Multiplicity of the maximum eigenvalue in structural optimization problems

Carl-Johan Thore[1]

Abstract Many problems in structural optimization can be formulated as a minimization of the maximum eigenvalue of a symmetric matrix. In practise it is often observed that the maximum eigenvalue has multiplicity greater than one close to or at optimal solutions. In this note we give a sufficient condition for this to happen at extreme points in the optimal solution set. If, as in topology optimization, each design variable determines the amount of material in a finite element in the design domain then this condition essentially amounts to saying that the number of elements containing material at a solution must be greater than the order of the matrix.

Keywords Maximum eigenvalue · Multiplicity · Structural optimization

1 Introduction

Many problems in structural optimization can be formulated as a minimization of the maximum eigenvalue of a symmetric matrix. Examples include maximization of the fundamental eigenfrequency (Seyranian et al. 1994; Du and Olhoff 2005), robust design optimization (Kobelev 1993; Takezawa et al. 2011; Brittain et al. 2012; Holmberg et al. 2015), and maximization of the fundamental buckling load (Cox and Overton 1992; Rahmatalla and Swan 2003; Stanford and Beran 2013).[1]

Let $\lambda_1(A)$ denote the largest eigenvalue of a symmetric matrix $A \in \mathbb{R}^{n \times n}$. In the present context A depends on the stiffness matrix arising from a finite element discretization of some structure. The stiffness matrix in turn depends on a set of design variables x, usually determining the amount of material present in each finite element. It is frequently observed that the multiplicity of $\lambda_1(A)$ is greater than one close to or at an optimal solution x^*, rendering the function $x \rightarrow \lambda_1(A(x))$ non-smooth (Overton and Womersley 1993). In this note we provide sufficient conditions for this to happen. The most important part of these is that the number of non-zero elements in x^* is greater than the order n of A. Since the conditions depend on the solution they cannot in general be checked *a priori*, but this is quite natural since one expects applied loads and boundary conditions to have a significant impact on the properties of the optimal solutions.

The work presented here builds on the analysis of Pataki (1998) and follows it closely in parts. Pataki (1998) considered unconstrained minimization of the maximum eigenvalue (so did Dedieu (2000) using different techniques), but for structural optimization it is necessary to (at least) introduce lower bounds on the design variables and a linear equality constraint limiting the total amount of material used. The main result here is Theorem 1 which gives a lower bound (greater than one) on the multiplicity of $\lambda_1(A(x^*))$ when A depends linearly on the vector of design variables

✉ Carl-Johan Thore
carl-johan.thore@liu.se

1 Division of Solid Mechanics, Department of Management and Engineering, Linköping University, SE 581 83, Linköping, Sweden

---

[1]To see that all of these problems can be formulated as a minimization of the maximum eigenvalue, note that $\lambda_1(A) = -\lambda_n(-A)$, where λ_1 and λ_n is the largest and, respectively, smallest eigenvalue, of the symmetric matrix A, and that, for a positive definite A, if λ_1 is the maximum eigenvalue for the generalized eigenvalue problem $(B - \lambda A)u = 0$, then it plays the same role in the eigenvalue problem $(A^{-1/2}BA^{-1/2} - \lambda I)u = 0$.

x, and x^* is an extreme point in the optimal solution set. In many problems of interest, however, A is a non-linear function of x. Fortunately, Lemma 1 shows that that such problems can be reduced to problems of the type treated by Theorem 1, and Theorem 2 then provides a lower bound on the multiplicity of $\lambda_1(A(x^*))$ in the non-linear case.

2 Notation and basic facts

Let $\mathbb{S}^n$ denote the space of $n \times n$ symmetric matrices with real entries. If a matrix in $\mathbb{S}^n$ is positive semi-definite we say that it belongs in $\mathbb{S}^n_+$. The eigenvalues of $A \in \mathbb{S}^n$ are denoted $\lambda_1 \geq \ldots \geq \lambda_n$, and we can write $A = Q\mathrm{Diag}\{\lambda_i\}Q^\mathsf{T}$ where $Q \in \mathbb{R}^{n \times n}$ is an orthonormal matrix and $\mathrm{Diag}\{\lambda_i\}$ is a diagonal matrix with $\lambda_1, \ldots, \lambda_n$ on the diagonal. The functions $\lambda_i : \mathbb{S}^n \to \mathbb{R}$ are Lipschitz continuous and convex. The multiplicity of λ_1 is the largest number p, $1 \leq p \leq n$, such that $\lambda_1 = \ldots = \lambda_p > \lambda_{p+1}$; we denote this number $\mathrm{mult}(\lambda_1)$. $\mathrm{tr}(A)$ is the trace of A.

A *face* of a convex set C is a non-empty, convex subset F of C such that if $F \ni x = \frac{1}{2}(x_1 + x_2)$ for some $x_1, x_2 \in C$, then $x_1, x_2 \in F$. A zero-dimensional face of a convex set, i.e. a point, is called an *extreme point* of that set.

We use the notation $t(n) = n(n+1)/2$ for the n-th triangular number and let $\mathbb{N}_0$ denote the set of non-negative integers.

3 The affine case

Let $A : \mathbb{R}^m \to \mathbb{S}^n$ be defined by

$$A(x) = A_0 + \sum_{i=1}^{m} x_i A_i,$$

where $A_i \in \mathbb{S}^n$ for $i = 0, \ldots, m$. Consider the optimization problem

$$\min_{x \in \mathcal{X}} \quad \lambda_1(A(x)), \tag{1}$$

where $\mathcal{X} = \{x \in \mathbb{R}^m \mid x \geq 0, \ a^\mathsf{T}x = b\}$. The vector $a \in \mathbb{R}^m$ is assumed to be such that this convex set is non-empty, compact and contains a point $x > 0$. The objective function in (1) is a composition of a convex and an affine function, hence convex. Since $\mathcal{X}$ is compact and $\lambda_1(\cdot)$ is lower semi-continuous, the set of optimal solutions to (1) is non-empty and compact (Andréasson et al. 2005, Theorem 4.7).

The results proved in this paper concern extreme points of the set of optimal solutions to (1) so it is of interest to know that such points exist. Let λ_1^* denote the optimal value of (1). The set of optimal solutions is then

$$\mathcal{O} = \left\{x \in \mathbb{R}^m \mid \lambda_1(A(x)) = \lambda_1^*\right\} \cap \mathcal{X}.$$

The first of the sets to the right is closed and convex (Pataki 1998, p. 348), so by Theorem 2.1 in (Rockafeller 1972) $\mathcal{O}$ is convex. Proposition 2.3.3 in Hiriart-Urruty and Lemarchal (1993) then ensures the existence of an extreme point in $\mathcal{O}$.

3.1 SDP-reformulation and the main theorem

There exists several different characterizations of the maximum eigenvalue of a symmetric matrix (Pataki 1998). Here we employ a version where the eigenvalue is the optimal value of a certain semi-definite program (SDP).

Let x be given. A symmetric matrix is positive semi-definite if and only if all its eigenvalues are non-negative (Horn and Johnson 1985, Theorem 7.2.1). It follows that $zI - A(x) \succeq 0 \Leftrightarrow z - \lambda_1(A(x)) \geq 0$, where I is an identity matrix. Therefore $\lambda_1(A(x))$ is the optimal value of the SDP

$$\min_{z \in \mathbb{R}, W \in \mathbb{S}^n_+} \quad z$$
$$\text{s. t.} \quad zI - W = A(x), \tag{2}$$

where W is a slack variable. Since $A(x)$ is symmetric, $A(x) = Q\mathrm{Diag}\{\lambda_i\}Q^\mathsf{T}$, with Q an orthonormal matrix of order n. Then (z^*, W^*) is an optimal solution to (2) if and only if $z^* = \lambda_1(A(x))$ and

$$W^* = Q\mathrm{Diag}\left\{0, z^* - \lambda_2, \ldots, z^* - \lambda_n\right\}Q^\mathsf{T}.$$

It follows from the last equality that

$$\mathrm{mult}(\lambda_1(A(x)) = n - \mathrm{rank}(W^*). \tag{3}$$

This implies that the lower the rank of W^*, the greater the multiplicity of $\lambda_1(A(x))$. Consequently, the proof of Theorem 1 below amounts to bounding the rank of W^*.

Now since x in (2) is arbitrary, problem (1) can be reformulated as

$$\min_{x \in \mathcal{X}} \quad \lambda_1(A(x)) = \min_{x \in \mathcal{X}, z \in \mathbb{R}, W \in \mathbb{S}^n_+} \quad z$$
$$\text{s. t.} \quad zI - W = A(x). \tag{4}$$

The key step in the proof of Theorem 1 below is the application of Theorem 2.2 in Pataki (1998), which provides an upper bound on the sum of the ranks of the matrix variables at faces of the feasible set of an SDP. Therefore we need to reformulate (4) into a problem to which the cited theorem applies. To this end, note that a scalar $x_j \geq 0$ can be seen as a symmetric, positive semi-definite matrix of order 1. This means that problem (4) can be written as

$$\min_{x_1 \in \mathbb{S}^1_+, \ldots, x_m \in \mathbb{S}^1_+, z \in \mathbb{R}, W \in \mathbb{S}^n_+} \quad z$$
$$\text{s. t.} \quad \begin{cases} zI - W = A(x) \\ a^\mathsf{T}x = b. \end{cases} \tag{5}$$

With this reformulation we can now prove the main theorem.

Theorem 1 *Let x^* be an extreme point of $\mathcal{O}$ such that $\sum_{i=1}^m r(x_i^*) > n$, where $r(x_i) = \text{rank}(x_i) \in \{0, 1\}$. Then*

$$\text{mult}(\lambda_1(A(x^*)) \geq$$
$$n - \max\left\{ w \in \mathbb{N}_0 \mid t(w) \leq t(n) - \sum_{i=1}^m r(x_i^*),\ w \leq n-2 \right\}.$$

Proof Since x^* is an extreme point of $\mathcal{O}$, it follows from Lemma 4.2 in Pataki (1998) (omit V and take $k = 1$) that (x^*, z^*, W^*), (z^*, W^*) solving (2) with $x = x^*$, is an extreme point of the feasible set of (5). Let $w = \text{rank}(W^*)$. Problem (5) has $q = 1$ unconstrained scalar variables and $t(n) + 1$ equality constraints (counting a symmetric matrix-constraint as $t(n)$ scalar constraints), and an extreme point is a face of dimension zero, so Theorem 2.2 from Pataki (1998) yields

$$\sum_{i=1}^m r(x_i^*) + t(w) \leq t(n) + 1 - q \quad \Leftrightarrow$$

$$t(w) \leq t(n) - \sum_{i=1}^m r(x_i^*). \quad (6)$$

To get a contradiction, assume that $w = n - 1$ [by (3) then, $\text{mult}(\lambda_1(A(x^*)) = 1$]. Straightforward calculations show that

$$t(n-1) = t(n) - n > t(n) - \sum_{i=1}^m r(x_i^*), \quad (7)$$

since by assumption, $\sum_{i=1}^m r(x_i^*) > n$. But now (7) contradicts (6) and it follows that $w < n - 1$; hence $w \leq n - 2$. This result together with (6) substituted in (3) then yields

$$\text{mult}\left(\lambda_1(A(x^*)) = n - w \geq\right.$$
$$n - \max\left\{ w \geq 0 \mid t(w) \leq t(n) - \sum_{i=1}^m r(x_i^*),\ w \leq n-2 \right\}.$$
$$\square$$

Remark If $x \geq 0$, $y \geq 0$ and $x + y = u$, then $0 \leq x \leq u$ and $0 \leq y \leq u$. Based on this observation we could account for an upper bound u_i on each of the x_i's in (5) by introducing additional variables $y_i \in \mathbb{S}_+^1$, $i = 1, \ldots, m$, and m linear equality constraints $x_i + y_i = u_i$. Inequality (6) in the proof above then becomes

$$t(w) \leq t(n) + m - \sum_{i=1}^m \left[r\left(x_i^*\right) + r\left(y_i^*\right) \right].$$

It holds that $r(x_i^*) + r(y_i)^* = 2$ if $0 < x_i^* < u$; 1 otherwise. Following the same arguments as in the proof one is thus led to a version of Theorem 1 where $\sum_{i=1}^m r(x_i^*) > n$ is replaced by

$$\left|\left\{ i \in \{1, \ldots, m\} \mid 0 < x_i^* < u \right\}\right| > n,$$

where $|\cdot|$ denotes the cardinality of a set. In terms of topology optimization this means that the number of "grey" elements in the solution should exceed the order of the matrix. This seems overly restrictive since researchers report coalescing of eigenvalues despite using penalization methods which lead to solutions where almost all variables attain their upper or lower bound (Holmberg et al. 2015).

Remark It is straightforward to modify the theorem to account for additional linear equality constraints – adding s such constraints leads to an additional term "$+s$" to the right in (6). Linear inequality constraints may be treated by introduction of (non-negative) slack variables. Adding one linear inequality and a slack variable s then implies that a term "$r(s)$" should be added to the left-hand side of (6) and "$+1$" to the right-hand side.

4 The non-linear case

Consider the optimization problem

$$\min_{x \in \mathcal{X}} \quad \lambda_1(A(x)), \quad (8)$$

where $x \to A(x)$ is a continuously differentiable, non-linear function. Lemma 1 below shows than any solution x^* to this problem is also a solution to a problem of type (1). Therefore it makes sense to assume that x^* is an extreme point in the optimal solution set of the latter problem. This is done in Theorem 2, the proof of which is then a direct application of Theorem 1.

Lemma 1 *Let x^* be an optimal solution to (8). Then x^* solves*

$$\min_{y \in \mathcal{X}} \quad \lambda_1\left(A^* + \sum_{i=1}^m (y_i - x_i^*) A_i^* \right), \quad (9)$$

where $A^* = A(x^*)$, and $A_i^* = \frac{\partial A}{\partial x_i}(x^*)$.

Proof We begin by noting that $\mathcal{X}$ satisfies Slater's constraint qualification (Andréasson et al. 2005, Def. 5.38), ensuring the necessity of the optimality conditions that follow.

Introduce the Lagrangian

$$L(x; \mu, \gamma) = \lambda_1(A(x)) + \mu(a^\mathsf{T} x - b) - \gamma^\mathsf{T} x.$$

The necessary first-order optimality conditions for a point $x^* \in \mathcal{X}$ in (8) is (Clarke 1983, Theorem 6.1.1) the existence of $(\mu, \gamma) \neq 0$, with $\gamma \geq 0$ such that

$$0 \in \partial L(x^*; \mu, \gamma)$$
$$\gamma_i x_i^* = 0, \quad i = 1, \ldots, m, \quad (10)$$

where $\partial L(\boldsymbol{x}^*; \mu, \boldsymbol{\gamma}) = \partial \lambda_1(\boldsymbol{A}^*) + \mu \boldsymbol{a} - \boldsymbol{\gamma}$ is the Clarke's generalized gradient of L. Let $p = \text{mult}(\lambda_1(\boldsymbol{A}^*))$. Theorem 3.9 in Overton and Womersley (1993) provides an expression for the generalized gradient $\partial \lambda_1(\boldsymbol{A}^*)$ using which the optimality conditions become the existence of $(\mu, \boldsymbol{\gamma}) \neq \boldsymbol{0}$, with $\boldsymbol{\gamma} \geq \boldsymbol{0}$, and $\boldsymbol{U} \in \mathbb{S}_+^p$ such that

$$
\begin{aligned}
0 &= \text{tr}\left(\boldsymbol{U} \boldsymbol{Q}_1^{\mathsf{T}} \boldsymbol{A}_i^* \boldsymbol{Q}_1\right) + \mu a_i - \gamma_i, & i &= 1, \dots, m \\
\gamma_i x_i^* &= 0, & i &= 1, \dots, m \quad (11) \\
\text{tr}(\boldsymbol{U}) &= 1,
\end{aligned}
$$

where the columns of $\boldsymbol{Q}_1 \in \mathbb{R}^{n \times p}$ are eigenvectors associated with $\lambda_1(\boldsymbol{A}^*)$.

Problem (9) is convex, so the generalized gradient of the associated Lagrangian coincide with the subdifferential (Clarke 1983, Proposition 2.2.7). The necessary and *sufficient*, first-order optimality conditions for (9) then follows from Theorem 28.3 in Rockafeller (1972), and are the existence of $(\mu, \boldsymbol{\gamma}) \neq \boldsymbol{0}$, with $\boldsymbol{\gamma} \geq \boldsymbol{0}$, and $\boldsymbol{U} \in \mathbb{S}_+^p$ such that

$$
\begin{aligned}
0 &= \text{tr}\left(\boldsymbol{U} \boldsymbol{Q}_1^{\mathsf{T}} \boldsymbol{A}_i^* \boldsymbol{Q}_1\right) + \mu a_i - \gamma_i, & i &= 1, \dots, m \\
\gamma_i y_i &= 0, & i &= 1, \dots, m \quad (12) \\
\text{tr}(\boldsymbol{U}) &= 1.
\end{aligned}
$$

But these conditions coincide with conditions (11). This shows that $\boldsymbol{y} = \boldsymbol{x}^*$ satisfies the optimality conditions (12) of problem (9). Since these are sufficient for optimality it follows that $\boldsymbol{x}^*$ solves (9). $\qquad \square$

Theorem 2 *Assume that $\sum_{i=1}^m r(x_i^*) > n$ and that $\boldsymbol{x}^*$ is an extreme point in the set of optimal solutions to problem* (9). *Then*

$$
\text{mult}(\lambda_1(\boldsymbol{A}^*)) \geq
$$
$$
n - \max\left\{ w \in \mathbb{N}_0 \mid t(w) \leq t(n) - \sum_{i=1}^m r(x_i^*), \ w \leq n-2 \right\}.
$$

Proof Since problem (9) is of the form (1) with optimal value $\lambda_1(\boldsymbol{A}^*)$ the assertion follows from Theorem 1. $\qquad \square$

5 Discussion

This paper has presented sufficient conditions for the maximum eigenvalue at an optimal solution to a type of problems occurring in structural optimization to have multiplicity greater than one. The conditions are given in Theorems 1 and 2, whose proofs are based on results obtained by Pataki (1998).

It is common in structural optimization to require that the design variables satisfy $x_j \geq \varepsilon$ for some small number $\varepsilon > 0$ to avoid singularity of the stiffness matrix (Christensen and Klarbring 2009). A simple change of variables $x_i \rightarrow$

$y_i + \varepsilon$ results in a problem with zero as the lower bound on the design variables to which the presented results apply.

In Holmberg et al. (2015) the author and co-workers considered minimization of the worst-case compliance for a structure subject to self-weight loadings. The problem was formulated as a non-linear SDP, which is likely the best way to treat it numerically, but it is not difficult to show that it is equivalent to

$$
\min_{\boldsymbol{x} \in \mathcal{H}} \lambda_1(\boldsymbol{H}(\boldsymbol{x})),
$$

where $\mathcal{H}$ is similar to $\mathcal{X}$, but includes upper bounds on $\boldsymbol{x}$, and $\boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{B}(\boldsymbol{x})\boldsymbol{K}(\boldsymbol{x})^{-1}\boldsymbol{B}(\boldsymbol{x})^{\mathsf{T}} \in \mathbb{S}^s$, in which $\boldsymbol{K}(\boldsymbol{x})$ is the global stiffness matrix, $\boldsymbol{B}(\boldsymbol{x})$ is a scaling matrix and $s \in \{2, 3\}$ is the number of spatial dimensions. For this problem, n in Theorem 2 equals s – a very small number – so the inequality $\sum_{i=1}^m r(x_i^*) > n$ will, in practise, always hold. Numerical experience (Holmberg et al. 2015) strongly corroborates the *suggestion* by Theorem 2 that the maximum eigenvalue at an optimal solution will have multiplicity greater than one.

6 Future work

There are at least three items that should be addressed in future work:

(i) Are numerical solutions to structural optimization problems typically extreme points in the optimal solution set? Is it possible to modify the results in this paper so that they apply also to non-extremal solutions?

(ii) Problems involving generalized eigenvalues fit into the present framework by assuming that they solve ordinary eigenvalue problems (see footnote in the Introduction). This hinges on the non-singularity of at least one of the involved matrices. This assumption is satisfied in the bulk of the relevant literature, but important exceptions exist (Achtziger and Kočvara 2007). Therefore direct treatment of generalized eigenvalues would be desirable.

(iii) The results given in this paper should be extended to account for upper bounds on the variables x_i. This is needed to analyze, for instance, three-dimensional topology optimization problems for continuum bodies. The issue was discussed in the remark following Theorem 1 above, but the obtained condition seems overly restrictive.

References

Achtziger W, Kočvara M (2007) On the maximization of the fundamental eigenvalue in topology optimization. Struct Multidiscip Optim 34(3):181–195

Andréasson N, Evgrafov A, Patriksson M (2005) An introduction to continuous optimization. Studentlitteratur

Brittain K, Silva M, Tortorelli DA (2012) Minmax topology optimization. Struct Multidiscip Optim 45(5):657–668

Christensen PW, Klarbring A (2009) An introduction to structural optimization. Springer, Berlin Heidelberg New York

Clarke FH (1983) Optimization and nonsmooth analysis. Wiley, New York

Cox SJ, Overton ML (1992) On the optimal design of columns against buckling. SIAM J Math Anal 23(2):287–325

Dedieu JP (2000) Does optimality imply ill-posedness? Some remarks about certain min-max optimization problems. Optimization: A Journal of Mathematical Programming and Operations Research 48(2):211–217

Du J, Olhoff N (2005) Topology optimization of continuum structures with respect to simple and multiple eigenfrequencies. In: 6th world congresses of structural and multidisciplinary optimization, Rio de Janeiro, Brazil

Hiriart-Urruty JB, Lemarchal C (1993) Convex analysis and minimization algorithms I. Springer, Berlin Heidelberg New York

Holmberg E, Thore C-J, Klarbring A (2015) Worst-case topology optimization with self-weight using semi-definite programming. Struct Multidiscip Optim 52(5):915–928

Horn RA, Johnson CR (1985) Matrix analysis. Cambridge University Press, Cambridge, UK

Kobelev V (1993) On a game approach to optimal structural design. Struct Optimization 6(3):194–199

Overton ML, Womersley RS (1993) Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. Math Program 62(1–3):321–357

Pataki G (1998) On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. Math Oper Res 23(2):339–358

Rahmatalla S, Swan CC (2003) Continuum topology optimization of buckling-sensitive structures. AIAA J 41(6):1180–1189

Rockafeller RT (1972) Convex Analysis. Princeton

Seyranian AP, Lund E, Olhoff N (1994) Multiple eigenvalues in structural optimization problems. Struct Optimization 8(4):207–227

Stanford B, Beran P (2013) Aerothermoelastic topology optimization with flutter and buckling metrics. Struct Multidiscip Optim 48:149–171

Takezawa A, Nii S, Kitamura M, Kogiso N (2011) Topology optimization for worst load conditions based on the eigenvalue analysis of an aggregated linear system. Comput Methods Appl Mech Eng 200(25):2268–2281