

# Topology optimization of planar shape memory alloy thermal actuators using element connectivity parameterization

M. Langelaar<sup>1,\*</sup>, G. H. Yoon<sup>2</sup>, Y. Y. Kim<sup>3</sup> and F. van Keulen<sup>1</sup>

<sup>1</sup>*Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands*

<sup>2</sup>*School of Mechanical Engineering, Hanyang University, Seoul, Republic of Korea*

<sup>3</sup>*School of Mechanical and Aerospace Engineering and National Creative Research Initiatives Center for Multiscale Design, Seoul National University, Seoul, Republic of Korea*

## SUMMARY

This paper presents the direct application of topology optimization to the design of shape memory alloy (SMA) thermal actuators. Because SMAs exhibit strongly nonlinear, temperature-dependent material behavior, designing effective multidimensional SMA actuator structures is a challenging task. We pursue the use of topology optimization to address this problem. Conventional material scaling topology optimization approaches are hampered by the complexity of the SMA constitutive behavior combined with large actuator deflections. Therefore, for topology optimization we employ the element connectivity parameterization approach, which offers improved analysis convergence and robustness, as well as an unambiguous treatment of nonlinear materials. A path-independent SMA constitutive model, aimed particularly at the NiTi R-phase transformation, is employed, allowing efficient adjoint sensitivity analysis. The effectiveness of the proposed SMA topology optimization is demonstrated by numerical examples of constrained and unconstrained formulations of actuator stroke maximization, which provide insight into the characteristics of optimal SMA actuators. Copyright © 2011 John Wiley & Sons, Ltd.

Received 24 September 2010; Revised 10 February 2011; Accepted 25 February 2011

**KEY WORDS:** shape memory alloys; NiTi; topology optimization; element connectivity parameterization; geometric nonlinearity; material nonlinearity

## 1. INTRODUCTION

### 1.1. Shape memory alloys and design challenges

Shape memory alloys (SMAs) are materials that exhibit a solid-state phase transformation under stress and/or temperature changes. The transformation is accompanied by a transformation strain, which can be used for actuation [1, 2]. Typically, SMAs offer very large strains and actuation stresses in comparison with other active materials; however, the fact that they are thermally controlled limits their dynamic performance [3, 4]. Nevertheless, they are applied in a wide variety of applications, including medical, aerospace and micro-systems [5–7].

It is expected that systematic, computer-assisted design techniques such as topology optimization will play an important role in further development of future SMA actuators, as their complexity increases. Many currently existing SMA actuators rely on a relatively simple layout, e.g. straight

\*Correspondence to: M. Langelaar, Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands.

†E-mail: M.Langelaar@tudelft.nl

wires or helical springs. For such layouts, understanding the behavior of a 1-D SMA wire or spring is sufficient for effective design [8]. However, this changes radically as soon as SMA actuators of more complex shape and topology have to be designed. As a consequence, it is necessary to understand and model the material behavior in a 2-D or 3-D setting. The complex material behavior and the fact that often several physical domains (mechanical, thermal, electrical) interact make it extremely difficult to design effective SMA actuators by intuition. These design challenges form a serious obstacle for the innovative use of the capabilities of SMA actuation in demanding and complex applications.

Systematic design optimization techniques based on computational SMA modeling promise to alleviate the above-mentioned difficulties in designing general SMA actuators. Motivated by this vision, various researchers have reported advances in the application of shape optimization to SMA problems. Among the notable contributions, we mention early work restricted to 1-D geometries and/or analytical modeling [9–11], heuristic stress-reduction-based SMA actuator improvement (e.g. [7, 12, 13]) and the use of all-purpose yet costly genetic algorithms (e.g. [14, 15]). Authors Langelaar and Van Keulen have also explored the topic of SMA shape optimization, and have reported on design-oriented SMA finite element modeling [16, 17], sensitivity analysis of a particular SMA model [18], and efficient gradient-based shape optimization of SMA devices (e.g. [19, 20]). However, even with many design variables, shape optimization is inherently restricted to the exploration of only a limited subset of all possible designs, dictated by the specified shape parameterization and initial design geometry. A more versatile and powerful approach is in contrast given by topology optimization.

### 1.2. Topology optimization

Topology optimization aims to find the optimal material distribution in a given domain to fulfill a certain objective, possibly subject to certain constraints. No input is generally required regarding the actual design geometry, as the method is capable of generating the most effective material layout with an arbitrary shape. Topology optimization was initially developed for the design of stiff light-weight geometries in linear elasticity [21], but has since been successfully applied to a wide variety of other engineering problems (see e.g. [22] for an overview). In the conventional *density-based material scaling* approach, the design is described in terms of density variables that are linked to the material properties of individual finite elements. Thus far, a topology optimization approach for SMA structures has however still been missing. The nonlinear material combined with the generally large deflections presents a difficult challenge to most conventional approaches for several reasons. First, treatment of nonlinear materials requires arbitrary choices in the density-based material scaling approach in defining the relation between material properties and density variables. The specific choices made can strongly affect the convergence behavior and final results [23]. Second, large deflections also lead to difficulties in the conventional topology optimization approach, for instance due to the excessive distortion of compliant low-density elements, leading to serious degradation of analysis convergence [24–26].

As an alternative to conventional material scaling, the *element connectivity parameterization* (ECP) approach, proposed previously by authors Yoon and Kim [25, 27], presents a new paradigm that avoids these difficulties in topology optimization involving material and geometrical nonlinearity. In the ECP scheme, design variables control the *connectivity between* the elements instead of their material properties. This is accomplished by introducing zero-length elastic links between elements. By judiciously changing the stiffness of these links, arbitrary topologies can be realized.

As a side note, on an abstract level, ECP bears some resemblance to Discontinuous Galerkin (DG) approaches, where differences in displacements of adjacent elements are accommodated within a variational formulation (see e.g. [28]). From this perspective, ECP is most similar to a local penalization-based DG approach (i.e. modulating discontinuity using link stiffnesses), implemented in a classical Galerkin finite element framework. In addition, the arrangement of elements, links and additional nodes used in ECP makes it computationally convenient and well suited for topology optimization purposes.

### 1.3. Approach

In this paper, we adopt the ECP approach in combination with a suitably simplified design-oriented SMA model, in order to solve SMA thermal actuator design problems through topology optimization. Preliminary results for this problem were reported in [29, 30]. Our approach in many ways must be regarded as a first exploratory step in this domain, and thus several simplifications have been made: without loss of generality our investigation focuses on the 2-D plane stress case to limit the required computational effort, using a particular SMA constitutive model that allows certain simplifications in the sensitivity analysis. Specifically, we focus on the R-phase transformation in NiTi, within a particular temperature range [16, 17]. We do not include dynamic aspects and impose uniform temperature distributions on the SMA material. Yet, what we want to emphasize is that even with these simplifications, the approach presented in this paper aims to provide a fruitful basis for further improvements and extensions. Moreover, in its current form this approach already may lead to novel and more effective SMA actuator designs.

### 1.4. Outline

This paper is organized as follows: first, in Section 2, a description of the applied SMA model is presented, followed by the formulation of the employed ECP approach. Section 3 subsequently introduces the considered SMA actuator optimization problem aimed at maximizing stroke, with optional requirements regarding structural stiffness. The sensitivity analysis of the response quantities in ECP setting is also presented. Section 4 discusses various aspects of the numerical implementation, including specific measures to increase analysis robustness and the design parameterization and regularization of the topology optimization problem. Subsequently, Section 5 presents several numerical examples. Next to illustrating the effectiveness of the proposed method and the effect of various parameters, these examples also provide insight into the characteristics of the considered SMA actuator design problem. At the end of Section 5, a way to interpret the obtained designs is discussed. Finally, conclusions are given in Section 6.

## 2. STRUCTURAL MODELING

### 2.1. Path-independent shape memory alloy constitutive modeling

For the investigation of topology optimization of SMA actuators, the focus is on a Ti-55.3wt%Ni alloy investigated earlier in e.g. [31]. This alloy exhibits the R-phase transformation, which is known for its relatively small hysteresis [2]. Figure 1(a) shows the stress–strain curves for this material at various temperatures. In the temperature range  $T = 328\text{--}343\text{ K}$ , the hysteresis is particularly small, which is desirable for responsive and controllable actuators. In this paper, we focus on this low-hysteresis temperature range.

Figure 1(b) shows an approximation of the stress–strain curves in the temperature range of interest, consisting of a three-section piecewise linear curve. These sections correspond to an initial austenite phase with modulus  $E_A(T)$ , a transformation section with an apparent modulus  $E_T(T)$ , and, finally, a section where the material has fully transformed into the R-phase with modulus  $E_R(T)$ . From the data in Figure 1(a) it is clear that the moduli and the transition strains ( $\varepsilon_1(T)$ ,  $\varepsilon_2(T)$ ) vary with temperature.

This typical three-section temperature-dependent piecewise linear stress–strain behavior forms the basis of the employed constitutive model. Our interest is primarily in the static solution, therefore dynamic and transient effects are not considered. Our simplified SMA model assumes isotropy, neglects the small hysteresis and incorporates the isochoric nature of the R-phase transformation. These choices result in a relatively simple temperature-dependent nonlinear elastic model, which is convenient for use in an optimization context. Given the narrow temperature range considered (15 K), conventional thermal expansion is not modeled as it is negligible in comparison with the shape memory effect. Here, we limit the discussion to the relevant aspects and main structure of the model in the 2-D plane stress case as it is used in the optimization, a full discussion of this model is presented in [17]. The phenomenological stress–strain relation for this plane stress

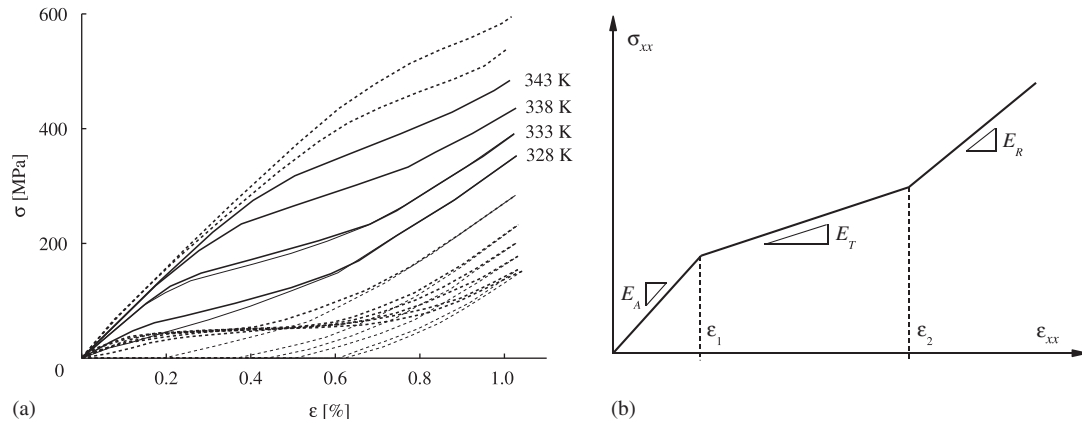


Figure 1. (a) Experimentally obtained stress–strain curves at various temperatures [31]. Thin lines are unloading curves and (b) schematic representation of a piecewise linear SMA stress–strain relation.

formulation is derived in [17] from the 3-D case as:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{Bmatrix} = \left( (1+\alpha)K \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{2G(\varepsilon_e, T)}{3} \begin{bmatrix} 2-\alpha & -(1+\alpha) & 0 \\ -(1+\alpha) & 2-\alpha & 0 \\ 0 & 0 & 3 \end{bmatrix} \right) \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{xy} \end{Bmatrix}, \quad (1)$$

where for ease of notation the term  $\alpha$  is defined as:

$$\alpha = \frac{2G(\varepsilon_e, T) - 3K}{4G(\varepsilon_e, T) + 3K}. \quad (2)$$

In Equation (1),  $K$  and  $G$  represent the bulk and shear moduli, respectively,  $\sigma$  the second Piola–Kirchhoff stress and  $\varepsilon$  the Green–Lagrange strain. To represent the isochoric R-phase transformation, the shear modulus  $G$  is modeled by a piecewise linear temperature-dependent function:

$$G(\varepsilon_e, T) = \frac{3K}{9K - A_i(\varepsilon_e, T)} \left( A_i(\varepsilon_e, T) + \frac{B_i(\varepsilon_e, T)}{\varepsilon_e} \right) \quad (i = 1, 2, 3). \quad (3)$$

Here  $A_i$  and  $B_i$  relate to material parameters with the indices referring to the sections of the piecewise linear approximation, which are given in Appendix A. Furthermore, on the basis of the distortional strain energy in a plane stress setting [17], the scalar effective strain measure  $\varepsilon_e$  is defined by:

$$\varepsilon_e^2 = \frac{4}{9}(\alpha^2 - \alpha + 1)(\varepsilon_{xx}^2 + \varepsilon_{yy}^2) + \frac{4}{9}(2\alpha^2 - 2\alpha - 1)\varepsilon_{xx}\varepsilon_{yy} + \frac{4}{3}\varepsilon_{xy}^2. \quad (4)$$

Since  $\alpha$  depends on  $G$  (see Equation (2)), the above equations for  $G$  and  $\varepsilon_e$  are solved simultaneously inside the constitutive routine using a Newton procedure. The constitutive tangent operator can be derived by differentiation of Equation (1) with respect to the strain, taking into account the strain-dependence of the shear modulus through implicit differentiation. For further details on the plane stress constitutive tangent operator, see [17].

This relatively simple SMA constitutive model gives a good representation of the experimentally observed thermomechanical shape memory behavior in the considered temperature range [17]. Moreover, because of its path-independent nature, computation of adjoint sensitivities remains relatively straightforward (cf. Section 3.2), making it a suitable model for exploration of SMA topology optimization problems.

## 2.2. ECP in nonlinear elasticity

In the element connectivity parameterization approach (ECP), finite elements are not connected to each other directly, but through 1-D zero-length springs and additional nodes, as illustrated in

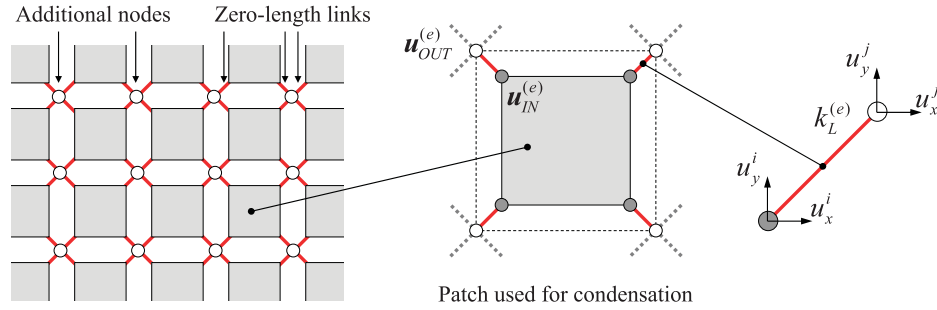


Figure 2. Mesh and patch layout used in the 2-D internal element connectivity parameterization approach. Outer and inner nodes are indicated by white resp. gray circles.

Figure 2 for the standard isoparametric quadrilateral elements employed here. This arrangement gives rise to attractive properties in topology optimization, such as a much reduced susceptibility to excessive element distortions in geometrically nonlinear problems [25], as well as advantages in problems involving material nonlinearity [32]. These properties make the ECP approach well suited for the present SMA topology optimization. In the following, the ECP formulation for the nonlinear elastic case is described, including aspects of the sensitivity analysis and the link stiffness interpolation function. We employ the *internal* ECP approach [33, 34], as opposed to the earlier *external* version. The main difference is in the link-and-elements arrangement, which in case of the internal ECP approach allows the use of local condensation that reduces the overall system size. A full validation of the outlined ECP topology optimization approach for various 2D- and 3-D linear and geometrically nonlinear topology optimization problems is presented elsewhere [33, 34].

To analyze the geometrically and materially nonlinear SMA actuators, the Total Lagrangian (TL) formulation is used (see e.g. [35]). The equilibrium equation is given by

$$\mathbf{R}(\hat{\mathbf{U}}) = \mathbf{0} \quad \text{with } \mathbf{R}(\mathbf{U}) = \text{int} \mathbf{F}(\mathbf{U}) - \text{ext} \mathbf{F}, \quad (5)$$

where  $\mathbf{R}$  is the residual force vector, given by the difference between the internal and external force vectors, and  $\hat{\mathbf{U}}$  is introduced to explicitly denote an equilibrium displacement vector. Cases with displacement-dependent external forces are not considered in this paper. Newton iterations are used to solve this nonlinear system of equations, resulting in

$$\mathbf{K}(\mathbf{U}^k) \Delta \mathbf{U}^{k+1} = -\mathbf{R}(\mathbf{U}^k), \quad \mathbf{U}^{k+1} = \mathbf{U}^k + \Delta \mathbf{U}^{k+1}, \quad (6)$$

where the system tangent stiffness matrix  $\mathbf{K}$  is given by the Jacobian  $\partial \mathbf{R} / \partial \mathbf{U}$ , and the superscript denotes the iteration number, omitted from here onward.

Subsequently, we describe how the internal ECP approach is combined with the above TL formulation. The employed link-and-elements arrangement is depicted in Figure 2. A single element together with its associated links and additional nodes is referred to as a *patch*. An individual zero-length link element in the considered 2-D setting, connecting nodes  $i$  and  $j$ , is governed by

$$k_L^{(e)} \begin{bmatrix} \mathbf{I}_2 & -\mathbf{I}_2 \\ -\mathbf{I}_2 & \mathbf{I}_2 \end{bmatrix} \{u_x^i \ u_y^i \ u_x^j \ u_y^j\}^T = \{F_x^i \ F_y^i \ F_x^j \ F_y^j\}^T, \quad (7)$$

where the nodal displacement components  $u_\alpha^\beta$ , where  $\alpha = x$  or  $y$ ,  $\beta = i$  or  $j$ , are indicated in Figure 2.  $F_\alpha^\beta$  represent the components of the associated force vector,  $\mathbf{I}_2$  is the  $2 \times 2$  identity matrix and  $k_L^{(e)}$  the link stiffness associated with the particular  $e^{\text{th}}$  patch this link belongs to. The contribution of an individual patch to Equation (6) is given by

$$\begin{bmatrix} k_L^{(e)} \mathbf{I}_8 & -k_L^{(e)} \mathbf{I}_8 \\ -k_L^{(e)} \mathbf{I}_8 & k_L^{(e)} \mathbf{I}_8 + \mathbf{K}_E^{(e)}(\mathbf{u}_{\text{IN}}^{(e)}) \end{bmatrix} \begin{Bmatrix} \Delta \mathbf{u}_{\text{OUT}}^{(e)} \\ \Delta \mathbf{u}_{\text{IN}}^{(e)} \end{Bmatrix} = - \begin{Bmatrix} \text{int} \mathbf{F}_{\text{OUT}}^{(e)} - \text{ext} \mathbf{F}_{\text{OUT}}^{(e)} \\ \text{int} \mathbf{F}_{\text{IN}}^{(e)} - \mathbf{0} \end{Bmatrix}, \quad (8)$$

where  $\mathbf{K}_E^{(e)}$  is the tangent stiffness matrix of the internal element, and  $\text{ext}\mathbf{F}_{\text{OUT}}^{(e)}$  represents the force applied to the outer nodes of this patch. By convention, no external forces are applied to inner nodes. The displacement components  $\mathbf{u}^{(e)}$  of the patch are partitioned into components associated with the outer and inner nodes, as indicated by the respective subscripts. The right-hand side vector contains the residual contributions of the outer and inner nodes, with

$$\text{int}\mathbf{F}_{\text{OUT}}^{(e)} = k_L^{(e)} (\mathbf{u}_{\text{OUT}}^{(e)} - \mathbf{u}_{\text{IN}}^{(e)}), \quad (9)$$

$$\text{int}\mathbf{F}_{\text{IN}}^{(e)} = \text{int}\mathbf{F}_E^{(e)}(\mathbf{u}_{\text{IN}}^{(e)}) - k_L^{(e)} (\mathbf{u}_{\text{OUT}}^{(e)} - \mathbf{u}_{\text{IN}}^{(e)}), \quad (10)$$

where  $\text{int}\mathbf{F}_E^{(e)}$  is the internal force vector of the  $e^{\text{th}}$  internal element.

In order to reduce the number of degrees of freedom (DoFs) at system level, in each iteration static condensation is applied to each patch contribution (Equation (8)), and only the DoFs of the outer nodes  $\mathbf{u}_{\text{OUT}}$  are used at system level. The condensed stiffness matrix  $\mathbf{K}_C^{(e)}$  and residual  $\mathbf{R}_C^{(e)}$  for the  $e^{\text{th}}$  patch are:

$$\mathbf{K}_C^{(e)} = k_L^{(e)} \mathbf{I} - (k_L^{(e)})^2 \left[ k_L^{(e)} \mathbf{I} + \mathbf{K}_E^{(e)} \right]^{-1}, \quad (11)$$

$$\mathbf{R}_C^{(e)} = \text{int}\mathbf{F}_{\text{OUT}}^{(e)} - \text{ext}\mathbf{F}_{\text{OUT}}^{(e)} + k_L^{(e)} \left[ k_L^{(e)} \mathbf{I} + \mathbf{K}_E^{(e)} \right]^{-1} \text{int}\mathbf{F}_{\text{IN}}^{(e)}. \quad (12)$$

Here and in the following, the dimension subscript of the identity matrix  $\mathbf{I}$  has been omitted. After assembling  $\mathbf{K}_C^{(e)}$  and  $\mathbf{R}_C^{(e)}$  into their global equivalents  $\mathbf{K}_C^{\text{sys}}$  and  $\mathbf{R}_C^{\text{sys}}$ , and solving the global condensed system  $\mathbf{K}_C^{\text{sys}} \Delta \mathbf{u}_{\text{OUT}} = -\mathbf{R}_C^{\text{sys}}$ , the inner displacement increments are obtained for each patch from the outer increments by:

$$\Delta \mathbf{u}_{\text{IN}}^{(e)} = \left[ k_L^{(e)} \mathbf{I} + \mathbf{K}_E^{(e)} \right]^{-1} \left\{ k_L^{(e)} \Delta \mathbf{u}_{\text{OUT}}^{(e)} - \text{int}\mathbf{F}_{\text{IN}}^{(e)} \right\}. \quad (13)$$

The assembled condensed system has exactly the same number of DoFs as the original mesh without additional nodes and zero-length links were added. In ECP implementations, the inclusion of link elements and additional nodes is therefore accounted for by modifications to the element stiffness and internal force vector routines, instead of actually constructing a modified mesh.

Extensive numerical benchmark tests have been carried out similar to those in [25], extended with SMA material modeling, to compare the ECP approach with more conventional stiffness scaling approaches (e.g. SIMP). Especially at larger deformations differences were striking, with ECP converging smoothly and quickly in all cases, while stiffness scaling resulted in either many more iterations (factor 4 on average) or in failure to converge (half of the studied cases). Space limitations prevent a more detailed discussion, but findings clearly confirmed the ECP advantages reported by previous studies and established their suitability for the present problem. Still, in certain cases during topology optimization, we have also experienced that the ECP approach at some point reached its limits. Section 4 discusses various measures we have applied to increase analysis robustness further to a level adequate for topology optimization.

### 2.3. ECP link stiffness interpolation

Performing ECP-based topology optimization directly using the patch link stiffnesses  $k_L$  as design variables often results in problematic convergence of the optimization process, caused by the strongly nonlinear relation between effective patch stiffness and  $k_L$  (see Figure 3(a), note the log scale). To overcome this problem, we introduce for every patch a new design variable  $\gamma$ , which determines the patch link stiffness  $k_L$  of all the links in the patch. The relation between design variables  $\gamma$  and link stiffnesses within a patch is chosen such that the resulting effective stiffness of the condensed patch  $\|\mathbf{K}_C\|$  varies smoothly with  $\gamma$ .<sup>‡</sup> The following link stiffness interpolation

<sup>‡</sup>Here the conventional matrix norm is used, which corresponds to the largest eigenvalue. Similar results are obtained using other matrix norms [34].

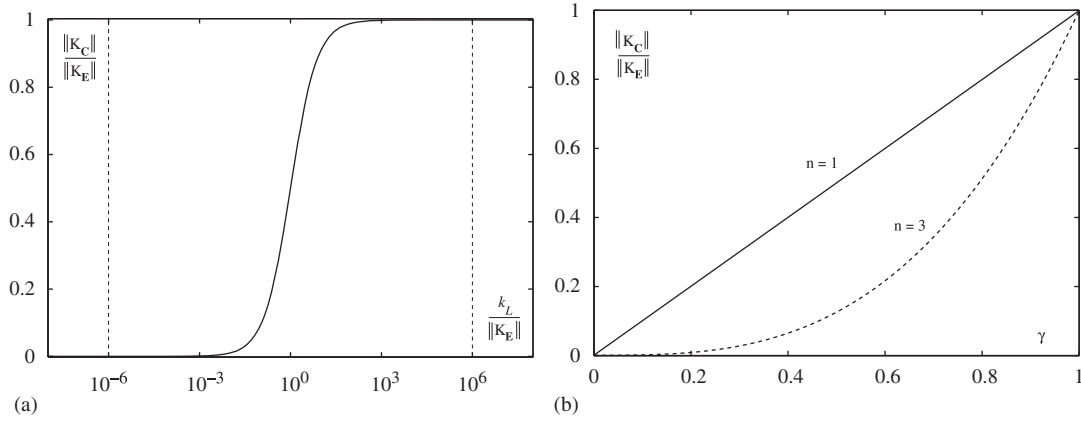


Figure 3. (a) Normalized condensed patch stiffness as a function of the normalized link stiffness (log scale) and (b) normalized condensed patch stiffness as a function of design variable  $\gamma$ , using the  $k_L(\gamma)$ -relation of Equation (14).

function  $k_L(\gamma)$ , derived in [30, 34], is used to transform the nonlinear  $\|\mathbf{K}_C\|(k_L)$ -relation into the desired linear  $\|\mathbf{K}_C\|(\gamma)$ -relation

$$k_L(\gamma) = \frac{\frac{\|\mathbf{K}_E\|}{1}}{\frac{1}{\delta + \gamma^n(1-2\delta)} - 1}, \quad 0 \leq \gamma \leq 1, \quad (14)$$

and its derivative as needed for sensitivity expressions reads as:

$$\frac{dk_L}{d\gamma} = k_L \frac{n\gamma^{n-1}(1-2\delta)}{(\delta + \gamma^n(1-2\delta))(1 - \delta - \gamma^n(1-2\delta))}. \quad (15)$$

Here the tolerance  $\delta$  controls lowest and highest link stiffness values, balancing stiffness errors between the void and solid case. The lowest normalized effective patch stiffness (void) will be  $\delta$ , the highest will equal  $1 - \delta$  (solid) [30]. A value of  $\delta = 10^{-6}$  was found to be adequate. The variable  $\gamma$  plays the same role as  $\rho$  in the density-based material scaling approach. An exponent  $n$  is applied to  $\gamma$  to introduce penalization, as illustrated in Figure 3(b), similar to the exponent used in SIMP [22]. In fact  $\gamma$  is used as a density indicator, and we refer to it as a density in the following, but continue to use the symbol  $\gamma$  instead of  $\rho$  to indicate the conceptual difference. In the case studies presented in this paper, a penalization exponent  $n=4$  has been applied to improve the convergence of the optimization process. The element stiffness matrix norm  $\|\mathbf{K}_E\|$  is evaluated only once at the start of the optimization process, in the undeformed configuration, to define an appropriate, fixed  $k_L(\gamma)$ -relation.

### 3. OPTIMIZATION PROBLEM AND SENSITIVITY ANALYSIS

#### 3.1. SMA actuator stroke maximization

This paper demonstrates ECP-based topology optimization for SMA structures on a number of 2-D plane stress examples of actuator stroke maximization under a constant external load  $^{\text{ext}}\mathbf{F}$ , with the SMA material being cycled between two temperature states  $T_i(\mathbf{x})$ . A generic problem setup is depicted in Figure 4, also involving displacement boundary conditions, output displacements  $^{\text{out}}\mathbf{U}$  and optional additional springs  $k$ . This generic problem can be used to model a wide variety of situations of practical interest.

The applied constant load  $^{\text{ext}}\mathbf{F}$  provides a prestress necessary to exploit the SMA material properties for actuation. Since at lower temperatures NiTi is more compliant than at higher temperatures (see Figure 1), a constant load will result in a larger deformation of the structure at the lower

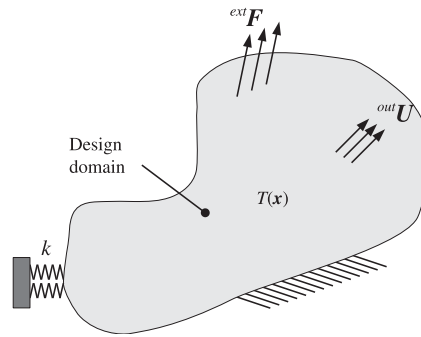


Figure 4. General problem setup used in the SMA thermal actuator topology optimization studies.

temperature  $T_1$  than at the higher temperature  $T_2$ . Optionally, the load  $\mathbf{F}^{\text{ext}}$  is also used to define a stiffness measure  $g$  in the final deformed configuration, based on the incremental compliance of the deformed structure with respect to this load, related to the measure introduced by Ramm *et al.* [36, 37]:

$$g = \mathbf{F}^{\text{ext}T} \mathbf{D} \quad \text{with} \quad \mathbf{K}(\hat{\mathbf{U}}) \mathbf{D} = \mathbf{F}^{\text{ext}}. \quad (16)$$

Here  $\mathbf{D}$  represents an additional displacement vector, used exclusively to define  $g$ . This quantity  $g = \mathbf{F}^{\text{ext}T} \mathbf{D}$  can be used to impose a minimum stiffness requirement on the actuator in its deformed state, thus ensuring well-defined positioning and avoiding overly compliant designs. Evaluation of  $g$  involves solving an additional linear problem using the (already decomposed) tangent stiffness matrix  $\mathbf{K}$  at the (fixed) final equilibrium configuration  $\hat{\mathbf{U}}$ , in order to obtain  $\mathbf{D}$ . Hereinafter, following Ramm *et al.* [36, 37], we continue to use the term ‘end stiffness’ for  $g$  (although in fact it represents a compliance-like quantity), to avoid confusion with the ‘end compliance’ defined by Buhl *et al.* [38]. We prefer this  $g = \mathbf{F}^{\text{ext}T} \mathbf{D}$  end stiffness criterion over, e.g. the end compliance or other possible stiffness definitions the nonlinear setting [37], since the characteristics of the deformed configuration under full load are of primary interest for the considered actuation.

Our aim is to find an SMA actuator design in terms of spatial density distribution  $\gamma(\mathbf{x})$  having the largest possible stroke  $s$  due to temperature variation, optionally with constraints regarding its end stiffness, as expressed by the following problem statement:

$$\begin{aligned} \max_{\gamma} \quad & s(\mathbf{U}^{\text{out}}(T_1, \gamma), \mathbf{U}^{\text{out}}(T_2, \gamma)) \\ \text{s.t.} \quad & g(T_l, \gamma) \leq g_{\max} \quad l = 1, 2. \quad (\text{optional}) \end{aligned} \quad (17)$$

The specific definition of the stroke  $s$  in case of multiple output displacement components is addressed in Section 5.1. For simplicity, the imposed temperature field is uniform and design-independent ( $T(\mathbf{x}) = T_l$ ); no thermal analysis is performed in the present study. Furthermore, we limit the discussion to two extreme temperature cases ( $l = 1, 2$ ), yet more complex design requirements can be readily composed of the building blocks presented in this paper. Note the absence of a volume constraint in Equation (17): unlike many other topology optimization problems, the SMA actuator stroke optimization turns out to have the property of selecting an optimal amount of material automatically. This is understood from the fact that both full void and full material designs do not yield effective actuators.

### 3.2. Sensitivity analysis

To enable computationally tractable topology optimization for our SMA actuator problem involving geometrically and material nonlinearities and thousands of design variables, adjoint sensitivity analysis is essential. Fortunately, because of the path-independent nature of the applied SMA model, the problem can be treated as a general nonlinear elastic one. This allows adjoint sensitivities to be computed by computationally relatively inexpensive operations at the final configuration only,

after the incremental-iterative solution process has converged. Far more involved back-integration, as required in the path-dependent case [39], is thus avoided.

Furthermore, because the ECP design variables do not affect the SMA material model but only the stiffnesses of the linearly elastic links between the SMA finite elements, the constitutive model does not have to be design-differentiated. This is a considerable advantage, since the sensitivity analysis of this relatively simple SMA model is already rather involved due to the lack of an explicit stress–strain relationship, as illustrated in [18]. Besides the already discussed robustness advantages, this aspect forms another reason to favor ECP for the present problem. The remainder of this section separately treats the sensitivity analysis of displacement-based responses (such as the stroke objective) and the end stiffness sensitivity, both in the context of ECP.

**3.2.1. Displacement-based responses.** In the considered topology optimization problems, a displacement-based response  $f$  can generally be written as:

$$f = \mathbf{L}^T \mathbf{U}. \quad (18)$$

Here  $\mathbf{L}$  is a constant vector that is used to select the displacement components of interest, and  $\mathbf{U}$  is the total displacement vector (the distinction between inner and outer displacements, and the effect of condensation is discussed later). Adjoint sensitivities of  $f$  can be derived by including the governing equations given in Equation (5) in the following Lagrangian:

$$\mathcal{L}^f = \mathbf{L}^T \mathbf{U} + \boldsymbol{\lambda}^T (\text{int} \mathbf{F}(\mathbf{U}, \gamma) - \text{ext} \mathbf{F}). \quad (19)$$

Here, a multiplier vector  $\boldsymbol{\lambda}$  has been introduced. The applied load  $\text{ext} \mathbf{F}$  is assumed to be configuration- and design-independent. Design differentiation of  $\mathcal{L}^f$  with respect to a design variable  $\gamma$  yields, in index notation:

$$\mathcal{L}_{,\gamma}^f = L_i U_{i,\gamma} + \lambda_i (\text{int} F_{i,\gamma} + \text{int} F_{i,U_j} U_{j,\gamma}) = (L_i + \lambda_j K_{ji}) U_{i,\gamma} + \lambda_i \text{int} F_{i,\gamma}. \quad (20)$$

Here  $\square_{,\gamma}$  denotes design differentiation, and the definition of the system tangent stiffness matrix  $\mathbf{K}$  introduced in Equation (6) has been used. Multiplier vector  $\boldsymbol{\lambda}$  is chosen such that the term with the state sensitivity  $U_{i,\gamma}$  in Equation (20) vanishes:

$$\mathbf{L}^T + \boldsymbol{\lambda}^T \mathbf{K} = \mathbf{0} \Rightarrow \mathbf{K} \boldsymbol{\lambda} = -\mathbf{L} \Rightarrow \boldsymbol{\lambda} = -\mathbf{K}^{-1} \mathbf{L}. \quad (21)$$

Symmetry of  $\mathbf{K}$  has been used here. Sensitivities are computed by subsequent substitution of  $\boldsymbol{\lambda}$  into Equation (20):

$$\frac{\partial \mathcal{L}^f}{\partial \gamma} = \frac{\partial f}{\partial \gamma} = \boldsymbol{\lambda}^T \frac{\partial \text{int} \mathbf{F}}{\partial \gamma}. \quad (22)$$

In the ECP implementation, condensation is applied for the adjoint sensitivity equations analogously as to the equilibrium equations, in order to reuse the decomposed system matrix based on the outer displacements  $\mathbf{K}_C^{\text{sys}}$ , when solving for  $\boldsymbol{\lambda}_{\text{OUT}}$ . This gives

$$\mathbf{K}_C^{\text{sys}} \boldsymbol{\lambda}_{\text{OUT}} = -\mathbf{L}_{\text{OUT}} - \sum_{e=1}^N \left( k_L^{(e)} \left[ k_L^{(e)} \mathbf{I} + \mathbf{K}_E^{(e)} \right]^{-1} \mathbf{L}_{\text{IN}}^{(e)} \right), \quad (23)$$

where the summation represents the assembly operator and  $N$  denotes the number of patches. Given the outer multipliers, the inner multipliers can subsequently be found by applying the equivalent of Equation (13) to each patch:

$$\boldsymbol{\lambda}_{\text{IN}}^{(e)} = \left[ k_L^{(e)} \mathbf{I} + \mathbf{K}_E^{(e)} \right]^{-1} \left\{ k_L^{(e)} \boldsymbol{\lambda}_{\text{OUT}}^{(e)} - \mathbf{L}_{\text{IN}}^{(e)} \right\}. \quad (24)$$

From Equations (9) and (10) it follows that the contribution of a particular patch to the partial derivatives of the internal forces used in Equation (22) is given by

$$\frac{\partial}{\partial \gamma} \begin{Bmatrix} \text{int} \mathbf{F}_{\text{OUT}}^{(e)} \\ \text{int} \mathbf{F}_{\text{IN}}^{(e)} \end{Bmatrix} = \frac{dk_L}{d\gamma} \frac{\partial}{\partial k_L} \begin{Bmatrix} \text{int} \mathbf{F}_{\text{OUT}}^{(e)} \\ \text{int} \mathbf{F}_{\text{IN}}^{(e)} \end{Bmatrix} = \frac{dk_L}{d\gamma} \begin{Bmatrix} (\mathbf{u}_{\text{OUT}}^{(e)} - \mathbf{u}_{\text{IN}}^{(e)}) \\ -(\mathbf{u}_{\text{OUT}}^{(e)} - \mathbf{u}_{\text{IN}}^{(e)}) \end{Bmatrix}, \quad (25)$$

where also the relation between the link stiffness and the design variable  $\gamma$  has been accounted for, using Equation (15). Combining the above with Equation (22) subsequently yields:

$$\frac{\partial \mathcal{L}^f}{\partial \gamma} = \frac{\partial f}{\partial \gamma} = \frac{dk_L}{d\gamma} (\boldsymbol{\lambda}_{\text{OUT}}^{(e)} - \boldsymbol{\lambda}_{\text{IN}}^{(e)})^T (\mathbf{u}_{\text{OUT}}^{(e)} - \mathbf{u}_{\text{IN}}^{(e)}). \quad (26)$$

Each design variable  $\gamma$  affects only the links within a single patch, allowing sensitivities to be evaluated on patch-level using local quantities.

**3.2.2. End-stiffness constraint sensitivity.** For the end-stiffness measure  $g$  defined in Equation (16), the adjoint sensitivity can be obtained by considering the following Lagrangian (see also [36, 37]):

$$\mathcal{L}^g = \text{ext} \mathbf{F}^T \mathbf{D} + \boldsymbol{\lambda}^T (\text{int} \mathbf{F}(\mathbf{U}, \gamma) - \text{ext} \mathbf{F}) + \boldsymbol{\mu}^T (\mathbf{K}(\mathbf{U}) \mathbf{D} - \text{ext} \mathbf{F}). \quad (27)$$

Here  $\boldsymbol{\lambda}$  and  $\boldsymbol{\mu}$  are two independent multiplier vectors related to the two governing equations involved (computation of  $\mathbf{U}$  and  $\mathbf{D}$ , respectively). Design differentiation of  $\mathcal{L}^g$  to a design variable  $\gamma$  yields, in index notation:

$$\begin{aligned} \mathcal{L}_{,\gamma}^g &= \text{ext} F_i D_{i,\gamma} + \mu_i (K_{ij,\gamma} D_j + K_{ij,U_k} U_{k,\gamma} D_j + K_{ij} D_{j,\gamma}) + \lambda_i (\text{int} F_{i,\gamma} + K_{ik} U_{k,\gamma}) \\ &= (\text{ext} F_i + K_{ij} \mu_j) D_{i,\gamma} + (\mu_i K_{ij,U_k} D_j + \lambda_i K_{ik}) U_{k,\gamma} + \mu_i K_{ij,\gamma} D_j + \lambda_i \text{int} F_{i,\gamma}. \end{aligned} \quad (28)$$

By choosing  $\boldsymbol{\mu} = -\mathbf{K}^{-1} \text{ext} \mathbf{F} = -\mathbf{D}$ , the term with design derivatives  $D_{i,\gamma}$  vanishes (symmetry of  $\mathbf{K}$  has been used). Subsequently, when choosing the second multiplier vector  $\boldsymbol{\lambda}$  to satisfy

$$K_{ki} \lambda_i = -\mu_i K_{ij,U_k} D_j = D_i K_{ij,U_k} D_j \quad (29)$$

also the derivatives of  $\mathbf{U}$  vanish, leaving the final sensitivity expression for  $g$ :

$$\frac{d\mathcal{L}^g}{d\gamma} = \frac{dg}{d\gamma} = -D_i K_{ij,\gamma} D_j + \lambda_i \text{int} F_{i,\gamma}. \quad (30)$$

To evaluate the right-hand side of Equation (29), in the ECP setting the term  $K_{ij,U_k}$  is computed by finite differences on the inner element level. Thus, no specific information of the inner element is needed or used. Note that for all ECP springs these terms are zero, as their stiffness is constant and independent of displacement.

The ECP implementation of the second term of the right-hand side of Equation (30) is given by Equation (26). For the first term,  $-D_i K_{ij,\gamma} D_j$ , design differentiation of Equation (8) shows that for an ECP patch the term  $K_{ij,\gamma}$  partitioned in inner and outer displacements is given by:

$$K_{ij,\gamma} = \frac{\partial k_L}{\partial \gamma} \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix}. \quad (31)$$

With this, the first term of Equation (30) can be evaluated on the ECP patch level by:

$$-D_i K_{ij,\gamma} D_j = -\frac{\partial k_L}{\partial \gamma} (D_{\text{OUT}}^e - D_{\text{IN}}^e)^T (D_{\text{OUT}}^e - D_{\text{IN}}^e) = -\frac{\partial k_L}{\partial \gamma} \|D_{\text{OUT}}^e - D_{\text{IN}}^e\|^2. \quad (32)$$

Thus, the sensitivity of the end stiffness measure  $g$  in the ECP setting with respect to an individual patch design variable  $\gamma$  reads as:

$$\frac{\partial g}{\partial \gamma} = \frac{\partial k_L}{\partial \gamma} (-\|D_{\text{OUT}}^e - D_{\text{IN}}^e\|^2 + (\lambda_{\text{OUT}}^{(e)} - \lambda_{\text{IN}}^{(e)})^T (\mathbf{u}_{\text{OUT}}^{(e)} - \mathbf{u}_{\text{IN}}^{(e)})), \quad (33)$$

which also can be evaluated on patch level.

## 4. NUMERICAL IMPLEMENTATION

### 4.1. Finite element solution approach

In spite of the superior robustness of ECP illustrated in various benchmark tests, we have experienced that during topology optimization of the considered SMA structures occasionally also in the case of the ECP formulation convergence difficulties occur. Since analysis failure means the premature end of the optimization process, analysis robustness is crucial. The nonlinear elastic problems are solved by an adaptive load-controlled incremental-iterative scheme as proposed in [40]:

$$\Delta \lambda^{k+1} = \Delta \lambda^k \sqrt{N_{\text{Desired}} / N_{\text{Actual}}}. \quad (34)$$

Starting with a given initial increment  $\Delta \lambda^0$  and a desired number of iterations  $N_{\text{Desired}}$ , this update rule sets new increments depending on the number of actually realized iterations  $N_{\text{Actual}}$ . The adaptivity increases the applicability of this scheme to the wide variety of designs that is generated during the optimization process, and also is able to increase efficiency when convergence is smooth. A 20% initial load factor increment and 10 desired iterations per increment are used. Convergence is reached when the normalized residual norm  $\|\mathbf{R}\| / \|\mathbf{F}^{\text{ext}}\|$  falls below  $10^{-4}$ , which provides a balance between accuracy required for topology optimization purposes and numerical efficiency. In addition, to improve robustness, when the solution process does not converge in a given increment within the maximum number of 25 iterations, a restart is automatically performed from the solution obtained at the preceding increment, with an increment size reduced by 50%.

To improve efficiency, use is made of the fact that the displacement fields of subsequent designs in the topology optimization process generally only differ slightly, given the gradual design changes. Thus, the final displacement of design  $i$  forms a good starting point for iterative solution of design  $i + 1$ . For each new analysis, this continued solution process is attempted first, and converges quickly for the majority of designs. If no convergence is reached within 25 iterations, the regular incremental-iterative procedure is attempted, starting from the undeformed configuration.

### 4.2. Selective adaptive SMA linearization

This section describes a material-model-specific adaptation that has proven quite useful in increasing both the efficiency and robustness of the considered SMA finite element analysis. Its effectiveness is probably related to the specific combination of the SMA stress–strain curve at small strains (decreasing stiffness) combined with the ECP formulation.

The first Newton iteration of an increment tends to overshoot the solution, often causing low density elements (which should hardly deform in the ECP setting) to enter the SMA transformation regime (section 2 in Figure 1(b)). A model with many low-density, highly compliant patches in this computationally expensive nonlinear range of the material behavior needs numerous iterations (and may even fail to converge), only to result in a final solution where all low-density elements have again returned to the initial linear part of the SMA stress–strain curve (Section 1 in Figure 1(b)). To avoid this inefficiency, we apply a selective adaptive linearization of the SMA material model in combination with the continued as well as regular solution process discussed in Section 4.1. First, the model is evaluated in the deformed configuration of the previous design or increment,

and based on this, for each element a phase fraction  $\phi_i$  is computed, defined as an indicator of the state of the material between the first and second transition points of the SMA stress–strain curve:

$$\phi_i = \int_{\Omega_i} \phi d\Omega \quad \text{with } \phi = \min \left( 1, \max \left( 0, \frac{\varepsilon_e - \varepsilon_1}{\Delta} \right) \right). \quad (35)$$

Here  $\Omega_i$  is the element domain, and  $\varepsilon_e$ ,  $\varepsilon_1$  and  $\Delta$  refer to the material model described in Section 2.1 and Appendix A. When the element phase fraction  $\phi_i$  is less than  $10^{-3}$ , the SMA material model for that element is temporarily replaced by a linear elastic model corresponding to the first section of the stress–strain curve, e.g.  $E = E_A(T)$  and  $\nu = 1/3$ . After convergence of this modified problem, all original element material models are restored and iterations are continued till convergence of the actual problem, so the correct solution is obtained. Generally, virtually all linearized elements remain in the linear range and the second solution process converges very quickly.

#### 4.3. Inner iterations

An additional ECP-specific technique that helps to converge in difficult cases is the use of inner iterations. This refers to performing Newton iterations on the inner DoFs of an ECP patch only, keeping the outer displacements temporarily fixed. Since the inner element has substantial freedom to orient itself and rotate within the ECP patch (cf. Figure 2), particularly when the link stiffnesses are low, under adverse circumstances it can occur that convergence to an equilibrium position becomes slow or difficult only within a single patch. This situation can be detected by monitoring the residual forces for relatively large contributions concentrated in only one or a few elements. Instead of performing numerous costly iterations on the full model, it is more efficient to iterate within the unbalanced patches themselves. From the second row of Equation (8), it follows that considering the outer nodes as fixed results in the following iterative scheme:

$$[k_L \mathbf{I} + \mathbf{K}_E(\mathbf{u}_{\text{IN}}^k)] \Delta \mathbf{u}_{\text{IN}}^{k+1} = k_L \Delta \mathbf{u}_{\text{OUT}} - \text{int} \mathbf{F}_{\text{IN}}(\mathbf{u}_{\text{IN}}^k), \quad (36)$$

$$\mathbf{u}_{\text{IN}}^{k+1} = \mathbf{u}_{\text{IN}}^k + \Delta \mathbf{u}_{\text{IN}}^{k+1}. \quad (37)$$

Here the element superscripts  $\square^{(e)}$  used in Equation (8) have been omitted for clarity, but it is emphasized that these relations refer to inner iterations for an individual patch. The low cost of the computations involving a single element also allows the use of a line-search procedure, to further improve robustness and convergence in difficult situations. Inner iterations are invoked when global convergence stalls, and when patches are detected that account for more than 20% of the total residual. After a number of inner iterations on the unbalanced patches, the normal global solution process is resumed.

Isolated single-patch unbalances have been responsible for analysis terminations in some of our numerical experiments, as they sporadically emerge in medium- to low-density elements with large inner element rotations in combination with severe compression. The strongly nonlinear and anisotropically weakening–stiffening SMA constitutive model may be an additional detrimental factor obstructing convergence. These proposed single-patch inner Newton iterations help to traverse such difficult points during the analysis.

#### 4.4. Design parametrization

To promote manufacturable results and to suppress numerical artifacts, a particular design parameterization has been applied. For historic reasons, we use design variables  $\alpha^{(i)}$  located at all additional ECP nodes—in principle these design variables could be located anywhere in the design domain. Their values and (undeformed) positions form the basis for a consistent density filtering scheme similar to that proposed by Bruns and Tortorelli [41], through which patch densities  $\gamma^{(j)}(\alpha)$  are obtained based at the element (patch) centroids, which in turn define the patch link stiffness  $k_L^{(j)}(\gamma^{(j)})$  (Equation (14)). This filtering serves to impose a minimum lengthscale on designs as well as to suppress undesired patterns in the topology (e.g. checkerboards). We apply a linear

radial basis function as weighting function, with filter radius  $r$ , as was also used in by Bruns and Tortorelli [41]. We refer to their original paper for further formulation details.

A well-known consequence of the applied linear density filtering is that the structural boundary smoothly varies from solid to void and thus shows intermediate densities, which could be considered undesirable from a manufacturing perspective. Convergence to pure black–white designs could possibly be achieved through the use of, e.g. the Heaviside projection approach by Guest *et al.* [42], but this requires a costly continuation approach. Moreover, given the limited accuracy obtained from computations using the relatively coarse mesh used during topology optimization design analysis, other forms of post-processing are better suited to arrive at realistic and safe production-ready designs (e.g. [19, 20]). We consider such post-processing outside the scope of the present investigation, and present only the designs as they have directly been obtained from the topology optimization procedure, showing the density-field  $\gamma$  in grayscale. Furthermore, to obtain consistently smoothed structural boundaries both within and at the border of the design domain, a ghost layer of zero-density design variables has been applied at the unsupported design domain borders.

## 5. NUMERICAL EXAMPLES AND DISCUSSION

This section presents numerical examples of topology optimization of SMA thermal actuators for maximum stroke, obtained using the discussed approaches. Next to illustrate the performance and potential of the developed topology optimization procedure, the results in this section also serve to provide some insight into the characteristics of the considered design problem.

### 5.1. Numerical examples problem formulation

The specific geometry and boundary conditions of the considered test problem are depicted in Figure 5(a), as an instance of the general SMA actuator optimization problem defined in Section 3.1. In the dark load-application region, distributed rather than concentrated loads are applied, to avoid the high gradients associated with the latter and to better represent physical reality. As output displacement  ${}^{\text{out}}U$ , we consider the  $y$ -displacements of the nodes where the distributed load  ${}^{\text{ext}}F$  is applied. Optional springs, acting in the  $y$ -direction with total stiffness  $k$  (also applied in a distributed sense), can also be included.

When dealing with multiple output displacement components, the optimization should not maximize the stroke at only one single node and disregard the rest, but instead a certain uniformity of  ${}^{\text{out}}U$  is desired. To achieve this, we define a measure  $\tilde{s}$  for the stroke  $s$  based on the *smallest* component of all output displacements in the low-temperature case (where the displacements are larger) and the *largest* component of all output displacements in the high-temperature case (where the displacements are smaller). This makes that the worst-performing nodes become most relevant

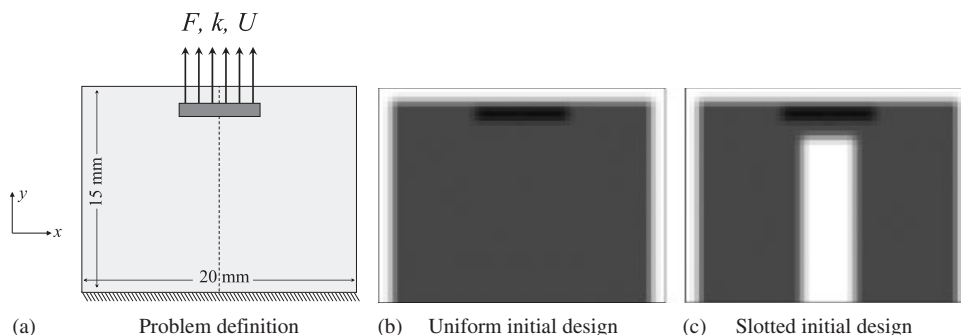


Figure 5. Problem definition (a) and initial designs (b, c) used for numerical SMA thermal actuator topology optimization studies.

to the optimizer. As differentiable substitutes for maximum and minimum operators, we use the  $P$ -norm, which leads to the following stroke-like measure  $\tilde{s}$ :

$$\tilde{s} = U_1 - U_2 \quad \text{with} \quad (38)$$

$$U_1 = \left( \frac{1}{m} \sum_{i=1}^m (\xi + u_i(T_1))^{-P} \right)^{-1/P} \approx \min^{\text{out}} U(T_1), \quad (39)$$

$$U_2 = \left( \frac{1}{m} \sum_{i=1}^m (u_i(T_2))^P \right)^{1/P} \approx \max^{\text{out}} U(T_2). \quad (40)$$

Here  $u_i$  represents one of the  $m$  components of  $^{\text{out}}U$ . This formulation assumes that output displacements are defined such, that they are always positive. The constant  $\xi=0.1$  has been introduced in the  $T_1$ -case to prevent division by zero; for the exponent,  $P=3$  proved adequate. Note that the objective  $\tilde{s}$  now no longer is the actual physical stroke, but a measure related to it, which also includes uniformity in displacement between all output nodes. The sensitivity analysis of  $\tilde{s}$  is discussed in Appendix B.

The problem defined in Figure 5(a) can be used to study two common application situations for SMA actuation (see e.g. [1, 7]): actuators preloaded by springs under pretension (bias springs,  $k>0$ ), and actuators preloaded by constant forces such as weights ( $k=0$ ). In the latter case, additional end stiffness constraints are included in the problem formulation since the unconstrained problem is not well defined and tends to generate overly compliant designs, as discussed in Section 3.1. The results for both cases are presented in Sections 5.2 and 5.3, respectively.

The general optimization problem definition is similar to Equation (17), with the stroke-like measure  $\tilde{s}$  defined in Equation (38) replacing the stroke objective  $s$ , and the nodal design variable vector  $\alpha$  (Section 4.4) replacing the ECP quasi-density variable vector  $\gamma$ , i.e.:

$$\begin{aligned} \max_{\alpha} \quad & \tilde{s} (^{\text{out}}U(T_1, \alpha), ^{\text{out}}U(T_2, \alpha)), \\ \text{s.t.} \quad & g(T_l, \alpha) \leq g_{\max} \quad l=1, 2 \quad (\text{optional}). \end{aligned} \quad (41)$$

In all considered cases, uniform temperatures  $T_1=328\text{ K}$  and  $T_2=343\text{ K}$  are applied, to fully exploit the low-hysteresis range of the material (cf. Figure 1). Symmetry of the analysis and the designs is imposed, which allows designs to be analyzed on only half the domain and optimized using only half the design variables, saving considerable computational effort. On the downside, imposing symmetry in the considered nonlinear setting could artificially hide potential asymmetric responses and designs; this topic is addressed in detail in Section 5.4 below. In the presented cases, meshes of  $30 \times 45$  elements were used, with a filter radius  $r$  of 0.8 mm. Section 5.5 briefly reports on numerical results using higher design resolutions and more refined meshes.

Furthermore, numerical experiments have shown that the present SMA actuator optimization problem is non-convex and possesses multiple local optima. To improve chances to find high-performing (local) optima, we have used multiple initial design configurations. The two considered initial configurations, denoted as the ‘uniform’ and ‘slotted’ case, are depicted in Figure 5(b) and (c). Note also the solid non-design load-application region as well as the zero-density ghost layer at the free domain boundary discussed earlier.

The optimization problems have been solved using Svanberg’s method of moving asymptotes optimizer [43] augmented with a move limit strategy, restricting the individual density design variable updates  $|\Delta\alpha^{(i)}|$  to 0.05 to ensure stable convergence. The optimization convergence criterion is based on the absolute change in the objective in two subsequent iterations. When this change is less than  $10^{-7}$  for a feasible design point, the process is considered to have reached convergence. This limit is deliberately chosen relatively small, in order not to stop the process prematurely.

Finally, after various demonstrations of the applicability and versatility of the proposed SMA actuator topology optimization, Section 5.6 discusses interpretation of the obtained results, and makes an attempt at explaining the implicit connection between the obtained designs and the underlying physical principles.

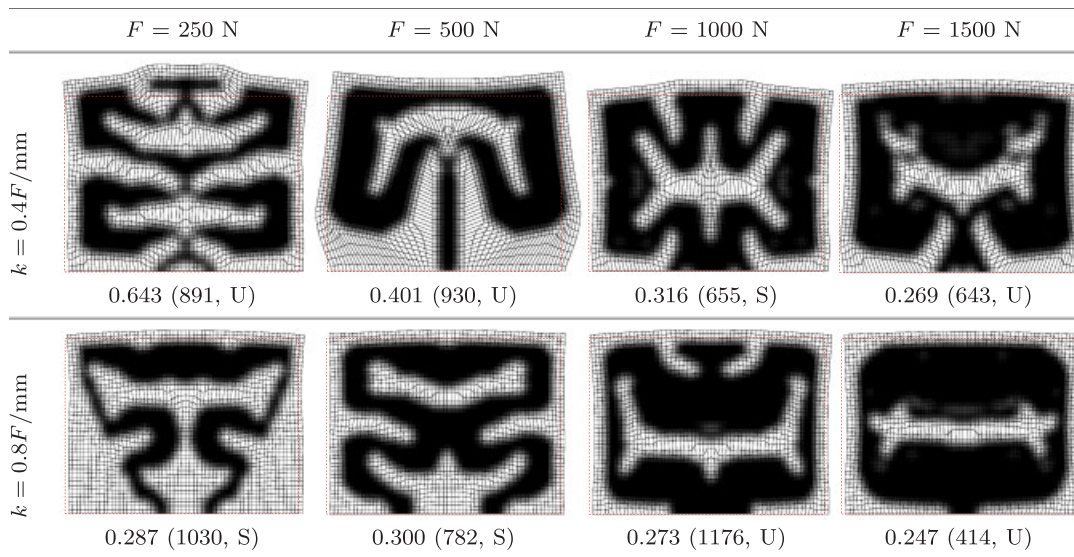


Figure 6. Overview of best results obtained for various combinations of applied load and bias spring stiffness. Depicted deformed designs are from the low temperature case. Obtained objective value, number of iterations and initial design (Slotted/Uniform) are indicated.

### 5.2. Case 1: Stroke maximization of spring-loaded SMA actuators

Obtained designs for unconstrained optimization according to Equation (41), for cases with various combinations of load magnitude  $F$  and spring stiffness  $k$ , are shown in Figure 6, together with the final strokes and number of iterations. The bias spring stiffness has been varied in proportion to  $F$ , corresponding to a certain preloading extension  $F/k$ , and for each load value a weaker and stiffer case has been considered. Individual spring-loaded cases are referenced by load and relative spring stiffness used, in the following notation:  $250 \text{ N}; 0.4F/\text{mm}$  for the  $F=250 \text{ N}$ ,  $k=0.4F/\text{mm}$ -case, etc. All cases have been optimized starting from both the uniform and slotted initial designs, and the best result is depicted as indicated. The results show the typical slightly fuzzy boundary caused by the applied density filter, and in some cases small intermediate-density compliant hinges or regions are observed, as no measure was included to actively suppress these. Nonetheless, we consider these results suitable for the present purposes: demonstration of the proposed optimization procedure and preliminary investigation of characteristics and trends in the considered SMA thermal actuator design problems.

Examination of Figure 6 shows that the actuator designs depend strongly on the applied load level and spring stiffness. Designs show considerable variation in shape, topology, slenderness and volume, where, qualitatively, in general designs grow more thick and massive as the applied load is increased. Furthermore, most designs fill the design domain as much as possible, and particularly at lower load magnitudes meandering structures emerge. Presumably, the optimizer modulates the stiffness of the structure to benefit the most from the applied loading, while simultaneously maximizing the amount of SMA material that can be recruited to create the actuation effect. A more in-depth discussion on design interpretation is given in Section 5.6 below.

Considering the objective values, the examples show that increasing spring stiffness leads to a reduction of the obtained strokes. This is directly related to the fact that the applied load reduces faster upon deformation of the structure when the bias spring is stiffer. Therefore, the optimization has less load and displacement ‘budget’ available to generate the actuation effect. In addition, at identical bias spring stiffness, the cases with higher loading also show higher objectives (cf.  $250 \text{ N}; 0.8F/\text{mm}$  versus  $500 \text{ N}; 0.4F/\text{mm}$ , etc.). This trend can be explained by the same reasoning.

Optimization runs have taken between 414 and 1176 iterations, in part due to the conservative termination criterion used combined with the move limit strategy. Yet, the nonlinearity of the

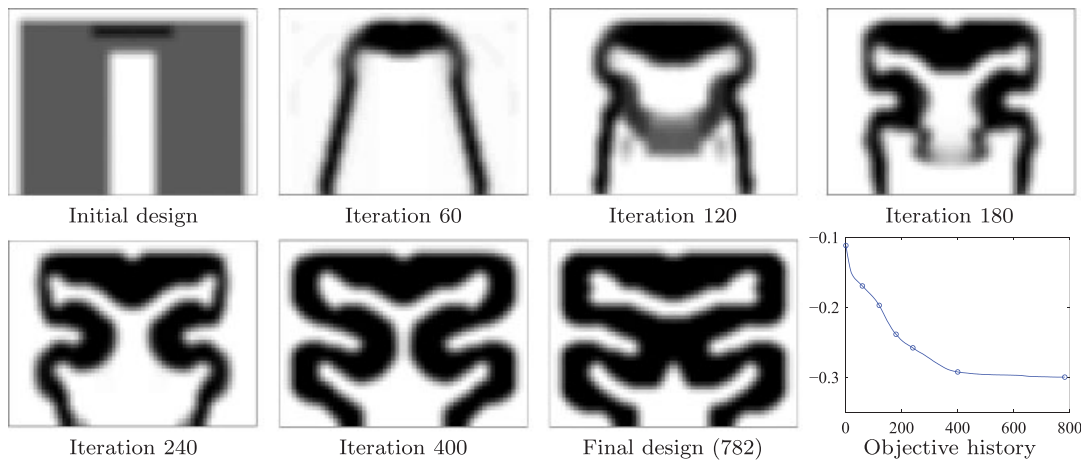


Figure 7. Optimization history for the 500N; 0.8F/mm case with slotted initial design, in terms of designs (undeformed) at seven instants and objective evolution. The design snapshots are indicated in the objective curve.

problem also strongly contributes to the relatively high number of iterations, as designs typically undergo various topological changes during the optimization process. To give an indication of the design evolution and optimization convergence, Figure 7 depicts several design snapshots of the 500N; 0.8F/mm case, as well as the objective history. The latter shows a smooth convergence, while the snapshots show how the design forms, breaks and forms connections and expands horizontally to exploit the full design domain.

### 5.3. Case 2: Stroke maximization of constant-force-loaded SMA actuators

In the case of SMA actuators loaded by constant force, and without bias springs ( $k=0$ ), end stiffness constraints are included the optimization problem of Equation (41), one for each temperature case. These require the actuator to have a minimum stiffness in its deformed state, which ensures well-defined positioning and helps to avoid overly compliant designs. Various combinations of applied load magnitude  $F$  and target end stiffness  $g_{\max}$  have been explored, and the best obtained designs are depicted in Figure 8. End stiffness is expressed in proportion to  $F$ , to define the positioning error under relative load fluctuations. Similar notation as for the spring-loaded results is used to reference individual cases, e.g. 250N; 1F mm denotes the  $F=250\text{N}$ ,  $g_{\max}=1F$ -case.

As in Section 5.2, designs vary considerably with applied load and required stiffness, although a number of cases show clear similarities. Comparing these results directly with the preceding spring-loaded case in terms of performance is not meaningful, as they are based on different situations and requirements. However, the general observations regarding filling of the domain and formation of meandering structures can be recognized in this case as well. Also, although to a lesser extent than in Figure 6, still the trend toward more massive, thicker designs at higher load levels can be distinguished. In addition, with the end stiffness requirement, the more compliant cases ( $g_{\max}=2F$  mm, bottom row in Figure 8) show larger objective values than their stiffer counterparts (top row). Flexibility promotes large displacements and potentially high strokes (i.e. displacement differences at different temperatures). Demands on structural stiffness therefore force the optimization process to find a compromise, which results in lower strokes. In relation to this, note that most of the final designs depicted in Figure 8 (as well as Figure 6) do not show excessive deformations. However, during the course of the optimization intermediate designs often exhibit far more compliant behavior and larger deformations, which necessitates the use of the described robust modeling and analysis approaches.

Figure 9 depicts the optimization process for the 250N; 1Fmm case, showing several design snapshots and objective and end stiffness constraint histories. The process spontaneously forms meandering structures that appear to modulate the compliance of the design. Again gradual and smooth convergence is observed, and only the low-temperature end stiffness constraint turns out to

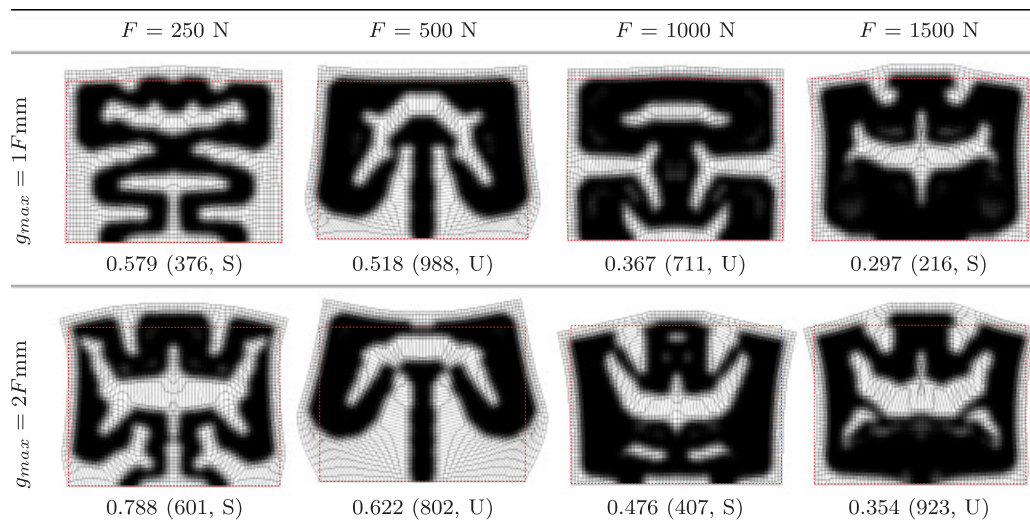


Figure 8. Overview of best results obtained for various combinations of applied load and end stiffness. Depicted deformed designs are from the low temperature case. Obtained objective value, number of iterations and initial design (Slotted/Uniform) are indicated.

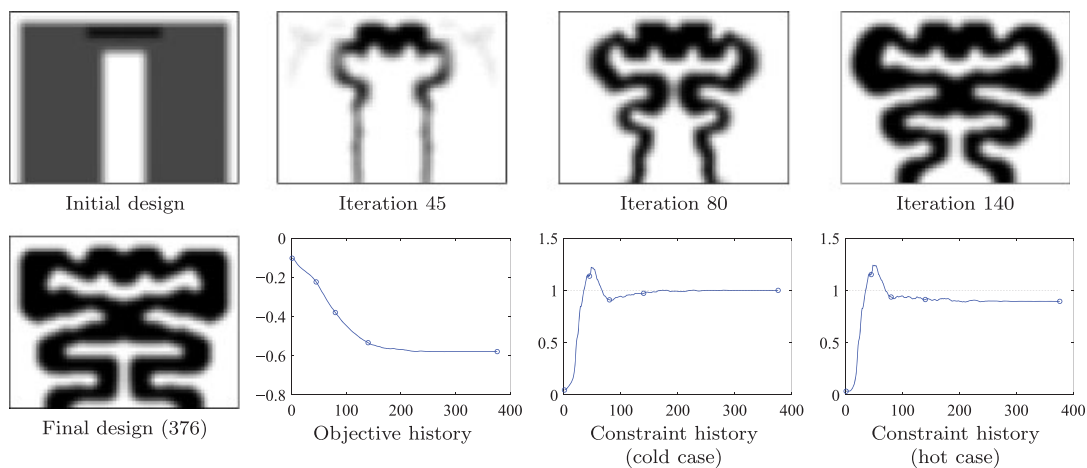


Figure 9. Optimization history for the 250N; 1Fmm case started with a nonuniform initial design, in terms of designs (undeformed) at five instants, objective and end stiffness constraint values. The design snapshots are indicated in the history curves.

be active at the optimum. This is not a universal rule, in other cases the high-temperature constraint was found to be limiting.

#### 5.4. Note regarding the use of symmetry

All the above results were obtained using symmetry conditions, in order to limit computational effort, both regarding the nonlinear analysis (fewer DoFs) and the optimization problem (fewer design variables). However, enforcing symmetry is not a recommended practice in general for nonlinear problems. In the analysis, it masks potential non-symmetric deformation modes (e.g. buckling), and regarding the optimization, the design space is restricted. To test the validity of using symmetric simulations in case of the obtained designs, full-domain simulations were performed of all final designs shown in Figures 6 and 8. In all the cases, the obtained deformations agreed with the results obtained using symmetry conditions. Even when including small load misalignments, no drastic changes occurred in design deformations, illustrating the stability of the symmetric

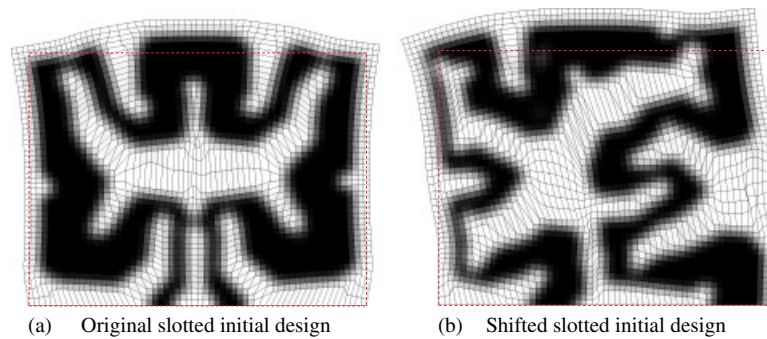


Figure 10. Deformed, low temperature case designs obtained in full domain optimization of the 250N; 2Fmm case, started with the original slotted design (a) and with the slotted design shifted by 1/3 mm to the right (b).

responses. Hence, in these cases, no asymmetric response occurs and their behavior is adequately represented by the applied symmetric analysis. This may however differ in other cases.

A second consideration is the enforced symmetry of the design. Possibly, not imposing symmetry conditions on design and analysis allows a non-symmetric and superior design to emerge during optimization. To get an indication of the validity of the optimization results obtained using enforced symmetry, we have optimized the constant-force 250N; 2F mm case on a full domain, without any symmetry conditions on design or analysis, starting from the symmetric slotted initial design. The obtained design, depicted in Figure 10(a), is virtually symmetric in shape and function, and shows strong resemblance to the design depicted in Figure 8. Most noticeable differences are found in the objective (0.836 versus 0.788 for the symmetric case) and the number of iterations (3816 versus 601, respectively). As the number of output nodes differs, it is better to compare the actual stroke of the center node, which shows a corresponding 0.755 versus 0.698 mm. It appears that in the full-domain case, the optimizer found a way to circumvent the local optimum which trapped it in the symmetric case. The end result is however still symmetric.

This may not be a general result, and in other cases more strongly asymmetric designs may emerge. One way to trigger asymmetric designs is starting from an asymmetric initial design. To illustrate this, Figure 10(b) shows a design obtained for the same constant-force 250N; 2F mm case on a full domain, started from the slotted initial design with the void center region shifted one element (1/3 mm) to the right. The slight initial asymmetry strongly affects the optimization process (as expected from the nonlinearity and non-convexity of the problem), and a radically different design emerges. In this particular case, an objective of 0.730 was reached in 5303 iterations (0.642 mm center node stroke), which is a local optimum given the superior symmetric design. However, this also may differ in other cases, and exploring multiple initial designs to increase the probability of reaching the best designs is recommended.

In summary, not imposing symmetry in nonlinear problems is in general the recommended approach. It potentially leads to better final designs, and does not mask otherwise hidden instabilities (asymmetric buckling modes). Certainly, the required computational effort is substantially larger, which is the main reason why the majority of the examples in this section nevertheless involves symmetry constraints.

### 5.5. Influence of the design resolution

This section explores the influence of the design resolution, which in the present setup depends on the element sizes together with the filter radius. It is expected that increasing the resolution leads to better optimization results, as the design freedom is increased. However, in this nonlinear problem, an increased resolution may also increase the number of local optima in which the optimization might get trapped. To explore this aspect, several experiments with the 250N; 1F mm case (with end stiffness constraints, see Figure 8) have been performed, using two meshes and three filter radii  $r$ . The results, again based on enforced symmetry for convenience, are summarized

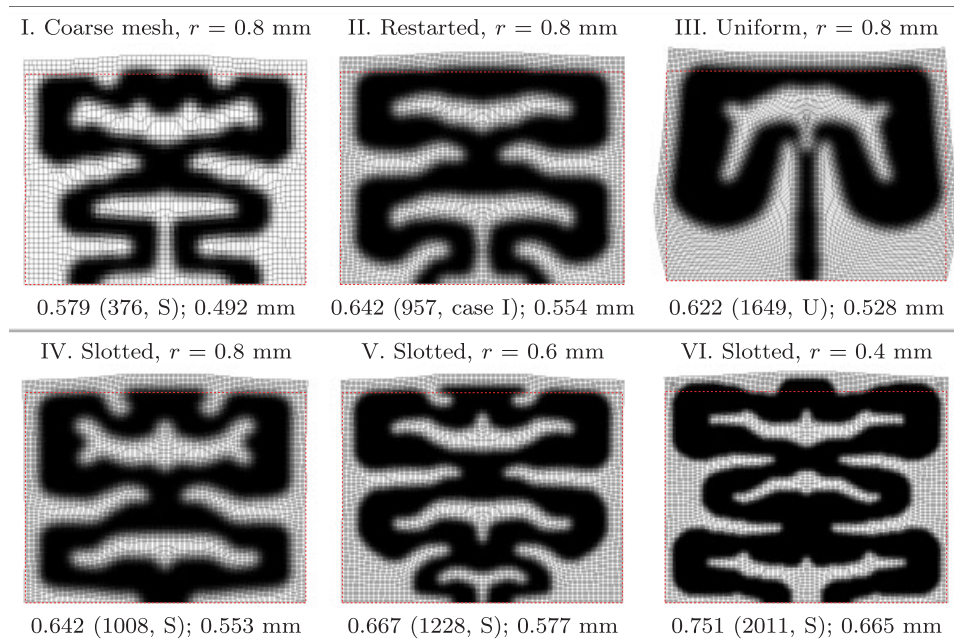


Figure 11. Influence of mesh refinement and filter size on obtained designs, for the constant-force 250 N; 1 F mm case. Shown are deformed designs in the low temperature case. Objective, iterations, initial design used, and center node stroke are indicated.

in Figure 11, showing cases I till VI. As the number of output nodes differs between the coarse and fine cases, center node strokes are also listed.

The coarse mesh ( $30 \times 45$  elements, 1406 design variables) result obtained with the slotted initial design and a filter radius  $r$  of 0.8 mm (I) is taken as the baseline case. The fine mesh consists of  $60 \times 90$  elements (5531 design variables), and filter radii 0.8, 0.6 and 0.4 mm are considered, the latter giving the same element size/filter radius ratio as the coarse problem. In addition to optimizations starting from the slotted initial design (IV–VI), a refined problem is started from the optimal coarse design mapped to the refined grid (II), and another is started from a uniform initial design (III).

From the results listed in Figure 11, it can be seen that, for this example, restarting the optimization from the coarse design on a finer grid with the same filter radius (II) resulted in a clear improvement. The finer grid allows a more precise design specification, so even though the filter size is the same a better design resolution is achieved. The refined design better fills the design domain, which may indicate that the coarse result may have converged prematurely.

The refined case started from a uniform initial design (III) yielded a structure similar to the 500 N cases shown in Figure 8. Its objective is however lower than that of the restarted case (II), signifying a local optimum. The refined 0.8 mm filter case started from the slotted initial design (IV) performs better, and although its shape and topology differs from the refined case restarted from the coarse design (II), the realized objectives and strokes are similar.

Finally, the refined cases with reduced filter radii indeed perform as expected: smaller  $r$  gives higher design freedom and yields higher strokes. The best case, with  $r = 0.4$  mm (VI), shows a 17% improvement over the  $r = 0.8$  mm refined case (II), and a 30% improvement over the coarse baseline result (I). In addition, the structural members seem to attain a certain width independent of the filter size, and as the filter radius reduces, the inherent density-filtering-induced intermediate-density boundary layer shrinks relative to the structural dimensions. This is seen most clearly in the  $r = 0.4$  mm case (VI), which has a much more crisp boundary than, e.g. the 0.8 mm cases. In addition to the increased design freedom and associated improved performance, this is another benefit of increasing the design resolution. A clear downside is however the computational cost: the increased number of DoFs results in considerably longer analysis times, and the increased number of design variables leads to a higher number of iterations.

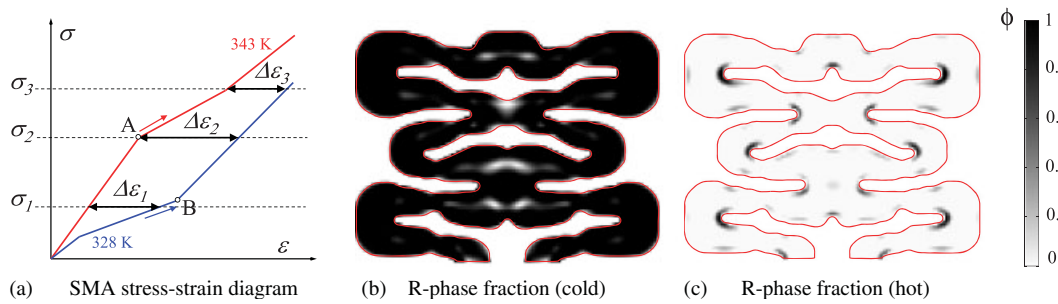


Figure 12. Schematic SMA stress–strain diagram (a), and R-phase fraction  $\phi$  distributions in the cold and hot case (b, c) of the refined 250N; 1Fmm design with filter radius 0.4 mm.

### 5.6. Interpretation of obtained designs

Given the multitude of different SMA actuator designs produced by our topology optimization, the question arises why they differ so much with the load level, and which phenomena drive the structures toward the observed designs. It is hard to give a specific answer to this question, but in qualitative terms, and discarding the influence of springs or stiffness constraints, a relatively simple interpretation can be given. With the schematic SMA stress–strain curves in Figure 12(a) taken from the actual material behavior (Figure 1(a)), we consider an SMA actuator operating under constant loading. Suppose that this constant loading results in local stress levels in the actuator that, for simplicity, do not change as the actuator deforms. Three such stress levels are indicated in Figure 12(a). Upon thermal cycling between low and high temperature levels (328–343 K), the applied stress induces different levels of strain in the material due to the temperature-dependent SMA material behavior. Ignoring geometrical effects depending on the shape of the structure, let us assume that a certain amount of strain difference  $\Delta\epsilon$  corresponds monotonically with a certain actuator stroke. Thus, stroke maximization equals maximization of local strain differences over temperature. By adapting the actuator design, the stress distribution can be modified, shifting local stress levels along the  $\sigma$ -axis. It is clear from Figure 12(a) that the optimal stress level is at  $\sigma_2$ , which yields a larger  $\Delta\epsilon$  than both the lower  $\sigma_1$  and the higher  $\sigma_3$ . In this situation, maximum use is made of the constitutive behavior of the SMA material. Note that for ease of presentation, this concept is presented in a 1-D representation, but it also holds in the multi-dimensional case.

This principle to make the best use of the SMA material behavior can also be qualitatively recognized in the obtained designs. As discussed, many designs contain meandering or bending-loaded structural parts, as well as thicker and thinner sections, which all serve to modulate the local stress levels to fully exploit the material behavior. These structures also help to increase the overall compliance and thus increase deflections at both temperatures, but without a constitutive temperature-driven effect as outlined here, compliance alone does not contribute to actuation capability.

Furthermore, note that in Figure 12(a) for the high-temperature curve the onset of the phase transformation is at point A, and for the low-temperature curve the transformation is completed at point B. Combined with the notion that at stress level  $\sigma_2$  (passing through A) the best performance is obtained, we can expect to see designs that show, in a large fraction of the structural domain, (1) very little phase transformation in the high-temperature case (since the best stress level lies at point A) and (2) virtually complete phase transformation in the low-temperature case (since the stress level  $\sigma_2$  lies above point B). Phase fractions can be computed quantitatively, through the computational R-phase fraction indicator  $\phi$  defined in Equation (35). Taking the  $r = 0.4$  mm design of Section 5.5 (250N; 1Fmm, Figure 11) as an example of a well-performing SMA actuator, Figure 12(b) and (c) show the phase fraction  $\phi$  in the low- and high-temperature case. Indeed the low-temperature case is nearly fully transformed, while the high-temperature case barely shows any R-phase material apart from a few small isolated regions.

Although this interpretation has its limitations (e.g. it does not consider other structural functions of the material such as in mechanisms, and does not account for compromises required to satisfy stiffness constraints), the phase fraction distributions provide an additional dimension for design

interpretation, and may facilitate the process of gaining further insight into the principles behind the actuator designs generated by the proposed SMA topology optimization.

## 6. CONCLUSIONS

The main challenge outlined in the Introduction of this paper was to formulate a method to make the creative power of topology optimization techniques available to the design of SMA actuation. Within the scope chosen for our initial investigation (plane stress, a specific material, uniform temperatures, quasi-static), the examples presented show that the proposed approach is indeed capable of designing effective SMA actuators. Specifically regarding the methodology, we conclude that:

- A robust analysis is crucial for the success of the optimization procedure. Our investigations reveal that in this regard, in a geometrically and materially nonlinear setting, the ECP approach is superior than the existing alternatives. In addition, in contrast to material-scaling methods (e.g. SIMP) it does not require arbitrary choices of material interpolations and its sensitivity analysis is independent of the material model.
- In spite of the above, we have found that also an ECP-based analysis can fail in extreme situations, for instance when the optimization creates inadequate and unrealistic designs. Multiple generic and problem-specific measures have been applied to maximize analysis robustness, which have made failures sufficiently rare to render the topology optimization procedure suited for practical use.
- We have been able to apply, in part due to the chosen material model, relatively inexpensive analysis and adjoint sensitivity analysis techniques, which have helped keeping the computational cost of the topology optimization procedure acceptable.

Furthermore, on the basis of the studied numerical examples, we conclude that in general, the numerical experiments confirm the intuitive notions that both larger loads and stricter stiffness requirements lead to SMA actuators with smaller strokes. A simplified interpretation in terms of SMA phase fractions could be linked to numerical observations on the obtained designs. The optimization itself was found to be rather sensitive to the initial design used, which is not surprising given the nonlinear nature of the problem. The use of multiple initial designs and, preferably, omission of symmetry constraints is recommended. As expected, when local minima can be avoided, both mesh refinement and increased design resolution enable better-performing designs, at the expense of increased analysis costs and more optimization iterations.

The methodology and results presented here form an encouraging first step toward a more sophisticated and powerful topology optimization-based design procedure for SMA structures. Next to computational improvements, one could consider alternate problem definitions and multiple load cases, the inclusion of electrical and thermal fields to simulate commonly applied resistive heating, 3-D geometries, etc. A significant further challenge is to replace the considered SMA material model by a more general path-dependent one, with all its computational consequences. In addition, the consideration of transient behavior and dynamic effects (e.g. latent heat of transformation) would add further realism. With all these possible improvements, nonlinear analysis stability under a wide variety of designs remains an extremely important aspect that requires adequate attention.

## APPENDIX A: MATERIAL PARAMETERS USED IN THE SMA CONSTITUTIVE MODEL

The temperature- and strain-dependent parameters  $A_i$  and  $B_i$  (Equation (3)) are defined by

$$A_i = \begin{cases} \varepsilon_{xx} \leq \varepsilon_1 : & A_1 = E_A, \\ \varepsilon_1 < \varepsilon_{xx} \leq \varepsilon_2 : & A_2 = E_T, \\ \varepsilon_{xx} > \varepsilon_2 : & A_3 = E_R, \end{cases} \quad (A1)$$

Table AI. Material constants used in the SMA constitutive model.

$K_\varepsilon$	$2.469 \times 10^{-4} \text{ K}^{-1}$	$K_E$	$677.9 \text{ MPa K}^{-1}$
$\varepsilon_0$	$5.889 \times 10^{-4}$	$\Delta$	$52.66 \times 10^{-4}$
$E_0$	18.82 GPa	$T_0$	328 K
$E_{A_0}$	66.26 GPa	$E_R$	45.56 GPa
$K_{E_A}$	$1.876 \times 10^2 \text{ GPa K}^{-1}$	$\nu$	0.333

$$B_i = \begin{cases} \varepsilon_{xx} \leq \varepsilon_1 : & B_1 = 0, \\ \varepsilon_1 < \varepsilon_{xx} \leq \varepsilon_2 : & B_2 = (E_A - E_T)\varepsilon_1, \\ \varepsilon_{xx} > \varepsilon_2 : & B_3 = (E_A - E_T)\varepsilon_1 + (E_T - E_R)\varepsilon_2 \end{cases} \quad (\text{A2})$$

with

$$\varepsilon_1(T) = K_\varepsilon(T - T_0) + \varepsilon_0, \quad \varepsilon_2(T) = \varepsilon_1(T) + \Delta, \quad (\text{A3})$$

$$E_A(T) = K_{E_A}(T - T_0) + E_{A_0}, \quad E_T(T) = K_E(T - T_0) + E_0, \quad (\text{A4})$$

where  $K_\varepsilon, K_E, \varepsilon_0, \Delta, E_0, T_0, K_{E_A}, E_{A_0}$  and  $E_R$  denote various material constants. Values as obtained by curve fitting to experimental data (Figure 1(a)) and used in this paper are listed in Table AI.

## APPENDIX B: SENSITIVITY ANALYSIS OF STROKE-LIKE MEASURE $\tilde{s}$

The sensitivity of  $\tilde{s}$  defined in Equation (38) is given by:

$$\frac{d\tilde{s}}{d\gamma} = \frac{dU_1}{d\gamma} - \frac{dU_2}{d\gamma} \quad \text{with} \quad (\text{B1})$$

$$\frac{dU_1}{d\gamma} = \left( \frac{1}{m} \sum_{i=1}^m (\xi + u_i(T_1))^{-P} \right)^{-1-1/P} \left( \frac{1}{m} \sum_{i=1}^m (\xi + u_i(T_1))^{-1-P} \frac{du_i(T_1)}{d\gamma} \right), \quad (\text{B2})$$

$$\frac{dU_2}{d\gamma} = \left( \frac{1}{m} \sum_{i=1}^m (u_i(T_2))^P \right)^{-1+1/P} \left( \frac{1}{m} \sum_{i=1}^m u_i(T_2)^{-1+P} \frac{du_i(T_2)}{d\gamma} \right). \quad (\text{B3})$$

The derivatives above contain summations involving terms  $du_i/d\gamma$ . These could all be computed separately using Equation (26) using an appropriate displacement-component selection vectors  $\mathbf{L}_i$  in Equation (18), and the summation could be performed as a post-processing operation. However, for larger  $m$ , this requires a considerable number of repeated computations. A more efficient way is to construct the  $\mathbf{L}$ -vector used in Equation (18) *within* the analysis, using the factors given in the second summation of Equations (B2) and (B3) (e.g.  $(\xi + u_i(T_1))^{-1-P}$  for the former) as entries in a selection vector  $\mathbf{L}$  constructed dynamically. Note that due to their dependence on  $u_i$  these factors cannot be supplied as constants in advance. This adaptation allows the efficient adjoint sensitivity analysis of  $U_1$  and  $U_2$  as single response quantities.

## ACKNOWLEDGEMENTS

The financial support of the Dutch MicroNed Programme as well as the National Creative Research Initiatives Program (Korea Science and Technology Foundation grant No. 2010-0019241), contracted through the Institute of Advanced Machinery and Design at Seoul National University is gratefully acknowledged. The authors thank Krister Svanberg for the use of his MMA code, and the reviewers for their constructive suggestions.

## REFERENCES

1. Duerig T, Melton K, Stöckel D, Wayman C. *Engineering Aspects of Shape Memory Alloys*. Butterworth-Heinemann: London, 1990.
2. Otsuka K, Wayman C (eds). *Shape Memory Materials*. Cambridge University Press: Cambridge, 1998.

3. Rohde M, Schüssler A. On the response-time behaviour of laser micromachined NiTi shape memory actuators. *Sensors and Actuators, A: Physical* 1997; **61**(1–3):463–468.
4. Peirs J, Reynaerts D, Van Brussel H. Scale effects and thermal considerations for micro-actuators. *Proceeding of the IEEE International Conference on Robotics and Automation*, Leuven, Belgium, 1998; **2**:1516–1521.
5. Duerig T, Pelton A, Stöckel D. An overview of NiTiNol medical applications. *Materials Science and Engineering: A—Structural Materials: Properties, Microstructure and Processing* 1999; **273–275**:149–160. DOI: 10.1016/S0921-5093(99)00294-4.
6. Van Humbeeck J. Non-medical applications of shape memory alloys. *Materials Science and Engineering: A—Structural Materials: Properties, Microstructure and Processing* 1999; **273–275**:134–148. DOI: 10.1016/S0921-5093(99)00293-2.
7. Kohl M, Brugger D, Ohtsuka M, Takagi T. A novel actuation mechanism on the basis of ferromagnetic SMA thin films. *Sensors and Actuators A: Physical* 2004; **114**(2–3):445–450. DOI: 10.1016/j.sna.2003.11.006.
8. Liang C, Rogers CA. Design of shape memory alloy actuators. *Journal of Mechanical Design* 1992; **114**:223–230. DOI: 10.1115/1.2916935.
9. Troisfontaine N, Bidaud P, Larnicol M. Optimal design of micro-actuators based on SMA wires. *Smart Materials and Structures* 1999; **8**(2):197–203. DOI: 10.1088/0964-1726/8/2/004.
10. Birman V, Chandrashekhara K, Sain S. Approach to optimization of shape memory alloy hybrid composite plates subjected to low-velocity impact. *Composites Part B: Engineering* 1996; **27**(5):439–446. DOI: 10.1016/1359-8368(96)00010-8.
11. Lu TJ, Hutchinson JW, Evans AG. Optimal design of a flexural actuator. *Journal of the Mechanics and Physics of Solids* 2001; **49**(9):2071–2093. DOI: 10.1016/S0022-5096(01)00024-2.
12. Fischer H, Vogel B, Pflöging W, Besser H. Flexible distal tip made of NiTiNol (NiTi) for a steerable endoscopic camera system. *Materials Science and Engineering: A—Structural Materials: Properties, Microstructure and Processing* 1999; **273–275**:780–783. DOI: 10.1016/S0921-5093(99)00415-3.
13. Kohl M, Just E, Pflöging W, Miyazaki S. SMA microgripper with integrated antagonism. *Sensors and Actuators A: Physical* 2000; **83**(1–3):208–213. DOI: 10.1016/S0924-4247(99)00385-4.
14. Lagoudas DC, Strelec JK, Yen J, Khan MA. Intelligent design optimization of a shape memory alloy actuated reconfigurable wing. *Proceeding of SPIE Smart Structures and Materials*, Newport Beach, CA, U.S.A., vol. 3984, 2000; 338–348. DOI: 10.1117/12.388778.
15. Dumont G, Kuhl C. Finite element simulation for design optimisation of shape memory alloy spring actuators. *Engineering Computations* 2005; **22**(7–8):835–848. DOI: 10.1108/02644400510619549.
16. Langelaar M, Van Keulen F. A simple R-phase transformation model for engineering purposes. *Materials Science and Engineering: A—Structural Materials: Properties, Microstructure and Processing* 2004; **378**(1–2):507–512. DOI: 10.1016/j.msea.2003.12.049.
17. Langelaar M, Van Keulen F. Modeling of shape memory alloy shells for design optimization. *Computers and Structures* 2008; **86**(9):955–963. DOI: 10.1016/j.compstruc.2007.04.017.
18. Langelaar M, Van Keulen F. Sensitivity analysis of shape memory alloy shells. *Computers and Structures* 2008; **86**(9):964–976. DOI: 10.1016/j.compstruc.2007.04.019.
19. Langelaar M, Van Keulen F. Sensitivity analysis and optimization of a shape memory alloy gripper. *Proceeding of the 8th International Conference on Computational Structures Technology*. Civil-Comp Press: Stirlingshire, U.K., 2006.
20. Langelaar M, Van Keulen F. Gradient-based design optimization of shape memory alloy active catheters. *Proceeding of the 15th AIAA/ASME/AHS Adaptive Structures Conference*, Waikiki, U.S.A., 2007.
21. Bendsoe MP, Kikuchi N. Generating optimal topologies in structural design using a homogenization method. *Computer Methods in Applied Mechanics and Engineering* 1988; **71**(2):197–224. DOI: 10.1016/0045-7825(88)90086-2.
22. Bendsoe MP, Sigmund O. *Topology Optimization—Theory, Methods and Applications*. Springer: Berlin, 2003.
23. Pedersen CBW. Revisiting topology optimization of continuum structures with elastoplastic response. *The 15th Nordic Seminar of Computational Mechanics*, Aalborg, Denmark, 18–19 October 2002.
24. Saxena A. Topology design of large displacement compliant mechanisms with multiple materials and multiple output ports. *Structural and Multidisciplinary Optimization* 2005; **30**(6):477–490. DOI: 10.1007/s00158-005-0535-z.
25. Yoon GH, Kim YY. Element connectivity parameterization for topology optimization of geometrically nonlinear structures. *International Journal of Solids and Structures* 2005; **42**(7):1983–2009. DOI: 10.1016/j.ijsolstr.2004.09.005.
26. Kawamoto A. Stabilization of geometrically nonlinear topology optimization by the Levenberg–Marquardt method. *Structural and Multidisciplinary Optimization* 2009; **37**(4):429–433. DOI: 10.1007/s00158-008-0236-5.
27. Yoon GH, Kim YY. The element connectivity parameterization formulation for the topology design optimization of multiphysics systems. *International Journal for Numerical Methods in Engineering* 2005; **64**(12):1649–1677. DOI: 10.1002/nme.1422.
28. Cockburn B. Discontinuous Galerkin methods. *ZAMM—Journal of Applied Mathematics and Mechanics* 2003; **83**(11):731–754.
29. Langelaar M, Yoon GH, Gurav S, Kim YY, Van Keulen F. Modeling and design of shape memory alloy actuators. *Proceeding of the 6th IEEE EuroSimE Conference*, Berlin, Germany, 2005; 626–633.

30. Langelaar M. Design optimization of shape memory alloy structures. *Ph.D. Thesis*, Delft University of Technology, 2006.
31. Tobushi H, Tanaka K, Kimura K, Hori T, Sawada T. Stress–strain–temperature relationship associated with the R-phase transformation in Ti–Ni shape memory alloy. *JSME International Journal Series I—Solid Mechanics Strength of Materials* 1992; **35**(3):278–284.
32. Yoon GH, Kim YY. Topology optimization of material-nonlinear continuum structures by the element connectivity parameterization. *International Journal for Numerical Methods in Engineering* 2007; **69**(10):2196–2218. DOI: 10.1002/nme.1843.
33. Yoon GH, Joung YS, Kim YY. Optimal layout design of three-dimensional geometrically non-linear structures using the element connectivity parameterization method. *International Journal for Numerical Methods in Engineering* 2007; **69**(6):1278–1304. DOI: 10.1002/nme.1808.
34. Yoon GH, Kim YY, Langelaar M, Van Keulen F. Theoretical aspects of the internal element connectivity parameterization approach for topology optimization. *International Journal for Numerical Methods in Engineering* 2008; **76**(6):775–797. DOI: 10.1002/nme.2342.
35. Bathe KJ. *Finite Element Procedures*. Prentice-Hall: Englewood Cliffs, 1995.
36. Ramm E, Kemmler R. Stability and large deformations in structural optimization. *Proceeding of the International Symposium on Lightweight Structures in Civil Engineering*, Warsaw, Poland, 2002.
37. Kemmler R, Lipka A, Ramm E. Large deformations and stability in topology optimization. *Structural and Multidisciplinary Optimization* 2005; **30**(6):459–476. DOI: 10.1007/s00158-005-0534-0.
38. Buhl T, Pedersen CBW, Sigmund O. Stiffness design of geometrically nonlinear structures using topology optimization. *Structural and Multidisciplinary Optimization* 2000; **19**(2):93–104. DOI: 10.1007/s001580050089.
39. Kleiber M, Antúnez H, Hien TD, Kowalczyk P. *Parameter Sensitivity in Nonlinear Mechanics: Theory and Finite Element Computations*. Wiley: New York, 1997.
40. Crisfield MA. *Non-linear Finite Element Analysis of Solids and Structures, Volume I: Essentials*. Wiley: New York, 1991.
41. Bruns TE, Tortorelli DA. Topology optimization of non-linear elastic structures and compliant mechanisms. *Computer Methods in Applied Mechanics and Engineering* 2001; **190**(26–27):3443–3459. DOI: 10.1016/S0045-7825(00)00278-4.
42. Guest JK, Prévost JH, Belytschko T. Achieving minimum length scale in topology optimization using nodal design variables and projection functions. *International Journal for Numerical Methods in Engineering* 2004; **61**(2):238–254. DOI: 10.1002/nme.1064.
43. Svanberg K. The method of moving asymptotes—a new method for structural optimization. *International Journal for Numerical Methods in Engineering* 1987; **24**:359–373. DOI: 10.1002/nme.1620240207.