}{\|G_k\|_{L^2(\mathcal{D})}} \quad \forall k \in \mathcal{N}, \end{aligned} \tag{4.28}$$

where $w \in (0, 1]$ is a step size determined by a line-search performed in order to decrease the value of the objective function $F(\rho)$. For more details, see the book by (Novotny and Sokolowski, 2020, Ch. 5), where the resulting topology design algorithm is presented in pseudo-code format.

5. Numerical examples

In all the following examples the material has elastic modulus $E = 210\text{x}10^9\text{N/m}^2$, Poisson coefficient $\nu = 0.3$ and specific weight $\rho_0 = 8000\text{kg/m}^3$. The structure is under plane stress considering thickness $t = 0.1\text{m}$. We also take $\gamma_0 = 0.001$, $\mu_2 = 1/\overline{\omega}_0$ and $\mu_1 = 1/V_0$, where $\overline{\omega}_0$ and V_0 are the volume and the regularized fundamental vibration frequency of the initial design (i.e. obtained with the domain completely filled with material). The relative bias of the regularization schemes is evaluated as

$$\epsilon = \frac{\overline{\omega} - \omega_1}{\omega_1}. \tag{5.1}$$

A sample of size $N = 10^5$ was employed for MCS required by the proposed Probabilistic Approach. The sample is kept fixed during the entire optimization procedure.

In this work the value of the parameters p and h are kept constant during the optimization procedure. In practice we recommend employment of continuation approaches to progressively increase/reduce the values of these parameters. Here we do not employ continuation approaches because we wish to compare only the aggregation schemes in its original forms.

5.1. **Example 1: Portal.** The first example was taken from the paper by Torii and de Faria (2017) and considers the structural domain from Fig. 1. The displacements at the lower border are constrained and a concentrated non structural mass equal to 10^4kg is included at the center of the upper border. A uniform mesh with 64,000 triangular finite elements is used. Symmetry of the structure is not enforced.

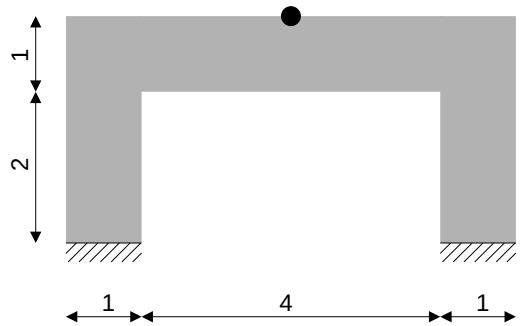

Figure 1. Portal

We observe two dominant closely spaced vibration frequencies, related to lateral and horizontal vibration modes (i.e., $m = 2$). In order to ensure a relative bias close to 1%

we then employ the p-norm approach with $p = 70$ (see Table 2). After 40 iterations we get the design presented in Figure 2 (top). The values of the objective function F, the relative volume V/V_0, the fundamental vibration frequency ω_1 and the relative bias ϵ are presented in Table 3. Note that the relative bias obtained with $p = 70$ is very close to the target of 1%. This confirms that the estimates from Table 2 are accurate enough for practical applications.

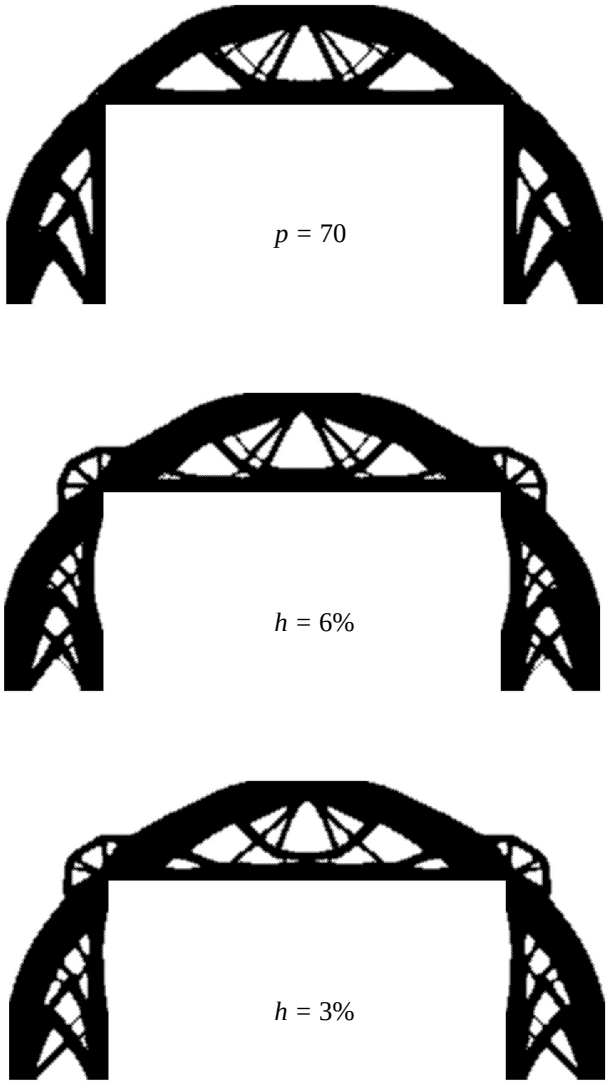

FIGURE 2. Portal designs

TABLE 3. Comparison of portal designs

Approach	Iterations	F	V/V_0	ω_1	ϵ
$p = 70$	40	1.6722	51.2104%	257.38	0.8436%
$h = 6.0\%$	40	1.6769	54.4729%	263.98	0.9506%
$h = 3.0\%$	52	1.6703	53.8104%	262.67	0.4439%

We now solve the problem with the proposed probabilistic approach. In order to ensure a relative bias close to 1% we take $h = 6.0\%$ (see Table 1). After 40 iterations we get the design presented in Figure 2 (middle). The information about this solution is also presented in Table 3. We observe that the design obtained is different from that obtained with the p-norm, even though they have very similar objective function. Thus, the designs seem to be equivalent local optima. We also observe that the relative bias of the regularization is again close to the target value of 1%, confirming accuracy of the estimates from Table 1.

If we wish to reduce the relative bias to 0.5% we should take $p = 140$. However, the optimization algorithm was not able to run with such a high exponent, due to numerical issues. We then employed the probabilistic approach with $h = 3.0\%$ (see Table 1). After 52 iterations we obtain the design presented in Figure 2 (bottom). Detailed information of this design is presented in Table 3. Indeed, the relative bias is very close to the target value of 0.5%.

From these results we observe that the probabilistic approach is able to give similar results to the p-norm approach. However, in the p-norm approach we are not able to reduce the relative bias beyond a certain level, because the exponent required becomes too large and causes numerical issues. The same difficulty does not occur for the probabilistic approach.

The evolution of the objective function during the optimization procedure is presented in Figure 3, for the three cases studied. We observe that no significant difference between the three cases can be observed, apart from the fact that the algorithm took more iterations to convergence for $h = 3\%$. This indicates that no significant differences between the p-norm approach and the proposed probabilistic approach are expected when equivalent regularization parameters are employed, namely $p = 70$ and $h = 6\%$.

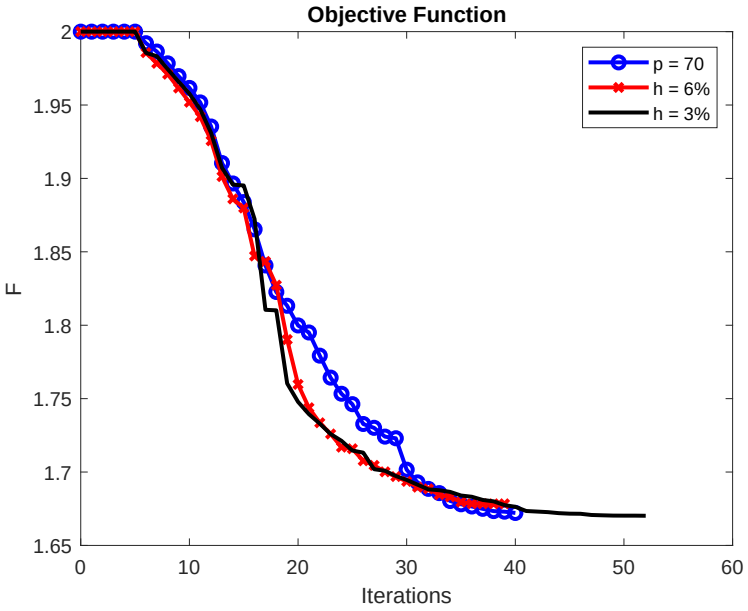

FIGURE 3. Convergence of the objective function for portal designs

5.2. **Example 2: Square domain.** In the second example we consider the square structural domain from Fig. 4. The displacements at the corners are constrained and concentrated non structural masses equal to 10^4kg are located at the middle of each side. A

uniform mesh with 57,600 triangular finite elements is used. Symmetry of the structure is not enforced.

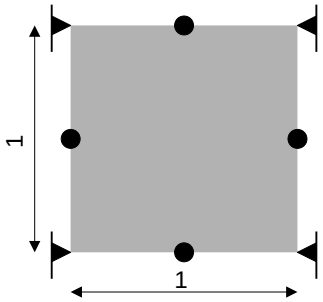

FIGURE 4. Square domain

Again, we start by solving the problem using the p-norm approach with $p = 70$ and the probabilistic approach with $h = 6.0\%$. It was observed that this problem also has two closely spaced vibration frequencies, related to vertical and horizontal vibration modes. Thus, these parameters should give a relative bias close to 1%.

The results are presented in Figure 5 (top and middle, respectively) and Table 4. The designs again have similar properties (although they look a little different), indicating that they seem to be equivalent local optima. We also observe that the relative bias is very close to the target value of 1%, indicating again that the estimates from Tables 1 and 2 are accurate.

We then solve the problem with the probabilistic approach using $h = 3.0\%$. This should produce a relative bias of the regularization close to 0.5%. The results are also presented in Figure 5 (bottom) and Table 4. We again observe that the relative bias is indeed close to the target value. As occurs in the previous example, the algorithm was not able to run with the p-norm approach for $p = 140$, because of numerical issues.

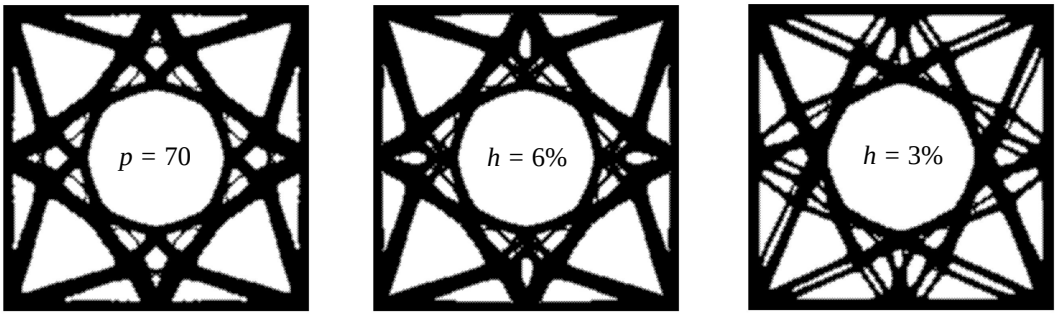

FIGURE 5. Square domain designs

TABLE 4. Comparison of square domain designs

Approach	Iterations	F	V/V_0	ω_1	ϵ
$p = 70$	17	1.6870	51.1065%	1035.7	1.1029%
$h = 6.0\%$	19	1.6762	49.9907%	1035.0	1.0097%
$h = 3.0\%$	25	1.6571	47.6759%	1032.1	0.4829%

The evolution of the objective function during the optimization procedure is presented in Figure 6, for the three cases studied. We observe that no significant difference between the three cases can be observed, apart from the fact that the algorithm took more iterations to convergence and was able to obtain a slightly better design for $h = 3\%$. Again, this indicates that the proposed probabilistic approach should be equivalent to the p-norm approach when equivalent regularization parameters are employed ($p = 70$ and $h = 6\%$).

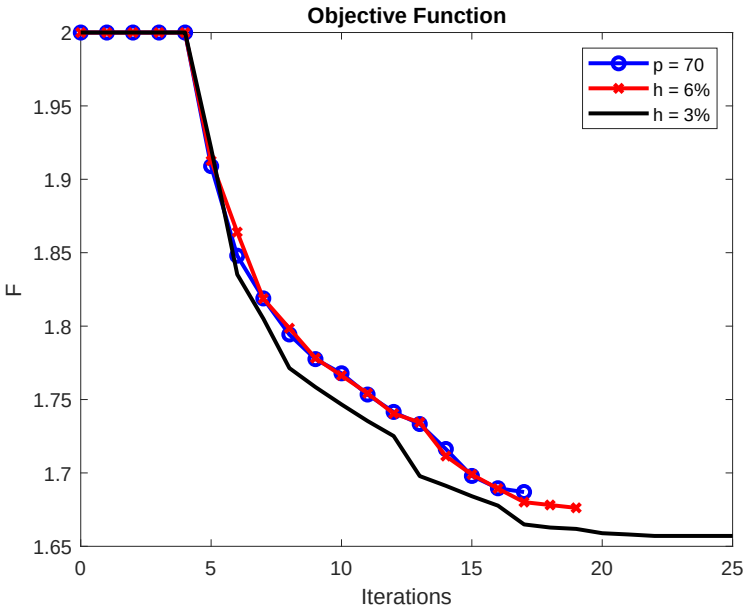

FIGURE 6. Convergence of the objective function for square domain

6. CONCLUSIONS

In this work we reviewed the p-norm and the KS regularization approaches and proposed a novel probabilistic approach. Sharpness estimates for the approaches are also provided. These estimates should be useful for practical choice of parameters.

The results of this work demonstrate that very sharp aggregation with the p-norm and the KS approaches is infeasible in the presence of closely spaced values. This occurs because very high exponents are required in this case, leading to numerical issues and blocking optimization algorithms from running. The proposed approach does not suffer from this issue, since the regularization parameter can be reduced to very small values to give very sharp regularization. For this reason, the proposed probabilistic approach is more appropriate for regularization in the presence of closely spaced extreme values.

The estimates from Eqs. (2.16) and (3.20) can also be applied for other purposes. First, these estimates can be employed to correct aggregation schemes, as proposed by Le et al. (2010). Besides, for given parameters p or h it is possible to identify the approximate number of closely spaced values (i.e., m) by checking the relative bias of the regularization scheme. This strategy may be useful to estimate the number of closely spaced values during the optimization procedure, allowing for adaptive choice of the regularization parameters during optimization.

Funding

This research was partly supported by CNPq (Brazilian Research Council) and FAPERJ (Research Foundation of the State of Rio de Janeiro). These financial support are gratefully acknowledged.

Conflict of interest

The authors declare that they have no conflict of interest.

Replication of results

The authors state that all the data necessary to replicate the results are presented in the manuscript.

References

S. Amstutz. Augmented Lagrangian for cone constrained topology optimization. *Computational Optimization and Applications*, 49:101–122, 2011.

S. Amstutz and H. Andrä. A new algorithm for topology optimization using a level-set method. *Journal of Computational Physics*, 216(2):573–588, 2006.

Gustavo Assis da Silva, Niels Aage, André Teófilo Beck, and Ole Sigmund. Local versus global stress constraint strategies in topology optimization: A comparative study. *International Journal for Numerical Methods in Engineering*, n/a(n/a):n/a, 2021. doi: 10.1002/nme.6781.

A. Dalklint, M. Wallin, and D.A. Tortorelli. Structural stability and artificial buckling modes in topology optimization. *Structural and Multidisciplinary Optimization*, 2021. doi: 10.1007/s00158-021-03012-z.

H. A. David and H. N. Nagaraja. *Order Statistics*. John Wiley & Sons, Hoboken, 3rd edition, 2003.

P. Duysinx and O. Sigmund. *New developments in handling stress constraints in optimal material distribution*. 1998. doi: 10.2514/6.1998-4906.

Eduardo Fernández, Maxime Collet, Pablo Alarcón, Simon Bauduin, and Pierre Duysinx. An aggregation strategy of maximum size constraints in density-based topology optimization. *Structural and Multidisciplinary Optimization*, 60:2113–2130, 2019. doi: s00158-019-02313-8.

Federico Ferrari and Ole Sigmund. Revisiting topology optimization with buckling constraints. *Structural and Multidisciplinary Optimization*, 59:1401–1415, 2019. doi: 10.1007/s00158-019-02253-3.

Huan-Huan Gao, Ji-Hong Zhu, Wei-Hong Zhang, and Ying Zhou. An improved adaptive constraint aggregation for integrated layout and topology optimization. *Computer Methods in Applied Mechanics and Engineering*, 289:387–408, 2015. doi: 10.1016/j.cma.2015.02.022.

Graeme J. Kennedy and Jason E. Hicken. Improved constraint-aggregation methods. *Computer Methods in Applied Mechanics and Engineering*, 289:332–354, 2015. doi: 10.1016/j.cma.2015.02.017.

G Kreisselmeier and R Steinhauser. Systematic control design by optimizing a vector performance indicator. In *Symposiumon Computer-Aided Design of Control Systems, IFAC*, 1979.

Chau Le, Julian Norato, Tyler Bruns, Christopher Ha, and Daniel Tortorelli. Stress-based topology optimization for continua. *Structural and Multidisciplinary Optimization*, 41: 605–620, 2010. doi: 10.1007/s00158-009-0440-y.

M. Loève. *Probability Theory I*. Springer-Verlag, New York, 4th edition, 1977.

A.A. Novotny and Jan Sokolowski. *An Introduction to the Topological Derivative Method*. SpringerBriefs in Mathematics. Springer International Publishing, 2020. doi: 10.1007/978-3-030-36915-6. URL `https://www.springer.com/gp/book/9783030369149`.

J. París, F. Navarrina, I. Colominas, and M. Casteleiro. Topology optimization of continuum structures with local and global stress constraints. *Structural and Multidisciplinary Optimization*, 39:419–437, 2009. doi: 10.1007/s00158-008-0336-2.

J. París, F. Navarrina, I. Colominas, and M. Casteleiro. Block aggregation of stress constraints in topology optimization of structures. *Advances in Engineering Software*, 41(3):433–441, 2010. doi: 10.1016/j.advengsoft.2009.03.006.

Nicholas M. K. Poon and Joaquim R. R. A. Martins. An adaptive approach to constraint aggregation using adjoint sensitivity analysis. *Structural and Multidisciplinary Optimization*, 34:61–73, 2007. doi: 10.1007/s00158-006-0061-7.

Leonel Quinteros, Viviana Meruane, and Eduardo Lenz Cardoso. Phononic band gap optimization in truss-like cellular structures using smooth p-norm approximations. *Structural and Multidisciplinary Optimization volume*, 64:113–124, 2021. doi: 10.1007/s00158-021-02862-x.

R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*. John Wiley & Sons, Hoboken, 2nd edition, 2008.

A. N. Shriryaev. *Probability*. Springer-Verlag, New York, 2nd edition, 1995.

A. J. Torii and J. R. de Faria. Structural optimization considering smallest magnitude eigenvalues: a smooth approximation. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 39(5):1745–1754, 2017. doi: 10.1007/s40430-016-0583-x.

A.J. Torii, R.H. Lopez, and L.F.F. Miguel. Modeling of global and local stability in optimization of truss-like structures using frame elements. *Structural and Multidisciplinary Optimization*, 51(6):1187–1198, 2015. doi: 10.1007/s00158-014-1203-y.

Alexander Verbart, Matthijs Langelaar, and Fred van Keulen. A unified aggregation and relaxation approach for stress-constrained topology optimization. *Structural and Multidisciplinary Optimization*, 55:663–679, 2017. doi: 10.1007/s00158-016-1524-0.

Cunfu Wang and Xiaoping Qian. Heaviside projection–based aggregation in stress-constrained topology optimization. *International Journal for Numerical Methods in Engineering*, 115(7):849–871, 2018. doi: 10.1002/nme.5828.

R. J. Yang and C. J. Chen. Stress-based topology optimization. *Structural Optimization*, 12:98–105, 1996. doi: 10.1007/BF01196941.

(A.J. Torii) Universidade Federal da Integração Latino-Americana (UNILA), Av. Tancredo Neves 6731, Foz do Iguaçu, 85867-970, Brazil

Email address: `andre.torii@unila.edu.br`

(J.R. de Faria) Universidade Federal da Paraíba (UFPB), Cidade Universitária, João Pessoa, PB, 58051-900, Brazil

Email address: `jairo@ci.ufpb.br`

(A.A. Novotny) Laboratório Nacional de Computação Científica LNCC/MCTI, Coordenação de Métodos Matemáticos e Computacionais, Av. Getúlio Vargas 333, 25651-075 Petrópolis - RJ, Brasil

Email address: `novotny@lncc.br`