Design Optimization
of Shape Memory Alloy Structures

Matthijs Langelaar

Design Optimization
of Shape Memory Alloy Structures

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op vrijdag 15 december 2006 om 10.00 uur
door

Matthijs LANGELAAR

werktuigkundig ingenieur
geboren te Apeldoorn.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. A. van Keulen

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. A. van Keulen	Technische Universiteit Delft, promotor
Prof. dr. Y.Y. Kim	Seoul National University
Prof. dr. V.V. Toropov	University of Leeds
Prof. dr. ir. A. de Boer	Universiteit Twente
Prof. dr. Z. Gürdal	Technische Universiteit Delft
Prof. dr. ir. S. van der Zwaag	Technische Universiteit Delft
Prof. ir. R.H. Munnig Schmidt	Technische Universiteit Delft

Printed in the Netherlands by *PrintPartners Ipskamp*

ISBN-13 978-90-9021347-7
ISBN-10 90-9021347-3

Cover: *Rendered three-dimensional model of the mechanical structure of a shape memory alloy active catheter, designed for controlled bending. Derived from an actual optimized finite element model (see Chapter 3 and Chapter 9).*

Contents

Chapter 1

Introduction

1.1 Motivation

This thesis considers design optimization of shape memory alloy (SMA) structures. This subject is motivated by the promising opportunities for miniature SMA actuators in medical applications.

1.1.1 The vision: microsubmarines

> CARTER: *"And take a little trip with them . . . "*
>
> GRANT (bewildered): *"Trip? Where to?"*
>
> CARTER (matter-of-fact): *"Well, the only way to reach that clot is from inside the Brain. So we've decided to put a Surgical Team and a Crew into a submarine – reduce it way down in size, and inject it into an Artery – "*
>
> GRANT (jolted): *"You mean I'm going along?"*
>
> CARTER: *"As part of the Crew."*
>
> GRANT: *"Wait a minute! They can't shrink me!"*
>
> CARTER (assuringly): *"Grant, our Miniaturizer can shrink anything."*

These lines are quoted from the 1965 screenplay for the movie *Fantastic Voyage* written by Harry Kleiner and directed by Richard Fleischer. In this science-fiction movie, miniaturization is accomplished by a technique to shrink objects temporarily by shrinking their atoms. The technique is not perfect, as objects

revert to their original size after merely one hour. A scientist called Benes has found a way to make the shrinkage permanent, but his life is in danger as he suffers from a bloodclot in his brain. Brain surgery is impossible, as it would destroy too much delicate brain tissue. In a desperate attempt to save his life, a crew of five attempts to remove the clot from the inside, by miniaturizing themselves and their submarine named Proteus and traveling to the brain via the bloodstream. While their time runs out, they make a "fantastic voyage" inside Benes' bloodvessels, fighting off attacks from white blood cells and antibodies.

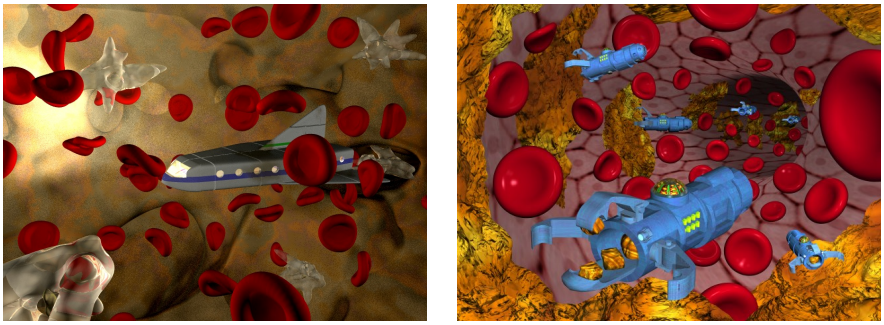


Figure 1.1: *Artist impressions of a "Fantastic Voyage" micro-submarine in a blood vessel (left, by David Morgan-Mar, with permission) and of nanorobots cleaning plaque from arterial walls (right, by Tim Fonseca, with permission).*

In spite of several strong arguments that render this concept of a micro-submarine impossible, the vision itself continues to inspire (Figure 1.1). The idea for this movie may in fact stem from another source often cited by authors addressing the prospects of miniaturization: in his famous lecture, Feynman (1960, 1992) also brings up the concept of a device that can operate on patients from inside their arteries:

> *"A friend of mine suggests a very interesting possibility for relatively small machines. He says that, although it is a very wild idea, it would be interesting in surgery if you could swallow the surgeon. You put the mechanical surgeon inside the blood vessel and it goes into the heart and looks around. It finds out which valve is the faulty one and takes a little knife and slices it out."* - R. Feynman, 1960.

That these inspiring visions actually lead to very real prototypes is illustrated by Figure 1.2, which shows a realization of a micro-submarine (hull dimensions 4 mm by 650 μm) presented by the German company microTEC in 1997 and then at the EXPO 2000 in Hannover (Moore, 2001). Next to the fact that this micro-submarine illustrated the capabilities of a highly precise UV curing additive fabrication technique for fabrication of complex three-dimensional microcomponents

(Reinhardt and Götzen, 2006), it also showed the possibilities of this technique for integration of various components of different materials at the microscale. This is illustrated by the fact that the propeller, which has a diameter of 600 μm and is attached to an axis with a diameter of merely 10 μm, can actually rotate freely. Note, however, that this prototype did not include sensors, actuators and control systems and a power source necessary to make it function autonomously, and the integration of those remains an enormous challenge. Still this micro-submarine received a lot of attention, and it even holds the world record for smallest medical submarine (Guinness, 2005).

Figure 1.2: *Conceptual colored scanning electron micrograph (SEM) of a micro-submarine in a human artery, made by German company MicroTEC (*www.microtec-d.com*), picture made by Oliver Meckes, Eye of Science (*www.eyeofscience.com*). Copyright: microTEC Germany, reprinted with permission.*

Micro-submarines in this form, however, will probably never be feasible for medical interventions such as actively hunting clots in bloodvessels. Scaling laws render propulsion in liquids increasingly inefficient, and speeds needed to overcome the blood flow velocity seem out of reach (see e.g. Avron *et al.*, 2004, Dario *et al.*, 1992, Edd *et al.*, 2003, Kosa *et al.*, 2005, Power, 1995, Vogel, 1994, Wautelet, 2001). Still, the potential uses for miniaturization in medicine are many, as discussed below.

1.1.2 The reality: minimally invasive therapy

Although miniaturization has been and is a strong trend in many industries nowadays, it is not accomplished by shrinking atoms, as suggested by Asimov (1966). Still, the idea of being able to operate from inside the body may have been the inspiration for many instruments currently used in so-called *minimally invasive* procedures (e.g. Fuchs, 2002, Mack, 2001, Rattner, 1999). An example is shown in Figure 1.3, which shows a comparison of conventional open heart surgery to a

minimally invasive version of the same procedure. Technical advancements have enabled invasive diagnostic studies or surgery to be carried out with minimal damage to healthy tissue, through the use of specialized instruments. This continuing trend from open surgery towards minimally invasive procedures offers clear advantages: less discomfort and complications, faster recovery times, shorter hospital stays, reduced treatment costs and higher survival rates.

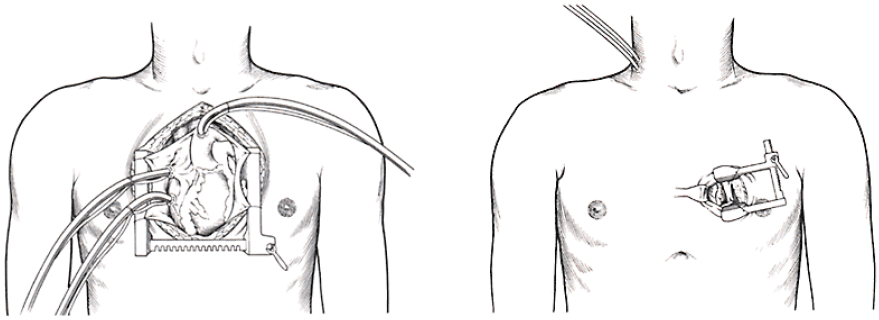


Figure 1.3: *Conventional (left) and minimally invasive (right) approaches to cardiac surgery. Image courtesy of the Department of Cardiothoracic Surgery of the University of Southern California, reprinted with permission.*

However, while the improvements for patients are clear, minimally invasive techniques make the task of physicians more difficult. Procedures must be carried out in an indirect way, which leads to hand-eye coordination problems and requires more extensive training. The maneuverability and dexterity of the instruments is often very limited. Moreover, much of the sensory feedback present in the "hands-on" approach in open surgery is lacking. Improved instruments can help to overcome these difficulties (Agrawal and Erdman, 2005, Mack, 2001, Tendick *et al.*, 1998).

Clearly, there is an incentive to make instruments used for minimally invasive therapy as small as possible, to achieve maximum access and minimal damage. But at the same time, as much relevant functionality and dexterity as possible should be implemented, to reduce the limitations imposed on the physicians. This can only be accomplished by extensive use of miniaturization. For example, diagnostics and sensory feedback can be enhanced by miniature sensors manufactured by microfabrication techniques adopted from IC processing. In addition, for the mechanical functionality of the instruments and mechanical interaction with tissues, micromanipulation and microactuation techniques are required (Fujimasa, 1996).

1.1.3 A challenge: the active catheter

An instrument that combines many aspects of instrument development for minimally invasive procedures is the *active catheter*. Catheters are long, thin, flexible tube-like instruments that are inserted in the groin, arm or neck, and advanced through bloodvessels to the site that requires investigation or treatment. Typical diameters range from 0.3 to 3 mm, depending on the target vessel. These instruments play an important role in the treatment of circulatory diseases, which are the leading cause of death in the European Union, accounting for over 40% of all deaths (Eurostat, 2005).

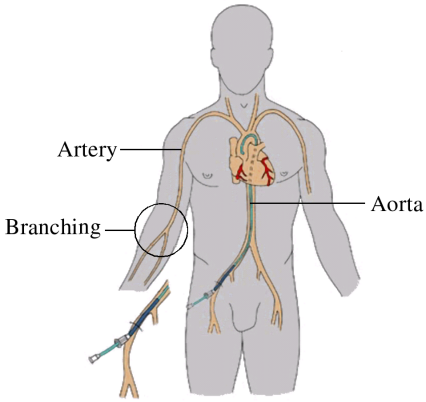


Figure 1.4: *Illustration of cardiac catheterization.*

Although existing catheters in the hands of experienced physicians already save many lives today, they still suffer from clear limitations. The inability to reach the target vessel due to navigation and maneuvering difficulties is the most common problem in cardiovascular catheter treatments, and difficult cases associated with tortuous or difficult anatomical obstacles still challenge even the most experienced physicians (Geske *et al.*, 2005). The placement of the catheter tip must be controlled from sometimes more than 1 meter away, by manipulating the distal end of the catheter, which often is a very challenging task. In addition, in tortuous venous anatomy, there is a risk of vessel wall penetration or other trauma inflicted by the catheter. Even slight damage can cause inflammation (*vasculitis*), which in turn can lead to potentially life-threatening vessel weakening (*aneurysms*) or blockage.

Enhanced catheters with steerable tips aim to alleviate the steering problem, but since these are operated by puller-wires, conventional steerable catheters require a high bending stiffness of the shaft to prevent buckling (Haga and Esashi, 2004). This high stiffness increases the risk of damaging the vessel wall in bent sections significantly. Recognizing this difficulty, research efforts are underway

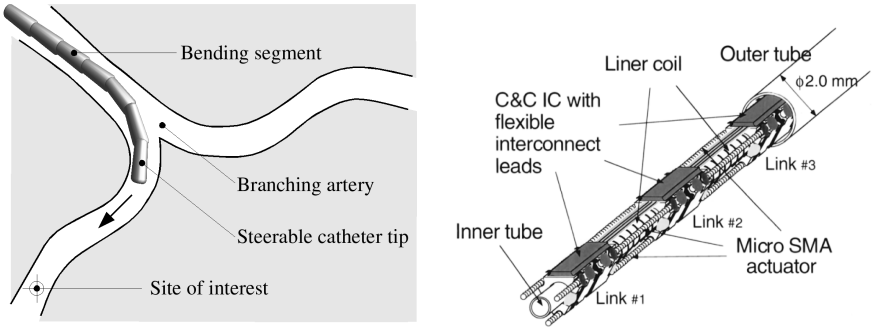


Figure 1.5: *Schematic illustration of an active catheter in use (left) and a design drawing of an active catheter prototype by Park and Esashi (1999b) (right).*

to integrate the capability of controlled steering into catheters, by exploiting microactuation techniques and "steer-by-wire" approaches. This is the essence of active catheters, as illustrated in Figure 1.5, which also shows the design of an active catheter by Park and Esashi (1999b). An overview of prototypes that have been realized in the past 10–15 years is given in Table 1.1. Note that many of these make use of shape memory alloy (SMA) microactuators, which are discussed in more detail in Section 1.2.1. A catheter that can be controlled to bend in any direction would also allow easier insertion, because it can follow the shape of the vessel precisely. Some groups have even managed to integrate tactile sensors along with the microactuators, that are able to detect contact with the vessel wall (e.g. Kaneko *et al.*, 1996). Although many prototypes have already been developed, active catheter design still remains challenging. Further design improvements are needed to achieve the required bending performance, rigidity and reliability, to avoid overheating and to reduce assembly complexity, before active catheters are technically and economically feasible.

The active catheter case illustrates the possibilities of miniaturization, microactuation and microsensors for medical applications. Note that this application shows a remote resemblance to the vision of microsubmarines discussed in Section 1.1.1, but in a more practical form. Controlled bending does not only allow steering and easier insertion, but also enables autonomous propulsion by means of crawling along vessel walls. This makes far more sense than swimming through the bloodstream. In fact, related concepts have been demonstrated already on a larger scale, for devices crawling on the outer surface of a beating heart (Patronik *et al.*, 2004) and for controlled motion of endoscopes through intestines (e.g. Dario *et al.*, 1997, 1999, Kim *et al.*, 2005, Peirs *et al.*, 2001). Conceptually, an active catheter could also be equipped with additional microactuated tools, e.g. grippers to take biopsies. Therefore, miniaturization could actually make Feynman's vision of a "surgeon in a vessel" a reality (Rebello, 2004).

Table 1.1: *Realized active catheter prototypes and related technologies for catheter steering/propulsion. In addition, large number of patents on catheter steering or propulsion has been filed (see e.g. Bakker, 2000, for an overview), but this table only includes demonstrated prototypes. Note that the existence of a patent does not imply that the described invention actually works.*

Publication	Outer diameter	Actuation principle	Remarks
Mineta *et al.* (2001)	1 mm	Shape memory alloy	Etched thin film actuators.
Haga *et al.* (2000)	1.5 mm	Shape memory alloy	Coil actuators.
Kaneko *et al.* (1996), Ohta (2001), Takizawa *et al.* (1999)	1.5 mm	Shape memory alloy	Equipped with tactile sensors to avoid wall contact.
Fukuda *et al.* (1994a,b)	1.7 mm	Shape memory alloy	Wire actuators embedded in polymer matrix.
Park and Esashi (1999a,b)	2 mm	Shape memory alloy	Coil actuators, integrated control circuits.
Chang *et al.* (2002)	2.1 mm	Shape memory alloy	Combined with ultrasound probe at tip.
Dario *et al.* (1991)	2.4 mm	Shape memory alloy	Uses shape memory alloy wires.
Lim *et al.* (1995, 1996)	2.5 mm	Shape memory alloy	Uses externally heated shape memory alloy coils.
Santa *et al.* (1996), Sewa *et al.* (1998)	0.8 mm	Conducting polymer actuator	Biocompatibility unknown, requires ion exchange with environment.
Guo *et al.* (1996)	2 mm	Polymer actuator	Biocompatibility unknown, requires ion exchange with environment.
Ruzzu *et al.* (1998)	2.5 mm	Pneumatic	Tip positioning by balloons.
Ikuta *et al.* (2002, 2003)	3 mm	Hydraulic	Bends only in one direction.
Ernst *et al.* (2005), Faddis *et al.* (2002), Schiemann *et al.* (2004)	-	Magnetic	Magnetic catheter guidance by permanent magnet tip and external magnetic field.
Sendoh *et al.* (1999)	0.6 mm	Magnetic	Propulsion concept: external rotating magnetic field rotates spiral.
Ishiyama *et al.* (2001)	2 mm	Magnetic	Propulsion concept: external rotating magnetic field rotates screw through gel and beef.
Behkam and Sitti (2004)	5 mm	Electromagnetic	Propulsion concept: swimming microrobot using flagellum (similar to E. Coli bacterium).

1.2 Background

In this section, preliminaries on two essential topics for this thesis will be briefly introduced. Section 1.2.1 discusses shape memory alloys, followed by an overview of design optimization in Section 1.2.2.

1.2.1 Shape memory alloys

Microactuation

The dominance of viscous and frictional forces at the microscale (Wautelet, 2001) makes that direct actuation principles are preferred for microactuation (Fearing, 1998). This means that actuator forces are directly transferred to their point of application, without using complex leverage or transformation mechanisms. Direct actuation can be achieved by the use of active materials such as piezoceramics or shape memory alloys (SMAs), and thermal actuation based on phase transformation or thermal expansion. Also electrostatic actuation is frequently used in microsystem applications (Fatikow and Rembold, 1997). Characteristic properties of various options for microactuation are listed in Table 1.2.

For active catheters (Table 1.1) and similar micromanipulator applications that interact with relatively large external forces of the macroscopic environment, SMA microactuators are generally preferred (see e.g. Stevens and Buckner, 2005). The scale of the forces involved requires powerful actuators, and considering the limited available volume, the energy density of the actuation is of prime importance (Fujita and Gabriel, 1991). Furthermore, relatively large deformations are required for steerable catheters, and SMAs offer a practical amount of actuation strain (Table 1.2). In addition, the typical low-frequency operation of active catheters and related minimally invasive tools (< 10 Hz) is compatible with SMA actuation. SMA response speed will improve further with miniaturization and the presence of active cooling by liquids (blood flow). Moreover, their excellent biocompatibility (Kapanen *et al.*, 2002) also makes SMAs excellent candidates for *in vivo* applications. The limited efficiency of SMA actuators is not a critical disadvantage here, as tethered operation is preferred for safety reasons, and therefore power supply is not a problem (Dario *et al.*, 1992). Finally, the relative mechanical simplicity of SMA actuation is a large advantage over approaches that require various additional components, such as relatively complex hydraulic/pneumatic solutions or complex and fragile linear actuators based on displacement accumulation (e.g. De Boer *et al.*, 2004, Shutov *et al.*, 2004, Yeh *et al.*, 2002).

Shape memory alloy characteristics

Many alloys exhibit shape memory behavior. The phenomenon was first discovered 1932, in a AgCd-alloy. Shape memory alloys started to be applied in a variety of applications after the discovery of the effect in NiTi alloys in 1959 (Kauffman and Mayo, 1997). In near-equiatomic composition, NiTi alloys exhibit

Table 1.2: *Comparison of various microactuation principles, sorted on energy density. Properties are indicative and often depend on actual actuator realization and scale. Data is collected from Dario* et al. *(1992), Huber* et al. *(1997), Jacobson* et al. *(1995), James and Rizzoni (2000), Madden* et al. *(2004), Otsuka and Wayman (1998), Pelrine* et al. *(2000).*

Actuator type	Strain [%]	Stress [MPa]	Energy density [J/m^3]	Response speed [Hz]	Efficiency [%]	Comments
Shape memory alloy (NiTi)	1–8	200–500	$6 \cdot 10^6$–> 10^8	1–100	< 5	Low voltages. Hysteresis. Gradual degradation at high strains ($> 4\%$). Faster at smaller scales.
Hydraulic (macroscale)	100	70	10^8	Fast	High	High pressure source, channels, valves required.
Thermal, solid-liquid phase change			$4.7{\cdot}10^6$			Modest temperature changes.
Pneumatic (macroscale)	100	1	10^6			High pressure source, channels, valves required.
Thermal expansion (Al, $\Delta T = 100$ K)	1	78	$4{\cdot}10^5$	1–1000	< 10	Low voltages. Faster at smaller scales. Large temperature changes.
Electrostrictive polymer (P(VDF-TrFE)	4	15	$3{\cdot}10^5$	Fast	-	High voltages.
Electrostatics (ideal)			$2{\cdot}10^5$	Fast	High	Voltage 100 V, gap 0.5 μm.
Dielectric elastomer (Silicone)	32	1.36	$2{\cdot}10^5$	Fast	90	High voltages.
Conducting polymer	2	5–34	$1{\cdot}10^5$	5	Low	Low voltages. Encapsulation typically needed.
Piezoelectric (PZT)	0.2	110	$1{\cdot}10^5$	>1000	>90	High voltages. Brittle in tension.
Natural muscle (human)	> 40	0.35	$7{\cdot}10^4$	-	40	Not an engineering material yet (with few exceptions (e.g Herr and Dennis, 2004)).
Electromagnetic	50	0.1	$2.5{\cdot}10^4$	Fast	>90	Increasingly less attractive at smaller scales.
Magnetostrictive (Terfenol-D)	0.2	70	$2.5{\cdot}10^4$	>1000	60	Requires additional means to locally change magnetic field.
Electrostatics (comb drive)	< 50		$3{\cdot}10^3$	Fast	High	
Piezo-polymer (PVDF)	0.1	4.8	$2.4{\cdot}10^3$	Fast	-	High voltages.
Electrostatics (force array)		0.008	$7{\cdot}10^2$	Fast	High	

very pronounced and relatively stable shape memory behavior, which is suited for practical applications. Also a variety of less expensive Cu-based SMAs exists, but their actuator properties are generally inferior to those of NiTi.

The actuator properties of SMAs originate from the diffusionless solid-state phase transformation that can occur in these materials (Funakubo, 1987, Otsuka and Wayman, 1998). In diffusionless transformations, no atomic bonds are broken or formed, so this transformation merely consists of a distortion of the atomic lattice. A transformation strain is associated with this distortion, which can be used for actuation. The transformation is triggered by changes in temperature, stress state and – in some alloys – magnetic field. In SMAs, one of the phases is a highly symmetric phase (cubic), that can only exist in one configuration. This phase is called the *austenite* phase, and is stable at higher temperatures. The other phase has lower symmetry, and can therefore exist in several *variants*. This *martensite* phase is stable at lower temperatures. Figure 1.6 schematically illustrates the transformation between these two phases. Note that martensite, due to its lower symmetry, can occur in a twinned state and a detwinned state.

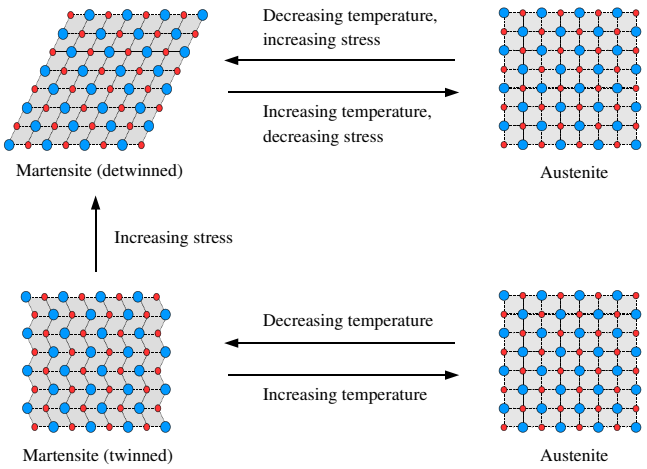


Figure 1.6: *Illustration of martensite-austenite forward and reverse transformations, schematically showing lattices as they typically occur in shape memory alloys.*

The characteristics of the temperature-induced phase transformation are shown in Figure 1.7. Upon heating from the martensite state, the material starts to transform locally into austenite at the austenite starting temperature A^S, and this process continues until at the austenite finish temperature A^F, all material has transformed to austenite. When cooling down, the reverse process takes place, and austenite transforms back to martensite. The temperatures where this trans-

formation starts and finishes are denoted by M^S and M^F, respectively. Note that the forward and reverse transformations do not occur at the same temperatures. The difference or *hysteresis* depends on the alloy and its composition, but typical values are 30–50 K. This hysteresis complicates the use of SMAs as continuously controlled actuators, and they are therefore most often used as discrete (on/off) actuators. However, in nickel-rich NiTi alloys, next to the austenite and martensite states, also a third rhombohedral phase occurs, which is called the *R-phase*. The phase transformation between austenite and this R-phase is known to have a very small hysteresis of approximately 2 K and an excellent resistance to fatigue (Otsuka and Wayman, 1998, Tobushi *et al.*, 1996). These properties make this transformation very attractive for actuation applications.

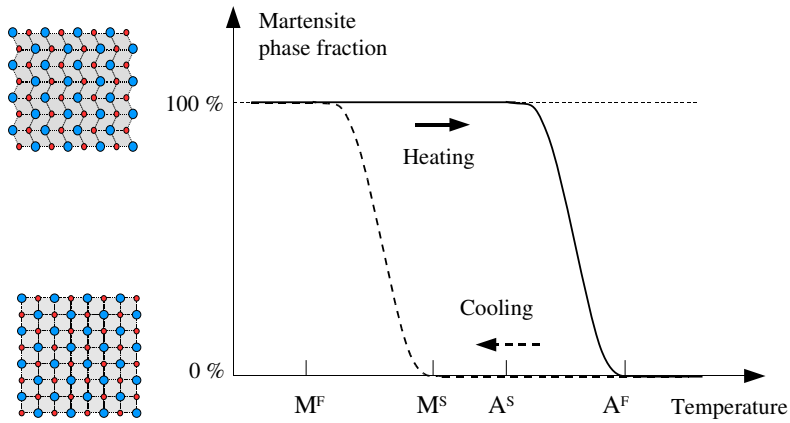


Figure 1.7: *Transformations in shape memory alloys induced by temperature changes, in a stress-free setting. M^S and A^S indicate the temperatures at which the transformations to martensite respectively austenite start, and likewise M^F and A^F indicate the temperatures at which they finish.*

Shape memory alloys have obtained their name from the curious property that these materials appear to "remember" their original shape. After a seemingly permanent deformation of an SMA structure, a temporary increase in temperature is sufficient to restore it back to its original state. This effect is called the one-way shape memory effect, and is illustrated in Figure 1.8 by the curve A-B-C-D-E-A. This behavior is observed when the material is initially in the twinned martensite state. The loading causes detwinning of the initially twinned martensite, which leads to a relatively flat section (B-C) in the loading stress-strain curve. When the load is decreased after further elastic deformation of the twinned martensite (C-D), the unloading curve (D-E) differs from the loading curve (A-B-C-D). Hence, also in the stress-strain space hysteresis is generally observed. After removing

the load, an apparently permanent deformation remains (E), as the detwinned martensite does not revert to the twinned state by itself, but remains in a new equilibrium configuration. By subsequent heating to a temperature above A^F (E-A), the phase transformation to austenite is triggered. As austenite has a cubic lattice which has no variants, this causes the material to recover its initial shape (A). This shape is maintained after returning to the (twinned) martensite state after cooling.

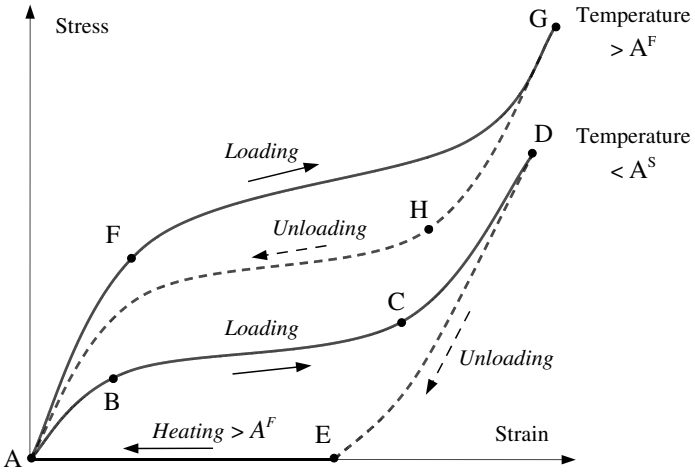


Figure 1.8: *Schematic loading and unloading stress-strain curves for a typical shape memory alloy, at two different temperatures. The one-way shape memory effect is shown by the sequence A-B-C-D-E-A, and the superelastic effect by A-F-G-H-A.*

Another effect, based on the same underlying phase transformation, is observed when the alloy is loaded in the austenite phase. The material in that case is able to elastically recover from quite large deformations (up to 8%). Normally, the elastic strain range for metals is nearly two orders less, at 0.1–0.2%. This remarkable effect is known as superelasticity, or more correctly transformation pseudo-elasticity. Pseudo-elastic behavior is also depicted in Figure 1.8, by the curve A-F-G-H-A. The stressing of the material in its austenite state triggers the stress-induced phase transformation (F), which shows a similar plateau in the stress-strain curve as observed before. At a certain load, all material has transformed into the detwinned martensite state (G). Upon unloading, however, the material transforms back to the austenite phase when the load has sufficiently decreased (H), and finally returns to its original shape when the load is completely removed (A). This ability to elastically recover large strains is often used in appli-

cations, next to the use of SMAs as actuator materials by the temperature-induced transformation. See for example Duerig *et al.* (1999) for an overview of medical applications of SMA superelasticity. An overview of non-medical applications is given by Van Humbeeck (1999). Upcoming applications also include the use of SMA actuators in microsystems, using thin films produced by sputtering techniques. An extensive overview of recent developments in this direction is given by Fu *et al.* (2004).

Shape memory alloy design

Designing SMA structures can be challenging, particularly in a two- or three-dimensional setting. The complex stress-strain-temperature behavior (Figure 1.7, Figure 1.8) of these materials makes design based on intuition alone very difficult. Moreover, SMA actuators are often controlled by direct resistive heating. This means that in the design process, electrical, thermal and mechanical aspects of the structure have to be considered simultaneously. For one-dimensional configurations (SMA wires), the design problem is more or less manageable, and various approaches have been proposed (Liang and Rogers, 1992a, 1997, Troisfontaine *et al.*, 1999). For more complex devices, numerical modeling is generally necessary to provide insight in the performance of a design. To this end, many models have been proposed to simulate the behavior of SMA materials. This is still an active field of research, and since the last review by Birman (1997), many improved or refined models have been proposed. A recent development is the use of micro-macro formulations, where the behavior of individual grains is considered and aspects such as crystallographic texture can be included (see e.g. Jung *et al.*, 2004, Nae *et al.*, 2003). These sophisticated models are however less useful for design purposes, as their complexity and the required computational effort is currently not suited for practical design iterations.

The ability to simulate SMA devices is of great importance to the design process of more complex SMA structures. By a cyclic process of model building/adaptation, simulation and evaluation, computational models can be of assistance in designing SMA structures. However, for an effective design process, manual and trial-and-error based iterative design adaptation (e.g. Büttgenbach *et al.*, 2001) is not the most suitable approach. Instead, use can be made of *design optimization* techniques. These techniques are discussed in more detail in the following section. Although design optimization has proven its value already in a wide variety of structural design problems (see e.g. Maute *et al.*, 1999, for an overview), application to the design of SMA structures has so far been limited to a few cases. Lu *et al.* (2001) have reported design optimization of a flexural actuator equipped with SMA wires, based on analytical modeling, and Birman (1996) optimized a panel with embedded SMA fibers. However, these approaches focus on one-dimensional SMA configurations and cannot deal with more general structures. Kohl and coworkers have considered peak stress reduction of two-dimensional SMA microactuators (e.g. Kohl *et al.*, 2000b, Kohl and Skrobanek,

1998, Skrobanek *et al.*, 1997), using finite element modeling. Although their approach bears some resemblance to an optimization procedure, it is not nearly as versatile as actual design optimization approaches based on the formulation of a formal optimization problem. Clearly, design optimization of SMA structures still is largely unexplored, although it has great potential for this application considering the challenging design problem. The aim of this thesis is therefore to develop and explore design optimization of SMA structures. However, before further discussion of the problem statement of this thesis, the following section first gives a brief overview of existing design optimization techniques.

1.2.2 Design optimization

Optimization-based design process

Design optimization provides a systematic approach to solve design problems (Arora, 2004). With the availability of affordable computer power and capable analysis software, this structured model-based approach has recently gained widespread acceptance in many branches of product development (Saitou *et al.*, 2005). The process starts by formulating the design problem that is considered, defining the model to evaluate designs and the parameters that will be considered as design variables, $\boldsymbol{x}$, and their domain $\mathcal{X}$. Generally, the model will be a numerical model. Subsequently, the design problem can be cast in the form of a formal optimization problem as follows:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \leq 0 \quad i = 1 \ldots n \\ & h_j(\boldsymbol{x}) = 0 \quad j = 1 \ldots m \\ & \boldsymbol{x} \in \mathcal{X} \end{array} \tag{1.1}$$

This problem consists of at least one objective f, i.e. the quantity that is to be minimized or maximized, and optionally a number of constraints $\boldsymbol{g}$ and $\boldsymbol{h}$, i.e. conditions that have to be satisfied. The objective and constraint values are functions of the design variables, and are collectively called the *responses* of the model. Next, explorative studies could be carried out, to gain more insight in the problem, and to possibly even adapt the initial problem formulation based on the findings. A well-posed design optimization problem often includes contradicting requirements between which the optimization process has to find the best compromise.

In order to solve the problem defined by Equation 1.1, an optimization algorithm is used. This algorithm is coupled to the computational model in order to find the optimal design, as illustrated in Figure 1.9. This results in an iterative process that usually requires many evaluations of the model. The efficiency of the process can usually be improved significantly when also design sensitivities, i.e. derivatives of the model responses with respect to the design variables, can be computed. The availability of sensitivity information allows the use of

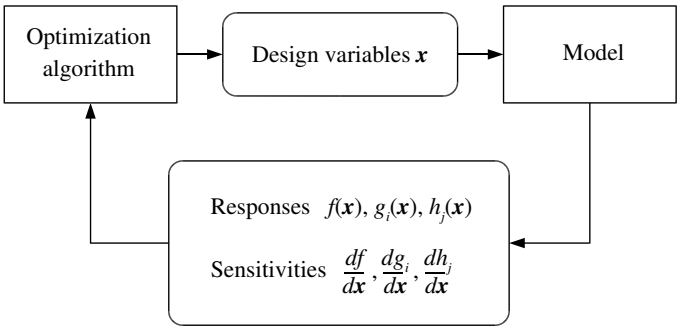


Figure 1.9: *Layout of the cyclic process for solving design optimization problems.*

gradient-based optimization algorithms, that generally have superior convergence characteristics compared to direct methods, that only use the values of the responses. For this reason, design sensitivity analysis has a central role in design optimization research (Haftka and Gürdal, 1992). In case of complicated, time-consuming models, optimization might not be feasible simply because it takes too long. In that case, more efficient simplified models can be used in the optimization process (Papalambros and Wilde, 2000). After an optimal design is obtained from the optimization process, a detailed analysis can be done.

The outlined process allows designers to ask specific design questions, and explore different design concepts in a systematic way. In finding the best design, this approach is far more effective than the traditional "trial-and-error" approach. It is both powerful and versatile, and can be used to simply find a design that "works", or to optimize performance to gain a competitive edge.

Structural optimization

The field of structural optimization covers the application of optimization techniques to the design of structures. Extensive information on structural optimization can, for example, be found in the textbooks by Haftka and Gürdal (1992), Arora (1997) and the review by Maute *et al.* (1999). Among other classifications, structural optimization approaches can be categorized in three categories of increasing versatility: *sizing*, *shape* and *topology* optimization.

Sizing optimization is generally applied to structures composed of members with certain cross-sectional properties, such as trusses, beams, plates or shells. The cross-sectional properties are used as design variables, for example the cross-sectional areas of trusses in a truss structure. The structural layout itself is not modified, only the "sizes" of the members are. Sizing optimization is illustrated in Figure 1.10(a).

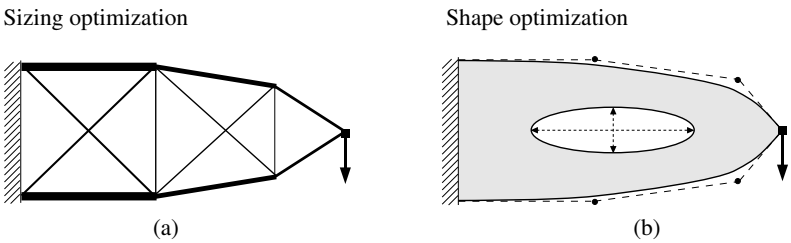


Figure 1.10: *Schematic illustration of sizing (a) and shape optimization (b). In sizing optimization, the cross-sectional areas of the trusses are used as design variables. In shape optimization, the coordinates of the Bézier control points and the dimensions of the elliptic hole are potential design variables.*

A more general approach is given by shape optimization, where the optimal shape of a design is determined. This method requires a geometrical description of the shape of the design, and parameters of this description are used as design variables. For example, for a design defined by Bézier curves, as depicted in Figure 1.10(b), the coordinates of the control points defining the curves could be used as design variables. Although shape optimization in principle allows more extensive modifications of the design than sizing optimization, the layout (topology) of the design remains unchanged. New holes or connections that were not present in the original parameterization can not be introduced during the optimization process, simply because such features are not described by the used design variables.

These limitations of shape optimization are overcome by topology optimization. This is a relatively recent technique, that finds its origin in the pioneering work by Bendsoe and Kikuchi (1988). In topology optimization, no *a priori* assumptions are made about the topological structure of the design. The design variables parameterize the discretized material distribution in a given design domain. This allows practically any conceivable arrangement of material, in contrast to the limited shapes achievable through shape optimization based on predefined geometrical features. Extensive overviews of topology optimization techniques can be found in the reviews by Eschenauer and Olhoff (2001), Hassani and Hinton (1998a,b,c) and the book by Bendsoe and Sigmund (2003).

Topology optimization is regarded to be particularly useful in the initial stages of a design process, since this technique is able to generate promising layouts without any prior assumptions. The accuracy that can be achieved in topology optimization is generally less than that of shape optimization. Constraints on stresses and strains are also easier to incorporate in shape optimization. Therefore, after an initial design has been suggested by topology optimization, shape optimization could be applied to the generated layout in a subsequent detailed design phase.

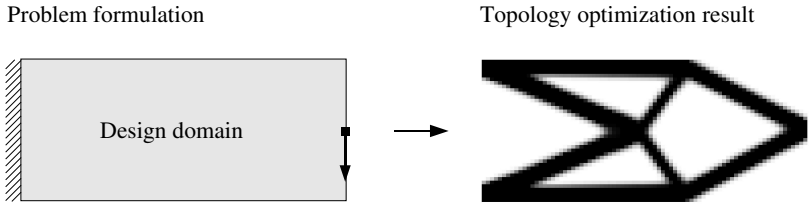


Figure 1.11: *Topology optimization problem definition including boundary and loading conditions (left) and the corresponding result (right) for the stiffest design using 40% of the design domain.*

1.3 Problem statement

After the preceding discussion on the underlying motivation as well as preliminaries on SMAs and design optimization, this section states the objective and scope of the research presented in this thesis.

1.3.1 Aim

The objective of this research is to develop design optimization techniques for the design of SMA structures. This objective is motivated by the challenging nature of SMA design problems, the lack of existing work on SMA design optimization and thirdly the expected benefit of structured and systematic design approaches. In particular, this work is aimed at emerging applications of SMAs as micro- or miniature actuators for minimally invasive applications, such as the active catheter case discussed in Section 1.1.3. The focus is however on the more generic design approaches, rather than on the specific applications themselves. The latter merely serve as carriers that give guidance to the design techniques that are developed. The main question that this research aims to answer is how and to what extent design optimization techniques can support the design of SMA structures in this context.

Sizing optimization has been demonstrated already by several studies on wire-based SMA active structures (e.g. Birman, 1996, Troisfontaine *et al.*, 1999). Therefore, the focus of this thesis is on shape and topology optimization, which so far have not been considered yet for SMA structures. In addition, SMA shape optimization involving uncertainties and sensitivity analysis of SMA computational models is studied.

The approach taken to investigate the merit of design optimization approaches for SMA design problems is to apply the mentioned techniques to representative practical cases. A first requirement therefore is the availability of an analysis model to evaluate the performance of SMA structures. After development of a

suitable model, the various optimization techniques are explored using an SMA active catheter and a miniature SMA gripper as carrier applications.

1.3.2 Scope

As mentioned, potential SMA applications in the field of medical instrumentation for minimally invasive procedures are used as relevant cases to achieve the necessary focus in the development and study of SMA design optimization techniques. The economic feasilibity, or an assessment of chances on clinical acceptance of these applications, such as active catheters, are outside the scope of this research. Moreover, this research is not intended to focus on technicalities, detailed manufacturing procedures and specific issues of a certain design concept. The applications merely serve as means to provide guidance to the design optimization techniques that are developed. In addition, the emphasis of the design problems is on the mechanical performance of the SMA structures, because this aspect is critical for their mechanical functionality.

The second limitation concerns the SMA material considered in this thesis. Because of its suitability for actuation applications and in vivo biomedical use, the focus is on the R-phase transformation. As mentioned earlier, typical of this R-phase transformation is its small hysteresis and its excellent resistance to fatigue (Otsuka and Wayman, 1998, Tobushi *et al.*, 1996). The small hysteresis makes that this transformation can be used for actuation in a narrow temperature range, making it suitable for in vivo applications.

Furthermore, for SMAs, many computational models have been proposed, targeted at various applications, describing various aspects of SMA behavior and having various levels of detail. In the modeling of SMA behavior in this thesis, the emphasis is on simplification, since design optimization itself is a computationally expensive process. The prime objective is to create a practical model that is suited for performing design optimization. This implies that it should be computationally efficient, sufficiently accurate, and able to capture the main characteristics of the material behavior. Subtleties and second-order effects, such as the detailed description of multi-axial and non-proportional loading, are not considered at this stage, for three reasons. Firstly, the existing models that allow inclusion of these effects currently are not suited for design optimization due to their complexity and computational cost. Secondly, the experimental validation of such models, particularly in the case of the R-phase transformation, is currently not conclusive (Raniecki *et al.*, 1999). Thirdly, it turns out that in practical designs, the influence of such higher-order effects is hardly relevant (e.g. Rejzner *et al.*, 2002). Therefore, although these aspects of SMA constitutive behavior are highly interesting, they are outside the scope of this thesis. In addition, since the considered applications generally are not critical with respect to the speed of operation, dynamic aspects are not taken into account either. This also helps to further simplify the modeling. The low requirements the medical applications impose on the dynamic behavior (1 Hz is expected to be sufficient) and the fact

that downscaling increases the bandwidth of SMA actuators justify ignoring the transient behavior in this thesis.

1.4 Outline

The majority of the chapters in this thesis are reproduced from journal or conference articles. This inevitably results in a certain degree of redundancy, but has the advantage that each chapter is self-contained and can be read independently. The structure of this thesis is illustrated by Figure 1.12. The underlying motivation, some preliminaries and the problem statement of this research have been given in the present chapter. Chapters 2 and 3 focus on the modeling of SMA material and devices. A simplified SMA material model, suited for design optimization purposes and aimed at (in vivo) SMA actuator applications, is described first in Chapter 2. Subsequently, Chapter 3 introduces a refined version of this model, and discusses the modeling of a steerable catheter equipped with integrated SMA microactuators.

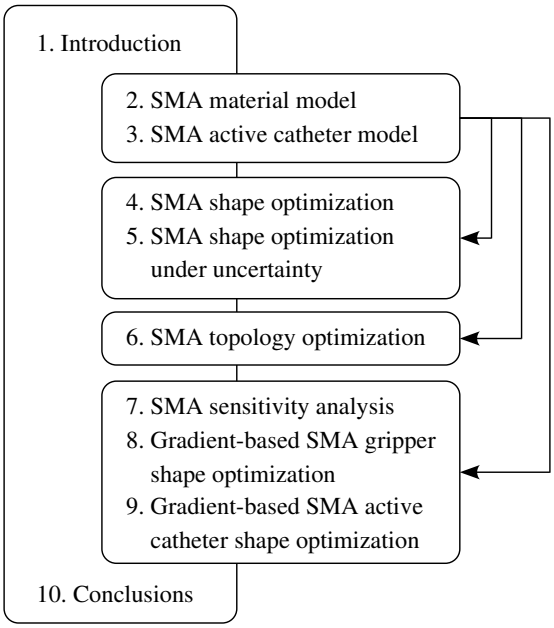


Figure 1.12: *Organization of this thesis.*

The models introduced in Chapters 2 and 3 form the foundation for the chapters dedicated to various aspects of design optimization that follow. A second

pair of chapters, Chapters 4 and 5, focus on the shape optimization of an SMA miniature gripper, which also potentially could be used in surgical procedures. Chapter 4 describes the shape optimization of two versions of this gripper. In Chapter 5, this work is combined with anti-optimization techniques developed by Gurav *et al.* (2003), which allow the efficient consideration of uncertainties in the design optimization.

Although shape optimization is a powerful design tool, it has its limitations due to the fixed layout of the designs that can be generated. The application of the more versatile topology optimization approach to the design of SMA actuators is studied in Chapter 6. This chapter adopts the parameterization technique recently proposed by Yoon and Kim (2005), which helps to overcome various difficulties in the topology optimization of SMA structures.

The 7^{th} chapter discusses various options for the sensitivity analysis of SMA computational models, and the model introduced in Chapter 3 in particular. The efficient computation of design sensitivities can dramatically reduce the computational expenses required for design optimization procedures. The availability of sensitivities allows the use of gradient-based optimization algorithms, which are in most cases superior to direct methods in terms of convergence and efficiency. The structure of the SMA material model and the fact that actuator models generally involve coupled electrical, thermal and mechanical simulations leads to several complications in the computation of the design sensitivities, which are resolved in this chapter. Chapter 8 subsequently demonstrates the effectiveness of gradient-based design optimization, using again the miniature gripper design problem considered before in Chapter 4 and Chapter 5. Chapter 9 does the same for the active catheter model introduced in Chapter 3 and adresses some specific challenges encountered in that case. Finally, conclusions, recommendations and future directions are given in Chapter 10.

Based on: Langelaar, M. and van Keulen, F (2004c). A simple R-phase transformation model for engineering purposes. *Materials Science and Engineering: A - Structural Materials: Properties, Microstructure and Processing.*, 378(1–2) pp. 507–512.

Chapter 2

A Simple R-Phase Transformation Model

For *in vivo* biomedical applications of shape memory alloys, actuators based on the R-phase/austenite transformation as present in certain TiNi alloys are particularly suited. Numerical modeling and optimization can improve the design of such actuators. In this chapter, a one- and three-dimensional constitutive model for this R-phase/austenite transformation is presented. In contrast to previous work, its formulation is particularly directed towards the efficiency in finite element based design optimization, next to its agreement with experimental data. Important features of this model are its simplicity, an explicit stress strain relationship and its path-independence.

2.1 Introduction

The development of medical instrumentation continues to offer opportunities for the application of shape memory alloys. Structures with integrated shape memory alloy actuators can enable further miniaturization of medical instruments and extension of their functionality. The application that motivated the present research is the use of shape memory alloy actuators in catheters. This significantly enhances the ability to position and steer a catheter, as has been demonstrated by various prototypes (Mineta *et al.*, 2002, Park and Esashi, 1999b). A more integrated design can further improve the reliability and manufacturability of these active catheters, but requires the capability to simulate the behavior of two- and three-dimensional structures with integrated actuators.

The R-phase/austenite transformation of TiNi alloys is particularly suited for this application, for three reasons: TiNi alloys are biocompatible, the narrow thermal operating range of the phase transformation does not violate restrictions for *in vivo* applications and its cyclic stability is excellent (Tobushi *et al.*, 1996). However, compared to the martensite/austenite transformation in TiNi, the recoverable strain of approximately 0.5% is clearly smaller.

Because the transformation strain is limited, it is important to exploit it as much as possible. In a multidimensional setting, this is not a trivial task. The combination of computational models with optimization techniques provides a structured process to find the best design. The application of this methodology on the design of structures with integrated shape memory alloy actuators holds the promise for further improvement of existing shape memory alloy devices as well as the discovery of new innovative designs.

The constitutive model may affect the efficiency of design optimization procedures to a large extent. This will be discussed in more detail in Section 2.2. With regard to this aspect, existing practical one-dimensional models (Ikuta *et al.*, 1991, Leclercq *et al.*, 1994, Lexcellent *et al.*, 1994, Tobushi *et al.*, 1992) for the R-phase/austenite transformation have been studied. In Section 2.3 a new model is proposed, which combines a simpler formulation with a better agreement with previously published experimental data, and which is well suited for design optimization. The generalization of this model to a three-dimensional setting is also affected by an emphasis on efficiency, which is why the approach proposed in Section 2.4 differs from the conventional one based on the Von Mises stress (Briggs and Ostrowski, 2002). An example illustrating the practical use of this model is discussed in Section 2.5, followed by some concluding remarks.

2.2 Design optimization considerations

In structural optimization, the best design is searched for by combining mathematical optimization techniques with a computational model parameterized by a set of design variables. This process typically requires a large number of evaluations of the model. The focus of the present work is initially directed towards a finite element model based on shell elements, which is computationally less costly than one based on fully three-dimensional elements. Still, due to the nonlinearities caused by the constitutive behavior and the expected large rotations, an incremental-iterative strategy is required to obtain a solution, which significantly increases the time of each design evaluation. Because the practicality and feasibility of the proposed design optimization approach depends on its computational costs, a computationally cheap constitutive model is preferred, since it directly affects the total cost of the procedure.

A second way the choice of the constitutive model influences the cost of design optimization is through its mathematical nature. Efficient optimization algorithms for computationally intensive models make use of gradient information,

i.e. derivatives of the objective and constraint functions to the design variables, to minimize the number of required evaluations. The computational effort to obtain these derivatives, also known as design sensitivities, depends on various factors. An important factor is whether the model is history-dependent. Models that make use of internal state variables belong to this category. In that case, computation of design sensitivities is considerably more involved (Kleiber *et al.*, 1997). Thus, from a design optimization point of view, next to their computational cost clearly the history-dependence of constitutive models is to be considered.

2.3 R-phase transformation modeling

For the intended applications, the Ti-55.3wt%Ni alloy used in an experimental study by Tobushi *et al.* (1992), which exhibits the R-phase/austenite transformation, is selected as a suitable material. Stress-strain curves from those experiments are shown in Figure 2.1. The focus will be particularly on the pseudo-elastic behavior observed for temperatures between 328 and 343 K, shown in Figure 2.1 by thick solid lines. The loading and unloading stress-strain curves in this temperature range are very similar, which allows approximating them by a single curve, i.e. neglecting the hysteresis. These temperatures are too high for *in vivo*

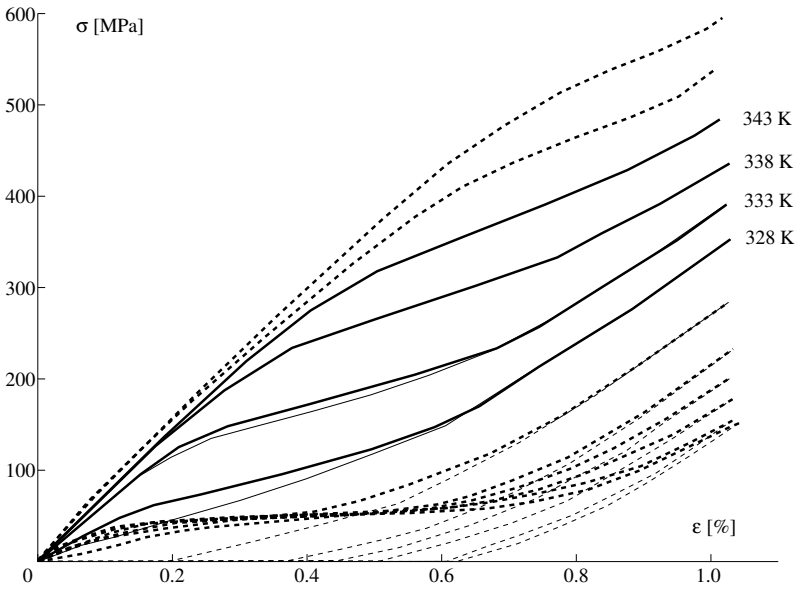


Figure 2.1: *Stress-strain curves at various temperatures from Tobushi* et al. *(1992) of a TiNi alloy. Thin lines are unloading curves.*

applications, but based on the fact that composition and heat treatment can influence transformation temperatures (Sawada *et al.*, 1993, Todoroki, 1990), it is reasonable to assume that shifting of this temperature range is possible.

Neglecting hysteresis is a common assumption in the models that have been published on the R-phase transformation. It turns out not only to be a convenient simplification, but due to this assumption the model also becomes history-independent. This can be understood by the fact that the phase fraction of martensite (or R-phase), which is an internal variable in Tanaka-based phenomenological models (Brinson and Huang, 1996, Tanaka *et al.*, 1986), becomes simply a function of the stress σ and temperature T due to this assumption: $\xi = \xi(\sigma, T)$. The simplifications discussed by Brinson and Huang (1996) result in the following general structure of the one-dimensional constitutive equation:

$$\sigma = E(\xi)(\varepsilon - \varepsilon_{\mathrm{tr}}^{\max}\xi). \tag{2.1}$$

Here ε represents the strain, $E(\xi)$ is the effective stiffness and $\varepsilon_{\mathrm{tr}}^{\max}$ is the maximum possible amount of transformation strain. The effective stiffness depends on the phase fraction because the material is generally a mixture of two phases with different properties. The actual transformation strain is assumed to be proportional to the phase fraction ξ. When hysteresis is ignored and $\xi = \xi(\sigma, T)$, an equation remains from which for a given strain and temperature the stress can be obtained, as required in a finite element model. However, generally the stress can not be expressed explicitly in terms of strain and temperature, and the resulting nonlinear equation has to be solved by a relatively costly iterative numerical procedure.

An example of this situation is found in the model proposed by Ikuta *et al.* (1991), which includes martensite, R-phase and austenite phases. Its basic structure is similar to the described Tanaka-based approach. Hysteresis is neglected for the R-phase transformation, but because its phase fraction depends on stress and temperature through a logistic function, in this model the stress cannot be solved directly.

In contrast, the model proposed by Lexcellent *et al.* (1994) and Leclercq *et al.* (1994) results in an explicit equation to obtain the stress. This model is based on a thermodynamics framework, and has a rather complex mathematical structure. However, with the parameters given by Leclercq *et al.* (1994), this model closely resembles a piecewise linear relationship between stress and strain, at a given temperature. This suggests that as an approximation to this model, in order to further simplify the formulation, piecewise linear functions can give good results. The model given by Tobushi *et al.* (1992) also results in an explicit piecewise linear stress-strain relationship. However, both these models assume that Young's moduli of austenite and R-phase material are equal, although the experimental data suggests otherwise.

In order to improve the agreement with the experimental data and at the same time simplify the model as much as possible, the following piecewise linear

stress-strain-temperature relationship is proposed to describe the one-dimensional constitutive behavior:

$$\sigma = \begin{cases} \varepsilon \leq \varepsilon_1 & : \quad \sigma_0 = E_A \varepsilon, \\ \varepsilon_1 < \varepsilon \leq \varepsilon_2 & : \quad \sigma_1 = E_T(\varepsilon - \varepsilon_1) + \sigma_0(\varepsilon_1), \\ \varepsilon > \varepsilon_2 & : \quad \sigma_2 = E_R(\varepsilon - \varepsilon_2) + \sigma_1(\varepsilon_2). \end{cases} \tag{2.2}$$

Here E_A and E_R are constant parameters, and $\varepsilon_1, \varepsilon_2$ and E_T are linear functions of temperature:

$$\begin{aligned} \varepsilon_1(T) &= K_\varepsilon(T - T_0) + \varepsilon_0, \\ \varepsilon_2(T) &= \varepsilon_1(T) + \Delta, \\ E_T(T) &= K_E(T - T_0) + E_0. \end{aligned} \tag{2.3}$$

In total this model contains eight parameters, $K_\varepsilon, K_E, \varepsilon_0, \Delta, E_0, T_0, E_A$ and E_R, which can be determined by curve fitting. Introducing the parameter T_0 is convenient for shifting the family of curves along the temperature axis. We chose to let ε_2 differ from ε_1 only by a constant, since the experimental data indicated that these parameters depend similarly on temperature.

The stress-strain relations are shown for the case of positive strains, but similar expressions can be formulated for negative strains. Although tension-compression asymmetry (TCA), the phenomenon that the magnitude of the stress inducing the phase transformation is different in tension and compression, has not so clearly been observed in the R-phase/austenite case (Raniecki *et al.*, 1999), in principle such an effect could be incorporated in this model by using different values for $K_\varepsilon, K_E, \varepsilon_0, \Delta$ and E_0 in compression, fitted to the appropriate experimental stress-strain curves.

In the following, TCA is not considered. By least squares curve fitting the model to the four experimental stress-strain curves from Tobushi *et al.* (1992) of $T = 328$–343 K, the eight constants could be determined. They are given in Table 2.1. The resulting curves generated by the model are shown together with the experimental data in Figure 2.2. It shows that with the proposed model a good representation of the experimental data can be constructed.

Note that the existing one-dimensional models (Ikuta *et al.*, 1991, Leclercq *et al.*, 1994, Lexcellent *et al.*, 1994, Tobushi *et al.*, 1992) aim to describe a broader range of phenomena than just the transformation pseudo-elasticity treated here.

Table 2.1: *Parameter values for the proposed model found by curve fitting to experimental data.*

K_ε	2.55×10^{-4} K^{-1}	K_E	619 MPa K^{-1}
ε_0	5.71×10^{-4}	Δ	54.2×10^{-4}
E_0	20.0 GPa	T_0	328 K
E_A	68.9 GPa	E_R	45.6 GPa

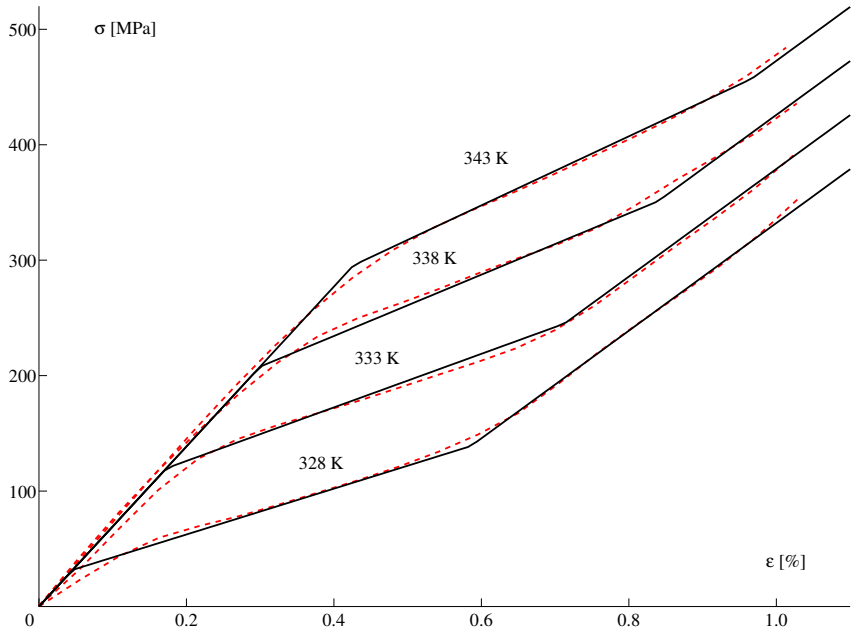


Figure 2.2: *Stress-strain curves at various temperatures from Tobushi* et al. *(1992) of a TiNi alloy (dashed lines) together with curves given by the proposed model (solid lines).*

Focusing on representing this particular aspect allows the proposed model to be both simple and more accurate. It is possible to extend this model to fit the experimental data even better by using additional terms and parameters, but in the current design optimization setting this simple piecewise linear formulation is a good compromise between accuracy and efficiency. It is emphasized that it is only meant to represent the experimentally observed behavior, although it could also be interpreted as a first order approximation of a more complex and general constitutive relation.

2.4 Generalization to 3-D

To generalize the proposed piecewise linear elastic model for transformation pseudo-elasticity to the three-dimensional case, a generalized version of the power-law nonlinear elastic model described by Pedersen and Taylor (1993) and Pedersen (1998) will be used. The fact that strains and particularly the transformation strains remain small allows the assumption that a linear elastic model with a varying Young's modulus can approximate the constitutive behavior. The model

uses a scalar effective strain based on the strain energy, defined by:

$$\varepsilon_e^2 = \{\varepsilon\}^T [C] \{\varepsilon\}. \tag{2.4}$$

Using this effective strain as a scalar strain measure, the secant stress-strain relation is given by:

$$\{\sigma\} = \frac{f(\varepsilon_e)}{\varepsilon_e} E [C] \{\varepsilon\}. \tag{2.5}$$

Here $\{\sigma\} = \{\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{23}, \sigma_{31}\}^T$ and $\{\varepsilon\} = \{\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33}, \gamma_{12}, \gamma_{23}, \gamma_{31}\}^T$ are energetically conjugated stress and strain tensors in vector notation, $[C]$ is the associated dimensionless, constant, symmetric and positive definite constitutive matrix, and E is a reference Young's modulus. The function $f(\varepsilon_e)$ should be chosen such that $f(\varepsilon_e)E$ represents the one-dimensional material behavior. For instance, in power-law nonlinear elasticity as discussed by Pedersen (1998), Pedersen and Taylor (1993), it is given by ε_e^p. In this way, the Young's modulus effectively depends on the value of the scalar strain measure. For $f = \varepsilon_e$ this model reduces to classical linear elasticity. Note that the way the stress and strain components relate to each other is fixed and controlled by the content of the matrix $[C]$, and that only their relative magnitudes are influenced by the term $f(\varepsilon_e)/\varepsilon_e$. Also note that in implementations, the situation at $\varepsilon_e = 0$ should clearly be dealt with properly. In the isotropic case $[C]$ is given by the well-known expression:

$$[C] = \frac{1}{(1-2\nu)(1+\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ \nu & 1-\nu & \nu & 0 & 0 & 0 \\ \nu & \nu & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix}, \tag{2.6}$$

where ν is the Poisson's ratio of the material. For TiNi, $\nu = 0.33$ is a commonly used value (Birman, 1997, Boyd and Lagoudas, 1994).

Differentiation and combination of Equation 2.4 and Equation 2.5 yields the following symmetric constitutive tangent matrix:

$$[K_T] = \left[\frac{E}{\varepsilon_e} [C] \Big(\Psi \{\varepsilon\} \{\varepsilon\}^T [C] + f [I] \Big) \right], \tag{2.7}$$

where $[I]$ is the unity matrix and Ψ is given by:

$$\Psi = \frac{1}{\varepsilon_e^2} \left(f' \varepsilon_e - f \right) \quad \text{with} \quad f' = \frac{df}{d\varepsilon_e}. \tag{2.8}$$

When the strain vector is equal to zero, the tangent matrix reduces to $f'E[C]$.

To generalize the proposed one-dimensional model, for f the discussed piecewise linear stress-strain relation is taken, divided by a reference Young's modulus,

for which the austenite Young's modulus is used ($E = E_A$). In addition, the one-dimensional strain is replaced by the effective strain ε_e, which gives:

$$f = \begin{cases} \varepsilon_e \leq \varepsilon_1 & : \quad f_0 = \varepsilon_e, \\ \varepsilon_1 < \varepsilon_e \leq \varepsilon_2 & : \quad f_1 = \frac{E_T}{E_A}(\varepsilon_e - \varepsilon_1) + f_0(\varepsilon_1), \\ \varepsilon_e > \varepsilon_2 & : \quad f_2 = \frac{E_R}{E_A}(\varepsilon_e - \varepsilon_2) + f_1(\varepsilon_2). \end{cases} \tag{2.9}$$

Note, using E_A as a reference does not mean that the change in stiffness between austenite and R-phase is ignored in the 3-D setting. When f, as given here, is substituted in the stress-strain relation given in Equation 2.2, it is clearly seen that, e.g. when $\varepsilon_e > \varepsilon_2$, the reference value E_A cancels out and E_R remains.

Another often used approach to generalize one-dimensional constitutive laws is to use the Von Mises stress as the equivalent stress (Briggs and Ostrowski, 2002), and this results in formulations which conveniently are quite similar to J_2-plasticity models. These can be further simplified by adopting a Hencky's stress-strain relation (Liang and Rogers, 1992b), justified considering the rather small strains in the current case. That combined with negligence of the hysteresis leads to a model without internal variables, with the associated favorable properties for design sensitivity analysis. However, it turns out that such a formulation leads to a model with a non-symmetric tangent operator and an implicit stress-strain relation, making it computationally much more expensive.

Regarding the multi-dimensional limit stress states at which the stress-induced transformation begins, not much experimental data was available for the R-phase/austenite case until recently. Torsion-tension (compression) tests have been carried out and analyzed by Raniecki *et al.* (1999), and their results seem to indicate higher limit stresses in pure shear loading compared to the Von Mises criterion. The presented model qualitatively agrees with this, giving a 15% higher limit stress in pure shear in comparison to the Von Mises equivalent stress. Neither a Von Mises based approach nor the proposed model includes TCA, but as mentioned previously, Raniecki *et al.* did not clearly observe TCA for this material. In spite of their efforts, the total picture of the transformation limit stress states for the R-phase/austenite transformation is still not completely clear. Moreover, regarding non-proportional loading, no results have been published for this material at all, to the authors' knowledge. A complete validation of the proposed three-dimensional model is therefore currently not possible.

On the other hand, not all aspects of the constitutive behavior are equally important, and second-order effects can often be ignored. For example, Rejzner *et al.* (2002) recently showed that even very pronounced TCA in martensite/austenite TiNi does not significantly affect the moment-curvature relation in bending. This means that for structures that are mainly loaded in bending, as the example presented in the following section, the effect of possible TCA can be neglected. In the design optimization context, aspects that do not significantly change the behavior of a structure are not likely to have an influence on the optimal design either.

2.5 Example

The proposed model has been implemented in a thin shell element as described in detail by Van Keulen and Booij (1996). For the present nonlinear material model numerical integration is used to evaluate the tangential stress resultants and couples and the tangent stiffness.

To demonstrate the proposed material model and to illustrate the opportunity for design optimization, a structure that could act as a gripper will be discussed, modeled by the material model described in this chapter, and using the parameter values as given in Table 2.1. It is a structure made from TiNi plate with a thickness of 1 mm. The undeformed configuration of the top half of this gripper is shown in Figure 2.3. The bottom half is similar, but mirrored in the x, y-plane. The purpose of the active structure is to grasp certain objects by moving the tip C of the gripper in the z-direction.

To induce an internal stress field, which is essential for the working principle of this device, the clamped ends A and B are pinched together. Then, it is assumed that by external heating the temperature of either the top plate (I) or bottom plate (II) can be raised uniformly from the initial 328 K to 338 K. Because the stress-strain relation of the material is temperature-dependent, different equilibrium positions are found, as shown in Figure 2.4. These configurations were found by an incremental-iterative nonlinear finite element analysis of a model consisting of about 1600 shell elements. Because of symmetry, only half the structure needed to be considered. In the closed configuration, a concentrated clamping force of 10 N was added at the tip C, acting in the positive z-direction. In both cases, the stresses and strains in the structure remained in the range covered by the constitutive model. The difference of the tip location in z-direction in the two configurations was 4.5 mm.

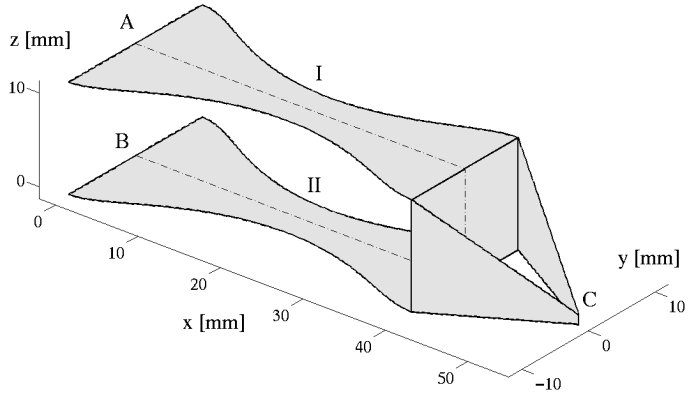


Figure 2.3: *Gripper geometry in undeformed configuration. The plate thickness is 1 mm.*

The geometry of this gripper was initially chosen randomly, and its range of motion was increased by adjusting the shape of Plates I and II by a trial-and-error approach. It is expected that the performance of this device could be further improved by applying a more structured approach, based on design optimization techniques. This topic will be explored further in Chapters 4, 5 and 8.

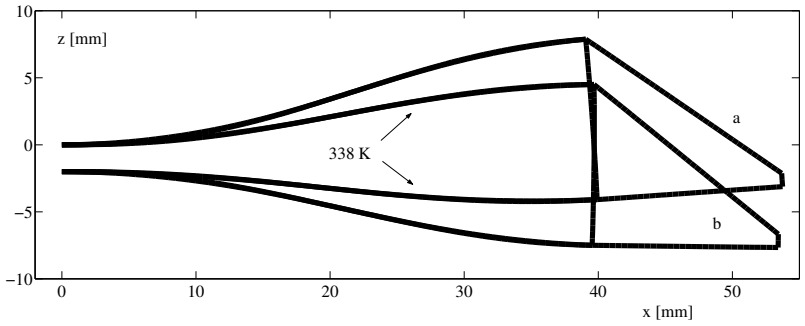


Figure 2.4: *Side view of gripper in open (a) and closed configuration (b) with a clamping force of 10 N.*

2.6 Concluding remarks

In this chapter, a constitutive model for transformation pseudo-elasticity involving the R-phase transformation is proposed. It adequately represents one-dimensional experimental data using a simple, computationally efficient formulation. An important feature of this material model is the absence of internal variables, which considerably reduces the effort required for design sensitivity computations.

Its generalization to a three-dimensional setting has been presented as well. Also here the aspects of computational efficiency and simplicity were given priority. When suitable experimental data becomes available, the accuracy of this three-dimensional version of the present model can be evaluated.

Utilization of structural optimization techniques in the design of shape memory alloy actuators has the potential to produce new, innovative solutions. The present model was developed with specifically the needs of design optimization in mind, and is therefore particularly suitable for exploration of this promising field.

Based on: Langelaar, M. and van Keulen, F. (2004b). Modeling of a shape memory alloy active catheter. In *12th AIAA/ASME/AHS Adaptive Structures Conference, Palm Springs, CA.*

Chapter 3

Modeling of a Shape Memory Alloy Active Catheter

Shape memory alloys (SMAs) have interesting properties for application in adaptive structures, and many researchers have already explored their possibilities. However, the complex behavior of the material makes the development of SMA adaptive structures a challenging task. It is generally accepted that systematic, model-based design approaches and design optimization techniques can be of great assistance in this case. Although some studies on design optimization of relatively simple SMA structures have been published, formal design optimization of more complex SMA devices still requires further exploration.

By considering a typical example, i.e. an active catheter, this chapter aims to provide new insights into and solutions for the problems encountered in the practical application of model-based design approaches to SMA adaptive structures. Active catheters are equipped with integrated micro-actuators that enable controlled bending, which yields enhanced maneuverability compared to conventional catheters. Next to a detailed discussion of an SMA active catheter finite element model, a novel SMA constitutive model is introduced. This model combines an adequate representation of the experimentally observed behavior with computational efficiency. Moreover, its history-independent nature significantly simplifies sensitivity analysis. Due to these features, application of optimization techniques to shape memory alloy adaptive structures becomes a realistic possibility.

3.1 Introduction

Due to their ability to generate relatively large strains and high stresses, shape memory alloys (SMAs) (Duerig *et al.*, 1990, Otsuka and Wayman, 1998) are important materials for adaptive structure design (Chopra, 2002). They can be used both as actuator and as structural material, which allows for designs with complex functions, and yet with a relatively simple structure. However, an obstacle for the widespread application of SMAs is given by the complexity of the behavior of the material. Realistic structures often do not allow a design approach based on intuition. Structured design approaches based on numerical models, combined with design optimization techniques, offer a way to overcome this difficulty.

While the constitutive modeling of shape memory alloys has received and continues to receive much attention, little has been reported on the subject of design optimization of SMA structures. Particularly, detailed finite element models of realistic complexity are rare in this context. Some work has been done on shape optimization of SMA structures through minimization of Von Mises stresses (Fischer *et al.*, 1999, Kohl and Skrobanek, 1998, Skrobanek *et al.*, 1997), but full design optimization based on detailed finite element models including electrical, thermal and mechanical effects has not been reported yet.

The aim of the present research is to fill this gap by exploring this topic. The final goal is to successfully apply gradient-based design optimization techniques to SMA adaptive structure design. In the present work, the focus is on the first step of this approach: the construction of a numerical simulation model suited for subsequent optimization. An active catheter design problem is taken as a typical example of a realistic shape memory alloy adaptive structure.

The new design concept for the active catheter considered in this chapter is introduced in Section 3.2.1. The design is aimed at using a small number of individual parts. In this concept, the shape memory alloy is not used for just one-dimensional actuation. This makes that it serves as a good example of a complicated, highly integrated shape memory alloy adaptive structure. The SMA material chosen for this application is a nickel-rich Ni-Ti alloy, which exhibits the R-phase transformation. Section 3.2.2 will elaborate on the motivation for this material selection, which is closely linked to restrictions imposed by the *in vivo* operating conditions of the catheter.

The third section focuses on the constitutive modeling of the considered material, and a new material model is introduced. Next to the sufficiently accurate representation of experimental data, its formulation is particularly aimed at its suitability for design optimization, as highlighted in the final part of this section.

The subject of the fourth section of this chapter is the finite element modeling of the active catheter structure. In this section, various aspects of the multidisciplinary finite element analysis, which consists of an electrical, thermal and mechanical component, are considered. Also, a subsection is devoted to the way symmetries in the design are exploited, which leads to unconventional constraints in the mechanical analysis. Subsequently, the techniques used to solve the re-

sulting nonlinear problem robustly and efficiently are discussed, followed by the presentation and discussion of results of the finite element simulation.

In the fifth and final section, conclusions are formulated and applications and extensions of the presented work are discussed. Some rather detailed but nonetheless important derivations related to the constitutive modeling and the symmetry constraint equations can be found in Appendices A to D at the end of this thesis.

3.2 Active catheter

3.2.1 Design concept

The continuing popularity of minimally invasive procedures in modern medicine leads to a growing demand for miniaturized medical instruments with added functionality. This trend is exemplified by efforts to develop *active catheters*, catheters equipped with actuators that enable bending to improve their maneuverability. In recent years, technologically advanced prototypes of such active catheters have been created by a number of researchers (Chang *et al.*, 2000, Haga *et al.*, 2000, Park and Esashi, 1999b), demonstrating the feasibility of the concept. All of them make use of SMA actuation elements, because these can deliver large actuation strains and/or stresses while the designs remain relatively simple. However, the prototypes presented in literature typically consist of a large number of individual parts that are connected together. This increases the complexity of the device and makes the assembly process involved and costly. Moreover, the large number of interconnections is likely to decrease reliability. In contrast to the published prototypes, the new active catheter design concept presented in this chapter aims at minimizing the number of individual parts, to overcome these disadvantages. The concept itself originates from manufacturability considerations: the proposed manufacturing method is to laser-cut or etch a pattern out of small-diameter shape memory alloy tubing. This procedure is currently routinely used for stents (e.g. Dickson *et al.*, 2002), but can be used for other structures as well. The pattern used in the present concept enables active bending in two orthogonal directions, and by combining these, the catheter can bend in any direction. The resulting structure is then subjected to a macroscopic strain, for example by applying a tensile axial load. This gives a deformation that allows the placement of *spacers*, which keep the structure in a partly deformed state. The internal stresses that are induced by these spacers are essential for actuation by means of the shape memory effect. A schematic illustration of part of the structure together with the spacers is shown in Figure 3.1. The fabrication procedure based on laser cutting is illustrated in Figure 3.2.

Finally, electrodes are attached to specific locations on the structure, for example by wrapping it with a thin flexible substrate with integrated wiring. It is even conceivable to integrate the spacers in this wrapping, thereby combining functions of wrapping fixation and actuator prestraining. By applying voltages to

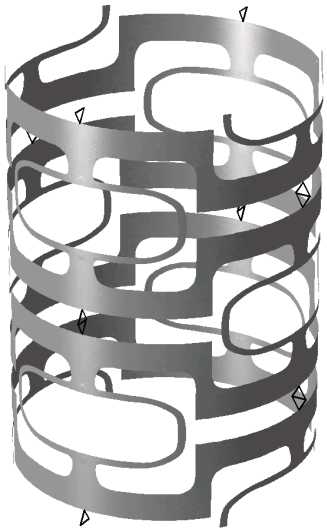


Figure 3.1: *Illustration of a segment of the proposed active catheter concept based on a small-diameter SMA tube patterned by laser cutting. The triangles schematically represent spacers. Tube wall thickness is not shown in the figure.*

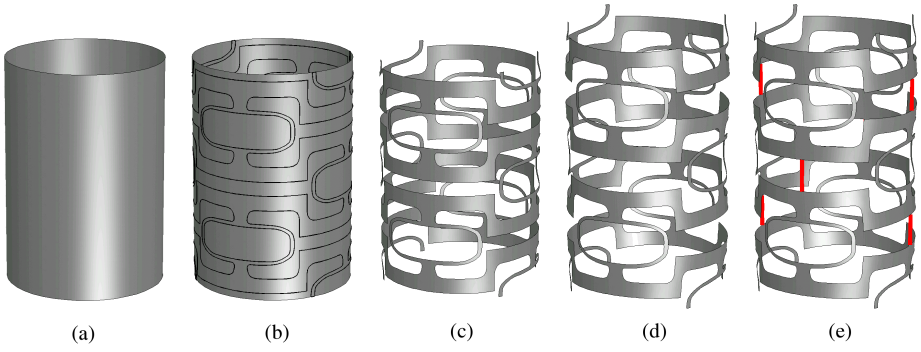


Figure 3.2: *Proposed fabrication of the new active catheter concept using laser cutting from a small diameter SMA tube (a). After laser cutting (b–c), the structure is stretched (d) and spacers are applied (e), in order to generate internal stresses. A full color version is given by Figure G.2 on page 250.*

the electrodes, electrical currents are generated in the structure, which produce heat due to the Joule effect. This generated heat leads to a local change in temperature, that locally influences the constitutive behavior of the SMA. This results in a redistribution of strains and a different equilibrium configuration. Given the right design, an active catheter can be constructed in this way, that bends locally under the influence of an applied voltage.

Although many other interesting sub-problems of active catheter design can be identified (e.g. sensor integration, control, isolation), the focus in this work is primarily on the basic structure of this conceptual active catheter. Note that the present design no longer consists of many individual parts with a single function that can be considered separately. Instead, many functions are integrated in a single structure, and the electrical, thermal and mechanical properties of the material combined with the specific geometry all simultaneously determine the total performance. Design of such a structure using intuition alone is very difficult, because of its integrated nature. Therefore, effective model-based design methods are essential for the development and realization of this concept. Results of finite element calculations, as presented in Section 3.4.4, will further illustrate the functioning of this design concept.

3.2.2 Material selection

Not all shape memory alloys are suited for use in active catheter applications as considered in this chapter. Since the catheter is inserted into blood vessels, an important restriction is that only a limited temperature range can be used, because most tissues are damaged by extended exposure to elevated temperatures.

An exact upper limit on the working temperature is difficult to specify, because the resulting damage depends on many factors, such as the amount of active convection provided by the blood flow, the type of tissue, and the duration and power of the exposure. A certain amount of local tissue degradation can also be tolerated, since due to the self-healing capabilities of the body it will not have any long-term effect.

In literature, various maximum tolerable temperatures are mentioned for the situation without active convection. Most publications are concerned with burning of skin tissue. Because the outer layer of skin, the epidermis, is composed of the same so-called squamous epithelial cells as the lining of blood vessels, it is assumed here, that these published findings are relevant for an assessment of blood vessel wall damage by high temperatures as well. Herzog *et al.* (1989) give 52°C as the contact temperature above which burns occur. Ng and Chua (2002a) mention 44°C as a starting temperature at which tissue degradation occurs with increasing severity. But to place this remark in some perspective, the severity of the damage should be considered as well. In another paper (Ng and Chua, 2002b), Ng and Chua also compare five published damage models. From these, it can be derived that for tissues held at 50°C, over 7 minutes are required to

inflict a first degree burn. A first degree burn causes swelling and redness of skin, and is comparable to a sunburn. It heals within a few days. At 49°C, this period required for a first degree burn is over 8 days, whereas at 60°C, with some variation, the models predict times around half a second. Based on these indications, in the present work an upper limit on the local temperature of the active catheter structure of 49°C (322 K) is used. Note that in these considerations, the effect of active convection through blood flow is not accounted for. This makes that this upper working temperature is still a conservative temperature limit.

The lower operating temperature limit of the catheter is imposed by the body temperature. This gives a temperature range of 37 to 49°C (310 to 322 K), which is too small for most shape memory alloys to be useful. The reason is that most SMAs exhibit *hysteresis*, meaning that the transformation temperatures for the forward and reverse transformation are different. Typically this hysteresis is 30 K, often even more. In practice this would mean that even when starting from 37°C a forward transformation still might be possible upon heating, a reverse transformation would not occur when the material returns to body temperature. This behavior is clearly not desirable for an active catheter.

A solution for this problem is provided by the R-phase transformation, which is a specific type of martensite transformation that occurs in certain nickel-rich Ni-Ti alloys. It is characterized by a very low hysteresis (typically 2 K) and has excellent cyclic stability (Tobushi *et al.*, 1996) and good biocompatibility (Duerig, 2002). However, the standard martensite/austenite transformation in Ni-Ti typically allows for 5 to 7% transformation strain (for a few cycles). This R-phase transformation in contrast only offers approximately 0.8%. Because the transformation strain is not very large in the R-phase case, it is important to exploit it as much as possible by carefully designing the structure of the actuator. This is possible by adopting a model-based design approach as proposed in this chapter. A key component in a computation model for SMA structures is the material model, and computational modeling of the R-phase transformation is the subject of the next section.

3.3 Shape memory alloy constitutive modeling

The formulation of a new constitutive model for R-phase transformation pseudo-elasticity in Ni-Ti is discussed here. First, the one-dimensional case is considered. Subsequently, the generalization of this model to the three-dimensional and plane stress case is presented.

3.3.1 One-dimensional case

For the intended active catheter and other *in vivo* biomedical applications, the Ti-55.3wt%Ni alloy used in an experimental study by Tobushi *et al.* (1992), which exhibits the R-phase/austenite transformation, is selected as a suitable material.

Stress-strain curves from those experiments are shown in Figure 3.3. The focus will be particularly on the pseudo-elastic behavior observed for temperatures between 328 and 343 K, shown in Figure 3.3 by thick solid lines. The loading and unloading stress-strain curves in this temperature range are very similar, which allows approximating them by a single curve, i.e. neglecting the hysteresis. This temperature range is too high for *in vivo* applications, but because transformation temperatures can be influenced by composition and heat treatment (Sawada *et al.*, 1993, Todoroki, 1990), shifting them to lower temperatures is possible.

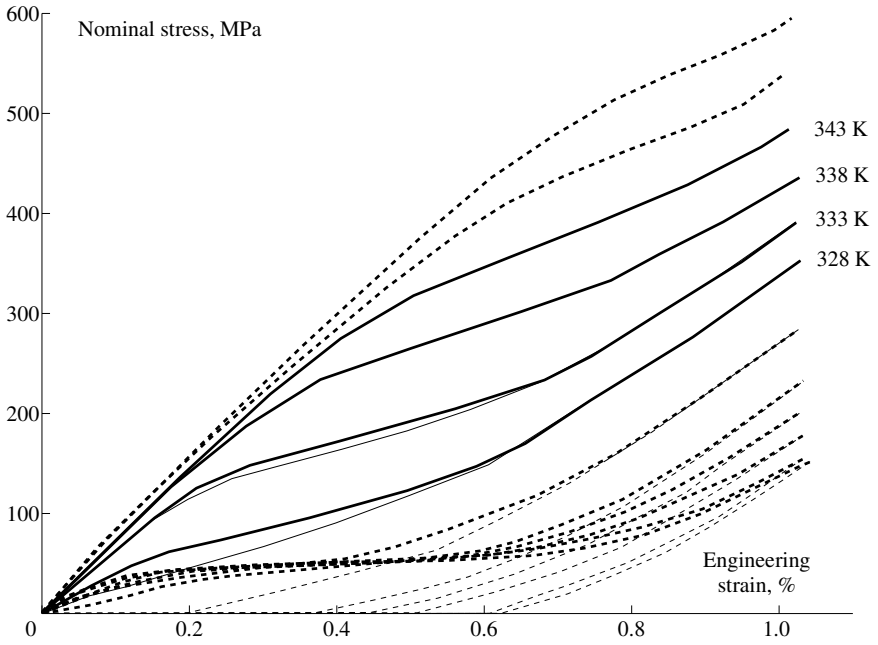


Figure 3.3: *Stress-strain curves at various temperatures from Tobushi* et al. *(1992) of a Ni-Ti alloy. Thin lines are unloading curves.*

For a review of R-phase transformation models previously presented in literature, and a more elaborate argumentation for the present one-dimensional model, see Chapter 2. An approximation used in the formulation of this model is that in the temperature range of interest, the hysteresis is neglected, since the small hysteresis of the R-phase transformation allows for this simplification. As a result, the model no longer requires internal variables and becomes history-independent. This significantly simplifies sensitivity analysis (Kleiber *et al.*, 1997), which increases the practical applicability of the intended model-based design optimization approach.

It was found in Chapter 2, that the following piecewise linear relation can

represent the experimental stress-strain curves sufficiently accurate:

$$\sigma = \begin{cases} \varepsilon \leq \varepsilon_1 & : \quad \sigma_0 = E_A \varepsilon, \\ \varepsilon_1 < \varepsilon \leq \varepsilon_2 & : \quad \sigma_1 = E_T(\varepsilon - \varepsilon_1) + \sigma_0(\varepsilon_1), \\ \varepsilon > \varepsilon_2 & : \quad \sigma_2 = E_R(\varepsilon - \varepsilon_2) + \sigma_1(\varepsilon_2), \end{cases} \tag{3.1}$$

where $\varepsilon_1, \varepsilon_2$ and E_T are linear functions of temperature:

$$\begin{aligned} \varepsilon_1(T) &= K_\varepsilon(T - T_0) + \varepsilon_0, \\ \varepsilon_2(T) &= \varepsilon_1(T) + \Delta, \\ E_T(T) &= K_E(T - T_0) + E_0, \end{aligned} \tag{3.2}$$

and E_A and E_R are constant parameters. These relations are only valid for positive strains, but similar relations can be formulated for negative strains. ε_1 and ε_2 are the transition strains at which the R-phase transformation starts and finishes. E_T is the apparent Young's modulus $d\sigma/d\varepsilon$ during the phase transition. By curve fitting, the parameters defining the material behavior can be determined.

Before the actual fitting, a conversion is applied to the experimental data. The finite element implementation is based on the Green-Lagrange strain $\varepsilon = \varepsilon_{GL}$ and the Second Piola-Kirchhoff stress $\sigma = \sigma_{PK2}$. The one-dimensional experimental data is given in terms of nominal stress σ_{nom} and engineering strain ε_{eng}. The equivalent Green-Lagrange strain and Second Piola-Kirchhoff stress can be derived from these quantities by:

$$\varepsilon_{GL} = \frac{1}{2}\varepsilon_{eng}^2 + \varepsilon_{eng}, \tag{3.3}$$

$$\sigma_{PK2} = \frac{\sigma_{nom}}{1 + \varepsilon_{eng}}. \tag{3.4}$$

The converted experimental data is used in the least squares fitting process. The parameter values that were subsequently obtained are listed in Table 3.1. Note, in the final simulation, $T_0 = 310$ K is used.

3.3.2 Three-dimensional case

The presented one-dimensional model can be generalized to three dimensions in various ways, by defining a link between the one- and three-dimensional case.

Table 3.1: *Parameter values for the proposed model found by curve fitting to experimental data.*

K_ε	2.55×10^{-4} K^{-1}	K_E	619 MPa K^{-1}
ε_0	5.71×10^{-4}	Δ	54.2×10^{-4}
E_0	20.0 GPa	T_0	328 K
E_A	68.9 GPa	E_R	45.6 GPa

The intended application of this model in design optimization procedures has been an important factor that has guided the choices made for the formulation applied here. The generalization preferably conserves the history-independence of the model in order to maintain a relatively simple sensitivity analysis. Also a symmetric tangent operator is preferred, to keep the computational costs low in comparison to a non-symmetric formulation, and to allow the use of standard solvers.

A simple and efficient approach has been presented in Chapter 2. There the link to the one-dimensional model was provided by a scalar strain definition related to strain energy. However, by this choice the fact that the R-phase transformation is pressure-insensitive and the associated transformation strain has no volumetric component (Bhattacharya and Kohn, 1996) was not accounted for. The model presented here instead uses an effective strain based on the distortional strain energy. Starting point for the derivation is Hooke's law in an isotropic three-dimensional setting. In Mandel notation, it can be expressed as:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sqrt{2}\sigma_{xy} \\ \sqrt{2}\sigma_{yz} \\ \sqrt{2}\sigma_{zx} \end{Bmatrix} = \left(K \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \dots \right. \\ \left. 2G \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right) \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \sqrt{2}\varepsilon_{xy} \\ \sqrt{2}\varepsilon_{yz} \\ \sqrt{2}\varepsilon_{zx} \end{Bmatrix}, \tag{3.5}$$

which is equivalent to, in short:

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2G\mathbf{G})\boldsymbol{\varepsilon}. \tag{3.6}$$

The strain energy Π_ε is associated with the tensor product of stress and strain. In vector notation, this becomes

$$\Pi_\varepsilon \cong \frac{1}{2}\boldsymbol{\sigma}^T\boldsymbol{\varepsilon} = \frac{1}{2}\boldsymbol{\varepsilon}^T(K\mathbf{K} + 2G\mathbf{G})\boldsymbol{\varepsilon} = \frac{1}{2}K\boldsymbol{\varepsilon}^T\mathbf{K}\boldsymbol{\varepsilon} + G\boldsymbol{\varepsilon}^T\mathbf{G}\boldsymbol{\varepsilon}. \tag{3.7}$$

The expression $\boldsymbol{\varepsilon}^T\mathbf{G}\boldsymbol{\varepsilon}$ is associated with the distortional part of the strain energy, $\boldsymbol{\varepsilon}^T\mathbf{K}\boldsymbol{\varepsilon}$ with the volumetric part. Based on this, a distortion energy related effective strain measure is defined as

$$\varepsilon_e^2 = \frac{2}{3}\boldsymbol{\varepsilon}^T\mathbf{G}\boldsymbol{\varepsilon}. \tag{3.8}$$

The factor 2/3 is included to make the energy-conjugated effective stress measure equal to the Von Mises stress, which is convenient. The scalar effective strain ε_e

defined in this way is an invariant of the strain tensor, which means its value does not depend on the coordinate system used to describe the strain components. This is an important requirement for a meaningful effective strain definition.

In order to extend Hooke's law to include the isochoric transformation strain $\boldsymbol{\varepsilon}_{tr}$, it is assumed that the transformation strain, which is a purely distortional strain (as it does not have a volumetric component), is proportional to the distortional part of the elastic strain. This assumption is plausible as it minimizes the total strain energy. The resulting stress-strain relation is given by:

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2G\mathbf{G})(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_{tr}) \quad \text{where} \quad \boldsymbol{\varepsilon}_{tr} = \beta\mathbf{G}\boldsymbol{\varepsilon}. \tag{3.9}$$

The symbol β is introduced here as a scaling factor. Accounting for the transformation strain $\boldsymbol{\varepsilon}_{tr}$ in this formulation is equivalent to allowing the shear modulus G of the material to change:

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2G\mathbf{G})(\boldsymbol{\varepsilon} - \kappa\mathbf{G}\boldsymbol{\varepsilon}) = (K\mathbf{K} + 2\hat{G}\mathbf{G})\boldsymbol{\varepsilon}. \tag{3.10}$$

Here $\hat{G} = G(1-\beta)$, and use has been made of the fact that $\mathbf{KG} = 0$ and $\mathbf{GG} = \mathbf{G}$.

In the temperature range of interest, the material is completely in the austenite phase when unloaded. Until the transformation begins, the behavior of the material during loading can be described exactly by Hooke's law, by just taking $\hat{G} = G$. As the transformation proceeds, the transformation strain is accounted for by adjusting this "effective" shear modulus $\hat{G}$. This formulation makes that the modification with respect to a linear elastic model is minimal. From this point onward this effective shear modulus $\hat{G}$ is used, and in order not to unnecessarily clutter the notation, the caret is omitted.

To complete the formulation of the stress-strain relation, the way the effective shear modulus depends on the effective strain has to be specified. This is done using the previously described piecewise linear one-dimensional stress-strain relation (Equation 3.1), which provides a relation $\sigma_{xx} = f(\varepsilon_{xx})$. In case of a tensile test, it can be derived that the proposed three-dimensional formulation yields:

$$\varepsilon_{yy} = \varepsilon_{zz} = \frac{2G - 3K}{2G + 6K}\varepsilon_{xx}, \tag{3.11}$$

$$\varepsilon_e = \frac{3K}{G + 3K}\varepsilon_{xx} \quad \Rightarrow \quad \varepsilon_{xx} = \frac{G + 3K}{3K}\varepsilon_e, \tag{3.12}$$

$$\sigma_{xx} = \frac{9GK}{G + 3K}\varepsilon_{xx} = 3G\varepsilon_e. \tag{3.13}$$

The last expression can be combined with Equation 3.1, which gives an equation from which the relation between the effective shear modulus and the effective strain can be found:

$$\sigma_{xx} = f(\varepsilon_{xx}) = f\left(\frac{G + 3K}{3K}\varepsilon_e\right) = 3G\varepsilon_e. \tag{3.14}$$

At a given temperature, the one-dimensional model gives a linear relationship between the uniaxial stress and strain component: its general form is $\sigma_{xx} = A\varepsilon_{xx} + B$. This leads to:

$$A\frac{G+3K}{3K}\varepsilon_e + B = 3G\varepsilon_e \quad \Rightarrow \quad G = \frac{3K}{9K-A}\left(A + \frac{B}{\varepsilon_e}\right). \tag{3.15}$$

Because the parameters A and B can be derived from the formulation of the one-dimensional model, this final expression relates the shear modulus directly to the effective strain.

To use this model in a finite element implementation, next to the stress-strain relation also the tangent operator is required. This operator can be found by differentiating the stress given by Equation 3.10 with respect to the strain. Details on its derivation can be found in Appendix A. The resulting tangent operator turns out to be symmetric.

3.3.3 Plane stress case

In the current setting of modeling an active catheter, it is attractive to use shell elements instead of solid elements. Particularly thin sections that are mainly loaded in bending can be modeled more efficiently by shells. In order to use this constitutive model in a thin shell element, a formulation for the plane stress case has to be derived. Using the general 3-D formulation, it can be derived that in plane stress, the transverse strain ε_{zz} is related to the in-plane strain components ε_{xx} and ε_{yy} by:

$$\varepsilon_{zz} = \frac{2G-3K}{4G+3K}(\varepsilon_{xx}+\varepsilon_{yy}) = \alpha(\varepsilon_{xx}+\varepsilon_{yy}), \tag{3.16}$$

where the symbol α represents $(2G-3K)/(4G+3K)$, to reduce the complexity of the expressions. Note that α depends on G. Using this relation, the transverse strain component can be eliminated from the stress-strain equation (Equation 3.5), leading to:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sqrt{2}\sigma_{xy} \end{Bmatrix} = \left(K \begin{bmatrix} 1+\alpha & 1+\alpha & 0 \\ 1+\alpha & 1+\alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} + \ldots \right. \\ \left. \frac{2G}{3} \begin{bmatrix} 2-\alpha & -(1+\alpha) & 0 \\ -(1+\alpha) & 2-\alpha & 0 \\ 0 & 0 & 3 \end{bmatrix} \right) \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix}. \tag{3.17}$$

The same can be done for the effective strain definition, which becomes:

$$\varepsilon_e^2 = \frac{4}{9}(\alpha^2-\alpha+1)(\varepsilon_{xx}^2+\varepsilon_{yy}^2) + \frac{4}{9}(2\alpha^2-2\alpha-1)\varepsilon_{xx}\varepsilon_{yy} + \frac{4}{3}\varepsilon_{xy}^2. \tag{3.18}$$

This indicates a slight complication: in the 3-D case, the effective strain only depended on the strain components (Equation 3.8). In the plane stress case, however, it turns out that by elimination of the transverse strain the effective strain also has become a function of α, which in turn depends on the shear modulus G. It turns out, that no convenient explicit expression can be found to express G as a function of ε_e, as in the 3-D case. The way the relation to the one-dimensional model is constructed is the same as discussed previously for the 3-D case, so Equation 3.15 still holds. Combining Equation 3.15, Equation 3.18 and the definition of α gives a complex equation in G, which can be solved numerically for a given temperature and strain situation. A Newton-iteration scheme is used to obtain the solution. Convergence is robust and usually requires only 3 to 4 iterations. Details on the equations can be found in Appendix B.

Also in the plane stress case a tangent operator is required in the finite element implementation. There are two ways to derive it: the first is to take the tangent operator of the three-dimensional case, and reduce it to a plane stress setting by eliminating the transverse strain component ε_{zz} from the equations. The second way to derive the tangent operator is to start from the stress-strain relation in the plane stress case (Equation 3.17) and differentiate it with respect to the strain. However, as in this case there exists no explicit relation between the shear modulus and the (effective) strain, use has to be made of implicit differentiation, which makes this approach slightly more cumbersome than the first one. The details of these derivations and expressions for the resulting tangent operators can be found in Appendix C.

Naturally, the resulting tangent operators are the same in both cases, regardless of the derivation. They are symmetric in case of a finite element formulation based on the stress vector $(\sigma_{xx}, \sigma_{yy}, \sigma_{xy})$ and strain vector given by $(\varepsilon_{xx}, \varepsilon_{yy}, \gamma_{xy})$. In fact, that is a formulation commonly used in implementations.

Both in the 3-D and the plane stress case, the stress-strain relations have been checked against the one-dimensional case and the tangent operators have been verified using global finite differences.

3.3.4 Discussion

The present model is certainly not the first model aimed at describing the constitutive behavior of a shape memory alloy. Neither does the proposed model cover effects such as the asymmetry of the material behavior in tension and compression, the dependence on a specific texture or the complex response to non-proportional loading. Partly this is because the experimental observations presently collected for these cases are not sufficiently conclusive. But another reason is that the purpose of this model is not to provide an exact description of the constitutive behavior in the finest possible detail, or to test certain theories regarding the nature of the material. Instead, the present model is formulated with the intent to be suited for and useful in an engineering and design optimization context. This

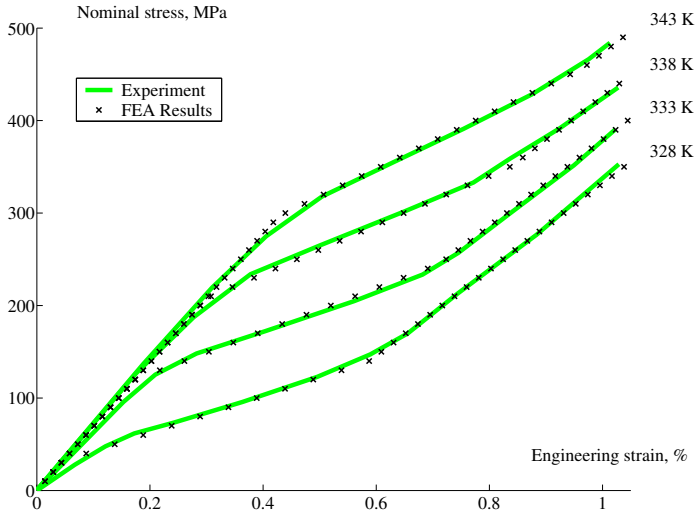


Figure 3.4: *Experimental stress-strain curves at various temperatures (Tobushi* et al.*, 1992) together with finite element results based on the proposed SMA material model.*

is achieved in this case by inclusion of the main characteristics of the R-phase pseudo-elastic response. The present model provides good correspondence with one-dimensional experimental data and accounts for the fact that the transformation strain is isochoric. The agreement with one-dimensional experimental data is illustrated in Figure 3.4, which shows results of a tensile test simulation using plate elements.

In comparison to the model presented earlier in Chapter 2, the present model requires iterations in the plane stress case. This is a consequence of the choice to include the isochoric transformation strain condition. It makes the present model more accurate, but also more expensive. Having these two models in fact can offer an advantage in an optimization setting, as it allows a trade-off between cost and accuracy. This can be exploited in multi-fidelity optimization approaches. Both models are history-independent, and therefore the sensitivity analysis is significantly less involved compared to the history-dependent case (Kleiber *et al.*, 1997). This, together with their relative simplicity and low computational costs, makes both these models well suited for use in design optimization of adaptive SMA structures.

The increased complexity of the present model in the plane stress case due to the need for internal Newton iterations can be reduced in several ways. For instance, when a slight reduction in accuracy is acceptable, a response surface method could be used to construct an approximation of the constitutive relations prior to the actual finite element computation. The shear modulus and effective

strain can then be evaluated by means of interpolation instead of iterations. It is however conceivable that such an approximate approach leads to convergence problems. In that case the interpolated data can be used as a starting point for the iterative process, which is likely to reduce the number of required iterations.

3.4 Finite element model

3.4.1 Electro-thermo-mechanical simulation

For simulation of active catheter concepts, in this work only steady state configurations are considered. To evaluate the capability of a design to bend, the equilibrium configuration in full bending is of interest, therefore the dynamics of the problem are not considered here. In the intended active catheter application, the speed of operation is also less important compared to the range of motion. Moreover, considering only steady-state processes keeps the simulation relatively simple as compared to a full dynamics simulation. With regard to the feasibility of design optimization, it is strongly preferred to not adopt unnecessary complicated modeling approaches.

The mechanical properties of SMAs are influenced by temperature and stress state. In the present design concepts, electrical resistive heating provides a local heat source. Evaluation of the behavior of an active catheter design therefore requires an electrical, thermal and mechanical simulation. The electrical simulation provides a dissipation density field which enters the thermal simulation as a body source term. Subsequently, the thermal simulation results in a temperature distribution. Local temperature values are used in the mechanical model where they influence the constitutive relations.

The simulation approach described here qualifies as a sequentially coupled multi-physics simulation. It might be argued that it is not correct to treat the present problem as a sequentially coupled one, because the resistivity of the material in the R-phase is approximately 20% higher (Funakubo, 1987) than in the austenite phase. As the temperature and mechanical loading state of the material are linked to the phase fraction of the R-phase, this means that when using the material itself via Joule heating, the electrical problem is affected by the other problems as well. Solution of the full problem then requires a fully coupled analysis.

However, in the present situation it has been found that the effect of accounting for this phase-dependent resistivity is not significant, and therefore sequential treatment of the problem is justified. In order to show this, it is assumed that the R-phase fraction γ varies from zero to one in the transition region of the stress-strain curve. In a first approximation a linear relationship is assumed:

$$\begin{cases} \varepsilon_e \leq \varepsilon_1 & : \quad \gamma = 0, \\ \varepsilon_1 < \varepsilon_e \leq \varepsilon_2 & : \quad \gamma = \frac{\varepsilon_e - \varepsilon_1}{\varepsilon_2 - \varepsilon_1}, \\ \varepsilon_e > \varepsilon_2 & : \quad \gamma = 1. \end{cases} \tag{3.19}$$

The dependence of the electrical conductivity K_E to γ is expressed by

$$K_E = (1-\gamma)K_E^A + \gamma K_E^R, \tag{3.20}$$

where K_E^A and K_E^R represent the electrical conductivity of pure austenite and R-phase material, respectively.

These relations have been included in the model, and fully coupled finite element simulations have been performed on the design. A staggered solution approach was adopted, consisting of repeated sequentially coupled electrical, thermal and mechanical simulations. The electrical conductivities of all elements in a new cycle were adapted according to the results in the previous cycle. This simple scheme converged fairly quickly, and it turned out that after 4 cycles a stable solution was reached. The difference between the obtained fully coupled solution and the sequentially coupled one was very small. The temperatures differed less than $0.5{\cdot}10^{-3}$ K, and the relative difference in the bending radius of the entire catheter was a few parts per million. Of course, the actual relation between the electrical conductivity and the phase fraction, temperature and mechanical loading state might differ in details from the approximations used for this test. However, considering the tiny differences in the results, this test clearly indicates that for the present concept it is not necessary to account for the phase-dependence of the electrical conductivity and to adopt a fully coupled solution strategy.

The electrical and thermal problems are both described by a Poisson equation. A standard finite element implementation for this equation has been applied here, based on a three-noded triangular element. Variations of the field variables through the thickness of the tube wall are neglected, based on the small thickness and relatively high conductivity. Isotropic thermal and electrical conductivity has been used. Both problems are linear, and computationally inexpensive to solve compared to the mechanical analysis. The heat generated by the resistive heating enters the thermal simulation as a body source term given by $Q = K_E||\nabla\phi||^2$, where K_E is the electrical conductivity and ϕ is the electrical potential. The temperature field that results from the thermal simulation is used in the mechanical simulation via the constitutive equations, as discussed in Section 3.3. For the mechanical simulation six-noded triangular shell elements are used, described by Van Keulen and Booij (1996). Because a nonlinear material model is used in the mechanical simulation, numerical integration is used in this shell element, with ten integration points over the thickness.

3.4.2 Load case, boundary conditions and symmetry

The active catheter design concept used in this chapter is a periodic structure consisting of many repetitions of the same basic unit, as shown in Figure 3.1. To limit the computational costs of the simulation it is attractive to exploit this periodicity in the geometry. For this reason, also a periodic load-case is chosen to

evaluate the performance of the design. The case that will be considered is full bending in a single plane, in this case the x, z-plane as indicated in Figure 3.5. The electrode activation pattern used to accomplish this is shown in Figure 3.5 for a certain voltage u. The electrodes for which no voltage is indicated remain at a reference zero potential. With this loading, the structure will bend in the positive rotation direction around the y-axis, using the right-hand rotation convention. Assuming the structure can be approximated by a sequence of an infinite number of basic segments, a symmetry condition applies between each consecutive pair of segments. In addition, the geometry and loading conditions are completely symmetric with respect to the x, z-plane. Therefore only a relatively small part of the whole structure, highlighted in blue in Figure 3.5, has to be used in the simulations.

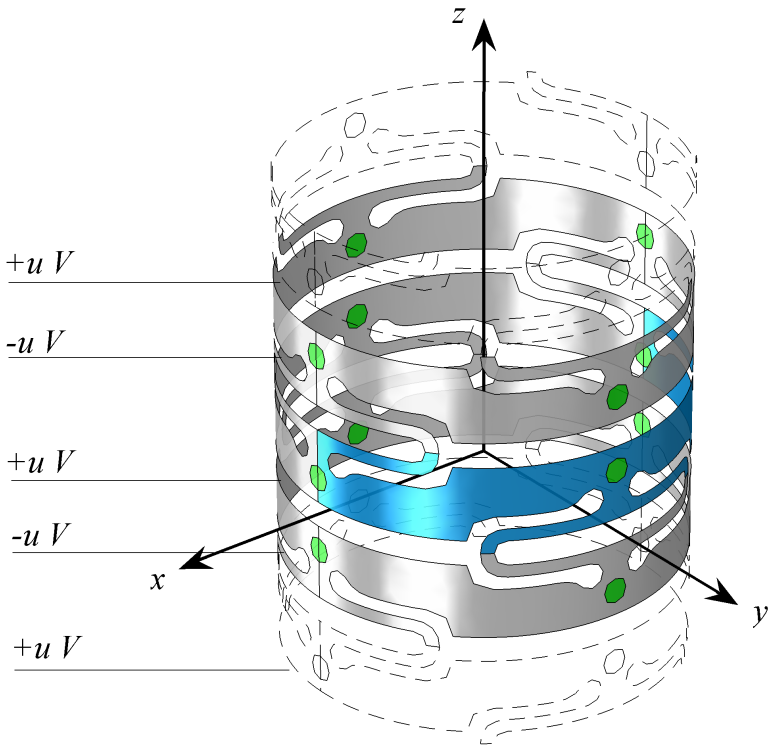


Figure 3.5: *Segmentation, applied voltage pattern and coordinate system. A full color version is given by Figure G.3 on page 250.*

In the electrical and thermal simulation the mechanical deformation is not accounted for, since the strains are too small to affect the outcome of these simulations. The application of symmetry boundary conditions therefore is straight-

forward. In the thermal case only convective boundary conditions are applied. In the mechanical case, a load is applied to the structure by the spacers. These spacers are implemented by means of truss elements under initial compression (cf. Figure 3.1).

An aspect that also requires some discussion is the implementation of the symmetry conditions in the mechanical case. The symmetry with respect to the x, z-plane is not difficult to enforce, but the symmetry planes between segments are more involved to deal with. The segment that will be analyzed is chosen to have one plane of symmetry parallel to the x, y-plane, which is easily translated into symmetry boundary conditions. However, the other symmetry plane is only initially parallel to the x, y-plane, but will rotate and shift as the structure deforms.

In this case, the symmetry plane is free to rotate around the y-axis, and to translate in x- and z-directions. Two points are sufficient to define this plane, provided these points do not differ only in their y-coordinates. In the derivation of the constraint equation to keep nodes on this symmetry plane, it is convenient to select two nodes as *master nodes*. These master nodes are considered to be the nodes that define the plane, and will be denoted by $\boldsymbol{M1}$ and $\boldsymbol{M2}$. A third node, the slave node, will be denoted by $\boldsymbol{S}$. The described situation is illustrated in Figure 3.6. Since the y-coordinates of points are not relevant for this discussion, locations of nodes are given by e.g. $\boldsymbol{M1} = (M1_x, M1_z)$.

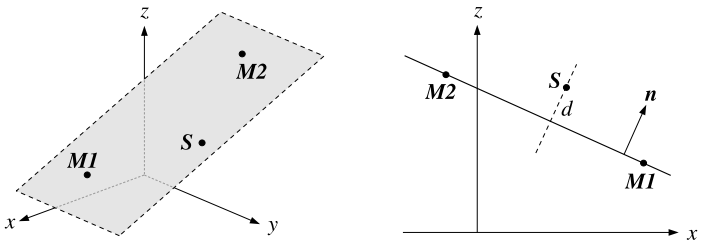


Figure 3.6: *Definition of the symmetry plane and the distance d.*

Keeping the slave node $\boldsymbol{S}$ on the plane defined by $\boldsymbol{M1}$ and $\boldsymbol{M2}$ is equivalent to the condition that the distance of $\boldsymbol{S}$ to the plane equals zero. This distance d is given by:

$$d = \boldsymbol{n} \cdot (\boldsymbol{S} - \boldsymbol{M1}) = \boldsymbol{n} \cdot (\boldsymbol{S} - \boldsymbol{M2}), \tag{3.21}$$

as illustrated in Figure 3.6. The normal vector of the plane $\boldsymbol{n}$ is given by:

$$\boldsymbol{n} = \frac{1}{\sqrt{(M2_x - M1_x)^2 + (M2_z - M1_z)^2}} \begin{Bmatrix} M2_z - M1_z \\ M1_x - M2_x \end{Bmatrix}. \tag{3.22}$$

Evaluation of Equation 3.21 yields, in both cases:

$$d = \frac{M2_z S_x - M1_z S_x + M1_z M2_x + M1_x S_z - M2_x S_z - M1_x M2_z}{\sqrt{(M2_x - M1_x)^2 + (M2_z - M1_z)^2}}. \quad (3.23)$$

The constraint equation that has to be satisfied for nodes on the symmetry plane now simply reads $h(\boldsymbol{M1}, \boldsymbol{M2}, \boldsymbol{S}) = d = 0$. For the implementation of this constraint, discussed briefly in the next subsection, its Jacobian and Hessian are needed as well. Their derivation is treated in Appendix D.

3.4.3 Solution process

The simulation aspects regarding the electro-thermo-mechanical coupling have been addressed in Section 3.4.1. Here the mechanical analysis itself is considered. The mechanical problem involves geometrical nonlinearities, physical nonlinearities and nonlinear constraints. Before discussing how the nonlinear constraints are accounted for, first an overview is given of the solution process itself.

Because the intent is to use the active catheter simulations in a design optimization setting, the solution process should be robust and should not require intervention from the user. For this reason, an incremental-iterative approach is used, with a simple adaptive algorithm (Crisfield, 1991) to determine the best increment size. The new increment size is determined after every increment by the following rule:

$$\Delta\mu^{k+1} = \Delta\mu^k \sqrt{\frac{N_D}{N_A}}. \quad (3.24)$$

Here $\Delta\mu$ represents the increment size used to scale loads, initial strains and prescribed displacements. N_D is the desired number of iterations per increment, and N_A is the actual number of iterations used in the previous increment. Starting cautiously with a small increment setting in combination with this adaptive approach gives a quite robust solution strategy. In order to increase the robustness even more, the possibility to automatically restart failed steps using half the increment size has been implemented as well. Extensive testing has shown that generally this strategy leads to satisfactory results.

To complete this discussion on the solution process, the implementation of the symmetry constraint has to be addressed. The symmetry constraint introduces a number of nonlinear constraint relations between the degrees of freedom (DOFs). Several approaches (Cheung *et al.*, 1996, Zienkiewicz and Taylor, 2000) exist to deal with these in a finite element setting:

1. Master-slave elimination, in which the constraint equations are used to eliminate slave DOFs out of the system equations. Master DOFs have to be chosen such that a consistent set of equations remains. This technique is not well suited for nonlinear cases.

2. Penalty formulation, where the constraint is enforced to some degree by penalizing the deviation with some penalty factor. It can be compared to adding stiffness to the system, which resists constraint violations. Constraints are therefore not satisfied exactly. Higher penalty values give greater accuracy, but lead to an increasing ill-conditioning of the system matrix.

3. Lagrange multiplier method. Constraint equations are included by defining an augmented unconstrained problem with the same solution as the original constrained problem. In this process, new variables, so-called *Lagrange multipliers*, are introduced in the system. In contrast to the penalty method, this approach can be compared to adding forces that enforce the constraints. An exact solution can be obtained. A drawback of this approach is that the system matrix of the augmented problem is no longer positive definite.

4. Augmented Lagrange multiplier method. This is basically a combination of the penalty formulation and the Lagrange multiplier method. Due to the addition of penalization terms the system matrix remains positive definite. Penalty factors do not have to be very high, since the accuracy is obtained by iteratively solving the Lagrange multipliers, and an exact solution can be obtained. Several multiplier update schemes are proposed, but generally quadratic convergence of the Newton method is lost.

After extensive testing of Methods 2 to 4, it was found that the augmented Lagrange multiplier method generally gives the most reliable results. The solution process is not very sensitive to the penalty factor used (in contrast to the penalty method) and results proved accurate. The fact that the system matrix is no longer positive definite in the Lagrange multiplier method led to problems with the Newton process used to solve the nonlinear equations, which is explained by the fact that convergence of Newton iterations is no longer guaranteed when positive definiteness is lost (Silva *et al.*, 2001).

The augmented Lagrange multiplier approach has been implemented using a specially formulated constraint element. The element stiffness matrix $\mathbf{k}^e$ of this constraint element is given by:

$$\mathbf{k}^e = p\frac{dh}{d\boldsymbol{U}}\frac{dh}{d\boldsymbol{U}}^T + (ph(\boldsymbol{U}) + \lambda)\frac{d^2h}{d\boldsymbol{U}^2}, \tag{3.25}$$

and its contribution to the load vector $\boldsymbol{f}^e$ reads:

$$\boldsymbol{f}^e = -(ph(\boldsymbol{U}) + \lambda)\frac{dh}{d\boldsymbol{U}}. \tag{3.26}$$

In these equations h represents the constraint, $\boldsymbol{U}$ is the vector of nodal displacements, p is a penalty factor and λ is the Lagrange multiplier variable associated with this constraint. The following conventional update rule is used:

$$\lambda^{k+1} = \lambda^k + ph(\boldsymbol{U}) \tag{3.27}$$

The multipliers are updated after every iteration. An alternative is to alternate multiplier updates and iterations with fixed multipliers until convergence, but generally that approach requires significantly more iterations. When starting a new increment, the Lagrange multipliers were incremented as well, by linear extrapolation from previous values. This proved to improve the convergence of the entire process.

3.4.4 Results

Finite element simulations were carried out for an SMA active catheter design model consisting of ca. 3500 elements. The input voltage was scaled to obey the maximum temperature limit. In the mechanical analysis, an initial increment size of 1% was used, in combination with the adaptive increment scaling algorithm. Physical constants and values of other parameters used in the simulation are listed in Table 3.2. Using a desired number of increments equal to five, the final solution was obtained after 14 increments. The augmented Lagrangian formulation was used for the symmetry plane constraint, with a penalty factor of 10^5 and using continuous multiplier updates. No line searches or restarts were required during this solution process.

Table 3.2: *Physical quantity values and settings used in the finite element simulation.*

Quantity	Value
Spacer Young's modulus	10 GPa
Spacer cross-sectional area	$1.0 \cdot 10^{-4}$ mm^2
Initial spacer space	25.4 μm
Spacer deployed length	50.8 μm
Tube wall thickness	45 μm
Tube diameter	1.50 mm
Electrical conductivity	$1.25 \cdot 10^6$ Sm^{-1}
Applied voltage	22 mV
Thermal conductivity	21 Wm^{-1}K^{-1}
Convection coefficient	$2 \cdot 10^3$ Wm^{-2}K^{-1}
Ambient temperature (T_0)	310 K
Maximum operating temperature	322 K

Some interesting results of the analysis are collected in Figure 3.7. In Figure 3.7-a the temperature distribution is shown, and it can be seen that one of the thin sections is clearly at a higher temperature than the one opposing it. This results in a different response to the load introduced by the spacers, as is shown in the deformed configuration in Figure 3.7-b. Without spacers, the bottom of the depicted segment in the unloaded configuration would be at $z = 0$. The top of the segment is supported in the x, y-plane and remains horizontal. But due

to the non-uniform temperature distribution the bottom of the segment is clearly turned in clockwise direction. The symmetry plane positioned at the bottom is visualized in Figure 3.7-b by a continuous line. Note that the effective strains, plotted on the deformed configuration in the same figure, are all below 1%. This means that they are within the range of validity of the present material model. But in addition, the highest values of the effective strain are very localized. Their contribution to the total deformation is therefore not very significant, and over-straining at isolated locations can be considered tolerable without affecting the validity of the overall result.

The deformation of the analyzed segment results in a total bending radius for this active catheter design of 35.2 mm, which is also shown in the composition of deformed segments in Figure 3.7-c. A close-up view of a section of the catheter in bent configuration is shown in Figure 3.8. The power consumption of a single segment is equal to 2.2 mW, and for the 90° turned catheter section shown in Figure 3.7-c the total power consumption is ca. 0.2 W. To give an indication of the effect of this power consumption, which is effectively equal to the generated heat, the following example can be considered. Given the volume of blood present in a vessel six times the diameter of the catheter and as long as the catheter section

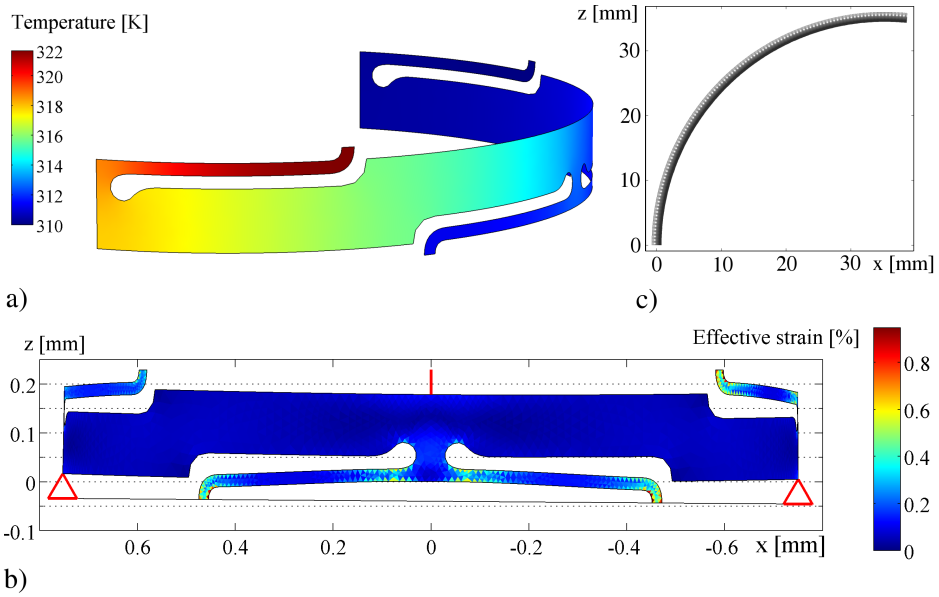


Figure 3.7: *Results of the finite element analysis. a) Temperature distribution, b) Deformed geometry and effective strain, c) Composition of catheter in bent configuration. A full color version is given by Figure G.1 on page 249.*

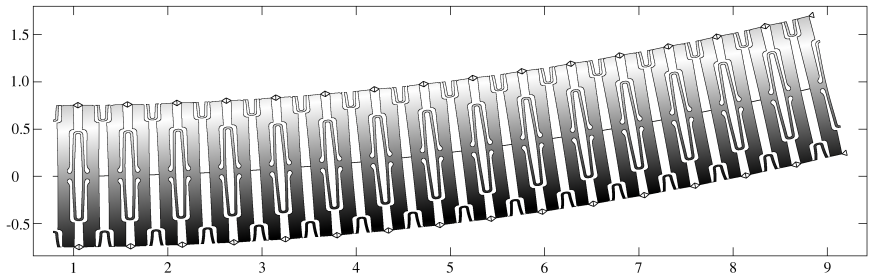


Figure 3.8: *Close-up of a catheter section in bent configuration (dimensions in mm).*

required for a 90° bend (ca. 5.5 cm), and assuming the heat capacitance of blood is equal to that of water at 310 K (i.e. 4.18 J/(gK)), then the time required to heat this blood by a single degree with the heat generated in full bending equals more than 70 seconds. In this calculation the effect of blood flow and conduction through the vessel wall has been ignored, which even more strongly shows that the present heat generation does not threaten the health of the patient.

The performance of the present design is promising, but at the same time it leaves room for further improvements. The bending radius of ca. 3.5 cm might be sufficient for most applications, but in other cases a smaller bending radius might be preferred. The preceding discussion on peak effective strains suggests that the present design concept still has some unused potential. And even without locally loading the material beyond the limits of the present material model, the deformations could possibly be increased by local shape adjustments that reduce the peak strains. Next to that, it is conceivable that different choices for the general shape of the segments, the tube diameter, wall thickness and the spacer dimensions have a combined effect such that the rotation per segment increases significantly, resulting in a smaller bending radius and therefore an even more agile active catheter. Potentially when bending radii become sufficiently small, it might even become possible to actuate the catheter in a kind of crawling motion, which could be used to propel it through the vessels, instead of advancing it by pushing from the remote insertion point. The prevention of buckling during insertion by pushing is the main reason why conventional catheters have to be fairly stiff. It would certainly be interesting to see whether other methods for catheter placement are feasible.

A second aspect to consider when exploring alternative or modified designs is reduction of the diameter of the catheter, as the 1.5 mm diameter used in the current design might be too large in some cases. Yet another objective for design improvement could be to reduce the sensitivity of the catheter performance to factors that are relatively uncertain or difficult to control, such as manufacturing inaccuracies, thermal convection coefficients or various aspects of the material behavior.

The results presented in this section demonstrate the basic functionality of the current design concept. They show that a respectable degree of steerability can be achieved by an R-phase transformation shape memory alloy, and that accompanying operating temperature and heat generation do not pose a serious threat to the patient's health. But next to that, these results also indicate the great potential of model-based design and the opportunities for design optimization. The present model can be used to evaluate the performance of every conceivable alternative active catheter design concept, and the kind of structures that can be explored in this way are only limited by the imagination of the designer.

3.5 Conclusions

Because it serves as a challenging and typical example of a complex SMA adaptive structure, the design of an active catheter is taken as the subject of this work. A new active catheter concept has been proposed, aimed to reduce the number of individual parts, to improve its reliability and simplify its assembly. The limitations imposed on the device by the *in vivo* operating conditions have been considered, and a specific shape memory alloy has been selected that is able to operate in a narrow temperature range. For this R-phase transformation shape memory alloy a new constitutive model has been presented. This model combines a sufficiently accurate description of the experimentally observed behavior with a good suitability for design optimization, through the fact that its history-independence allows for efficient computation of design sensitivities. This suitability for optimization is also the motivation for the efforts to reduce the computational cost of the finite element model, by exploiting symmetry as much as possible. For this reason also a special constraint element to implement a rotating symmetry plane condition has been developed. The performance of the present active catheter design has been evaluated by the finite element model, and the results confirmed its potential.

However, this work does not only apply to catheters. The finite element modeling concept and the constitutive model can be used to model many other shape memory alloy adaptive structures as well. The selected material is very well suited for *in vivo* biomedical applications, where the continuing trend toward less invasive procedures stimulates the development of enhanced and miniaturized instrumentation. But also outside the medical field many (potential) shape memory alloy applications can be identified, and also for those the modeling and design approaches presented in this chapter can be of great use.

Returning to the active catheter design case, it is expected to continue to serve as a fruitful example to develop and study design optimization techniques for shape memory alloy adaptive structures. A number of directions for further improvement of the present active catheter design have been suggested at the discussion of the finite element results. Of course, it is possible to attempt to achieve these improvements by a process of repeated modification and evaluation of many designs. However, it is much more efficient to adopt a more systematic

approach, which is provided by structural design optimization techniques. The present model has been constructed with practical use in an optimization setting in mind, and is particularly suited for further exploration of this topic.

Based on: Langelaar, M. and van Keulen, F (2004a). Design optimization of shape memory alloy structures. In *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY.*

Chapter 4

Shape Optimization of an SMA Gripper

Due to their unique properties, shape memory alloys (SMAs) are well suited for a wide variety of (micro-)actuation applications. However, the complexity of their constitutive behavior complicates the design process. Much effort has therefore been spent on formulating mathematical models to describe SMA phenomena and analyze SMA structures. A few researchers have actually tried to use this knowledge in the design process by combining SMA models with optimization techniques. However, only cases with significant limitations have been published: some are restricted to one-dimensional wire models, others use assumed optimality criteria that might not always yield the optimal design.

This chapter presents shape optimization of arbitrarily shaped SMA structures based on the actual material behavior itself. This allows for a clear and unambiguous formulation of the design problem, and provides a structured approach for the design of shape memory alloy structures. A novel material model is used for this purpose. This model is specifically tailored for transformation pseudoelasticity based on the R-phase in Ni-Ti, a popular shape memory alloy. To illustrate the effectiveness of the proposed approach, a shape optimization study of a miniature gripper is discussed.

4.1 Introduction

Shape memory alloys (SMAs) are materials which exhibit a diffusionless solid-state phase transformation. This transformation is associated with a certain transformation strain that can be used for actuation. Local stress state, temperature and sometimes also magnetic field affects the transformation and therefore

these quantities can be used to control the behavior. The stresses or strains that can be obtained from SMAs are relatively large as compared to other actuator materials, particularly in nickel-titanium alloys, and this makes these materials very interesting for many applications. More information on SMAs can be found in Duerig *et al.* (1990) and Otsuka and Wayman (1998), among others.

An important difficulty with shape memory alloys is the fact that it is hard to design effective devices due to the complex behavior of these materials. Computational modeling of the constitutive behavior has received a lot of attention, as clearly the ability to analyze models of SMA structures helps to understand and improve their design. Yet, the combination of this analysis capability with optimization procedures leads to a much more powerful and efficient design tool. Some researchers have already applied optimization in various SMA design problems, but to the authors' knowledge never before has a formal design optimization method been applied to a generic finite element (FE) model based on the actual SMA constitutive behavior. A reason for this might be that existing constitutive models suited for finite element analysis are often too computationally intensive or complex for use in an optimization setting. However, recently, constitutive models have been developed with specifically the suitability for use in design optimization in mind (Langelaar and Van Keulen, 2004b,c) (Chapter 2 and Chapter 3 of this thesis). Not only the agreement with experimental observations, but also computational efficiency and the possibility to efficiently compute design sensitivities has been considered. These models therefore enable a generic design optimization procedure of shape memory alloy structures. In this chapter, such a procedure is explored and presented for the first time.

The structure of this chapter is as follows: first, in Section 4.2 an overview is given of research related to design optimization of SMA structures and the approach presented in this chapter is discussed. The constitutive model for the shape memory alloy material used in the finite element analysis is the subject of Section 4.3. This is followed by Section 4.4 in which the design concept, parameterization, modeling and the formulation of the design optimization showcase of a miniature SMA gripper is presented. Results of the optimization are the topic of Section 4.5. Finally, conclusions and future directions are given in Section 4.6.

4.2 Design optimization of SMA structures

In recent literature, there have been a few publications related to the formal design optimization of SMA structures. Some of these limit themselves to structures based on wires (Troisfontaine *et al.*, 1999) or composites with embedded wires (Birman, 1996). A thorough optimization study based on an analytical model of a flexural actuator with corrugated core has been presented by Lu *et al.* (2001). However, this analytical approach only applies to a limited set of design problems where the geometry and loading pattern of the structure remain relatively simple.

Research focusing on more general structures is described in the publications by Kohl and coworkers on microvalves (Kohl *et al.*, 2000a, Skrobanek *et al.*, 1997), linear actuators (Kohl and Skrobanek, 1998) and grippers (Kohl *et al.*, 2000b). The approach to the optimal design of SMA structures in these papers, although it is not elaborated in great detail, is based on the assumption that the optimal design has a homogeneous stress distribution. This is motivated by the idea that peak stresses will lead to undesirable degradation of shape memory properties, and that the fraction of the material participating in the transformation depends on the stress level. A design with homogeneous stresses is therefore assumed to offer the highest resistance to fatigue, while achieving the maximum material participation for actuation.

Morgan and Friend show that stress concentrations can indeed accelerate the degradation of SMA structures (Morgan and Friend, 2001), although this is certainly not the only factor in the complicated mechanism of degradation of shape memory alloy properties. Although fatigue resistance is important, it does seem more logical to take the actual performance of a device as the objective for the design optimization, and take the fatigue resistance into account as a constraint, instead of optimizing the fatigue resistance alone.

Skrobanek *et al.* (1997) consider a microvalve that is actuated by bending beams laser-cut out of a thin sheet. The width of each beam varies along its length, and the profile used has been found analytically by demanding a constant stress for a given load pattern. Kohl *et al.* (2000a,b), Kohl and Skrobanek (1998) use an heuristic peak stress reduction algorithm developed by Mattheck and Burkhardt (Mattheck and Burkhardt, 1990), which basically reduces notch stresses by simulating the growth of trees. The same algorithm is used by Fischer *et al.* (1999) in the design of a Ni-Ti flexible endoscope tip. In all these works, no specific details are given on the optimization procedure and the material model used, but the lack of additional information seems to indicate the use of standard linear elasticity.

In this chapter, we propose and demonstrate an alternative approach, inspired by methods used in the field of structural optimization. In the proposed method, the performance of a design is analyzed by a parameterized finite element model. This model uses a newly developed constitutive law, that describes the SMA material behavior. The model is combined with an optimization algorithm, in this case the multi-point approximation method (MAM, Toropov *et al.*, 1993a), in order to iterate towards the optimal design. An advantage of this approach is that it is very versatile, since the use of finite element modeling allows it to deal with a great variety of designs, unlike the mentioned approaches based on analytical models. Another advantage is that the optimization problem is formulated based on the *actual* performance of the structure, without the need for additional assumptions, unlike e.g. the more heuristic methods based on assumed optimality criteria reported in literature. This makes this optimization-based approach transparent, versatile and unambiguous. Finally, a clear advantage is that also additional constraints can be added easily.

By its nature, the finite element model used to analyze the design is rather expensive in terms of computation time, and since in this case it involves highly nonlinear models. Moreover, the finite error remaining after the iterative solution process used for these models can lead to numerical noise in the responses, and this can hamper the optimization process. In this setting, the multi-point approximation method is well suited to perform the design optimization. For recent developments in the multi-point approximation method, the interested reader is referred to Van Keulen and Toropov (1997a) and Toropov *et al.* (1999a). The basic concept of this method is to perform the optimization by solving a sequence of sub-problems. A certain number of design points are generated and evaluated using the analysis model, and with these points a response surface is constructed, i.e. an analytical model is fitted to the generated points. The optimal point on this response surface in a certain trust region is located using another optimization algorithm: since the response surface is an analytical expression, this optimization sub-problem can be solved with little computational effort compared to the expensive analysis model. Next to that, the conversion to a response surface also can average out the noise in the responses. In order to check the quality of the response surface the optimal point found on the response surface is evaluated with the analysis model, and based on the resulting degree of correspondence the boundaries of the trust region are adjusted. This process is repeated until convergence, which can be defined by the degree of improvement, the amount of change in the design point or the size of the trust region.

4.3 Constitutive modeling

The focus in this chapter is on the pseudoelastic behavior in Ni-Ti due to the R-phase/austenite transformation. This is motivated by its unique small hysteresis and high fatigue resistance, which makes this an interesting material to use in (micro-)actuators (Otsuka and Wayman, 1998). Also the limited temperature range in which the transformation takes place makes it feasible to use this material in actuated devices for medical applications, which offer interesting opportunities for SMA technology.

The constitutive model used in the finite element computations for this chapter to describe the SMA material behavior is a refined version of a hyperelastic model by Langelaar and Van Keulen (2004c) (Chapter 2). This model is suited for quasistatic analysis of pseudoelastic behavior with a small hysteresis. The refinements are found in the formulation of the effective strain that now leads to a pressure-insensitive behavior and the fact that the transformation strain is fully isochoric, which leads to a closer agreement with the experimental observations. A full description of this refined model is given by Langelaar and Van Keulen (2004b) (Chapter 3). In brief, the model is based on the following effective strain

definition related to the distortional strain energy:

$$\varepsilon_e^2 = \frac{2}{3}\boldsymbol{\varepsilon}^T \mathbf{G}\boldsymbol{\varepsilon}, \tag{4.1}$$

where ε_e is the effective strain and $\boldsymbol{\varepsilon}$ denotes the Green-Lagrange strain in vector notation. The stress-strain relation reads:

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2G\mathbf{G})\boldsymbol{\varepsilon}. \tag{4.2}$$

Here $\boldsymbol{\sigma}$ is the second Piola-Kirchhoff stress in vector notation, K is the bulk ratio and G is the effective shear ratio. The matrices $\mathbf{G}$ and $\mathbf{K}$ in the preceding equations are constant and given by:

$$\mathbf{G} = \tfrac{1}{3}\begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}, \qquad \mathbf{K} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.3}$$

For a given value of the shear ratio G, this stress-strain relation corresponds to the well-known Hooke's Law for linear elasticity. However, in this case, in order to account for the isochoric transformation strains of the R-phase/austenite transformation, the value of G is taken as a function of the effective strain ε_e. The relation between G and ε_e is found by using data obtained from experimental tests. For the current model, data presented by Tobushi *et al.* (1992) have been used, in combination with a piecewise linear formulation. Results of finite element simulations of a tensile test specimen using this model are shown in Figure 4.1, together with the experimental results. It can be seen that they are in good agreement. As the model is intended to represent the experimentally observed behavior, its range of validity is limited to effective strains up to 1%. At higher effective strain levels, a stress-induced transformation to the martensite phase might occur, which is not described by the present model.

The tangent operator of this material model is obtained by differentiation of the stress-strain relation in Equation 4.2, which yields:

$$\frac{d\boldsymbol{\sigma}}{d\boldsymbol{\varepsilon}} = K\mathbf{K} + 2G\mathbf{G} + \frac{4}{3\varepsilon_e}\frac{dG}{d\varepsilon_e}\mathbf{G}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{G}. \tag{4.4}$$

It can be seen that this tangent operator is symmetric, since $\mathbf{K}$, $\mathbf{G}$ and also the product $\mathbf{G}\boldsymbol{\varepsilon}(\mathbf{G}\boldsymbol{\varepsilon})^T$ are symmetric. The term $dG/d\varepsilon_e$ follows from the relation between the shear ratio G and the effective strain mentioned earlier. Further details of these expressions, as well as the formulation of this material model in the plane stress case, can be found in Langelaar and Van Keulen (2004b) (Chapter 3).

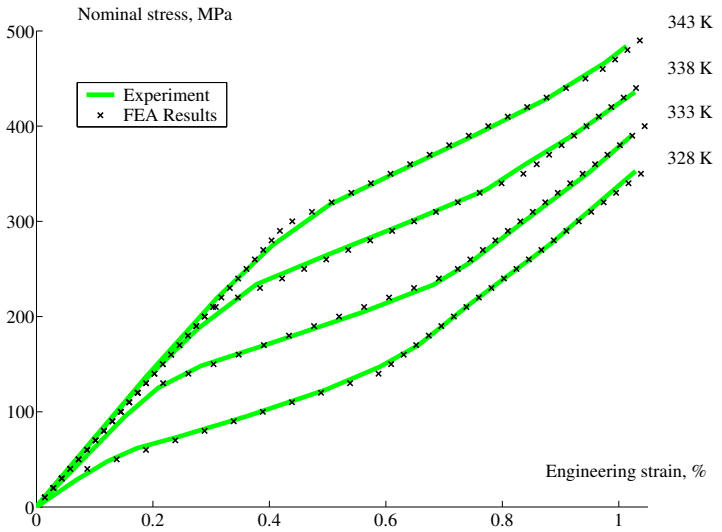


Figure 4.1: *Experimental stress-strain curves at various temperatures as determined by Tobushi* et al. *(1992) together with results obtained by finite element analyses using the proposed SMA material model.*

Note that the design optimization approach discussed and demonstrated in this chapter is not restricted to the present material model; in principle any SMA model that can be implemented in a finite element software package can be used. However, the practical feasibility of the procedure depends on the computational effort required for the function evaluations, relative to the available computing resources. The present model provides robust and reasonably efficient computations, and therefore fulfills this requirement.

4.4 Gripper design case

To demonstrate the effectiveness of the proposed FE-based design optimization approach, the design of a miniature gripper will be discussed. The conceptual design of this gripper is shown in Figure 4.2. It consists of an identical top and bottom arm made of folded Ni-Ti plates. In order to generate internal stresses in the material, starting from the undeformed configuration in Figure 4.2, the end of the top plate of the top arm is pinched towards the bottom plate, and similarly the bottom plate of the bottom arm is pinched towards the top plate of that arm. In this situation, the equilibrium configuration of each arm can be changed by either heating the top or bottom plate. Heating the inner plates (i.e. the bottom plate of the top arm and the top plate of the bottom arm) will cause the tip ends

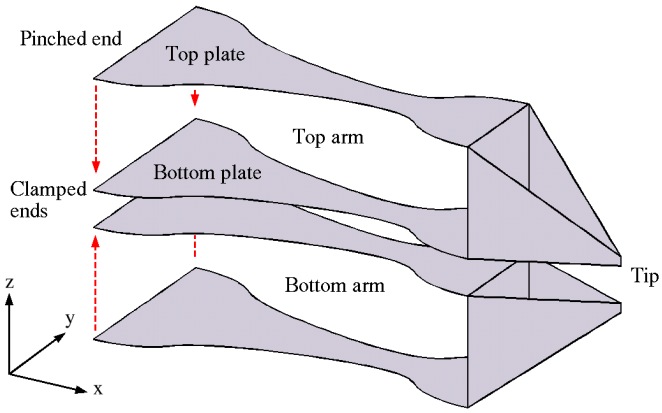


Figure 4.2: *Gripper geometry in undeformed configuration.*

to move apart, opening the gripper. Likewise, heating the outer plates will make them move towards each other, closing the gripper. In the closing configuration, clamping forces of 100 mN are applied in z-direction at the tips of the gripper.

Because of symmetry, only a quarter of the gripper needs to be modeled: in this case half the top arm is used. This part together with the parameterization of the geometry is shown in Figure 4.3. Dimensions in this chapter are given in millimeters, unless stated otherwise. In the finite element modeling, symmetry boundary conditions are used in the x,z-plane. The parameters chosen for this design study are the plate thickness t, the undeformed arm height H, the actuation plate begin and end width W_1 and W_2, and the shape of the actuation plate. This shape is described by a quadratic B-spline (see e.g. Farin, 2002), and the y-coordinates of the two middle control points are used as design variables: Y_1 and Y_2. Further geometrical details of the miniature gripper are listed in Table 4.1.

In the finite element modeling of this gripper, two cases are considered: in the first case, it is assumed that the plates used for actuation are heated externally,

Table 4.1: *Significant coordinates of B-spline control points and other points defining the geometry of the miniature gripper.*

Point	x [mm]	y [mm]	Point	x [mm]	y [mm]
Control point 1	0	W_1	Control point 2	0.5	W_1
Control point 3	1	Y_1	Control point 4	3	Y_2
Control point 5	3.5	W_2	Control point 6	4	W_2
Tip	5.377	0	Slit end	3.8	0

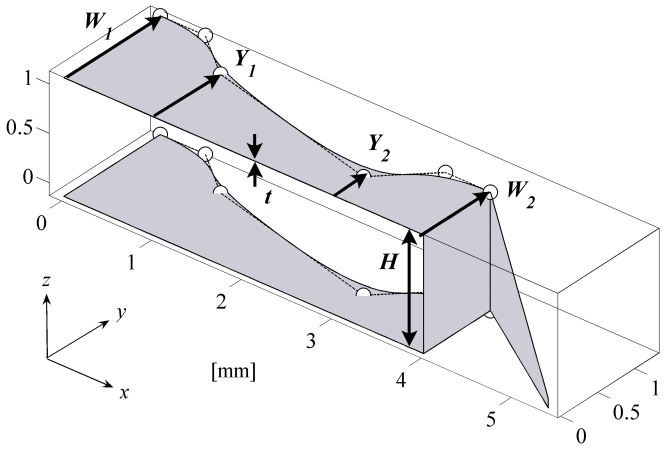


Figure 4.3: *Gripper design parameterization. Because of symmetry, only a quarter of the full gripper needs to be considered.*

and their temperature is chosen to be 10 degrees above the reference temperature. As the intention is to use this gripper in biomedical applications, 310 K is chosen as a reference temperature instead of 328 K shown in Figure 4.1, as in Chapter 3. Geometrical nonlinearities are considered in the simulation, and also the constitutive model is strongly nonlinear, which makes it necessary to use Newton iterations to find the new equilibrium configuration. An incremental-iterative approach is used, with rather conservative settings for the increment size. The reason for using small increments is that the analysis has to finish successfully for all designs in the design range: next to efficiency, robustness is very important for models used in optimization. A triangular shell element is used for the mechanical analysis (Van Keulen and Booij, 1996). For every design, a new mesh is generated automatically based on its geometry using a preprocessor: no mesh morphing techniques are used. Although this is one of the sources of noise in the responses of the model, this approach is preferred for its flexibility, ease of use and robustness.

In the second case, use is made of the Joule effect to heat the plate. This means that heat is generated internally by sending an electrical current through the structure by applying a voltage over the end of the plate. The design shown in Figure 4.2 does physically not allow for individual heating of top or bottom plates in this way, therefore a narrow slit is added in the middle of each plate in the direction of the x-axis, as shown in Figure 4.4. The width of the slit equals 200 micrometer. Figure 4.4 also shows the way the voltage is applied, in this case to the top plate. To simulate the performance of the gripper, an electrical, thermal and mechanical finite element analysis is needed. Dissipated heat from

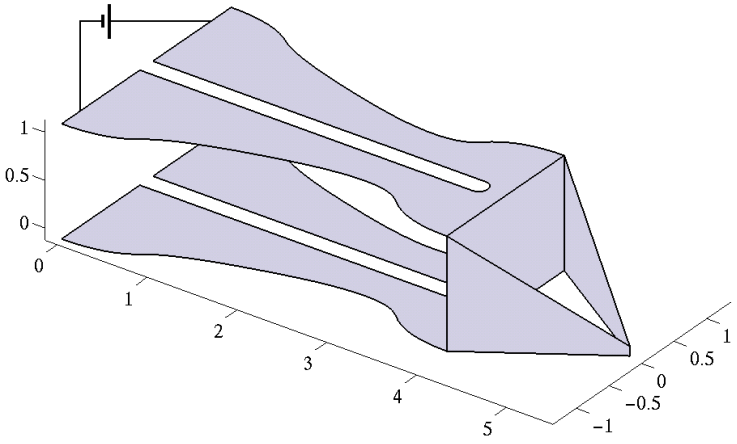


Figure 4.4: *Gripper geometry in the Joule heated case, in undeformed configuration. The voltage difference is applied to the top plate in order to close the gripper, or to the bottom plate in order to open it.*

the electrical analysis is used as a heat source in the thermal analysis, and the resulting temperature distribution is used in the mechanical analysis. Physical constants used in the simulations are collected in Table 4.2. The input voltage for every design is adjusted iteratively in order to meet a required maximum temperature of 320 K (same as in the externally heated case), since on physical grounds it is likely that this is the most favorable case. But of course, it is also possible to include the input voltage as a design variable as well, and add the temperature restriction as a constraint.

The objective for this design study is to maximize the range of motion of the gripper tips. Therefore, the difference between the z-coordinates of the gripper tip in open and closed configurations is taken as the objective function. The material model is limited to a maximum effective strain of 1%, and therefore a constraint on the effective strain is added in both the open and closed configuration. This

Table 4.2: *Physical constants used in the finite element modeling.*

Quantity	Value
Electrical conductivity	$1.25 \cdot 10^6$ Sm^{-1}
Thermal conductivity	21 $\mathrm{Wm}^{-1}\mathrm{K}^{-1}$
Thermal convection coefficient	$2.0 \cdot 10^3$ $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$
Ambient temperature	310 K

yields the following design optimization problem:

$$\begin{aligned} &\max_{\boldsymbol{x}} f(\boldsymbol{x}) \\ &\text{subject to:} \\ &\quad \max\left(\varepsilon_e^{\text{open}}\right) \leq 0.01 \\ &\quad \max\left(\varepsilon_e^{\text{closed}}\right) \leq 0.01 \\ &\quad x_i^{\text{lower}} \leq x_i \leq x_i^{\text{upper}}, \quad i = 1 \ldots 6 \end{aligned} \tag{4.5}$$

Here $\boldsymbol{x}$ is the vector of design variables $\{t, H, Y_1, Y_2, W_1, W_2\}$ and $f(\boldsymbol{x})$, the objective function, equals the difference between the z-coordinate of the tip of the gripper in opened and closed position. The last constraints represent the bounds used on the design variables. The values for the lower and upper bounds for each design variable are listed in Table 4.3. To illustrate the effect of the effective strain constraints on the optimal design, optimization studies have been performed with and without these constraints.

To shorten the time consumed by the optimization process, a parallel computing approach has been used, as described by Gurav *et al.* (2003). In the multi-point approximation method, the evaluation of design points for the response surface can be carried out in parallel quite easily. In this case, a heterogeneous Linux cluster consisting of 9 dual-processor PCs has been used, yielding 18 computational nodes. To give an indication of the computational effort involved in an analysis: the mechanical finite element model has about 10,000 degrees of freedom (order of magnitude - the exact number depends on the specific values of the design variables), and typically 60 Newton iterations are required to perform a single analysis. Note, both for the open and closed configuration an analysis is required. The electrical and thermal analyses hardly contribute to the total time, because they have much fewer degrees of freedom and do not require iterations.

4.5 Results

In total, four design optimization cases have been considered: an unconstrained and constrained case using external heating, and an unconstrained and constrained case using Joule heating. The multi-point approximation method turned out to be an effective tool and finished the optimization in 20 to 30 iterations. The number of function evaluations (i.e. finite element analyses) for a complete

Table 4.3: *Bounds used for the design variables.*

Variable	t	H	Y_1	Y_2	W_1	W_2
Lower bound [mm]	0.05	0.3	0.01	0.01	0.1	0.1
Upper bound [mm]	0.3	2	1.5	1.5	1.5	1.5

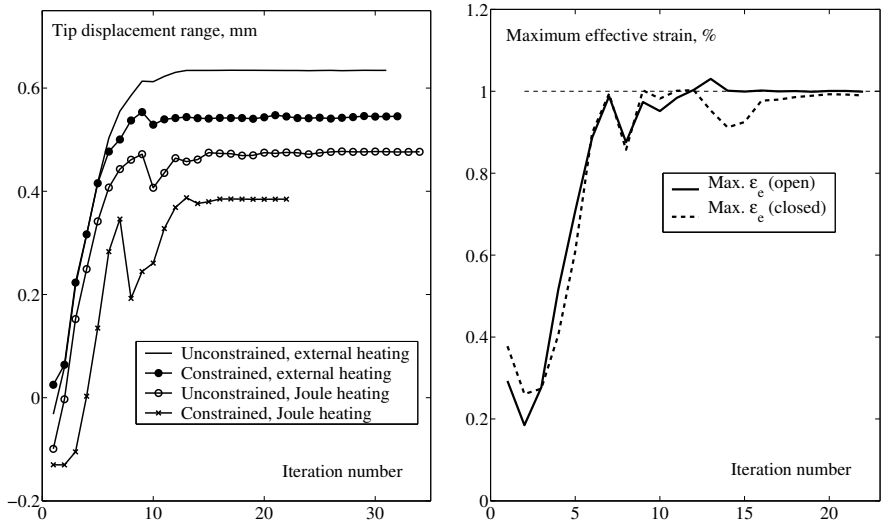


Figure 4.5: *Evolution of the objective (left) and constraint values (right) for various cases.*

optimization varied between roughly 700 and 1000, and thanks to the parallel computing approach the whole procedure took a modest 10 to 15 hours. The evolution of the objective function, the tip displacement range, throughout the optimization history is shown in Figure 4.5(a), for all four cases. The evolution of the constraint values for the Joule heated case are shown in Figure 4.5(b); for the externally heated case the trend was similar.

The optimal designs that have been found in all four cases are listed in Table 4.4, and the corresponding values of objective and constraints in Table 4.5. As could be expected, the addition of constraints on the maximum effective strain

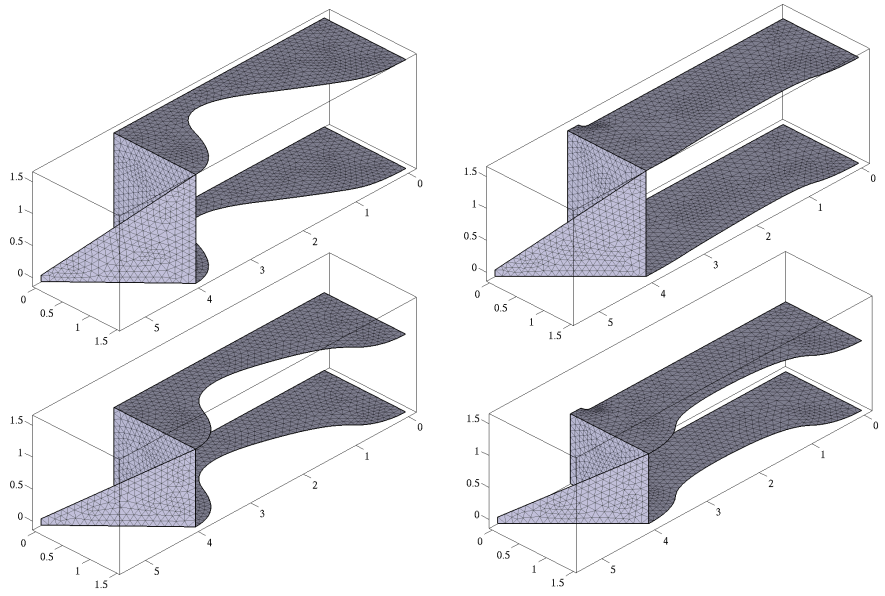
Table 4.4: *Design variable values at the optimum design for various cases.*

	External heating		Joule heating	
	Unconstrained	Constrained	Unconstrained	Constrained
t [mm]	0.08792	0.09601	0.09182	0.10058
H [mm]	1.7036	1.2168	1.6570	1.0941
Y_1 [mm]	1.2965	0.8956	1.3271	1.0276
Y_2 [mm]	0.0100	0.3177	1.3972	0.9789
W_1 [mm]	1.5000	1.4965	1.4960	1.5000
W_2 [mm]	1.5000	1.4988	1.4432	1.4423

Table 4.5: *Objective (displacement range) and constraint values (maximum strain) at the optimum design for various cases.*

Case	Displacement range [mm]	Max. strain (open/closed) [%]
Unconstrained, external heating	0.6345	-
Constrained, external heating	0.5453	0.9988 / 0.9235
Unconstrained, Joule heating	0.4765	-
Constrained, Joule heating	0.3847	0.9990 / 0.9896

lowers the range of motion of the gripper tip, and also it is seen, that the range of motion in the Joule heated case is smaller than in the externally heated case. This can be explained by the fact that in the Joule heated case, the optimal design somehow has to strike a balance between the shape that gives the best temperature distribution and that with the best mechanical performance in terms of tip displacement: these two aspects are interrelated. Therefore the final design will be a compromise. In the externally heated case, the shape of the structure



(a) Externally heated cases (b) Joule heated cases

Figure 4.6: *Optimal designs in the unconstrained (top) and constrained (bottom) case, for both types of heating.*

only influences the mechanical aspects, which leaves more freedom to improve the range of motion.

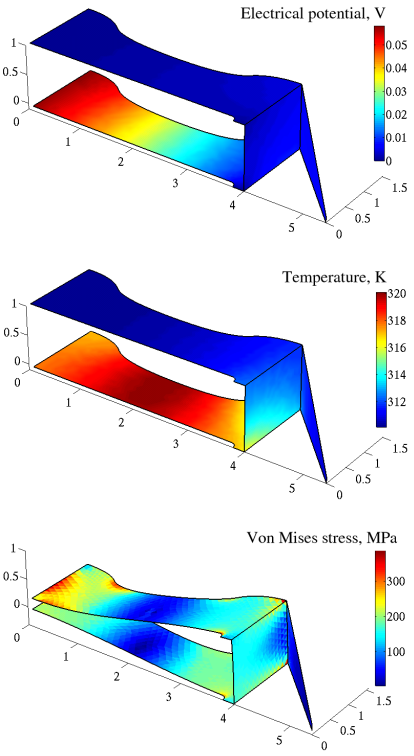
It is interesting to investigate the optimal designs not only by considering the numerical values shown in Table 4.4, but also by looking at the actual geometries. To this end, the optimal structures are presented in Figure 4.6(a) for the externally heated case, and in Figure 4.6(b) for the Joule heated case. It is seen, that in both cases the addition of the effective strain constraints leads to a design that is less high. From the numerical data it can be seen that in addition the plates are slightly thicker. Another interesting observation is that the shape of the plates is quite different depending on the heating conditions: in the externally heated case, locally the width of the plates is reduced strongly, whereas in the Joule heated case, the width remains much more constant along the length of the plate. This illustrates the fact that in the latter case a trade-off must be made between the electrothermal and mechanical properties of the structure.

More insight into the way this gripping device works can be gained by looking more closely at the analysis results. Most interesting is the Joule heated case, as it is a genuine multi-disciplinary problem. In Figure 4.7 results of the electrical, thermal and mechanical analysis are shown for the opened respectively closed gripper configuration. The electrical potential shows the way the device is operated: a voltage difference is applied over the ends of either the top or bottom plate. The resulting temperature distributions show that due to the generated heat, the temperature of the bottom, respectively, top plate is raised. The upper limit imposed on the temperature is respected. The Von Mises stress distribution found in the mechanical analysis is shown on the deformed configurations. The stress distributions differ considerably in opened and closed configurations, but note that in the latter situation also a 100 mN clamping force is acting on the structure.

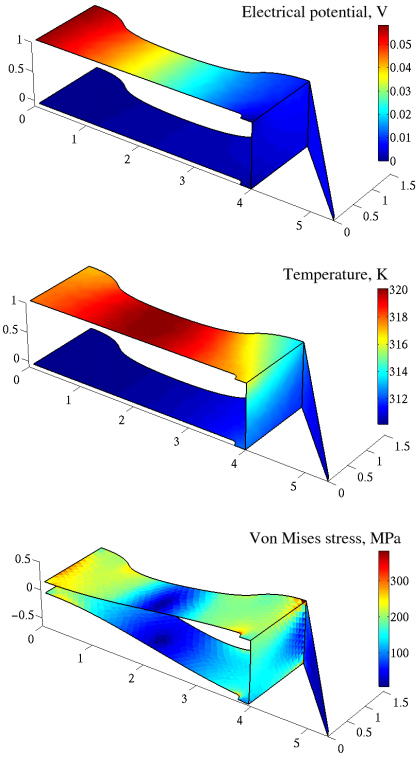
Finally, to show even more clearly how this gripper works, the contours of the open and closed configurations of the top gripper arm in the constrained Joule heated case are shown together in side view in Figure 4.8. Due to the changes in temperature distribution, the mechanical behavior of the material is changed locally, which leads to the two different equilibrium configurations.

4.6 Conclusion and outlook

In this chapter, structured design optimization of shape memory alloy devices is presented. Finite element modeling using a novel constitutive model for the SMA material behavior has been used in combination with a parameterized geometric model. A formal optimization problem was formulated and the optimization was carried out using the multi-point approximation method. To the authors knowledge, this is the first time a formal design optimization approach has been applied to a general finite element based model of such a device. Its viability and effectiveness for practical SMA structure design is demonstrated by means of an



(a) Analysis results for the opened configuration.



(b) Analysis results for the closed configuration.

Figure 4.7: *Computed electrical potential (top), temperature distribution (middle) and Von Mises stress distribution on the deformed structure (bottom) for the optimal design in the constrained Joule heated case, in opened (left) and closed (right) gripper configurations. A full color version is given by Figure G.4 on page 251.*

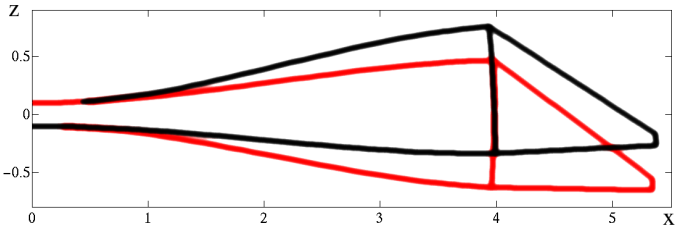


Figure 4.8: *Gripper top arm in open (black) and closed (red) position, side view. A full color version is given by Figure G.5 on page 251.*

example. It is emphasized that the presented approach is not limited to specific shapes or material models such as those used in this study. The generality of this approach is one of its main advantages. Another advantage is the ability to formulate clear and unambiguous design objectives and constraints based on the actual performance of the device.

Overall computing time was reduced by the use of parallel computing, and decreased to a level that makes this approach suitable for practical use. Further reductions are possible when design sensitivities are available: for example, the multi-point approximation method would then require less function evaluations for the construction of a response surface (Van Keulen and Vervenne, 2002). For this aspect the material model *does* make a difference: the history-independent nature of the material model used in this chapter makes that sensitivity analysis is considerably less involved compared to other, usually history-dependent SMA models (Kleiber *et al.*, 1997, Langelaar and Van Keulen, 2004c). This means that sensitivities could be computed with relatively little computational effort, which makes using sensitivities an attractive option to increase the practical applicability of the presented approach, particularly in larger design problems. This topic is explored further in Chapter 7, 8 and 9 of this thesis.

Based on: Gurav, S.P., Langelaar, M. and van Keulen, F. Cycle-based alternating anti-optimization combined with nested parallel computing: application to shape memory alloy microgripper. *Computers and Structures*, in review.

Chapter 5

Shape Optimization under Uncertainty

In this chapter, a new method for uncertainty-based design optimization based on an anti-optimization approach using Bounded-But-Unknown (BBU) uncertainties is studied on the basis of a practical application. The basic anti-optimization technique looks at the worst case scenario by finding the worst settings of the uncertainties for each constraint evaluation separately. This *Rigorous* anti-optimization technique involves two-level optimization in which anti-optimization is nested within the main optimization, making it computationally exhaustive. In the alternative *Lombardi-Haftka* approach, anti- and main optimization are carried out alternately avoiding the nested approach, which is quite efficient. A new Cycle-based Alternating technique based on a similar idea is studied in this chapter. In this technique, anti-optimization is carried out at the end of every cycle of the main optimization. The above anti-optimization techniques are studied and compared on the basis of an illustrative elastically supported beam example. Additionally, in the present chapter, a nested parallel computing strategy is developed in order to make the Cycle-based Alternating technique computationally efficient when a cluster of computers is available for function evaluation. This is particularly essential in case of practical problems involving expensive function evaluations, e.g., using Finite Element Analysis. The effectiveness of the Cycle-based Alternating technique combined with nested parallel computing is demonstrated by application to the uncertainty-based shape optimization of a shape memory alloy microgripper.

5.1 Introduction

5.1.1 Optimization involving bounded-but-unknown uncertainties

Many practical design optimization tasks involve uncertainties. In case statistical data on uncertainties is available, it can be used to construct statistical distributions for uncertainties. If such distributions are sufficiently reliable, a reliability-based design can be obtained by using probabilistic methods, see e.g. the textbook by Elishakoff (1983). In general, probabilistic methods require an abundance of experimental data, and even small inaccuracies in the statistical data can lead to large errors in the computed probability of failure (Elishakoff, 1999). However, in case of practical applications, it often happens that there is insufficient data available to construct reliable distributions. This particularly occurs in early stages of a design process. In such situations, the computed reliability of a structure can exhibit large errors. This can be crucial in applications which are required to perform without failure, or with very small probability of failure for the entire lifespan, for example, in case of space applications. On the other hand, the available data, in combination with engineering experience, can be used to set *tolerances* or *bounds* on uncertainties, within which the distribution is unknown, thus identifying uncertainties as Bounded-But-Unknown (BBU) (Ben-Haim, 1996, Ben-Haim and Elishakoff, 1990)).

The anti-optimization technique described in Elishakoff *et al.* (1994) tackles the BBU uncertainties using the worst case approach. This technique involves vertex checking of the uncertainty domain, in order to obtain the worst response of the structure. In comparison to other approaches, this technique is computationally very efficient for problems that are monotonic with respect to the uncertainties, however its application is limited to that category of problems. A more generalized approach, which can handle non-monotonicities, is adopted by Van Keulen *et al.* (2001) and Gurav *et al.* (2002). Here, this generalized or *Rigorous* anti-optimization is applied to the uncertainty-based design optimization of a car deck floor of a ferry. However, this technique suffers from the required large number of expensive function evaluations, due to the underlying two-level nested optimization. The *Enhanced* anti-optimization technique, in which sensitivities, a database and parallel computing are used to improve the computationally efficiency, is studied in Gurav *et al.* (2005) on the basis of a practical application related to microsystems. However, the approach still becomes computationally expensive for an increasing number of design variables and uncertainties.

5.1.2 Cycle-based approaches

The need for and lack of a computationally efficient method to handle uncertainties in design optimization problems has motivated the search for alternative approaches, which avoid the expensive two-level nested optimization used in the mentioned techniques. In one such approach, proposed by Lombardi and Haftka

(1998), anti- and main optimization are carried out alternately, thereby avoiding the nested approach. This approach can converge very rapidly in case of monotonic problems and is quite efficient in terms of the required number of function evaluations. However, in case of non-monotonic problems, for which the worst case can fluctuate from design to design, a large number of optimization cycles may be required to achieve convergence. Inspired by the Lombardi-Haftka technique, a slightly modified approach is adopted in the present chapter. Here, in the proposed *Cycle-based Alternating* technique, anti-optimization is carried out at the end of each cycle of the main optimization, to obtain the worst sets of uncertainties. These worst sets of uncertainties are subsequently used during the next cycle of the main optimization. For the initial cycle of the main optimization, uncertainties are set to zero. In case of fluctuating uncertainties, the Cycle-based Alternating technique is expected to converge faster than the Lombardi-Haftka technique, since the worst case is re-evaluated after each cycle, instead of after each complete sub-optimization process. Also in comparison to the previously proposed Enhanced anti-optimization technique, in case of fluctuating worst cases this newly proposed approach is expected to be more efficient. Additionally, the use of response derivatives with respect to uncertainties for estimation of worst uncertainties during the cycle can improve the convergence of the Cycle-based Alternating technique.

In many practical situations, the function evaluations of optimization problems involve computationally expensive finite element analysis (FEA). Moreover, the number of such FEAs required in the uncertainty-based design optimization can be quite high, depending on the problem at hand, resulting in impractical computation times. For such problems, it is necessary to perform function evaluations in parallel, e.g. using a cluster of computers. In the present chapter, a nested parallel computing approach is used in combination with the proposed cycle-based technique, in order to optimize a shape memory alloy (SMA) microgripper, simulated by computationally expensive FEA. The parallel computing framework used here is implemented in Python (Lutz, 2001).

The techniques for uncertainty-based optimization discussed above are embedded in a structural optimization setting using the Multipoint Approximation Method (MAM) (Toropov *et al.*, 1993b, 1999b, Van Keulen and Toropov, 1998). The different techniques will first be studied and compared on the basis of an elastically supported beam problem, considered earlier by Lombardi and Haftka (1998). To demonstrate the ability of the proposed Cycle-based Alternating technique to solve practical problems involving expensive FEA, the uncertainty-based shape optimization of an SMA microgripper is considered in the current chapter. In this study, uncertainties affect relevant environmental operating conditions as well as parameters in the SMA material model. Non-deterministic design optimization of SMA structures has thus far not been reported in the literature. By means of the considered microgripper example, this study aims to demonstrate the practical applicability of the proposed technique for this class of applications.

5.1.3 Outline

This chapter is organized as follows: first, the basic optimization problem formulation, together with a short description of the MAM optimization procedure, is given in Section 5.2. The BBU description of uncertainties is the subject of Section 5.3. Subsequently, in Section 5.4, various uncertainty-based design optimization techniques using BBU uncertainties are described. In Section 5.5, the discussed anti-optimization techniques are studied on the basis of an elastically supported beam problem. After having verified the newly proposed method on these test problems, its application in the uncertainty-based design optimization of an SMA microgripper is described in Section 5.6, followed by the conclusions in Section 5.7.

5.2 Multipoint Approximation Method

5.2.1 Introduction

In the present chapter the Multipoint Approximation Method (MAM) is used as a basis for optimization. Often practical applications involve numerical evaluation of response functions. From an optimization point of view, these types of problems can either suffer from numerical noise or the large computational time involved. The MAM, which is based on the sequential application of Response Surface Methodology (Khuri and Cornell, 1996, Myers and Montgomery, 1995), is well suited for dealing with those difficulties. The interested reader is referred to the studies by Toropov *et al.* (1993b, 1999b) and Van Keulen and Toropov (1997b, 1998). The MAM uses sequential response surface approximations to the responses, in order to reduce the number of expensive numerical response evaluations. However, it should be noted here, that particularly in combination with higher-order response surfaces, the method becomes increasingly expensive when the number of design variables increases.

5.2.2 Optimization problem formulation

Designing a structure implies that a design concept has to be selected, which subsequently has to be optimized. The latter involves the selection of design variables, which determine, among other features, the dimensions, shapes and materials to be used. This set of n design variables is denoted as $\mathbf{x}$, with

$$\mathbf{x} = (x_1 \dots x_n). \tag{5.1}$$

Throughout this chapter, it is assumed that all design variables are continuous. The behavior of the structure is described by the response functions, which are functions of the design variables. These response functions are denoted as $\mathbf{f}$ with

$$\mathbf{f} = (\mathrm{f}_0 \dots \mathrm{f}_m), \tag{5.2}$$

which may reflect, for example, weight, cost, buckling loads, maximum equivalent stress, or strain levels. Now the optimization problem can be formulated mathematically as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathrm{f}_o(\mathbf{x}) \\ \text{s.t.} \quad & \mathrm{f}_i(\mathbf{x}) \leq 1, \quad i = 1, \ldots, m, \\ & \mathrm{A}_j \leq x_j \leq \mathrm{B}_j, \quad j = 1, \ldots, n. \end{aligned} \tag{5.3}$$

Here, f_0 is the objective function and f_i are constraints. The design space is represented by the upper and lower limits on x_j, A_j and B_j, respectively.

The MAM is based on a sequential replacement of the actual optimization problem, as described by Equation 5.3, by a series of approximate optimization problems as schematically illustrated in Figure 5.1. The approximate optimization problem for a cycle p can be formulated as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \tilde{\mathrm{f}}_o^{(p)}(\mathbf{x}) \\ \text{s.t.} \quad & \tilde{\mathrm{f}}_i^{(p)}(\mathbf{x}) \leq 1, \quad i = 1, \ldots, m, \\ & \mathrm{A}_j^{(p)} \leq x_j \leq \mathrm{B}_j^{(p)}, \quad j = 1, \ldots, n, \\ & \mathrm{A}_j^{(p)} \geq \mathrm{A}_j, \quad \mathrm{B}_j^{(p)} \leq \mathrm{B}_j. \end{aligned} \tag{5.4}$$

Here, the response functions are replaced with approximate functions over the sub-domain for a cycle. For the current approximate optimization problem, $\tilde{\mathrm{f}}_i^{(p)}(\mathbf{x})$ are considered as adequate approximations of $\mathrm{f}_i(\mathbf{x})$ over the sub-domain (p), represented by the move limits $\mathrm{A}_j^{(p)}$ and $\mathrm{B}_j^{(p)}$, see Figure 5.1. It should be noted

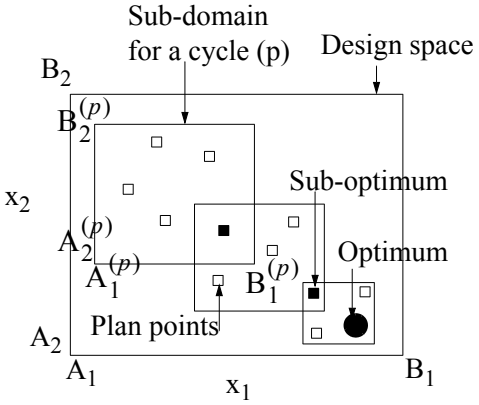


Figure 5.1: *Optimization using the MAM for a problem of two design variables (x_1 and x_2).*

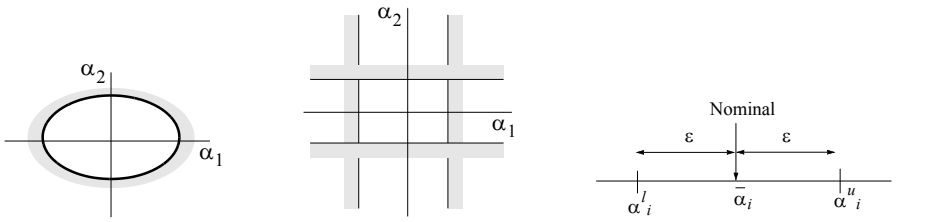
here, that the move limits for the initial cycle ($A_j^{(0)}$ and $B_j^{(0)}$), can be chosen either arbitrarily or based on engineering experience. Many times this can significantly influence the convergence. For example, if the initial domain includes the optimum, then the optimization can converge quite rapidly.

5.3 Bounded-But-Unknown uncertainty

If the problem at hand is non-deterministic, i.e. there are uncertainties that play a non-negligible role, the response functions also depend on the uncertainties. The set of uncertainty variables is denoted by $\boldsymbol{\alpha}$, with

$$\boldsymbol{\alpha} = (\alpha_1 \dots \alpha_u). \tag{5.5}$$

Consequently, the response functions depend on both design variables and uncertainties, hence $\mathbf{f}(\mathbf{x}, \boldsymbol{\alpha})$. In the present chapter, uncertainties are modeled using



(a) Ellipsoidal bound for all uncertainties together.

(b) Simple box bound with separate bound for each uncertainty.

(c) Simple box bounds on uncertainties: $\bar{\alpha}_i$ are nominal values of uncertainties with lower (α_i^l) and upper (α_i^u) bounds and ε as the dimension of the space of the feasible uncertainty.

Figure 5.2: *Bounds on uncertainties.*

the BBU approach. Thus, several bounds can be introduced, each providing a bound for a group of uncertainty variables or all uncertainty variables simultaneously, as illustrated in Figure 5.2. At the same time, we may want to measure the amount of uncertainty. Thus, measures for the dimensions of the subspace containing all possible selections of uncertainty variables are desired. This can be cast into a mathematical framework as follows. Assuming a set with b bounds, then a possible or *feasible* selection of $\boldsymbol{\alpha}$ satisfies, see Van Keulen *et al.* (2001):

$$B_i(\boldsymbol{\alpha}, \boldsymbol{\varepsilon}) \leq 0, \quad \text{for} \quad i = 1, \dots, b. \tag{5.6}$$

Otherwise the selection of the uncertainty variables $\boldsymbol{\alpha}$ is *infeasible*. Here, the components of $\boldsymbol{\varepsilon}$ are used to specify the dimensions of the space of feasible uncertainty variables. We will therefore refer to these components as the *levels of uncertainty*.

In the present chapter, simple box bounds, see Figure 5.2 (b) and (c), are used to specify uncertainties as

$$\left(\alpha_i - \bar{\alpha}_i\right)^2 - \varepsilon^2 \leq 0. \tag{5.7}$$

This type of bounds generally come from a tolerance specified on a nominal value, for example due to manufacturing inaccuracies. These bounds can be alternatively represented in terms of lower (α_i^l) and upper bounds (α_i^u) on uncertainties as

$$\begin{array}{rcl} \alpha_i^l & = & \bar{\alpha}_i - \varepsilon, \\ \alpha_i^u & = & \bar{\alpha}_i + \varepsilon. \end{array} \tag{5.8}$$

Note that the bounds used in this study were chosen based on engineering intuition, rather than a detailed analysis, since the purpose of the present examples is mainly to illustrate the proposed optimization technique. However, in practical situations where more detailed data is available, the same procedure can be applied with different bounds.

5.4 Uncertainty-based optimization using Anti-optimization

5.4.1 Anti-optimization

The anti-optimization technique to tackle BBU uncertainties consists of two levels of optimization. The outer level is given by the main optimization, and the inner level consists of the nested anti-optimization. The main optimization here is a standard minimization problem, that searches for the best design in the design domain. Anti-optimization is performed for every constraint, in order to obtain the worst values of constraints for each design within the main optimization. The anti-optimization problem using BBU uncertainties can be formulated mathematically as:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \mathrm{f}_o(\mathbf{x}) \\ \text{s.t.} & \mathrm{f}_i(\mathbf{x}; \boldsymbol{\alpha}_i^*) \leq 1, \quad i = 1, \ldots, m, \end{array} \tag{5.9}$$

where $\boldsymbol{\alpha}_i^*$ is the maximizer of

$$\begin{array}{ll} \max\limits_{\boldsymbol{\alpha}_i} & \mathrm{f}_i^*(\mathbf{x}; \boldsymbol{\alpha}_i) \\ \text{s.t.} & \mathrm{B}_j(\boldsymbol{\alpha}_i, \boldsymbol{\varepsilon}) \leq 0, \quad j = 1, \ldots, b. \end{array} \tag{5.10}$$

Here, $\mathrm{f}_0(\mathbf{x})$ is the objective function and $\mathrm{f}_i(\mathbf{x}, \boldsymbol{\alpha}_i)$ are constraints, whereas $\mathrm{B}_j(\boldsymbol{\alpha}_i, \boldsymbol{\varepsilon})$ are bounds on uncertainties. The minimization as defined in Equation 5.9 will be referred to as the main optimization. Notice that, in general, the evaluation of the constraints involves, for each set of design variables, one full anti-optimization for every individual constraint. This anti-optimization is reflected by Equation 5.10. The anti-optimization technique in the setting of the

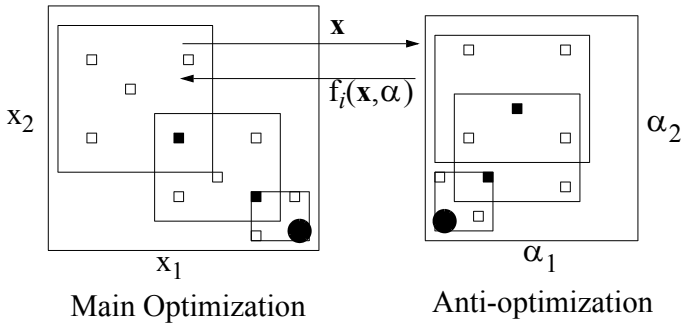


Figure 5.3: *Anti-optimization technique in the MAM setting for a problem of two design variables (x_1 and x_2) and two uncertainties (α_1 and α_2). The big boxes indicate the search (sub-)domains. The small open boxes indicate sets of design variables (left) or uncertainty variables (right) for which function evaluations are carried out. The small solid boxes indicate solutions of the approximate optimization problems.*

MAM is depicted in Figure 5.3. For the applications studied in the present chapter, uncertainties bounded by simple box bounds (see Figure 5.2(b) and (c) and Equation 5.8) are adopted. Therefore, the constrained maximization problem, as defined by Equation 5.10, reduces to an unconstrained maximization problem given by:

$$\begin{array}{ll} \max\limits_{\boldsymbol{\alpha}_i} & \mathrm{f}_i^*(\mathbf{x};\boldsymbol{\alpha}_i) \\ \text{s.t.} & \boldsymbol{\alpha}_i^l \leq \boldsymbol{\alpha}_i \leq \boldsymbol{\alpha}_i^u. \end{array} \tag{5.11}$$

The above *Rigorous* anti-optimization technique can handle large uncertainties safely. Moreover, it can account for discontinuities, if any. The price paid for this flexibility is the large computing effort required for the anti-optimization processes. In case of practical problems involving large numbers of design variables and uncertainties, anti-optimization can become very computationally expensive. In order to reduce the total number of expensive numerical response evaluations, the anti-optimization technique is modified in Gurav *et al.* (2005) by making use of database techniques and sensitivities.

In this *Enhanced* anti-optimization, derivative information, if available, is utilized to decrease the required total number of expensive function evaluations. In many cases of computational response analysis, gradient information can often be obtained at a fraction of the computing time as compared to the analysis itself (De Boer and Van Keulen, 2000a, Van Keulen and De Boer, 1998a, Van Keulen *et al.*, 2005). This sensitivity information can be used in addition to the function values to construct Gradient Enhanced Response Surfaces (GERS) (Van Keulen and Vervenne, 2004, Vervenne and Van Keulen, 2002). This incorporation of sensitivities can improve the quality of the response surfaces. Alternatively, fewer

response evaluations may be required to construct the approximations. Thus, using derivative information may decrease the total number of expensive function evaluations and hence may speed up the numerical optimization process. Similarly, a database technique is used to modify the anti-optimization technique in order to reduce the number of expensive function evaluations required for each anti-optimization. For this purpose, the worst sets of uncertainties obtained by anti-optimizations are stored in a database. When there is enough data available in the database, it is used to generate starting points for the anti-optimizations. Often this can speed up the anti-optimization processes significantly. Additionally, a parallel computing strategy is combined with anti-optimization in Gurav *et al.* (2005) in order to speed up the whole procedure.

5.4.2 Lombardi-Haftka Alternating anti-optimization

Although the Enhanced anti-optimization provides a quite efficient way of handling uncertainties, it becomes increasingly impractical with an increasing number of design variables and uncertainties. Therefore, it is necessary to think of alternative approaches. In the Lombardi-Haftka approach (Lombardi and Haftka, 1998), nesting of anti-optimization within the main optimization is avoided. In this method, the main and anti-optimization is carried out alternately as follows:

$$\text{(Main optimization)} \quad \begin{array}{ll} \min\limits_{\mathbf{x}} & \mathrm{f}_o(\mathbf{x}) \\ \text{s.t.} & \mathrm{f}_i(\mathbf{x};\boldsymbol{\alpha}_i^*) \leq 1, \quad i = 1,\ldots,\mathrm{n}. \end{array} \tag{5.12}$$

for given worst set of uncertainties $\boldsymbol{\alpha}_i^*$. These uncertainties are kept constant for each of the constraints until convergence of the main optimization, Equation 5.12, has been reached. Thereafter, new settings of the uncertainties $\boldsymbol{\alpha}_i^*$ are determined through anti-optimizations for the optimum $\mathbf{x}^*$ obtained by Equation 5.12, given by:

$$\text{(Anti-optimization)} \quad \begin{array}{ll} \max\limits_{\boldsymbol{\alpha}_i} & \mathrm{f}_i(\mathbf{x}^*;\boldsymbol{\alpha}_i) \\ \text{s.t.} & \boldsymbol{\alpha}_i^l \leq \boldsymbol{\alpha}_i \leq \boldsymbol{\alpha}_i^u. \end{array} \tag{5.13}$$

These cycles are repeated until convergence. For the initial iteration uncertainties are chosen arbitrarily or as $\boldsymbol{\alpha}^* = \bar{\boldsymbol{\alpha}}$, see Figure 5.2(c). This process is considered to have converged when:

$$\left|\left| \frac{\mathrm{f}_0 - \mathrm{f}_{0_{\mathrm{prev}}}}{\mathrm{f}_0} \right|\right| \leq \text{tolerance}, \quad \text{and} \tag{5.14}$$

$$\left|\left| \frac{\alpha_{ik}^* - \alpha_{ik_{\mathrm{prev}}}^*}{\alpha_{ik}^*} \right|\right| \leq \text{tolerance} \qquad i = 1,\ldots,\mathrm{n}; \quad k = 1,\ldots,\mathrm{u}. \tag{5.15}$$

Thus, convergence is defined based on relative change in the objective function, Equation 5.14, and the worst set of uncertainties corresponding to each constraint, Equation 5.15.

5.4.3 Cycle-based Alternating anti-optimization

The Lombardi-Haftka technique can be quite efficient in terms of number of expensive function evaluations in case of problems involving monotonicities. However, it can suffer from bad convergence in cases for which worst uncertainties fluctuate from design to design. Inspired by Lombardi and Haftka (1998), in the present chapter a slightly modified approach is presented. The idea is to solve

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \mathrm{f}_o^{(p)}(\mathbf{x}) \\ \text{s.t.} & \mathrm{f}_i^{(p)}(\mathbf{x};\boldsymbol{\alpha}_i^p) \leq 1, \quad i = 1,\ldots,\mathrm{n}, \end{array} \tag{5.16}$$

for given $\boldsymbol{\alpha}_i^p$. This set of uncertainties consists of the maximizers of

$$\begin{array}{ll} \max\limits_{\boldsymbol{\alpha}_i} & \mathrm{f}_i^{(p)}(\mathbf{x^p};\boldsymbol{\alpha}_i) \\ \text{s.t.} & \boldsymbol{\alpha}_i^l \leq \boldsymbol{\alpha}_i \leq \boldsymbol{\alpha}_i^u. \end{array} \tag{5.17}$$

Here, Equation 5.16 represents the main optimization for the p^{th} cycle which is solved first to obtain corresponding sub-optimum $\mathbf{x}^p$. For the initial cycle, the worst set of uncertainties $\boldsymbol{\alpha}_i^p$ needed in Equation 5.16 are chosen arbitrarily or as $\boldsymbol{\alpha}^{\mathrm{p}} = \bar{\boldsymbol{\alpha}}$, see Figure 5.2(c). Then, anti-optimization represented by Equation 5.17 is carried out at the sub-optimum ($\mathbf{x}^p$) to obtain the worst set of uncertainties ($\boldsymbol{\alpha}_i^p$), which will be used in the next cycle of the main optimization. During every cycle of the main optimization depending on the optimization history, the size and direction of the subdomain keeps changing until convergence is reached, as depicted in Figure 5.4.

In the Cycle-based Alternating technique, during each cycle of the main optimization, evaluation of constraints does not involve expensive anti-optimization,

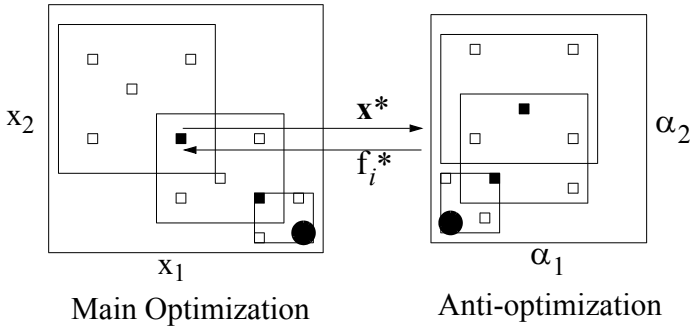


Figure 5.4: *Cycle-based Alternating anti-optimization technique in the MAM setting for a problem of two design variables (x_1 and x_2) and two uncertainties (α_1 and α_2). Here the anti-optimization is carried out only at the end of every cycle that is for the sub-optimal designs indicated by the solid boxes.*

which makes it computationally less expensive. Even though the technique suffers from slower convergence in case of fluctuating uncertainties, it shows better convergence as compared to the Lombardi-Haftka technique. Moreover, with the increase in number of iterations, the Lombardi-Haftka technique becomes computationally very expensive in terms of total function evaluations as compared to the Cycle-based Alternating technique. Additionally, in case additional information is available, such as estimates of worst uncertainties through derivative information, this information can be utilized during the cycle. Such additions are expected to further improve the convergence of the proposed technique.

5.4.4 Combined Cycle-based Alternating and Asymptotic method

The Asymptotic method uses derivatives of the response functions with respect to uncertainties to estimate worst set of uncertainties. In case of fluctuating uncertainties, when the Asymptotic method is combined with the Cycle-based Alternating technique, it tremendously improves the convergence of the Cycle-based Alternating technique. The Asymptotic method is fully described in Van Keulen *et al.* (2001) and discussed here for the type of problems dealt with in the present chapter.

Here, approximations for responses are constructed using Taylor series around $\boldsymbol{\alpha} = \bar{\boldsymbol{\alpha}}$ for a given design $\mathbf{x}$. These Taylor series read

$$\Delta \mathrm{f}_i = \frac{\partial \bar{\mathrm{f}}_i}{\partial \alpha_k} \Delta\alpha_k + \frac{1}{2} \frac{\partial^2 \bar{\mathrm{f}}_i}{\partial \alpha_k \, \partial \alpha_l} \Delta\alpha_k \Delta\alpha_l + \ldots, \tag{5.18}$$

with

$$\Delta \mathrm{f}_i = \mathrm{f}_i(\mathbf{x}, \boldsymbol{\alpha}) - \bar{\mathrm{f}}_i = \mathrm{f}_i(\mathbf{x}, \boldsymbol{\alpha}) - \mathrm{f}_i(\mathbf{x}, \bar{\boldsymbol{\alpha}}). \tag{5.19}$$

It is important to emphasize that it is not always possible to construct the above Taylor series. This is, for example, the case when a response function is continuous but its derivatives are discontinuous. If the above Taylor series can be created, then first-order approximations for the response functions are obtained by dropping all higher-order terms, giving

$$\Delta \mathrm{f}_i = \frac{\partial \bar{\mathrm{f}}_i}{\partial \alpha_k} \Delta\alpha_k. \tag{5.20}$$

It is important to realize that the derivatives of the response functions with respect to the uncertain variables are relatively inexpensive to calculate, provided efficient algorithms for sensitivity analysis are available (Van Keulen *et al.*, 2005). Using the approximations for the response functions given by Equation 5.20, the maximization problem given by Equation 5.17, which is used to find the worst set of uncertainties for a given design $\mathbf{x}$, is replaced by

$$\begin{aligned} &\max_{\Delta\boldsymbol{\alpha}_i} \quad \mathbf{h}_i^{\mathrm{T}} \Delta\boldsymbol{\alpha} \\ &\text{s.t.} \quad \boldsymbol{\alpha}_i^l \leq \boldsymbol{\alpha}_i \leq \boldsymbol{\alpha}_i^u, \end{aligned} \tag{5.21}$$

with

$$\mathrm{h}_i^{\mathrm{T}} = \left(\frac{\partial \bar{\mathrm{f}}_i}{\partial \alpha_1}, \ldots, \frac{\partial \bar{\mathrm{f}}_i}{\partial \alpha_u} \right). \tag{5.22}$$

As discussed before, in case of the Cycle-based Alternating technique, during the cycle worst sets of uncertainties are kept constant while computing responses. These worst sets of uncertainties are updated at the end of the optimization cycle by anti-optimization carried out for the sub-optimum. In the combined technique described here, during each cycle estimation of worst uncertainties by means of the Asymptotic method (Equation 5.21) is used. However, the rigorous check by means of full anti-optimization Equation 5.17 at the end of cycle, to update the worst set of uncertainties, is still kept in place. Additionally, this worst set of uncertainties obtained at the end of a cycle is used as a basis for the Taylor series approximation for the next cycle, i.e. $\mathrm{h}_i^{\mathrm{T}}|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*}$. The Asymptotic evaluation of worst uncertainties during the cycle can be solved computationally inexpensively, whereas it can improve the convergence significantly in case of fluctuating uncertainties. In the present chapter, the Cycle-based Alternating technique combined with the Asymptotic method is studied using an elastically supported beam example, in Section 5.5.

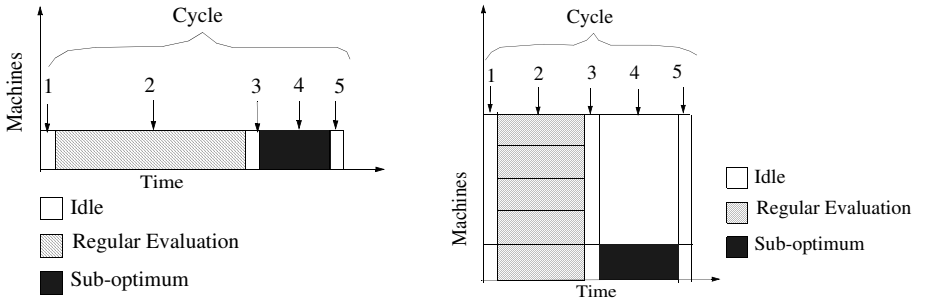
5.4.5 Parallel computing

Optimization using Parallel Computing

In many practical problems, evaluation of response functions is based on computationally expensive finite element analysis (FEA). For such problems, the time required for design optimization involving a large number of such expensive FEAs will easily become impractical. The duration of the optimization process can be reduced to practical levels through the use of parallel computing. To evaluate responses involving FEAs in parallel, clusters of multiple processors can be utilized (see e.g. Van Keulen and Toropov, 1999).

Optimization using the MAM involves various steps, such planning of experiments, response evaluation, response surface approximation, nonlinear minimization and updating of the move limits, see Section 5.2. The computing times required for these steps for a typical optimization involving expensive FEAs are illustrated schematically in Figure 5.5 (a). It can be clearly seen there, that evaluation of the plan points is computationally the most expensive phase, whereas the nonlinear minimization and the move limit updates are relatively inexpensive. When a cluster of several processors is available for computation, it can be used to evaluate the expensive response evaluations in parallel, see Figure 5.5 (b).

It should be noted here, that a single additional response evaluation is required at the end of the cycle, to evaluate the sub-optimum of the approximate optimization problem. In the present setting for response evaluation, splitting of this single response evaluation is not possible, therefore this sub-optimal point can only be evaluated on a single processor, keeping the other processors idle. This



(a) Plan points evaluated sequentially.

(b) Plan points evaluated using parallel computing.

Figure 5.5: *A cycle in optimization using the MAM: Steps involved are 1. Planning of new points, 2. Evaluation of plan points, 3. Nonlinear minimization, 4. Evaluation of sub-optimum, 5. Updating of move limits.*

increases the overall idle time significantly. The parallel computing framework used in the present research is developed in Python, see the textbook by Lutz (2001). To start multiple threads in parallel, the Threading module from Python is used in the current framework, which is schematically shown in in Figure 5.6. Here, each job involves evaluation of response functions, using, for example, FEA. The number of parallel jobs that can be started simultaneously depends on the number of processors available for computation.

During the evaluation of an individual job, first the design parameters are sent to the remote processor, see Figure 5.6. Then, the actual evaluation of responses is started on the remote processor by the associated thread. When the evaluations for the job are finished, corresponding responses are received back and are associated with the job. As soon as a processor finishes response evaluation and becomes again available for computation, the next job in the queue is submitted

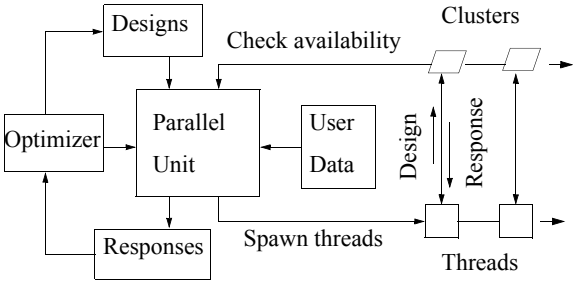


Figure 5.6: *A framework for parallel computing using the Threading module of Python.*

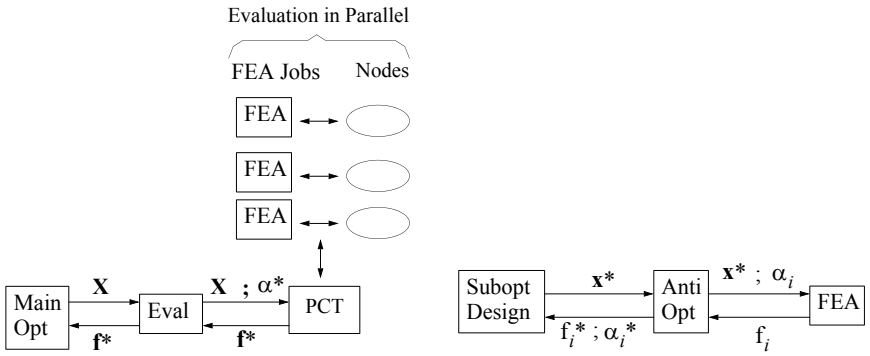
to it. This procedure is repeated until all the jobs have been evaluated.

It should be mentioned here, that the communication between the master and slave computers is implemented through files containing data, for example, flags indicating whether the submitted job is finished or not. The communication using files does increase the overhead time, however in practical cases this is negligible compared to the time required for an individual response evaluation.

Uncertainty-based Optimization using Nested Parallel Computing

In the case of uncertainty-based design optimization using anti-optimization, the overall number of required FEAs is quite high compared to the deterministic optimization. Therefore, to have a practical technique that can tackle uncertainties in practical design optimization problems, the use parallel computing is essential. In Gurav *et al.* (2005), parallel computing is combined with Enhanced anti-optimization to carry out uncertainty-based design optimization of a practical microstructure analyzed using expensive FEA. However, the strategy used there for the parallel computing is rather simple, and does not lead to the best utilization of the available computing power. In this chapter, an improved strategy for parallel computing is combined with the Cycle-based Alternating technique to carry out uncertainty-based design optimization efficiently.

For combining parallel computing together with the Cycle-based Alternating technique, various strategies have been considered. The effectiveness of these strategies for parallel computing is discussed by means of a virtual problem involving four constraints and three processors. It should be noted here, that the different strategies differ only in the evaluation of the suboptimum at the end of the cycle of the main optimization, see Figure 5.7(b). The evaluation of responses



(a) Response evaluation during the cycle of the main optimization.

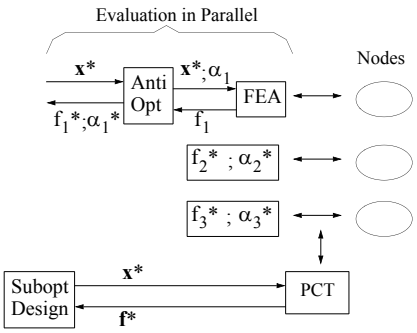
(b) Response evaluation at end of the cycle of the main optimization for the suboptimum.

Figure 5.7: *Response evaluation for the main optimization for the Cycle-based Alternating technique. Here worst uncertainties obtained from the previous cycle are used for the evaluation. Response evaluation involves FEA in parallel.*

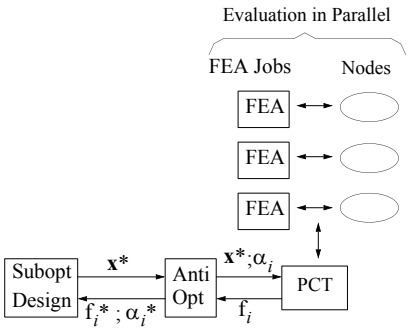
during the cycle of the main optimization involving expensive FEAs is carried out in parallel using available nodes as shown in Figure 5.7(a) in the same way for all strategies.

In case of Strategy I, see Figure 5.8(a), all anti-optimizations can be started in parallel, running one anti-optimization on one node. Each anti-optimization requires several cycles for convergence, and each iteration involves evaluation of constraints by means of FEAs in series on the corresponding node. It can be easily seen here, that for the current fictitious problem, the number of anti-optimizations mismatch the number of available nodes. For this problem three anti-optimizations corresponding to three constraints can be started in parallel on three nodes. However, during the evaluation of the fourth anti-optimization, only one node is utilized whereas all other nodes remain idle. The overall idle time can significantly increase with the increase in the difference between the number of constraints and number of nodes. Furthermore, increase in the computational time for a single FEA can add up to this idle time for such problems. However, in case the number of constraints matches the number of nodes, this strategy can be efficient.

For Strategy II, see Figure 5.8(b), anti-optimizations are carried out in series. However, the FEAs involved in each of these anti-optimizations are carried out in parallel, utilizing all available nodes. At the end of every cycle of the anti-optimization, a single FEA needs to be carried out. As before, this additional FEA can increase the overall idle time substantially. Additionally, during the cycle of each anti-optimization, if the number of FEAs are a multiple of the number of available nodes, then these nodes can be utilized efficiently. However, if there is a mismatch, then this will further increase the overall idle time. Moreover, the increase in computational time for the individual FEA adds up to the overall



(a) Parallel computing Strategy I: anti-optimizations are carried out in parallel, one on each node.



(b) Parallel computing Strategy II: FEAs for each single anti-optimization are carried out in parallel.

Figure 5.8: *Various strategies in combining parallel computing with the Cycle-based Alternating technique.*

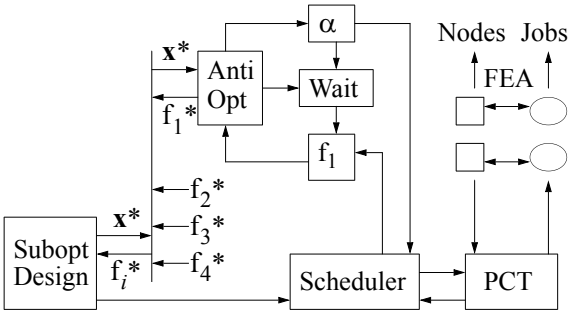
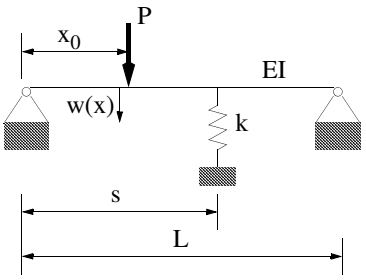


Figure 5.9: *Evaluation of responses at the end of the cycle of the main optimization, i.e. at the suboptimum. Here anti-optimization for each constraint is carried out, in order to obtain the worst values. Corresponding worst uncertainties are used for the next cycle. Anti-optimizations are carried out in parallel using nested parallel computing.*

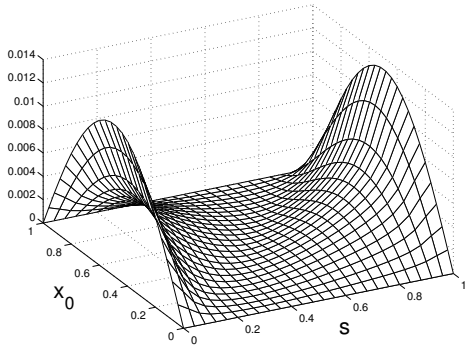
idle time, making the strategy computationally less efficient for such problems. In the present chapter, Strategy III, referred here as nested parallel computing due to the involved two levels in parallel computing, is combined with the Cycle-based Alternating technique, and applied for the SMA microgripper optimization involving uncertainties. This third strategy overcomes the disadvantages of the preceding two approaches, and is schematically depicted in Figure 5.9. In the nested parallel computing approach as proposed here, at the end of each cycle of the main optimization, the anti-optimizations are first started in parallel. Next, a program called "Scheduler" is started in the background. This Scheduler collects all the designs to be evaluated within each cycle of every anti-optimization, thus synchronizing all anti-optimizations. All designs collected by the Scheduler are subsequently evaluated in parallel. This nested parallel computing approach involves two levels. In the outer level, anti-optimization processes are started in parallel, whereas in the inner level, the actual evaluation of designs within all anti-optimizations is carried out in parallel as well. In this nested approach, efficiency is not affected seriously when the numbers of anti-optimizations and nodes do not match. Moreover, making a common list of all FEAs and then evaluating them in parallel utilizes the available nodes more efficiently. A disadvantage of this approach is that the implementation is significantly more complex, which is the price for the better utilization of the available computing power.

5.5 Test case

An elastically supported beam, see Figure 5.10, has been used by Lombardi and Haftka (1998) to test the anti-optimization technique in the presence of nonlinearities. One of the features of this problem is the strong dependence of the worst



(a) Problem setting: here L = 1 m, P = 1000 N, k = 500000 N/m, EI = 1000 Nm^2.



(b) Plot of the integrated displacement as a function of x_0 and s.

Figure 5.10: *Elastically supported beam example.*

case uncertainty values on the design variable. The different anti-optimization techniques discussed in Section 5.4 are studied on the basis of this example. For comparative study, for this test example similar notation as in Lombardi and Haftka (1998) is used.

In this problem, a beam loaded by a concentrated force P is supported elastically by a support with stiffness k, to limit its vertical displacement w, as shown in Figure 5.10 (a). The goal of the optimization here is to optimally place the elastic support, in order to minimize the integral of the displacement over the length of the beam. The location of the concentrated force P is uncertain, and can vary over the whole length of the beam. Assuming the nominal location of the concentrated force at the center of the beam ($\bar{x}_0 = 0.5L$), with $\varepsilon = 0.5L$, the lower and upper bounds on uncertainty (x_0) can be given as

$$\begin{array}{rcl} x_0^l & = & \bar{x}_0 - \varepsilon \;=\; 0, \\ x_0^u & = & \bar{x}_0 + \varepsilon \;=\; L. \end{array} \tag{5.23}$$

Here, the nonlinear displacement function $w(x, s, x_0)$ is the objective of the anti-optimization. It can be determined analytically by integration of the fourth-order differential equation governing the deformation of the beam:

$$EI\frac{\partial^4 w}{\partial x^4} = P\delta(x - x_0) - kw(x)\delta(x - s). \tag{5.24}$$

The main and anti-optimization problems are formulated as follows: solve the main optimization to obtain minimizers s for given x_0

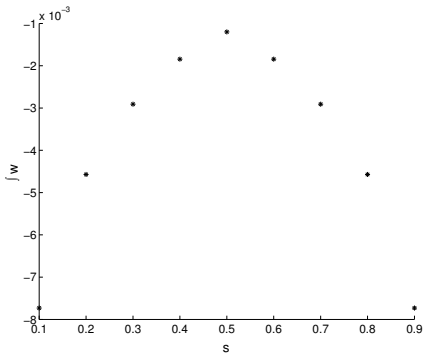
$$\min_s \int_0^L w(x, s, x_0)\, dx \quad \text{s.t.} \quad 0 \le s \le L, \tag{5.25}$$

and solve the anti-optimization to obtain maximizers x_0 for given s as

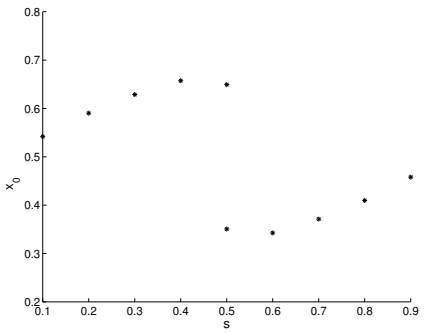
$$\max_{x_0} \int_0^L w(x, s, x_0)\, dx \quad \text{s.t.} \quad 0 \leq x_0 \leq L. \tag{5.26}$$

The actual integrated displacement is plotted in Figure 5.10 (b). It can be clearly seen here, that the displacement function is highly nonlinear, and displays a saddle-like response formed due to the discontinuity depending upon the location of the load (x_0) with respect to the location of the support (s). As a result, the worst location of the load (x_0) strongly depends on the location of the elastic support (s) and can fluctuate from point to point. In a first study, for selected designs (s) at regular intervals, anti-optimization is carried out to find the worst location of uncertainty (x_0) and the corresponding worst response ($\int w$), see Figure 5.11. The worst uncertainty (x_0) can be clearly seen to vary with respect to the design (s), and in addition a discontinuity can be identified at $s = 0.5$L. Here, the optimum location of the support is at 0.5m, whereas the worst location of the load is either at 0.35m or at 0.65m.

The uncertainty-based optimization problem as described by Equation 5.25 and Equation 5.26 is carried out here using three different techniques. Results are compared in terms of convergence and total number of function evaluations. A comparison of the optimization history of the different techniques is shown in Figure 5.12, and the obtained optimum design and the total number of function evaluations are compared in Table 5.1. It is clear, that the alternating techniques require more steps until convergence compared to the Rigorous approach, as shown in Figure 5.12 and Figure 5.13. In the Lombardi-Haftka case, due to the fluctuating uncertainties the process did not converge. The convergence of the Rigorous technique requires only a small number of steps, however, the total number of



Plot for the worst response ($\int w$) obtained by anti-optimization.



Plot for the worst uncertainty (x_0) obtained by anti-optimization.

Figure 5.11: *Anti-optimization results as a function of the design s for the elastically supported beam test case.*

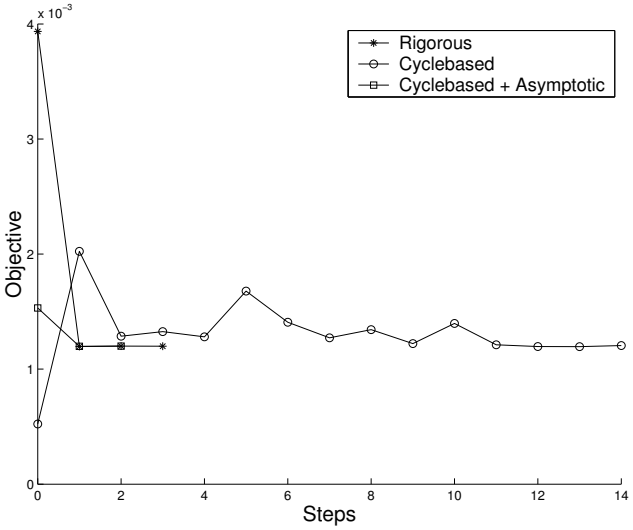
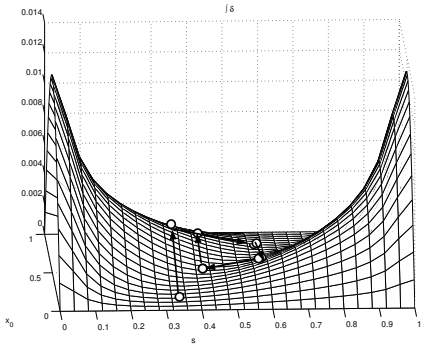
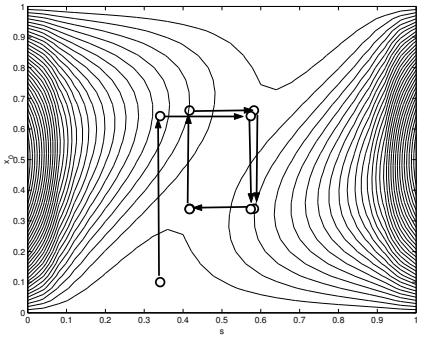


Figure 5.12: *Optimization history for the elastically supported beam problem: comparison of uncertainty-based optimization using different techniques.*



(a) Plot of the integrated displacement function together with the optimization history for the Lombardi-Haftka Method.



(b) Contour plot for the integrated displacement function together with the optimization history for the Lombardi-Haftka Method.

Figure 5.13: *Optimization history using the Lombardi-Haftka method for the elastically supported beam problem.*

Table 5.1: *Comparison between uncertainty-based optimization results using different methods for the elastically supported beam test problem.*

Method	Design variable s	Objective $\int w$	Function evaluations
Rigorous	0.4987	1.1993e-03	720
Cyclebased	0.4985	1.2048e-03	676
Cyclebased + Asymptotic	0.4997	1.1997e-03	100

function evaluations is the highest for this approach. Moreover, this number can rapidly increase further when the number of design variables, uncertainties and optimization cycles increases. Table 5.1 clearly shows, that in spite of requiring a larger number of steps, the cycle-based techniques result in fewer function evaluations. Moreover, if the number of steps required for the convergence is reduced, the gain in the number of function evaluations can be substantial. In case of the combined cycle-based and asymptotic technique, a further spectacular reduction in the number of function evaluations is realized. This is due to the incorporation of derivative information for predicting the worst uncertainties during each cycle. It should be noted here, that the required derivatives are obtained analytically for this problem. For practical problems, sensitivity analysis should be carried out, which should be available at low computational cost in order to keep this method efficient.

5.6 SMA Microgripper optimization under uncertainty

5.6.1 Introduction

The problem considered in this section is the shape optimization of a shape memory alloy microgripper. Shape memory alloys (SMAs) are materials in which a solid-state phase transformation can occur under the influence of a change in temperature or stress state. Internally, the lattice structure of the alloy changes from one configuration to another. The transformation is accompanied by a transformation strain, that can be used for actuation. Compared to other actuator materials, SMAs are capable of generating relatively large strains and stresses. This makes these materials very interesting for many applications. For further information about SMAs, see e.g. Otsuka and Wayman (1998) or Duerig *et al.* (1990). The focus of the present study is on the so-called R-phase transformation in nickel-rich NiTi alloys. The thermomechanical behavior of this material has been studied experimentally by Tobushi *et al.* (1992), and stress-strain curves at various temperatures are shown in Figure 5.14. Unlike most SMAs, this R-phase transformation is characterized by a small hysteresis and a relatively narrow

thermal operating range. These properties are attractive for actuator applications (Kohl *et al.*, 2004). The fact that in this case relatively small temperature changes can still induce significant SMA effects, makes that also *in vivo* medical applications might be possible using this alloy. The present microgripper design study is situated in that context.

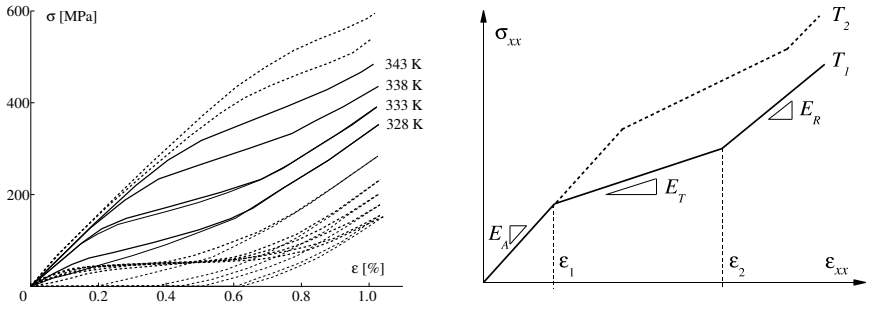


Figure 5.14: *Experimental stress-strain data at different temperatures for a NiTi alloy (Tobushi* et al.*, 1992). Thick and thin lines represent loading and unloading curves, respectively (left). The right diagram shows a schematic stress-strain diagram illustrating the piecewise linear approximation used in the SMA model.*

The constitutive model used to describe this SMA behavior focuses on the temperature range of 328–343 K, where the hysteresis is sufficiently small to be neglected. It is based on a piecewise linear approximation, which is fitted to the stress-strain curves, as schematically illustrated in Figure 5.14. A more detailed description of this model and its generalization to a three dimensional setting can be found elsewhere (Langelaar and Van Keulen, 2004b,c) (Chapter 2 and Chapter 3). Note that the selected temperature range is not directly suited for *in vivo* applications, but because the transformation temperatures can be influenced by heat treatments and alloy composition (Sawada *et al.*, 1993), lowering this range to compatible temperatures is possible.

5.6.2 Microgripper model

Before discussing the formulation of the optimization problem, the microgripper design concept is presented here. The conceptual design of this gripper is shown in Figure 5.15. It consists of an identical top and bottom arm made of folded Ni-Ti plates. An initial deformation is applied in order to generate internal stresses in the material, which are required to make use of the shape memory effect. Starting from the undeformed configuration in Figure 5.15, the ends of the outer plates are pinched toward the inner plates. In this situation, the equilibrium configuration of each arm can be changed by changing the temperature of either the inner or

outer plates. Resistive heating is used for this purpose, and to guide the electrical current through individual plates a slit is present along the length of each plate. Heating the inner plates will cause the tip ends to move apart, opening the gripper. Similarly, heating the outer plates will make them move toward each other, closing the gripper. In the closing configuration, clamping forces of 100 mN are applied in z-direction at the tips of the gripper, acting against the closing forces. A related microgripper design problem has been studied before by Langelaar and Van Keulen (2004a) (Chapter 4). However, in that case, uncertainties were not considered in the shape optimization. In the present problem, uncertainties in both the operating conditions and the SMA material properties are taken into account.

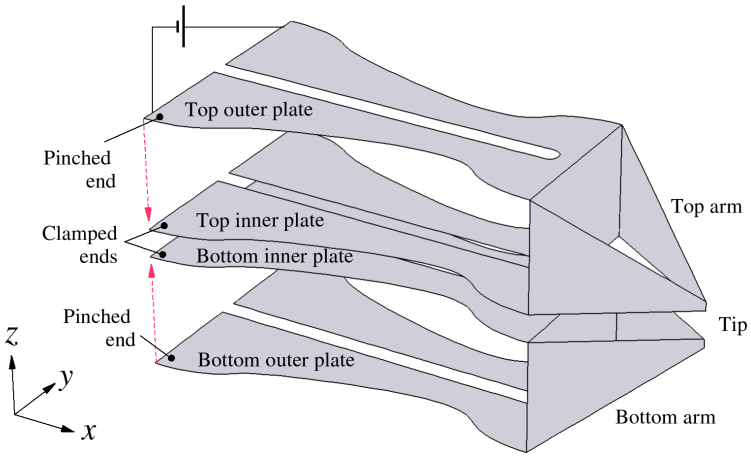


Figure 5.15: *Conceptual gripper geometry in the undeformed configuration.*

Because of symmetry, only a quarter of the gripper needs to be modeled: in this case half the top arm is used. This part, together with the parameterization of the geometry, is shown in Figure 5.16. The design variables chosen for this design problem are the plate thickness t, the undeformed arm height H, the actuation plate front width W_2, and the shape of the actuation plate. This shape is described by a quadratic B-spline (Farin, 2002), and the y-coordinates of the two middle control points are used as design variables: Y_1 and Y_2. Previous design studies have shown that the plate end width W_1 always remains at its upper bound, therefore it is excluded from the present design problem and set to 15 mm, based on previous studies. Further geometrical details of the miniature gripper can be found in Table 5.2.

The gripper is simulated by finite element analysis of the parameterized design shown in Figure 5.16. For both the opened and closed case, a quasi-static electrical, thermal and mechanical analysis is performed, to simulate the SMA behavior

Table 5.2: *Significant coordinates of B-spline control points and other points defining the geometry of the miniature gripper.*

Point	x [mm]	y [mm]	Point	x [mm]	y [mm]
Control point 1	0	15	Control point 2	0.5	15
Control point 3	1	Y_1	Control point 4	3	Y_2
Control point 5	3.5	W_2	Control point 6	4	W_2
Tip	5.377	0	Slit end	3.8	0

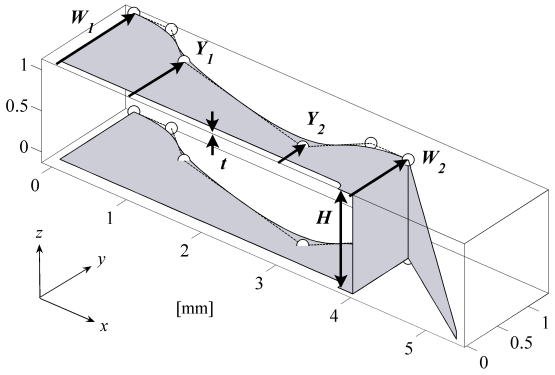


Figure 5.16: *Design parameterization of the gripper. Because of symmetry, only a quarter is considered.*

under the influence of Joule heating. Dissipated heat from the electrical analysis is used as a heat source in the thermal analysis, and the resulting temperature distribution is used in the mechanical analysis. Physical constants used in the simulations are listed in Table 5.3. Particularly the mechanical analysis is computationally intensive, because of the nonlinear SMA material model as well as the consideration of geometrical nonlinearities. An adaptive incremental-iterative scheme is used to ensure robust convergence. A triangular shell element is used for the mechanical analysis (Van Keulen and Booij, 1996).

5.6.3 Optimization problem formulation

The objective chosen for this design study is to maximize the range of motion of the gripper tips, i.e. the stroke of the gripper. Therefore, the difference between the z-coordinates of the gripper tip displacement u_z^{tip} in open and closed configurations is taken as the objective function:

$$\max_{\mathbf{x}} u_z^{\text{tip, open}}(\mathbf{x}) - u_z^{\text{tip, closed}}(\mathbf{x}) \tag{5.27}$$

Table 5.3: *Physical constants used in the finite element modeling.*

Quantity	Value
Electrical conductivity	$1.25 \cdot 10^6$ Sm^{-1}
Thermal conductivity	21 $Wm^{-1}K^{-1}$
Thermal convection coefficient	$2.0 \cdot 10^3$ $Wm^{-2}K^{-1}$
Ambient temperature	328 K

Here $(\mathbf{x})$ represents the vector of design variables. The design variables for the current problem, see Figure 5.16, are listed in Table 5.4.

Table 5.4: *Design variables* $(\mathbf{x})$ *for the design optimization of the SMA gripper.*

Design variable	Symbol	Lower bound	Upper bound	Unit
Plate thickness	t	0.05	0.3	mm
Gripper arm height	H	0.3	2	mm
Plate shape control point 1	Y_1	0.01	1.5	mm
Plate shape control point 2	Y_2	0.01	1.5	mm
Plate front width	W_2	0.1	1.5	mm
Applied voltage	V	0.001	0.5	V

The considered gripper optimization problem also involves a number of constraints. The validity of the material model is limited to a certain strain range, therefore a constraint on the effective strain ε_e is added in both the open and closed configuration. In addition, motivated by the possibility to use this SMA material for *in vivo* active devices, the thermal operating range is limited to 10 K. This means that per element in the finite element mesh, the following constraints are added to the optimization problem:

$$g_{\varepsilon}^{(i)} = \frac{\varepsilon_e^{(i)}}{\varepsilon_e^{(max)}} \leq 1 \tag{5.28}$$

$$g_{T}^{(i)} = \frac{T^{(i)} - T_{min}}{T_{max} - T_{min}} \leq 1 \tag{5.29}$$

The maximum effective strain $\varepsilon_e^{(max)}$ is set to 1%, and the minimum and maximum temperature values used are 328 and 338 K, respectively. A minimum value is included in the formulation in order to scale the temperature constraint properly.

To reduce the number of individual constraints and to make the anti-optimization approach feasible, a Kreisselmeier-Steinhauser (Kreisselmeier and Steinhauser, 1983) constraint aggregation function is used. This aggregation function in its

standard form is given by:

$$KS(\boldsymbol{g}) = \frac{1}{\rho} \ln \left(\sum_{i=1}^{N} e^{\rho g_i} \right) \tag{5.30}$$

where N is the number of individual constraints g_i (e.g. the number of elements) and ρ is a parameter that determines the bias of the aggregation. A higher value of ρ puts larger weight on higher constraint values, tending toward a maximum-operator, whereas a lower weight has the opposite effect, tending towards an averaging operator. In the present study, a modified Kreisselmeier-Steinhauser function is used. Here, in order to reduce the contribution of local violations, the individual element constraint values are weighted by the associated element area A_i as

$$KS_A(\boldsymbol{g}) = \frac{1}{\rho} \ln \left(\frac{1}{\sum_{i=1}^{N} A_i} \cdot \sum_{i=1}^{N} A_i e^{\rho g_i} \right) \tag{5.31}$$

Note that the use of this Kreisselmeier-Steinhauser function can not prevent a small number of isolated individual violations of the aggregated constraints. However, for a suitable choice of the parameter ρ the violations remain very small and limited in number. A value of 40 has been used here, as this turned out to give a satisfactory behavior.

The resulting optimization problem is now given by:

$$\begin{aligned}
&\max_{\mathbf{x}} \ \left(u_z^{\text{tip, open}}(\mathbf{x}) - u_z^{\text{tip, closed}}(\mathbf{x})\right)^{(p)} \\
&\text{Subject to:} \\
&\quad KS_A^{(p)}(g_\varepsilon^{\text{open}}(\mathbf{x}; \boldsymbol{\alpha}_1^{(p)})) \leq 1 \\
&\quad KS_A^{(p)}(g_\varepsilon^{\text{closed}}(\mathbf{x}; \boldsymbol{\alpha}_2^{(p)})) \leq 1 \\
&\quad KS_A^{(p)}(g_T^{\text{open}}(\mathbf{x}; \boldsymbol{\alpha}_3^{(p)})) \leq 1 \\
&\quad KS_A^{(p)}(g_T^{\text{closed}}(\mathbf{x}; \boldsymbol{\alpha}_4^{(p)})) \leq 1 \\
&\quad \mathbf{x}^l \leq \mathbf{x} \leq \mathbf{x}^u
\end{aligned} \tag{5.32}$$

The lower and upper bounds of the design variables are represented by $\mathbf{x}_l$ and $\mathbf{x}_u$, respectively, and $\boldsymbol{\alpha}_i^{(p)}$ denotes the worst settings for the uncertainties corresponding to each of the constraints in optimization cycle p, obtained as the maximizers of the following anti-optimization problems:

$$\begin{aligned}
&\max_{\boldsymbol{\alpha}_1} KS_A^{(p)}(g_\varepsilon^{\text{open}}(\mathbf{x}^{(p)}; \boldsymbol{\alpha}_1)) \quad \text{s.t.} \quad \boldsymbol{\alpha}^l \leq \boldsymbol{\alpha}_1 \leq \boldsymbol{\alpha}^u \\
&\max_{\boldsymbol{\alpha}_2} KS_A^{(p)}(g_\varepsilon^{\text{closed}}(\mathbf{x}^{(p)}; \boldsymbol{\alpha}_2)) \quad \text{s.t.} \quad \boldsymbol{\alpha}^l \leq \boldsymbol{\alpha}_2 \leq \boldsymbol{\alpha}^u \\
&\max_{\boldsymbol{\alpha}_3} KS_A^{(p)}(g_T^{\text{open}}(\mathbf{x}^{(p)}; \boldsymbol{\alpha}_3)) \quad \text{s.t.} \quad \boldsymbol{\alpha}^l \leq \boldsymbol{\alpha}_3 \leq \boldsymbol{\alpha}^u \\
&\max_{\boldsymbol{\alpha}_4} KS_A^{(p)}(g_T^{\text{closed}}(\mathbf{x}^{(p)}; \boldsymbol{\alpha}_4)) \quad \text{s.t.} \quad \boldsymbol{\alpha}^l \leq \boldsymbol{\alpha}_4 \leq \boldsymbol{\alpha}^u
\end{aligned} \tag{5.33}$$

Here $\boldsymbol{x}^{(p)}$ represents the suboptimal design obtained in the main optimization problem Equation 5.32 and anti-optimizations are carried out for this design for

each constraint, as specified in Equation 5.33. For evaluation of the objective, a complete electro-thermo-mechanical analysis is required for both the opened and closed configuration. However, for the constraint values, which are the objectives in the anti-optimization problems, no full gripper simulation is required. For instance, for $KS_A(g_\varepsilon^{\text{open}})$, only an electro-thermo-mechanical analysis is required for the opened configuration. Likewise for $KS_A(g_T^{\text{open}})$, an electro-thermal analysis for the opened configuration is sufficient. In relation to the computational effort required for the nonlinear mechanical analysis involving an incremental-iterative solution process, the computational effort required for this electro-thermal analysis is virtually zero. Therefore, significant computational savings were possible by exploiting these observations in the practical implementation, leading in this case to a reduction of the computational effort by approximately a factor 4.

The uncertainty variables contained in the set $\boldsymbol{\alpha}$ that are selected for the present design problem are listed in Table 5.5, together with their nominal values and their bounds. The ambient temperature is considered uncertain because it is hard to control. The convection coefficient is difficult to determine unless the environmental conditions are well known and stable, which is not likely to be the case. The other uncertainty variables are parameters of the SMA constitutive model, and these account for any inaccuracy that might be present in the measurements, as well as unknown aspects of the SMA behavior that have not been included in the modeling. One could think of, for instance, the minor hysteresis that has been neglected in the formulation of the SMA model. The range for these parameters has been chosen such, that it covers a substantial deviation of the modeled material behavior. This is illustrated by the stress-strain diagram shown in Figure 5.17, which visualizes the effect of the uncertainties in E_A, E_R,

Table 5.5: *Uncertainty variables $\boldsymbol{\alpha}$ for the gripper problem, their deterministic or nominal values together with their upper and lower bounds.*

Uncertainty variable	Symbol	Nominal value	Lower bound	Upper bound
Face convection coefficient [kWm^{-2}K^{-1}]	h	2.0	1.8	2.2
Ambient temperature [K]	T_a	328	327.6	328.4
Austenite Young's modulus [GPa]	E_A	68.939	63.0	75.0
R-phase apparent Young's modulus [GPa]	E_R	45.612	43	48
Initial apparent transition Young's modulus [GPa]	E_0	20.006	19.5	20.5
Poisson ratio	ν	0.3333	0.3	0.36

E_0 and T_a on the one-dimensional material model. The variation of the ambient temperature is assumed to directly affect the temperature of the SMA material, which is the worst case situation.

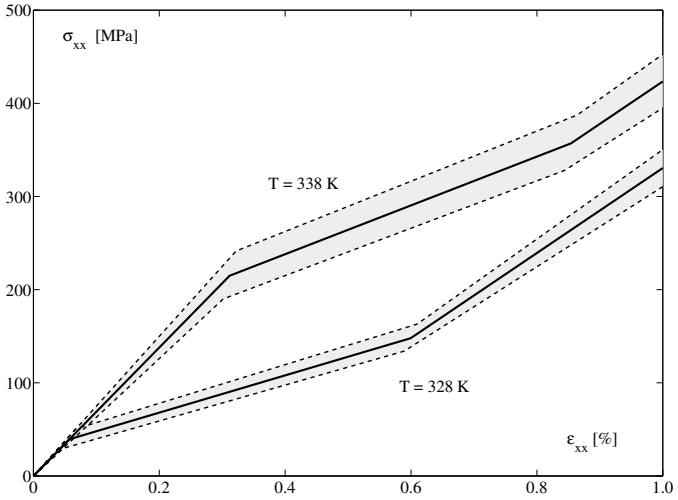


Figure 5.17: *One-dimensional stress-strain curves according to the deterministic model (thick lines) and the range covered by the uncertain material parameters (gray), at different temperatures.*

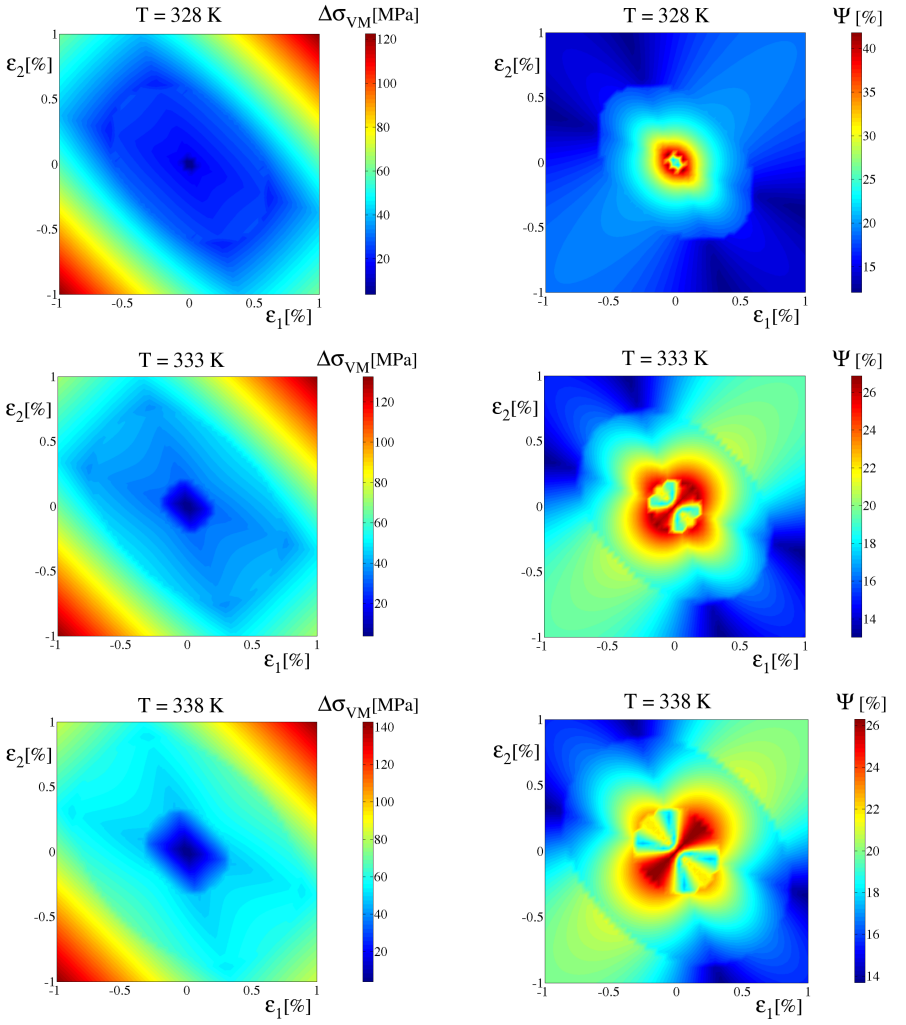
In order to visualize the effect of the uncertainties on the material model in the plane stress setting used in the gripper model itself, and to include the effect of the uncertainty in the Poisson ratio ν, two new quantities are defined. A first measure to illustrate the effect is the difference between the largest and smallest Von Mises stress value $\Delta\sigma_{VM}$ at a certain strain state, for any combination of uncertainty values in the defined ranges:

$$\Delta\sigma_{VM}(\varepsilon_1, \varepsilon_2) = \max_{\boldsymbol{\alpha}} \sigma_{VM}(\varepsilon_1, \varepsilon_2; \boldsymbol{\alpha}) - \min_{\boldsymbol{\alpha}} \sigma_{VM}(\varepsilon_1, \varepsilon_2; \boldsymbol{\alpha}) \tag{5.34}$$

where ε_1 and ε_2 are the principal strains, and σ_{VM} is the Von Mises stress. In order to be able to judge the relative magnitude of this difference, also a quantity Ψ is introduced where $\Delta\sigma_{VM}$ is normalized by the nominal value of the Von Mises stress at the considered strain state:

$$\Psi(\varepsilon_1, \varepsilon_2) = \frac{\Delta\sigma_{VM}(\varepsilon_1, \varepsilon_2)}{\sigma_{VM}(\varepsilon_1, \varepsilon_2)} \tag{5.35}$$

These two quantities are visualized in Figure 5.18 (a) and (b), for various values of the nominal temperature. Again the variation of the ambient temperature is



(a) Maximum absolute differences in Von Mises stress.

(b) Maximum differences in Von Mises stress relative to the nominal values.

Figure 5.18: *Maximum absolute (left) and relative (right) differences in Von Mises stress at different strain states due to the effect of the uncertainties, at different nominal temperatures. A full color version is given by Figure G.6 on page 252.*

assumed to directly affect the temperature of the material. Note that the relative effect of the uncertainties is quite large, roughly 20–25% on average, and that the uncertainties affect different strain states differently. Because of this, the equilibrium configuration of the gripper will therefore most likely be affected by the uncertainties. Hence it is hard to make a statement about which combination of uncertainty variables will result in the worst value for strain constraints.

In contrast to the strain constraint, the temperature constraint is affected only by the uncertainties in the ambient temperature and the convection coefficient. In fact, on physical grounds it is clear that the largest value for the temperature constraint is obtained when the ambient temperature uncertainty is at its upper bound and the convection coefficient is at its lower bound. But in this study, no use is made of this knowledge, and the temperature constraints are treated in the most general way.

Actually, it might be argued that the uncertainties affecting the temperature distribution are not very relevant, since in practice a temperature-controlled voltage could be applied, that ensures that the temperature limit is respected. However, this example is mainly meant to illustrate the versatility and effectiveness of the proposed technique in dealing with uncertain factors. Eliminating or circumventing some of the uncertainties might in fact be possible for this design problem, but in many other problems this might not be the case at all.

5.6.4 Results

Using the Cycle-based Alternating anti-optimization technique proposed in this chapter, design optimization of the SMA microgripper has been performed. The even more efficient combined Cycle-based Alternating and Asymptotic method could not be employed because sensitivity information was not yet available. Linear approximations have been used for response surfaces, both in main and anti-optimization. The Python-based nested parallel computing framework reduced the total time of the optimization process, and a cluster of 14 CPUs (1 GHz Pentium) was used. Both the deterministic and uncertainty-based optimization converged after ca. 20 iterations, and the evolution of the objective and constraint values are depicted in Figure 5.19 and Figure 5.20, respectively.

The objective history of the deterministic case shows a sharp spike at the fourth step, where the stroke of the gripper even becomes negative. This is clearly undesirable, and possibly is caused by the inability of that design to generate the required clamping force. However, the optimizer recovers in the subsequent step. The activity of the constraints shown in Figure 5.20 shows that all constraints are relevant to the design problem. This is confirmed by the fact that the final constraint values shown in Table 5.6 are all very close to 1.

The design variable values and the responses of the final designs are listed in Table 5.6. The stroke in case of the uncertainty-based design is ca. 15% smaller

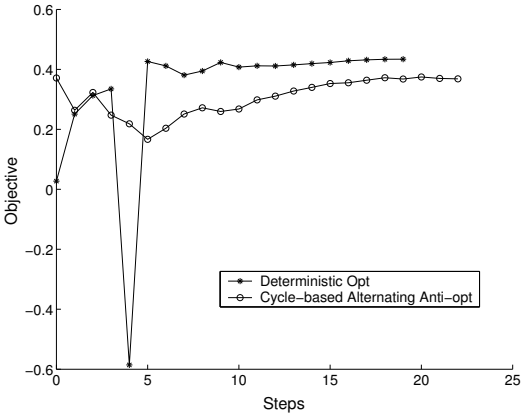
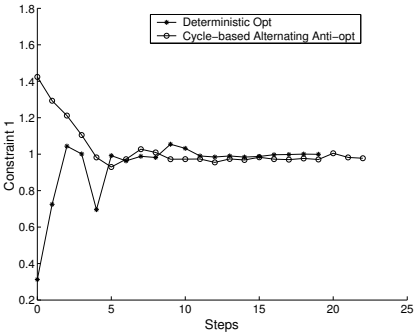
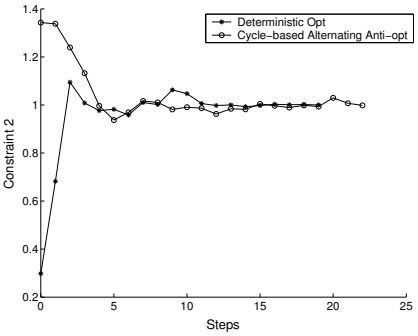


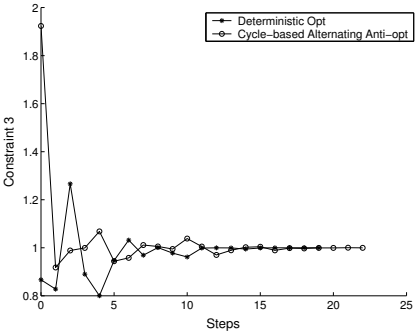
Figure 5.19: *Optimization history: objective function.*



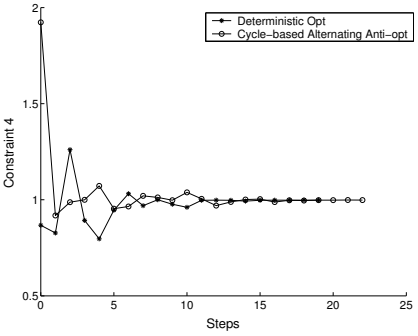
(a) Evolution of $KS_A(g_\varepsilon^{\text{open}})$.



(b) Evolution of $KS_A(g_\varepsilon^{\text{closed}})$.



(c) Evolution of $KS_A(g_T^{\text{open}})$.



(d) Evolution of $KS_A(g_T^{\text{closed}})$.

Figure 5.20: *Optimization history: constraints.*

than that of the deterministic design. Accounting for uncertainties essentially requires the design to move further away from the deterministic constraints, resulting in a reduction of the objective. Note that in the present study, only the effect of uncertainties on constraint functions is considered due to the present setting of the optimizer. In contrast, the effect of uncertainties on the objective function value is considered separately at the end of both optimizations. This is done by means of a separate anti-optimization for the optimal design, in order to anticipate the influence of uncertainties on the objective function. This anti-optimization is carried out separately for both the deterministic and uncertainty-based optimal design. In case of the deterministic optimization, the stroke reduces from 0.4341 mm (Table 5.6) to 0.3259 mm, whereas in the uncertainty-based case it reduces from 0.3684 mm (Table 5.6) to 0.2728 mm. Note that the presented technique can in principle be extended to include this influence of uncertainties on the objective in a similar way as this has been done for constraints.

The total number of FEAs and the wall clock time required for both deterministic and uncertainty-based optimization are also listed in Table 5.6. The uncertainty-based case requires 30 times as many evaluations as the deterministic case, which clearly shows the need of sophisticated techniques such as Nested Par-

Table 5.6: *Comparison between deterministic and uncertainty-based optimization results using Cycle-based Alternating anti-optimization for the SMA microgripper problem.*

	Deterministic	Uncertainty-based
Optimal Response		
f [mm]	0.4341	0.3684
$KS_A(g_\varepsilon^{\text{open}})$	0.9992	0.9776
$KS_A(g_\varepsilon^{\text{closed}})$	0.9999	0.9983
$KS_A(g_T^{\text{open}})$	0.9999	0.9998
$KS_A(g_T^{\text{closed}})$	0.9979	0.9982
Optimal Design		
t [mm]	0.0974	0.1022
H [mm]	1.1703	1.0463
Y_1 [mm]	1.2485	1.3047
Y_2 [mm]	1.1125	1.0208
W_2 [mm]	1.4001	1.2703
V [V]	0.0603	0.0553
Number of FEA	221	6645
Number of FEA (relative)	1	30
Wall clock time [hours:min]	8:17	62:12
Wall clock time (relative)	1	7.5

allel Computing in order to make the uncertainty-based optimization practical. Furthermore, the fact that the anti-optimizations involve only parts of the total analysis, combined with the high utilization of the parallel computing resources by the use of Nested Parallel Computing, makes that the uncertainty-based optimization procedure is 4 times more efficient, relative to the deterministic case. This is illustrated by the fact that the duration of the uncertainty-based case equals merely 7.5 times the duration of the deterministic optimization, while 30 times as many function evaluations are involved.

The geometries of the SMA gripper corresponding to the optimal designs obtained in the deterministic and uncertainty-based optimization are depicted in Figure 5.21. From the top and side views, it can clearly be seen that the deterministic design is wider at the front and also higher. A geometrical difference that cannot be seen in Figure 5.21 is that the plates are 5% thicker in the uncertainty-based design. The operation of the gripper is demonstrated in Figure 5.22, which shows a side view of the optimal uncertainty-based gripper in the open and closed configuration.

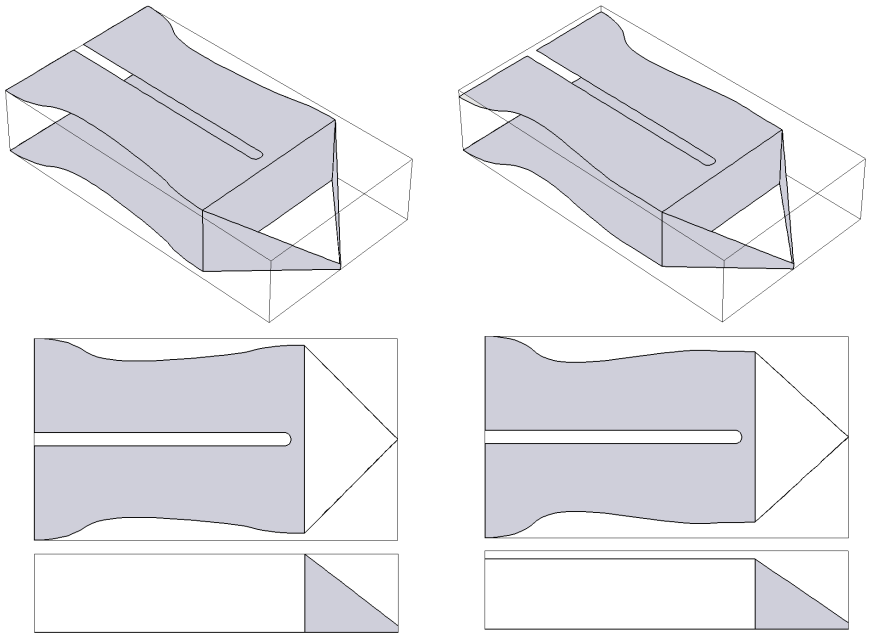


Figure 5.21: *Undeformed geometries of the optimal designs obtained by deterministic optimization (left) and optimization considering bounded-but-unknown uncertainties (right).*

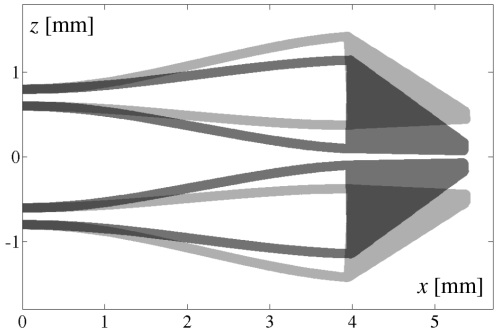


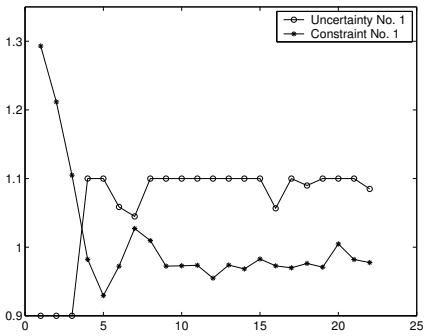
Figure 5.22: *Side view of the optimized gripper in opened (light gray) and closed (dark gray) configuration.*

In the case of uncertainty-based optimization, the values of the uncertainty variables that yield the worst (i.e. largest) constraint values for the final design are listed in Table 5.7 for all four constraints. For the temperature constraints, only the uncertainties in the thermal quantities are meaningful, and it turns out that indeed their worst case values are as predicted earlier. Inspection shows that all of the uncertainty variables are at either their lower or upper bound for the final design, except the face convection coefficient h at the strain constraint for the opened case $KS_A(g_\varepsilon^{\text{open}})$. Evaluation of this constraint with h at its lower and upper bound confirmed that indeed the worst value is obtained for h at an interior value.

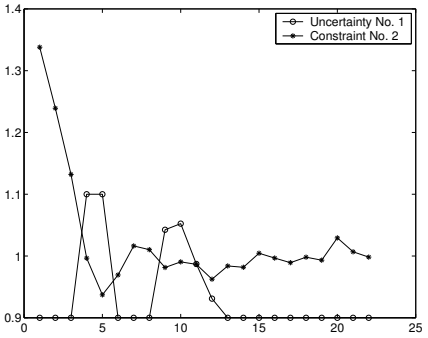
When considering the evolution of this uncertainty variable h during the optimization process, as depicted in Figure 5.23, it turns out that in many other steps also interior worst-case values are found, in both strain constraints. Also, for the

Table 5.7: *Worst set of uncertainties obtained for the uncertainty-based optimum for each constraint for the SMA microgripper problem.*

Normalized uncertainty	$KS_A(g_\varepsilon^{\text{open}})$	$KS_A(g_\varepsilon^{\text{closed}})$	$KS_A(g_T^{\text{open}})$	$KS_A(g_T^{\text{closed}})$
h/h	1.08485	0.9	0.9	0.9
$T_a/\bar{T}_a$	1.00122	1.00122	1.00122	1.00122
$E_A/\bar{E}_A$	1.08792	1.08792	-	-
$E_R/\bar{E}_R$	0.94273	0.94273	-	-
$E_0/\bar{E}_0$	0.97470	0.97470	-	-
$\nu/\bar{\nu}$	1.08	1.08	-	-



Plot for the normalized worst value of the uncertainty in face convection coefficient $h/\bar{h}$ and the worst value of constraint $KS_A(g_\varepsilon^{\text{open}})$ obtained by anti-optimization corresponding to suboptimal design at each step.



Plot for the normalized worst value of the uncertainty in face convection coefficient $h/\bar{h}$ and the worst value of constraint $KS_A(g_\varepsilon^{\text{closed}})$ obtained by anti-optimization corresponding to suboptimal design at each step.

Figure 5.23: *Anti-optimization results in terms of worst set of uncertainties and constraint corresponding to suboptimal design at each step for the SMA microgripper problem.*

strain constraint in the opened case $KS_A(g_\varepsilon^{\text{open}})$, the uncertainty variable h initially stays at the lower bound, but changes to values equal or close to the upper bound. From this observation, it can be concluded that the present uncertainty-based SMA gripper optimization problem indeed exhibits fluctuating uncertainties. The worst case for uncertainty variable h clearly is design-dependent, and also takes interior values. The complexity and nonlinearity of the model leads to non-monotonicity in the uncertainty-based design optimization. In general, for complex models, it is hard to predict which combination of uncertainties will result in the worst constraint values, and whether even interior worst cases are possible. Therefore, the proposed anti-optimization technique is a good choice for such problems, even though it is computationally more involved than approaches that rely on, e.g., vertex checking.

5.7 Discussion and conclusions

The Cycle-based Alternating anti-optimization technique combined with nested parallel computing has successfully been applied to the design optimization of an SMA microgripper involving bounded-but-unknown uncertainties. This allows for practical optimization-based design of SMA devices in situations where, for example, material properties and operating conditions are not exactly known, but where bounds on these unknown quantities can be specified. The worst set of uncertainties turned out be design-dependent, and the Cycle-based Alternating technique proved to be able to handle this situation effectively.

The optimization problem involving bounded-but-unknown uncertainties is solved using an anti-optimization technique. To demonstrate and compare various anti-optimization approaches, the elastically supported beam problem from literature has been used. For this test problem, the worst set of uncertainties is also strongly dependent on the design, and fluctuates considerably from design to design. This situation is a challenging case for efficient anti-optimization, however, many problems of realistic complexity, such as the SMA microgripper problem, share this characteristic. For such problems exhibiting fluctuating worst uncertainties, it is shown that the Cycle-based Alternating technique is more efficient than the existing Lombardi-Haftka Alternating technique in terms of convergence and required computational effort. Rigorous anti-optimization proves to be better than the cycle-based techniques in terms of convergence properties. However, since the number of function evaluations required for the rigorous approach can quickly become prohibitive as the problem size increases, in terms of the number of design variables, uncertainties and constraints, the rigorous approach quickly becomes impractical for problems involving computationally expensive function evaluations.

Further improvement of the effectiveness of the Cycle-based Alternating technique is possible with the use of derivative information. It is demonstrated by means of the elastically supported beam problem that combining the Cycle-based Alternating technique with the Asymptotic method, which uses derivative information for prediction of worst cases, is significantly more efficient than all other examined techniques, in terms of both the convergence and number of function evaluations. In the present study however, this combined Cycle-based Alternating and Asymptotic method is not applied in the SMA microgripper optimization, because of the lack of the required derivative information.

For the problems involving computationally expensive response evaluations, use of parallel computing is essential to make the above optimization techniques computationally feasible. For the SMA microgripper problem, which involves computationally expensive finite element simulations, the Cycle-based Alternating technique combined with nested parallel computing enabled the practical application of anti-optimization. For this problem, the total number of FEAs required for uncertainty-based optimization is 30 times that of the deterministic optimization. However, a high utilization of the computing cluster due to Nested Parallel Computing strategy in combination with a smart treatment of function evaluations in the anti-optimization procedures reduced the total wall clock time for uncertainty-based optimization to merely 7.5 times that of deterministic optimization. Hence, for this modest computational cost, the proposed techniques enable designers to account for bounded-but-unknown uncertainties in the design optimization process, and it has been shown that this approach is indeed feasible even for relatively complex and large problems as the considered SMA structure. It is therefore expected, that by making it feasible to account for uncertainties in the design process in a systematic way, the proposed methodology contributes to the development of better, safer and more reliable SMA structures and devices.

Based on: Langelaar, M., Yoon, G.H., Kim, Y.Y. and van Keulen, F. Topology optimization of shape memory alloy thermal actuators using element connectivity parameterization. *Computer Methods in Applied Mechanics and Engineering*, in review.

Chapter 6

Topology Optimization of SMA Actuators

This chapter presents the first application of topology optimization to the design of shape memory alloy actuators. Shape memory alloys (SMAs) exhibit strongly nonlinear, temperature-dependent material behavior. The complexity in the constitutive behavior makes it difficult to apply the conventional element density-based topology optimization to the topology design of SMA structures. Therefore, in the present study, the recently developed element connectivity parameterization (ECP) formulation is applied, which offers important advantages for complex nonlinear topology optimization problems. A history-independent constitutive model of SMAs is employed which allows efficient adjoint sensitivity analysis. The effectiveness of the proposed technique is illustrated by several numerical examples.

6.1 Introduction

6.1.1 Shape memory alloys

This chapter presents a procedure to perform topology optimization of shape memory alloy (SMA) structures. In SMA materials, a solid state phase transformation can take place under local temperature change and/or stress state variation. This transformation is associated with a transformation strain that can be utilized for actuation. SMA actuators are applicable in situations requiring relatively large deflections, combined with substantial mechanical loading (Duerig *et al.*, 1990, Otsuka and Wayman, 1998). SMAs are used in a wide variety of ap-

plications, and present developments particularly focus on emerging applications in medicine and microsystems (Büttgenbach *et al.*, 2001, Duerig *et al.*, 1999, Kohl, 2004, Van Humbeeck, 1999).

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 presently existing SMA actuators rely on a relatively simple layouts, e.g. straight wires or helical springs. Clearly, for such layouts, understanding the behavior of a one-dimensional SMA wire or spring is sufficient for effective design (see e.g. Liang and Rogers, 1992a). This changes radically as soon as SMA actuators of more complex geometries have to be designed. As a consequence, it is necessary to understand and model the material behavior in a 2D or 3D setting. The complex material behavior and the fact that often several physical domains (mechanical, thermal, electrical) are involved make it very difficult to design effective SMA actuators by intuition. Therefore, utilization of powerful systematic design techniques such as topology optimization is crucial for the realization of these structures.

6.1.2 Topology optimization

Topology optimization aims to find the optimal material distribution in a given domain to fulfill a certain objective. The technique, initially developed for the design of linear elastic structures (Bendsoe and Kikuchi, 1988), has been applied to a wide variety of engineering problems (see e.g. Bendsoe and Sigmund, 2003). In comparison to the alternative shape and sizing design optimization approaches, the unique strength of topology optimization is that it can generate arbitrary topologies. The parameterization used in shape and sizing optimization is typically much more restrictive. On the other hand, in topology optimization the analysis is generally less accurate, and it is harder to include e.g. stress or buckling constraints. These properties make that topology optimization is most applicable in the initial design stage. However, since decisions made in that stage have a strong effect on the final performance of a design, the merit of topology optimization is evident.

The conventional approach basically consists of controlling the material parameters of elements in a finite element model, e.g. the Young's moduli, using density design variables. In this way, in e.g. structural problems, the presence of void regions in the structure can be approximated by very low stiffness elements. The majority of the reported applications is restricted to linear analysis models. However, next to the expansion of topology optimization to other physical domains, in recent years, researchers have started to explore its application to nonlinear problems. The focus is mainly on structural problems involving geometrical nonlinearities and/or elastoplasticity (Buhl *et al.*, 2000, Jung and Gea, 2004, Maute *et al.*, 1998). These nonlinear problems pose new challenges to the conventional topology optimization approach, for instance due to the excessive distortion of highly compliant elements (Yoon and Kim, 2005). In addition, in

case of complex nonlinear materials such as SMAs, the conventional approach of relating material properties to design variables easily leads to complications and ambiguity (Pedersen, 2002).

The Element Connectivity Parameterization (ECP) approach proposed by Yoon and Kim (2004, 2005) presents a new paradigm that avoids these difficulties. In this ECP scheme, not the properties of elements, but the connectivity between the elements is controlled by design variables. This is accomplished by introducing zero-length links between elements. By changing the stiffness of these links, different topologies can be realized. A drawback of the originally proposed "external" ECP method (Yoon and Kim, 2004, 2005) is, that the number of degrees of freedom increases considerably. However, a recently developed "internal" ECP version (Yoon *et al.*, 2005) is employed here, in which the introduction of additional nodes, a modified arrangement of elements and links, and the application of condensation completely eliminates this disadvantage.

6.1.3 SMA material modeling

The application of topology optimization to the design of SMA structures has not been reported before. Next to the outlined difficulties involved in topology optimization of nonlinear problems in general, this is most likely due to the lack of sufficiently simple, though practical SMA material models that are suited for optimization, in spite of the large number of models that have already been developed (Birman, 1997). In this chapter, a new SMA model is presented, which is well suited for (topology) optimization. The shape memory effect considered is superelasticity due to the R-phase/austenite transformation in NiTi. Generally, SMAs exhibit hysteresis in the stress-strain-temperature behavior. However, in case of this transformation, the hysteresis is particularly small, and the cyclic stability of the transformation is excellent. These are attractive properties for actuator applications (Kohl, 2004, Otsuka and Wayman, 1998). However, the reversible transformation strain of the R-phase transformation is considerably less (ca. 0.8%) than that of the martensite transformation in NiTi (up to ca. 7%). Therefore, it is crucial to optimize designs to utilize the relatively small transformation strain in the most effective way.

6.1.4 Outline

This chapter is organized as follows: in Section 6.2, the new SMA material model is presented, together with a comparison to experimental data. A one-dimensional constitutive model is generalized to a three-dimensional setting, and subsequently also the formulation in a plane stress situation is discussed. Detailed derivations can be found in Appendices A till D. In addition, in Section 6.2.6, the strategies used to ensure robust convergence of the SMA finite element model used in the topology optimization procedure are discussed.

Section 6.3 is devoted to the modified Element Connectivity Parameterization. After discussion of the difficulties encountered in the conventional density-based topology optimization approach in nonlinear problems, the ECP concept is presented as well as its detailed formulation. An important advantage of the ECP method is that, in contrast to density-based approaches using design-dependent material properties, the sensitivity analysis does not require differentiation of the material model, as shown in Section 6.3.4. This aspect is particularly important in case of complex nonlinear material models, where differentiation can quickly become quite involved. Section 6.3 is concluded by a description of a new interpolation function used in the proposed ECP formulation, that allows for the use of penalization similar to the popular SIMP model (Solid Isotropic Material with Penalization, Bendsoe and Sigmund, 2003).

In Section 6.4, regularization techniques are studied in order to avoid certain numerical artifacts of the topology optimization process. It is shown, that limiting the design space by the use of nodal design variables results in the desired behavior. Subsequently, Section 6.5 presents the results of the application of topology optimization to the design of SMA thermal actuator structures. Various results of a study on the characteristics of the SMA topology optimization problem are presented, including the effect of the load magnitude and direction, the mesh density, and the initial design. Finally, based on the observations made in this study, conclusions are drawn with regard to the potential and effectiveness of the developed techniques for SMA topology optimization in Section 6.6.

6.2 SMA modeling

6.2.1 Design optimization considerations

In order to assist the design process of SMA structures, and to obtain a deeper understanding of the underlying phenomena that lead to the peculiar behavior of SMA materials, much research has already been done on the computational modeling of SMA constitutive behavior (Birman, 1997). In contrast to existing models, the SMA constitutive model presented here is developed with an emphasis on its suitability for design optimization, as well as on its ability to adequately represent the observed material behavior. Design optimization generally is an iterative process that requires many evaluations of the analysis model. Thus, to make design optimization practically feasible, it is important to consider the efficiency of the analysis model. In the present work, as a first simplification, the dynamic response of the actuator is neglected, and a quasistatic simulation is performed. This is sufficient in many applications where the response speed is not critical.

A second aspect that is of great importance in design optimization – and particularly topology optimization – is the computational complexity of the sensitivity analysis. In history-dependent models, sensitivity analysis is much more

involved than in the history-independent case, and adjoint sensitivity analysis is impractical (Kleiber *et al.*, 1997). Internal variables, such as phase fractions used in most of the existing SMA models, render those models history-dependent. By neglecting the (rather small) hysteresis of the R-phase transformation, it turns out to be possible to eliminate internal state variables, resulting in a history-independent model. In that case, (adjoint) sensitivity analysis can be carried out at the end of the analysis, without the need to account for every increment during the evolution of the model over time.

6.2.2 One-dimensional R-phase transformation modeling

Figure 6.1 shows experimental data obtained by Tobushi *et al.* (1992) for the stress-strain-temperature behavior of a Ti-55.3wt%Ni alloy exhibiting the R-phase transformation. The modeling procedure described in the present chapter focuses on this particular alloy, but it can similarly be applied to other alloys in which the R-phase transformation occurs. The hysteresis between loading and unloading behavior is practically negligible at temperatures above 328 K. The focus of the present model is mainly on applications with a limited thermal operating range, such as (*in vivo*) actuator applications, and the range 328–343 K has been selected as the temperature range of interest. Given this small range, strains due to thermal expansion are ignored in this formulation. The temperatures are still high for *in vivo* use, but the range can be shifted by changes in the alloy composition or thermal processing (Gyobu *et al.*, 1996, Sawada *et al.*, 1993).

Several one-dimensional models for the stress-strain-temperature behavior for the R-phase transformation in NiTi have been published, and a review is given by Langelaar and Van Keulen (2004c) (Chapter 2). By focusing on a temperature range where the hysteresis is negligible, simplifications can be applied. The one-dimensional model used in the present research is similar to the model proposed earlier (Langelaar and Van Keulen, 2004c) (Chapter 3), and is summarized here. Based on the shape of the experimental curves, the stress-strain relations are approximated by a three-part piecewise linear function, as depicted in Figure 6.1:

$$\sigma_{xx} = \begin{cases} \varepsilon_{xx} \leq \varepsilon_1 & : \quad E_A \varepsilon_{xx}, \\ \varepsilon_1 < \varepsilon_{xx} \leq \varepsilon_2 & : \quad E_T(\varepsilon_{xx} - \varepsilon_1) + E_A \varepsilon_1, \\ \varepsilon_{xx} > \varepsilon_2 & : \quad E_R(\varepsilon_{xx} - \varepsilon_2) + E_T(\varepsilon_2 - \varepsilon_1) + E_A \varepsilon_1, \end{cases} \tag{6.1}$$

where E_R is a constant parameter. In order to represent the experimental data well, $\varepsilon_1, \varepsilon_2$, E_T and E_A are chosen to be linearly dependent on the temperature T, in the following way:

$$\begin{aligned} \varepsilon_1(T) &= K_\varepsilon(T - T_0) + \varepsilon_0, \\ \varepsilon_2(T) &= \varepsilon_1(T) + \Delta, \\ E_T(T) &= K_E(T - T_0) + E_0, \\ E_A(T) &= K_A(T - T_0) + E_A^0. \end{aligned} \tag{6.2}$$

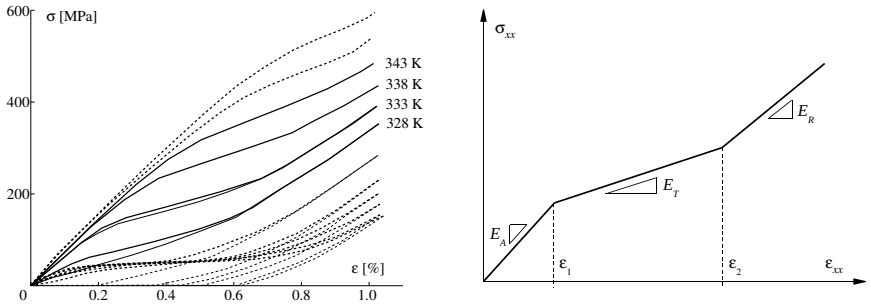


Figure 6.1: *Experimentally obtained stress-strain curves at various temperatures (Tobushi* et al.*, 1992). Thin lines are unloading curves (left). Schematic representation of piecewise linear one-dimensional SMA stress-strain relation (right).*

ε_1 and ε_2 are the transition strains that mark the points where the piecewise linear function changes its slope. In this approximation, these points correspond to the strains at which the R-phase transformation starts and finishes. E_T is the apparent Young's modulus $d\sigma_{xx}/d\varepsilon_{xx}$ during the phase transition. E_A can be interpreted as the Young's modulus of the austenite phase, and was taken as a constant in the previous model (Langelaar and Van Keulen, 2004c) (Chapter 3). However, for the topology optimization problems considered in the present chapter, it has been found that making E_A depend on temperature as well provides more useful sensitivity information in low-stressed areas of the design. This idea has been implemented as follows:

$$E_A(T) = K_A(T - T_0) + E_A^0 \tag{6.3}$$

Here E_A^0, K_A and T_0 are parameters of the material model, and T represents the temperature. This modification improves the convergence of the optimization procedure. Next to this, it also provides a slightly better approximation of the experimental data.

The proposed model is calibrated to experimental data by determining the material parameters using curve fitting. Before the actual fitting, a conversion is applied to the experimental data. The finite element implementation is based on the Green-Lagrange strain ε_{GL} and the Second Piola-Kirchhoff stress σ_{PK2}. The one-dimensional experimental data is given in terms of nominal stress σ_{nom} and engineering strain ε_{eng}. The equivalent Green-Lagrange strain and Second Piola-Kirchhoff stress can be obtained from these quantities by:

$$\varepsilon_{GL} = \frac{1}{2}\varepsilon_{eng}^2 + \varepsilon_{eng}, \tag{6.4}$$

$$\sigma_{PK2} = \frac{\sigma_{nom}}{1 + \varepsilon_{eng}}. \tag{6.5}$$

Table 6.1: *Parameter values for the SMA model.*

K_ε	2.47×10^{-4} K^{-1}	K_E	678 MPa K^{-1}	ε_0	5.89×10^{-4}
Δ	52.7×10^{-4}	E_0	18.8 GPa	T_0	328 K
E_A^0	66.3 GPa	E_R	45.6 GPa	K_A	188 MPa K^{-1}

The converted experimental data is used in the least squares fitting process, and the parameter values that were subsequently obtained are listed in Table 6.1.

This one-dimensional model will be referred to by $\sigma_{xx}(\varepsilon_{xx}, T)$. For subsequent derivations, it is useful to present an alternative description of the piecewise linear stress-strain equations in the following generalized form:

$$\sigma_{xx} = A_i(T)\varepsilon_{xx} + B_i(T). \tag{6.6}$$

The terms $A_i(T)$ and $B_i(T)$ follow from the original formulation given in Equation 6.1, and the indices refer to the three segments of the piecewise linear relation:

$$A_i(T) = \begin{cases} \varepsilon_{xx} \le \varepsilon_1 & : \quad A_1 = E_A, \\ \varepsilon_1 < \varepsilon_{xx} \le \varepsilon_2 & : \quad A_2 = E_T, \\ \varepsilon_{xx} > \varepsilon_2 & : \quad A_3 = E_R, \end{cases} \tag{6.7}$$

$$B_i(T) = \begin{cases} \varepsilon_{xx} \le \varepsilon_1 & : \quad B_1 = 0, \\ \varepsilon_1 < \varepsilon_{xx} \le \varepsilon_2 & : \quad B_2 = (E_A - E_T)\varepsilon_1. \\ \varepsilon_{xx} > \varepsilon_2 & : \quad B_3 = (E_A - E_T)\varepsilon_1 + (E_T - E_R)\varepsilon_2. \end{cases} \tag{6.8}$$

6.2.3 Three-dimensional modeling

The transformation pseudoelasticity due to the R-phase transformation is described by a path-independent nonlinear thermo-elastic formulation. Use is made of the experimentally observed characteristics of the R-phase transformation, namely that it is associated with an isochoric transformation strain, and that the transformation behavior is insensitive to hydrostatic pressure (Kakeshita *et al.*, 1992, Miyazaki and Wayman, 1988). The starting point of this formulation is Hooke's Law for linear elasticity, split into volumetric and deviatoric parts:

$$\boldsymbol{\sigma} = K(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I} + 2G\left(\boldsymbol{\varepsilon} - \frac{1}{3}(\operatorname{tr}\boldsymbol{\varepsilon})\mathbf{I}\right), \tag{6.9}$$

where K and G represent the shear and bulk moduli, respectively.

In Mandel notation, this can also be written as:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sqrt{2}\sigma_{xy} \\ \sqrt{2}\sigma_{yz} \\ \sqrt{2}\sigma_{zx} \end{Bmatrix} = \left(K \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \dots \right. \\ \left. 2G \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right) \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \sqrt{2}\varepsilon_{xy} \\ \sqrt{2}\varepsilon_{yz} \\ \sqrt{2}\varepsilon_{zx} \end{Bmatrix}, \tag{6.10}$$

which is equivalent to, in short:

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2G\mathbf{G})\boldsymbol{\varepsilon}. \tag{6.11}$$

Upon transformation, the isochoric phase transformation strain must be accounted for. In this nonlinear elastic model, it is found that the distortional strain energy is minimized when the transformation strain $\boldsymbol{\varepsilon}_{tr}$ is aligned with the distortional strain component $\mathbf{G}\boldsymbol{\varepsilon}$. This can equivalently be represented by a variable shear modulus $\tilde{G}$:

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2G\mathbf{G})(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_{tr}) \quad \text{where} \quad \boldsymbol{\varepsilon}_{tr} = \beta\mathbf{G}\boldsymbol{\varepsilon} \quad \Rightarrow \tag{6.12}$$

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2\tilde{G}\mathbf{G})\boldsymbol{\varepsilon} \quad \text{with} \quad \tilde{G} = G(1-\beta). \tag{6.13}$$

Here use has been made of $\mathbf{KG} = [\mathbf{0}]$ and $\mathbf{GG} = \mathbf{G}$. To relate the value of the effective shear modulus $\tilde{G}$ to the strain state, an scalar equivalent strain measure is used based on the deviatoric strain energy, given by:

$$\varepsilon_e^2 = \frac{2}{3}\boldsymbol{\varepsilon}^T\mathbf{G}\boldsymbol{\varepsilon}. \tag{6.14}$$

This scalar distortion measure is invariant, and equal to $4/3(I_2 + I_1^2/3) = 4/3J_2$, where I_1 and I_2 are the first and second invariant of the Green-Lagrange strain tensor, and J_2 is the second invariant of its distortional part. The factor 2/3 is included to make this strain measure energy-conjugated to the Von Mises stress, which covers the pressure-insensitive property of the transformation. The constitutive stress-strain relation for R-phase transformation pseudoelasticity now simply takes the following form:

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2G(\varepsilon_e)\mathbf{G})\boldsymbol{\varepsilon}. \tag{6.15}$$

Note that here the shear modulus can vary depending on the value of the effective strain, and the tilde used before in Equation 6.13 to indicate this variability

will be omitted from this point onwards. The relation between the value of the shear modulus and the effective strain $G(\varepsilon_e)$ is found by considering a tensile test. In that case, the stress-strain behavior is described by the one-dimensional model $\sigma_{xx}(\varepsilon_{xx}, T)$ presented in Equation 6.1. In addition, the proposed three-dimensional model given by Equation 6.10 in combination with Equation 6.14 yields:

$$\varepsilon_{yy} = \varepsilon_{zz} = \frac{2G - 3K}{2G + 6K}\varepsilon_{xx} \quad \Rightarrow \tag{6.16}$$

$$\varepsilon_e = \frac{3K}{G + 3K}\varepsilon_{xx}, \quad \varepsilon_{xx} = \frac{G + 3K}{3K}\varepsilon_e \tag{6.17}$$

$$\sigma_{xx} = \frac{9GK}{G + 3K}\varepsilon_{xx} = 3G\varepsilon_e \tag{6.18}$$

Combining these expressions with the one-dimensional model results in a single equation that can be used to determine G for a given value of ε_e:

$$\sigma_{xx}(\varepsilon_{xx}, T) = \sigma_{xx}\left(\frac{G + 3K}{3K}\varepsilon_e, T\right) = 3G\varepsilon_e. \tag{6.19}$$

Using the generalized notation introduced in Equation 6.6 subsequently gives:

$$A_i(T)\frac{G + 3K}{3K}\varepsilon_e + B_i(T) = 3G\varepsilon_e \quad \Rightarrow \tag{6.20}$$

$$G = \frac{3K}{9K - A_i(T)}\left(A_i(T) + \frac{B_i(T)}{\varepsilon_e}\right). \tag{6.21}$$

This final expression relates the shear modulus directly to the effective strain. Finally, the tangent operator of this constitutive model is found by differentiation of Equation 6.15 to the strain, which yields:

$$[K]_T = K\mathbf{K} + 2G\mathbf{G} + \frac{4}{3\varepsilon_e}\frac{dG}{d\varepsilon_e}\mathbf{G}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{G}. \tag{6.22}$$

The detailed derivation is given in Appendix A.

6.2.4 Plane stress case

Using the general 3-D formulation, it can be derived that in plane stress, the transverse strain ε_{zz} is related to the in-plane strain components ε_{xx} and ε_{yy} by:

$$\varepsilon_{zz} = \frac{2G - 3K}{4G + 3K}(\varepsilon_{xx} + \varepsilon_{yy}) = \alpha(\varepsilon_{xx} + \varepsilon_{yy}), \tag{6.23}$$

where the symbol α is introduced to represent the term $(2G - 3K)/(4G + 3K)$, to reduce the complexity of the expressions. Note that α depends on G. Using this

relation, the transverse strain component can be eliminated from the stress-strain equation (Equation 6.10), leading to:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sqrt{2}\sigma_{xy} \end{Bmatrix} = \left(K \begin{bmatrix} 1+\alpha & 1+\alpha & 0 \\ 1+\alpha & 1+\alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} + \ldots \right. \\ \left. \frac{2G}{3} \begin{bmatrix} 2-\alpha & -(1+\alpha) & 0 \\ -(1+\alpha) & 2-\alpha & 0 \\ 0 & 0 & 3 \end{bmatrix} \right) \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix}. \tag{6.24}$$

The same can be done for the effective strain definition, which becomes:

$$\varepsilon_e^2 = \frac{4}{9}(\alpha^2 - \alpha + 1)(\varepsilon_{xx}^2 + \varepsilon_{yy}^2) + \frac{4}{9}(2\alpha^2 - 2\alpha - 1)\varepsilon_{xx}\varepsilon_{yy} + \frac{4}{3}\varepsilon_{xy}^2. \tag{6.25}$$

This indicates a slight complication: in the 3-D case, the effective strain only depends on the strain components (Equation 6.14). In the plane stress case, however, by elimination of the transverse strain, the effective strain also has become a function of α, which in turn depends on the shear modulus G. It turns out, no convenient explicit expression can be found to express G as a function of ε_e, as in the 3-D case. Combining Equation 6.21, Equation 6.25 and the definition of α gives a nonlinear equation in G, which can be solved numerically for a given temperature and strain situation. Newton-Raphson iterations are used to obtain the solution. Convergence is robust and usually requires only 3 to 4 iterations. Details on the equations can be found in Appendix B.

Also in the plane stress case a tangent operator is required in the finite element implementation. There are two ways to derive it: the first is to take the tangent operator of the three-dimensional case, and reduce it to a plane stress setting by eliminating the transverse strain component ε_{zz} from the equations. This is worked out in more detail in Appendix C.1. The second way to derive the tangent operator is to start from the stress-strain relation in the plane stress case (Equation 6.24) and to differentiate it with respect to the strain. However, as in this case there exists no explicit relation between the shear modulus and the (effective) strain, use has to be made of implicit differentiation, which makes this a rather complicated derivation. The details are given in Appendix C.2. Naturally, the resulting tangent operator is the same in both cases, regardless of the derivation. It is symmetric in case of a finite element formulation based on the stress vector $(\sigma_{xx}, \sigma_{yy}, \sigma_{xy})$ and strain vector given by $(\varepsilon_{xx}, \varepsilon_{yy}, \gamma_{xy})$, as commonly used in implementations.

6.2.5 Verification and discussion

The presented SMA constitutive model in the plane stress formulation has been implemented in a standard four-noded quadrilateral element (see e.g. Bathe,

1995). This element is most popular in topology optimization, and is also used in this research. The implementation of the material model has been verified by simulating a tensile test. Figure 6.2 shows the good agreement between experimental stress-strain data from Tobushi *et al.* (1992) and results obtained using the finite element analysis.

More particular constitutive aspects such as the influence of texture, multiaxial non-proportional loading effects and tension-compression asymmetry are not considered in the present model, for several reasons. On the one hand these effects are considered to be refinements of the principal material behavior, and of lesser importance for the present application. The present model is mainly aimed at an efficient formulation that is able to capture the first-order characteristics, with a sufficient degree of accuracy for design optimization in an early design stage. Also, firm and conclusive experimental evidence that allows incorporation of these effects in the formulation is currently still lacking.

On the other hand, including more detailed constitutive effects will almost certainly require the use of additional internal variables in the model, which will render the adjoint sensitivity computation essential for topology optimization intractable. A simpler and even more computationally efficient model has been formulated before (Langelaar and Van Keulen, 2004c) (Chapter 2), however that model did not account for the isochoric transformation strain, and the transformation was not pressure-independent. Therefore the present model is a practical compromise between representing the physical characteristics of the material as good as possible and maintaining suitability for (topology) optimization.

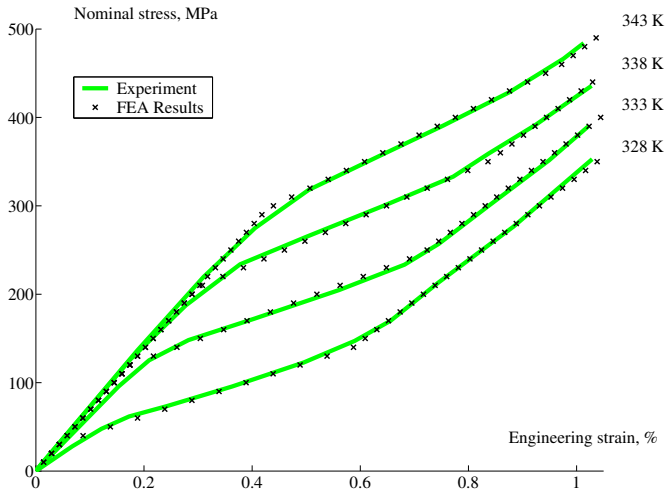


Figure 6.2: *Simulated and experimentally determined stress-strain curves at various temperatures.*

6.2.6 Robust analysis techniques for optimization

A vital property of the simulation model used for the topology optimization procedure is robustness. The SMA constitutive model is strongly nonlinear, and also geometrical nonlinearities are considered in the finite element analysis. To evaluate the responses necessary for the next step in the optimization process, the analysis is performed using an incremental approach combined with Newton iterations. Newton iterations have a tendency to overshoot the solution, which can in turn lead to excessive corrections and subsequent divergence. Closer to the equilibrium solution, the Newton process is stable and converges quadratically. By applying the load in a number of smaller increments, the deviation from the equilibrium path is reduced and the Newton iterations in each increment are more likely to converge. However, the total computational effort required for the analysis generally increases with the number of increments. Since topology optimization itself is already a time-consuming iterative process, it is important to strive towards using the minimum number of increments the analysis requires.

It has also been observed in case of the present SMA finite element model, when the configuration of the system deviates too far from the equilibrium path, the tangent operator in some cases can lose its positive definiteness during the Newton iterations. This subsequently lead to divergence (when using a general-purpose solver for the non-positive definite system). We believe this behavior is the result of a combination of the strong and sharp nonlinearities in the SMA model, the inclusion of geometrical nonlinearities and the overshoot and overcorrection tendency of the Newton iterations. The present material model involves significant changes in effective stiffness due to the phase transformation, which lead to numerical convergence problems. Conventional remedies, such as using relaxed or damped Newton variants worked to some degree, but the parameters involved turned out to be strongly case-dependent. Adding a line search to the Newton iteration was found to work robustly, but was prohibitively expensive. Therefore, to prevent this type of numerical problem, sufficiently small increments turned out to be the most practical solution.

To combine the two objectives, robustness and efficiency, an adaptive solution strategy is applied to the SMA analysis used for topology optimization, schematically depicted in Figure 6.3. Since in topology optimization, changes in the design are generally gradual, the changes in the response are generally small as well, in consecutive optimization steps. Starting the Newton iterations from the solution obtained for the preceding design therefore can lead to very fast convergence. This procedure is referred to in Figure 6.3 as “Continued analysis”. To increase the chances of success, it is only attempted under certain conditions, as described in Figure 6.3. If in spite of these precautions this continued analysis fails, a full analysis is performed using N increments. If this full analysis also does not converge, the setting for the number of increments N is increased and another attempt is made. This continues until a certain maximum, so that the process can not continue indefinitely.

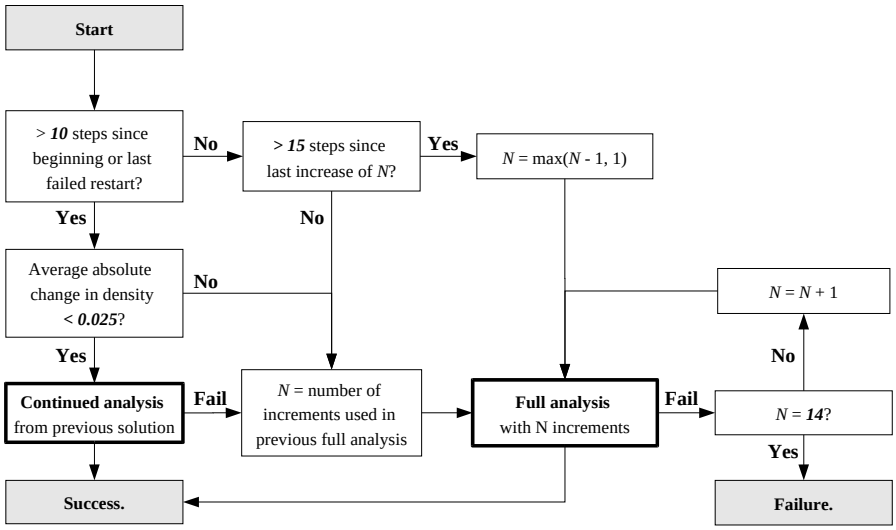


Figure 6.3: *Solution strategy used in the topology optimization analysis.*

This strategy proved to work very well in practice. Particularly the continued analysis, possible because a history-independent model is used, resulted in swift convergence. No detailed study was made to obtain the optimal settings for this strategy, but using the constants shown in Figure 6.3 proved to result in satisfactory behavior. Note that although this strategy dramatically improved the robustness of the analysis, it cannot guarantee convergence in all cases. In fact, in the Results section, a case will be described where the analysis eventually failed. But in the majority of practical cases the proposed strategy resulted in a robust as well as efficient solution process.

6.3 Element Connectivity Parameterization method

6.3.1 Difficulties in conventional method (density approach)

Three difficulties arise when the conventional density-based topology optimization method is applied to nonlinear problems. First, an interpolation needs to be defined between the material properties of elements and their densities. The behavior of the optimization strongly depends on this interpolation, and in case of complex nonlinear material models, finding a proper interpolation can be very difficult (Pedersen, 2002). The second problem is that because material properties are functions of the design variables, the sensitivity analysis can require a full

differentiation of the material model, which can become very involved for complex material models. Finally, in the density-based approach, elements that should not be part of the design are represented by very compliant material. As noted before by Yoon and Kim (2005), among others, in nonlinear finite element models these compliant elements can suffer from excessive distortion, due to which their tangent stiffness matrices can lose positive definiteness. This seriously affects stability and convergence of the entire model.

6.3.2 Basic idea and mesh layout

Because of the observed difficulties of the conventional density-based method, the recently developed element connectivity parameterization (ECP) formulation (Yoon and Kim, 2004, 2005, Yoon *et al.*, 2006b) is used here. Unlike the density based approach, the layout is not described by varying the densities and modifying the material properties of elements, but by varying the connectivity between elements. Computationally, this concept is implemented using zero-length links that connect the nodes of neighboring elements with additional nodes, see Figure 6.4. The internal element can involve geometrical and material nonlinearities, however the connecting links behave linearly. The stiffnesses of the links connected to an individual element are controlled by a single design variable.

Note that parameterization of element connectivity requires separation of elements, which increases the number of degrees of freedom of the problem significantly. This has been identified as a disadvantage of the original external ECP approach (Yoon and Kim, 2004, 2005). The present arrangement with additional nodes is chosen because it allows for condensation of the element-link patch (see Figure 6.4), which reduces the number of degrees of freedom to that of the original mesh with directly connected elements. In the condensed patch, the element connectivity is dealt with internally, hence this approach is denoted as internal ECP.

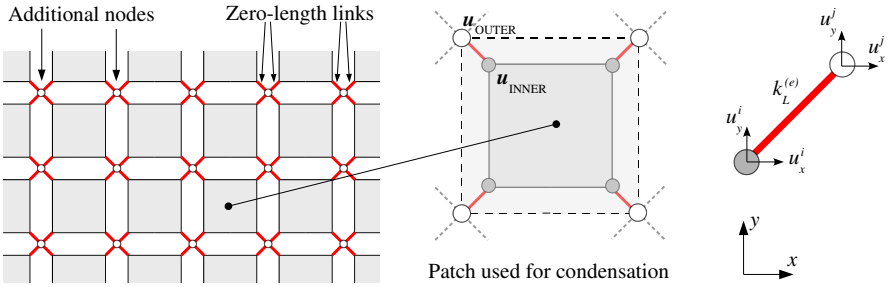


Figure 6.4: *Mesh and cell layout used in the internal element connectivity parameterization approach.*

External forces and boundary conditions are applied to the additional nodes.

Because element connectivity instead of element properties is parameterized, clearly there is no need to define a material model interpolation. In addition, sensitivity analysis for ECP is straightforward, because only the zero-length links are affected by the design variables. Finally, since elements remain stiff, and only the stiffness of the one-dimensional links between the elements is modified, numerical problems due to excessive distortion of compliant elements are prevented (Yoon and Kim, 2005). For these reasons, the ECP approach is very suitable for topology optimization of structures involving complex nonlinear materials, such as the SMA actuators studied in this thesis.

6.3.3 Governing equations

Without loss of generality, the focus of the present research is on two-dimensional plane stress problems, to limit the required computational effort. A zero-length link element in the considered 2-D setting, connecting nodes i and j is governed by:

$$\mathbf{K}_L^{(e)} \left\{ u_x^i \quad u_y^i \quad u_x^j \quad u_y^j \right\}^T = \left\{ F_x^i \quad F_y^i \quad F_x^j \quad F_y^j \right\}^T, \tag{6.26}$$

where the nodal displacement components u_α^β, where $\alpha = x$ or y, $\beta = i$ or j, are indicated in Figure 6.4. F_α^β represent the components of the associated force vector, and the link stiffness matrix $\mathbf{K}_L^{(e)}$ is given by:

$$\mathbf{K}_L^{(e)} = k_L^{(e)} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} = k_L^{(e)} \begin{bmatrix} \mathbf{I}_2 & -\mathbf{I}_2 \\ -\mathbf{I}_2 & \mathbf{I}_2 \end{bmatrix}, \tag{6.27}$$

where $\mathbf{I}_2$ represents the 2-by-2 unity matrix, and $k_L^{(e)}$ is the link stiffness associated to the particular e^{th} element this link belongs to. Subsequently, the governing equation of the patch of the e^{th} element and its associated links is given by:

$$\begin{bmatrix} k_L^{(e)}\mathbf{I}_8 & -k_L^{(e)}\mathbf{I}_8 \\ -k_L^{(e)}\mathbf{I}_8 & k_L^{(e)}\mathbf{I}_8 + \mathbf{K}_E^{(e)} \end{bmatrix} \cdot \begin{Bmatrix} \boldsymbol{u}_{\text{OUTER}}^{(e)} \\ \boldsymbol{u}_{\text{INNER}}^{(e)} \end{Bmatrix} = \begin{Bmatrix} {}^{\text{ext}}\boldsymbol{F}_{\text{OUTER}}^{(e)} \\ \mathbf{0} \end{Bmatrix}, \tag{6.28}$$

where $\mathbf{K}_E^{(e)}$ is the stiffness matrix of the internal element, and ${}^{\text{ext}}\boldsymbol{F}_{\text{OUTER}}^{(e)}$ represents the force applied to the outer nodes of this patch. The displacement components of the patch are partitioned in components associated with the outer and inner nodes. In the following, the dimension subscript of the unity matrix $\mathbf{I}$ will be omitted. Equation 6.28 is valid in the linear setting. To analyze the geometrically and materially nonlinear SMA actuators, the Total Lagrangian (TL) formulation is used. A full description of the TL formulation can be found in reference works (e.g. Bathe, 1995). The final system equations are given by:

$$\boldsymbol{R}(\boldsymbol{U}) = {}^{\text{ext}}\boldsymbol{F} - {}^{\text{int}}\boldsymbol{F}(\boldsymbol{U}) = \mathbf{0}, \tag{6.29}$$

where $\boldsymbol{R}$ is the residual force vector, given by the difference between the external and internal force vectors. Cases with displacement-dependent external forces are not considered in this chapter. Newton iterations are used to solve this nonlinear system of equations, resulting in:

$$\mathbf{K}(\boldsymbol{U}^k)\Delta\boldsymbol{U}^{k+1} = \boldsymbol{R}(\boldsymbol{U}^k), \qquad \boldsymbol{U}^{k+1} = \boldsymbol{U}^k + \Delta\boldsymbol{U}^{k+1}, \tag{6.30}$$

where the tangent stiffness matrix $\mathbf{K}$ is given by the negative Jacobian $-\partial\boldsymbol{R}/\partial\boldsymbol{U}$, and the superscript denotes the iteration number. The contribution of an individual patch to this system is given by:

$$\begin{bmatrix} k_L^{(e)}\mathbf{I} & -k_L^{(e)}\mathbf{I} \\ -k_L^{(e)}\mathbf{I} & k_L^{(e)}\mathbf{I} + \mathbf{K}_E^{(e)}(\boldsymbol{u}_{\text{INNER}}^{(e)}) \end{bmatrix} \cdot \begin{Bmatrix} \Delta\boldsymbol{u}_{\text{OUTER}}^{(e)} \\ \Delta\boldsymbol{u}_{\text{INNER}}^{(e)} \end{Bmatrix} = \dots \\ \begin{Bmatrix} {}^{\text{ext}}\boldsymbol{F}_{\text{OUTER}}^{(e)} - {}^{\text{int}}\boldsymbol{F}_{\text{OUTER}}^{(e)} \\ \mathbf{0} - {}^{\text{int}}\boldsymbol{F}_{\text{INNER}}^{(e)} \end{Bmatrix}. \tag{6.31}$$

The right-hand side vector contains the residual contributions of the outer and inner nodes, where:

$${}^{\text{int}}\boldsymbol{F}_{\text{OUTER}}^{(e)} = k_L^{(e)}\left(\boldsymbol{u}_{\text{OUTER}}^{(e)} - \boldsymbol{u}_{\text{INNER}}^{(e)}\right), \tag{6.32}$$

$${}^{\text{int}}\boldsymbol{F}_{\text{INNER}}^{(e)} = {}^{\text{int}}\boldsymbol{F}_{E}^{(e)} - k_L^{(e)}\left(\boldsymbol{u}_{\text{OUTER}}^{(e)} - \boldsymbol{u}_{\text{INNER}}^{(e)}\right), \tag{6.33}$$

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

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

$$\mathbf{K}_C^{(e)} = k_L^{(e)}\mathbf{I} - \left(k_L^{(e)}\right)^2\left[k_L^{(e)}\mathbf{I} + \mathbf{K}_E^{(e)}\right]^{-1}, \tag{6.34}$$

$$\boldsymbol{R}_C^{(e)} = {}^{\text{ext}}\boldsymbol{F}_{\text{OUTER}}^{(e)} - {}^{\text{int}}\boldsymbol{F}_{\text{OUTER}}^{(e)} - k_L^{(e)}\left[k_L^{(e)}\mathbf{I} + \mathbf{K}_E^{(e)}\right]^{-1}{}^{\text{int}}\boldsymbol{F}_{\text{INNER}}^{(e)}. \tag{6.35}$$

After solving the condensed system, the inner displacement increments are computed for each patch from the outer increments by:

$$\Delta\boldsymbol{u}_{\text{INNER}}^{(e)} = \left[k_L^{(e)}\mathbf{I} + \mathbf{K}_E^{(e)}\right]^{-1}\left\{k_L^{(e)}\Delta\boldsymbol{u}_{\text{OUTER}}^{(e)} - {}^{\text{int}}\boldsymbol{F}_{\text{INNER}}^{(e)}\right\}. \tag{6.36}$$

The condensed system has exactly the same number of degrees of freedom as the original mesh before additional nodes and zero-length links were added. In the implementation of the ECP method, 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.

6.3.4 Sensitivity analysis

In the topology optimization problems considered here, the response of interest f can generally be written as:

$$f = \boldsymbol{L}^T \boldsymbol{U}. \tag{6.37}$$

Here $\boldsymbol{L}$ is a constant vector that is used to select the displacement components of interest, and $\boldsymbol{U}$ is the total displacement vector. The distinction between inner and outer displacements, and the effect of condensation will be considered later. The sensitivity of this response f is determined by the adjoint method. Adjoint sensitivity analysis has large computational advantages over direct sensitivity analysis, when dealing with large numbers of design variables and relatively few response quantities, as in topology optimization. The derivation involves including the effect of the state equations given in Equation 6.29 by the formulation of the following Lagrangian:

$$\mathcal{L} = \boldsymbol{L}^T \boldsymbol{U} + \boldsymbol{\lambda}^T \boldsymbol{R}(\boldsymbol{U}(\boldsymbol{\gamma}), \boldsymbol{\gamma}). \tag{6.38}$$

In the equilibrium configuration, $\mathcal{L}$ is equivalent to the response f, since in that case $\boldsymbol{R}$ equals zero. The design variables in the ECP method, given by $\boldsymbol{\gamma}$, determine the link stiffness $k_L^{(e)}$ of the links in each patch. Since different expressions can be used to relate the link stiffnesses to the design variables, the sensitivity of the response to the link stiffness is considered first. Differentiation of $\mathcal{L}$ with respect to $k_L^{(e)}$ yields:

$$\frac{\partial \mathcal{L}}{\partial k_L^{(e)}} = \boldsymbol{L}^T \frac{\partial \boldsymbol{U}}{\partial k_L^{(e)}} + \boldsymbol{\lambda}^T \left(\frac{\partial \boldsymbol{R}}{\partial k_L^{(e)}} + \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{U}} \frac{\partial \boldsymbol{U}}{\partial k_L^{(e)}} \right). \tag{6.39}$$

This equation holds for any linear or nonlinear *history-independent* model, such as the presented SMA model. Rearranging yields:

$$\frac{\partial \mathcal{L}}{\partial k_L^{(e)}} = \left(\boldsymbol{L}^T - \boldsymbol{\lambda}^T \mathbf{K} \right) \frac{\partial \boldsymbol{U}}{\partial k_L^{(e)}} + \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{R}}{\partial k_L^{(e)}}, \tag{6.40}$$

where the definition of the system tangent stiffness matrix $\mathbf{K}$ introduced in Equation 6.30 has been used. In the adjoint sensitivity analysis technique, the vector of multipliers $\boldsymbol{\lambda}$ is now chosen such that the term with the sensitivities of the displacements $\partial \boldsymbol{U} / \partial k_L^{(e)}$ in Equation 6.40 vanishes:

$$\boldsymbol{L}^T - \boldsymbol{\lambda}^T \mathbf{K} = \mathbf{0} \;\Rightarrow\; \mathbf{K}\boldsymbol{\lambda} = \boldsymbol{L}. \tag{6.41}$$

The symmetry of $\mathbf{K}$ has been used here. When using a direct solver, $\mathbf{K}$ is available in decomposed form after the analysis, and the multipliers $\boldsymbol{\lambda}$ can be obtained at a fraction of the computational effort required for the nonlinear analysis. Sensitivities are obtained by subsequent substitution of $\boldsymbol{\lambda}$ into Equation 6.40:

$$\frac{\partial \mathcal{L}}{\partial k_L^{(e)}} = \boldsymbol{\lambda}^T \frac{\partial \boldsymbol{R}}{\partial k_L^{(e)}}. \tag{6.42}$$

In the implementation, condensation is applied to the adjoint sensitivity equations analogously as to the equilibrium equations, in order to re-use the decomposed system matrix based on the outer displacements, denoted by $\mathbf{K}_{\text{OUTER}}$. This gives:

$$\mathbf{K}_{\text{OUTER}}\boldsymbol{\lambda}_{\text{OUTER}} = \boldsymbol{L}_{\text{OUTER}}. \tag{6.43}$$

Given the outer multipliers, the inner multipliers can subsequently be found by applying the equivalent of Equation 6.36 to each patch:

$$\boldsymbol{\lambda}^{(e)}_{\text{INNER}} = \left[k^{(e)}_L\mathbf{I} + \mathbf{K}^{(e)}_E\right]^{-1}\left\{k^{(e)}_L\boldsymbol{\lambda}^{(e)}_{\text{OUTER}} - \boldsymbol{L}^{(e)}_{\text{INNER}}\right\}. \tag{6.44}$$

From Equation 6.31, it can be derived that the contribution of a particular patch to the partial derivatives of the residual used in Equation 6.42 is given by:

$$\frac{\partial \boldsymbol{R}^{(e)}}{\partial k^{(e)}_L} = \begin{Bmatrix} -\left(\boldsymbol{u}^{(e)}_{\text{OUTER}} - \boldsymbol{u}^{(e)}_{\text{INNER}}\right) \\ \left(\boldsymbol{u}^{(e)}_{\text{OUTER}} - \boldsymbol{u}^{(e)}_{\text{INNER}}\right) \end{Bmatrix}. \tag{6.45}$$

Combining Equation 6.45 and Equation 6.42 subsequently yields:

$$\frac{\partial \mathcal{L}}{\partial k^{(e)}_L} = \left(\boldsymbol{\lambda}^{(e)}_{\text{INNER}} - \boldsymbol{\lambda}^{(e)}_{\text{OUTER}}\right)\left(\boldsymbol{u}^{(e)}_{\text{OUTER}} - \boldsymbol{u}^{(e)}_{\text{INNER}}\right) = \frac{\partial f}{\partial k^{(e)}_L}. \tag{6.46}$$

Finally, for a given relation between design variables $\boldsymbol{\gamma}$ and link stiffnesses $\boldsymbol{k}_L$, the sensitivity with respect to the design variables can be evaluated by applying the chain rule:

$$\frac{\partial f}{\partial \gamma_i} = \frac{\partial f}{\partial \boldsymbol{k}_L}^T \frac{\partial \boldsymbol{k}_L}{\partial \gamma_i}. \tag{6.47}$$

In the implementation, the sparsity of $\partial k^{(e)}_L/\partial\gamma_i$ can be exploited. The presented sensitivity analysis has been verified by finite differences. Note that nowhere in the computation of the adjoint sensitivities, differentiation of the material model is required. This is a clear advantage of the ECP approach.

6.3.5 Interpolation function

In order to understand the effect of the link stiffness on the stiffness of the patch as a whole, it is insightful to rewrite the condensed stiffness matrix given in Equation 6.34 using the Woodbury formula (see e.g. Golub and Loan, 1996). In its most general form, the Woodbury formula states that:

$$\left[\mathbf{A} + \mathbf{U}\mathbf{V}^T\right]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left[\mathbf{I} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}\right]^{-1}\mathbf{V}^T\mathbf{A}^{-1}. \tag{6.48}$$

Using $\mathbf{U} = \mathbf{K}^{(e)}_E$, $\mathbf{V} = \mathbf{I}$ and $\mathbf{A} = k^{(e)}_L\mathbf{I}$, it follows that:

$$\left[k^{(e)}_L\mathbf{I} + \mathbf{K}^{(e)}_E\right]^{-1} = \frac{1}{k^{(e)}_L}\mathbf{I} - \left(\frac{1}{k^{(e)}_L}\right)^2\mathbf{K}^{(e)}_E\left[\mathbf{I} + \frac{1}{k^{(e)}_L}\mathbf{K}^{(e)}_E\right]^{-1}. \tag{6.49}$$

Substitution in Equation 6.34 yields, while dropping the $^{(e)}$ superscripts for readability:

$$\mathbf{K}_C = \mathbf{K}_E \left[\mathbf{I} + \frac{1}{k_L} \mathbf{K}_E \right]^{-1}. \tag{6.50}$$

From this equation, it is clearly seen how the link stiffness k_L influences the total condensed stiffness of the patch, $\mathbf{K}_C$: when the link stiffness approaches zero, $\mathbf{K}_C$ will tend toward a very compliant stiffness matrix ($k_L\mathbf{I}$). On the other hand, when k_L becomes large, the condensed stiffness matrix tends toward the original element stiffness $\mathbf{K}_E$. Equation 6.50 can be reformulated into:

$$\mathbf{K}_C^{-1} = \mathbf{K}_E^{-1} \left[\mathbf{I} + \frac{1}{k_L} \mathbf{K}_E \right] = \mathbf{K}_E^{-1} + [k_L \mathbf{I}]^{-1}. \tag{6.51}$$

This last relation shows that the condensed stiffness of the patch essentially is the result of Reuss (series) homogenization of the original element and a material with a stiffness matrix $k_L\mathbf{I}$. A more elaborate investigation of the relation between homogenization techniques and the present ECP approach is presented by Yoon *et al.* (2006b). To determine the way the link stiffness affects the condensed stiffness of the patch, a matrix norm can be used to turn this relation into a scalar expression. In this case, the largest eigenvalue norm is used, but similar results are obtained using other matrix norms (Yoon *et al.*, 2006b). It follows that:

$$\frac{1}{||\mathbf{K}_C||} = \frac{1}{||\mathbf{K}_E||} + \frac{1}{k_L} \;\Rightarrow\; \frac{||\mathbf{K}_C||}{||\mathbf{K}_E||} = \frac{1}{1 + \frac{||\mathbf{K}_E||}{k_L}}. \tag{6.52}$$

The quantity $||\mathbf{K}_C||/||\mathbf{K}_E||$ expresses the relative stiffness of the (condensed) patch with respect to the stiffness of the original element, in terms of the chosen matrix norm. It is referred to as the *stiffness ratio* in this chapter, and Figure 6.5(a) depicts the way this quantity varies with the normalized link stiffness. Figure 6.5(a) appears to show a smooth transition from a very low stiffness to a stiffness ratio equal to 1, when the link stiffness varies from low to high. At the point where the link stiffness equals the norm of the element stiffness matrix $||\mathbf{K}_E||$, the stiffness ratio equals 0.5. Note, however, that the normalized link stiffness is shown on a logarithmic scale. On a linear scale, the transition from low to high stiffness is very abrupt and nearly step-wise.

If the link stiffness would directly be used as a design variable in the topology optimization problem, the strong non-smoothness would make convergence of the process very difficult. Clearly, a more gradual relation between the stiffness ratio and the design variable is desirable. This can be accomplished by introducing an interpolation function based on the inverse of the stiffness ratio/link stiffness relation of Equation 6.52, and use that to define a relation between design variable γ and link stiffness k_L:

$$k_L = \frac{||\mathbf{K}_E||}{\frac{1}{\gamma} - 1} \qquad 0 \leq \gamma \leq 1. \tag{6.53}$$

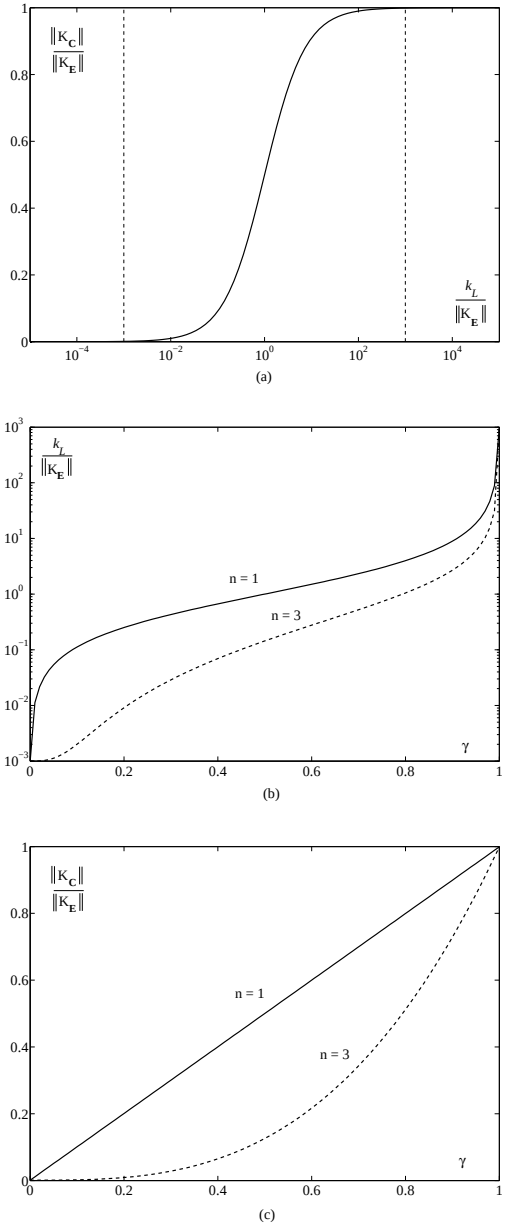


Figure 6.5: *Graphs showing the stiffness ratio versus the normalized link stiffness (a), the normalized link stiffness interpolation function (b), and the stiffness ratio versus design variable γ (c). Dashed lines in (b) and (c) illustrate the effect of penalization with different exponents n.*

Ideally, to disconnect elements, the link stiffness should be zero, and to fully connect elements, the link stiffness should be infinite, as is the case with the relation given by Equation 6.53. In practice, however, this approach would result in various numerical problems, such as divisions by zero and ill-conditioning or non-positive definiteness of the system matrix. For this reason, the link stiffness relation needs to be modified and a suitable lower and upper limit for k_L has to be determined. Given the monotonic influence of the link stiffness on the stiffness ratio shown in Figure 6.5(a), suitable limits can be found by allowing deviation by a certain tolerance ε from the ideal situation:

$$\frac{||\mathbf{K}_C||}{||\mathbf{K}_E||}\left(k_L^{(min)}\right) = \varepsilon, \qquad \frac{||\mathbf{K}_C||}{||\mathbf{K}_E||}\left(k_L^{(max)}\right) = 1 - \varepsilon. \tag{6.54}$$

Using Equation 6.52, it follows that:

$$k_L^{(min)} = \frac{||\mathbf{K}_E||}{\frac{1}{\varepsilon} - 1}, \qquad k_L^{(max)} = \frac{||\mathbf{K}_E||}{\frac{1}{1-\varepsilon} - 1}. \tag{6.55}$$

In this chapter, $\varepsilon = 0.001$ has been used. This results in normalized link stiffnesses in the range of 10^{-3} to 10^{3}, as indicated in Figure 6.5(a). Inspired by Equation 6.55, the design variable/link stiffness interpolation function proposed in Equation 6.53 can be altered as:

$$k_L = \frac{||\mathbf{K}_E||}{\frac{1}{\varepsilon+\gamma(1-2\varepsilon)} - 1} \qquad 0 \leq \gamma \leq 1. \tag{6.56}$$

The advantage of this formulation is that the domain of the design variable γ is still exactly $[0, 1]$. This relation is shown in Figure 6.5(b), and the resulting relation between the stiffness ratio and the design variable γ is depicted in Figure 6.5(c). Clearly a linear interpolation is achieved. Subsequently, penalization can be added to this interpolation, which helps designs to converge toward black-white designs by making intermediate design variable values less efficient. In this chapter, an exponential penalization function is used, which finally results in the following interpolation function:

$$k_L = \frac{||\mathbf{K}_E||}{\frac{1}{\varepsilon+\gamma^n(1-2\varepsilon)} - 1} \qquad 0 \leq \gamma \leq 1. \tag{6.57}$$

The element stiffness matrix norm $||\mathbf{K}_E||$ is evaluated once at the start of the optimization process. In the case studies presented in this chapter, using a constant penalization exponent $n = 3$ or $n = 4$ turned out to improve the convergence of the optimization process. Validation of the outlined ECP topology optimization approach for various two- and three-dimensional linear and geometrically nonlinear compliance minimization problems is presented elsewhere (Yoon *et al.*, 2006a,b).

6.4 Problem formulation and regularization

6.4.1 Problems considered

The aim of this research is to develop a topology optimization procedure for the design of SMA actuators. In this case, SMA structures are considered that can generate a displacement under a constant load. The actuation is triggered by a change in the temperature of the structure. This study is limited to two-dimensional plane stress structures, that are subjected to uniform changes in temperature. The geometry, applied load and boundary conditions of the two types of design problems considered in this chapter are depicted in Figure 6.6. Both problems involve a constant load F acting on the SMA structure from the top of the design domain, and supports are defined at either the whole bottom or only part of it. The structures are of unit thickness, and symmetry is enforced. In the analysis, only the left half of the structures is considered. Unless mentioned otherwise, the topology optimization process is started with a uniform 50% density distribution. The optimization is performed using the Method of Moving Asymptotes by Svanberg (1987), which is often claimed to be the most effective optimizer for topology optimization problems.

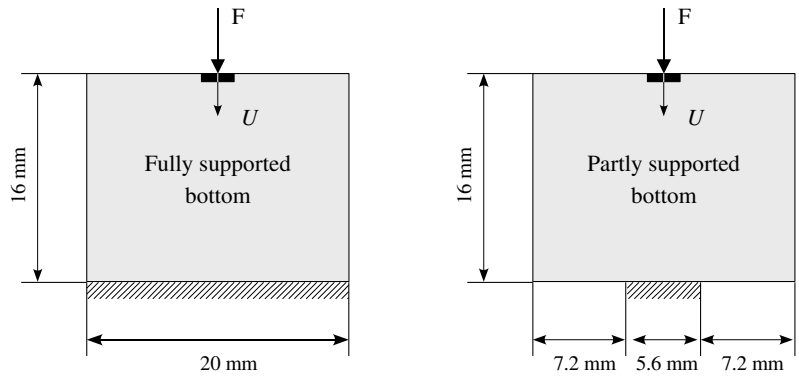


Figure 6.6: *Geometry, applied load and boundary conditions of problems considered for SMA topology optimization.*

The objective of the design problem is to maximize the stroke of the actuator. The stroke is evaluated by taking the difference in displacement of the output point at the lowest and highest temperature:

$$\text{Objective:} \quad \max_{\boldsymbol{\gamma}} |U_{T=T_A} - U_{T=T_B}|. \tag{6.58}$$

In all cases studied here, $T_A = 328K$ and $T_B = 343K$ has been used. In the

presented examples, the point where the load is applied is selected as the output point. In general, other choices are possible. In many topology optimization problems, a mass constraint is commonly added to the problem formulation. However, it has turned out that for this SMA actuator design problem, the mass constraint is not necessary. Unlike in e.g. compliance minimization, using more mass does not by definition result in a better objective value. However, using no material at all obviously also does not result in an effective actuator. Somewhere inbetween those two extremes, an optimal amount of mass exists, that can be found by the optimization process. It is not necessary to constrain the mass using an explicit constraint, in order to create a meaningful and well-behaving optimization problem.

6.4.2 Numerical artifacts in topology optimization

It was found that in case of SMA actuator design, possibly because of the objective to increase the displacement difference, the optimization tends to exploit numerical artifacts such as one-node hinges and associated mechanism deformation modes. A clear case of such a mechanism mode is shown in Figure 6.7, which was encountered during the topology optimization process. From an optimization point of view, for the given discretized model, exploiting mechanism modes indeed can lead to very good objective values in certain cases. However, in the actual design problem, hinges are usually not desired, as particularly for micro-actuators they simply cannot be manufactured.

Another well-known artifact often encountered in topology optimization is the

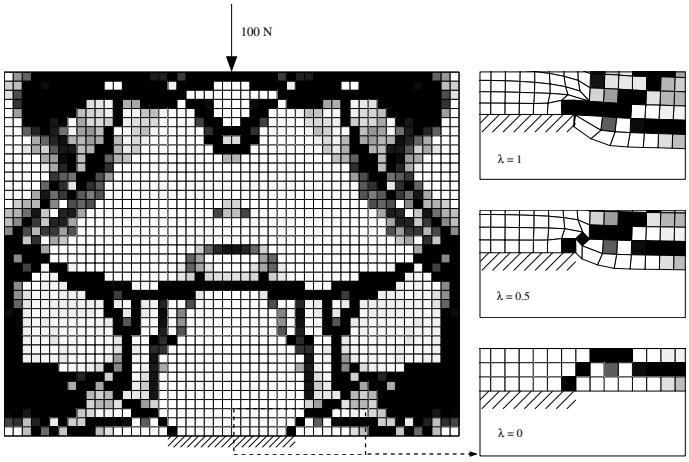


Figure 6.7: *Mechanism deformation mode occurring during the design process. The details on the right show the deformation at a sequence of load factors* λ.

so-called checkerboard pattern (see e.g. Bendsoe and Sigmund, 2003). This is an alternating pattern of elements with maximum and minimum density, that in the finite element discretization results in a favorable stiffness-to-weight ratio. Checkerboard pattern problems were encountered less frequently in trials than mechanism modes, possibly because in contrast to the classical compliance minimization problem, the stiffness-to-weight ratio is not an important factor in the design process. Regularization of the topology optimization problem is required to prevent the optimizer to pursue and find unphysical and meaningless solutions.

6.4.3 Effect of filtering

To regularize the topology optimization process and to avoid these problems, filtering is often applied. This is a heuristic technique that consists of replacing the sensitivity of each element $\partial f/\partial \rho_i$ by a weighted spatial average of the sensitivities in its neighborhood. Filtering has been introduced by Sigmund (1997), and usually a weighting function H_i is used that decays linearly with the distance to the center of the element:

$$\frac{\tilde{\partial f}}{\partial \gamma_p} = \frac{1}{\gamma_p} \frac{\sum_{q=1}^{N} H_q \gamma_q \frac{\partial f}{\partial \gamma_q}}{\sum_{q=1}^{N} H_q}, \tag{6.59}$$

with

$$H_i = \begin{cases} R - \mathrm{dist}(p,q) & \mathrm{dist}(p,q) \leq R, \\ 0 & \mathrm{dist}(p,q) > R. \end{cases} \tag{6.60}$$

Here $\mathrm{dist}(p,q)$ is an operator that produces the distance between the center of elements p and q, and R represents the radius of the domain of influence of the filter. However, the addition of a filtering step in the optimization procedure basically means that the design description and the sensitivities are no longer consistent. This can result in rather erratic changes in the objective, systematic worsening or even a total failure. For the SMA topology optimization, these effects were frequently encountered, as shown in Figure 6.8 and Figure 6.9. Moreover, the Sigmund filtering turned out to be unable to prevent the one-node connections and mechanism modes. A reason why filtering appears to work well for e.g. compliance minimization problems, but fails to regularize the SMA actuator topology optimization problem, could be that in compliance minimization all sensitivities have the same sign. Adding material reduces the compliance, or at least does not increase it. In the presently studied SMA actuator design problems however, the objective is to maximize the stroke of the actuator. It was observed that both positive and negative sensitivities of this objective can occur. We speculate that this fact might be related to the reason why the heuristic filtering approach is not effective in this class of problems.

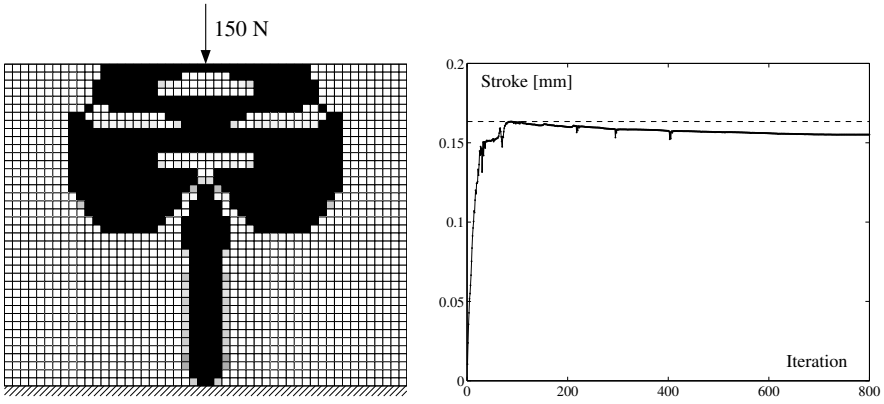


Figure 6.8: *Obtained design (left) and evolution of the objective function during the optimization process (right), for a case using a sensitivity filter technique. Note the fact that the stroke actually* decreases *as the optimization progresses.*

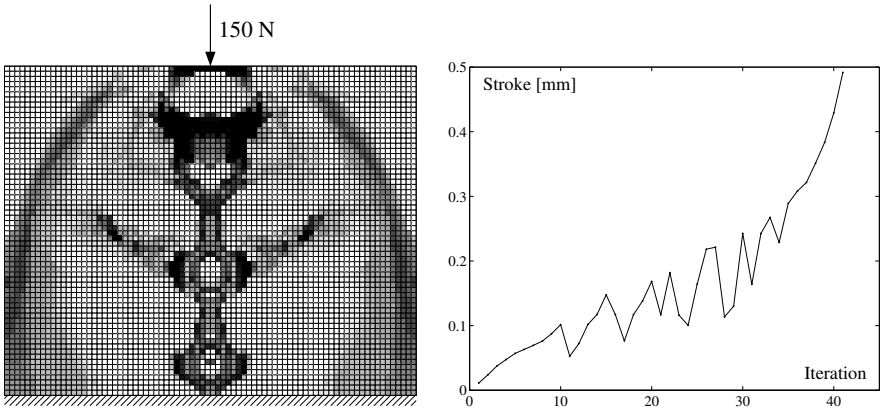


Figure 6.9: *Obtained design (left) and rather erratic evolution of the objective function during the optimization process (right), for a case using a sensitivity filter technique. In this case, the design gradually dissolved and the analysis failed.*

6.4.4 Nodal design variables

The use of nodal design variables was found to be a simple and effective method to regularize the SMA actuator topology optimization problem. Unlike the conventional element-based approach, where design variables are assigned to elements, the design variables are assigned to the nodes in the finite element mesh, as illustrated in Figure 6.10. The design variable values for each element are then determined by the average of the nodal design variables of its nodes:

$$\gamma_i = \frac{1}{n} \sum_{j=1}^{n} \eta_j. \tag{6.61}$$

Here γ_i represents the density associated with element i, n equals the number of nodes of the element and η_j are the design variables associated with these nodes.

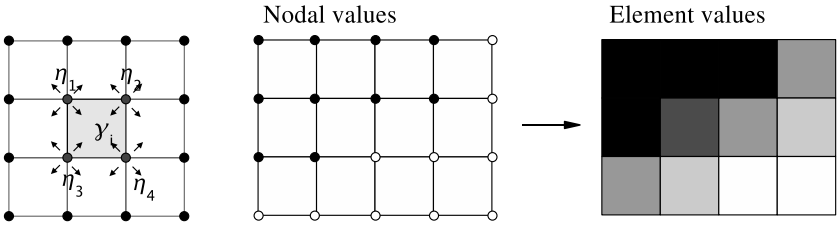


Figure 6.10: *Nodal design variables and corresponding element values for a given layout of nodal design variables.*

In this way, the design space is restricted, and undesirable numerical artifact patterns, such as checkerboards and one-node hinges, can no longer be represented in this formulation. Thus, they are banned from the set of possible solutions of the optimization problem, and physically more meaningful solutions can be generated. For this reason, nodal variable approaches have recently gained popularity (Bendsoe and Sigmund, 2003, Guest *et al.*, 2004, Matsui and Terada, 2004, see e.g.). Several alternative regularization methods have been proposed (see e.g. Poulsen, 2003, Sigmund and Petersson, 1998, Yoon *et al.*, 2004, among others), but in general these methods tend to be rather involved. In contrast, the nodal variable approach is simple and straightforward to implement. And unlike in the case of filtering approaches, the design description and the sensitivities are formulated in a fully consistent way:

$$\frac{\partial f}{\partial \eta_j} = \sum_{k=1}^{m} \frac{\partial f}{\partial \gamma_k} \frac{\partial \gamma_k}{\partial \eta_j} = \frac{1}{n} \sum_{k=1}^{m} \frac{\partial f}{\partial \gamma_k}. \tag{6.62}$$

Here m is the number of elements node j is associated with, here conceptually represented by the set $k = 1 \ldots m$.

The continuity of the density field, that is inherent to the nodal variable approach, results in the advantage that it prohibits undesired density patterns such as checkerboards and one-node connections. On the other hand, a disadvantage of this approach is that it unavoidably implies that edges of the design will consist of elements with intermediate density values, as shown for an example layout in Figure 6.10. This makes that optimization results appear to be slightly blurred. However, by using a suitable interpolation function, the effect of the elements with intermediate densities on the response of the structure can be controlled. Experience has shown that adequate penalization, such that elements with intermediate density have a less-than-proportional stiffness, generally improves convergence and leads to sufficiently clear results.

6.5 SMA topology optimization results

As this is the first research where topology optimization is applied to SMA thermal actuator design, several studies have been conducted to investigate the nature and characteristics of this class of problems. In this section, the results of the following studies are reported:

- Effect of the load magnitude
- Effect of mesh refinement
- Effect of the load direction
- Effect of the initial design

Although this limited study does not intend to give a full characterization of the SMA topology optimization problem, several interesting aspects are revealed by these investigations, that illustrate the complex nature of this strongly nonlinear problem. All results are obtained using the nodal design variable formulation presented in the previous section, and for the problem formulations given in Section 6.4.1.

6.5.1 Effect of load magnitude

Unlike in linear problems, the SMA topology optimization problem shows a clear effect of the load magnitude on the resulting design. Both the fully and partly supported design problems, introduced before in Figure 6.6, have been evaluated for loads of 50, 75, 100 and 150 N. Results for the fully supported bottom case are shown in Figure 6.11, and for the partly supported case in Figure 6.12. Note that the designs obtained in the various cases are considerably different, clearly showing the influence of the load magnitude on the optimization process.

There appears to be a tendency to use more material in case of larger loads. This can be understood from the fact that the SMA material is most effective

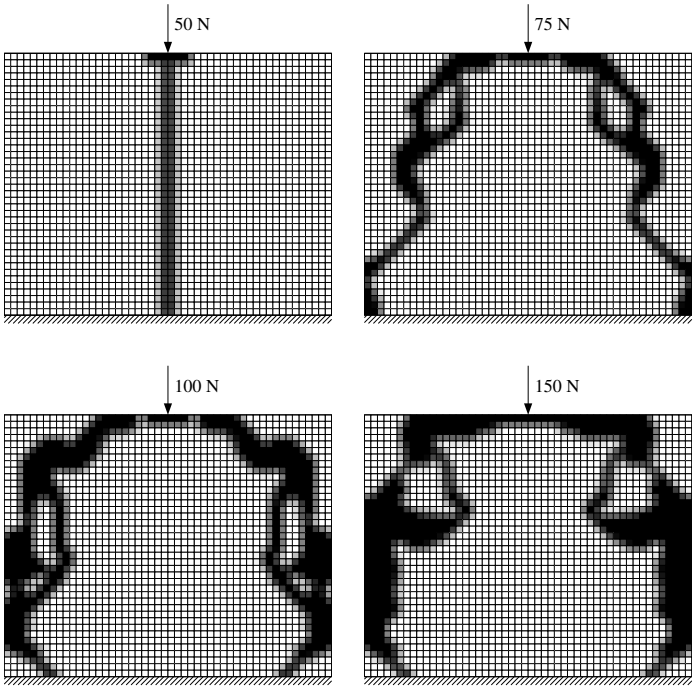


Figure 6.11: *Topologies found by topology optimization in the fully supported case, for load magnitudes of 50, 75, 100 and 150 N.*

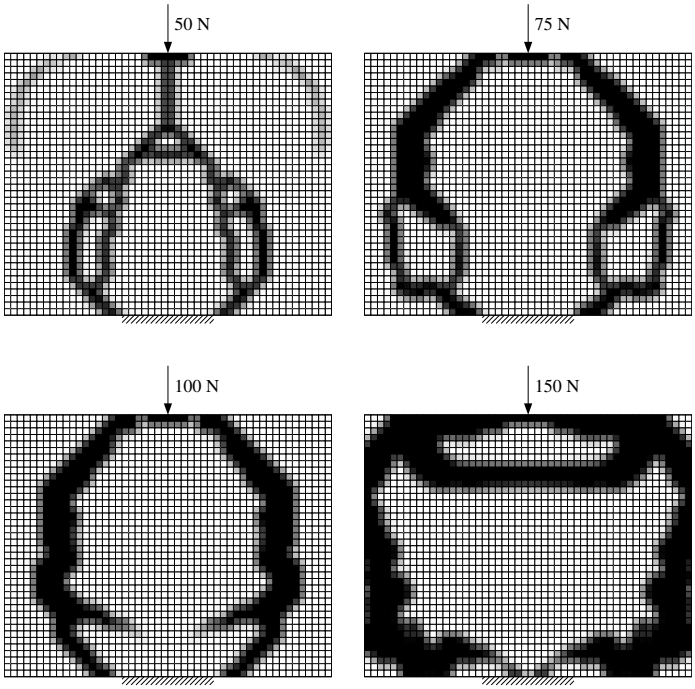


Figure 6.12: *Topologies found by topology optimization in the partly supported case, for load magnitudes of 50, 75, 100 and 150 N.*

when in the low temperature situation, it is in the transformed, R-phase state while in the high temperature situation, it should be preferably in the austenite phase. That situation leads to the largest difference in mechanical properties of the structure, which will result in the largest stroke. Inspection shows that in optimized structures, the strain distribution in large parts of the structures indeed is as described here. This is shown in Figure 6.13, where the volume fraction of the R-phase is shown together with the deformed structures, in both temperature cases for the final design in the 75 N fully supported case. The volume fraction of the R-phase is used here as a comparison measure, and is defined as:

$$\phi_R = \begin{cases} \varepsilon_e \leq \varepsilon_{e(1)} & 0 \\ \varepsilon_{e(1)} < \varepsilon_e \leq \varepsilon_{e(2)} & (\varepsilon_e - \varepsilon_{e(1)})/(\varepsilon_{e(2)} - \varepsilon_{e(1)}) \\ \varepsilon_e > \varepsilon_2 & 1 \end{cases} \tag{6.63}$$

The transition strains $\varepsilon_{e(1)}$ and $\varepsilon_{e(2)}$ are the strain levels at which the piecewise linear stress-strain relation switches from one branch to another, as derived in Appendix B. For the definition of this comparative phase fraction measure ϕ_R, it is assumed that a linear relation exists between the effective strain and the R-phase volume fraction, which is not necessarily correct. However, it is a useful measure to compare to what extent the material is in a state where the transformation from austenite to R-phase is ongoing or has completed, or in a state where it is still about to begin. As can be seen in Figure 6.13, in the high temperature situation hardly any material is in the transformed state, whereas in the low temperature situation practically the complete structure has transformed into the R-phase.

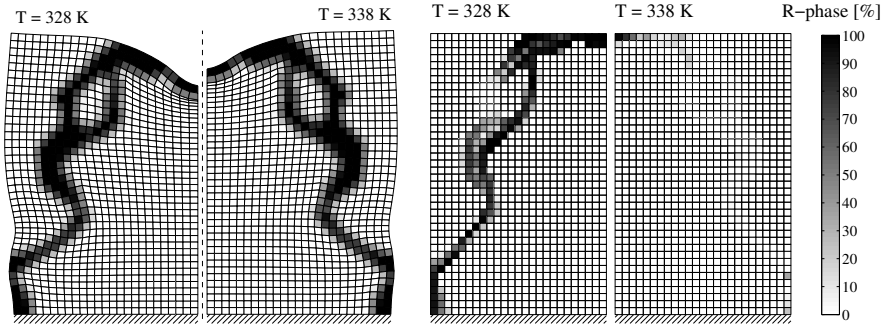


Figure 6.13: *Deformed structures in the low and high temperature situation for the fully supported 75 N case (left), and volume fractions of R-phase in both situations (right). Displacements have been scaled by a factor 5.*

The physical explanation for the designs that are obtained by the topology optimization process is that structures are generated for which the effective strains are such, that in the low temperature phase nearly all material transforms into

the R-phase, whereas in the high temperature phase nearly all material remains in the austenite phase. This apparently results in the most effective structure. This interpretation also explains why generally with higher load levels designs with thicker structural members are obtained: in order to keep the strain in this "most effective range", the structure has to become stiffer when the load is increased. Next to that, it is observed that often winding or meandering structures are generated, and configurations are obtained that show a considerable amount of bending deformation. Strains in the right magnitude help to increase the difference between the low and high temperature configuration, whereas bending deformations in many cases can lead to larger displacements at the point of interest. Note that particularly the winding structures, such as, for example, those seen in the 150 N partly supported case, are rather uncommon in topology optimization results.

In the partly supported 50 N case (Figure 6.12), some excess material remains on the top left and right side, that is not connected to the final structure. This is possible because no mass constraint is used, so there is no motivation for the optimizer to remove parts that do not contribute to the objective. Because the void elements are not exactly void numerically, but have some finite stiffness, generally material that is not actively contributing but obstructing the actuator by adding stiffness, is removed during the optimization. In this 50 N case, apparently, the disconnected parts remaining from earlier design configurations in the optimization process do not adversely contribute to the resulting stroke. Likewise, in the 100 N partly supported case, two inward-pointing "spikes" remain at the bottom of the structure, apparently because of the same reason.

A final aspect that deserves attention is a problem that occurred in the 150 N fully supported case, but a tendency towards this problem was also observed in other cases. The original result for this case is depicted in Figure 6.14, which shows a structure in which the part where the load is applied and the rest of the structure (the "arms") are nearly disconnected. No one-node hinges are formed, since that is impossible in the nodal design variable setting. However, effectively still hinge-like structures are created, by areas of very low stiffness, as is clearly shown in the deformed configuration in Figure 6.14. This is an undesired effect, because the aim is to obtain a clear topology with distinct solid and void regions, that actually is manufacturable. Possibly, the very compliant regions acting as hinges could be realized by using a secondary material or by using porous microstructures, but this is not the objective of the present design study. In this case, fortunately, the tendency of this problem to form such hinges could be suppressed by slightly extending the non-design domain. The result obtained with that modification is the one shown in Figure 6.12. Strikingly, the performance of the modified case actually turned out to be better than that of the structure with the hinges (a stroke of 0.3758 versus 0.3401 mm).

As a comparative study, it is interesting to investigate the performance of all obtained designs at various load levels for *different* load magnitudes than the design load. For this purpose, scans have been made at load levels ranging from 0 to 200 N, with increments of 25 N, and the strokes of all actuators under

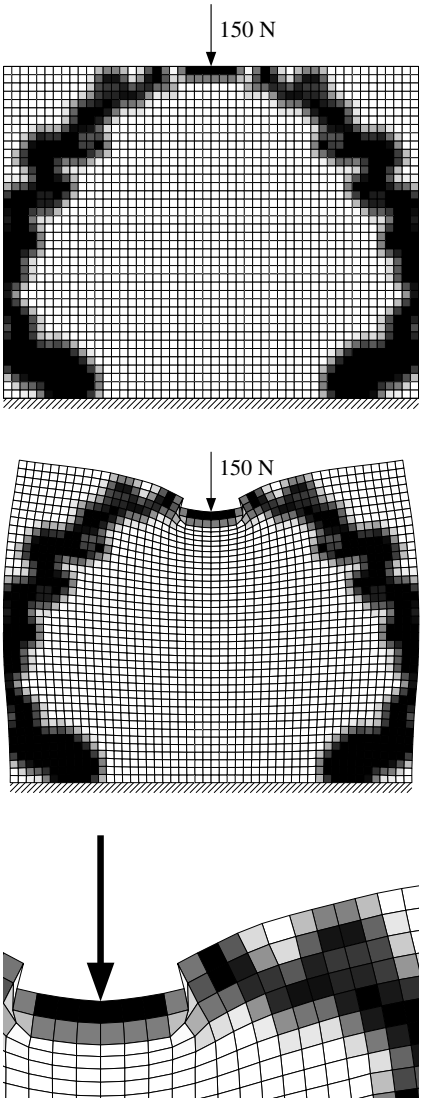


Figure 6.14: *Optimal design obtained in the 150 N fully supported case with the original non-design domain. Note the fact that the structure is nearly disconnected.*

these loads have been plotted in Figure 6.15. It can be seen that for both the fully and partly supported case, the 50 N designs are clearly local optima, since all structures designed for larger loads outperform these designs. This could be related to the strongly nonlinear nature of the material model, which possibly at lower load levels results in a problem that has different characteristics than at higher load levels. For example, the fact whether or not the phase transformation is induced by the loading condition significantly affects the local apparent stiffness. This can definitely influence the sensitivities, and, accordingly, affect the direction the structural optimization process takes in the design space.

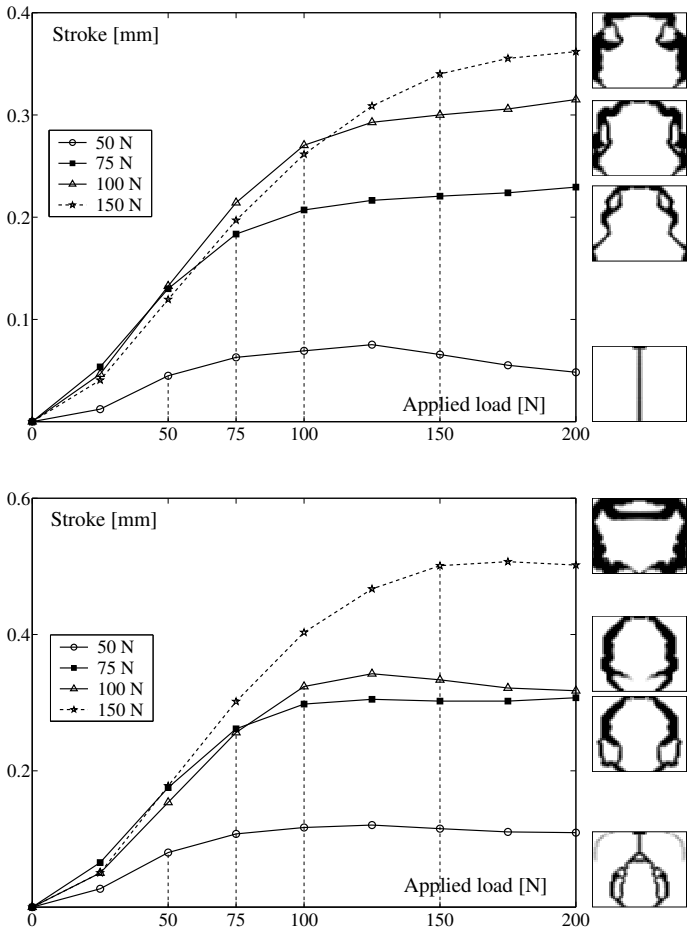


Figure 6.15: *Actuator stroke as a function of the applied load, for the designs found by topology optimization in case of a fully supported (top) and partly supported (bottom) bottom edge.*

For larger loads, some other local optima are revealed, and particularly the partly supported 100 N case is clearly inferior to the 150 N design loaded at 100 N. The nonlinearity of the SMA actuator design problem is likely to result in a nonconvex topology optimization problem, where it cannot be guaranteed that a global optimum is obtained. However, performing the design at different loads and making load-scans, as shown in Figure 6.15, increases the chance of finding the optimal design layout.

6.5.2 Effect of mesh refinement

In case of the previously shown results, a mesh of 25 by 40 elements has been used, which is the default mesh for all problems studied in this chapter. Only the left half of the structure is simulated, as symmetry is assumed and enforced, without loss of generality. In this subsection, the same design problems have been evaluated using a refined mesh of 50 by 80 elements. Results for loads of 50, 75 and 100 N for the fully and partly supported bottom case are shown in Figure 6.16. The strokes for coarse and fine cases are listed in Table 6.2.

Table 6.2: *Strokes for optimal designs for various load cases, on coarse and fine meshes.*

Bottom boundary condition	Load magnitude [N]	Stroke [mm] (25×40 mesh)	Stroke [mm] (50×80 mesh)
Fully supported	50	0.0449	0.0800
Fully supported	75	0.1835	0.3423
Fully supported	100	0.3237	0.3269
Fully supported	150	0.3758	(Analysis failed)
Partly supported	50	0.0826	0.1189
Partly supported	75	0.2617	0.4515
Partly supported	100	0.2704	0.5507
Partly supported	150	0.5010	(Analysis failed)

Clearly the finer mesh allows for much more detail in the final design, since no mesh-independence measures or enforced length scales have been used in these studies, in order to reveal the effect of mesh refinement. For the 50 N cases, the topologies show some resemblance to the coarse mesh solutions, although some more detail and thinner structures have developed. The refinement has resulted in all cases in a clear increase in performance, as shown by the strokes listed in Table 6.2. Unlike in the 50 N cases, the other cases show no correspondence in optimal topologies between the coarse and refined meshes. Interesting winding structures are again obtained, particularly in the 100 N cases. Generally, the design evolution during the optimization process is more complex and sometimes exhibits drastic changes during optimization. This was also observed for the coarse mesh case, but is even more apparent in the refined case. Possibly the

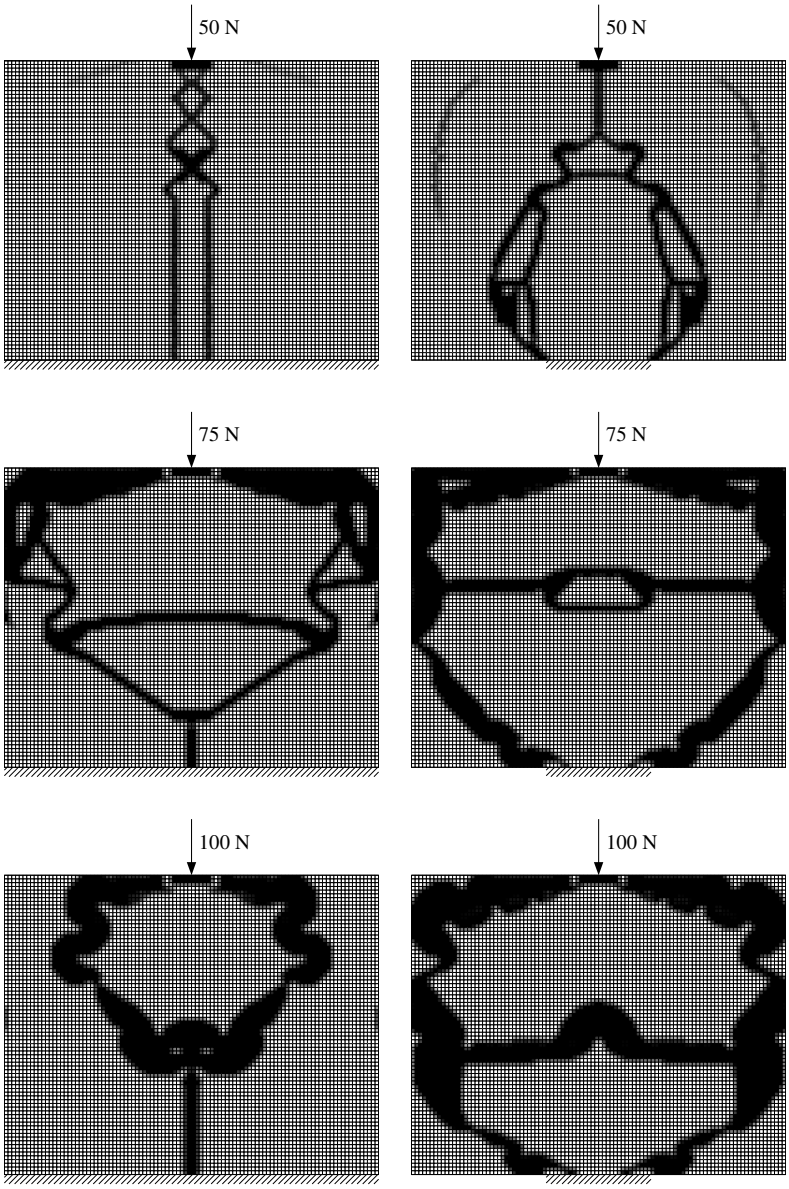


Figure 6.16: *Topologies found by topology optimization in the fully supported (left column) and partly supported (right column) cases, using a refined mesh, for load magnitudes of 50, 75, 100 N.*

fact that the design space is greatly enlarged contributes to this effect. As an illustrative example of the interesting design evolution that occurs during the optimization process, the objective history of the 75 N fully supported case is depicted in Figure 6.17, together with some intermediate designs. The objective curve, although it varies smoothly with the iteration number (due to the use of consistent sensitivities), shows several "bumps" at points where drastic topological changes occurred.

Note that the number of iterations is rather large in this case, i.e. over 800. Most coarse mesh design problems converged in 200-300 iterations, but in the refined mesh cases, the number of design variables is larger. This apparently makes that the optimizer needs more iterations to determine the optimal topology. In addition, optimizer settings were chosen such that the process would not stop prematurely. At roughly iteration 400 the main features of the final design were already clear in the present case.

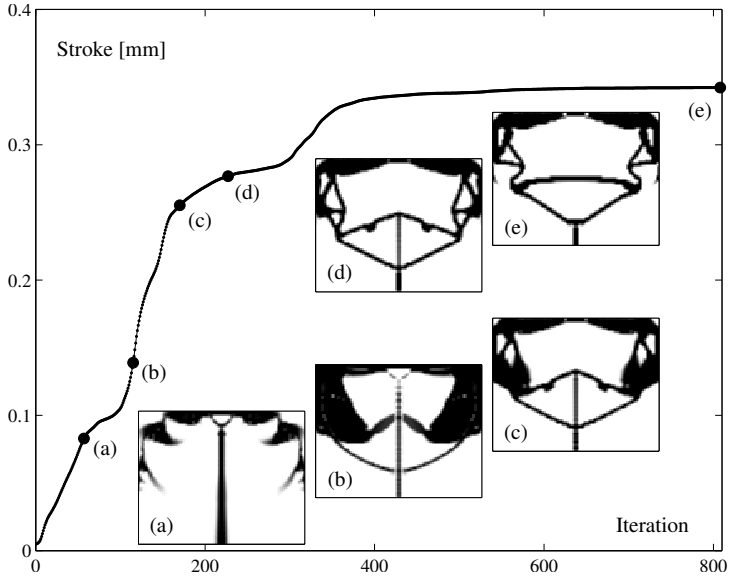


Figure 6.17: *Evolution of the objective and the design for the 75 N fully supported case on a refined mesh. Note the rather drastic changes in the topology during the optimization process.*

No results are shown for the 150 N cases on the refined meshes, because convergence problems were encountered during the optimization process. It turns out, that these cases experience buckling in the low temperature case, which is subsequently exploited by the optimizer. The process is driven towards a very unstable structure which is close to a bifurcation point. This eventually results in a failure to converge within the maximum number of increments. An unconverged result,

obtained after 335 iterations in the fully supported case, is shown in Figure 6.18, together with the evolution of the objective function during the optimization. At roughly iteration 150, the buckling phenomenon starts to develop in the low temperature structure. Very high objective values are obtained, compared to those found in comparable cases: at its peak value in iteration 287, the stroke equals an unparalleled 8.5 mm. Deformed structures for the low and high temperature situation are shown in Figure 6.18 as well. In spite of the fact that nodal design variables are used, a structure that effectively acts as a hinge has formed by an aggregate of elements with intermediate density. It should be noted, however, that this concerns an unconverged design. Clearly, the low temperature case shows a very large deformation, due too the buckling that has occurred. Note, by the way, that the void elements seem to be excessively distorted, but this does not lead to numerical problems because of the ECP approach. The sudden drop in objective in the next step is caused by the fact that the structure no longer buckles, due to the change in design suggested by the optimization algorithm. Although in subsequent iterations the objective starts to improve again, at iteration 335 convergence can not be reached and the process is halted.

In these 150 N cases, a fundamental difficulty of nonlinear problems is encountered: the possibility of numerical complexities such as multiple solutions, instability, buckling and bifurcation points. There are two approaches to deal with this situation: one possibility is to try to avoid these phenomena by, for example, including a buckling constraint, or a constraint on the displacement. Another option, however, is to embrace these features of the problem, since a careful exploitation of buckling phenomena could be very interesting for the design of high-performance SMA binary actuators. The objective values found in the present case clearly illustrate the potential. Care should be taken to make sure that, when making use of buckling, the structure indeed can be cycled between the two configurations that are found by the analysis used here; possibly another simulation approach is required in that case. However, this subject is not pursued further in this thesis.

6.5.3 Effect of load direction

To study the effect of the load direction, the fully and partly supported cases with a 150 N load were also evaluated for upward loads. This resulted in completely different topologies, as shown in Figure 6.19. Possibly, geometrical nonlinearities, in combination with the complex nonlinear SMA material model, are the cause of these differences.

It is interesting to subsequently evaluate the designs obtained with a downward load using an upward load, and vice versa, to check whether any local optima can be identified. The results of these cross-examination checks are listed in Table 6.3.

It can be seen that the solution obtained for the fully supported downward case is a local optimum, since the structure designed with an upward load, loaded with the downward load, performs 4% better. In all other cases, the design load clearly

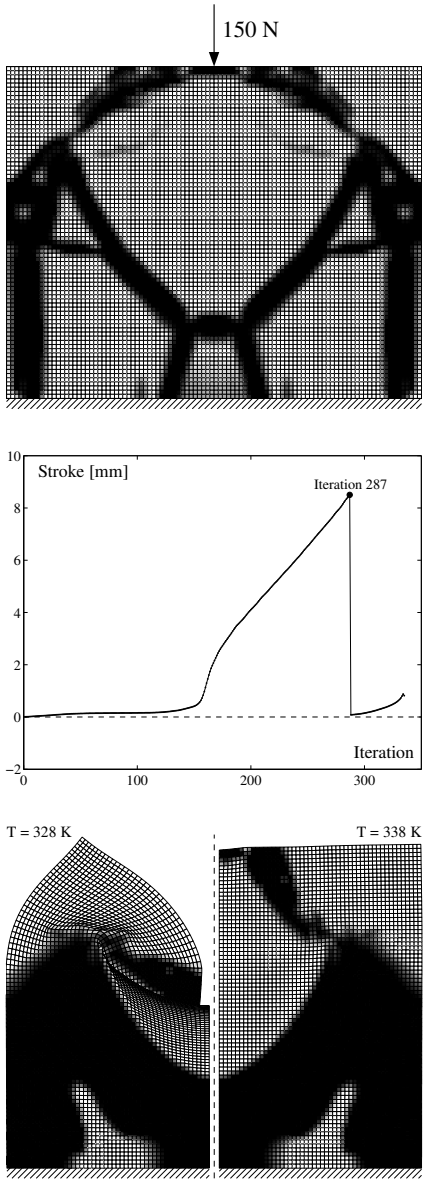


Figure 6.18: *Unconverged result of topology optimization on a refined mesh in case of a 150 N load in the fully supported case (left), together with the evolution of the objective (middle) and the deformed structures at iteration 287.*

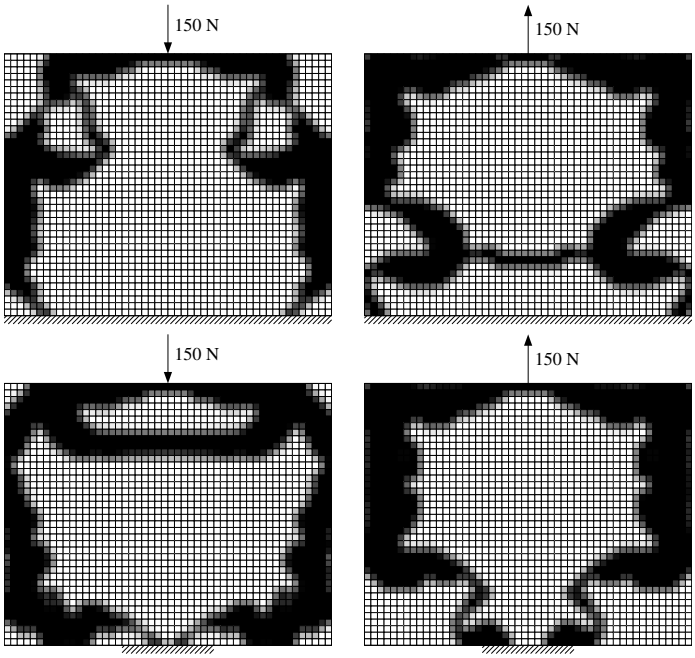


Figure 6.19: *Obtained designs for a 150 N downward and upward loads, for cases with fully (left) and partly supported bottom (right).*

Table 6.3: *Stroke of SMA actuator structures evaluated for different loading directions.*

Design case	Design load	Applied load	Stroke [mm]
Fully supported	150 N down	150 N down	0.3758
Fully supported	150 N up	150 N down	**0.3919**
Fully supported	150 N up	150 N up	0.3117
Fully supported	150 N down	150 N up	0.1867
Partly supported	150 N down	150 N down	0.5010
Partly supported	150 N up	150 N down	0.4329
Partly supported	150 N up	150 N up	0.3635
Partly supported	150 N down	150 N up	0.3418

gave the best performance. It is also interesting to note that there are striking differences in the performance of designs under upward or downward loading. This also illustrates the nonlinearity of this design problem. Geometrically nonlinear effects can lead to differences in the deformation behavior with regard to the loading direction. For instance, the design obtained for the 150 N downward load in the fully supported situation is found to have a stroke of 0.3758 mm in case of downward loading, but only yields 0.1867 mm (less than half) when loaded in the upward direction.

6.5.4 Design improvement from a baseline design

The effect of the initial design used in the topology optimization process is studied by a practical problem concerning the improvement of an existing design, referred to as the *baseline design*. The baseline design used in this study is depicted in Figure 6.20, and is a classical meander-like structure as conventionally used for planar SMA actuators. It is attempted to improve this design using the developed topology optimization procedure. In order to make a fair comparison, the performance of the baseline design is optimized first, by selecting the load magnitude for which it delivers its maximum stroke. For the present design, this optimal load turns out to be 200 N (see Figure 6.20).

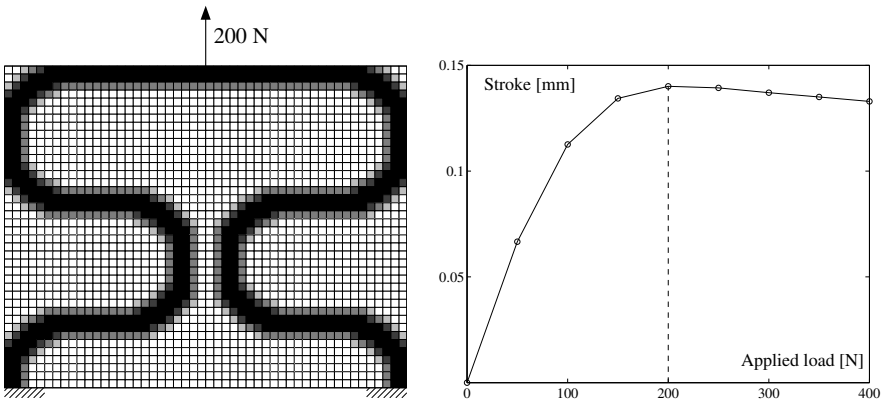


Figure 6.20: *Baseline actuator design and its performance versus various applied loads. The maximum stroke is obtained for a load of 200 N.*

Instead of starting with a uniform density distribution of 50%, the topology optimization procedure was started with the baseline design. Generally, for non-convex problems such as the present SMA topology optimization problem, the starting configuration can have a large effect on the final solution. Many local optima might exist, and which one the process will converge to, depends on the chosen initial design. To avoid any bias, normally uniform density distributions

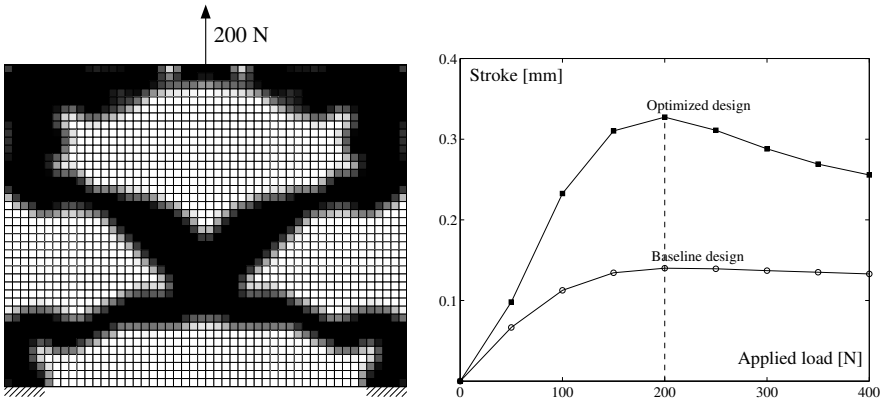


Figure 6.21: *Actuator design obtained by topology optimization starting from the baseline design, and its performance versus various applied loads, together with the performance curve of the baseline design. The optimized design outperforms the baseline designs for the entire range of loads considered.*

are used in topology optimization. In this case, however, the process is deliberately started from the baseline design in order to investigate the ability of the topology optimization procedure to improve the performance of an existing design.

The resulting design is shown in Figure 6.21. The topology of the structure has changed, the two sides have merged in the center. In addition, the structure has become less slender, and some slight winding can be recognized. But most interesting is the striking difference in performance: at the design load, the optimized design shows a 134% better performance compared to the baseline design. Moreover, although this is indeed the peak value of the optimized design, also at other load levels the structure obtained by topology optimization clearly outperforms the baseline design, as shown by the load-stroke data in Figure 6.21.

It is interesting to examine whether such well-performing designs can also be found when starting from a uniform density distribution, in the present case. Using boundary conditions similar to those used in the baseline design, as well as for the fully and partly supported bottom cases used in the previous studies, topology optimization has been performed as well for this 200 N upward load, and resulting topologies are depicted in Figure 6.22.

The strokes of all cases relevant to this problem are listed in Table 6.4. All designs obtained by topology optimization show a large improvement with respect to the baseline design, of the order of a factor 2 or better. The design obtained with the partly supported bottom boundary condition has the smallest stroke of the optimized designs, and clearly is a local optimum. The structure obtained in the fully supported case could also be realized in the partly supported case,

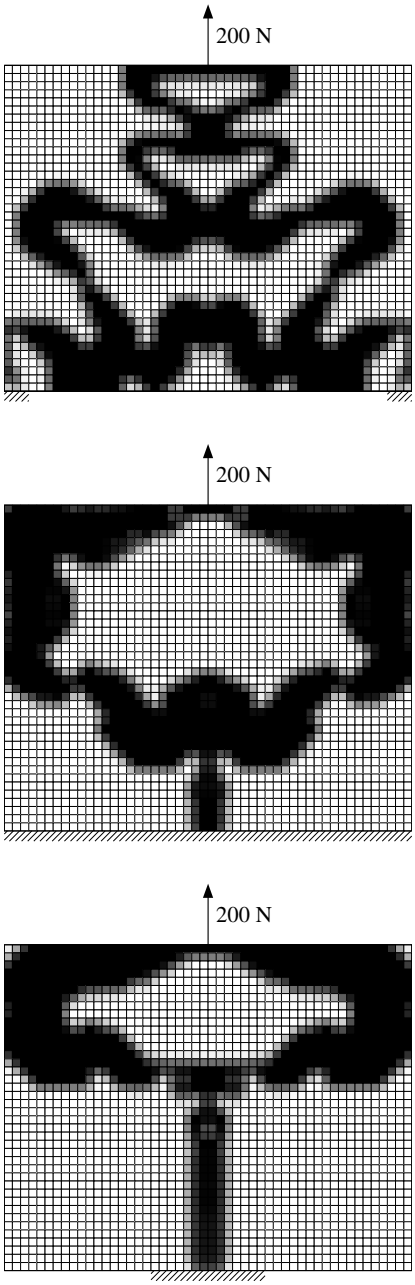


Figure 6.22: *Optimal designs obtained for a 200 N upward load, for different boundary conditions. Optimization is started from a uniform density distribution.*

Table 6.4: *Stroke of SMA actuator structures.*

Design case	Stroke [mm]
Baseline design	0.1401
Optimized design starting from baseline	0.3273
Optimized design starting from uniform density distribution	0.3297
200 N upward load, fully supported bottom	0.3321
200 N upward load, partly supported bottom	0.2780

and shows a better objective (+19%). The design obtained in the case that used the baseline boundary conditions and that was started from a uniform density distribution (Figure 6.22, top figure) has a very interesting topology, of which the upper part actually resembles the shape obtained by the optimization starting from the baseline design. Compared to the latter, it shows an even slightly better performance, but it is debatable whether this difference is significant. It can be argued that the less than perfect approximation of the boundary shape of the structure due to the coarseness of the mesh and the presence of elements with intermediate density values make that the accuracy of the analysis is not optimal. However, this comment is not unique to the present SMA optimization problem, but applies to most topology optimization procedures in general.

It is conceivable, that SMA topology optimization is more sensitive to analysis inaccuracies than topology optimization problems involving linear materials, since the material behavior varies significantly with the strain state and magnitude, and the performance of the active structures is closely linked to this SMA material behavior. However, this difference in sensitivity is difficult to quantify, since there is no clear way to compare such different problems. Still, it would be of interest to study the effect of using different analysis and design meshes, and to investigate how the results of procedures using more refined analysis meshes compare to the results obtained in the present study. Moreover, given the finite accuracy of topology optimization design models in general, also for the present SMA case it is recommended to extract the shape of the design obtained by the topology optimization process, and to perform a subsequent shape optimization using a high-fidelity analysis model, as demonstrated before in e.g. Chapter 4. Such subsequent detailed design, as well as the suggested refinements, are however not the focus of the present chapter, and can be seen as extensions of the procedures introduced here. The redesign example discussed in this subsection convincingly illustrates the merit of the proposed topology optimization procedure, by more than doubling the performance, and in addition the results in Figure 6.21 and Figure 6.22 show the influence of initial design and the boundary conditions on the topology optimization results.

6.6 Conclusions

The successful application of a topology optimization method to the design of shape memory alloy thermal actuators has been presented for the first time. The results presented in this chapter illustrate that the proposed procedure is capable of generating a wide variety of innovative, non-trivial, well-performing designs, and that in addition it is able to improve existing structures significantly. The effectiveness of the procedure relies on two aspects: the efficient and history-independent SMA material model presented in Section 6.2 (see also Chapter 3), and the new efficient and robust ECP topology optimization paradigm. The presented SMA model exploits the small hysteresis of the pseudoelastic behavior of the R-phase/austenite transformation. The resulting formulation combines an adequate representation of experimental observations with excellent suitability for optimization. The history-independent nature of the model enables adjoint sensitivity analysis, which is essential for performing topology optimization. The strong nonlinearity of the model nonetheless requires an elaborate adaptive solution strategy in order to ensure the robustness of the finite element simulations. Also, it should be noted that the SMA model does not cover intricate phenomena such as the effect of crystallographic texture or peculiarities of the material behavior under non-proportional multi-dimensional loading. However, as explained, such detailed modeling would be neither justified nor practical, given the present state of experimental findings and the intended use of the model for (topology) optimization.

Difficulties that have seriously complicated the application of topology optimization to nonlinear problems have been shown to be alleviated by the introduced Element Connectivity Parameterization method. Due to the radically different concept of parameterizing the connectivity between elements instead of modifying the properties of elements themselves, the problem of the conventional density-based method with e.g. excessive distortion of compliant elements is avoided. In addition, in the ECP method there is no need to arbitrarily define a relation between the design variables and the material properties, and the associated complexities in sensitivity analysis and convergence of the optimization process are consequently prevented. These aspects are of particular importance for problems involving complex nonlinear material models, as in the present case of shape memory alloys. Since the ECP method presented in this chapter does not have the disadvantage of increasing the number of degrees of freedom, and given its effectiveness and advantages, it is expected to be an attractive option for many nonlinear topology optimization problems.

It was found that restriction of the design space by the use of nodal design variables was effective in avoiding undesired numerical artifacts, such as unrealistic designs containing one-node hinge patterns. In contrast, the heuristic filtering technique, that has been applied successfully in compliance minimization problems, proved to be ineffective and unable to regularize the topology optimization problem for SMA structures. The strong nonlinearity of the SMA material

combined with the geometrically nonlinear formulation result in a nonconvex optimization problem, which poses a challenge for the optimization algorithm. In several cases, clearly local optima were found. A study of the characteristics of the SMA topology optimization problem has revealed that it shows a strong dependence on the magnitude and direction of the applied load. Interestingly, a volume constraint turned out to be unnecessary, as there apparently exists an optimal amount of material for a given problem setting.

The displacements and strokes of the actuators remained relatively modest in most of the design cases. This is directly related to the fact that the transformation strain in case of the R-phase transformation is rather small. In addition, the considered problems all involved planar structures. In configurations where also transverse bending is possible, as in shell structures, much larger displacements can be realized. In one case, the optimization process was halted because the analysis failed to converge. The convergence difficulties were related to a buckling phenomenon that was observed in the structure. A methodology how to avoid or rather exploit this behavior is yet to be realized. Also, many existing SMA actuators are actuated by a nonuniform temperature field, generated by resistive heating. The problems considered in the present chapter were limited to a uniform temperature, but following the approach of Sigmund (2001) an electrothermomechanical formulation is also possible. These and other extensions of the presented topology optimization technique will be the subject of future investigations.

6.7 Acknowledgments

The authors would like to thank Krister Svanberg for giving permission to use his MMA code for this research.

Based on: Langelaar, M. and van Keulen, F (2006). Sensitivity analysis of shape memory alloy shells. *Computers and Structures*, in review.

Chapter 7

SMA Sensitivity Analysis

This chapter presents procedures to perform efficient design sensitivity analysis for the shape memory alloy (SMA) actuators considered in this thesis. Design sensitivities are derivatives of response quantities with respect to the design variables. Availability of sensitivity information at low computational cost can dramatically improve the efficiency of the optimization process, as it enables use of efficient gradient-based optimization algorithms. The formulation and computation of design sensitivities of SMA shell structures using the direct differentiation method is considered, in a steady state electro-thermo-mechanical finite element context. Finite difference, semi-analytical and refined semi-analytical sensitivity analysis approaches are discussed and compared in terms of efficiency, accuracy and implementation effort, based on a representative finite element model of a miniature SMA gripper.

7.1 Introduction

Shape memory alloys (SMAs) are active materials with a high power density, capable of producing comparatively large actuation strains and stresses (Otsuka and Wayman, 1998). Their actuation properties originate from a solid-state phase transformation, which is affected by changes in temperature or stress, and strains associated with this transformation can be used for actuation. SMA actuators are widely used in wire or spring configurations, but upcoming applications in, e.g., medical instrumentation or microsystems also demand more complex shapes. However, designing effective multi-dimensional SMA actuators is a challenging task, due to the complex behavior of the material and the fact that often electrical, thermal and mechanical aspects have to be considered simultaneously. For

this reason, interest in the application of systematic computational design approaches, such as design optimization techniques, to the design of SMA structures is increasing.

Design optimization has been applied to SMA wire-based configurations (Birman *et al.*, 1996, Troisfontaine *et al.*, 1999) and to SMA structures modeled by analytical models (Lu *et al.*, 2001). However, the models used in these studies cannot be extended to more general SMA structures. In addition, others have applied peak stress reduction algorithms (Mattheck and Burkhardt, 1990) to more complex SMA designs (Kohl *et al.*, 1999, Skrobanek *et al.*, 1997). However, their approach is less versatile than the more general and systematic design optimization techniques developed in the structural optimization community, based on a formal mathematical problem formulation combined with optimization algorithms (see e.g. Haftka and Gürdal (1992) for an overview).

To realize efficient SMA design optimization suited for a wide range of problems of realistic complexity, theavailability of sensitivity information is crucial. Various approaches exist to perform sensitivity analysis, and the available techniques and their characteristics are discussed extensively in dedicated books and review papers (Choi and Kim, 2005a,b, Haftka and Gürdal, 1992, Kleiber *et al.*, 1997, Van Keulen *et al.*, 2005). An essential aspect is that with the appropriate techniques, design sensitivities can often be obtained at low computational cost, compared to the response evaluation itself. This advantage is particularly evident in the case of history-independent nonlinear models (Kleiber *et al.*, 1997). In that case the analysis itself is quite expensive, since the nonlinearity usually requires an incremental-iterative solution strategy. In comparison to this significant computational effort, the sensitivity analysis for path-independent models is far less demanding.

The sensitivity analysis presented in this chapter is based on a simple constitutive model for SMA behavior based on the R-phase transformation in NiTi (Langelaar and Van Keulen, 2004b) (Chapter 3). In contrast to the majority of existing SMA models, this model is history-independent and therefore well suited for use in sensitivity analysis and design optimization. This chapter starts with an overview of various sensitivity analysis approaches in Section 7.2. The present work is aimed particularly at SMA shell structures, as these can generate large actuator displacements through bending deformation. The most general case of actuation by means of resistive heating is considered, which requires a sequentially coupled electrical, thermal and mechanical finite element analysis. Simpler situations, e.g. a given temperature distribution, are also covered by this general formulation. Section 7.3 discusses the derivation and computation of design sensitivities for SMA shell structures in this setting. Numerical results based on finite difference, semi-analytical and refined semi-analytical sensitivity analysis approaches are subsequently presented and discussed in Section 7.4, using a representative case study of a miniature SMA gripper, followed by conclusions.

7.2 Sensitivity analysis approaches

Several approaches exist to perform sensitivity analysis, and the available techniques and their characteristics are briefly reviewed here. For more detailed information, the interested reader is referred to dedicated books by Haftka and Gürdal (1992), Kleiber *et al.* (1997), Choi and Kim (2005a,b) and also the recently published review by Van Keulen *et al.* (2005).

In the following subsections, the system response of interest will be denoted by f and the state variables by $\boldsymbol{u}$. For simplicity, only a single design variable s is considered, without loss of generality. The response is considered to be a function of both $\boldsymbol{u}$ and s, where the state variables also implicitly depend on the design variable, i.e. $f = f(\boldsymbol{u}(s), s)$. Adjoint formulations are not considered, since for the intended shape optimization problems they are not expected to offer significant advantages over the direct differentiation method.

7.2.1 Variational approaches

Figure 7.1 shows a schematic overview that illustrates how the various approaches to perform sensitivity analysis are related to the governing equations. In so-called *continuum* or *variational* approaches, first the continuum governing equations are differentiated with respect to the design variables, and the resulting sensitivity equations are subsequently discretized. In the discrete approaches, on the other hand, the design differentiation is applied to the discretized governing equations.

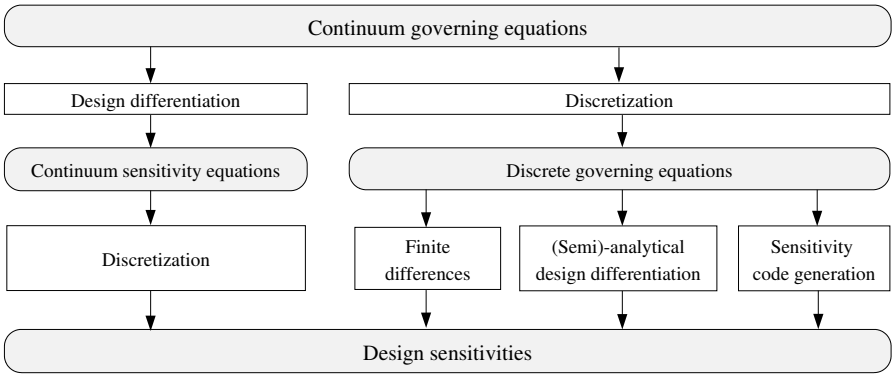


Figure 7.1: *Overview of design sensitivity analysis techniques.*

In the variational approach, the sensitivity equations are not linked to the discretization used for the analysis. This means that theoretically, a different discretization can be used, which makes this approach more flexible than the other methods. However, in practice, in order to make the method efficient and

to limit inconsistencies, it is usually beneficial to use the same discretization. Another advantage is that in case internal routines and quantities of an analysis program can not be accessed, the variational approach could be considered for sensitivity analysis (Choi and Duan, 2000). For the present SMA simulation, the variational approach will not be considered, since other methods are at least equally efficient, and require less implementation effort.

7.2.2 Discrete approaches: finite differences

Discrete design sensitivity approaches are based on the discretized system equations. In the following subsections, the system response of interest will be denoted by f and the state variables by $\boldsymbol{u}$. For simplicity, only a single design variable s is considered, without any loss of generality. The response is considered to be a function of both $\boldsymbol{u}$ and s, where the state variables also implicitly depend on the design variable, i.e. $f = f(\boldsymbol{u}(s), s)$.

A popular sensitivity analysis approach is the use of finite differences to approximate the design derivatives. One or more additional simulations are required to evaluate the responses in perturbed design configurations $f(s \pm \Delta s)$, after completing the nominal analysis $f(s)$, e.g.:

$$\text{Forward finite difference:} \quad \frac{df}{ds} \approx \frac{f(s + \Delta s) - f(s)}{\Delta s} \tag{7.1}$$

$$\text{Central finite difference:} \quad \frac{df}{ds} \approx \frac{f(s + \Delta s) - f(s - \Delta s)}{2\Delta s} \tag{7.2}$$

Critical for the accuracy of the obtained derivatives is the selection of the proper design perturbation Δs. A larger perturbation worsens truncation errors due to the truncation of the Taylor series, a smaller perturbation can lead to round-off or condition errors. In practical situations, it can be difficult to select a suitable perturbation, particularly in case of "noisy" numerical models. With noise is meant the erratic variation of the response due to numerical reasons, for example due to adaptive mesh refinement or finite convergence tolerances for iterative procedures. When due to this noisy character the response is not a smooth function of the design variable, the finite difference approach can produce highly inaccurate results.

Next to these weaknesses, the fact that the finite difference method requires one or more additional function evaluations for every design variable makes it rather costly. To reduce the computational effort, use can be made of efficient reanalysis techniques, that can evaluate or approximate the perturbed responses at a fraction of the computational cost of a full analysis. Particularly in case of path-independent nonlinear problems, the efficiency of the finite difference approach can be improved significantly, when after the nominal analysis the perturbed analysis is started from the nominal solution, instead of starting from the initial configuration. In this setting, care must be taken to correct for the finite residual that remains after convergence of the solution procedure, to ensure accurate

sensitivity results when using reanalysis. An effective remedy has been suggested by Haftka (1985), who has shown that subtracting the nominal residual from the applied load in the perturbed problem alleviates this problem.

An additional improvement of the basic finite difference method has initially been suggested by Lyness (1967), Lyness and Moler (1967), and recently described by Squire and Trapp (1998). By using complex arithmetic, the finite difference approach can be made insensitive to the numerical round-off error, which allows the use of very small perturbations. The complex forward difference equation for the design sensitivity with respect to s is given by:

$$\frac{df}{ds} = \mathrm{Im}(f(s + i\Delta s)) \tag{7.3}$$

Drawbacks of this approach are that the entire analysis has to be performed using complex operations, which is up to four times as expensive as a normal analysis using only real numbers, and that twice the amount of memory is required. Naturally, full access to the source code is also necessary for this approach. The implementation effort, on the other hand, is reported to be relatively modest (Anderson and Nielsen, 2001).

7.2.3 Discrete approaches: semi-analytical design sensitivities

An apparently straightforward way to compute sensitivity information based on the discretized system equations is to differentiate the discrete equations analytically with respect to the design variables. However, it turns out that for most models of realistic complexity, this differentiation is tedious, error-prone and requires a large implementation effort. Therefore, the popular semi-analytical methods use analytical design differentiation only at the global level of the discretized equations, and the lower level derivatives are approximated by finite difference techniques. To illustrate this method, a linear finite element setting is considered, governed by:

$$\mathbf{K}\boldsymbol{u} = \boldsymbol{f}. \tag{7.4}$$

This is a common equation in a solid mechanics context, where $\mathbf{K}$ represents the tangent stiffness matrix, $\boldsymbol{u}$ the displacement vector and $\boldsymbol{f}$ the load vector. The load vector and the stiffness matrix are considered to be independent of the displacements. In a general design optimization setting, $\mathbf{K}$ and $\boldsymbol{f}$ can depend on design variable s, hence also the state variable vector $\boldsymbol{u}$ implicitly depends on s. The sensitivity of a response $f(\boldsymbol{u}(s), s)$ is now given by:

$$\frac{df}{ds} = \frac{\partial f}{\partial s} + \frac{\partial f}{\partial \boldsymbol{u}}^T \frac{d\boldsymbol{u}}{ds}. \tag{7.5}$$

It is seen, that in order to evaluate this response sensitivity, the sensitivity of the state variables $d\boldsymbol{u}/ds$ is required. An equation to evaluate this quantity is

obtained by design differentiation of the governing equations (Equation 7.4):

$$\frac{d\mathbf{K}}{ds}\boldsymbol{u} + \mathbf{K}\frac{d\boldsymbol{u}}{ds} = \frac{d\boldsymbol{f}}{ds} \;\Rightarrow\; \mathbf{K}\frac{d\boldsymbol{u}}{ds} = \left(\frac{d\boldsymbol{f}}{ds} - \frac{d\mathbf{K}}{ds}\boldsymbol{u}\right). \tag{7.6}$$

This equation for the design sensitivity has the same structure as the original governing equation (Equation 7.4). The right-hand side is referred to as the *pseudo-load*, since "loading" the structure with this "load" results in the sensitivities of the displacements. In the semi-analytical method, the design derivatives $d\boldsymbol{f}/ds$ and $d\mathbf{K}/ds$ needed to compute this pseudo-load are not evaluated analytically, but by finite differences. Usually forward finite differences are used:

$$\frac{d\mathbf{K}}{ds} \approx \frac{\mathbf{K}(s+\Delta s) - \mathbf{K}}{\Delta s}, \qquad \frac{d\boldsymbol{f}}{ds} \approx \frac{\boldsymbol{f}(s+\Delta s) - \boldsymbol{f}}{\Delta s}. \tag{7.7}$$

When a direct solver is used, the system matrix $\mathbf{K}$ has already been decomposed, and $d\boldsymbol{u}/ds$ can be computed by backsubstitution, at a fraction of the computational cost of the analysis itself. When an iterative solver is used, the preconditioning effort can be reused.

Slender structures where elements undergo significant rotations have shown reduced accuracy of semi-analytical sensitivities. This problem is attributed to increased errors in the finite difference components used for the pseudo-load calculation. Various remedies have been suggested, including the use of higher order finite difference formulae, an "exact" formulation (Olhoff *et al.*, 1993) and the so-called "refined" semi-analytical formulation, which is based on exact differentiation of the rigid body modes of elements (De Boer and Van Keulen, 2000b, Van Keulen and De Boer, 1998b). Element rigid body modes are displacement fields that do not generate internal forces, i.e. on element level the following holds:

$$\mathbf{K}\boldsymbol{u}_r = \mathbf{0}. \tag{7.8}$$

Design differentiation of this condition gives

$$\frac{d\mathbf{K}}{ds}\boldsymbol{u}_r + \mathbf{K}\frac{d\boldsymbol{u}_r}{ds} = \mathbf{0} \;\Rightarrow\; \frac{d\mathbf{K}}{ds}\boldsymbol{u}_r = -\mathbf{K}\frac{d\boldsymbol{u}_r}{ds}. \tag{7.9}$$

The sensitivity of rigid body modes $\boldsymbol{u}_r$ is easy to compute analytically, and the part of the pseudo-load vector term associated with these modes $(d\mathbf{K}/ds)\boldsymbol{u}_r$ can therefore be evaluated exactly, using the relation above. An alternative to avoid the finite difference inaccuracies might also be to use the complex variable approach discussed in the previous section.

7.2.4 Discrete approaches: sensitivity code generation

A third discrete approach is to consider not even the discrete governing equations, but operate directly on the source code by which they are implemented. Programs have been developed that parse this source code and generate derived code that, when compiled, is able to compute the design derivatives directly. Some well known programs are for example ADIFOR (Bischof *et al.*, 1996) and ADOL-C (Griewank *et al.*, 1996), among others. This technique could be regarded as automated analytical differentiation, and therefore the resulting sensitivities are traditionally referred to as "automated derivatives" or "computational derivatives".

Sensitivity code generation techniques cannot exploit the structure of the governing equations, such as, for example, the reuse of the decomposed tangent stiffness matrix in semi-analytical approaches. Also, fully analytical partial design derivatives not seldomly require considerably more computational effort compared to the finite difference approach used in semi-analytical methods. Still, potentially this approach allows for the implementation of design sensitivities with minimal human implementation effort. However, generally for larger programs, the application of sensitivity code generators is not straightforward, and human intervention is often needed to obtain reasonably efficient derivative code (Bartholomew-Biggs *et al.*, 2000, Coleman and Jonsson, 1999, Sherman *et al.*, 1994). Using sensitivity code generation selectively for parts of a code, possibly in combination with a semi-analytical formulation, seems a better approach than applying it to complete, complex programs at once.

Typically, a design sensitivity computation for a case of realistic complexity takes the time of several function evaluations, in the case of straightforward application of these methods (Bartholomew-Biggs *et al.*, 2000, Coleman and Jonsson, 1999). For the present SMA case, these code generation techniques have not been used, because manual differentiation provides more insight and control, and because no sensitivity code generator is currently available for the Pascal programming language used for the SMA finite element analysis implementation.

7.2.5 Discussion

Having briefly reviewed the options for design sensitivity analysis, it can be concluded that the most appropriate methods to implement for the present SMA case are the direct semi-analytical method, as well as the finite difference method combined with a fast reanalysis technique. The latter could be particularly useful in case of the active catheter model discussed in Chapter 3, because the symmetry constraint in that model complicates the semi-analytical design sensitivity analysis. Basically this constraint makes that a directional derivative should be used. When the constraint would be enforced with a penalty method, there is no problem, and the standard semi-analytical formulation can be used. However, with the employed augmented Lagrangian method, with its iteratively updated

Lagrange multipliers, the situation is more complicated. A possible solution is to revert to the standard Lagrangian formulation after convergence, but using finite differences in combination with fast reanalysis is expected to require far less implementation effort. This problem is addressed in more detail in Chapter 9.

Concerning the SMA model, it is observed that the constitutive equations are based on a piecewise linear stress-strain relation (see e.g. Chapter 2 and Chapter 3). This leads to the situation that non-differentiable states exists in the simulated structures when the material is exactly at one of the transition points, which could yield complications in the sensitivity analysis. In these points, formally only the existence of directional derivatives is well-defined. Similar problems occur in e.g. elastoplastic models, but investigations have shown that, although theoretically this non-differentiability presents a problem, in practical numerical cases, it is highly unlikely that material states in the integration points *exactly* correspond to these transition points, and therefore the practical effect on the accuracy of the sensitivity is negligible (Cho and Choi, 2000, Schwarz and Ramm, 2001, Vidal and Haber, 1993). For this reason, similar to the elasto-plastic case, no special treatment is applied to the transition points in the sensitivity analysis of the present SMA model.

7.3 Sensitivity analysis of SMA finite element model

This section treats the formulation of semi-analytical design sensitivities for electro-thermally actuated SMA structures, as the comparison of approaches given in Section 7.2 showed this to be an attractive approach for such problems. The other attractive method, finite differences in combination with fast reanalysis, is applied as described by Haftka (1985), and requires no further explanation. This section will also briefly outline essential details of the coupled finite element simulation of the SMA structures, to the extent that is required to adequately specify the formulation of the sensitivity equations.

For the electrical and thermal simulation, the field equation formulation and the associated three-noded triangular element as given by Fagan (1992) is used. The mechanical analysis is performed using triangular shell elements developed by Van Keulen and Booij (1996), in combination with the SMA constitutive model described in Chapter 3. Details of the used elements are not given here, but can be found in the mentioned references. Topologically equivalent meshes are used for the electrical, thermal and mechanical cases, thus no mapping of quantities between disciplines is required, which simplifies the coupling in the multidisciplinary analysis. However, the approach outlined here basically also applies to cases using different elements and meshes. The Einstein summation convention is used in this section, and partial derivatives are abbreviated by $\partial f/\partial x_i = f_{,i}$.

7.3.1 Electrical and thermal case

The isotropic electrical and thermal steady-state problems are governed by the same partial differential equation, known as the Poisson equation:

$$K\phi_{,ii} + Q = 0. \tag{7.10}$$

Here ϕ represents the potential field, which corresponds to the electrical potential in the electrical case, and the temperature in the thermal case. K is the isotropic conductivity, a material property, and Q is a local source term. The boundary conditions (BCs) for Equation 7.10 can be the following:

$$\begin{array}{ll} \text{Dirichlet BCs:} & \phi = \hat{\phi}, \\ \text{Neumann BCs:} & K\phi_{,i}n_i = \hat{q}, \\ \text{Robin BCs:} & K\phi_{,i}n_i = h(\phi - \phi_\infty). \end{array} \tag{7.11}$$

The Dirichlet and Neumann BCs prescribe either the potential $\hat{\phi}$ at or the flux $\hat{q}$ through the boundary, given by the outward boundary normal $\boldsymbol{n}$. The Robin BC only applies to the thermal case, where it describes the convection at the surface. Here h is the convection coefficient and ϕ_∞ represents the ambient temperature.

Discretization of Equation 7.10 using the finite element method results in a system of linear equations. To distinguish between the electrical and thermal cases, subscripts e and t are used:

$$\begin{array}{ll} \text{Electrical case:} & \mathbf{K}_e(s)\boldsymbol{u}_e(s) = \boldsymbol{f}_e(s), \\ \text{Thermal case:} & \mathbf{K}_t(s)\boldsymbol{u}_t(s) = \boldsymbol{f}_t(\boldsymbol{u}_e(s)). \end{array} \tag{7.12}$$

Note, in the thermal case, the load vector is a function of the electrical variables, since resistive heating is used to generate internal heat sources in the structure. In the finite element setting, the locally dissipated energy q can be evaluated by:

$$q = \boldsymbol{u}_e^T \mathbf{K}_e \boldsymbol{u}_e. \tag{7.13}$$

This equation is to be evaluated on element level, and each contribution is added to the thermal load vector $\boldsymbol{f}_t$ in the thermal analysis.

The sensitivities for the electrical case can be obtained directly using the direct semi-analytical approach given in Equation 7.6:

$$\frac{d\boldsymbol{u}_e}{ds} = \mathbf{K}_e^{-1}\left(\frac{d\boldsymbol{f}_e}{ds} - \frac{d\mathbf{K}_e}{ds}\boldsymbol{u}_e\right), \tag{7.14}$$

Forward finite differences are used to approximate the derivatives of $\mathbf{K}_e$ and $\boldsymbol{f}_e$. Similarly, the thermal case yields:

$$\frac{d\boldsymbol{u}_t}{ds} = \mathbf{K}_t^{-1}\left(\frac{d\boldsymbol{f}_t}{ds} - \frac{d\mathbf{K}_t}{ds}\boldsymbol{u}_t\right). \tag{7.15}$$

However, the thermal load vector $\boldsymbol{f}_t$ is a function of the electrical variables, and its design derivative $d\boldsymbol{f}_t/ds$ is used in this expression. It is possible to evaluate this term by means of finite differences, but this would require an additional perturbed electrical analysis for every design variable. A more efficient approach is to compute the analytic design derivative of q (Equation 7.13):

$$\frac{dq}{ds} = 2\boldsymbol{u}_e^T \mathbf{K}_e \frac{d\boldsymbol{u}_e}{ds} + \boldsymbol{u}_e^T \frac{d\mathbf{K}_e}{ds} \boldsymbol{u}_e. \tag{7.16}$$

The design derivative of the electrical tangent matrix $\mathbf{K}_e$ in this expression can again be approximated by finite differences. This term is also used in the sensitivity analysis for the electrical state variables, given in Equation 7.14. Therefore, in the present implementation, the design sensitivity of the dissipated heat given in Equation 7.16 and the electrical state variable sensitivity are evaluated simultaneously, to avoid repeated calculations. The procedure has been verified by comparing the results to finite difference computations.

7.3.2 Mechanical case

In the modeling of the mechanical case, geometrical and physical nonlinearities are included. A Total Lagrangian approach is adopted, and the SMA material behavior is represented by the constitutive model described in Chapter 3. The coupling between the thermal and mechanical analysis originates from the temperature-dependence of the SMA model. Finite element discretization yields the following discrete nonlinear governing equations:

$$\boldsymbol{R}(\boldsymbol{u}(s), \boldsymbol{u}_t(s), s) = \boldsymbol{f}(s) - \boldsymbol{f}_{int}(\boldsymbol{u}, \boldsymbol{u}_t(s), s) = \mathbf{0} \tag{7.17}$$

Here, the residual $\boldsymbol{R}$ is a function of the nodal displacement vector $\boldsymbol{u}$ and the temperatures $\boldsymbol{u}_t$, and is given by the difference between the external force vector $\boldsymbol{f}$ and the internal force vector $\boldsymbol{f}_{int}$. Configuration- or temperature-dependent external forces are not accounted for in the considered problem, hence only $\boldsymbol{f}_{int}$ is a function of $\boldsymbol{u}$ and $\boldsymbol{u}_t$. This set of nonlinear equations is solved using an incremental-iterative method. Design differentiation of Equation 7.17 yields:

$$\frac{d\boldsymbol{R}}{ds} = \frac{\partial \boldsymbol{R}}{\partial s} + \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}} \frac{d\boldsymbol{u}}{ds} + \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}_t} \frac{d\boldsymbol{u}_t}{ds}. \tag{7.18}$$

The quantity $-\partial \boldsymbol{R}/\partial \boldsymbol{u}$ is referred to as the tangent stiffness matrix $\mathbf{K}$. Rearranging and using this definition of $\mathbf{K}$ gives:

$$\frac{d\boldsymbol{u}}{ds} = \mathbf{K}^{-1} \left(\frac{\partial \boldsymbol{R}}{\partial s} + \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}_t} \frac{d\boldsymbol{u}_t}{ds} \right). \tag{7.19}$$

The design sensitivities of the temperature field $d\boldsymbol{u}_t/ds$ have already been computed after the thermal analysis, and the explicit partial derivative of the residual

with respect to the design variable $\partial \boldsymbol{R}/\partial s$ can be computed conveniently using finite differences, following the semi-analytical approach. The remaining term in the pseudo-load vector contains the expression $\partial \boldsymbol{R}/\partial \boldsymbol{u}_t$, which represents the coupling between the thermal and mechanical problem. This expression can be interpreted as a matrix with n_m rows and n_t columns, where n_m is the number of mechanical degrees of freedom and n_t the number of thermal degrees of freedom. Fortunately, this matrix is very sparse, and in practice it is never formed explicitly. Moreover, from Equation 7.17, it follows that when considering the influence of the temperature distribution $\boldsymbol{u}_t$ on the residual $\boldsymbol{R}$, only the internal load component of the residual has to be considered:

$$\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}_t} = -\frac{\partial \boldsymbol{f}_{int}}{\partial \boldsymbol{u}_t}. \tag{7.20}$$

Note that this term needs to be computed only once, regardless of the number of responses or design variables. Two approaches have been explored and implemented to compute this coupling term: one using finite differences and another based on a fully analytical derivation, which both are discussed hereafter.

Finite difference approach

It is important to reduce the quantity used in the finite difference calculation as much as possible, to avoid unnecessary computations. Straightforward application of finite differences to the complete internal force vector would require as many internal force evaluations as there are thermal degrees of freedom. Given the complexity of the material model and its non-negligible evaluation time, this would be a rather inefficient procedure. Therefore, before discussing the finite difference approach to evaluate the thermo-mechanical coupling term, the internal force contribution of an individual element is considered more closely.

The internal force contribution of an individual finite element is given by:

$$\boldsymbol{f}_{int} = \mathbf{D}^T \boldsymbol{\sigma} \quad \text{with} \quad D_{ij} = \frac{\partial \varepsilon_i}{\partial u_j}. \tag{7.21}$$

Here $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ are vectors of generalized stresses and strains, respectively. All quantities in this subsection are element-based quantities, unless stated otherwise. From Equation 7.21, and given the fact that $\mathbf{D}$ does not depend on the temperature, it follows that:

$$\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}_t} = -\frac{\partial \boldsymbol{f}_{int}}{\partial \boldsymbol{u}_t} = -\mathbf{D}^T \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{u}_t}. \tag{7.22}$$

Hence, by evaluating the term $\partial \boldsymbol{\sigma}/\partial \boldsymbol{u}_t$, the coupling term can be obtained. Note that the element stress will only depend on the local temperature. The shell element used in this study uses only a single integration point, and its temperature

is given by the average value T_{avg} of the nodal temperatures of its three corner nodes $u_t^{(i)}$:

$$T_{avg} = \frac{1}{3}\sum_{i=1}^{3} u_t^{(i)}. \tag{7.23}$$

From this, it follows that the design sensitivity of the average element temperature equals:

$$\frac{dT_{avg}}{ds} = \frac{1}{3}\sum_{i=1}^{3} \frac{du_t^{(i)}}{ds}. \tag{7.24}$$

Subsequently, the coupling term contribution for a given element becomes:

$$\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}_t}\frac{d\boldsymbol{u}_t}{ds} = \mathbf{D}^T\left(-\frac{\partial \boldsymbol{\sigma}}{\partial T_{avg}}\right)\frac{dT_{avg}}{ds}. \tag{7.25}$$

The total thermo-mechanical sensitivity coupling term in the system-level pseudo-load vector given in Equation 7.19 is obtained by assembling these element contributions. The derivative of the element stress vector $\boldsymbol{\sigma}$ with respect to the average element temperature T_{avg} is evaluated using forward finite differences:

$$\frac{\partial \boldsymbol{\sigma}}{\partial T_{avg}} \approx \frac{\boldsymbol{\sigma}(T_{avg} + \Delta T_{avg}) - \boldsymbol{\sigma}(T_{avg})}{\Delta T_{avg}}. \tag{7.26}$$

Note that the cost of this procedure is comparable to a single stress update as performed in every Newton iteration, and that these terms can be reused for multiple design variables.

Analytical approach

As shown in the previous section, the derivative of the element stress vector $\boldsymbol{\sigma}$ with respect to the average element temperature T_{avg} is the only nonstandard term that needs to be computed in order to evaluate the thermo-mechanical coupling sensitivity related term of the pseudo-load vector (Equation 7.19). Instead of using finite differences to compute the this term, as in Equation 7.26, it is also possible to use a fully analytical approach. The advantage is that possible accuracy problems of the finite difference approach are avoided, which is known to suffer from sensitivity to the perturbation size, and poor performance with noisy numerical models. The SMA material model internally uses an iterative solution process, which is terminated once a certain tolerance criterion is satisfied. This is a typical source of numerical noise, that could spoil the accuracy of the finite difference approach.

To derive the term $\partial\boldsymbol{\sigma}/\partial T_{avg}$ analytically, it is necessary to consider the SMA material model in more detail. In-depth information on this model can be found in Chapter 3 and Chapter 6. The focus here is on the derivatives of the components

of the stress tensor, since the derivatives of the generalized stresses easily follow once these stress tensor component derivatives have been obtained.

In the current plane stress setting, the SMA constitutive equations are formulated by:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sqrt{2}\sigma_{xy} \end{Bmatrix} = \left(K \begin{bmatrix} 1+\alpha & 1+\alpha & 0 \\ 1+\alpha & 1+\alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} + \ldots \right. \\ \left. \frac{2G}{3} \begin{bmatrix} 2-\alpha & -(1+\alpha) & 0 \\ -(1+\alpha) & 2-\alpha & 0 \\ 0 & 0 & 3 \end{bmatrix} \right) \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix}, \tag{7.27}$$

with $\alpha = (2G-3K)/(4G+3K)$, where K is the bulk modulus. The temperature-dependent quantity here is the shear modulus G, and since α is a function of G, it depends on T as well. From the definition of α, it follows that:

$$\frac{\partial \alpha}{\partial T} = \frac{\partial}{\partial T}\left(\frac{2G-3K}{4G+3K}\right) = \frac{18K}{(4G+3K)^2}\frac{\partial G}{\partial T}. \tag{7.28}$$

With this, differentiation of Equation 7.27 with respect to the temperature T yields:

$$\begin{Bmatrix} \frac{\partial \sigma_{xx}}{\partial T} \\ \frac{\partial \sigma_{yy}}{\partial T} \\ \frac{\partial \sqrt{2}\sigma_{xy}}{\partial T} \end{Bmatrix} = \frac{\partial G}{\partial T} \left(\frac{6K(3K-2G)}{(4G+3K)^2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \ldots \right. \\ \left. \frac{2}{3} \begin{bmatrix} 2-\alpha & -(1+\alpha) & 0 \\ -(1+\alpha) & 2-\alpha & 0 \\ 0 & 0 & 3 \end{bmatrix} \right) \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix}. \tag{7.29}$$

The remaining unknown in this equation is $\partial G/\partial T$. This term is not easily evaluated, because no explicit equation is available for $G(T)$. Therefore, implicit differentiation has to be used. The equation that is solved iteratively in order to obtain G and the effective strain ε_e is given in Chapter 6 by:

$$Z = C_i(T) + \frac{D_i(T)}{\varepsilon_e} - G = 0, \tag{7.30}$$

where the effective strain is given by

$$\varepsilon_e = \sqrt{\frac{4}{9}(\alpha^2-\alpha+1)(\varepsilon_{xx}^2+\varepsilon_{yy}^2) + \frac{4}{9}(2\alpha^2-2\alpha-1)\varepsilon_{xx}\varepsilon_{yy} + \frac{4}{3}\varepsilon_{xy}^2}. \tag{7.31}$$

C_i and D_i are quantities that depend on the temperature and the material parameters. Their exact form as well as their derivatives are given in Appendix E.

Note also the presence of α in the effective strain definition, which makes Z a nonlinear function of G. Implicit differentiation of Equation 7.30 yields:

$$Z(T, G(T)) = 0 \Rightarrow \frac{\partial Z}{\partial T} + \frac{\partial Z}{\partial G}\frac{\partial G}{\partial T} = 0 \Rightarrow \frac{\partial G}{\partial T} = -\frac{\frac{\partial Z}{\partial T}}{\frac{\partial Z}{\partial G}}. \tag{7.32}$$

Using this relation, the temperature derivative of G is found to be given by:

$$\frac{\partial G}{\partial T} = \frac{\frac{dC_i}{dT} + \frac{1}{\varepsilon_e}\frac{dD_i}{dT}}{\frac{4D_iK(\varepsilon_{xx}+\varepsilon_{yy})^2(2\alpha-1)}{(4G+3K)^2\varepsilon_e^3} + 1}. \tag{7.33}$$

With this expression, the temperature sensitivity of the stress components given in Equation 7.29 can be evaluated analytically. No iterations are required, in contrast to the normal stress update procedure. No special measures are taken at the transition points, where strictly speaking only directional derivatives are defined. However, as discussed previously, it is unlikely that this will result in a severe loss of accuracy. In spite of the expectation that the finite difference appraoch discussed before might suffer from a possible noisy behavior of the stress function, no indications of such problems were found in practical tests. A comparison between this analytical approach and the finite difference option revealed no significant differences in accuracy or evaluation time. In the implementation of the semi-analytical sensitivity analysis, the analytical approach as discussed here has been used.

7.3.3 Effective strain sensitivity

An important response quantity used in optimization problems involving the present SMA model is the effective strain. To make sure the model is not used outside its range of applicability, constraints are included that limit the maximum effective strain to $\varepsilon_e^{max} = 1\%$:

$$\max_z \varepsilon_e - \varepsilon_e^{max} \leq 0. \tag{7.34}$$

Here z is the transverse coordinate of the shell element. Because the maximum effective strain within an element occurs at the outside layer (see Appendix F), in practice only the case $|z| = t/2$, where t represents the shell thickness, needs to be considered. Since including this maximum effective strain condition for every individual element leads to a large number of constraints, and since in practice a small number of isolated and limited constraint violations can be tolerated, the following Kreisselmeier-Steinhauser (Kreisselmeier and Steinhauser, 1983) constraint aggregation function can be used to combine all element constraints $g_i = \varepsilon_e^{(i)} - \varepsilon_e^{max}$ into a single constraint, given by:

$$KS_A(\boldsymbol{g}) = \frac{1}{\rho}\ln\left(\frac{1}{\sum_{i=1}^N A_i}\cdot\sum_{i=1}^N A_i e^{\rho g_i}\right). \tag{7.35}$$

Here the element contributions are weighted by the element area A_i, to reduce the significance of small local violations. The parameter ρ controls the tolerance of the aggregated constraint and can be chosen depending on the demands of the application. This approach has also been used in the SMA gripper optimization described in Chapter 5. A larger ρ-value allows smaller isolated violations. The effective strain constraint is now given by:

$$KS_A(\boldsymbol{g}) - 1 \leq 0 \tag{7.36}$$

Since the Kreisselmeier-Steinhauser aggregation of effective strain constraints is an important response quantity, and because obtaining the design sensitivity of this quantity is nontrivial, detailed expressions for this sensitivity are derived here. The sensitivity of $KS_A(\boldsymbol{g})$ is given by:

$$\frac{dKS_A(\boldsymbol{g})}{ds} = \frac{\sum_i \rho A_i e^{\rho g_i} \frac{dg_i}{ds} + \sum_i e^{\rho g_i} \frac{dA_i}{ds} - \frac{\sum_i \frac{dA_i}{ds}}{\sum_i A_i} \cdot \sum_i A_i e^{\rho g_i}}{\rho \sum_i A_i e^{\rho g_i}}. \tag{7.37}$$

Note that not only the sensitivity of the constraint dg_i/ds is required, but also the sensitivity of the element area dA_i/ds plays a role. This term is relevant when dealing with shape design variables. In the implementation, it is evaluated using forward finite differences.

Remains the sensitivity of the maximum effective strain constraint g_i itself, which is equal to the sensitivity of the effective strain at a given position z. Since the maximum occurs at the outside layer of the shell (see Appendix E), only the effective strains at both outside layers are evaluated, and also only the sensitivities of these quantities are relevant. Analytical or semi-analytical evaluation of the effective strain sensitivity is preferred over finite differences, because the latter would require solving the perturbed displacement vector for every design variable. These analytical sensitivities are discussed here, as several nontrivial steps are involved.

In the present shell element setting, the effective strain is given by

$$\varepsilon_e = \sqrt{\begin{array}{l} \frac{4}{9}A_1((\varepsilon_{xx} + z\kappa_{xx})^2 + (\varepsilon_{yy} + z\kappa_{yy})^2) + \dots \\ \qquad \frac{4}{9}A_2(\varepsilon_{xx} + z\kappa_{xx})(\varepsilon_{yy} + z\kappa_{yy}) + \frac{4}{3}(\varepsilon_{xy} + z\kappa_{xy})^2 \end{array}}, \tag{7.38}$$

where

$$\alpha = \frac{2G - 3K}{4G + 3K}, \qquad A_1 = \alpha^2 - \alpha + 1, \qquad A_2 = 2\alpha^2 - 2\alpha - 1. \tag{7.39}$$

The ε_{ij} and κ_{ij} quantities are the membrane strain and curvature components at the midplane of the shell. Design differentiation of ε_e yields:

$$\frac{d\varepsilon_e}{ds} = \frac{\partial \varepsilon_e}{\partial \boldsymbol{\varepsilon}}^T \frac{d\boldsymbol{\varepsilon}}{ds} + \frac{\partial \varepsilon_e}{\partial z}\frac{dz}{ds} + \frac{\partial \varepsilon_e}{\partial \alpha}\left(\frac{\partial \alpha}{\partial s} + \frac{\partial \alpha}{\partial G}\frac{dG}{ds}\right). \tag{7.40}$$

Here the term

$$\frac{\partial \varepsilon_e}{\partial \boldsymbol{\varepsilon}}^T \frac{d\boldsymbol{\varepsilon}}{ds} = \frac{\partial \varepsilon_e}{\partial \varepsilon_{xx}} \frac{d\varepsilon_{xx}}{ds} + \ldots + \frac{\partial \varepsilon_e}{\partial \kappa_{xy}} \frac{d\kappa_{xy}}{ds} \tag{7.41}$$

stands for the complete summation of products of partial effective strain sensitivities with respect to the strain components and their respective total sensitivities. These strain component sensitivities are derived from the sensitivities of the generalized strains $\boldsymbol{\varepsilon}$, computed using

$$\frac{d\boldsymbol{\varepsilon}}{ds} = \frac{d\mathbf{D}}{ds}\boldsymbol{u} + \mathbf{D}\frac{d\boldsymbol{u}}{ds}. \tag{7.42}$$

The midplane membrane strains ε_{ij} correspond directly to the generalized membrane strains, but the generalized curvature components $\mathfrak{k}_{ij}$ are related to the midplane curvatures κ_{ij} by $\mathfrak{k}_{ij} = \kappa_{ij} A$, where A is the element area. It follows that the sensitivity of the curvature components is given by:

$$\frac{d\kappa_{ij}}{ds} = \frac{1}{A}\frac{d\mathfrak{k}_{ij}}{ds} - \frac{\mathfrak{k}_{ij}}{A^2}\frac{dA}{ds}. \tag{7.43}$$

Terms $d\mathbf{D}/ds$ and dA/ds are evaluated using forward finite difference. Full expressions for the partial derivatives of ε_e with respect to the strain components and α required in Equation 7.40 are not all shown here, but follow directly from Equation 7.38, for example:

$$\frac{\partial \varepsilon_e}{\partial \varepsilon_{xx}} = \frac{2}{9\varepsilon_e}\left(2A_1(\varepsilon_{xx} + z\kappa_{xx}) + A_2(\varepsilon_{yy} + z\kappa_{yy})\right), \text{ etc.} \tag{7.44}$$

The sensitivity of the shear modulus dG/ds, also needed in Equation 7.40, is computed similarly to dG/dT in the previous section, using implicit differentiation of Equation 7.30:

$$\frac{dZ}{ds} = \frac{\partial Z}{\partial s} + \frac{\partial Z}{\partial G}\frac{dG}{ds} = 0 \quad \Rightarrow \quad \frac{dG}{ds} = -\frac{\frac{\partial Z}{\partial s}}{\frac{\partial Z}{\partial G}}, \tag{7.45}$$

where

$$\frac{\partial Z}{\partial G} = -\frac{D_i}{\varepsilon_e^2}\frac{\partial \varepsilon_e}{\partial \alpha}\frac{\partial \alpha}{\partial G} - 1, \tag{7.46}$$

and

$$\begin{aligned}\frac{\partial Z}{\partial s} =& \frac{\partial C_i}{\partial T}\frac{dT}{ds} + \frac{\partial C_i}{\partial s} + \left(\frac{\partial D_i}{\partial T}\frac{dT}{ds} + \frac{\partial D_i}{\partial s}\right)\frac{1}{\varepsilon_e} - \ldots \\ & \frac{D_i}{\varepsilon_e^2}\left(\frac{\partial \varepsilon_e}{\partial \boldsymbol{\varepsilon}}^T \frac{d\boldsymbol{\varepsilon}}{ds} + \frac{\partial \varepsilon_e}{\partial z}\frac{dz}{ds} + \frac{\partial \varepsilon_e}{\partial \alpha}\frac{\partial \alpha}{\partial s}\right).\end{aligned} \tag{7.47}$$

Here T again denotes the temperature in the integration point, which for the present shell element equals the averaged element temperature. The terms $\partial C_i/\partial s$, $\partial D_i/\partial s$, dz/ds and $\partial \alpha/\partial s$ are all evaluated by forward finite difference. The effective strain is non-differentiable at $\varepsilon_e = 0$, and for this situation the average of the directional sensitivities is used, which results in a value of zero.

7.4 Evaluation

7.4.1 Numerical results

The various sensitivity analysis procedures discussed in this chapter for the electro-thermo-mechanical SMA finite element model are applied to the SMA gripper structure used in Chapter 5. This model represents the most general and complete case that can be considered with the modeling approach presented in this research, and therefore serves as a representative problem. The configuration considered is the deterministic optimal design, and sensitivities computed using global finite difference (GFD), direct semi-analytical (SA) and direct refined semi-analytical (RSA, (De Boer and Van Keulen, 2000b, Van Keulen and De Boer, 1998b)) methods are compared. In the mechanical finite difference case, the perturbed analysis is started from the final nominal configuration, which raises the efficiency of this method significantly. The correction for finite difference sensitivities of iteratively solved problems suggested by Haftka (1985) is used to obtain accurate results. The response quantities and design variables that are considered are listed in Table 7.1 and Table 7.2, respectively. The first 6 design variables control the gripper shape and the applied voltage, the latter 6 in fact are uncertainty variables that are not directly controlled by the designer. Still, in an optimization procedure, sensitivities of both types of variables are needed, and in this section no further distinction is made. Nominal values are used for the uncertainty variables.

Design sensitivities have been evaluated using a number of relative design perturbations, ranging from 10^{-14} to 10^{-2}, to compare the stability and accuracy of the results. This is more informative and relevant than the numerical values of the sensitivities by themselves. A large quantity of sensitivity data has been generated, therefore only a selection of the most interesting results is discussed here. These results of interest are shown in Figure 7.2, Figure 7.3 and Figure 7.4.

Table 7.1: *Response quantities considered in the numerical study of SMA design sensitivity analysis.*

Response quantity name	Symbol	Nominal value
Tip displacement [mm]	u_z	0.4341
Aggregated maximum effective strain constraint, open configuration	$KS_A(g_\varepsilon^{open})$	0.9992
Aggregated maximum effective strain constraint, closed configuration	$KS_A(g_\varepsilon^{closed})$	0.9999
Aggregated maximum temperature constraint, open configuration	$KS_A(g_T^{open})$	0.9999
Aggregated maximum temperature constraint, closed configuration	$KS_A(g_T^{closed})$	0.9979

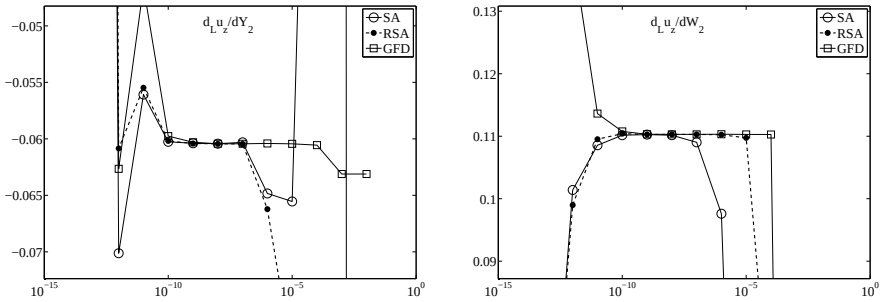


Figure 7.2: *Logarithmic sensitivities $d_L u_z/dY_2$ (left) and $d_L u_z/dW_2$ (right), computed using various methods, as a function of the relative design perturbation.*

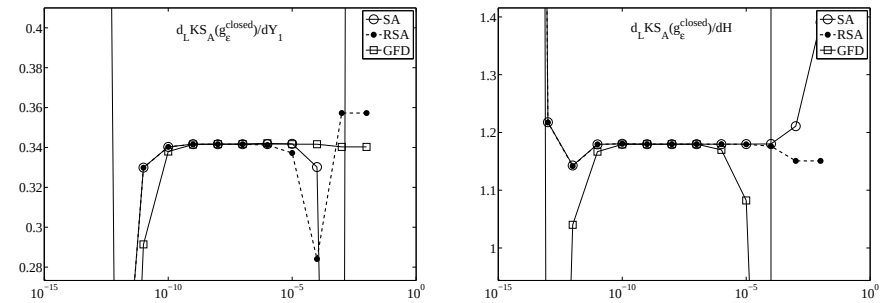


Figure 7.3: *Logarithmic sensitivities $d_L KS_A(g_\varepsilon^{closed})/dY_1$ (left) and $d_L KS_A(g_\varepsilon^{closed})/dH$ (right), computed using various methods, as a function of the relative design perturbation.*

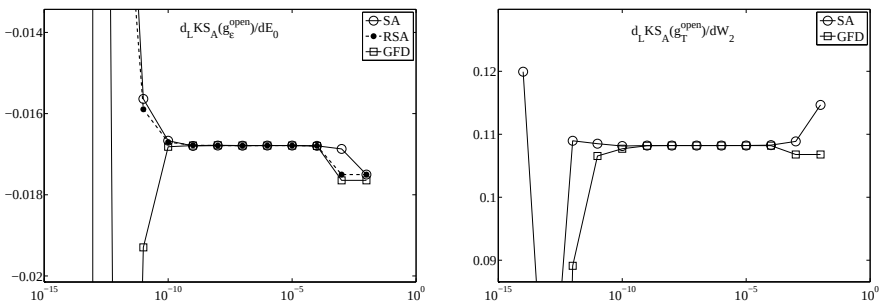


Figure 7.4: *Logarithmic sensitivities $d_L KS_A(g_\varepsilon^{open})/dE_0$ (left) and $d_L KS_A(g_T^{open})/dW_2$ (right), computed using various methods, as a function of the relative design perturbation.*

Table 7.2: *Design (and uncertainty) variables considered in the numerical study of SMA design sensitivity analysis.*

Variable name	Symbol	Nominal value
Plate thickness	t	0.0974 mm
Gripper arm height	H	1.1703 mm
Gripper shape control point 1	Y_1	1.2485 mm
Gripper shape control point 2	Y_2	1.1125 mm
Plate front width	W_2	1.4001 mm
Applied voltage	V	0.0603 V
Face convection coefficient	h	2 $\mathrm{kWm^{-2}K^{-1}}$
Ambient temperature	T_a	328 K
Austenite Young's modulus	E_A	68.939 GPa
R-phase apparent Young's modulus	E_R	45.612 GPa
Initial apparent transition Young's modulus	E_0	20.006 GPa
Poisson ratio	ν	0.333333

7.4.2 Discussion

In all results, the influence of the relative design perturbation can clearly be seen. For both small and large values of the perturbation, the accuracy of the sensitivities deteriorates. This is in agreement with the expectations discussed earlier: reducing the perturbation increases the round-off error, while increasing the perturbation also increases the error due to the truncation of the approximating Taylor series used in the finite difference formulation.

Regarding the semi-analytical approaches, the range of perturbations that yield stable results varies depending on the response quantity and variable considered. Generally, it is observed that this range is smaller for sensitivities of mechanical responses (u_Z, ε_e-constraints) with respect to shape design variables. Typical examples are shown in Figure 7.2 and Figure 7.3, and typically accurate sensitivities are only obtained for relative perturbations in the order of 10^{-9} up to 10^{-7}. As sensitivities become smaller, this range even reduces to 10^{-9}–10^{-8} (Figure 7.2). In case of either non-shape variables or non-mechanical responses, the relative perturbation range for stable sensitivities is considerably larger, approximately 10^{-9} up to 10^{-4}. A possible explanation for this observation is that the nonlinearity of the relationship between the pseudo-load and the shape design variables is considerably stronger in comparison to the other cases. This is indicated by the fact that the truncation error already becomes evident at relatively small perturbations of approximately 10^{-7}. No indications have been found that the (electro-)thermo-mechanical coupling adds to the inaccuracy, as voltage sensitivity of mechanical quantities turned out to be fairly stable, comparable to results shown in Figure 7.4.

Computation of accurate shape design sensitivities, particularly in case of slender structures, is known to be challenging, as discussed in Section 7.2.3. The refined semi-analytical (RSA) method has been proposed to reduce errors in pseudo-load vectors, which increased with increasing rigid body motions of elements. In the present numerical example, it is not clearly observed that this RSA method improves the accuracy of the sensitivity computation. In one case, concerning $d_L u_z/dW_2$ in Figure 7.2, the RSA shows clearly a larger range of stable perturbations. However, in several other cases there is no improvement over the conventional semi-analytical (SA) method. In fact, some cases (e.g. $d_L u_z/dY_2$ in Figure 7.2) show a slight decrease in accuracy when applying the RSA method. This could be attributed to the increased number of numerical operations that have to be performed, compared to the SA case. Apparently, the plausible source of the inaccuracy in the present case, the strong nonlinearity of the problem and hence its susceptibility to truncation errors, is not affected by the RSA method. Possibly for cases in which elements are subjected to larger rigid body motions, results might be different.

The finite difference approach, due to the use of the nominal configuration as a starting point for the perturbed analyses in the mechanical case, proves to be a competitive approach for the present problem. In most cases only one or two iterations are necessary to converge to the perturbed solution. Given the fact that the mechanical analysis itself requires roughly 100 iterations, this is a relatively low computational cost, even when it has to be repeated for every of the 12 considered design variables. For the linear electrical and thermal analysis, the GFD approach in contrast is relatively costly, since for each design variable a complete additional analysis is required. However, compared to the mechanical problem, the computational effort spent on the linear electrical and thermal problem is hardly significant (<1%). Therefore, considering the SMA analysis as a whole, the computational cost of GFD-based sensitivity analysis for the present number of variables is still relatively modest.

The range of relative design perturbations that yield accurate results for the GFD method is generally comparable to that of both semi-analytical cases, although differences can be found in either direction. For example, for the sensitivities in Figure 7.2, the GFD method shows the most stable behavior, whereas for $d_L KS_A(g_\varepsilon^{closed})/dH$ in Figure 7.3, the opposite is true. The smallest stable perturbation is approximately equal to that of the (R)SA cases or slightly larger. The latter might be due to the finite accuracy of the analysis, as a tolerance is used to terminate the Newton iterations. Apart from this, no indication was found that the accuracy of the GFD approach suffered from numerical noise of the model, although it cannot be ruled out that the present case just represents a favorable design point. The largest stable perturbation is in this case influenced by the truncation error in the Taylor approximation of the response quantity, instead of that of the pseudo-load terms. This explains the observed differences, but the fact remains that *a priori* selection of an appropriate relative perturbation is difficult.

Table 7.3: *Time consumption of sensitivity analysis of an SMA miniature gripper shell finite element model, with 12 design variables.*

Method	No sensitivity analysis	SA	RSA	GFD
Total time [s]	226	476	521	552
Index w.r.t. SA [%]	-	100	109	116
Sensitivity analysis time [s]	-	250	295	326
Time per variable [s]	-	20.8	24.6	27.2

Finally, discussion of the computational cost of sensitivity analysis is also of interest. Table 7.3 shows the time consumption of the various sensitivity analysis methods considered in this chapter. It turns out, that the SA method is computationally the most efficient, followed by the RSA method. The least efficient method is the restarted GFD approach, but to put this in perspective, it turns out that for the considered case with twelve design variables, this method is only 16% slower than the best-performing SA method. Given the considerable difference in implementation effort, this makes that for many practical cases the GFD approach would be preferred. Note also that per design variable, the cost of sensitivity analysis roughly corresponds to 10% of the cost of a single analysis.

7.5 Conclusions

In this chapter, the sensitivity analysis of SMA structures actuated by Joule heating has been considered. The behavior of these structures is described by a nonlinear path-independent model, which allows for sensitivity analysis procedures that require significantly less computational effort than the analysis itself. A restarted global finite difference (GFD) approach as well as direct semi-analytical (SA) and refined semi-analytical (RSA) procedures have been studied in this context. As the number of design variables in the intended SMA shape optimization applications is generally modest, an adjoint formulation has not been considered in this chapter.

Numerical testing on an SMA miniature gripper model has revealed that particularly for shape design sensitivities of mechanical response quantities, the selection of a proper relative design perturbation is critical. By testing a range of perturbations, it was found that in certain cases only a small interval exists for which accurate sensitivities are obtained. The semi-analytical approaches appeared to be slightly more sensitive to the perturbation, compared to the GFD method. The reported improved accuracy of the RSA method over the SA formulation was not observed in all cases, probably because the dominant errors in the present case are not related to large rigid body motions, but rather to the nonlinearity of the SMA model.

Generally, one or two iterations proved sufficient to obtain a perturbed solution in the GFD case, starting from the nominal final configuration. For the linear electrical and thermal analyses, the superior efficiency of the semi-analytical approach is undisputed. However, considering the entire SMA analysis as a whole, the GFD method turns out to be quite competitive for the considered class of problems, due to the relatively high cost of the incremental-iterative scheme used to solve the nonlinear mechanical problem. For the representative miniature gripper example, the GFD approach proved to be only 16% slower than the SA case, for 12 design variables. This makes it an attractive option for sensitivity analysis, particularly when its comparatively straightforward and generic implementation is taken into account.

In contrast, the complexity of the computation of the pseudo-load in the semi-analytical approaches was found to increase significantly due to the fact that the problem involves a coupled electrical, thermal and mechanical analysis. Particularly the term associated with the thermo-mechanical coupling required additional implementation effort. A fully analytical as well as a finite-difference-based formulation has been implemented, and it was found that both procedures perform well. In spite of the fact that an iterative process is used inside the material model, inaccuracies due to numerical noise have not been detected in case of the finite-difference approach, and results obtained with both methods were in good agreement.

An additional consideration is the sensitivity analysis of derived quantities. An important response quantity in the used SMA model is the maximum effective strain. In case of the finite difference method, evaluation of its sensitivity is straightforward, as the finite difference formula can simply be applied, using the computed nominal and perturbed responses. In contrast, because of the complex SMA material model, in the semi-analytical cases a rather lengthy and complex procedure is required to compute the sensitivity of the effective strain from the state variable sensitivities.

In conclusion, it can be stated that for the present class of SMA problems, given a suitable ratio of design variables versus response quantities, the restarted finite difference approach is a viable option for sensitivity analysis, particularly considering the significantly smaller implementation effort. However, for larger problems with considerably more degrees of freedom, the advantage of semi-analytical methods will increase, as the cost of system matrix decomposition will increase sharply. Also, when the number of design variables increases, the GFD approach will become increasingly unattractive, as the computational cost per design variable is higher than in the (R)SA methods. For large numbers of design variables, an adjoint semi-analytical approach is likely to be the most efficient. The developed sensitivity analysis methods are expected to be of great use in further work on design optimization of SMA structures.

Based on: Langelaar, M. and van Keulen, F. (2006). Sensitivity analysis and optimization of a shape memory alloy gripper. *Computers and Structures*, in review.

Chapter 8

Gradient-based Shape Optimization of an SMA Gripper

Optimization techniques can be used to improve the design process of shape memory alloy (SMA) structures. This chapter presents the shape optimization of a miniature SMA gripper using two gradient-based optimization algorithms, Sequential Quadratic Programming (SQP) and the Method of Moving Asymptotes (MMA). The use of gradient information enables faster optimization or allows more design variables, in comparison to direct methods. To obtain gradient information, the sensitivity analysis of the sequentially coupled electro-thermo-mechanical SMA gripper problem is investigated. Finite difference and semi-analytical sensitivity analysis techniques for this problem are compared on accuracy, stability, efficiency and implementation aspects. Furthermore, using the SMA gripper as a representative example, the effectiveness and computational efficiency of the gradient-based optimization methods is compared to a direct response-surface-based approach.

8.1 Introduction

Shape memory alloys (SMAs) are interesting materials for many (micro-)actuation applications, because of their unusually high energy density (Otsuka and Wayman, 1998). SMAs can generate relatively large actuation strains and stresses, controlled by modest changes in temperature. Often SMAs are applied in the form of wires or springs, but in many emerging applications in, e.g., microsystems or medical instruments, more complex structures are considered. However, designing effective planar or three-dimensional SMA actuators is no easy task, due to the

complexity of the thermo-mechanical behavior. Moreover, often the temperature of SMA actuators is controlled by direct resistive heating of the material itself, which makes that its geometry as well as its electrical, thermal and mechanical properties are all interacting in a complicated way. For this reason, efforts are made to develop systematic model-based design procedures for SMA structures. Particularly design optimization techniques are promising, as these methods offer an approach to handle design problems in a structured and formal way (Haftka and Gürdal, 1992).

Several examples of the application of design optimization concepts to SMA design problems can be found in recent literature. Skrobanek *et al.* (1997) have reported the use of a heuristic peak stress reduction algorithm in the design of SMA microvalves, which can be viewed as a kind of shape optimization. The same procedure has been used by Fischer *et al.* (1999) to reduce peak stresses in the design of a flexible SMA endoscope tip. Finite element modeling is used, so this method can be applied to a wide variety of structures. However, this approach does not involve a formal optimization problem formulation. This significantly limits its applicability, as constraints and objectives regarding other aspects than peak stress reduction cannot be included.

Lu *et al.* (2001) applied a combination of graphical optimization and sequential quadratic programming (SQP) to an SMA design problem, using an elaborate analytical model of an SMA actuator with a corrugated core. Unfortunately, analytical modeling is not practical for more general SMA structures. Recently, Dumont and Kuhl (2005) reported the optimization of an SMA spring actuator using a genetic algorithm. Genetic algorithms are stochastic optimization procedures that are typically quite robust, but also very costly in terms of the number of required evaluations. Therefore, this approach is clearly limited in terms of the complexity of the problem and the number of design variables that can be considered.

In contrast, this chapter presents gradient-based design optimization of a shape memory alloy (SMA) structure, a miniature gripper. In an earlier article, we have reported the shape optimization of a similar SMA gripper using a direct response surface method (Langelaar and Van Keulen, 2004a) (Chapter 4). However, gradient-based methods are potentially superior to direct approaches in terms of efficiency, in case the design sensitivities of the problem can be computed at relatively low cost. This implies that solutions can be obtained faster and with less computational effort, or that more design variables can be included. The presented approach is not limited to design of grippers, since finite element modeling is employed, which can be used to model virtually any kind of SMA structure. To allow for efficient sensitivity analysis, the finite element model uses a dedicated history-independent SMA constitutive model.

This article is organized as follows: first, the SMA gripper design concept and modeling is presented in Section 8.2, together with the formulation of the design optimization problem. Subsequently, Section 8.3 discusses the sensitivity analysis of this gripper, using finite difference and semi-analytical methods. The

shape optimization results for the SMA gripper are presented in Section 8.4. The performance of the previously used response surface method (Multi-point Approximation Method (Toropov *et al.*, 1993a)) is compared to two gradient-based methods, sequential quadratic programming (SQP) and Svanberg's Method of Moving Asymptotes (Svanberg, 1987), followed by conclusions.

8.2 SMA miniature gripper

8.2.1 Concept and modeling

The design concept for the miniature SMA gripper used in this chapter is shown in Figure 8.1. The gripper consists of an identical top and bottom arm made of folded Ni-Ti plates. The gripper is intended to exploit the shape memory behavior due to the R-phase transformation in Ni-Ti, as described in an experimental study by Tobushi *et al.* (1992). This transformation has a very small hysteresis, and allows this gripper to operate in a relatively narrow temperature window. An initial deformation is applied in order to generate internal stresses in the material, which are required to make use of the shape memory effect. Starting from the undeformed configuration in Figure 8.1, the ends of the outer plates are pinched toward the inner plates during assembly. In this situation, the equilibrium configuration of each arm can be changed by changing the temperature of either the inner or outer plates. Resistive heating is used for this purpose, and to guide the electrical current through individual plates a slit is present along the length of each plate. Heating the inner plates will cause the tip ends to move apart, opening the gripper. Similarly, heating the outer plates will make them move towards each other, closing the gripper. In the closed configuration, clamping forces of 100 mN are applied in z-direction at the tips of the gripper.

Because of symmetry, only a quarter of the gripper needs to be modeled: in this case half the top arm is used. This part together with the parameterization of the geometry is shown in Figure 8.2. Seven design variables have been selected for this design problem: the plate thickness t, the undeformed arm height H, the actuation plate front and end width W_1 and W_2, and the shape of the actuation plate. This shape is described by a quadratic B-spline (see Farin, 2002), and the y-coordinates of the two middle control points are used as design variables: Y_1 and Y_2. Finally, the voltage V to operate the gripper is used as the seventh design variable. Further geometrical details of the miniature gripper are listed in Table 8.1.

The gripper is simulated by finite element analysis of the parameterized design shown in Figure 8.2. For both the opened and closed case, a quasi-static sequentially coupled electrical, thermal and mechanical analysis is performed, to simulate the SMA behavior under the influence of Joule heating. Dissipated heat from the electrical analysis is used as a heat source in the thermal analysis, and the resulting temperature distribution is used in the mechanical analysis. Physical

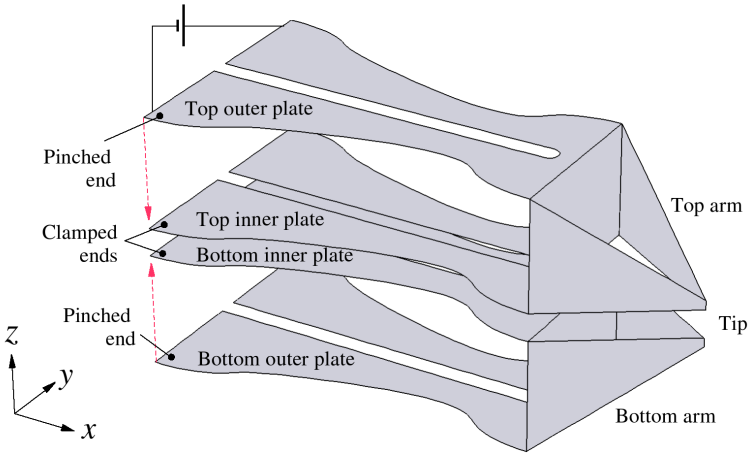


Figure 8.1: *Conceptual gripper geometry in the undeformed configuration.*

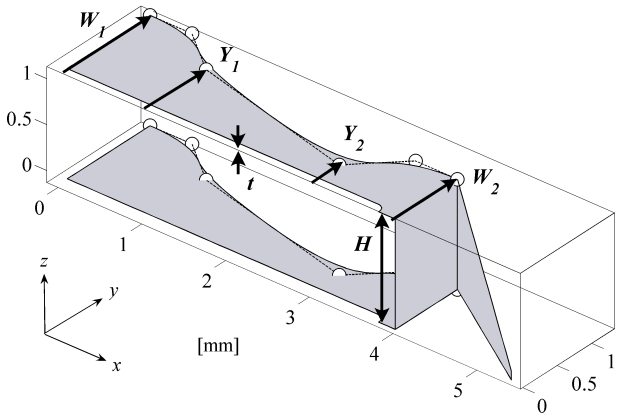


Figure 8.2: *Design parameterization of the gripper. Because of symmetry, only a quarter is considered.*

Table 8.1: *Significant coordinates of B-spline control points and other points defining the geometry of the miniature gripper.*

Point	x [mm]	y [mm]	Point	x [mm]	y [mm]
Control point 1	0	W_1	Control point 2	0.5	W_1
Control point 3	1	Y_1	Control point 4	3	Y_2
Control point 5	3.5	W_2	Control point 6	4	W_2
Tip	5.377	0	Slit end	3.8	0

Table 8.2: *Physical constants used in the finite element modeling.*

Quantity	Value
Electrical conductivity	$1.25 \cdot 10^6$ Sm^{-1}
Thermal conductivity	21 $\mathrm{Wm}^{-1}\mathrm{K}^{-1}$
Thermal convection coefficient	$2.0 \cdot 10^3$ $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$
Ambient temperature	328 K

constants used in the simulations are listed in Table 8.2.

Particularly the mechanical analysis is computationally intensive, because of the nonlinear SMA material model as well as the consideration of geometrical nonlinearities. A dedicated SMA material model is employed (Langelaar and Van Keulen, 2004b) (Chapter 3), aimed at the R-phase transformation in Ni-Ti, which is used in this gripper. In the considered temperature range (328–343 K), hysteresis is practically absent, which allows simplifications in the constitutive modeling. Neglecting hysteresis furthermore renders the model history-independent, which simplifies the sensitivity analysis considerably (Kleiber *et al.*, 1997). Triangular shell elements are used for the mechanical analysis (Van Keulen and Booij, 1996), and the nonlinear SMA constitutive model is included using numerical integration over the thickness of the shells. To ensure robust convergence for all combinations of design variables, an adaptive incremental-iterative scheme is used.

8.2.2 Optimization problem

The aim of the design problem is to maximize the displacement range of the tips of this miniature gripper, so it can be used for gripping and clamping a wide range of objects. This translates into maximizing the difference between the gripper tip z-positions in the opened and closed configurations. Furthermore, constraints are added that restrict both the strain and the temperature of the structure. Strain constraints are necessary to remain within the range of applicability of the material model, which has been based on experimental data for strains up to 1%. Temperature constraints are added to limit the thermal operating range of the

gripper, which is an important requirement in, e.g., biomedical applications. In this way, this gripper design problem serves as a realistic example of SMA shape optimization.

It is not practical to include these strain and temperature constraints for every single finite element in the mesh, as the number of constraints would become prohibitively large. Furthermore, in a practical setting, limited and small constraint violations are tolerable, and strictly constraining every element individually would be overly stringent. Therefore, in this chapter, the individual element constraints are combined into global constraints using the Kreisselmeier-Steinhauser constraint aggregation function (Kreisselmeier and Steinhauser, 1983). A slightly adapted version is used here, that also accounts for the area A of each element. In this way, small local violations are tolerated to some extent, but still the constraint restricts the design from becoming unacceptable in a global sense. The design optimization problem can now formally be written as:

$$\begin{aligned}
&\max_{\mathbf{x}} f(\mathbf{x}) = 2\left(u_z^{\text{tip, open}}(\mathbf{x}) - u_z^{\text{tip, closed}}(\mathbf{x})\right)\\
&\text{Subject to:}\\
&\quad g_1(\mathbf{x}) = KS_A(g_\varepsilon^{\text{open}}(\mathbf{x})) - 1 \leq 0,\\
&\quad g_2(\mathbf{x}) = KS_A(g_\varepsilon^{\text{closed}}(\mathbf{x})) - 1 \leq 0,\\
&\quad g_3(\mathbf{x}) = KS_A(g_T^{\text{open}}(\mathbf{x})) - 1 \leq 0,\\
&\quad g_4(\mathbf{x}) = KS_A(g_T^{\text{closed}}(\mathbf{x})) - 1 \leq 0,\\
&\quad \underline{\mathbf{x}} \leq \mathbf{x} \leq \overline{\mathbf{x}}.
\end{aligned} \tag{8.1}$$

The vector $\mathbf{x}$ represents the design variables, and $\underline{\mathbf{x}}$ and $\overline{\mathbf{x}}$ contain their lower and upper bounds, respectively. The individual variables and their bounds are listed in Table 8.3. The aggregated strain constraints are given by g_1 and g_2, one for each configuration. Likewise g_3 and g_4 represent the temperature constraints of the open and closed gripper. The individual element constraints and the aggregation function are given by:

$$g_\varepsilon^{(i)} = \frac{\varepsilon_e^{(i)}}{\varepsilon_e^{(max)}} \leq 1, \tag{8.2}$$

$$g_T^{(i)} = \frac{T^{(i)} - T_{min}}{T_{max} - T_{min}} \leq 1, \tag{8.3}$$

$$KS_A(\boldsymbol{g}) = \frac{1}{\rho} \ln\left(\frac{1}{\sum_{i=1}^{N} A^{(i)}} \cdot \sum_{i=1}^{N} A^{(i)} e^{\rho g^{(i)}}\right). \tag{8.4}$$

For this gripper, $T_{min} = 328$ K and $T_{max} = 338$ K have been used. The maximum effective strain $\varepsilon_e^{(max)}$ is set to 0.01. The definition of the effective strain ε_e is given in Chapter 3.

To get a good impression of the various optimization methods studied in this chapter, each method will be applied to this problem starting from three different initial designs, as listed in Table 8.4. All these designs are infeasible starting

Table 8.3: *Design variables* ($\mathbf{x}$) *for the design optimization of the SMA gripper.*

Design variable	Symbol	Lower bound	Upper bound	Unit
Plate thickness	t	0.05	0.3	mm
Gripper arm height	H	0.3	2	mm
Plate shape control point 1	Y_1	0.01	1.5	mm
Plate shape control point 2	Y_2	0.01	1.5	mm
Plate front width	W_1	0.1	1.5	mm
Plate end width	W_2	0.1	1.5	mm
Applied voltage	V	0.01	0.5	V

Table 8.4: *Starting points used for the design optimization of the SMA gripper.*

Case	t [mm]	H [mm]	Y_1 [mm]	Y_2 [mm]	W_1 [mm]	W_2 [mm]	V [V]
A	0.08464	0.452	1.018	0.965	1.390	1.379	0.06411
B	0.09645	1.415	0.816	0.912	1.151	0.590	0.05032
C	0.20712	0.459	1.307	0.661	0.714	1.212	0.06810

points, and they have a relatively poor objective function value. In fact, these points were found as sub-optimal points in the first iteration of three different runs of the multi-point approximation method, and they have subsequently been used as starting points for the SQP and MMA studies.

Before proceeding with the discussion of the results of the design optimization, first the sensitivity analysis of this problem is treated in the following section.

8.3 Sensitivity analysis

8.3.1 Methods

In order to be able to use the efficient gradient-based optimization methods, it is necessary to compute the derivatives of the responses (objective, constraints) with respect to the design variables, i.e. the design sensitivities. For the present sequentially coupled gripper model involving a complex nonlinear SMA material model, the sensitivity analysis is not trivial, and therefore some selected aspects are highlighted in this section. An in-depth discussion of the sensitivity analysis for this type of problems is given in Chapter 7.

Various methods exist for evaluating design sensitivities, and extensive discussions can be found in, e.g., books by Haftka and Gürdal (1992), Kleiber *et al.* (1997) and the recent review by Van Keulen *et al.* (2005). Considering the fact that the SMA gripper is modeled with a history-independent nonlinear model,

and that the number of design variables is not far greater than the number of responses, appropriate methods for sensitivity analysis are either the restarted Global Finite Difference (GFD) approach or the direct Semi-Analytical approach (SA). As the model involves shell elements, for which the SA method is known to be unreliable for large rotations, an improved SA formulation such as the Refined Semi-Analytical (RSA) method (Van Keulen and De Boer, 1998b) should also be considered.

The GFD method used here is based on forward finite difference, i.e. the design sensitivities of a response f with respect to a design variable s are approximated by:

$$\frac{df}{ds} \approx \frac{f(s+\Delta s) - f(s)}{\Delta s}, \tag{8.5}$$

where Δs is a small design perturbation. Since the present problem is history-independent, the evaluation of the mechanical problem can be carried out by restarting iterations from the final unperturbed solution, thereby saving a costly incremental-iterative solution process. This reduces the computational cost significantly, and in combination with its relatively straightforward implementation, GFD sensitivity analysis becomes an attractive option. The residual correction proposed by Haftka (1985) has been applied to improve the accuracy of the computed sensitivities.

The SA and RSA approaches, in contrast, utilize analytical differentiation of the discretized equations. For example, for a typical linear finite element problem given by $\mathbf{KU} = \mathbf{F}$, design differentiation results in:

$$\frac{d\mathbf{K}}{ds}\mathbf{U} + \mathbf{K}\frac{d\mathbf{U}}{ds} = \frac{d\mathbf{F}}{ds} \;\Rightarrow\; \frac{d\mathbf{U}}{ds} = \mathbf{K}^{-1}\left(\frac{d\mathbf{F}}{ds} - \frac{d\mathbf{K}}{ds}\mathbf{U}\right). \tag{8.6}$$

At this point however, the quantities $d\mathbf{F}/ds$ and $d\mathbf{K}/ds$ are evaluated using finite difference, similar to Equation 8.5. Note that the expression for $d\mathbf{U}/ds$ involves the inverse of the system matrix $\mathbf{K}^{-1}$, which can be reused when a direct solver is used. This makes that sensitivities can be obtained at a fraction of the cost of a full analysis. In case of RSA, inaccuracies in the finite difference approximation of $d\mathbf{K}/ds$ that occur in mechanical problems with large rotations are reduced by exact differentiation of rigid body modes (Van Keulen and De Boer, 1998b). The following subsections focus particularly on the non-standard aspects of the (R)SA sensitivity analysis for the present problem.

8.3.2 Handling of interdisciplinary coupling

The analysis of the SMA gripper involves a sequentially coupled electrical, thermal and mechanical problem. This implies that in the sensitivity analysis, as in the analysis itself, the coupling between the disciplinary models has to be accounted for. The first coupling is found between the electrical and thermal problem, where the heat generated by the electrical current serves as a body source term for the

thermal simulation. In the SA setting, this results in an electrical contribution to the load design derivative $d\mathbf{F}/ds$ in Equation 8.6, that accounts for the effect of a changing voltage on the thermal load. This contribution could be evaluated by finite differences, but this would require an additional perturbed electrical simulation for every design variable. In the SA context, the aim is to avoid such expensive solutions, and therefore an analytical approach is considered. The dissipated heat q on element level is given by:

$$q = \boldsymbol{u}_e^T \mathbf{K}_e \boldsymbol{u}_e. \tag{8.7}$$

Here $\boldsymbol{u}_e$ represents the electrical state variables (i.e. the electrical potential) and $\mathbf{K}_e$ the element conductivity matrix. Design differentiation of q subsequently results in:

$$\frac{dq}{ds} = 2\boldsymbol{u}_e^T \mathbf{K}_e \frac{d\boldsymbol{u}_e}{ds} + \boldsymbol{u}_e^T \frac{d\mathbf{K}_e}{ds} \boldsymbol{u}_e. \tag{8.8}$$

The design derivative of the electrical tangent matrix $\mathbf{K}_e$ can again be approximated by finite differences. This term is also used in the SA sensitivity analysis for the electrical case, and can therefore be reused. The procedure has been verified by comparing the results to finite difference computations.

The second coupling in the present SMA actuator problem exists between the thermal and mechanical problem, where the local temperature distribution affects the mechanical material behavior. The discrete governing equations for the nonlinear mechanical problem are given by:

$$\boldsymbol{R} = \boldsymbol{f}_{ext}(s) - \boldsymbol{f}_{int}(\boldsymbol{u}_m(s), \boldsymbol{u}_t(s), s) = \mathbf{0}. \tag{8.9}$$

Here $\boldsymbol{R}$ is the global unbalance or residual load vector, which is iteratively reduced to zero in the solution process. The terms $\boldsymbol{f}_{ext}$ and $\boldsymbol{f}_{int}$ represent the external and internal load vectors, respectively, and $\boldsymbol{u}_m$ and $\boldsymbol{u}_t$ contain the system displacements and temperatures. The SA approach now yields the following equation for the displacement sensitivity:

$$\frac{d\boldsymbol{u}_m}{ds} = \mathbf{K}_m^{-1} \left(\frac{\partial \boldsymbol{R}}{\partial s} + \frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}_t} \frac{d\boldsymbol{u}_t}{ds} \right). \tag{8.10}$$

Here use is made of the fact that the mechanical tangent matrix $\mathbf{K}_m$ is given by $-\partial \boldsymbol{R}/\partial \boldsymbol{u}_m$. The term $\partial \boldsymbol{R}/\partial s$ is evaluated using finite differences, and the temperature sensitivity $d\boldsymbol{u}_t/ds$ has already been calculated at this point. To evaluate Equation 8.10, still the term $\partial \boldsymbol{R}/\partial \boldsymbol{u}_t$ needs to be computed, which expresses the interdisciplinary coupling between the mechanical and thermal problem. A finite difference approach is possible, however this requires as many perturbations as the number of thermal degrees of freedom. It is more attractive to first further specify this term analytically. From the fact that the external load is independent of temperature, it is found that

$$\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}_t} = -\frac{\partial \boldsymbol{f}_{int}}{\partial \boldsymbol{u}_t}. \tag{8.11}$$

Furthermore, since on element level the internal force is given by:

$$\boldsymbol{f}_{int} = \mathbf{D}^T \boldsymbol{\sigma} \qquad \text{with} \quad D_{ij} = \frac{\partial \varepsilon_i}{\partial u_j}, \tag{8.12}$$

for each individual element, with $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ as the element generalized stress and strain vectors, Equation 8.11 eventually becomes:

$$\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{u}_t} = -\mathbf{D}^T \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{u}_t}. \tag{8.13}$$

Hence, by evaluating the term $\partial \boldsymbol{\sigma} / \partial \boldsymbol{u}_t$, the coupling term can be obtained. As only local temperatures affect element stress values, using this element-level formulation the cost of computing this term using finite differences is comparable to merely a single stress update. However, a fully analytical treatment is possible as well, by differentiation of the SMA constitutive model. A specific challenge in this approach is considered in the following subsection.

8.3.3 Handling of the implicit SMA material model

When considering differentiation of the SMA constitutive relations used in the present model, a challenge is encountered in the fact that the SMA material model itself is not formulated in a fully explicit way. This differentiation is necessary in for example the discussed thermo-mechanical sensitivity coupling term, or the sensitivities of derived quantities such as stresses or effective strains. Without addressing the full details of its formulation (these can be found elsewhere (Langelaar and Van Keulen, 2004b, 2006) (Chapter 3 and Chapter 6)), by discussing the structure of the equations this subsection aims to make clear how this problem is resolved.

The SMA stress-strain relation used in the present shell elements takes the following form:

$$\boldsymbol{\sigma} = \mathbf{C}(G(T, \boldsymbol{\varepsilon}))\boldsymbol{\varepsilon}. \tag{8.14}$$

Here $\mathbf{C}$ represents the constitutive tangent matrix, that relates the generalized stresses $\boldsymbol{\sigma}$ to the generalized strains $\boldsymbol{\varepsilon}$. In this SMA model, $\mathbf{C}$ is not constant, but changes with the value of the effective shear modulus G. This shear modulus in turn is a function of the temperature T and the current strains $\boldsymbol{\varepsilon}$. However, the way G depends on T and $\boldsymbol{\varepsilon}$ is not formulated explicitly, but is given by an implicit equation of the following structure:

$$G = C(T) + \frac{D(T)}{f(G, \boldsymbol{\varepsilon})}. \tag{8.15}$$

Here C and D are governed by material properties and vary with temperature, and f is a complex function of G itself as well as the strain vector $\boldsymbol{\varepsilon}$. During

finite element calculations, in order to evaluate the material response using Equation 8.14, the shear modulus G is determined from Equation 8.15 by solving the following nonlinear equation using Newton iterations:

$$Z(G) = C(T) + \frac{D(T)}{f(G, \boldsymbol{\varepsilon})} - G = 0. \tag{8.16}$$

Note that T and $\boldsymbol{\varepsilon}$ are given at this point, and only G is still to be determined.

The fact that G is not formulated as an explicit function leads to a slight complication in the analytical sensitivity analysis involving this material model. For example, in order to analytically evaluate the term $\partial\boldsymbol{\sigma}/\partial\boldsymbol{u}_t$ used in the thermomechanical sensitivity coupling (Equation 8.13), differentiation of Equation 8.14 to T yields:

$$\left.\frac{\partial\boldsymbol{\sigma}}{\partial T}\right|_{\boldsymbol{u}_m=\text{constant}} = \left(\frac{\partial\mathbf{C}}{\partial G}\frac{\partial G}{\partial T}\right)\boldsymbol{\varepsilon}. \tag{8.17}$$

Since G is not explicitly formulated, the term $\partial G/\partial T$ can not be evaluated directly. To deal with this situation, implicit differentiation must be applied to Equation 8.16:

$$Z(T, G(T)) = 0 \;\Rightarrow\; \frac{\partial Z}{\partial T} + \frac{\partial Z}{\partial G}\frac{\partial G}{\partial T} = 0 \;\Rightarrow\; \frac{\partial G}{\partial T} = -\frac{\partial Z}{\partial T}\left(\frac{\partial Z}{\partial G}\right)^{-1}. \tag{8.18}$$

In a similar way, analytical expressions can be derived for sensitivities of quantities derived from the state vector, the displacement vector $\boldsymbol{u}_m$. Further detailed expressions can be found in Chapter 7.

8.3.4 Comparative evaluation of sensitivity analysis techniques

The discussed finite difference, semi-analytical and refined semi-analytical sensitivity analysis methods have been implemented to perform the sensitivity analysis of the considered SMA gripper. In terms of implementation, GFD requires the least effort, followed by SA and finally RSA. Particularly the analytical treatment of the interdisciplinary coupling terms and the SMA material model make the implementation of (R)SA rather involved in this case. However, next to implementation effort it is important to evaluate how these three sensitivity analysis methods perform in terms of accuracy and efficiency.

All methods utilize finite difference approximations at some point, and this involves a relative design perturbation. Accuracy of finite difference derivatives is known to depend on the choice of this relative perturbation, and therefore sensitivities have been computed for relative perturbations ranging from 10^{-14} to 10^{-2}. The design point used for this sensitivity analysis was the optimal design found in a previous study (Chapter 5), given by $\mathbf{x} = (t, H, Y_1, Y_2, W_1, W_2, V)$ =(0.0974, 1.1703, 1.2485, 1.1125, 1.5000, 1.4001, 0.0603). The results of two typical cases are shown in Figure 8.3, which clearly shows that a perturbation

range exists where stable values are obtained, and that both too large and too small perturbations lead to increasing deviations. In general it was found that for mechanical responses the stable range was smaller than for thermal ones, and that shape design variables were more sensitive to the perturbation used. Considering all results, no method was found to consistently outperform the others in terms of accuracy. RSA did not result in clear improvements over SA for the present problem.

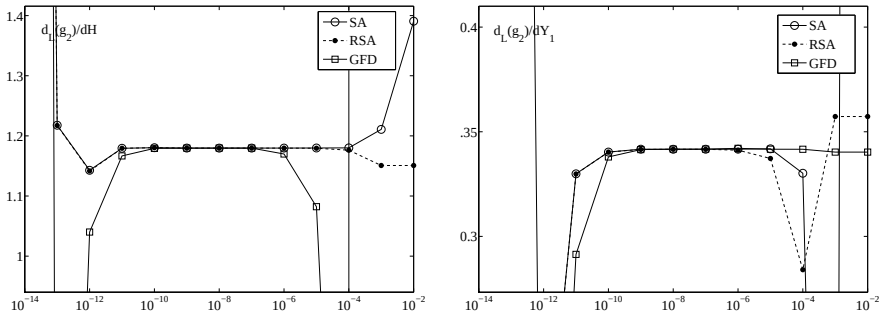


Figure 8.3: *Logarithmic sensitivities $d_L g_2/dH$ (left) and $d_L g_2/dY_1$ (right), computed using various methods, as a function of the relative design perturbation.*

The three sensitivity analysis methods have also been compared in terms of their computational efficiency. For this comparison, the same design was used as for the perturbation study, and a machine equipped with a 2 GHz AMD Athlon processor was used for the evaluation. Table 8.5 lists the resulting time consumption of each method for an analysis of the SMA gripper in both the open and closed case, combined with a full sensitivity analysis of the five responses of the optimization problem given in Equation 8.1. It can be seen that the SA method is the most efficient for the present problem, followed by RSA and GFD. Note however also that in terms of total evaluation time, GFD only requires 12% more time than SA. When also considering the fact that the implementation of the GFD method is far less involved than the SA method for the present problem, the GFD method in fact is an attractive option.

Still, based on its superior efficiency for the present problem, and since it had been implemented already, the semi-analytical method was chosen for use in the gradient-based design optimization of the SMA gripper discussed in Section 8.4. Furthermore, considering the results of the perturbation variation study, a relative perturbation of 10^{-7} was selected.

8.3.5 Impact of numerical noise on design sensitivities

The extensive numerical studies of the sensitivities of this SMA gripper problem computed by the GFD and (R)SA methods have shown that in most cases, the

results are in good agreement. However, due to the numerical noise that can be observed for certain responses, in rare cases the GFD approach sometimes produces inaccurate results. Two interesting examples have been encountered at the (feasible) design point $\mathbf{x}^*$ =(0.1546, 0.8774, 0.8977, 0.5614, 1.2687, 1.1719, 0.036085), for the strain constraint in the open case as a function of variations in H and Y_2, respectively. The behavior of this response is illustrated in Figure 8.4.

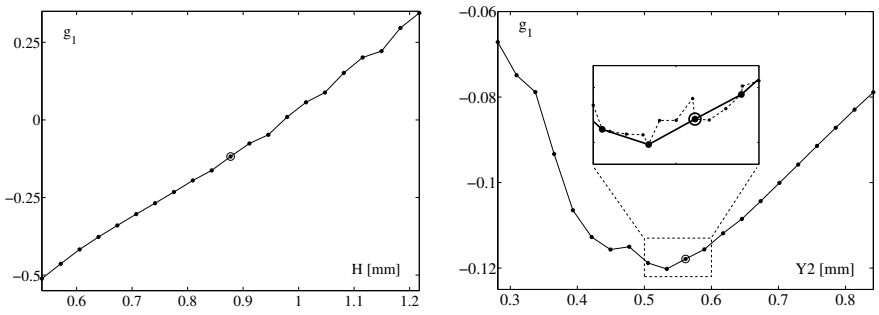


Figure 8.4: *Response data for the strain constraint in the open case (g_1), against variations in H (left) and Y_2 (right, with closeup) at a certain design point $\mathbf{x}^*$ (o).*

The responses in Figure 8.4 show nonsmooth variations, linked to numerical noise in the response evaluation. This noise could be caused by remeshing effects, or by the fact that an iterative solution process is used for solving the nonlinear mechanical problem, which is terminated at a certain finite error. Still, clearly a trendline can be identified from the graphs. Ideally, the gradients used in the optimization process should follow this trendline, but in reality sensitivity information only has a local validity. When a model suffers from noise, it can be debated whether it makes sense to use derivative information and gradient-based optimization methods. Response surface based methods (such as the Multi-point Approximation Method used in this chapter) might be better suited for optimization of such noisy functions, as the approximations tend to dampen the effect of noise.

Table 8.5: *Time consumption of sensitivity analysis of the SMA miniature gripper finite element model, with 7 design variables.*

Method	SA	RSA	GFD	Analysis only
Total time [s]	372	398	416	226
Index w.r.t. SA	100	107	112	
Sensitivity analysis time [s]	146	172	190	

Table 8.6: *Sensitivity values of g_1 at $\mathbf{x}^*$ obtained by various methods, with respect to H and Y_2.*

Sensitivity	$\partial g_1/\partial H$ [mm^{-1}]	$\partial g_1/\partial Y_2$ [mm^{-1}]
From trendline (approx.)	1.211	79.514×10^{-3}
GFD	-29.318	-4.520×10^{-3}
SA	1.213	-4.523×10^{-3}

Table 8.6 illustrates the effect numerical noise can have on the sensitivities. It lists sensitivity values based on the trendlines taken from Figure 8.4, together with results obtained by SA and GFD approaches. In the case of $\partial g_1/\partial H$, the SA result matches the trendline well, but the GFD result is clearly inaccurate, both in sign and in magnitude. Apparently the perturbation used in the GFD computation has been such, that noise has affected the accuracy. The authors want to emphasize that this is the only case of this type of inaccuracy of GFD results that has been encountered so far, in spite of extensive testing. Still, it illustrates the potential susceptibility of GFD sensitivities to large inaccuracy due to numerical noise.

In case of $\partial g_1/\partial Y_2$, SA and GFD results are in good agreement, but fail to capture the slope of the trendline that appears to be present in Figure 8.4, which again clearly differs both in sign and in magnitude. As shown in the inset in Figure 8.4, the response $g_1(Y_2)$ is locally rather noisy, showing fast variations. Apparently, the SA and GFD results are indeed accurate *locally*, but they fail to represent the response behavior on a slightly larger scale. This phenomenon could clearly reduce the effectiveness of gradient-based optimization methods. Again, the authors want to stress that observations of this phenomenon have been very rare for the present problem, but still the presented case illustrates why caution should be taken when using gradient-based algorithms in combination with noisy response functions.

8.4 Gripper design optimization

8.4.1 Optimization methods

The optimization of the design problem defined in Equation 8.1 has been performed by three different methods, one direct and two gradient-based methods. Each of the methods has been tested using the three different starting points given in Table 8.4. The direct method used is the Multi-point Approximation Method (MAM) (Toropov *et al.*, 1993a), which essentially is a combination of design of experiments techniques, response surface building and a move limit strategy. After evaluating various designs at different sampling points, a response surface is con-

structed and optimization is performed within the current trust region using this approximation. The point that is obtained is called the *sub-optimal* point, and this design is subsequently evaluated using the model, in order to judge the quality of the approximation. Depending on the accuracy and several other heuristic indicators, the trust region is adapted and new sampling points are planned for the next iteration.

The strengths of the MAM are its robustness in case of noisy function evaluations, through the use of approximations. Further advantages are its versatility and ease of application, as it allows treating the model as a "black box". However, generally considerable numbers of sampling points and function evaluations are needed, and for nontrivial problems parallel computing is generally required to obtain results within a reasonable timeframe. Furthermore, both as advantage and disadvantage, the MAM involves a large number of optimizer parameter settings that can be tuned to the specific problem. This includes the order and shape of the response surface, parameters affecting the move limit strategy, and parameters related to the design of experiments. In this study, linear response surfaces were used to limit the number of required function evaluations, and a few trial runs have been performed in order to find practical settings for other parameters. Note that more extensive tuning might further improve the performance of this method, but as this requires a significant additional effort, such tuning was not considered fair in comparison to the other methods that hardly require any tuning.

The gradient-based methods that are considered are the Method of Moving Asymptotes (MMA) by Svanberg (1987) and the Sequential Quadratic Programming method (SQP), (see e.g. Fletcher, 1980). MMA approximates the responses by smooth monotonic approximations with asymptotes, which leads to convex approximations. MMA has proven its effectiveness for problems with large numbers of design variables, such as topology optimization problems, and it is interesting to test its performance in the present SMA shape optimization problem. SQP is generally accepted as the best general-purpose gradient-based optimization algorithm, and the specific implementation used here is the one implemented in the Matlab Optimization Toolbox (Version 2.1).

8.4.2 Optimization results

A complete overview of the performance of the three methods and the designs obtained is given in Table 8.7. It can be seen that the designs obtained with the MAM method are consistently inferior to those found with the gradient-based methods. In part, this is due to the convergence settings used. If convergence tolerances would have been tightened, it is possible that better objectives could be obtained with the MAM approach, however at the cost of additional iterations and function evaluations. Furthermore, more extensive and elaborate tuning of the parameters and settings of the MAM method might improve the optimal designs as well, but to allow a fair comparison between the methods in a practical setting,

Table 8.7: *Results overview of the optimization study of the SMA miniature gripper.*

Method	MAM			MMA			SQP		
Case	A	B	C	A	B	C	A	B	C
Objective									
f [mm]	0.8916	0.7975	0.8541	0.9164	0.9166	0.9166	0.9149	0.9137	0.9159
Constraints									
g_1	-0.005	-0.014	-0.010	-0.001	-0.007	-0.000	-0.004	-0.004	0.000
g_2	0.003	-0.002	0.001	-0.041	-0.050	-0.040	-0.044	-0.050	-0.042
g_3	0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	0.001	-0.000
g_4	-0.003	-0.002	-0.003	-0.001	-0.016	-0.002	-0.002	-0.001	-0.002
Variables									
t	0.0957	0.1030	0.0858	0.0959	0.0959	0.0960	0.0958	0.0962	0.0960
H	1.2130	1.0910	1.2550	1.3023	1.3010	1.3020	1.3030	1.3250	1.3120
Y_1	1.4970	1.1090	1.3530	1.0132	1.0180	1.0130	1.0070	0.9701	1.0060
Y_2	1.1580	0.8910	1.2360	1.0935	1.0970	1.0970	1.0860	1.0190	1.0460
W_1	1.5000	1.3660	1.4960	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000
W_2	1.4960	1.0860	1.3200	1.5000	1.5000	1.5000	1.5000	1.5000	1.5000
V	0.0605	0.0585	0.0647	0.0600	0.0600	0.0599	0.0599	0.0595	0.0597
Cost									
Iterations	24	28	34	22	20	29	13	6	16
Evaluations	348	310	569	22	20	29	17	11	22
Time (rel.)	264	309	373	200	182	264	155	100	200
Averaged time (rel.)		208			142			100	

we have limited the tuning effort for this study. Moreover, linear response surfaces have been used in the MAM studies, to limit the number of sampling points required in every iteration. When quadratic response surfaces would have been used, it might be that the higher quality of the approximation would lead to faster convergence. However, the number of sampling points required for using quadratic approximations depends quadratically on the number of design variables, which makes this option unattractive for larger design problems.

The designs obtained by the gradient-based methods are in good agreement, regardless of the starting point used. This could indicate that the present problem has convex properties, although this cannot be concluded with certainty based on this modest number of trials. In comparison, it turns out that MMA is more consistent in finding the same optimum than SQP, but it also uses more iterations. Variables W_1 and W_2 are at their upper bounds, indicating that when the application would allow wider gripper plates, further performance improvement could be possible. Active constraints that can be identified in various cases are either the strain or temperature constraints of the opened case, indicating that this case is apparently the most critical for this design problem.

There is some variation in the number of iterations between the different cases and methods, and generally the gradient-based methods converge slightly faster. SQP shows the least iterations, but in some iterations a line search is used, which

makes that the number of function evaluations is close to that used by MMA. Furthermore, the MAM method clearly requires much more function evaluations, as it is based on response surface building on sets of sampling points. However, next to comparing this number of evaluations, in practical situations it can also be relevant to consider the time required for the optimization process. Assuming sufficient computing facilities and software licenses are available to evaluate all sampling points of every MAM iteration simultaneously, and given the fact that evaluation of the responses requires less time than evaluating the responses and their gradients, it is found that in terms of time, the difference between the MAM approach and the gradient-based methods is much smaller than just a comparison of the number of function evaluations indicates. These time comparisons have been normalized with respect to the fastest case (set to 100). Note that every MAM iteration in this view requires the time of two function evaluations (sampling and sub-optimal point evaluation), which for the present problem both require 60% of the time of a simulation with sensitivity analysis (see also Table 8.5). On average, in terms of time between starting the process and obtaining an answer, the gradient-based approaches beat the MAM method by only a factor 2 or less. Considering the fact that setting up a sensitivity analysis for a given problem can be a very complex and time-consuming task, this shows that for practical purposes response-surface-based methods are in fact quite competitive, for problems with a modest number of design variables. On the other hand, purely based on the computational effort and the quality of the results, the gradient-based methods clearly turn out to be superior in this case, showing a reduction of 94 to 96% of the average number of function evaluations in comparison to MAM.

Note that comparing optimization algorithms is not as exact as it may seem. Much depends on the settings for certain parameters of the optimizer, the starting point used, and the termination criteria. For the gradient-based cases, the optimization process was terminated when the norm of relative variable change became smaller than 10^{-3}, or when the absolute objective change dropped below 10^{-4}. Constraint violations smaller than 10^{-3} were tolerated. For the MAM, similar but slightly less strict convergence criteria were used.

Figure 8.5 and Figure 8.6 show the history of the objective function and the thickness design variable t during the optimization process, for all three algorithms, for the cases using Starting point C. The objective initially worsens, as the starting point is infeasible, and the optimizers initially try to achieve feasibility. It can also be seen that the MAM objective does not improve significantly anymore after 25 iterations, and its final objective is lower than that of the gradient-based methods. This could be caused by the fact that linear approximations are used, while close to convergence quadratic approximations would probably perform better. An interesting extension of the MAM method as used here might therefore be, to include an algorithm that allows adaptive switching between approximations, based on the convergence characteristics.

The evolution of the plate thickness illustrates key differences between the three methods. The MAM method makes use of an adaptive move limit strategy,

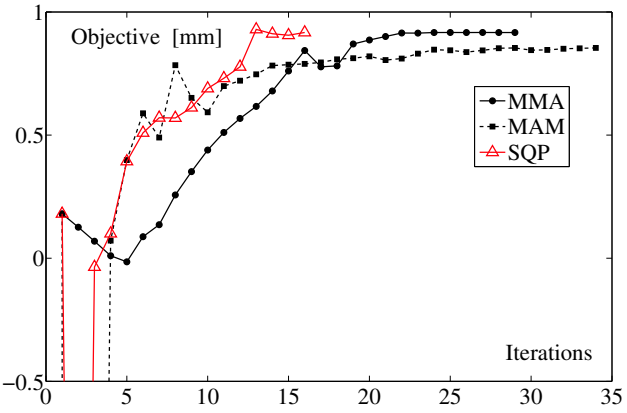


Figure 8.5: *History of the objective value for the three different optimizers used, in case of Starting point C.*

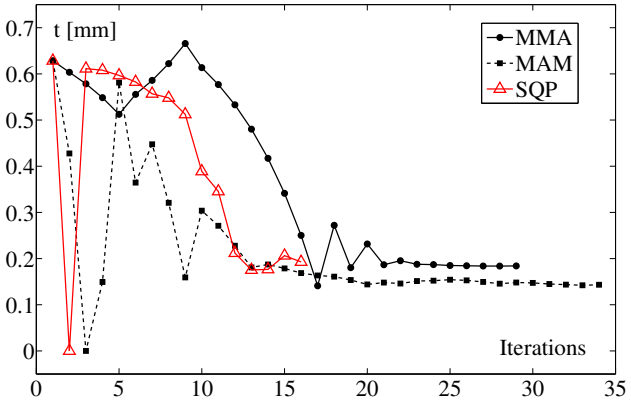


Figure 8.6: *History of the thickness design variable for the three different optimizers used, in case of Starting point C.*

which makes that initially rather large design changes are possible, but gradually the move limits are tightened as the process zooms in toward the optimum. The MMA method intrinsically limits the changes of design variables, and no large steps are observed here. The thickness gradually evolves to its final optimal value. Finally, in case of SQP, no explicit limits are imposed on the changes of design variables, which is reflected in the fairly large thickness changes observed both at the beginning and at a later stage of the process. A full discussion of all the convergence histories for all responses, variables, methods and starting points is outside the scope of this chapter, but the two examples shown in Figure 8.5 and Figure 8.6 give a good representation of the typical observations also found in other cases.

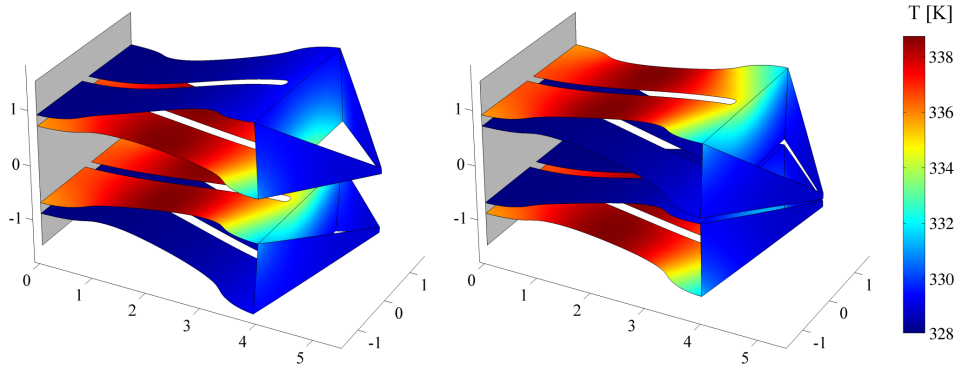


Figure 8.7: *Temperature distribution [K] on the optimal miniature SMA gripper in the open and closed configuration. A full color version is given by Figure G.7 on page 253.*

Finally, the geometry of the optimal SMA gripper design is shown in Figure 8.7, in both the open and closed configuration. The design obtained by MMA for Starting point C has been used here. Figure 8.7 also shows the temperature distributions in both configurations, and it can clearly be seen that either the inner or outer plates are activated by the applied voltage. It can also be seen that the maximum temperature exceeds the value used in the individual element constraints (338 K), since constraint aggregation was used. The maximum temperature in the optimal design was found to be 338.76 K. However, the excessive temperatures only are present in a small area of the complete gripper. It depends on the demands of a specific application whether the given degree of constraint violation is acceptable. If not, the constraint violation can be reduced easily by e.g. increasing the ρ-parameter in the aggregation function (Equation 8.4) or by reducing the limits used in the individual constraints. In a trial run, it was found that increasing the ρ-parameter from 40 to 80 resulted in a 40 to 50% decrease of the maximum violations. However, the tightening of the constraints also decreased the objective by 6%.

8.5 Conclusions

As recognized by other authors, the application of design optimization techniques shows great promise for the design of SMA structures, as the complexity of the problem often defies trial-and-error or intuition-based approaches. The research presented in this chapter extends the available techniques for SMA design optimization in several ways. Firstly, use is made of a versatile multi-disciplinary finite element model, that considers electrical, thermal and mechanical aspects in order to evaluate the performance of a design. Next to this, the design problem is cast into a formal optimization problem, which allows the consideration of clear performance-based objectives and constraints. Furthermore, an efficient semi-analytical sensitivity analysis procedure has been developed for the finite element model. This finally allows the use of gradient-based optimization algorithms, that are clearly more effective than methods that only utilize the responses. A key enabler of this approach is the history-independent SMA constitutive model used in this work, which allows efficient sensitivity analysis.

The comparison of sensitivity analysis approaches revealed that a restarted finite difference approach only required slightly more computation time compared to the semi-analytical approach, for the miniature SMA gripper case. Since implementation of the finite difference method is much less involved, this method turns out to be an attractive option. Some caution must however be taken with the use of this method in combination with noisy responses, as it was found that in rare cases it can produce inaccurate sensitivity results. Furthermore, the presented method is restricted to history-independent models.

Comparison of the SQP, MMA and MAM optimization approaches clearly show that the response-surface-based MAM method required far more function evaluations, and resulted in an optimum design of lesser quality. Here it should be noted that no extensive tuning of the optimizer settings was performed beyond what could be considered reasonable, in order to allow a fair comparison. Furthermore, if the algorithm would be able to adaptively switch between the type of approximations that are used (e.g. linear or quadratic), improvements in convergence and efficiency are certainly possible.

The gradient-based SQP and MMA methods yielded more or less comparable results. MMA obtained slightly more consistent results, whereas SQP converged faster. Based on the results of this study, the most efficient approach in terms of performance is to use semi-analytical sensitivity analysis in combination with the SQP optimizer. From another point of view however, when merely the time needed to obtain a design optimization result is of interest and sufficient parallel computing power is available, the MAM method only requires double the time of the fastest gradient-based method, while not requiring a sensitivity analysis implementation. Furthermore, the MAM approach can also be applied to history-dependent problems. As long as the number of design variables remains modest, it should certainly be considered.

In conclusion, the presented results clearly demonstrate the potential of gradient-based design optimization for SMA structures. Extension of this approach to models involving history-dependent material behavior, such as the majority of existing SMA constitutive models, remains as a challenge for the future.

8.6 Acknowledgments

The authors would like to thank Krister Svanberg for giving permission to use his MMA code for this research.

Based on: Langelaar, M. and van Keulen, F. (2007). Gradient-based Design Optimization of Shape Memory Alloy Active Catheters. *15th AIAA/ASME/AHS Adaptive Structures Conference, Waikiki, Hawaii, April 23–26, 2007*, accepted.

Chapter 9

Gradient-based Optimization of SMA Active Catheters

The design of an active catheter is an example of a challenging design problem of a Shape Memory Alloy (SMA) adaptive structure. The objective is to find a geometry that combines the electrical, thermal and mechanical properties of the structure in such a way that optimal bending performance is achieved. This chapter introduces the application of an efficient gradient-based design optimization procedure to this design problem. The specific model used focuses on the R-phase transformation in NiTi, and involves multi-point constraints to implement symmetry conditions. The nonlinear mechanical analysis is carried out using an incremental-iterative approach in combination with an augmented Lagrangian technique to account for the nonlinear constraints. Sensitivity analysis is performed using finite differences in combination with fast reanalysis, where a new correction term is applied to the multi-point constraints that significantly improves the accuracy. The proposed gradient-based optimization approach is compared to an alternative direct method, and a clear advantage in terms of the number of required function evaluations is achieved. The application of design optimization yields active catheter designs that clearly outperform previous versions. It is expected that the presented method will prove useful for the design of other SMA adaptive structures as well.

9.1 Introduction

The improvement of the steerability and controllability of catheters by integrating miniaturized distributed actuators has recently attracted considerable interest

(see e.g. Dario *et al.*, 1991, Kaneko *et al.*, 1996, Lim *et al.*, 1995, Mineta *et al.*, 2001, Park and Esashi, 1999b), and also Chapter 3. By the integration of actuators, catheters can be turned from passive devices into adaptive instruments, whose shape can be modified in a controllable way. This capability is very helpful in accurately navigating to remote treatment locations of the body in a non-invasive manner (Tendick *et al.*, 1998).

Most of the active catheter prototypes that have been developed make use of shape memory alloys (SMAs) for actuation. SMA actuators are an attractive choice in comparison to other options, because they offer the highest work density, relatively large actuation strains and stresses, and good biocompatibility (Humbeeck, 2001). Moreover, because SMAs can be used for direct actuation, the designs can remain relatively simple. However, since the geometry of a design is directly linked to its electrical, thermal and mechanical performance, which in turn all affect the overall functionality of SMA devices, designing effective structures is a difficult task, particularly in a three-dimensional setting. This chapter therefore introduces the use of gradient-based design optimization applied to the design of an SMA active catheter, and shows that this offers a systematic and effective procedure to design such SMA adaptive structures.

Several examples of the application of design optimization concepts to SMA design problems can be found in recent literature (e.g Dumont and Kuhl, 2005, Fischer *et al.*, 1999, Langelaar and Van Keulen, 2004a, Lu *et al.*, 2001, Skrobanek *et al.*, 1997). However, none of these studies used efficient gradient-based optimization algorithms. This limits their applicability to relatively simple problems with few design variables. The aim of the present chapter is to develop and demonstrate an approach that is more scalable, and better suited for design optimization of SMA structures of higher complexity and with many design variables. The case of an active catheter design problem is used as a carrier, as it involves all the aspects relevant to the more complex type of design problems we intend to solve.

First, Section 9.2 discusses briefly the active catheter concept and the designs that will be considered in the present study, as well as the computational modeling and analysis procedure. Note that the proposed optimization-based design approach is not limited to the specific SMA active catheters used in this chapter, but that it can be applied to other design problems involving SMA structures as well. In order to enable the use of efficient gradient-based optimization algorithms, design sensitivities need to be computed, and Section 9.3 treats the sensitivity analysis used for the present problem by a finite difference approach involving reanalysis. As the finite element model of the present problem involves multi-point constraints that are handled by an augmented Lagrangian approach, a novel correction is used to ensure the accuracy of the sensitivities when using reanalysis techniques, inspired by the correction for nonlinear elastostatic problems previously proposed by Haftka (1985). After validation of the sensitivity analysis, Section 9.4 presents the results of various design optimization studies, and a comparison is made between a direct method, the Multipoint Approxima-

tion Method (MAM, Toropov *et al.*, 1993a) and the gradient-based Sequential Quadratic Programming (SQP) algorithm, followed by the conclusions.

9.2 Problem formulation

9.2.1 Active catheter design concept

Most active catheter prototypes reported in the literature consist of an assembly of many individual parts, which increases the complexity and cost of the manufacturing process and reduces the overall reliability and robustness of the devices. Therefore, in a previous paper, we have proposed an active catheter design aimed at maximum integration of electrical, thermal and structural functions (see Chapter 3). In the present chapter, this concept is further extended and optimized for various diameters. The active catheter concept and its proposed fabrication process are shown in Figure 9.1. A pattern is laser-cut or etched in a small diameter SMA tube, and subsequently initial strains are introduced in the resulting structure by inserting spacers. Straining the structure is essential in order to make use of the SMA effect for actuation. The pattern is chosen such, that the spacers can be inserted easily, while the structure as a whole is stretched axially.

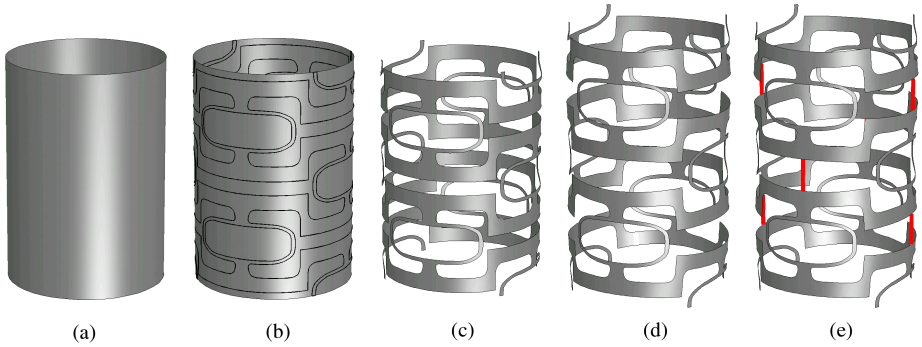


Figure 9.1: *Proposed fabrication of the new active catheter concept using laser cutting from a small diameter SMA tube (a). After laser cutting (b–c), the structure is stretched (d) and spacers are inserted (e), in order to generate internal stresses. A full color version is given by Figure G.2 on page 250.*

The final structure of the schematic concept shown in Figure 9.1 contains orthogonal bending sections, and by combining these, the catheter can bend in any direction. The actuation takes place by sending a current through parts of the structure, which locally heats the SMA structure because of its inherent resistivity. Figure 9.2 illustrates a voltage pattern that could be applied to this structure in order to induce bending. A more elaborate description of this active

catheter concept can be found in Chapter 3. Note that although the design of an active catheter typically also involves other aspects as well, such as a control and sensory system, this chapter focuses on the design of its mechanical structure, since that is the part that is essential for achieving the desired functionality.

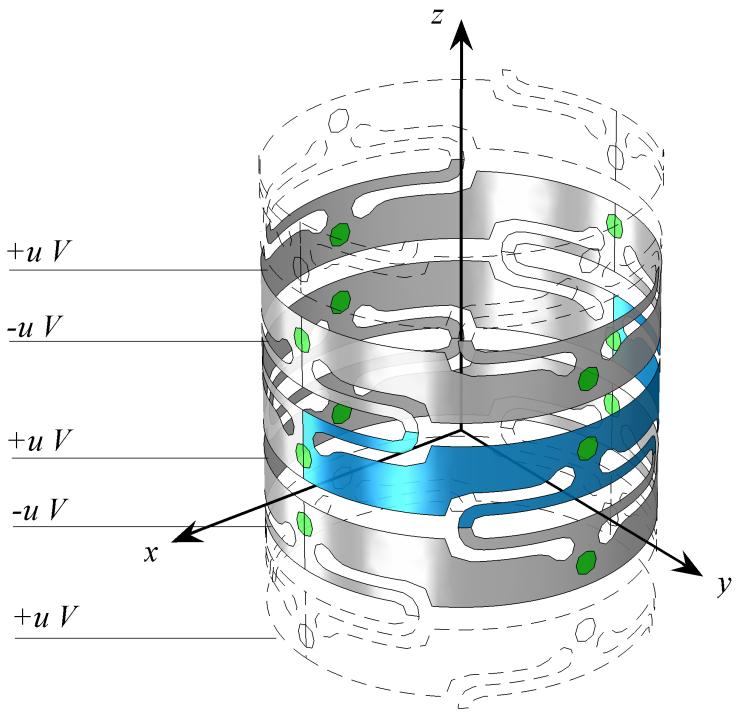


Figure 9.2: *Segmentation, applied voltage pattern and coordinate system. A full color version is given by Figure G.3 on page 250.*

Given the limitations imposed by the *in vivo* working environment of devices such as active catheters, the temperature range in which the actuation units are operated is limited. For human applications, the ambient temperature is given by the body temperature, i.e. 37°C. Based on various clinical models on tolerable temperatures, a conservative estimation for the upper limit was determined to be 49°C (see Chapter 3). This small temperature window does not allow for an effective use of conventional SMAs. However, it was found that the R-phase transformation, which occurs in certain NiTi alloys, can be applied in this setting. In addition, the hysteresis typical for SMAs is particularly small in case of the R-phase transformation, which is an attractive property for actuator applications (Otsuka and Wayman, 1998, Tobushi *et al.*, 1996). The inherent drawback of using the R-phase transformation is that the actuation strain is considerably smaller

than that of the generally used martensite transformation of NiTi, offering only 0.7% recoverable strain versus 5–7% for the martensite case. This limited amount of actuation strain makes it important to optimize the design as much as possible, to achieve maximum functionality and the largest possible amount of bending.

9.2.2 Active catheter modeling

The active catheter is controlled by applying voltage to electrodes on the SMA structure, which results in local temperature changes that influence the material behavior, resulting in a change in the equilibrium configuration of the structure. In this way, by choosing a certain voltage pattern, the catheter can be bent in any direction, as illustrated in Figure 9.2 for positive bending in the x, z-plane. Therefore, to simulate the bending performance of SMA active catheter designs, a sequentially coupled electrical, thermal and mechanical analysis is required. Finite element modeling is used to construct a computational model. The electrical and thermal problems are linear, whereas the mechanical problem involves geometrical and physical nonlinearities. The latter is solved by an incremental-iterative solution strategy, and involves shell elements (Van Keulen and Booij, 1996), truss elements with initial strains to model the spacers, and special constraint elements to enforce a symmetry constraint discussed below (see also Chapter 3).

For the present active catheter application, a suitable alloy has been selected, which exhibits the R-phase transformation. This Ti-55.3wt%Ni alloy has been experimentally characterized by Tobushi *et al.* (1992), and based on the experimental data and known properties of the R-phase transformation, a constitutive model has been developed as described in Chapter 3. The details of this model are not repeated here, as the focus of the present chapter is on the sensitivity analysis and gradient-based design optimization of the SMA structure. However, it is important to note that in contrast to many conventional constitutive models for SMA behavior, the used model is history-independent, since the hysteresis of the R-phase transformation is sufficiently small to be neglected. In this way, the model essentially classifies as a temperature-dependent nonlinear elastic model. The critical advantage of the history-independence of the model is that it significantly simplifies the sensitivity analysis (see e.g. Kleiber *et al.*, 1997).

In an optimization process, typically many function evaluations of different designs are required, therefore it is important to exploit methods to limit the computational costs. For the present active catheter model, use is made of the symmetry and periodicity of the structure and its loading, by considering only the segment highlighted in Figure 9.2. The sides of this segment must remain in the x, z symmetry plane, and the top and bottom also act as symmetry planes with respect to the connecting segments. However, the top and bottom symmetry planes, both initially parallel to the x, y-plane, must be free to rotate around the y-axis, and to translate in the x, z plane, as the catheter bends. This condition has been implemented by multi-point constraint equations, which enforce that a set of nodes remains in a single plane, which is free to rotate and translate as described.

A full discussion of the formulation of these constraints can again be found in Chapter 3 and is not repeated here. An augmented Lagrangian approach is used to account for the constraints in the solution process, as this method turned out to offer the highest robustness and efficiency. However, this has implications for the sensitivity analysis, as is discussed in more detail in Section 9.3.

9.2.3 Design optimization formulation

In order to find the active catheter design that offers maximum bending, design optimization is applied. The objective of the present optimization problem is to maximize the bending, which is equivalent with minimization of the bending radius. Two constraints are included in the optimization problem: firstly, the SMA material is restricted to a certain strain range, as the range of validity of the used material model is limited to 1% effective strain. Secondly, the operating temperatures are limited to the range of 37–49°C as discussed before. Including strain and temperature constraints for every individual element would lead to a prohibitively large number of constraints, therefore a Kreisselmeier-Steinhauser (KS) constraint aggregation procedure is applied (Kreisselmeier and Steinhauser, 1983). This leads to the following formulation of the design optimization problem:

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{x}) = R(\mathbf{x}) \\ &\text{Subject to:} \\ &\quad g_1(\mathbf{x}) = KS_A(g_\varepsilon(\mathbf{x})) - 1 \leq 0, \\ &\quad g_2(\mathbf{x}) = KS_A(g_T(\mathbf{x})) - 1 \leq 0, \\ &\quad \underline{\mathbf{x}} \leq \mathbf{x} \leq \overline{\mathbf{x}}. \end{aligned} \tag{9.1}$$

Here R represents the bending radius measured from the centerline of the tube, ε_e is the effective strain, a scalar measure for the strain used in the SMA constitutive model (see e.g. Chapter 3), and T is the element temperature. The vector $\mathbf{x}$ represents the design variables, and $\underline{\mathbf{x}}$ and $\overline{\mathbf{x}}$ contain their lower and upper bounds. The aggregated strain and temperature constraints are given by g_1 and g_2, respectively. The individual element constraints and the used aggregation function are given by:

$$g_\varepsilon^{(i)} = \frac{\varepsilon_e^{(i)}}{\varepsilon_e^{(max)}} \leq 1, \tag{9.2}$$

$$g_T^{(i)} = \frac{T^{(i)} - T_{min}}{T_{max} - T_{min}} \leq 1, \tag{9.3}$$

$$KS_A(\boldsymbol{g}) = \frac{1}{\rho} \ln \left(\frac{1}{\sum_{i=1}^{N} A^{(i)}} \cdot \sum_{i=1}^{N} A^{(i)} e^{\rho g^{(i)}} \right). \tag{9.4}$$

Here A is the element area, N is the number of elements and ρ is a parameter used in the KS function, which is set to 40 in this study. Furthermore, the strain

Table 9.1: *Design variables and their bounds.*

Variable	Lower bound	Upper bound
Wall thickness ratio t/D	0.03	0.2
Shape parameter H/H_{ref}	0.8	2
Shape parameter W/W_{ref}	0.1	2
Spacing factor ratio Δ/Δ_{ref}	0.5	3
Applied voltage V [V]	0.005	0.05

and temperature limits are 1% and 37–49°C, respectively. The KS_A aggregation function is a slightly modified version of the standard KS function, which reduces the importance of small localized constraint violations by including the element areas. The degree of violation can be controlled by the parameter ρ as well as by adjusting the constraint limits, depending on the demands of the application.

For the present optimization study, three related but different design concepts are considered, as illustrated in Figure 9.3. The geometries in Figure 9.3 show only a quarter of the analysis segment highlighted in Figure 9.2, since the full segment is constructed by connecting four of the basic units shown in Figure 9.3. The concepts differ mainly in the design of the flexible section, which contains one, two or three full loops, hence the designs will be referred to as the 1-loop, 2-loop or 3-loop design. The diameter D of the tube is considered a parameter, since certain medical applications will require a certain diameter depending on vessel size and required working channel. The design variables selected for the present optimization study are the wall thickness t, the two shape variables H and W as indicated in Figure 9.3, the spacing-factor Δ that indicates with what factor the spacers will increase the associated gaps in the structure, and finally the applied voltage V. The wall thickness is normalized by the tube diameter, and the other geometrical variables are normalized with respect to their initial value in the reference designs shown in Figure 9.3, after those designs have been scaled according to the required design diameter. The bounds used for these design variables are listed in Table 9.1.

9.3 Sensitivity analysis

9.3.1 Method

In order to enable the use of efficient gradient-based optimization algorithms, it is necessary to compute design sensitivities for the present active catheter model. Various approaches exist to perform this sensitivity analysis, which are reviewed by e.g. Haftka and Gürdal (1992), Kleiber *et al.* (1997), Van Keulen *et al.* (2005). The choice of the most efficient method depends on the nature of the problem.

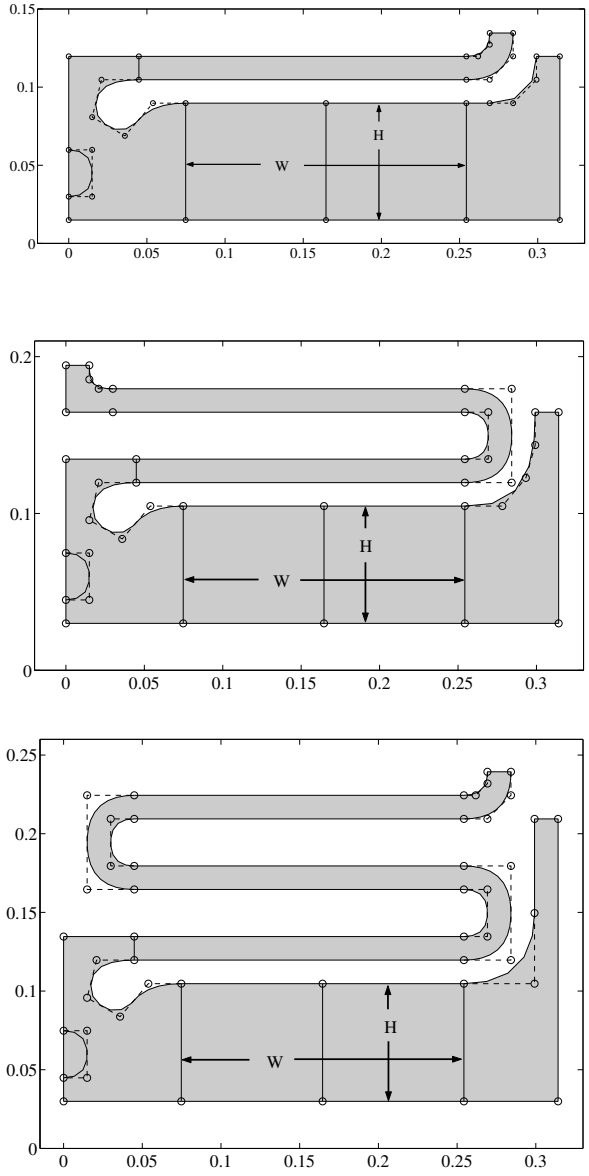


Figure 9.3: *Default designs for the 1-loop, 2-loop and 3-loop case (dimensions in mm).*

In the present case, the model is a path-independent nonlinear model involving nonlinear constraints. However, the solution technique that is used in the simulation, the augmented Lagrangian approach, in fact changes the nature of the problem. Auxiliary variables called Lagrange multipliers are introduced as internal variables, and these are updated in each iteration by an additional update rule, given by:

$$\lambda^{k+1} = \lambda^k + ph(\mathbf{u}^k) \tag{9.5}$$

Here λ^k represents the multiplier value at iteration k, p is the penalty factor used and h is the constraint that needs to be satisfied, which is a function of the displacement field $\mathbf{u}$. When internal variables are present in a model, the model has to be treated as a path-dependent, transient model (Kleiber *et al.*, 1997). As the solution is determined by both the displacements and the multiplier values at the final stage, the design derivatives of the multipliers need to be computed as well in order to evaluate the total design sensitivities of the system. In order to do this, in fact after each iteration a sensitivity analysis has to be carried out, since Equation 9.5 links new multiplier values to previous multiplier and displacement values.

Based on these observations, it appears that a costly transient sensitivity analysis procedure has to be applied. However, in fact the underlying problem that is being solved still is a path-independent nonlinear model with nonlinear constraints. Because of the solution approach that was chosen, it has been turned into a transient type of problem. In case a penalty or Lagrange multiplier method would be applied, the problem would remain path-independent and sensitivities could be computed efficiently. Therefore, we observe that the choice of the solution method influences the required sensitivity analysis. However, when the final solution of the original problem is the same, it is justified to use any method that is applicable to either approach, since the underlying problem that has been solved remains unchanged. The penalty method results in approximate solutions, since the constraints are not satisfied exactly when a finite penalty value is used. The Lagrange multiplier method in contrast yields the same solution as the adopted augmented Lagrange multiplier technique. In this method, the auxiliary multipliers are not used as internal variables but are treated as degrees of freedom, in the same way as the displacements are treated. Hence, the problem remains path-independent and efficient sensitivity analysis techniques suited for this type problems can be used.

A previous study of options for design sensitivity analysis of similar SMA design optimization problems without constraints has shown that attractive options for these nonlinear path-independent problems are either the semi-analytical method or finite differences in combination with reanalysis (see Chapter 7). The semi-analytical method was found to be slightly more efficient, but this was offset by the fact that its implementation is considerably more costly. For the present case, one option might be to transform the problem into the Lagrange multiplier formulation, after obtaining the solution using the augmented Lagrangian

approach, and to subsequently use the semi-analytical approach to compute the sensitivities. Using the Lagrange multiplier method for the analysis itself is not feasible, as it proved to suffer from convergence problems due to the loss of positive definiteness of the tangent system matrix (see Chapter 3). However, implementation of the proposed transformation and the subsequent semi-analytical sensitivity analysis procedure will require a substantial effort, and in addition the Lagrange multiplier formulation requires a solver that is able to handle the non-positive definite matrix. Furthermore, as the problem is transformed, the original decomposed system matrix can not be reused, because both the formulation and the number of degrees of freedom are modified. Therefore, this approach would reduce the efficiency of the semi-analytical approach considerably.

A more attractive option is to make use of the finite difference method in combination with reanalysis. Finite differences involve a nominal and a perturbed solution, and by starting the perturbed analysis from the nominal solution, the cost of this method approaches that of the semi-analytical method (see Chapter 7). And in fact, there is no restriction on using the same augmented Lagrangian solution approach for the perturbed case as well. Implementation of this finite difference approach is not as complex as the described semi-analytical option, and moreover this method is far more efficient than adopting the alternative total transient or global finite difference techniques. Note again that in the present problem, the transient nature is only a result of the adopted solution strategy, and not an inherent property of the problem itself. The presented reasoning therefore does not hold for cases where the problem itself is path-dependent.

9.3.2 Implementation and validation

The proposed sensitivity analysis procedure consisting of forward finite differences in combination with reanalysis has been implemented in an in-house dedicated finite element code. The reanalysis only applies to the mechanical part of the simulation, the electrical and thermal problems are linear and a full analysis is used to solve the associated perturbed cases. A correction is applied to the residual of the perturbed mechanical problem, in order to ensure accurate sensitivities, as proposed by Haftka (1985). The finite residual remaining from the nominal case is subtracted from the residual of the perturbed case. This improves the accuracy, particularly in case of small perturbations, since it isolates the influence of the design variable perturbation from further iterative improvement of the solution. But the residual is not the only finite error that exists in the nominal solution. Since also the constraint equations $\mathbf{h} = \mathbf{0}$ are solved by an iterative process in the augmented Lagrangian approach, a small but finite error remains and the constraints are not exactly satisfied when the iterative process is terminated. In case of the constraints in the present SMA active catheter model, the error is expressed by the distances of nodes to the defined symmetry plane. These generally finite distances result in nonzero equality constraint values $\Delta\mathbf{h}$ at the final solution of the nominal case. By subtracting these $\Delta\mathbf{h}$ values from the corresponding

constraints in the perturbed case, again the effect of the design perturbation is isolated from further improvement of the nominal solution, in the same way the residual is corrected as proposed by Haftka (1985). This correction is expected to result in more accurate sensitivities, particularly for smaller perturbations. To the best knowledge of the authors, improving the accuracy of sensitivities by this constraint correction for problems with multi-point constraints has not been reported before.

In order to investigate and validate the proposed procedure and the effect of the constraint correction, first a small test problem will be considered. This simplified test problem has the same components as the full catheter problem, i.e. a sequentially coupled electrical, thermal and mechanical simulation, where the mechanical problem involves the same shell, truss and constraint elements as the full model. This test problem is depicted in Figure 9.4, which shows four shell elements using the SMA material model also used for the catheter (in the rectangle ACDF), one linear elastic bar element (CG) and two constraint elements (ABG and ACG). These constraint elements restrict the motion of nodes B and C to the x, z-plane. The edge AF is clamped, and a load is applied to the side CD as indicated. Edge DEF is constrained in y- and z-direction, and also the position of

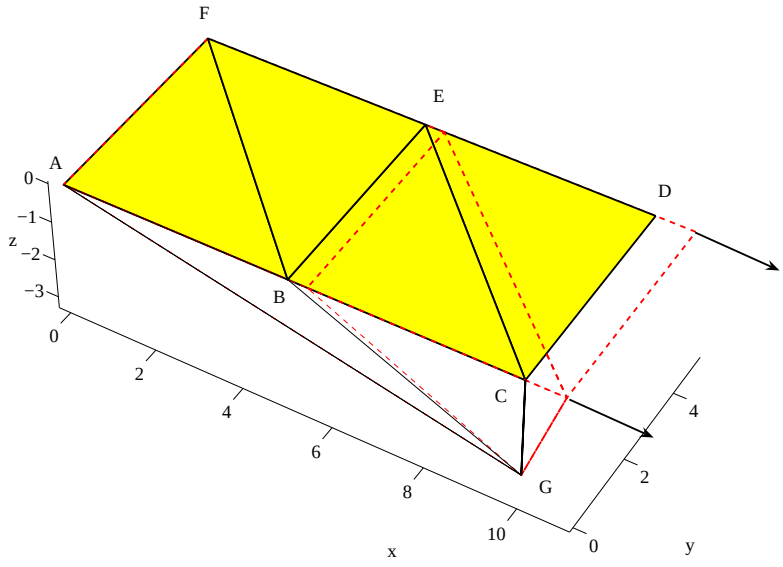


Figure 9.4: *Test problem used to test the sensitivity analysis procedure. Dashed lines show the deformed configuration.*

node G is fixed. Without constraints, the loading would result in a contraction in y-direction, causing nodes B and C to move away from the x, z-plane. Note that in this test problem it would have been much simpler to constrain the y-displacement of these nodes instead of using the constraint elements, but this is not relevant for the purpose of this test model. The aim is to verify the sensitivity analysis procedure using an equivalent but simpler model. The electrical and thermal analysis only involve the rectangle ACDF, where they affect the SMA material. The electrical boundary conditions are given by prescribed voltages on edges AF and CD, and the thermal case involves only surface convection.

In order to validate the proposed sensitivity analysis procedure and the correction applied to the constraints, three cases are considered:

1. Finite differences using a *full* analysis for the perturbed case (no constraint correction necessary!), providing reference values

2. Finite differences using a *restarted* perturbed analysis, *with* constraint correction

3. Finite differences using a *restarted* perturbed analysis, *without* constraint correction

In both cases using reanalysis, the proven residual correction proposed by Haftka (1985) has been applied. Tests have been done with various settings of the convergence tolerance, to also examine their effect on the accuracy of the resulting finite difference sensitivities. The nominal convergence tolerance equalled 10^{-5}. This tolerance relates to the sum of the norm of the unbalanced forces with respect to the norm of the total applied forces and the norm of the displacement increment normalized by the norm of the total displacement. The response considered is the aggregated strain constraint value $KS_A(g_\varepsilon)$ as also used in Equation 9.1. Various design variables have been used, and sensitivities have been computed for a range of relative design perturbations, in order to evaluate the accuracy and stability of the different approaches.

A typical result is shown in Figure 9.5 for the case of the shell thickness design variable t. It is seen that when using tolerance settings similar to the nominal case, the effect of the constraint correction in the restarted case is minor. In that situation, both corrected and uncorrected restarted cases clearly lack accuracy as the relative design perturbation is reduced, in comparison to the full finite difference approach. This reduced range of stable and accurate sensitivity results for the restarted case is presumably due to the fact that for increasingly smaller perturbations, the restarted iterative solution process terminates more quickly, since the unbalance terms generated by the design perturbation also reduce. When the same tolerance level is used in the perturbed case as is used in the nominal analysis, the perturbed result might not fully include the (subtle) effect of the change in the design variable, which leads to inaccurate sensitivity results. This explanation is supported by the fact that the accuracy improves drastically when additional

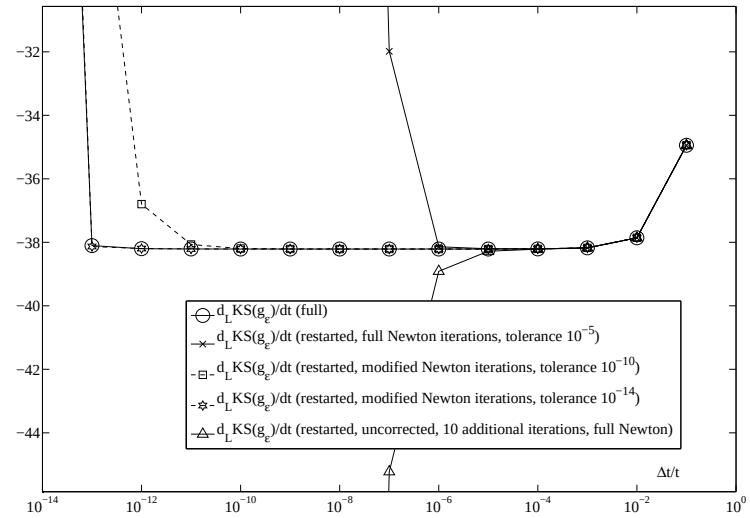


Figure 9.5: *Logarithmic sensitivities of the aggregated effective strain constraint for the test problem, as a function of the relative design perturbation in the thickness t.*

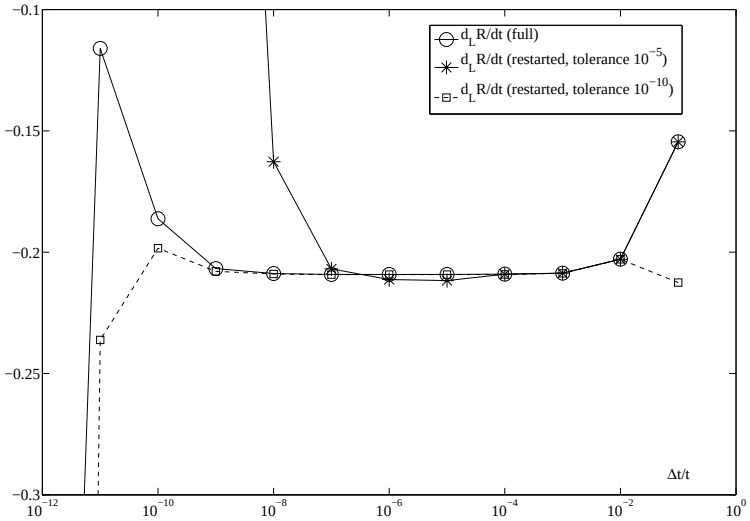


Figure 9.6: *Logarithmic sensitivities of the bending radius R for the 1-loop catheter case, as a function of the relative design perturbation in the thickness t.*

iterations are performed after the restarted perturbed analysis has converged to the nominal tolerance level, or equivalently when tighter tolerance settings are applied. This is illustrated in Figure 9.5 by the results of the restarted cases with constraint correction with reduced tolerance levels, which approximate the full finite difference results very closely. Modified Newton iterations were used in these cases, since these turned out to be as effective as full Newton iterations. Particularly for large problems, this will considerably reduce the required computational effort. To allow a fair judgement of the effect of leaving out the constaint correction, 10 additional full Newton iterations were performed in that case. This however did not improve the accuracy, as is seen in Figure 9.5, and it can be concluded that this is due to the fundamental error of not accounting for the finite accuracy of the constraint conditions for the nominal case.

Sensitivity analysis results for the active catheter case are depicted in Figure 9.6, and here the constraint correction as it was verified by the test problem study has been included. The results shown are for the sensitivity of the bending radius R with respect to the catheter wall thickness variable t for the 1-loop design concept, but for other concepts, design variables and responses, similar results have been obtained. Also here the beneficial effect of tightening the tolerance settings of the perturbed analysis can be observed. The results of the restarted case with a tolerance of 10^{-10} approximate the full finite difference sensitivity results very closely, while when using the nominal tolerance setting of 10^{-5}, acceptable, but clearly less accurate results are found. Due to the use of modified Newton iterations, the cost of the sensitivity analysis is still relatively low. This is illustrated by the fact that in this active catheter case, the unavoidable mesh genereration for the perturbed design takes twice the time required for the modified Newton process to obtain the perturbed solution. In comparison with the full finite difference approach, the computational time for a single evaluation including sensitivities is reduced by a factor 2.5, for the present case with 5 design variables. When the problem is scaled up to larger models with more design variables, this factor will increase even further. Based on the findings obtained in this sensitivity analysis study, the restarted finite difference approach using modified Newton iterations and a tolerance level of 10^{-10} is selected to evaluate the sensitivities in the optimization problems, using a relative perturbation of 10^{-6}.

9.4 Optimization results

The design optimization of the proposed SMA active catheter concepts is performed using two different approaches, the Multi-point Approximation Method originally proposed by Toropov *et al.* (1993a) and Sequential Quadratic Programming (see e.g. Fletcher, 1980). The MAM is a direct method which relies on building a series of approximate subproblems by response surface building, in combination with a move limit strategy. In every iteration, optimization of the subproblem results in a sub-optimal point, which is subsequently evaluated using

the full model in order to determine the accuracy of the response surface. The sample points required for the response surfaces are evaluated in parallel, in order to reduce the duration of the optimization process. Note that evaluations for optimization using MAM do not involve sensitivity analysis. The second algorithm, SQP, is generally regarded as the best general-purpose gradient-based optimization algorithm, and the specific implementation used here is the one found in the Matlab Optimization Toolbox (version 3.0.3). The gradients are computed using the finite difference approach in combination with reanalysis, as discussed in Section 9.3.

The used SQP implementation sometimes activates a line search procedure within an iteration after obtaining a search direction from the solution of the QP subproblem. This is why the number of iterations in general does not match the number of function evaluations. The SQP runs are started using the first sub-optimal point found in the corresponding MAM run as initial point. The MAM involves various heuristic parameters and settings that affect the generation of sampling points, the response surface fitting process and the move limit strategy. A moderate effort to tune these parameters has been made, and it was found that using linear response surface based on 10 sampling points yielded robust and steady convergence. Possibly further experimentation with the parameters might further reduce the total number of required function evaluations, but at the expense of further computational investments.

A full listing of the results of the design optimization for the 1-loop, 2-loop and 3-loop concepts using these two algorithms is given in Table 9.2. The obtained optimal bending radii are in close agreement in most cases, only in the 2-loop case with a 0.5 mm diameter tube the MAM converges to a local optimum. Typical convergence histories for both SQP and MAM are shown in Figure 9.7, which concerns the 2-loop 1.5 mm case. It is seen that the MAM requires many iterations particularly when it approaches the optimum, and possibly changing the move limit strategy parameters or the approximation settings may improve its convergence rate further. Note also that the obtained bending radius for the 1-loop 1.5 mm case of 29.9 mm is a clear improvement over the 35 mm obtained by a previous design (see Chapter 3). However, more impressive improvements are made when the design concept itself is changed as well instead of merely optimizing it, as illustrated by the 2-loop 1.5 mm case, which manages to produce a bending radius of 20.5 mm, an improvement of over 40%.

In general, Table 9.2 clearly shows that SQP requires less iterations and less function evaluations than the MAM, but its disadvantage is that sensitivity analysis is required, which increases the cost of the evaluations. Accounting for the fact that evaluations including sensitivity analysis for the considered cases are roughly twice as costly as a simulation alone, the advantage in total computational time of SQP over MAM is a factor 4 on average. When considering instead the duration of the optimization procedure in a parallel computing environment, the number of MAM iterations is relevant. Note that each iteration of MAM takes the time of two simulations because of the evaluation of the sub-optimal point, and that

these MAM iterations should be compared to the evaluations used in the SQP case, which cannot exploit the parallel infrastructure (unless the simulation itself is parallellized, but this is not the case here). A direct comparison can be made, since the SQP evaluations including sensitivity also take the time of roughly two simulations. When comparing the average total duration, it turns out that the MAM is only 7% slower than SQP, assuming the evaluation of sampling points can be fully parallellized. When instead looking at the average ratio over all cases, the difference between MAM and SQP is still only 18%, in favor of the latter.

When considering the actual optimal designs that were obtained, as listed in Table 9.3, it turns out that, surprisingly, there are considerable differences in the values of the design variables obtained by MAM and SQP for the same cases, even though the objective and constraint values listed in Table 9.2 are in close agreement. The largest variation is found in the shape variables H/H_{ref} and W/W_{ref}. Investigation of the logarithmic sensitivities of the final designs reveals that indeed these variables have only a weak relation to the response values. Apparently the shape of these designs is less important than the used diameter, tube wall thickness, spacer length and applied voltage, at least in case of the current parameterization. Possibly a finer parameterization allowing more shape variations might change this situation, but this question is not explored further in the present study. What is clear however, is that considerable improvements

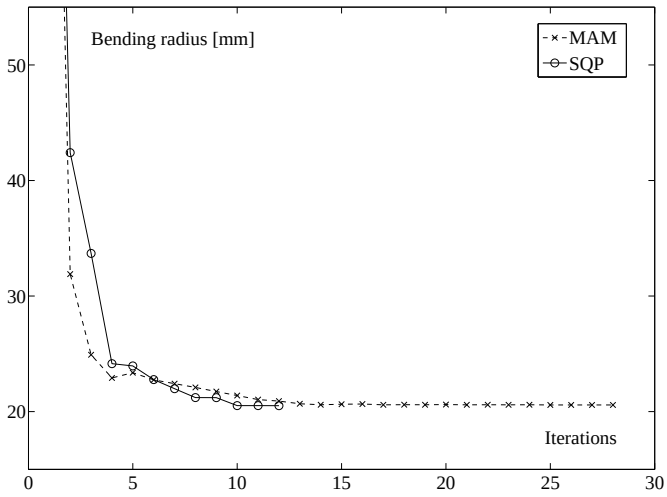


Figure 9.7: *Successful convergence of the objective value for the 1.5 mm 2-loop MAM and SQP cases.*

have been made with respect to an earlier design presented in Chapter 3, and that various design concepts can be explored and compared in a systematic way, using the proposed design optimization procedure.

A final aspect to note is that depending on the selected catheter diameter, a different design concept turns out to yield the best performance in terms of bending ratio. Table 9.4 lists the bending radii of all the considered cases, normalized by the catheter diameter. For a diameter of 0.5 mm, the 3-loop design results in the best performance. For the 1.0 mm case, the 2-loop design performs best, although the margin with the 3-loop case is small. For the 1.5 mm diameter catheter, again the 2-loop case comes first, with this time a clearer difference from the 3-loop design. These observations can be explained by the fact that the physical aspects that determine the performance of these active catheter structures are not linearly dependent on their dimensions. An optimal design at a

Table 9.2: *Optimal responses for various design concepts and cases, as well as the number of iterations and function evaluations, for the direct MAM method and the gradient-based SQP method.*

Case	Bending radius [mm]	$KS_A(g_\varepsilon)$	$KS_A(g_T)$	Iterations	Evaluations
1-loop design					
Diam. 0.5 mm, MAM	11.43	1.0000	1.0000	28	206
Diam. 0.5 mm, SQP	11.43	1.0000	1.0000	11	19
Diam. 1.0 mm, MAM	20.16	1.0000	1.0000	27	199
Diam. 1.0 mm, SQP	20.00	1.0000	1.0000	23	34
Diam. 1.5 mm, MAM	29.94	1.0000	0.9998	17	129
Diam. 1.5 mm, SQP	29.92	1.0000	1.0000	12	18
2-loop design					
Diam. 0.5 mm, MAM	**10.47**	0.9999	0.7650	19	143
Diam. 0.5 mm, SQP	**7.921**	1.0000	0.8551	9	15
Diam. 1.0 mm, MAM	14.08	0.9997	0.9999	14	108
Diam. 1.0 mm, SQP	14.07	1.0000	1.0000	11	17
Diam. 1.5 mm, MAM	20.58	1.0000	1.0000	27	199
Diam. 1.5 mm, SQP	20.51	1.0000	1.0000	12	23
3-loop design					
Diam. 0.5 mm, MAM	7.201	1.0000	0.9998	19	143
Diam. 0.5 mm, SQP	7.199	1.0000	1.0000	11	12
Diam. 1.0 mm, MAM	14.17	1.0000	0.9999	22	164
Diam. 1.0 mm, SQP	14.17	1.0000	1.0000	9	12
Diam. 1.5 mm, MAM	22.29	0.9313	0.9999	28	206
Diam. 1.5 mm, SQP	22.28	0.9241	1.0000	18	38

given scale can not simply be scaled by a factor in order to obtain an optimal design at another scale. However, as shown here, with the help of optimization well-performing designs can be generated for any scale, in a systematic manner.

In order to illustrate and analyze the way the SMA active catheter designs function, the temperature distributions of the 1.0 mm diameter 1-, 2- and 3-loop optimal designs (SQP) are plotted in Figure 9.8. Clear differences can be seen

Table 9.3: *Optimal values of the design variables for various design concepts and cases, as obtained by the direct MAM method and the gradient-based SQP method.*

Case	t/D	H/H_{ref}	W/W_{ref}	Δ/Δ_{ref}	V [V]
1-loop design					
Diam. 0.5 mm, MAM	0.0362	1.3722	1.5140	2.7287	0.0175
Diam. 0.5 mm, SQP	0.0351	1.2378	1.1401	2.8051	0.0176
Diam. 1.0 mm, MAM	0.0500	1.4282	1.1868	2.2421	0.0208
Diam. 1.0 mm, SQP	0.0431	0.8552	1.9255	2.4549	0.0215
Diam. 1.5 mm, MAM	0.0502	0.8732	1.1871	2.2550	0.0242
Diam. 1.5 mm, SQP	0.0483	1.5418	0.8434	2.3039	0.0244
2-loop design					
Diam. 0.5 mm, MAM	0.1149	1.0979	0.6036	1.3908	0.0164
Diam. 0.5 mm, SQP	0.0350	1.4491	0.9224	1.9342	0.0230
Diam. 1.0 mm, MAM	0.0414	1.3605	0.4790	1.8190	0.0312
Diam. 1.0 mm, SQP	0.0399	1.4852	1.2347	1.8490	0.0315
Diam. 1.5 mm, MAM	0.0500	1.5050	0.1000	1.7473	0.0348
Diam. 1.5 mm, SQP	0.0441	1.3614	1.2267	1.8160	0.0362
3-loop design					
Diam. 0.5 mm, MAM	0.0347	1.2096	0.9962	2.8024	0.0273
Diam. 0.5 mm, SQP	0.0356	1.2534	1.0450	2.7617	0.0271
Diam. 1.0 mm, MAM	0.0374	1.6041	0.7755	2.8235	0.0351
Diam. 1.0 mm, SQP	0.0376	1.3356	0.6253	2.8147	0.0350
Diam. 1.5 mm, MAM	0.0389	0.8000	0.6661	2.6511	0.0410
Diam. 1.5 mm, SQP	0.0377	1.6335	2.0000	2.6695	0.0414

Table 9.4: *Optimal bending radii normalized by the catheter diameter, for all considered SQP cases.*

Case	1-loop	2-loop	3-loop
Diam. 0.5 mm, SQP	22.8600	15.8420	**14.3974**
Diam. 1.0 mm, SQP	20.0000	**14.0700**	14.1700
Diam. 1.5 mm, SQP	19.9467	**13.6733**	14.8533

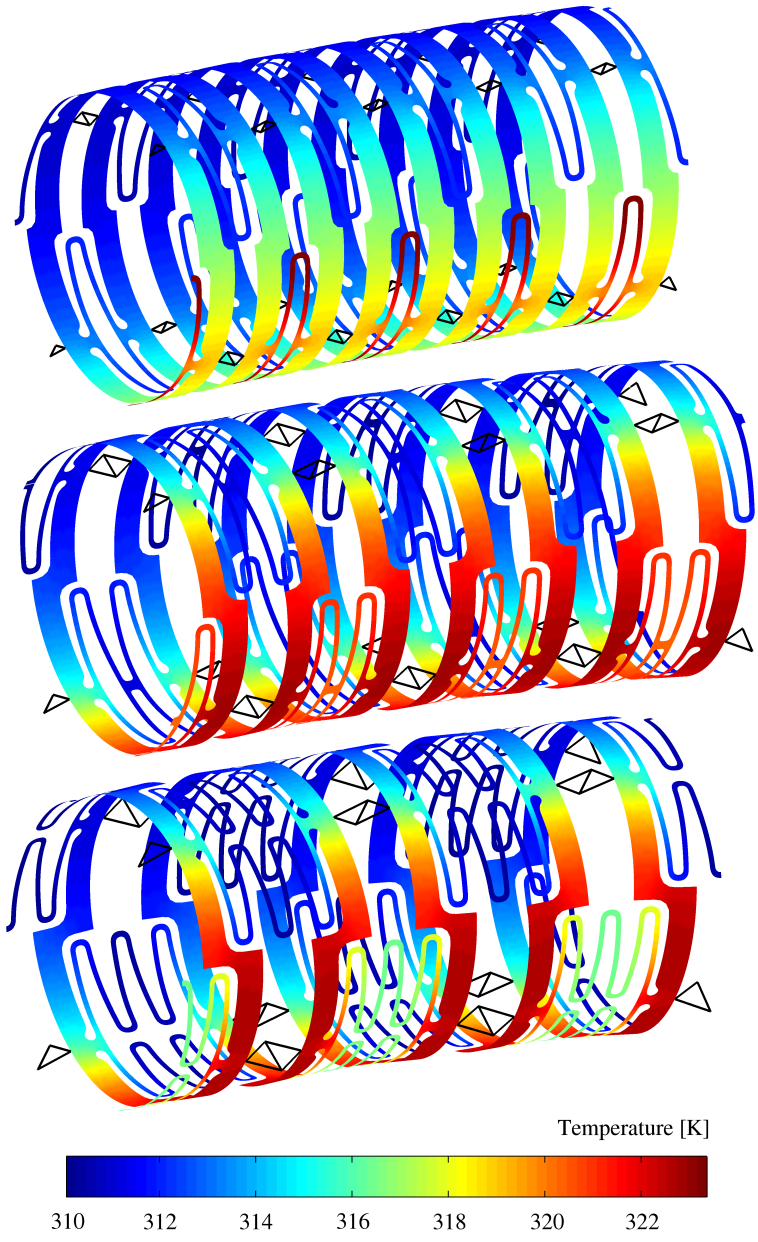


Figure 9.8: *Temperature distributions on parts of the deformed catheter structure for the best designs of all three concepts, for a 1.0 mm diameter. A full color version is given by Figure G.8 on page 254.*

between the different concepts. In the 1-loop case, the highest temperatures are found in the activated winding section. In the 2-loop case, the winding section has been increased by adding a second loop, and as a result its resistivity has increased as well. Consequently, in comparison to the 1-loop case, a larger part of the current is not flowing through the winding section, but through the support structure itself to a nearby electrode. This is reflected in the temperature distribution, which now shows considerable heating of the support structure as well. Note that this does not render this design less effective: in fact, the bending performance of this 2-loop design exceeds that of the 1-loop design by 30%. Finally, in the 3-loop design the addition of yet another loop results in a further change of the temperature distribution. In this case, the resistivity of the winding section has increased up to the point where apparently most of the heat is generated in the support structure instead of the part intended for actuation. Still the activated winding section shows higher temperatures than the opposing inactive side, but the difference is smaller than the operating limits would allow. Maximum temperatures instead occur in the support structure. In spite of this, the performance is still very close to that of the 2-loop case.

These observations provide hints for further improvement of these designs. By modifying the electrode arrangement, e.g. by adding an electrode in the middle of the winding section and letting the current flow from there, the heating could be localized more. Potentially this will further improve the performance of these active catheters, and new optimal designs could be found using the same optimization procedure. Note also that the temperature range shown in Figure 9.8 exceeds the maximum temperature of 322 K mentioned earlier. Due to the type of aggregated constraint that has been used in the formulated optimization problem, limited local violations can occur. In this case, the maximum violation is found in the 1-loop design, of close to 1.4 K in a spot on the winding section. The violations in the other cases are smaller, up to a maximum of around 0.7 K. Overall, typically only a small fraction of the total structure exceeds the 322 K temperature limit. If this is unacceptable for the application, adjusted designs can be generated by performing the optimization with tightened constraints, e.g. by increasing the Kreisselmeier-Steinhauser ρ parameter or modifying the constraint limits.

Finally, Figure 9.9 illustrates the improvements the best designs at each diameter have achieved over the previous design given in Chapter 3. The obtained bending radii are visualized in Figure 9.9 by plotting each optimized catheter in its deformed configuration. Since the design obtained in the previous study already appeared potentially useful for certain applications, these improved designs show even greater promise. Clearly also reducing the diameter improves the bending radius of the instruments, however at some point manufacturing limitations will prevent further miniaturization. Also, for catheter applications, a certain inner diameter is required for the working channel. In guidewire applications, no working channel is required, and as far as manufacturing techniques allow, further diameter reduction and improved bending behavior are certainly possible.

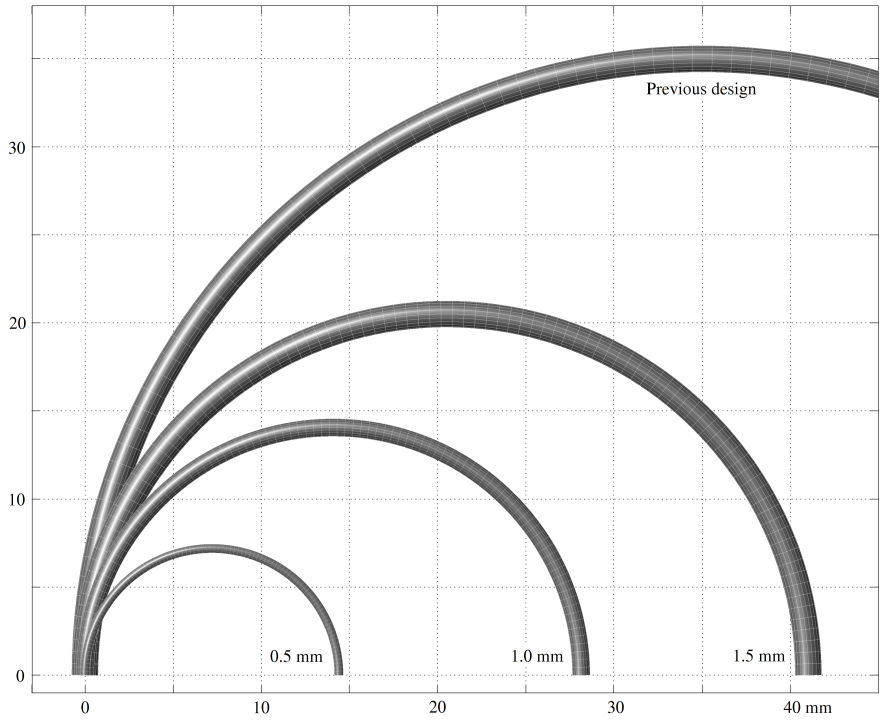


Figure 9.9: *Deformed optimal active catheters for various diameters, together with the deformed design presented earlier in Chapter 3.*

9.5 Discussion and conclusions

An efficient optimization-based design procedure for SMA adaptive structures has been developed and demonstrated in this chapter. The design of an active catheter has been chosen as a carrier application, as it presents a complex and nontrivial SMA design problem. A sequentially coupled electro-thermo-mechanical finite element model has been used to evaluate the performance of the catheter designs. The use of finite element analysis yields a versatile design approach that can be applied to a large variety of SMA design problems. Three different active catheter design concepts have been considered in the present study.

The thermal operating range of active catheters is limited, and hence use is made of the R-phase transformation. A simplified constitutive model is used, that renders the analysis path-independent, which significantly simplifies sensitivity analysis. The computation of sensitivities is necessary to use efficient and scalable gradient-based optimization algorithms. Specifically in the present design problem, multi-point constraints were included in the mechanical model, in

order to reduce its complexity by exploiting symmetries. A novel sensitivity analysis procedure has been proposed and verified, which relies on finite differences in combination with fast reanalysis of the perturbed design, using an augmented Lagrangian approach to handle the constraint equations. A tighter convergence tolerance was in the perturbed cases than for the nominal analysis, to improve the accuracy of the design sensitivities. In combination with a constraint correction for the finite accuracy involved in solving the nominal nonlinear constraint equations, this was demonstrated to yield an accurate and efficient sensitivity analysis procedure.

The proposed gradient-based design optimization procedure has been shown to be very effective and efficient, requiring far less function evaluations and computational time than an alternative direct method. However, when sufficient computational resources are available for parallel computation, the used Multi-point Approximation Method turned out to be a close match to the gradient-based SQP algorithm in terms of total duration of the optimization procedure, even without extensive tuning of its settings. Particularly for cases where implementation of the sensitivity analysis is difficult or costly, or for cases involving different SMA models that cannot exploit the path-independent properties used in the present case, this type of direct optimization procedures form a practical alternative. It has been found that when considering active catheters with different diameters, simple scaling of designs results in suboptimal performance. Instead, with the help of the proposed design optimization procedure, designs can easily and systematically be tailored to specific applications, yielding optimal performance.

The designs obtained from the optimization studies at various diameters show large improvements in bending capacity in comparison to previously published results. An analysis of the obtained optimal designs however also pointed out, that a different arrangement of the electrodes used to activate the SMA actuation sections may lead to further improvements. Also a more fine-grained shape parameterization of the design concepts could lead to higher performance, as this would enlarge the design space. The studied cases involving only two shape variables turned out to be rather insensitive to shape changes. However, an efficient optimization method such as the presented gradient-based optimization procedure is essential when dealing with design problems with many design variables. Given its effectiveness and versatility, it is expected that the application of design optimization to complex problems such as the design of adaptive SMA structures will increase further, and the presented approach could form a powerful tool for the design of many future SMA devices.

Chapter 10

Conclusions and Future Directions

10.1 Conclusions

The conclusions discussed in this final chapter have been organized according to the four major topics treated in this thesis: modeling of shape memory alloy (SMA) behavior, shape optimization without using sensitivities, topology optimization and finally gradient-based shape optimization. These topics all relate to the central question of this research, which is how and to what extent design optimization techniques can be utilized for designing SMA structures.

10.1.1 SMA modeling

Consideration of SMA modeling is essential, since the complexity of a model is inversely proportional to its suitability for design optimization. Since the target applications for this research are miniaturized actuators for medical instruments and microsystems, the focus has been on the SMA behavior due to the R-phase transformation in NiTi. The R-phase transformation is well suited for actuation applications due to its small hysteresis, and the modest required thermal operating range is compatible with *in vivo* use.

Chapter 2 and Chapter 3 both propose new constitutive models to describe the behavior of interest. By selecting a specific operating range, and due to the negligible hysteresis of the material in this range, it was possible to formulate models that are considerably less complex than conventional SMA models. It should be noted, however, that conventional models are generally more widely applicable, and many of the recently proposed models also attempt to include the influence of crystallographic texture, non-proportional loading and tension-

compression asymmetry. These aspects have been neglected in the presented models, which are formulated with the intent to balance between accuracy and suitability for design optimization. The models proposed in Chapter 2 and Chapter 3 represent different compromises between these two aspects. Both models result in a correct representation of known one-dimensional experimental data, and the main difference is found in the way the one-dimensional formulation is generalized to a three-dimensional setting. In comparison, the approach taken in Chapter 2 is computationally simpler, it results in an explicit formulation, but the isochoric nature of the R-phase transformation is not strictly accounted for. Chapter 3 extends the approach followed in the formulation of this model, and includes the known aspects of the transformation. This however leads to a more complex model, which in the plane stress case involves an implicit equation, that needs to be solved iteratively. However, considering its less approximate nature, this model has subsequently been used for optimization in the rest of this thesis.

A key property of both models is their history-independence, which is accomplished through neglecting the small hysteresis of the R-phase transformation. This no longer makes it necessary to use internal variables to keep track of the state of the material. Essentially, in this way both models approximate the considered material as a temperature-dependent nonlinear elastic solid. The advantage of this history-independence is that it significantly simplifies sensitivity analysis. For history-dependent materials, the sensitivity analysis must account for the entire loading path, hence time integration of sensitivities is necessary. However, in the present history-independent case, sensitivities can be computed in one step, after completing the full analysis. This advantage has been essential for the topology optimization in Chapter 6 and the sensitivity analysis and gradient-based shape optimization of Chapters 7, 8 and 9.

> Therefore, in summary, it can be concluded that the specific SMA models introduced in this thesis are very well suited for design optimization of the class of SMA structures of interest, and their suitability for design optimization forms a key element in enabling the realized optimization techniques.

Furthermore, Chapter 3 introduced a novel design for an active catheter, which by means of integration is expected to be more reliable and easier to assemble than the designs used in recent prototypes. The design was simulated using the formulated SMA material model, and showed promising performance. Although the transformation strain offered by the R-phase transformation is limited, it has been demonstrated that due to the structural design of the active catheter, still a respectable degree of bending could be achieved. Nonetheless, further improvement proved to be possible when shape optimization was applied, as shown in Chapter 9. This case clearly illustrates the potential for design optimization for this type of SMA applications involving complex structures.

10.1.2 Direct shape optimization

Chapter 4 demonstrates the viability and effectiveness of shape optimization of an SMA gripper, using the response-surface-based multi-point approximation method (MAM, Toropov *et al.*, 1993a). This is a direct method, which implies that it does not require design sensitivity information. This is generally not the most efficient approach, but it has the advantages that it is robust, requires no sensitivity analysis and has the property to smoothen noisy responses. Parallel computing has been used to reduce the duration of the optimization procedure to a practical level. The shape optimization of the gripper was based on a clearly formulated optimization problem, and finite element modeling was used to evaluate the performance of designs. This versatile and generic approach applies to many devices, and all kinds of objectives and constraints can be accounted for, in contrast to more limited analytic or heuristic methods previously reported in literature.

Chapter 5 extends the gripper optimization considered in Chapter 4 with the consideration of bounded-but-unknown uncertainties. These are uncertainties in e.g. loading conditions or material properties, characterized by their bounds. For innovative products or in an early design stage, precise statistical data on uncertainties is generally not available. Probabilistic approaches therefore cannot be used. However, often it is possible to define bounds on the values of the variable parameters, based on expert opinions. Considering the effect of uncertainties is important in order to generate robust and reliable designs. Parallel computing has been used to make the optimization feasible for practical applications. Instead of rigorous anti-optimization, an alternating technique was used that yielded substantial efficiency improvements. It was also shown, that the availability of sensitivity information potentially could improve the efficiency even more, but this has not yet been demonstrated for an SMA problem.

In order for the gripper design to remain feasible, a decrease in performance of 15% was observed compared to the deterministic design. In general, when accounting for uncertainties, the performance of designs will decrease. However, in return the optimized design will perform reliably even under varying conditions. This approach to optimization is believed to be of great practical value, as in many realistic design situations uncertainty and variability cannot be ignored.

> It can be concluded that deterministic and non-deterministic shape optimization of SMA structures of realistic complexity is feasible, even without the use of gradient information. However, in order to reduce the duration of the optimization process to a practical level, parallel computing facilities and optimization algorithms that can exploit these are indispensable.

10.1.3 Topology optimization

A procedure for topology optimization of SMA structures has been described in Chapter 6. In fact, this is the first application of topology optimization techniques to SMA structural design. Given the facts that topology optimization relies on adjoint sensitivity analysis, and that this adjoint sensitivity analysis is very cumbersome for history-dependent models (Kleiber *et al.*, 1997), the models described in Chapters 2 and 3 turn out to be very well suited for use in SMA topology optimization. To realize SMA topology optimization, use has been made of a novel approach to parameterize the design, the so-called Element Connectivity Parameterization (ECP, Yoon and Kim, 2005). Although the ECP approach used in Chapter 6 does not require the sensitivities of the material model itself, its history-independence still is essential. The proposed topology optimization procedure has been demonstrated on problems involving planar SMA structures under constant loading, and actuated by uniform temperature changes, but the proposed method applies to more general problems as well.

The ECP approach differs from conventional approaches to topology optimization in the aspect that not the material properties of elements, but the connections between elements are controlled by the design variables. This approach offers various advantages, such as improved stability of geometrically nonlinear calculations and absence of arbitrariness in material property interpolation functions. Furthermore, as already noted, sensitivity analysis is straightforward and does not require differentiation of the underlying material model. Unlike earlier versions of the element connectivity parameterization concept (e.g. Yoon and Kim, 2004), the improved ECP approach used for SMA topology optimization does not increase the number of degrees of freedom of the system. Hence the solution time and complexity remains comparable to that of a standard finite element model of similar dimensions.

A heuristic algorithm was developed to improve the robustness of the SMA finite element analyses during topology optimization. The drastic design changes possible in the topology optimization procedure combined with the strongly nonlinear behavior of the SMA structure require an adaptive incremental-iterative solution strategy. Except for one case where the topology optimization resulted in an instable structure, the solution strategy proved to be effective and reliable. It was found in that particular case, that exploiting instability and buckling phenomena resulted in a large improvement of performance, and this is in fact an interesting direction for further research.

Nodal design variables were used to counteract the tendency of the optimization problem to generate artificial solutions containing checkerboard patterns or one-node hinges. In contrast, conventional heuristic filtering approaches to achieve regularization of the design problem turned out to be ineffective in case of SMA structures. The reason for this is the fact that sensitivities in the latter case can have positive or negative values, while in case of compliance minimization the sensitivities are always positive. The use of node-based instead of element-based

design variables resulted in a slight increase in the total number of design variables, however the used MMA optimizer was able to deal with this without any problems.

It was found that mesh density and load intensity strongly affected the topology optimization results. The latter is a result of the nonlinearity of the SMA material behavior, which results in different optimal designs for different load levels. The fact that increasing mesh densities lead to designs with finer details is common to topology optimization, and several remedies have been proposed to impose a minimum lengthscale (Guest *et al.*, 2004, Poulsen, 2003, Zhou *et al.*, 2001). Furthermore, results were found to depend on the initial design used, which is an indication that the optimization problem is nonconvex, and has multiple local optima. Nonlinearity of SMA model and the associated nonconvexity of the SMA design problem are a likely cause for the existence of local minima. Hence, achieving global optimality remains a fundamental challenge in topology optimization of nonlinear problems. However, this does not mean that the proposed optimization technique has no practical value. The technique is able to generate original and effective design concepts with very little input from the designer, apart from specifying the problem. Furthermore, as has been shown by an example, the proposed method is able to significantly improve the performance of existing designs.

> In conclusion, the combination of a dedicated SMA model, adaptive nonlinear solution techniques, Element Connectivity Parameterization, nodal design variables and the Method of Moving Asymptotes has enabled topology optimization of SMA structures. Obtaining globally optimal topologies however remains a challenge.

10.1.4 Gradient-based shape optimization

The use of sensitivity information can drastically improve the efficiency of optimization procedures. However, this improvement can only be realized when the sensitivities can be computed with little additional computational effort. For this reason, Chapter 7 reviewed the options for sensitivity analysis for the SMA model introduced in Chapter 3, and described the formulation of restarted finite difference and semi-analytical sensitivities in detail, for a general electro-thermo-mechanical setting. The history-independent nature of the constitutive model allowed for efficient sensitivity analysis. The implementation of the semi-analytical approaches turned out to be considerably more involved than that of the finite difference alternative, because of the coupling between the electrical, thermal and mechanical analysis and the implicit nature of the SMA material model. In order to evaluate the required coupling terms, both analytical and finite difference approaches were studied. No clear difference in performance or accuracy was observed, however the analytical approach required considerably more implementation effort.

The same observation was made for the calculation of sensitivities of quantities derived from the state variables, such as e.g. equivalent strains. After evaluation of the state vector sensitivity, in the semi-analytical approach a long analytical calculation is required to evaluate the equivalent strain sensitivity, which in addition needs to be implemented separately for every derived response quantity of interest. Hence, the rather complex and implicit nature of the material model makes that the derivation and implementation of sensitivities of these derived quantities requires a significant effort. In contrast, in the restarted finite difference case, the perturbed solution is available. Therefore, evaluating sensitivities of derived quantities is straightforward, since the same finite difference approach can be applied as is used for the state variables. Of course, the accuracy of finite difference sensitivities also depends on the design perturbation used, but this also holds for semi-analytical formulations.

For a studied SMA gripper case, it was found that stable sensitivity values were obtained for relative design perturbations of 10^{-7}. Furthermore, the more elaborate refined semi-analytical approach showed no clear accuracy improvement, and the range for which stable sensitivity values were obtained was more or less similar for all considered methods. For the considered problem, the computational cost of a complete evaluation including restarted finite difference sensitivity analysis was only slightly higher (16%) than an evaluation involving semi-analytical sensitivity analysis. When also considering its relatively modest implementation effort, it was concluded that the restarted finite difference approach is an attractive option for the present electro-thermo-mechanical SMA model.

Chapters 8 and 9 subsequently demonstrate the use of design sensitivities in representative SMA design optimization studies, using gradient-based optimization methods. Chapter 8 reports on shape optimization of an SMA miniature gripper, and convincingly demonstrates the advantage of gradient-based methods over direct approaches. An approximately twenty-fold reduction in the number of function evaluations was realized, through the use of gradient-based optimizers (MMA and SQP), in comparison to optimization using the response-surface based MAM method. Next to this, the consistency and accuracy of the results obtained using the gradient-based methods clearly was superior. In all fairness, it should be noted that the quality of the MAM results can possibly be improved by spending further effort on finding better settings for the optimizer parameters, for this specific problem. It should also be noted that implementation of sensitivity analysis requires additional effort, and that for practical cases with only few design variables, this effort may not be justified. Particularly, when sufficient parallel computing facilities are available, the versatility and ease of use of direct approaches that can benefit from parallellization, such as the MAM method, could be preferred over the efficiency offered by gradient-based methods.

Chapter 9 finally demonstrated the use of a gradient-based optimization method on the SMA active catheter design problem, introduced earlier in Chapter 3. This chapter illustrated, that with the help of efficient design optimization techniques, different design concepts can be evaluated and compared in a systematic way.

Considerable improvements in bending performance were realized, in comparison to the baseline design studied in Chapter 3. The sensitivity analysis was performed using finite differences, in combination with fast reanalysis. By restarting perturbed analyses from the nominal solution and using modified Newton iterations, the computational cost of the sensitivity analysis was minimized. A new correction was applied to the multi-point constraints present in the model, which was shown to improve the accuracy of the sensitivities. The gradient-based SQP optimizer was used to perform the optimization, and compared to the direct MAM optimizer. Although the latter was able to complete the design optimization process in nearly the same timespan, given sufficient parallel computing resources, clearly also in this case gradient-based design optimization proved to be the most efficient way to optimize designs of SMA structures.

> Gradient-based optimization techniques can most efficiently optimize designs of SMA structures, and the required cost-effective sensitivity analysis is enabled by the SMA models proposed in this thesis. Sensitivity analysis by means of finite differences in combination with fast reanalysis offers a good compromise between efficiency and implementation effort.

10.1.5 Overall conclusion

The research presented in this thesis clearly illustrates the possibilities that design optimization techniques offer for improving and enhancing design procedures of SMA structures. Improved design capabilities are essential to further exploit the potential of SMAs in applications such as instruments for minimally invasive procedures or micro-actuators. Shape optimization was demonstrated with and without the use of sensitivity information, and with and without accounting for uncertainties in loading conditions and/or material properties. Furthermore, in order to generate innovative shapes and design concepts, a topology optimization procedure has also been developed for the design of SMA structures. This technique can be applied in the early stages of a design process, where design decisions have the largest impact. Being able to utilize optimization already at that stage is certainly expected to have a significant impact on the quality and performance of the final designs. Substantial performance improvements were demonstrated in various design problems, obtained using the structured and systematic methods of design optimization. It is therefore expected that designers of future SMA devices will increasingly make use of the demonstrated numerical modeling and optimization techniques, in order to improve the effectiveness and reliability of their designs.

However, it should be mentioned that a key factor in enabling these design optimization procedures has been the history-independent SMA model used in this study, targeted specifically at the R-phase transformation in NiTi. This makes that parts of the presented work can not be directly extended to design cases of

SMA structures involving different material behavior and significant hysteresis. Particularly, the sensitivity analysis and topology optimization procedures relied on the history-independence of the problem, and generalizing these to history-dependent cases is still a challenge. However, the direct methods shown for deterministic and uncertainty-based shape optimization can directly be applied to history-dependent problems, and sensitivity analysis for history-dependent problems, although more involved, is certainly possible. Chapter 8 and 9 clearly show the benefits of exploiting gradient information, for problems within the scope of this thesis. Hence, possibly also for history-dependent problems, gradient-based design optimization could be of interest, although the increased cost of the sensitivity analysis will reduce its computational advantages to some extent. The following section will elaborate further on these and other future directions.

10.2 Future directions

Although this thesis aimed to cover a wide range of aspects related to design optimization of SMA structures, there are still various topics that could be explored further. Also, many of the techniques considered in this research showed possibilities for further development. An overview of several interesting directions for future work is given here, divided between activities aimed at validation and realization, and opportunities and suggestions to further extend the presented approaches.

10.2.1 Validation and realization

The following suggestions can be made with regard to further validation or practical realization of aspects of this research:

- The material models presented in Chapter 2 and Chapter 3 are based on one-dimensional experimental data, in combination (in case of Chapter 3) with knowledge of the underlying physical characteristics of the R-phase transformation. Further experimental validation and assessment of the accuracy of these models is recommended.
- In relation to the previous point, there is a clear opportunity for more extensive experimental investigation of R-phase-transformation-based material behavior. Little is know of, e.g., the effect of crystallographic texture, nonproportional loading, complex multi-dimensional loading patterns, complex thermomechanical loading, and the extent and implications of tension-compression asymmetry. Findings could be used to further refine the material model, as long as this does not unnecessarily complicate the design optimization.
- Strong indications exist that shifting of transformation temperatures of the material studied in this thesis is possible, which allows its application for in

vivo medical applications. However, this has not been experimentally verified. Furthermore, it is likely that changing alloy composition to shift transformation temperatures also affects the shape of the stress-strain curves. Further experimental studies are required to explore this aspect in more detail.

- The active catheter design presented in Chapter 3, and the optimized designs found in Chapter 9, could be worked out in more detail and realized in practice, using e.g. laser cutting or chemical etching to manufacture the structure from an SMA tube. Once fabricated, its performance could be evaluated and the possibility to steer and even propel the catheter by means of crawling motions could be assessed.

- Although a careful study has been made in Chapter 3 on what temperature and heat output is tolerable inside arteries, based on information that currently is available in literature, it is advisable to in addition carry out a clinical study to collect more experimental data and knowledge on how living tissues react to thermal loading. Furthermore, for an effective design of an SMA active catheter, it is important to determine key properties such as the thermal convection coefficient h between the catheter and the blood more accurately, through in vivo measurements. Tangwongsan *et al.* (2004) already demonstrated that this is possible, by performing such measurements in living pigs. Compared to their results, the values used in this research were adequate, but slightly conservative. More detailed and precise information for the human case would certainly be useful.

- Furthermore, regarding all optimal designs that have been described in the chapters focusing on shape or topology optimization, the generated designs could actually be realized and their performance could be evaluated and compared to the predictions made by the models.

10.2.2 Extensions

The scope of the research presented in this thesis has been clearly defined in Chapter 1. It has been necessary to narrow down the subject of this work, in order to achieve a sufficient focus and depth. However, here in this final chapter there is an opportunity to look beyond the outlined research area, and consider the possibilities and future extensions that are indicated by the obtained results.

- In this research, dynamic effects have been ignored, since in the intended applications the dynamic performance of devices was not considered to be a critical aspect. However, there may be other applications where in fact the dynamic response of devices is very relevant. In that case, the dynamic and transient behavior of SMA actuators can no longer be ignored, and has

to be described by appropriate models. When combined with the demonstrated optimization techniques, it will become possible not only to optimize actuators for stroke or force, but also for e.g. their response or cycle time.

Note, however, that sensitivity analysis for transient problems is more involved than for the case considered in this work, and that proper modeling of thermal and mechanical dynamics of SMA materials is a challenging task in itself.

- The optimization approaches presented in this thesis could in principle also be applied to SMA design problems involving more general SMA behavior, such as the martensite transformation in NiTi, which offers a much larger actuation strain (up to 7%), but also involves considerable hysteresis. Note here, that since the hysteresis is no longer negligible, a history-independent formulation as used for the R-phase transformation is no longer adequate. Hence, the associated advantages of using a history-independent model with respect to sensitivity analysis will also vanish. Still, it is expected that also for this case the use of design optimization techniques offers considerable benefits.

- Regarding the optimization under uncertainty discussed in Chapter 5, the combination of the Asymptotic approach with the Cycle-based Alternating technique for optimization under Bounded-But-Unknown uncertainties showed great promise in a small analytical beam example. It is interesting to further study the performance of this approach on a problem of realistic complexity, such as the presented SMA miniature gripper. Sensitivities have been derived and implemented, as described in Chapter 7, therefore basically the key ingredients are available.

- Although Chapter 6 already showed quite a number of topology optimization results, from another perspective the topic has only been touched upon in this research, and several interesting challenges remain:

  - To avoid the mesh-dependency of results, measures to control the minimum member size in the generated structures could be included.
  - The topology optimization formulation could be extended to the more general electro-thermo-mechanical case, to deal with Joule-heated actuators with nonuniform temperature distributions.
  - The topology optimization procedure could be extended to shell structures and three-dimensional solid structures. This would allow application of topology optimization to the active catheter design problem studied in Chapter 3 and Chapter 9, and possibly totally different design concepts may emerge from this.
  - In one case, the topology optimization failed because of buckling instabilities, therefore measures to avoid these instabilities are of interest.

This aspect has already been the subject of a recent study by Kemmler *et al.* (2005). Furthermore, once the stability problems have been solved, it could be attractive to try to exploit buckling and postbuckling effects, as a means to increase the displacement range of the generated SMA actuators. Also the generation of bistable structures is of interest.

- The accuracy of the analysis used in topology optimization could be improved, by using a different discretization and geometry description technique. The jagged edges, combined with the fact that SMA actuator topology optimization often tends to form very thin, detailed and curved structures makes that the accuracy of the analysis is suboptimal. An improvement could be found in the use of recently proposed *level-set techniques* (see e.g. De Ruiter, 2005, Kwak and Cho, 2005, Liu *et al.*, 2005, Wang and Wang, 2006, Wang *et al.*, 2004), which allow the separation of geometrical description and the analysis mesh, and which are able to generate structures with smooth boundaries.
- Furthermore, to ensure that designs remain within the range of applicability of the constitutive model, strain constraints could be included in the topology optimization problem formulation. However, such constraints require special measures to handle the optimization, as they are known to lead to hard-to-find singular optima and spoil the convergence characteristics of topology optimization problems (Duysinx and Bendsoe, 1998, Guo *et al.*, 2001, Stolpe and Svanberg, 2003). Possibly, the constraint aggregation approach used for the microgripper case (Chapter 5, Chapter 8) could be a viable solution, particularly when combined with recently proposed adaptive approaches for the penalty parameter (see Martins and Poon, 2005).

  Note, however, that on the other hand, the designs resulting from topology optimization appeared to avoid excessive strains. It seems that it is more advantageous to use the material in a moderate strain range, in order to improve the performance of the designs. Therefore, the question whether there is a real need to include strain constraints in topology optimization of SMA actuators should also be addressed critically.
- Nonconvexity of topology optimization problems involving nonlinear models leads to possible convergence to local minima. Measures to reduce this nonconvexity, or to somehow escape local optima could be studied, in order to decrease the chance that optimization processes converge prematurely to inferior designs. Possibly also the interpolation function used provides a sort of design-dependent scaling of the problem, and this may affect its convexity properties.
- Considering the mentioned extension of this work to more general history-dependent SMA models, a fundamental challenge is to apply topology optimization to problems involving such models. Topology

optimization, with usually thousands of design variables, relies heavily on adjoint sensitivity analysis. However, for history-dependent problems, such sensitivity analysis is much more involved than for the history-independent case considered in this thesis. Therefore, application of topology optimization to history-dependent problems still presents a challenge.

- Regarding the sensitivity analysis presented in Chapter 7, it could be considered to also formulate and implement an adjoint sensitivity analysis procedure for the electro-thermo-mechanical case. This would be of interest for problems where the number of design variables is significantly larger than the number of responses.

- Slightly outside the scope of the present thesis, the recommendation could be made to include more adaptive algorithms in the Multi-point Approximation Method used in Chapter 4, 5, 7 and 8, or to include a procedure that is able to determine important properties of a problem prior to the actual optimization process, which would subsequently be able to adjust the optimizer settings accordingly. This would relieve the user from the task of finding proper settings for the many parameters of this optimizer, and it would also allow the algorithm to change its behavior during the optimization process, and consequently to converge faster.

- Finally, based on the findings and the analysis results of the optimized active catheter designs given in Chapter 9, it is of interest to create and optimize modified design concepts. It is expected that this will lead to further performance improvements. The modeling might also be further refined, for example by using solid instead of shell elements, for designs with relatively large tube wall thicknesses. Note, however, that for a more realistic active catheter design study, the formulation of the design problem should be based on the requirements of the actual clinical procedure for which the instrument is designed. Next to its dimensions and active bending performance, also aspects such as its bending stiffness and buckling resistance might be of interest. However, all these design aspects can be included in the formulation of the design optimization problem.

The length of these lists of suggestions for future work indicates that still enough challenges remain for further research on this topic. Hopefully, the design optimization results from this thesis provide inspiration to confront these challenges, and to make further contributions to improve the procedures for designing SMA structures.

Appendix A

Derivation of 3-D Tangent Operator

The tangent operator $d\boldsymbol{\sigma}/d\boldsymbol{\varepsilon}$ for the three-dimensional constitutive model can be found by differentiating the stress-strain relation with respect to the strain. This three-dimensional constitutive relation is given by:

$$\boldsymbol{\sigma} = (K\mathbf{K} + 2G\mathbf{G})(\boldsymbol{\varepsilon} - \kappa\mathbf{G}\boldsymbol{\varepsilon}) = (K\mathbf{K} + 2\hat{G}\mathbf{G})\boldsymbol{\varepsilon}. \tag{A.1}$$

Using G instead of $\hat{G}$ to simplify the notation, it follows that:

$$\frac{d\boldsymbol{\sigma}}{d\boldsymbol{\varepsilon}} = K\mathbf{K} + 2G\mathbf{G} + 2\mathbf{G}\boldsymbol{\varepsilon}\frac{dG}{d\varepsilon_e}\frac{d\varepsilon_e}{d\boldsymbol{\varepsilon}}. \tag{A.2}$$

The derivative of the effective strain ε_e to the strain itself is found from its definition:

$$\varepsilon_e^2 = \frac{2}{3}\boldsymbol{\varepsilon}^T\mathbf{G}\boldsymbol{\varepsilon} \quad \Rightarrow \quad \varepsilon_e = \sqrt{\frac{2}{3}\boldsymbol{\varepsilon}^T\mathbf{G}\boldsymbol{\varepsilon}} \quad \Rightarrow \quad \frac{d\varepsilon_e}{d\boldsymbol{\varepsilon}} = \frac{2}{3}\frac{1}{\varepsilon_e}\boldsymbol{\varepsilon}^T\mathbf{G}. \tag{A.3}$$

This leads to the final expression of the tangent operator in the three-dimensional case:

$$\frac{d\boldsymbol{\sigma}}{d\boldsymbol{\varepsilon}} = K\mathbf{K} + 2G\mathbf{G} + \frac{4}{3\varepsilon_e}\frac{dG}{d\varepsilon_e}\mathbf{G}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{G}. \tag{A.4}$$

The term $dG/d\varepsilon_e$ is easily found from the expression for G:

$$G = \frac{3K}{9K - A}\left(A + \frac{B}{\varepsilon_e}\right) \Rightarrow \frac{dG}{d\varepsilon_e} = -\frac{3KB}{9K - A}\frac{1}{\varepsilon_e^2}. \tag{A.5}$$

The tangent operator is symmetric, since $\mathbf{K}$, $\mathbf{G}$ and also the product $\mathbf{G}\boldsymbol{\varepsilon}(\mathbf{G}\boldsymbol{\varepsilon})^T$ are symmetric.

Appendix B

Shear Modulus/Effective Strain Relation in Plane Stress Case

The starting point for defining the equations to solve G and ε_e numerically is given by:

- the relation between G and ε_e obtained from the one-dimensional behavior:

$$G = \frac{3K}{9K - A_i(T)} \left(A_i(T) + \frac{B_i(T)}{\varepsilon_e} \right) = C_i(T) + \frac{D_i(T)}{\varepsilon_e}, \tag{B.1}$$

  where $C_i(T)$ and $D_i(T)$ are introduced for convenience, and the index $i = 1, 2, 3$ refers to the corresponding part of the piecewise linear approximation to the one-dimensional stress-train curve used in the material model,

- the expression for the effective strain in the plane stress case:

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

- and the definition of α:

$$\alpha = \frac{2G - 3K}{4G + 3K}. \tag{B.3}$$

In order to solve this set of equations, Equation B.1 is rewritten to define a function Z:

$$Z = C_i(T) + \frac{D_i(T)}{\varepsilon_e} - G = 0. \tag{B.4}$$

The effective strain is eliminated by substituting Equation B.2 in Equation B.4, resulting in a nonlinear equation $Z(G) = 0$. To solve this equation with Newton iterations, the derivative of Z with respect to G is required. Differentiation gives:

$$\frac{dZ}{dG} = -\frac{4(G - C_i(T))K(\varepsilon_{xx} + \varepsilon_{yy})^2(2\alpha - 1)}{(4G + 3K)^2\varepsilon_e^2} - 1. \tag{B.5}$$

Note, α is to be replaced by Equation B.3, but is maintained in this expression for readability. Convenient starting points for the iterations are the transition points of the piecewise linear model. These transition points differ from the transition strains ε_1, ε_2 used in the one-dimensional formulation, since in general $\varepsilon_{xx} \neq \varepsilon_e$. Using the relations between ε_{xx} and ε_e derived before, it is found that the transition points in terms of the effective strain ε_e are given by:

$$\varepsilon_{e(1)}(T) = \frac{9K - E_A(T)}{9K}\varepsilon_1(T), \tag{B.6}$$

$$\varepsilon_{e(2)}(T) = \varepsilon_{e(1)}(T) + \frac{9K - E_T(T)}{9K}\Delta(T). \tag{B.7}$$

These transition points subdivide the effective strain range in three intervals. In the austenite interval, i.e. $\varepsilon_e \leq \varepsilon_{e(1)}$, no iterations are required since $D_1 = 0$ and $G = C_1$. Here only ε_e needs to be evaluated (Equation B.2), and if indeed $\varepsilon_e \leq \varepsilon_{e(1)}$ the procedure is finished. In the second and third interval, iterations are necessary, and likewise the subsequent solution is checked against the relevant transition strains.

Appendix C

Plane Stress Constitutive Tangent Operator

Two approaches are described in this appendix for the derivation of the constitutive tangent operator in the plane stress case. The first is based on reduction of the 3-D tangent operator to the plane stress case, the second on differentiation of the stress-strain relation in the plane stress setting. The results are identical.

C.1 Method 1: based on the 3-D tangent operator

The tangent operator in the three-dimensional case, which is taken as a starting point for this derivation of the tangent operator in plane stress, is given by:

$$\frac{d\boldsymbol{\sigma}}{d\boldsymbol{\varepsilon}} = K\mathbf{K} + 2G\mathbf{G} + 2\mathbf{G}\boldsymbol{\varepsilon}\frac{dG}{d\varepsilon_e}\frac{d\varepsilon_e}{d\boldsymbol{\varepsilon}} = K\mathbf{K} + 2G\mathbf{G} + \frac{4}{3\varepsilon_e}\frac{dG}{d\varepsilon_e}\mathbf{G}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{G}. \tag{C.1}$$

Combining this with the the relation between G and ε_e, obtained from the one-dimensional behavior and given by

$$G = \frac{3K}{9K - A}\left(A + \frac{B}{\varepsilon_e}\right) = C + \frac{D}{\varepsilon_e}, \tag{C.2}$$

yields

$$\frac{d\boldsymbol{\sigma}}{d\boldsymbol{\varepsilon}} = K\mathbf{K} + 2G\mathbf{G} - \frac{4D}{3\varepsilon_e^3}\mathbf{G}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{G} = K\mathbf{K} + 2G\mathbf{G} - \frac{4(G-C)^3}{3D^2}\mathbf{G}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{G} = \mathbf{T}. \tag{C.3}$$

In Equation C.2, C and D are introduced for convenience. These are constants in this setting, which depend on the temperature. In the case of vector notation, as used here, the tangent operator takes the form of a square matrix. In this appendix, this tangent matrix will be denoted by $\mathbf{T}$, and its components by $T_{[i,j]}$.

These expressions were derived for stress and strain vectors given by

$$\boldsymbol{\sigma} = (\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sqrt{2}\sigma_{xy}, \sqrt{2}\sigma_{yz}, \sqrt{2}\sigma_{zx})^T \tag{C.4}$$

and

$$\boldsymbol{\varepsilon} = (\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{zz}, \sqrt{2}\varepsilon_{xy}, \sqrt{2}\varepsilon_{yz}, \sqrt{2}\varepsilon_{zx})^T. \tag{C.5}$$

In the plane stress case, reduced vectors can be used: $\boldsymbol{\sigma} = (\sigma_{xx}, \sigma_{yy}, 0, \sqrt{2}\sigma_{xy})^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{zz}, \sqrt{2}\varepsilon_{xy})^T$. However, the vectors commonly used in finite element implementations for plane stress and shell elements are $\boldsymbol{\sigma} = (\sigma_{xx}, \sigma_{yy}, \sigma_{xy})^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_{xx}, \varepsilon_{yy}, \gamma_{xy})^T$. This means it is necessary to eliminate the row and column associated with ε_{zz} from the matrix, and to apply other transformations to convert the expressions based on the original stress-strain vectors to the situation using the new vectors. First, starting from the reduced version of Equation C.1, the transverse strain component ε_{zz} is eliminated, which results in the expressions given in Equation C.6 till Equation C.9. Next, the expression in Equation C.9 is transformed to a new set of stress and strain vectors as used in finite element formulations, which yields the final symmetric plane stress tangent operator given in Equation C.11.

It can be seen from this last expression that when the original tangent operator matrix $\mathbf{T}$, that was used as a starting point, was symmetric, this reduced tangent operator for the plane stress case will also be symmetric, when using the specified stress and strain vectors are used. For the present material model, this is the case, since the tangent operator given by Equation C.1 indeed is symmetric.

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ 0 \\ \sqrt{2}\sigma_{xy} \end{Bmatrix} = \begin{bmatrix} T_{[1,1]} & T_{[1,2]} & T_{[1,3]} & T_{[1,4]} \\ T_{[2,1]} & T_{[2,2]} & T_{[2,3]} & T_{[2,4]} \\ T_{[3,1]} & T_{[3,2]} & T_{[3,3]} & T_{[3,4]} \\ T_{[4,1]} & T_{[4,2]} & T_{[4,3]} & T_{[4,4]} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix} \Rightarrow \tag{C.6}$$

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ 0 \\ \sqrt{2}\sigma_{xy} \end{Bmatrix} = \begin{bmatrix} T_{[1,1]} & T_{[1,2]} & T_{[1,3]} & T_{[1,4]} \\ T_{[2,1]} & T_{[2,2]} & T_{[2,3]} & T_{[2,4]} \\ -\frac{T_{[3,1]}}{T_{[3,3]}} & -\frac{T_{[3,2]}}{T_{[3,3]}} & -1 & -\frac{T_{[3,4]}}{T_{[3,3]}} \\ T_{[4,1]} & T_{[4,2]} & T_{[4,3]} & T_{[4,4]} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix} \Rightarrow \tag{C.7}$$

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ 0 \\ \sqrt{2}\sigma_{xy} \end{Bmatrix} = \begin{bmatrix} T_{[1,1]} - T_{[1,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[1,2]} - T_{[1,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & 0 & T_{[1,4]} - T_{[1,3]}\frac{T_{[3,4]}}{T_{[3,3]}} \\ T_{[2,1]} - T_{[2,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[2,2]} - T_{[2,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & 0 & T_{[2,4]} - T_{[2,3]}\frac{T_{[3,4]}}{T_{[3,3]}} \\ 0 & 0 & -1 & 0 \\ T_{[4,1]} - T_{[4,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[4,2]} - T_{[4,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & 0 & T_{[4,4]} - T_{[4,3]}\frac{T_{[3,4]}}{T_{[3,3]}} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix} \Rightarrow \tag{C.8}$$

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sqrt{2}\sigma_{xy} \end{Bmatrix} = \begin{bmatrix} T_{[1,1]} - T_{[1,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[1,2]} - T_{[1,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & T_{[1,4]} - T_{[1,3]}\frac{T_{[3,4]}}{T_{[3,3]}} \\ T_{[2,1]} - T_{[2,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[2,2]} - T_{[2,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & T_{[2,4]} - T_{[2,3]}\frac{T_{[3,4]}}{T_{[3,3]}} \\ T_{[4,1]} - T_{[4,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[4,2]} - T_{[4,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & T_{[4,4]} - T_{[4,3]}\frac{T_{[3,4]}}{T_{[3,3]}} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix} \tag{C.9}$$

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{Bmatrix} = \begin{bmatrix} T_{[1,1]} - T_{[1,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[1,2]} - T_{[1,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & T_{[1,4]} - T_{[1,3]}\frac{T_{[3,4]}}{T_{[3,3]}} \\ T_{[2,1]} - T_{[2,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[2,2]} - T_{[2,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & T_{[2,4]} - T_{[2,3]}\frac{T_{[3,4]}}{T_{[3,3]}} \\ \frac{1}{\sqrt{2}}\left(T_{[4,1]} - T_{[4,3]}\frac{T_{[3,1]}}{T_{[3,3]}}\right) & \frac{1}{\sqrt{2}}\left(T_{[4,2]} - T_{[4,3]}\frac{T_{[3,2]}}{T_{[3,3]}}\right) & \frac{1}{\sqrt{2}}\left(T_{[4,4]} - T_{[4,3]}\frac{T_{[3,4]}}{T_{[3,3]}}\right) \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix} \Rightarrow \tag{C.10}$$

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{Bmatrix} = \begin{bmatrix} T_{[1,1]} - T_{[1,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[1,2]} - T_{[1,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & \frac{1}{\sqrt{2}}\left(T_{[1,4]} - T_{[1,3]}\frac{T_{[3,4]}}{T_{[3,3]}}\right) \\ T_{[2,1]} - T_{[2,3]}\frac{T_{[3,1]}}{T_{[3,3]}} & T_{[2,2]} - T_{[2,3]}\frac{T_{[3,2]}}{T_{[3,3]}} & \frac{1}{\sqrt{2}}\left(T_{[2,4]} - T_{[2,3]}\frac{T_{[3,4]}}{T_{[3,3]}}\right) \\ \frac{1}{\sqrt{2}}\left(T_{[4,1]} - T_{[4,3]}\frac{T_{[3,1]}}{T_{[3,3]}}\right) & \frac{1}{\sqrt{2}}\left(T_{[4,2]} - T_{[4,3]}\frac{T_{[3,2]}}{T_{[3,3]}}\right) & \frac{1}{2}\left(T_{[4,4]} - T_{[4,3]}\frac{T_{[3,4]}}{T_{[3,3]}}\right) \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \gamma_{xy} \end{Bmatrix}. \tag{C.11}$$

C.2 Method 2: based on the plane stress equations

Starting point for the derivation of the tangent operator are the following equations:

- The stress-strain relation in the plane stress setting:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sqrt{2}\sigma_{xy} \end{Bmatrix} = \left(K \begin{bmatrix} 1+\alpha & 1+\alpha & 0 \\ 1+\alpha & 1+\alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} + \ldots \right. \\ \left. \frac{2G}{3} \begin{bmatrix} 2-\alpha & -(1+\alpha) & 0 \\ -(1+\alpha) & 2-\alpha & 0 \\ 0 & 0 & 3 \end{bmatrix} \right) \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \sqrt{2}\varepsilon_{xy} \end{Bmatrix}, \tag{C.12}$$

  where $\alpha = (2G - 3K)/(4G + 3K)$.

- The definition of the effective strain:

$$\varepsilon_e^2 = \frac{4}{9}(\alpha^2 - \alpha + 1)(\varepsilon_{xx}^2 + \varepsilon_{yy}^2) + \frac{4}{9}(2\alpha^2 - 2\alpha - 1)\varepsilon_{xx}\varepsilon_{yy} + \frac{4}{3}\varepsilon_{xy}^2. \tag{C.13}$$

- The relation between G and ε_e obtained from the one-dimensional behavior:

$$G = \frac{3K}{9K - A}\left(A + \frac{B}{\varepsilon_e}\right) = C + \frac{D}{\varepsilon_e}. \tag{C.14}$$

  where C and D are introduced for convenience. These are constants in this setting, which depend on the temperature.

The tangent operator basically is the derivative of the stress with respect to the strain. However, the stress as given in Equation C.12 is not only a direct function of the strain, but also of α (which depends on G) and G, which in turn depends on ε_e, where ε_e is a function of the strain components (Equation C.13). So, in vector index notation, the tangent operator is basically given by:

$$\frac{d\sigma_j}{d\varepsilon_i} = \frac{\partial \sigma_j}{\partial \varepsilon_i} + \left(\frac{\partial \sigma_j}{\partial \alpha}\frac{d\alpha}{dG} + \frac{\partial \sigma_j}{\partial G}\right)\frac{\partial G}{\partial \varepsilon_i}. \tag{C.15}$$

The terms $\partial G/\partial \varepsilon_i$ are not obvious to obtain, as an explicit relation is not present. This difficulty can be resolved as follows: the shear modulus G is linked to the strains through the effective strain, in the equation that relates G to the one-dimensional stress-strain data (Equation C.14), given by the general expression $G = C + D/\varepsilon_e$. This function is implicit, because also the expression for ε_e involves G, as is clear from Equation C.13. To determine the terms $\partial G/\partial \varepsilon_i$, a function Z is introduced defined as $Z = \varepsilon_e^2(G-C)^2 - D^2 = 0$. Using this function,

by means of implicit differentiation the $\partial G/\partial\varepsilon_i$ terms can be determined. The basic idea is:

$$\frac{\partial Z}{\partial \varepsilon_i} = 0 = 2\varepsilon_e \left(\frac{\partial \varepsilon_e}{\partial \varepsilon_i} + \frac{\partial \varepsilon_e}{\partial G}\frac{\partial G}{\partial \varepsilon_i} \right)(G-C)^2 + 2\varepsilon_e^2 (G-C)\frac{\partial G}{\partial \varepsilon_i} \Rightarrow \tag{C.16}$$

$$\frac{\partial G}{\partial \varepsilon_i} = \frac{-2\varepsilon_e \frac{\partial \varepsilon_e}{\partial \varepsilon_i}(G-C)^2}{2\varepsilon_e \frac{\partial \varepsilon_e}{\partial G}(G-C)^2 + 2\varepsilon_e^2 (G-C)} = \frac{-\frac{\partial \varepsilon_e}{\partial \varepsilon_i}(G-C)}{\frac{\partial \varepsilon_e}{\partial G}(G-C) + \varepsilon_e}. \tag{C.17}$$

The derivatives needed to evaluate this expression are easily found:

$$\frac{\partial \varepsilon_e}{\partial \varepsilon_{xx}} = \frac{1}{2\varepsilon_e}\frac{\partial(\varepsilon_e^2)}{\partial \varepsilon_{xx}} = \frac{2}{9\varepsilon_e}\left(2(\alpha^2-\alpha+1)\varepsilon_{xx} + (2\alpha^2-2\alpha-1)\varepsilon_{yy}\right), \tag{C.18}$$

$$\frac{\partial \varepsilon_e}{\partial \varepsilon_{yy}} = \frac{1}{2\varepsilon_e}\frac{\partial(\varepsilon_e^2)}{\partial \varepsilon_{yy}} = \frac{2}{9\varepsilon_e}\left(2(\alpha^2-\alpha+1)\varepsilon_{yy} + (2\alpha^2-2\alpha-1)\varepsilon_{xx}\right), \tag{C.19}$$

$$\frac{\partial \varepsilon_e}{\partial \varepsilon_{xy}} = \frac{1}{2\varepsilon_e}\frac{\partial(\varepsilon_e^2)}{\partial \varepsilon_{xy}} = \frac{4}{3\varepsilon_e}\varepsilon_{xy} = \frac{2}{9\varepsilon_e}6\varepsilon_{xy}, \tag{C.20}$$

and

$$\frac{\partial \varepsilon_e}{\partial G} = \frac{1}{2\varepsilon_e}\frac{\partial(\varepsilon_e^2)}{\partial G} = \frac{1}{2\varepsilon_e}\frac{\partial(\varepsilon_e^2)}{\partial \alpha}\frac{\partial \alpha}{\partial G} = \frac{2}{9\varepsilon_e}(2\alpha-1)(\varepsilon_{xx}+\varepsilon_{yy})^2\frac{\partial \alpha}{\partial G} \tag{C.21}$$

$$= \frac{2}{9\varepsilon_e}(2\alpha-1)(\varepsilon_{xx}+\varepsilon_{yy})^2\frac{18K}{(4G+3K)^2}. \tag{C.22}$$

This leads to the following $dG/d\varepsilon_i$ derivatives:

$$\frac{\partial G}{\partial \varepsilon_{xx}} = -\frac{(G-C)\left(2(\alpha^2-\alpha+1)\varepsilon_{xx} + (2\alpha^2-2\alpha-1)\varepsilon_{yy}\right)}{(2\alpha-1)(\varepsilon_{xx}+\varepsilon_{yy})^2\frac{18K}{(4G+3K)^2}(G-C) + \frac{9}{2}\varepsilon_e^2}, \tag{C.23}$$

$$\frac{\partial G}{\partial \varepsilon_{yy}} = -\frac{(G-C)\left(2(\alpha^2-\alpha+1)\varepsilon_{yy} + (2\alpha^2-2\alpha-1)\varepsilon_{xx}\right)}{(2\alpha-1)(\varepsilon_{xx}+\varepsilon_{yy})^2\frac{18K}{(4G+3K)^2}(G-C) + \frac{9}{2}\varepsilon_e^2}, \tag{C.24}$$

$$\frac{\partial G}{\partial \varepsilon_{xy}} = -\frac{(G-C)}{(2\alpha-1)(\varepsilon_{xx}+\varepsilon_{yy})^2\frac{18K}{(4G+3K)^2}(G-C) + \frac{9}{2}\varepsilon_e^2}6\varepsilon_{xy}. \tag{C.25}$$

The tangent operator for the plane stress case is now given by:

$$\begin{aligned}\frac{d\sigma}{d\varepsilon} =& K\begin{bmatrix}1+\alpha & 1+\alpha & 0\\ 1+\alpha & 1+\alpha & 0\\ 0 & 0 & 0\end{bmatrix} + \dots \\ & \frac{2G}{3}\begin{bmatrix}2-\alpha & -(1+\alpha) & 0\\ -(1+\alpha) & 2-\alpha & 0\\ 0 & 0 & 3\end{bmatrix} + \begin{Bmatrix}Q_{xx}\\ Q_{yy}\\ 2\varepsilon_{xy}\end{Bmatrix}\cdot\begin{Bmatrix}\frac{\partial G}{\partial \varepsilon_{xx}}\\ \frac{\partial G}{\partial \varepsilon_{yy}}\\ \frac{\partial G}{\partial \varepsilon_{xy}}\end{Bmatrix}^T,\end{aligned} \tag{C.26}$$

where

$$
\begin{aligned}
Q_1 &= \begin{cases} K\frac{d\alpha}{dG}(\varepsilon_{xx}+\varepsilon_{yy}) + \frac{2}{3}(2-\alpha)\varepsilon_{xx} - \dots \\ \qquad \frac{2}{3}(1+\alpha)\varepsilon_{yy} - \frac{2}{3}G(\varepsilon_{xx}+\varepsilon_{yy})\frac{d\alpha}{dG} \end{cases} && \text{(C.27)} \\
&= \left(K - \frac{2}{3}G\right)(\varepsilon_{xx}+\varepsilon_{yy})\frac{d\alpha}{dG} + \frac{2}{3}(2-\alpha)\varepsilon_{xx} - \frac{2}{3}(1+\alpha)\varepsilon_{yy} && \text{(C.28)} \\
&= \left[\left(K - \frac{2}{3}G\right)\frac{d\alpha}{dG} - \frac{2}{3}(1+\alpha)\right](\varepsilon_{xx}+\varepsilon_{yy}) + 2\varepsilon_{xx}, && \text{(C.29)} \\
Q_2 &= \begin{cases} K\frac{d\alpha}{dG}(\varepsilon_{xx}+\varepsilon_{yy}) + \frac{2}{3}(2-\alpha)\varepsilon_{yy} - \dots \\ \qquad \frac{2}{3}(1+\alpha)\varepsilon_{xx} - \frac{2}{3}G(\varepsilon_{xx}+\varepsilon_{yy})\frac{d\alpha}{dG} \end{cases} && \text{(C.30)} \\
&= \left[\left(K - \frac{2}{3}G\right)\frac{d\alpha}{dG} - \frac{2}{3}(1+\alpha)\right](\varepsilon_{xx}+\varepsilon_{yy}) + 2\varepsilon_{yy}, && \text{(C.31)}
\end{aligned}
$$

and

$$
\frac{d\alpha}{dG} = \frac{18K}{(4G+3K)^2}. \tag{C.32}
$$

The expression in Equation C.26 is the correct equation for the tangent operator based on the stress-strain vectors $(\sigma_{xx}, \sigma_{yy}, \sigma_{xy})$ and $(\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{xy})$. It is generally nonsymmetric. To obtain the symmetric tangent operator related to the conventional stress-strain vectors $(\sigma_{xx}, \sigma_{yy}, \sigma_{xy})$ and $(\varepsilon_{xx}, \varepsilon_{yy}, \gamma_{xy})$, which mostly used in finite element implementations, a modification is required: the third column needs to be divided by two. This finally leads to the following expression:

$$
\begin{aligned}
\frac{d\sigma}{d\varepsilon}_{(\gamma)} =& K\begin{bmatrix} 1+\alpha & 1+\alpha & 0 \\ 1+\alpha & 1+\alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} + \dots \\
& \frac{2G}{3}\begin{bmatrix} 2-\alpha & -(1+\alpha) & 0 \\ -(1+\alpha) & 2-\alpha & 0 \\ 0 & 0 & 3/2 \end{bmatrix} + \begin{Bmatrix} Q_{xx} \\ Q_{yy} \\ 2\varepsilon_{xy} \end{Bmatrix} \cdot \begin{Bmatrix} \frac{\partial G}{\partial \varepsilon_{xx}} \\ \frac{\partial G}{\partial \varepsilon_{yy}} \\ \frac{1}{2}\frac{\partial G}{\partial \varepsilon_{xy}} \end{Bmatrix}^T .
\end{aligned} \tag{C.33}
$$

It turns out that this tangent operator is, again, symmetric, and equal to Equation C.11.

Appendix D

Jacobian and Hessian of Symmetry Plane Constraint

To solve a finite element problem containing multi-point constraints using an Lagrangian or augmented Lagrangian approach, the Jacobian and Hessian of the constraints are required. This appendix presents the derivation of these quantities for the following symmetry plane constraint h, introduced in Chapter 3:

$$h = \frac{M2_z S_x - M1_z S_x + M1_z M2_x + M1_x S_z - M2_x S_z - M1_x M2_z}{\sqrt{(M2_x - M1_x)^2 + (M2_z - M1_z)^2}} = 0. \quad \text{(D.1)}$$

Here $\boldsymbol{M1}$, $\boldsymbol{M2}$ and $\boldsymbol{S}$ represent the location of two master nodes and a slave node, respectively. The constraint equation restricts the slave node to a plane parallel to the y-axis, defined by the two master nodes. In order to clarify the notation, the original coordinates of the nodes are collected in a single vector $\boldsymbol{X}$, and the associated displacements are collected in the displacement vector $\boldsymbol{U}$:

$$\boldsymbol{X} = (M1_x, M1_z, M2_x, M2_z, S_x, S_z)^T, \quad \text{(D.2)}$$

$$\boldsymbol{U} = (M1_u, M1_w, M2_u, M2_w, S_u, S_w)^T, \quad \text{(D.3)}$$

$$\boldsymbol{x} = \boldsymbol{X} + \boldsymbol{U} = (a, b, c, d, e, f)^T. \quad \text{(D.4)}$$

Subscript u and w denote a displacement component in x- and z-direction, respectively. The components of the actual nodal location vector $\boldsymbol{x}$ are represented by simple symbols to clarify the notation. For example, a represents $M1_x + M1_u$.

The constraint equation in this notation reads

$$h = \frac{e(d-b) + f(a-c) + bc - ad}{\sqrt{(c-a)^2 + (d-b)^2}} = \frac{R}{\sqrt{Q}}, \quad \text{(D.5)}$$

where Q and R are introduced to simplify the following expressions. It can be seen that the constraint is a nonlinear function of the nodal displacements. Differentiating the constraint h to the displacement vector $\boldsymbol{U}$ yields the Jacobian $dh/d\boldsymbol{U}$:

$$\frac{dh}{d\boldsymbol{U}} = \frac{d}{d\boldsymbol{U}}\left(\frac{R}{\sqrt{Q}}\right) = \frac{\frac{dR}{d\boldsymbol{U}}Q - \frac{1}{2}R\frac{dQ}{d\boldsymbol{U}}}{Q\sqrt{Q}}. \tag{D.6}$$

Also, the Hessian $d^2h/d\boldsymbol{U}^2$ can be expressed as

$$\frac{d^2h}{d\boldsymbol{U}^2} = \frac{d}{d\boldsymbol{U}}\left(\frac{dh}{d\boldsymbol{U}}\right) = \begin{array}{l}\dfrac{\frac{d^2R}{d\boldsymbol{U}^2}Q^2 - \frac{1}{2}\frac{dR}{d\boldsymbol{U}}\frac{dQ}{d\boldsymbol{U}}^T Q - \frac{1}{2}\frac{dQ}{d\boldsymbol{U}}\frac{dR}{d\boldsymbol{U}}^T Q}{Q^2\sqrt{Q}} + \dots \\ \dfrac{\frac{3}{4}R\frac{dQ}{d\boldsymbol{U}}\frac{dQ}{d\boldsymbol{U}}^T - \frac{1}{2}R\frac{d^2Q}{d\boldsymbol{U}^2}Q}{Q^2\sqrt{Q}}\end{array}. \tag{D.7}$$

In these equations, the derivatives of R and Q are given by:

$$\frac{dR}{d\boldsymbol{U}} = (f-d, c-e, b-f, e-a, d-b, a-c)^T, \tag{D.8}$$

$$\frac{dQ}{d\boldsymbol{U}} = 2(a-c, b-d, c-a, d-b, 0, 0)^T, \tag{D.9}$$

$$\frac{d^2R}{d\boldsymbol{U}^2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{D.10}$$

$$\frac{d^2Q}{d\boldsymbol{U}^2} = \begin{bmatrix} 2 & 0 & -2 & 0 & 0 & 0 \\ 0 & 2 & 0 & -2 & 0 & 0 \\ -2 & 0 & 2 & 0 & 0 & 0 \\ 0 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{D.11}$$

The Jacobian and Hessian are used in the implementation of this constraint condition.

Appendix E

Derivatives of Material Parameters C_i and D_i

As also discussed in Chapter 6, the piecewise linear stress-strain equations describing the one-dimensional SMA constitutive behavior can be cast in the following generalized form:

$$\sigma_{xx} = A_i(T)\varepsilon_{xx} + B_i(T). \tag{E.1}$$

The terms $A_i(T)$ and $B_i(T)$ follow from the original formulation, and the indices refer to the three segments of the piecewise linear relation:

$$A_i(T) = \begin{cases} \varepsilon_{xx} \leq \varepsilon_1 & : \quad A_1 = E_A, \\ \varepsilon_1 < \varepsilon_{xx} \leq \varepsilon_2 & : \quad A_2 = E_T, \\ \varepsilon_{xx} > \varepsilon_2 & : \quad A_3 = E_R. \end{cases} \tag{E.2}$$

$$B_i(T) = \begin{cases} \varepsilon_{xx} \leq \varepsilon_1 & : \quad B_1 = 0, \\ \varepsilon_1 < \varepsilon_{xx} \leq \varepsilon_2 & : \quad B_2 = (E_A - E_T)\varepsilon_1, \\ \varepsilon_{xx} > \varepsilon_2 & : \quad B_3 = (E_A - E_T)\varepsilon_1 + (E_T - E_R)\varepsilon_2. \end{cases} \tag{E.3}$$

The generalization to a three-dimensional setting yields the following relation between the effective strain ε_e and the shear modulus G:

$$G = \frac{3K}{9K - A_i(T)} \left(A_i(T) + \frac{B_i(T)}{\varepsilon_e} \right) = C_i(T) + \frac{D_i(T)}{\varepsilon_e}. \tag{E.4}$$

The quantities $C_i(T)$ and $D_i(T)$ are defined here for convenience. Their derivatives are used in the computation of $\partial G/\partial T$, and are given by:

$$\frac{dC_i}{dT} = \frac{27K^2 \frac{dA_i}{dT}}{(9K - A_i)^2}, \qquad \frac{dD_i}{dT} = \frac{3K(9K - A_i)\frac{dB_i}{dT} + 3KB_i \frac{dA_i}{dT}}{(9K - A_i)^2}, \tag{E.5}$$

where

$$\frac{dA_i}{dT} = \begin{cases} \varepsilon_e \leq \varepsilon_{e(1)} & : \quad \frac{dA_1}{dT} = 0, \\ \varepsilon_{e(1)} < \varepsilon_e \leq \varepsilon_{e(2)} & : \quad \frac{dA_2}{dT} = K_E, \\ \varepsilon_e > \varepsilon_{e(2)} & : \quad \frac{dA_3}{dT} = 0. \end{cases} \tag{E.6}$$

$$\frac{dB_i}{dT} = \begin{cases} \varepsilon_e \leq \varepsilon_{e(1)} & : \quad \frac{dB_1}{dT} = 0, \\ \varepsilon_{e(1)} < \varepsilon_e \leq \varepsilon_{e(2)} & : \quad \frac{dB_2}{dT} = -K_E\varepsilon_1 + (E_A - E_T)K_\varepsilon, \\ \varepsilon_e > \varepsilon_{e(2)} & : \quad \frac{dB_3}{dT} = K_E\Delta + (E_A - E_R)K_\varepsilon. \end{cases} \tag{E.7}$$

Appendix F

Maximum Effective Strain Values Occur at Outer Layers

In shell elements, the effective strain used in the SMA material model described in Chapter 3 can vary throughout the thickness of the element. Since the maximum effective strain is relevant to optimization problems, it is of interest to investigate whether the maximum always occurs at the outer layers of the shell. This would reduce the number of quantities that need to be considered. In this appendix, it is shown that this is indeed the case.

The strain components in every layer $\tilde{\varepsilon}_{ij}$ are a function of the transverse coordinate z (with $z = 0$ at the midplane) and the midplane membrane strains ε_{ij} and curvatures κ_{ij}:

$$\begin{Bmatrix} \tilde{\varepsilon}_{xx} \\ \tilde{\varepsilon}_{yy} \\ \tilde{\varepsilon}_{xy} \end{Bmatrix}_{\text{Layer}} = \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{xy} \end{Bmatrix}_{\text{Midplane}} + z \begin{Bmatrix} \kappa_{xx} \\ \kappa_{yy} \\ \kappa_{xy} \end{Bmatrix}_{\text{Midplane}} . \tag{F.1}$$

Note that the distinction between layer and midplane components is only made in this appendix. Since the strain components can vary from layer to layer, the effective strain ε_e can vary as well. It is defined in terms of layer strain components as:

$$\varepsilon_e = \sqrt{\frac{4}{9}(\alpha^2 - \alpha + 1)(\tilde{\varepsilon}_{xx}^2 + \tilde{\varepsilon}_{yy}^2) + \frac{4}{9}(2\alpha^2 - 2\alpha - 1)\tilde{\varepsilon}_{xx}\tilde{\varepsilon}_{yy} + \frac{4}{3}\tilde{\varepsilon}_{xy}^2}, \tag{F.2}$$

where

$$\alpha = \frac{2G - 3K}{4G + 3K}. \tag{F.3}$$

By substitution of Equation F.1, ε_e can be expressed in midplane quantities and z. The square root complicates the further derivations, hence the square of the effective strain will be considered:

$$\begin{aligned}\varepsilon_e^2 = g(z) = &\frac{4}{9}(\alpha^2-\alpha+1)((\varepsilon_{xx}+z\kappa_{xx})^2+(\varepsilon_{yy}+z\kappa_{yy})^2)+\dots \\ &\frac{4}{9}(2\alpha^2-2\alpha-1)(\varepsilon_{xx}+z\kappa_{xx})(\varepsilon_{yy}+z\kappa_{yy})+\frac{4}{3}(\varepsilon_{xy}+z\kappa_{xy})^2.\end{aligned} \tag{F.4}$$

Note that since the effective strain is nonnegative, the square operator preserves the monotonicity. In the remainder of this appendix, ε_e^2 is therefore considered instead of ε_e itself, and is referred to as $g(z)$, which is a quadratic function of z. To proof that this nonnegative function reaches its extreme values for the outer layers, it has to be verified that the second order derivative is nonnegative for all possible strains and curvatures, such that the quadratic function $g(z)$ cannot have an internal maximum in an interval. This second order derivative is given by:

$$\frac{d^2g}{dz^2} = \frac{8}{9}(\alpha^2-\alpha+1)(\kappa_{xx}^2+\kappa_{yy}^2)+\frac{8}{9}(2\alpha^2-2\alpha-1)\kappa_{xx}\kappa_{yy}+\frac{8}{3}\kappa_{xy}^2, \tag{F.5}$$

which simplifies to

$$\frac{d^2g}{dz^2} = \frac{8}{9}(\alpha^2-\alpha+1)(\kappa_{xx}+\kappa_{yy})^2+\frac{8}{3}\kappa_{xy}^2-\frac{8}{3}\kappa_{xx}\kappa_{yy}. \tag{F.6}$$

Subsequent substitution of the definition of α gives:

$$\frac{d^2g}{dz^2} = \frac{8}{3}\left(\underbrace{\Psi(\kappa_{xx}+\kappa_{yy})^2+\kappa_{xy}^2}_{\geq 0}-\underbrace{\kappa_{xx}\kappa_{yy}}_{\leq 0}\right) \tag{F.7}$$

with

$$\Psi = \frac{4G^2+6GK+9K^2}{(4G+3K)^2} > 0. \tag{F.8}$$

In order to determine the sign of d^2g/dz^2, the magnitudes of the terms larger than and smaller than zero have to be compared. The term Ψ involving G and K plays a crucial role in this comparison. This Ψ turns out decrease when the ratio G/K increases, as shown in Figure F.1. Because the shear modulus has its largest value before the onset of the phase transformation, the critical situation to consider is the austenite state. In that state, the moduli G and K are related to the Young's modulus and Poisson ratio in the following way:

$$G = \frac{E}{2(1+\nu)}, \qquad K = \frac{E}{(3(1-2\nu)} \;\Rightarrow\; \frac{G}{K} = \frac{3(1-2\nu)}{2(1+\nu)}. \tag{F.9}$$

Combining this relation with the definition of Ψ given in Equation F.8, Ψ can be expressed as a function of the Poisson ratio ν. As shown in Figure F.1, Ψ increases with an increase in ν.

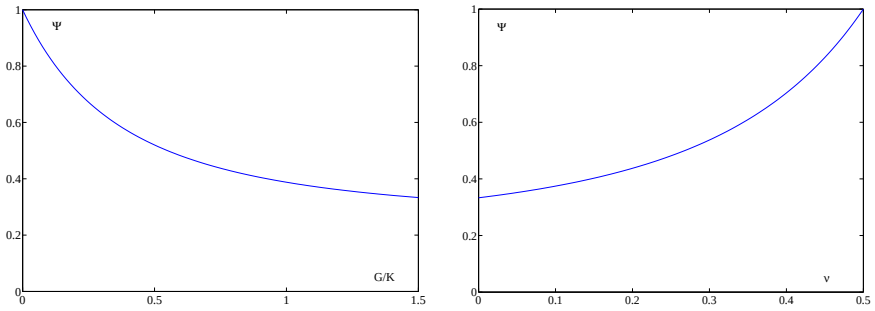


Figure F.1: *The term Ψ as defined in Equation F.8, as a function of G/K (left) and of ν (right).*

As an (unrealistic) worst case, the case $\nu = 0$ is considered, which gives:

$$\frac{8}{3}\frac{d^2g}{dz^2} = \frac{1}{3}(\kappa_{xx} + \kappa_{yy})^2 + \kappa_{xy}^2 - \kappa_{xx}\kappa_{yy} \tag{F.10}$$

$$= \frac{1}{3}\left(\kappa_{xx}^2 - \kappa_{xx}\kappa_{yy} + \kappa_{yy}^2\right) + \kappa_{xy}^2. \tag{F.11}$$

Since

$$\kappa_{xx}^2 - \kappa_{xx}\kappa_{yy} + \kappa_{yy}^2 \geq (\kappa_{xx}^2 - 2\kappa_{xx}\kappa_{yy} + \kappa_{yy}^2) = (\kappa_{xx} + \kappa_{yy}) \geq 0, \tag{F.12}$$

it follows that even in this worst case, d^2g/dz^2 cannot become negative. Therefore it can be concluded that the maximum value of the effective strain ε_e must be located at the outer layer of the shell.

Appendix G

Full Color Illustrations

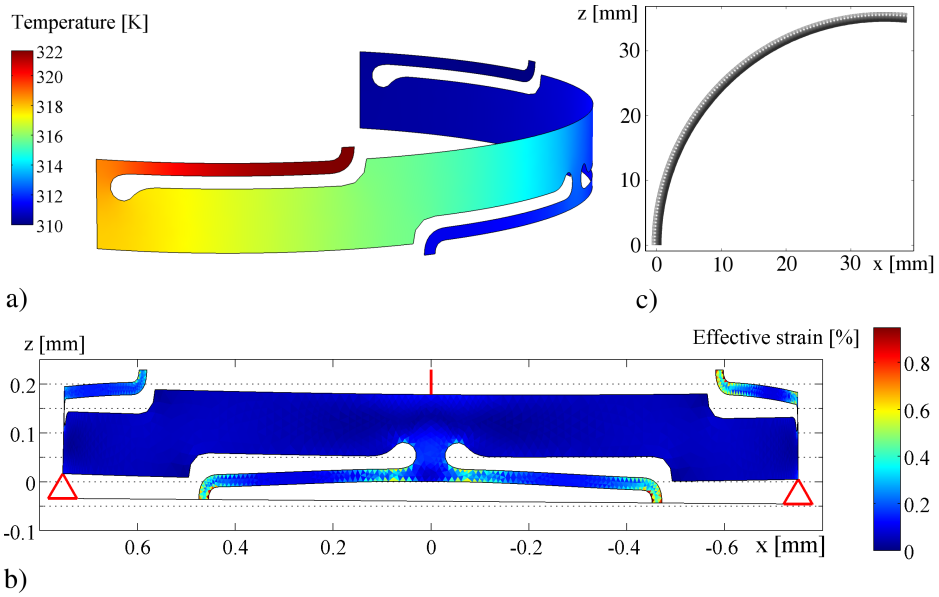


Figure G.1: *(Figure 3.7) Results of the finite element analysis. a) Temperature distribution, b) Deformed geometry and effective strain, c) Composition of catheter in bent configuration.*

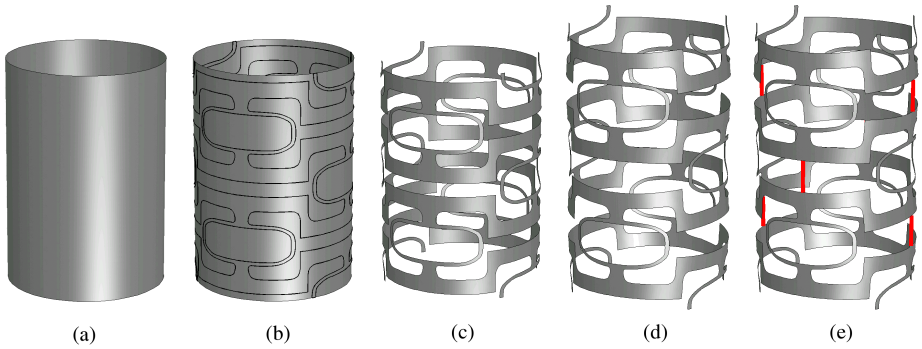


Figure G.2: *(Figure 3.2) Proposed fabrication of the new active catheter concept using laser cutting from a small diameter SMA tube (a). After laser cutting (b–c), the structure is stretched (d) and spacers are applied (e), in order to generate internal stresses.*

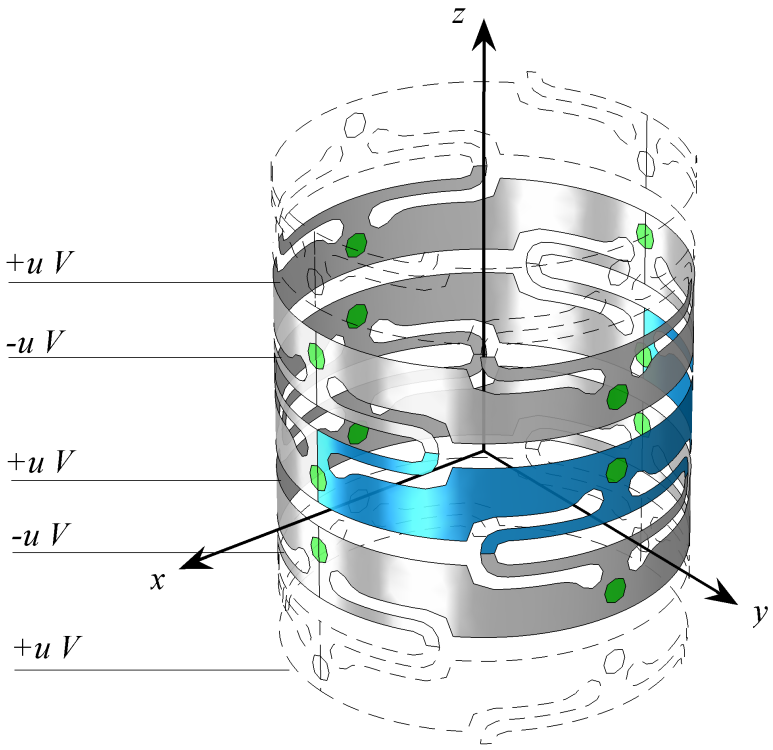
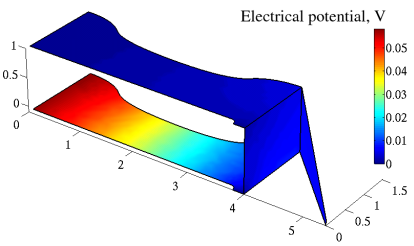
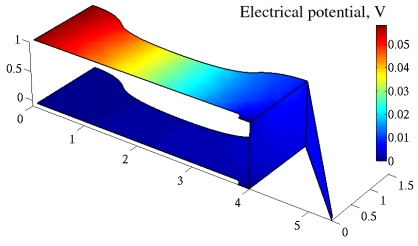


Figure G.3: *(Figure 3.5) Segmentation, applied voltage pattern and coordinate system.*



(a) Analysis results for the opened configuration.



(b) Analysis results for the closed configuration.

Figure G.4: *(Figure 4.7) Computed electrical potential (top), temperature distribution (middle) and Von Mises stress distribution on the deformed structure (bottom) for the optimal design in the constrained Joule heated case, in opened (left) and closed (right) gripper configurations.*

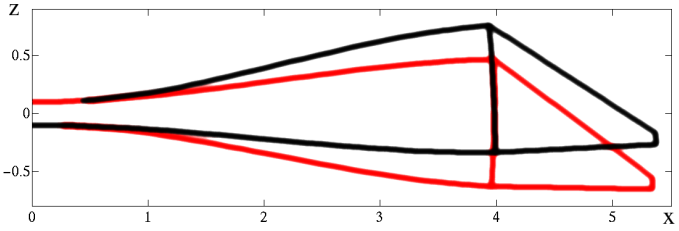
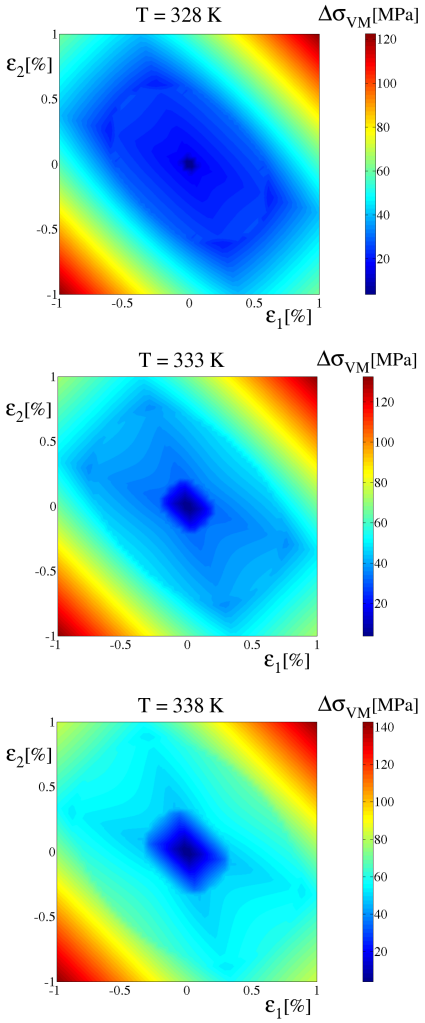
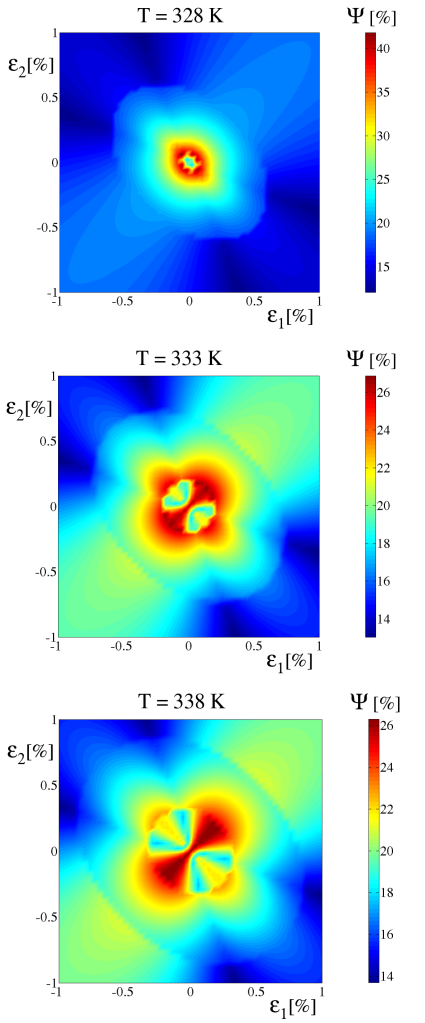


Figure G.5: *(Figure 4.8) Gripper top arm in open (black) and closed (red) position, side view.*



(a) Maximum absolute differences in Von Mises stress.



(b) Maximum differences in Von Mises stress relative to the nominal values.

Figure G.6: *(Figure 5.18) Maximum absolute (left) and relative (right) differences in Von Mises stress at different strain states due to the effect of the uncertainties, at different nominal temperatures.*

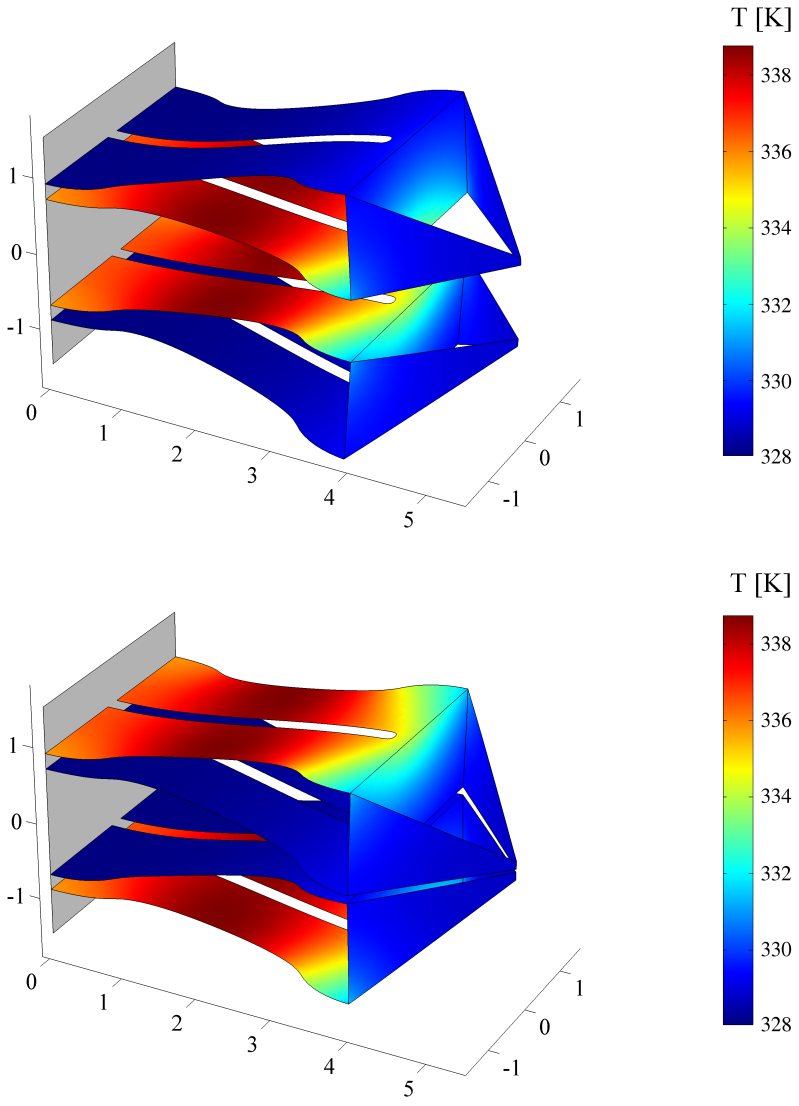


Figure G.7: *(Figure 8.7) Temperature distribution [K] on the optimal miniature SMA gripper in the open and closed configuration.*

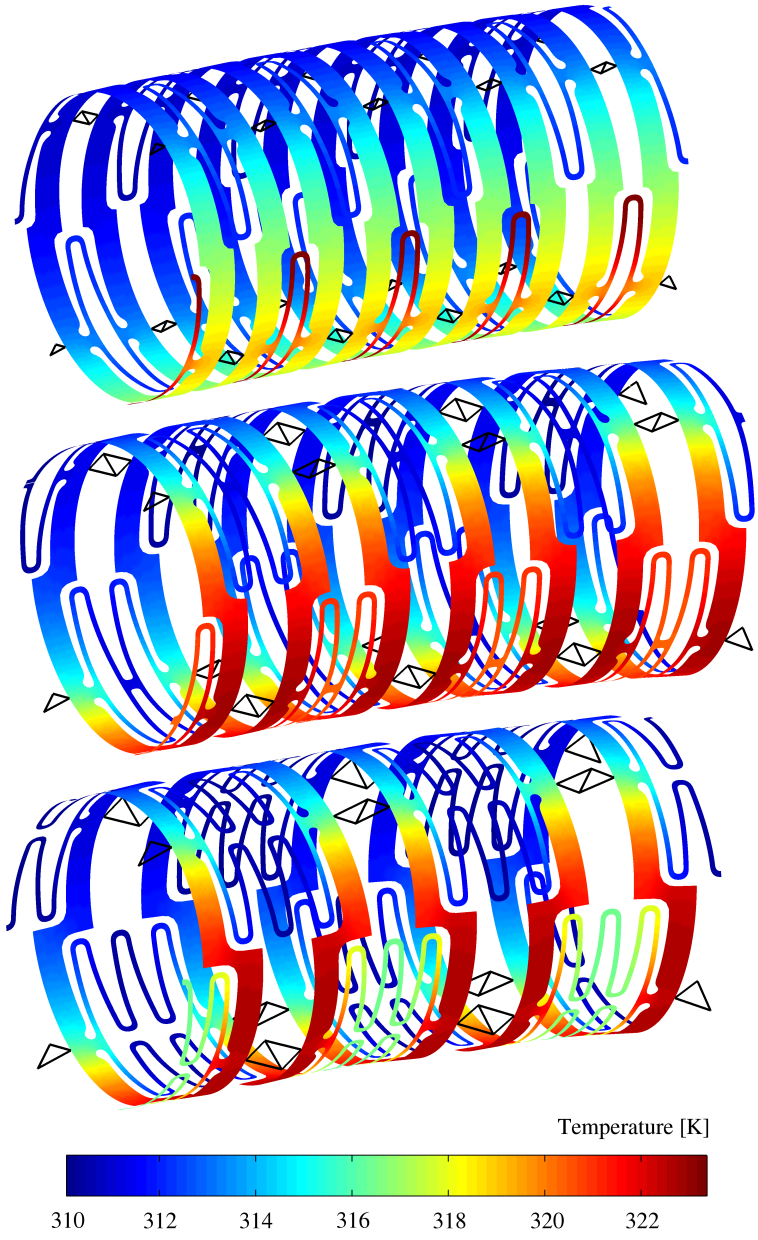


Figure G.8: *Temperature distributions on parts of the deformed catheter structure for the best designs of all three concepts, for a 1.0 mm diameter.*

Bibliography

Agrawal, S. and Erdman, A. (2005). Biomedical assist devices and new biomimetic machines - a short perspective. *Journal of Mechanical Design*, 127(4):799–801.

Anderson, W. and Nielsen, E. (2001). Sensitivity analysis for navier-stokes equations on unstructured meshes using complex variables. *AIAA Journal*, 39(1):56–63.

Arora, J. (1997). *Guide to structural optimization*. ASCE, New York.

Arora, J. (2004). *Introduction to Optimum Design*. Elsevier Academic Press, San Diego, CA.

Asimov, I. (1966). *Fantastic Voyage*. Bantam Books.

Avron, J., Gat, O., and Kenneth, O. (2004). Optimal swimming at low reynolds numbers. *Physical Review Letters*, 93:186001.

Bakker, N. (2000). Steering principles for catheters and guidewires. Technical Report Report N561, Delft University of Technology.

Bartholomew-Biggs, M., Brown, S., Christianson, B., and Dixon, L. (2000). Automatic differentiation of algorithms. *Journal of Computational and Applied Mathematics*, 124(1–2):171–190.

Bathe, K. (1995). *Finite Element Procedures*. Prentice-Hall, Englewood Cliffs.

Behkam, B. and Sitti, M. (2004). E. Coli inspired propulsion for swimming microrobots. In *Proceedings of 2004 ASME International Mechanical Engineering Conference and Exposition, November 2004, Anaheim, CA*.

Ben-Haim, Y. (1996). *Robust reliability in the mechanical sciences*. Springer.

Ben-Haim, Y. and Elishakoff, I. (1990). *Convex models of uncertainty in applied mechanics*. Elsevier.

Bendsoe, M. and Kikuchi, N. (1988). Generating optimal topologies in structural design using a homogenization method. *Computer Methods in Applied Mechanics and Engineering*, 71(2):197–224.

Bendsoe, M. and Sigmund, O. (2003). *Topology Optimization - Theory, Methods and Applications*. Springer-Verlag, Berlin.

Bhattacharya, K. and Kohn, R. (1996). Symmetry, texture and the recoverable strain of shape-memory polycrystals. *Acta Materialia*, 44(2):529–542.

Birman, V. (1996). Optimum design of hybrid shape memory alloy sandwich panels for maximum natural frequencies. In *Proceedings of SPIE, The International Society for Optical Engineering, Smart Structures and Materials 1996: Industrial and Commercial Applications of Smart Structures Technologies. San Diego, CA, USA*, volume 2721, pages 263–272.

Birman, V. (1997). Review of mechanics of shape memory alloy structures. *Applied Mechanics Reviews*, 50(11):629–645.

Birman, V., Chandrashekhara, K., and Sain, S. (1996). Approach to optimization of shape memory alloy hybrid composite plates subjected to low-velocity impact. *Composites Part B: Engineering*, 27(5):439–446.

Bischof, C., Carle, A., Khademi, P., and Mauer, A. (1996). The ADIFOR 2.0 system for the automatic differentiation of Fortran 77 programs. *IEEE Computational Science & Engineering*, 3:18–32.

Boyd, J. and Lagoudas, D. (1994). Thermomechanical response of shape memory composites. *Journal of Intelligent Material Systems and Structures*, 5(3):333–346.

Briggs, J. and Ostrowski, J. (2002). Experimental feedforward and feedback control of a one-dimensional sma composite. *Smart Materials and Structures*, 11:9–23.

Brinson, L. and Huang, M. (1996). Simplifications and comparisons of shape memory alloy constitutive models. *Journal of Intelligent Material Systems and Structures*, 7(1):108–114.

Buhl, T., Pedersen, C., and Sigmund, O. (2000). Stiffness design of geometrically nonlinear structures using topology optimization. *Structural and Multidisciplinary Optimization*, 19(2):93–104.

Büttgenbach, S., Bütefisch, S., Leester-Schädel, M., and Wogersien, A. (2001). Shape memory microactuators. *Microsystem Technologies*, 7(4):165–170.

Chang, J., Chung, S., Lee, Y., Park, J., Lee, S., Yang, S., Moon, S., Tschepe, J., Chee, Y., and Han, D. (2002). Development of endovascular microtools. *Journal of Micromechanics and Microengineering*, 12(6):824–831.

Chang, J. K., Chung, S., Lee, Y., Park, J., Lee, S.-K., Yang, S. S., Moon, S.-Y., and Han, D.-C. (2000). Intravascular micro active catheter for minimal invasive surgery. In Dittmar, A. and Beebe, D., editors, *1st Annual International, Conference on Microtechnologies in Medicine and Biology, 2000*, pages 243–246. IEEE Catalog Number = 00EX451.

Cheung, Y., Lo, S., and Leung, A. (1996). *Finite Element Implementation*. Blackwell Science.

Cho, S. and Choi, K. (2000). Design sensitivity analysis and optimization of non-linear transient dynamics. part I - sizing design. *International Journal for Numerical Methods in Engineering*, 48(3):351–373.

Choi, K. and Duan, W. (2000). Design sensitivity analysis and shape optimization of structural components with hyperelastic material. *Computer Methods in Applied Mechanics and Engineering*, 187(1):219–243.

Choi, K. and Kim, N. (2005a). *Structural Sensitivity Analysis and Optimization 1 - Linear Systems*. Springer-Verlag.

Choi, K. and Kim, N. (2005b). *Structural Sensitivity Analysis and Optimization 2 - Nonlinear Systems and Applications*. Springer-Verlag.

Chopra, I. (2002). Review of state of art of smart structures and integrated systems. *AIAA Journal*, 40(11):2145–2187.

Coleman, T. and Jonsson, G. (1999). The efficient computation of structured gradients using automatic differentiation. *SIAM Journal on Scientific Computing*, 20(4):1430–1437.

Crisfield, M. (1991). *Non-linear Finite Element Analysis of Solids and Structures*, volume 1: Essentials. John Wiley & Sons Ltd.

Dario, P., Carrozza, M., Lencioni, L., Magnani, B., and d'Attanasio, S. (1997). Micro robotic system for colonoscopy. In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation, ICRA, Albuquerque, NM, USA*, volume 2, pages 1567–1572. IEEE Catalog Number = 97CH35992.

Dario, P., Carrozza, M., and Pietrabissa, A. A. (1999). Development and in vitro testing of a miniature robotic system for computer-assisted colonoscopy. *Computer Aided Surgery*, 4(1):1–14.

Dario, P., Valleggi, R., Carrozza, M., Montesi, M., and Cocco, M. (1992). Microactuators for microrobots: a critical survey. *Journal of Micromechanics and Microengineering*, 2(3):141–157.

Dario, P., Valleggi, R., Pardini, M., and Sabatini, A. (1991). A miniature device for medical intracavitary intervention. In *IEEE MEMS '91, Nara, Japan*, pages 171–175.

De Boer, H. and Van Keulen, F. (2000a). Refined semi-analytical design sensitivities. *Int. J. Solids and Structures.*, 37(46-47):6961–6980.

De Boer, H. and Van Keulen, F. (2000b). Refined semi-analytical design sensitivities. *International Journal of Solids and Structures*, 37(46–47):6961–6980.

De Boer, M., Luck, D., Ashurst, W., Maboudian, R., Corwin, A., Walraven, J., and Redmond, J. (2004). High-performance surface-micromachined inchworm actuator. *Journal of Microelectromechanical Systems*, 13(1):63–74.

De Ruiter, M. (2005). *Topology Optimization using a Topology Description Function Approach.* PhD thesis, Delft University of Technology.

Dickson, T., Moore, B., and Toyama, N. (2002). Innovations: Laser-cutting Nickel-Titanium. In *Materials Science Forum*, volume 394–395, pages 309–312. Ütikon-Zürich Trans Tech Publications. Proceedings of the International Conference on Shape Memory and Superelastic Technologies and Shape Memory Materials (SMST-SMM 2001), Kunming, China.

Duerig, T. (2002). The use of superelasticity in modern medicine. *MRS Bulletin*, 27(2):101–104.

Duerig, T., Melton, K., Stöckel, D., and Wayman, C., editors (1990). *Engineering Aspects of Shape Memory Alloys.* Butterworth-Heinemann, London.

Duerig, T., Pelton, A., and Stöckel, D. (1999). An overview of nitinol medical applications. *Materials Science and Engineering: A - Structural Materials: Properties, Microstructure and Processing*, 273-275(Special Issue):149–160.

Dumont, G. and Kuhl, C. (2005). Finite element simulation for design optimisation of shape memory alloy spring actuators. *Engineering Computations*, 22(7–8):835–848.

Duysinx, P. and Bendsoe, M. (1998). Topology optimization of continuum structures with local stress constraints. *International Journal for Numerical Methods in Engineering*, 43(8):1453–1478.

Edd, J., Payen, S., Rubinsky, B., Stoller, M., and Sitti, M. (2003). Biomimetic propulsion for a swimming surgical micro-robot. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, October 27–31, Las Vegas, U.S.A.*

Elishakoff, I. (1983). *Probabilistic methods in the theory of structures.* John Wiley and Sons.

Elishakoff, I. (1999). *Whys and hows in uncertainty modelling: Chap. 5 - Are probabilistic and anti-optimization approaches compatible?* SpringerWienNewYork.

Elishakoff, I., Haftka, R. T., and Fang, J. J. (1994). Structural design under bounded uncertainty-optimization and anti-optimization. *International Journal of Computers & Structures*, 53:1401–1405.

Ernst, S., Hachiya, H., Chun, J., and Ouyang, F. (2005). Remote catheter ablation of parahisian accessory pathways using a novel magnetic navigation system - a report of two cases. *Journal of Cardiovascular Electrophysiology*, 16(6):659–662.

Eschenauer, H. and Olhoff, N. (2001). Topology optimization of continuum structures: A review. *Applied Mechanics Reviews*, 54(4):331–390.

Eurostat (2005). *Health in Europe - Data 1998-2003*. Office for Official Publications of the European Communities, Luxembourg.

Faddis, M., Blume, W., Finney, J., Hall, A., Rauch, J., Sell, J., Bae, K., Talcott, M., and Lindsay, B. (2002). Novel, magnetically guided catheter for endocardial mapping and radiofrequency catheter ablation. *Circulation*, 106:2980–2985.

Fagan, M. (1992). *Finite Element Analysis; Theory and Practice*. Harlow: Longman.

Farin, G. (2002). *Curves and surfaces for CAGD : a practical guide*. Kaufmann, San Francisco.

Fatikow, S. and Rembold, U. (1997). *Microsystem Technology and Microrobotics*. Springer-Verlag.

Fearing, R. (1998). Powering 3 dimensional microrobots: Power density limitations. In *IEEE International Conference on Robotics and Automation, May 16–21, Leuven, Belgium*.

Feynman, R. (1960). There's plenty of room at the bottom. *Engineering and Science*, 23:22–36.

Feynman, R. (1992). There's plenty of room at the bottom. *Journal of Microelectromechanical Systems*, 1(1):60–66.

Fischer, H., Vogel, B., Pfleging, W., and Besser, H. (1999). Flexible distal tip made of nitinol (NiTi) for a steerable endoscopic camera system. *Materials Science and Engineering: A - Structural Materials: Properties, Microstructure and Processing*, 273-275(Special Issue):780–783.

Fletcher, R. (1980). *Practical Methods of Optimization, Vol. 2 - Constrained Optimization*. John Wiley and Sons.

Fu, Y., Du, H., Huang, W., Zhang, S., and Hu, M. (2004). TiNi-based thin films in MEMS applications: a review. *Sensors and Actuators A - Physical*, 112(2–3):395–408.

Fuchs, K. (2002). Minimally invasive surgery. *Endoscopy*, 34(2):154–159.

Fujimasa, I. (1996). *Micromachines; a new era in mechanical engineering*. Oxford University Press, Oxford.

Fujita, H. and Gabriel, K. (1991). New opportunities for microactuators. In *Transducers '91, International Conference on Solid-State Sensors and Actuators*.

Fukuda, T., Guo, S., Kosuge, K., Arai, F., Negoro, M., and Nakabayashi, K. (1994a). Micro active catheter system with multi degrees of freedom. In *IEEE International Conference onRobotics and Automation, 1994. Proceedings*, volume 3, pages 2290–2295. IEEE Catalog Number = 94CH3375-3.

Fukuda, T., xiang Guo, S., Arai, F., Kosuge, K., and Negoro, M. (1994b). Study on active catheter system (1st report, structure, experimental results and characteristic evaluation of multiunits using SMA active catheter with multi D.O.F.). *Nippon Kikai Gakkai Ronbunshu, C Hen / Transactions of the Japan Society of Mechanical Engineers, Part C*, 60(573):1727–1734.

Funakubo, H., editor (1987). *Shape Memory Alloys*, volume 1 of *Precision Machinery and Robotics*. Gordon and Bleach, New York.

Geske, J., Goldstein, R., and Stambler, B. (2005). Novel steerable telescoping catheter system for implantation of left ventricular pacing leads. *Journal of Interventional Cardiac Electrophysiology*, 12(1):83–89.

Golub, G. and Loan, C. V. (1996). *Matrix Computations*. Johns Hopkins, Baltimore.

Griewank, A., Juedes, D., and Utke, J. (1996). ADOL-C, a package for the automatic differentiation of algorithms written in C/C++. *ACM Transactions on Mathematical Software*, 22(2):131–167.

Guest, J., Prévost, J., and Belytschko, T. (2004). Achieving minimum length scale in topology optimization using nodal design variables and projection functions. *International Journal for Numerical Methods in Engineering*, 61(2):238–254.

Guinness, editor (2005). *Guinness World Records 2005*. Guinness.

Guo, S., Fukuda, T., Nakamura, T., Arai, F., Oguro, K., and Negoro, M. (1996). Micro active guide wire catheter system - characteristic evaluation, electrical model and operability evaluation of micro active catheter. In *IEEE International Conference on Robotics and Automation, 1996. Proceedings.*, volume 3, pages 2226–2231. IEEE Catalog Number = 96CH35857.

Guo, X., Cheng, G., and Yamazaki, K. (2001). A new approach for the solution of singular optima in truss topology optimization with stress and local buckling constraints. *Structural and Multidisciplinary Optimization*, 22(5):364–372.

Gurav, S., Langelaar, M., Goosen, J., and Van Keulen, F. (2003). Bounded-but-unknown uncertainty optimization of micro-electro-mechanical-systems. In *Proceedings of the Second M.I.T. Conference on Computational Fluid and Solid Mechanics, June 17-20, 2003, M.I.T. Cambridge, MA 02139 U.S.A.* CD-ROM.

Gurav, S. P., Goosen, J. F. L., and Van Keulen, F. (2005). Uncertainty-based design optimization of structures using bounded-but-unknown uncertainties, design sensitivities and parallel computing: Application to MEMS. *Computers & Structures*, 83/14:1134–1149.

Gurav, S. P., Vervenne, K., Damveld, H. J., and Van Keulen., F. (2002). Bounded-but-unknown uncertainties in design optimization by combining the multipoint approximation method and design sensitivities. In *Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit, Denver, CO, U.S.A., Paper 1759, 22-25 April*.

Gyobu, A., Kawamura, Y., Horikawa, H., and Saburi, T. (1996). Martensitic transformations in sputter-deposited shape memory Ti-Ni films. *Materials Transactions, JIM*, 37(4):697–702.

Haftka, R. (1985). Sensitivity calculations for iteratively solved problems. *International Journal for Numerical Methods in Engineering*, 21(8):1535–1546.

Haftka, R. and Gürdal, Z. (1992). *Elements of structural optimization*. Kluwer, Dordrecht.

Haga, Y. and Esashi, M. (2004). Biomedical microsystems for minimally invasive diagnosis and treatment. *Proceedings of the IEEE*, 92(1):98–114.

Haga, Y., Esashi, M., and Maeda, S. (2000). Bending, torsional and extending active catheter assembled using electroplating. In *The Thirteenth Annual International Conference on Micro Electro Mechanical Systems, 2000. MEMS 2000*, pages 181–186. IEEE Catalog Number = 00CH36308.

Hassani, B. and Hinton, E. (1998a). A review of homogenization and topology optimization I - homogenization theory for media with periodic structure. *Computers & Structures*, 69(6):707–717.

Hassani, B. and Hinton, E. (1998b). A review of homogenization and topology optimization II - analytical and numerical solution of homogenization equations. *Computers & Structures*, 69(6):719–738.

Hassani, B. and Hinton, E. (1998c). A review of homogenization and topology optimization III - topology optimization using optimality criteria. *Computers & Structures*, 69(6):739–756.

Herr, H. and Dennis, B. (2004). A swimming robot actuated by living muscle tissue. *Journal of NeuroEngineering and Rehabilitation*, 1(6).

Herzog, I., Reul, K., and Jenninger, W. (1989). *Verbrennungen*. Kohlhammer, Stuttgart, Berlin, Köln.

Huber, J., Fleck, N., and Ashby, M. (1997). The selection of mechanical actuators based on performance indices. *Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, 453(1965):2185–2205.

Humbeeck, J. V. (2001). Shape memory alloys: A material and a technology. *Advanced Engineering Materials*, 3(11):837–850.

Ikuta, K., Ichikawa, H., and Suzuki, K. (2002). Safety-active catheter with multiple-segments driven by micro-hydraulic actuators. *Lecture Notes in Computer Science*, 2488:182–191.

Ikuta, K., Ichikawa, H., Suzuki, K., and Yamamoto, T. (2003). Safety active catheter with multi-segments driven by innovative hydro-pressure micro actuators. In *IEEE MEMS '03, Kyoto, Japan*.

Ikuta, K., Tsukamoto, M., and Hirose, S. (1991). Mathematical model and experimental verification of shape memory alloy for designing micro actuator. In *Proceedings of the 1991 IEEE Micro Electro Mechanical Systems - MEMS '91*, volume 2957, pages 103–108. IEEE. IEEE Catalog Number = 91CH2957.

Ishiyama, K., Sendoh, M., Yamazaki, A., and Arai, K. (2001). Swimming micromachine driven by magnetic torque. *Sensors and Actuators A: Physical*, 91(1–2):141–144.

Jacobson, J., Goodwinjohansson, S., Bobbio, S., Bartlett, C., and Yadon, L. (1995). Integrated force arrays - theory and modeling of static operation. *Journal of Microelectromechanical Systems*, 4(3):139–150.

James, R. and Rizzoni, R. (2000). Pressurized shape memory thin films. *Journal of Elasticity*, 59(1–3):399–436.

Jung, D. and Gea, H. (2004). Topology optimization of nonlinear structures. *Finite Elements in Analysis and Design*, 40(11):1417–1427.

Jung, Y., Papadopoulos, P., and Ritchie, R. (2004). Constitutive modelling and numerical simulation of multivariant phase transformation in superelastic shape-memory alloys. *International Journal for Numerical Methods in Engineering*, 60(2):429–460.

Kakeshita, T., Shimizu, K., Nakamichi, S., Tanaka, R., Endo, S., and Ono, F. (1992). Effect of hydrostatic pressures on thermoelastic martensitic transformations in aged Ti-Ni and ausaged Fe-Ni-Co-Ti shape memory alloys. *Materials Transactions, JIM*, 33(1):1–6.

Kaneko, S., Aramaki, S., Arai, K., Takahashi, Y., Adachi, H., and Yanagisawa, K. (1996). Multi-freedom tube type manipulator with SMA plate. *Journal of Intelligent Material Systems and Structures*, 7(3):331–335.

Kapanen, A., Ilvesaro, J., Danilova, A., Ryhänen, J., Lehenkari, P., and Tuukkanen, J. (2002). Behaviour of Nitinol in osteoblast-like ROS-17 cell cultures. *Biomaterials*, 23(3):645–650.

Kauffman, G. and Mayo, I. (1997). The story of Nitinol: The serendipitous discovery of the memory metal and its applications. *The Chemical Educator*, 2(2):1–21.

Kemmler, R., Lipka, A., and Ramm, E. (2005). Large deformations and stability in topology optimization. *Structural and Multidisciplinary Optimization*, 30(6):459–476.

Khuri, A. I. and Cornell, J. A. (1996). *Response Surfaces, Design and Analyses, Second Edition, Revised and Expanded*. Marcel Dekker, Inc., New York.

Kim, B., Lee, S., Park, J., and Park, J. (2005). Design and fabrication of a locomotive mechanism for capsule-type endoscopes using shape memory alloys (SMAs). *IEEE/ASME Transactions on Mechatronics*, 10(1):77–86.

Kleiber, M., Antúnez, H., Hien, T., and Kowalczyk, P. (1997). *Parameter Sensitivity in Nonlinear Mechanics: Theory and Finite Element Computations*. John Wiley & Sons Ltd.

Kohl, M. (2004). *Shape Memory Microactuators*. Springer-Verlag Berlin Heidelberg.

Kohl, M., Brugger, D., Ohtsuka, M., and Takagi, T. (2004). A novel actuation mechanism on the basis of ferromagnetic SMA thin films. *Sensors and Actuators A: Physical*, 114(2–3):445–450. Selected papers from Transducers 03.

Kohl, M., Dittmann, D., Quandt, E., and Winzek, B. (2000a). Thin film shape memory microvalves with adjustable operation temperature. *Sensors and Actuators A: Physical*, 83(1-3):214–219.

Kohl, M., Just, E., Pfleging, W., and Miyazaki, S. (2000b). SMA microgripper with integrated antagonism. *Sensors and Actuators A: Physical*, 83(1-3):208–213.

Kohl, M. and Skrobanek, K. (1998). Linear microactuators based on the shape memory effect. *Sensors and Actuators A: Physical*, 70(1-2):104–111.

Kohl, M., Skrobanek, K., and Miyazaki, S. (1999). Development of stress-optimized shape memory microvalves. *Sensors and Actuators, A: Physical*, 72(3):243–250.

Kosa, G., Shoham, M., and Zaaroor, M. (2005). Propulsion of a swimming micro medical robot. In *Proceedings of the IEEE International ConferenceRobotics and Automation, April 18–22, Barcelona, Spain*.

Kreisselmeier, G. and Steinhauser, R. (1983). Application of vector performance optimization to a robust control loop design for a fighter aircraft. *International Journal of Control*, 37:251–284.

Kwak, J. and Cho, S. (2005). Topological shape optimization of geometrically nonlinear structures using level set method. *Computers & Structures*, 83(27):2257–2268.

Langelaar, M. and Van Keulen, F. (2004a). Design optimization of shape memory alloy structures. In *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY*. AIAA paper 2004-4414.

Langelaar, M. and Van Keulen, F. (2004b). Modeling of a shape memory alloy active catheter. In *12th AIAA/ASME/AHS Adaptive Structures Conference, Palm Springs, CA*. AIAA paper 2004-1653.

Langelaar, M. and Van Keulen, F. (2004c). A simple R-phase transformation model for engineering purposes. *Materials Science and Engineering: A - Structural Materials: Properties, Microstructure and Processing*, 378(1–2):507–512.

Langelaar, M. and Van Keulen, F. (2006). Sensitivity analysis of shape memory alloy shells. In *Third European Conference on Computational Mechanics, Lisbon, Portugal, 5–8 June*.

Leclercq, S., Lexcellent, C., Tobushi, H., and Lin, H. P. (1994). Thermodynamical modelling of recovery stress associated with R-phase transformation in TiNi shape memory alloys. *Materials Transactions, JIM*, 35(5):325–331.

Lexcellent, C., Tobushi, H., Ziolkowski, A., and Tanaka, K. (1994). Thermodynamical model of reversible R-phase transformation in TiNi shape memory alloy. *International Journal of Pressure Vessels and Piping*, 58(1):51–57.

Liang, C. and Rogers, C. (1992a). Design of shape memory alloy actuators. *Journal of Mechanical Design*, 114:223–230.

Liang, C. and Rogers, C. (1992b). A multidimensional constitutive model for shape memory alloys. *Journal of Engineering Mathematics*, 26(3):429–443.

Liang, C. and Rogers, C. (1997). Design of shape memory alloy actuators (reprinted from journal of mechanical design, vol 114, pg 223-230, 1993). *Journal of Intelligent Material Systems and Structures*, 8(4):303–313.

Lim, G., Minami, K., Sugihara, M., Uchiyama, M., and Esashi, M. (1995). Active catheter with multi-link structure based on silicon micromachining. In *Micro Electro Mechanical Systems, 1995, MEMS '95, Proceedings*, pages 116–121. IEEE Catalog Number = 95CH35754.

Lim, G., Park, K., Sugihara, M., Minami, K., and Esashi, M. (1996). Future of active catheters. *Sensors and Actuators A: Physical*, 56(1–2):113–121.

Liu, Z., Korvink, J., and Huang, R. (2005). Structure topology optimization: fully coupled level set method via FEMLAB. *Structural and Multidisciplinary Optimization*, 29(6):407–417.

Lombardi, M. and Haftka, R. T. (1998). Anti-optimization technique for structural design under load uncertainties. *Computer Methods in Applied Mechanics and Engineering*, 157:19–31.

Lu, T., Hutchinson, J., and Evans, A. (2001). Optimal design of a flexural actuator. *Journal of the Mechanics and Physics of Solids*, 49(9):2071–2093.

Lutz, M. (2001). *Programming Python*. O'Reilly.

Lyness, J. (1967). Numerical algorithms based on the theory of complex variables. In *Proceedings of the 22nd ACM National Conference, Thomas Book, Co., Washington DC*, pages 124–134.

Lyness, J. and Moler, C. (1967). Numerical differentiation of analytic functions. *SIAM Journal on Numerical Analysis*, 4:202–210.

Mack, M. (2001). Minimally invasive and robotic surgery. *Journal of the American Medical Association*, 285(5):568–572.

Madden, J., Vandesteeg, N., Anquetil, P., Madden, P., Takshi, A., Pytel, R., Lafontaine, S., Wieringa, P., and Hunter, I. (2004). Artificial muscle technology: Physical principles and naval prospects. *IEEE Journal of Oceanic Engineering*, 29(3):706–728.

Martins, J. and Poon, N. (2005). On structural optimization using constraint aggregation. In *6th World Congress on Structural and Multidisciplinary Optimization, Rio de Janeiro, Brazil, May 30 - June 3*.

Matsui, K. and Terada, K. (2004). Continuous approximation of material distribution for topology optimization. *International Journal for Numerical Methods in Engineering*, 59(14):1925–1944.

Mattheck, C. and Burkhardt, S. (1990). A new method of structural shape optimization based on biological growth. *International Journal of Fatigue*, 12(3):185–190.

Maute, K., Schwarz, S., and Ramm, E. (1998). Adaptive topology optimization of elastoplastic structures. *Structural Optimization*, 15(2):81–91.

Maute, K., Schwarz, S., and Ramm, E. (1999). Structural optimization - the interaction between form and mechanics. *Zeitschrift für Angewandte Mathematik und Mechanik*, 79(10):651–673.

Mineta, T., Mitsui, T., Watanabe, Y., Kobayashi, S., Haga, Y., and Esashi, M. (2001). Batch fabricated flat meandering shape memory alloy actuator for active catheter. *Sensors and Actuators A: Physical*, 88(2):112–120.

Mineta, T., Mitsui, T., Watanabe, Y., Kobayashi, S., Haga, Y., and Esashi, M. (2002). An active guide wire with shape memory alloy bending actuator fabricated by room temperature process. *Sensors and Actuators A-Physical*, 97(8):632–637.

Miyazaki, S. and Wayman, C. (1988). The R-phase transition and associated shape memory mechanism in Ti-Ni single crystals. *Acta Metallurgica*, 36(1):181–192.

Moore, A. (2001). Of silicon and submarines. *EMBO Reports*, 2:367–370.

Morgan, N. and Friend, C. (2001). A review of shape memory stability in NiTi alloys. *Journal De Physique IV*, 11(8):325–332. In: Proceedings of Fifth European Symposium on Martensitic Transformations and Shape Memory Alloys - ESOMAT 2000 - Villa Olmo, Como, Italy.

Myers, R. H. and Montgomery, D. C. (1995). *Response Surface Methodology; Process and Product Optimization using Designed Experiments*. John Wiley & Sons Inc., New York.

Nae, F., Matsuzaki, Y., and Ikeda, T. (2003). Micromechanical modeling of polycrystalline shape-memory alloys including thermo-mechanical coupling. *Smart Materials & Structures*, 12:6–17.

Ng, E. and Chua, L. (2002a). Prediction of skin burn injury. part 1: Numerical modelling. *Proceedings of the Institution of Mechanical Engineers Part H - Journal of Engineering in Medicine*, 216(3):157–170.

Ng, E. and Chua, L. (2002b). Prediction of skin burn injury. part 2: Parametric and sensitivity analysis. *Proceedings of the Institution of Mechanical Engineers Part H - Journal of Engineering in Medicine*, 216(3):171–183.

Ohta, R. (2001). Results of R&D on catheter-type micromachine. In *Proceedings of International Symposium on Micromechatronics and Human Science, Nagoya, Japan*, pages 5–12.

Olhoff, N., Rasmussen, J., and Lund, E. (1993). A method of "exact" numerical differentiation for error elimination in finite-element-based semi-analytical shape sensitivity analyses. *Mechanics of Structures & Machines*, 21:1–66.

Otsuka, K. and Wayman, C., editors (1998). *Shape Memory Materials*. Cambridge University Press.

Papalambros, P. Y. and Wilde, D. J. (2000). *Principles of Optimal Design*. Cambridge University Press.

Park, K. and Esashi, M. (1999a). An active catheter with integrated circuit for communication and control. In *Twelfth IEEE International Conference on Micro Electro Mechanical Systems, 1999. MEMS '99*, pages 400–405. IEEE Catalog Number = 99CH36291.

Park, K. and Esashi, M. (1999b). A multilink active catheter with polyimide-based integrated CMOS interface circuits. *Journal of Microelectromechanical Systems*, 8(4):349–357.

Patronik, N., Zenati, M., and Riviere, C. (2004). Crawling on the heart: A mobile robotic device for minimally invasive cardiac interventions. *Lecture Notes in Computer Science*, 3217:9–16.

Pedersen, C. (2002). Revisiting topology optimization of continuum structures with elastoplastic response. In *15th Nordic Seminar of Computational Mechanics, 18–19 October, Aalborg, Denmark*.

Pedersen, P. (1998). Some general optimal design results using anisotropic, power law nonlinear elasticity. *Structural Optimization*, 15(2):73–80.

Pedersen, P. and Taylor, J. (1993). Optimal design based on power law non-linear elasticity. In *Optimal design with advanced materials: the Frithiof Niordson volume*, pages 51–66. Elsevier, Amsterdam. Proceedings of the IUTAM Symposium on Optimal Design with Advanced Materials, Lyngby, Denmark, 18-20 August, 1992.

Peirs, J., Reynaerts, D., and Van Brussel, H. (2001). A miniature manipulator for integration in a self-propelling endoscope. *Sensors and Actuators A: Physical*, 92(1–3):343–349.

Pelrine, R., Kornbluh, R., Joseph, J., Heydt, R., Pei, Q., and Chiba, S. (2000). High-field deformation of elastomeric dielectrics for actuators. *Materials Science and Engineering: C*, 11(2):89–100.

Poulsen, T. (2003). A new scheme for imposing a minimum length scale in topology optimization. *International Journal for Numerical Methods in Engineering*, 57(6):741–760.

Power, H. (1995). *Bio-fluid mechanics*. Computational Mechanics Publications, Southampton.

Raniecki, B., Miyazaki, S., Tanaka, K., Dietrich, L., and Lexcellent, C. (1999). Deformation behavior of Ti-Ni shape memory alloy undergoing R-phase reorientation in torsion-tension (compression) tests. *Archives of Mechanics*, 51(5):745–784.

Rattner, D. (1999). Future directions in innovative minimally invasive surgery. *The Lancet*, 353(1001):S12–S15.

Rebello, K. (2004). Applications of MEMS in surgery. *Proceedings of the IEEE*, 92(1):43–55.

Reinhardt, A. and Götzen, R. (2006). 3D-CSP and RMPD technologies. See *http://www.microtec-d.com*.

Rejzner, J., Lexcellent, C., and Raniecki, B. (2002). Pseudoelastic behaviour of shape memory alloy beams under pure bending: experiments and modelling. *International Journal of Mechanical Sciences*, 44(4):665–686.

Ruzzu, A., Bade, K., Fahrenberg, J., and Maas, D. (1998). Positioning system for catheter tips based on an active microvalve system. *Journal of Micromechanics and Microengineering*, 8(2):161–164.

Saitou, K., Izui, K., Nishiwaki, S., and Papalambros, P. (2005). A survey of structural optimization in mechanical product development. *Journal of Computing and Information Science in Engineering*, 5(3):214–226.

Santa, A. D., Mazzoldi, A., and DeRossi, D. (1996). Steerable microcatheters actuated by embedded conducting polymer structures. *Journal of Intelligent Material Systems and Structures*, 7(3):292–300. Presented at the International Symposium of Microsystems, Intelligent Materials and Robots, Sendai, Japan, Sept. 27–29, 1995.

Sawada, T., Tobushi, H., Kimura, K., Hattori, T., Tanaka, K., and Lin, P. (1993). Stress-strain-temperature relationship associated with the R-phase transformation in TiNi shape memory alloy (influence of shape memory processing temperature). *JSME International Journal Series A: Mechanics and Material Engineering*, 36(1):395–401.

Schiemann, M., Killmann, R., Kleen, M., Abolmaali, N., Finney, J., and Vogl, T. (2004). Vascular guide wire navigation with a magnetic guidance system: Experimental results in a phantom. *Radiology*, 232:475–481.

Schwarz, S. and Ramm, E. (2001). Sensitivity analysis and optimization for nonlinear structural response. *Engineering Computations*, 18(3):610–641.

Sendoh, M., Ajiro, N., Ishiyama, K., Inoue, M., Arai, K., Hayase, T., and Akedo, J. (1999). Effect of machine shape on swimming properties of the spiral-typemagnetic micro-machine. *IEEE Transactions on Magnetics*, 35(5, Part 2):3688–3690.

Sewa, S., Onishi, K., Asaka, K., Fujiwara, N., and Oguro, K. (1998). Polymer actuator driven by ion current at low voltage, applied to catheter system. In *The Eleventh Annual International Workshop on Micro Electro Mechanical Systems, 1998. MEMS 98. Proceedings*, pages 148–153. IEEE Catalog Number = 98CH36176.

Sherman, L., Taylor, A., Green, L., Newman, P., Hou, G., and Korivi, V. (1994). First- and second-order aerodynamic sensitivity derivatives via automatic differentiation with incremental iterative methods. In *5th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analaysis and Optimization*. AIAA 94–4262, Panama City Beach, Florida, September 7–9.

Shutov, M., Howard, D., Sandoz, E., Sirota, J., Smith, R., and Collins, S. (2004). Electrostatic inchworm microsystem with long range translation. *Sensors and Actuators A: Physical*, 114(2–3):379–386. Selected papers from Transducers 03.

Sigmund, O. (1997). On the design of compliant mechanisms using topology optimization. *Mechanics of Structures and Machines*, 25(4):495–526.

Sigmund, O. (2001). Design of multiphysics actuators using topology optimization - Part I: One-material structures. *Computer Methods in Applied Mechanics and Engineering*, 190(49–50):6577–6604.

Sigmund, O. and Petersson, J. (1998). Numerical instabilities in topology optimization: A survey on procedures dealing with checkerboards, mesh-dependencies and local minima. *Structural Optimization*, 16(1):68–75.

Silva, J., Medeiros, W., and Barbosa, V. (2001). Pitfalls in nonlinear inversion. *Pure and Applied Geophysics*, 158(5/6):945–964.

Skrobanek, K., Kohl, M., and Miyazaki, S. (1997). Stress-optimized shape memory microvalves. In *Proceedings of the 1997 10th Annual International Workshop on Micro Electro Mechanical Systems, MEMS, Nagoya, Japan*, pages 256–261.

Squire, W. and Trapp, G. (1998). Using complex variable to estimate derivatives of real functions. *SIAM Rev.*, 40(1):110–112.

Stevens, J. and Buckner, G. (2005). Actuation and control strategies for miniature robotic surgical systems. *Journal of Dynamic Systems Measurement and Control - Transactions of the ASME*, 127(4):537–549.

Stolpe, M. and Svanberg, K. (2003). A note on stress-constrained truss topology optimization. *Structural and Multidisciplinary Optimization*, 25(1):62–64.

Svanberg, K. (1987). MMA - method of moving asymptotes - a new method for structural optimization. *International Journal for Numerical Methods in Engineering*, 24:359–373.

Takizawa, H., Tosaka, H., Ohta, R., Kaneko, S., and Ueda, Y. (1999). Development of a microfine active bending catheter equipped with MIF tactile sensors. In *Twelfth IEEE International Conference on Micro Electro Mechanical Systems, 1999. MEMS '99*, pages 412–417. IEEE Catalog Number = 99CH36291.

Tanaka, K., Kobayashi, S., and Sato, Y. (1986). Thermomechanics of transformation pseudoelasticity and shape memory effect in alloys. *International Journal of Plasticity*, 2:59–72.

Tangwongsan, C., Will, J., Webster, J., Meredith, Jr., K., and Mahvi, D. (2004). In vivo measurement of swine endocardial convective heat transfer coefficient. *IEEE Transactions on Biomedical Engineering*, 51(8):1478–1486.

Tendick, F., Sastry, S., Fearing, R., and Cohn, M. (1998). Applications of micromechatronics in minimally invasive surgery. *IEEE-ASME Transactions on Mechatronics*, 3(1):34–42.

Tobushi, H., Tanaka, K., Kimura, K., Hori, T., and Sawada, T. (1992). 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*, 35(3):278–284.

Tobushi, H., Yamada, S., Hachisuka, T., Ikai, A., and Tanaka, K. (1996). Thermomechanical properties due to martensitic and R-phase transformations of TiNi shape memory alloy subjected to cyclic loadings. *Smart Materials & Structures*, 5(6):788–795.

Todoroki, T. (1990). Shape memory sensor and actuator for air conditioners. In Duerig, T., Melton, K., Stöckel, D., and Wayman, C., editors, *Engineering aspects of shape memory alloys*, pages 315–329. Butterworth-Heinemann, London.

Toropov, V. V., Filatov, A. A., and Polynkine, A. A. (1993a). Multiparameter structural optimization using FEM and multipoint explicit approximations. *Structural Optimization*, Vol. 6:pp. 7–14.

Toropov, V. V., Filatov, A. A., and Polynkine, A. A. (1993b). Multiparameter structural optimization using fem and multipoint explicit approximations. *Structural Optimization*, 6:7–14.

Toropov, V. V., Van Keulen, F., Markine, V. L., and Alvarez, L. F. (1999a). Multipoint approximations based on response surface fitting: a summary of recent results. In *In: V.V. Toropov (ed.) Engineering Design Optimization, Proceedings of 1st, ASMO UK /ISSMO conference, 8-9 July*, pages 371–380, Bradford, UK. University Press.

Toropov, V. V., Van Keulen, F., Markine, V. L., and Alvarez, L. F. (1999b). Multipoint approximations based on response surface fitting: a summary of recent results. In *In: V.V. Toropov (ed.) Engineering Design Optimization, Proceedings of 1st, ASMO UK /ISSMO conference, 8-9 July*, pages 371–380, Bradford, UK. University Press.

Troisfontaine, N., Bidaud, P., and Larnicol, M. (1999). Optimal design of micro-actuators based on SMA wires. *Smart Materials & Structures*, 8(2):197–203.

Van Humbeeck, J. (1999). Non-medical applications of shape memory alloys. *Materials Science and Engineering: A - Structural Materials: Properties, Microstructure and Processing*, 273-275(Special Issue):134–148.

Van Keulen, F. and Booij, J. (1996). Refined consistent formulation of a curved triangular finite rotation shell element. *International Journal for Numerical Methods in Engineering*, 39(16):2803–2820.

Van Keulen, F., Damveld, H. J., Elishakoff, I., and Toropov, V. V. (2001). Bounded-but-unknown uncertainties in design optimization. In *Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit, Seattle, WA, U.S.A., Paper 1242, 16-19 April*. CD-ROM.

Van Keulen, F. and De Boer, H. (1998a). Rigorous improvement of semi-analytical design sensitivities by exact differentiation of rigid body motions. *Int. J. Numerical Methods in Engineering*, 42:71–91.

Van Keulen, F. and De Boer, H. (1998b). Rigorous improvement of semi-analytical design sensitivities by exact differentiation of rigid body motions. *International Journal for Numerical Methods in Engineering*, 42:71–91.

Van Keulen, F., Haftka, R., and Kim, N. (2005). Review of options for structural design sensitivity analysis. Part 1: Linear systems. *Computer Methods in Applied Mechanics and Engineering*, 194(30–33):3213–3243.

Van Keulen, F. and Toropov, V. (1997a). New developments in structural optimization using adaptive mesh refinement and multi-point approximations. *Engineering Optimization*, 29:217–234.

Van Keulen, F. and Toropov, V. V. (1997b). New developments in structural optimization using adaptive mesh refinement and multipoint approximations. *Engineering Optimization*, 29:217–234.

Van Keulen, F. and Toropov, V. V. (1998). Multipoint approximations for structural optimization problems with noisy response functions. In *ISSMO/NASA/AIAA First Internet Conference on Approximations and Fast Reanalysis in Engineering Optimization, 14-27 June*. CD-ROM.

Van Keulen, F. and Toropov, V. V. (1999). The multipoint approximation method in a parallel computing environment. *ZAMM*, 79(S1):S67–S70.

Van Keulen, F. and Vervenne, K. (2002). Gradient-enhanced response surface building. In *Proceedings 9th AIAA/ISSMO Symposium and Exhibit on Multidisciplinary Analysis and Optimization, 4-6 September 2002, Atlanta, Georgia, USA, paper AIAA-2002-5455*.

Van Keulen, F. and Vervenne, K. (2004). Gradient-enhanced response surface building. *Struct. Multidisc. Optim.*, 27(5):337–351.

Vervenne, K. and Van Keulen, F. (2002). An alternative approach to response surface building using gradient information. In *Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit, Denver, CO, USA, Paper 1584, 22-25 April*. CD-ROM.

Vidal, C. and Haber, R. (1993). Design sensitivity analysis for rate-independent elastoplasticity. *Computer Methods in Applied Mechanics and Engineering*, 107(3):393–431.

Vogel, S. (1994). *Life in moving fluids*. Princeton University Press, Princeton.

Wang, S. and Wang, M. (2006). Radial basis functions and level set method for structural topology optimization. *International Journal for Numerical Methods in Engineering*, 65(12):2060–2090.

Wang, X., Wang, M., and Guo, D. (2004). Structural shape and topology optimization in a level-set-based framework of region representation. *Structural and Multidisciplinary Optimization*, 27(1–2):1–19.

Wautelet, M. (2001). Scaling laws in the macro-, micro- and nanoworlds. *European Journal of Physics*, 22(6):601–611.

Yeh, R., Hollar, S., and Pister, K. (2002). Single mask, large force, and large displacement electrostatic linear inchworm motors. *Journal of Microelectromechanical Systems*, 11(4):330–336.

Yoon, G., Joung, Y., and Kim, Y. (2005). Optimal layout design using the element connectivity parameterization method: Application to three dimensional geometrical nonlinear structures. In *6th World Congress on Structural and Multidisciplinary Optimization, Rio de Janeiro, Brazil, May 30 - June 3*.

Yoon, G., Joung, Y., and Kim, Y. (2006a). Optimal layout design for three dimensional geometrical nonlinear structures using the element connectivity parameterization. Submitted.

Yoon, G. and Kim, Y. (2004). Topology optimization for multiphysics problems using the element connectivity parameterization method. In *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 30 August - 1 September 2004, Albany, New York*. AIAA 2004-4413.

Yoon, G. and Kim, Y. (2005). Element connectivity parameterization for topology optimization of geometrically nonlinear structures. *International Journal of Solids and Structures*, 42(7):1983–2009.

Yoon, G., Kim, Y., Bendsoe, M., and Sigmund, O. (2004). Hinge-free topology optimization with embedded translation-invariant differentiable wavelet shrinkage. *Structural and Multidisciplinary Optimization*, 27(3):139–150.

Yoon, G., Langelaar, M., Joung, Y., Van Keulen, F., and Kim, Y. (2006b). The theoretical aspects of the internal element connectivity parameterization: The application to two dimensional linear and nonlinear structures. In preparation.

Zhou, M., Shyy, Y., and Thomas, H. (2001). Checkerboard and minimum member size control in topology optimization. *Structural and Multidisciplinary Optimization*, 21(2):152–158.

Zienkiewicz, O. and Taylor, R. (2000). *The Finite Element Method*, volume Volume 2, Solid Mechanics. Butterworth-Heinemann.

Summary

Design Optimization of Shape Memory Alloy Structures

This thesis presents various methods to apply design optimization principles to the design of shape memory alloy (SMA) structures. SMAs are materials that exhibit a solid state phase transformation, influenced by changes in temperature and stress, which can be used for actuation. In comparison to other active materials, SMAs can generate relatively large strains and stresses, and are therefore of interest for a variety of (micro-)actuation applications. Designing an effective SMA structure, however, presents a challenge, because of the complexity of the SMA material behavior and the fact that often the electrical, thermal and mechanical properties of a structure have to be considered simultaneously. For this reason, this research aims to explore how and to what extent design optimization techniques can contribute to the solution of SMA design problems. Little previous work has been reported on this topic.

The approach taken to investigate the applicability and effectiveness of various design optimization approaches for SMA structures is to apply these to representative practical cases. In this thesis, the focus is on applications emerging from the medical field, where the popularity of minimally invasive therapy leads to a demand for miniaturized instruments with enhanced functionality. Because of their mentioned properties, SMAs are promising candidates for use as active materials in these instruments. The shape memory effect due to the so-called austenite/R-phase transformation in nickel-rich NiTi alloys is considered in particular, because it is characterized by a small hysteresis and excellent cyclic stability, making it well suited for the targeted applications. Specifically, SMA miniature grippers and steerable catheters are used in the case studies.

In order to develop a generic and versatile design methodology, finite element modeling is used to evaluate the functionality of SMA structures. In this thesis, a nonlinear quasi-static finite element model involving a sequentially coupled electrical, thermal and mechanical analysis is applied. First, a novel three-dimensional model is developed to describe the SMA material behavior due to the austenite/R-phase transformation. An empirical approach is used, combined with the fundamental properties of the underlying phase transformation, to capture the essence of the material behavior. The limited thermal operating range required for the applications allows substantial simplifications, which result in a history-independent model. This history-independent nature reduces the computational cost and complexity of design sensitivity analysis in comparison to conventional SMA models, which allows the use of efficient gradient-based optimization procedures. The model is illustrated by the analysis of a novel concept for a steerable catheter, equipped with integrated SMA actuation sections.

Subsequently, shape optimization studies of a miniature SMA gripper are presented. A direct optimization approach is first explored, which does not require the computation of design sensitivities. The stroke of gripper jaws is maximized, subject to constraints on the operating temperature and the maximum strain, in order to remain within the range of validity of the constitutive model. The Multi-point Approximation Method is used to perform the optimization, which relies on response surface fitting in combination with a move limit strategy. In combination with parallel computing to reduce the duration of the procedure, this method is demonstrated to be an effective way to generate optimized gripper designs.

The same miniature gripper is also used in an optimization study in which the influence of uncertainties is considered as well. Since usually in the initial design stage reliable probabilistic data to describe the nature of uncertainties is not available, the uncertainties considered in this study are characterized only by the bounds of the domain in which the uncertain parameters may vary. The design optimization problem is formulated such, that the optimization aims to maximize the expected performance, while the worst combination of uncertainties still yields a feasible design. By means of an adapted anti-optimization approach, which avoids the use of an expensive nested optimization procedure by alternating between main- and anti-optimization cycles, and in combination with parallel computing, also this non-deterministic SMA shape optimization procedure was performed successfully against strongly reduced computational cost. Again the Multi-point Approximation Method was used as the optimizer.

In order to efficiently handle optimization problems involving more design variables and expensive evaluations, gradient-based methods are preferred over direct approaches. Therefore, various options for sensitivity analysis of the SMA finite element models used in this thesis are investigated. A comparison is made between full finite differences, finite differences in combination with fast reanalysis, semi-analytical and refined semi-analytical direct differentiation methods. The implicit nature of the SMA material model and the coupling between the electrical,

thermal and mechanical problems require special attention in the semi-analytical cases. The study reveals that all methods yield sensitivities of comparable accuracy, and that the semi-analytical approaches have only a slight advantage in terms of computational effort over finite differences with fast reanalysis, while the implementation of the latter is considerably less involved. The relative efficiency of the sensitivity analysis for the considered class of SMA models is to a large extent due to the history-independent nature of the used SMA material model.

The availability of design sensitivities enables shape optimization of SMA structures using gradient-based algorithms. The effectiveness of this approach is demonstrated using optimization case studies involving the SMA miniature gripper, as well as the SMA steerable catheter model introduced earlier. The latter contains nonlinear multi-point constraints that are used to enforce symmetry and periodicity conditions, which are handled in the model using an augmented Lagrangian approach. For this situation, finite differences in combination with fast reanalysis is selected as the preferred method for sensitivity analysis. A correction is introduced to account for the finite error in the constraint satisfaction in the nominal case. In combination with stricter convergence settings for the perturbed cases, this correction is shown to clearly improve the accuracy of the results. Modified Newton iterations are used in the reanalysis, which further reduces the required computational effort. The algorithms considered are Sequential Quadratic Programming and the Method of Moving Moving Asymptotes. All gradient-based optimization procedures turn out to require far less function evaluations in comparison to the previously used direct method, and the optimum design is generally located more precisely.

Finally, also topology optimization of SMA thermal actuators is considered in this thesis. In topology optimization, no prior assumptions are made regarding the geometrical layout of the design, and only a design domain and boundary conditions are defined. In this work, a novel approach to parameterize the layout design problem is applied, based on modifying the connectivity between finite elements instead of changing their material properties. This Element Connectivity Parameterization concept avoids difficulties that conventional approaches encounter in this setting, such as the ambiguity of defining an interpolation of material properties, loss of numerical stability due to excessive distortion of elements with low stiffness, and complex sensitivity analysis. With this technique, in combination with the Method of Moving Asymptotes, topology optimization of SMA structures could be performed successfully. Next to the capacity to generate effective SMA actuator structures from a neutral initial configuration, it was also found that this topology optimization procedure was able to achieve large performance improvements by adapting a given design.

The combined modeling and design approaches reported in this thesis constitute a set of versatile and effective design procedures for SMA structures, that shows great potential for application in a wide variety of specific applications. Particularly in the development of future SMA devices with complex geometries, such as, for example, enhanced miniature medical instruments, it is expected

that designers can benefit considerably from systematic design optimization procedures as developed in this research. Furthermore, the element connectivity parameterization approach used in the topology optimization of SMA thermal actuators also offers advantages over conventional approaches in other topology optimization problems involving nonlinear material behavior. Finally, the novel active catheter designs proposed in this research showed good performance after optimization, and could very well serve as a basis for future prototypes.

Samenvatting

Ontwerpoptimalisatie van Geheugenmetaalconstructies

Dit proefschrift presenteert verscheidene methoden om ontwerpoptimalisatie toe te passen op geheugenmetaalconstructies. Geheugenmetalen zijn metaallegeringen die een fasetransformatie kunnen ondergaan in de vaste toestand, onder invloed van veranderingen in temperatuur en mechanische spanning. De rekken die met deze transformatie gepaard gaan kunnen gebruikt worden voor actuatie.

In vergelijking met andere actieve materialen zijn geheugenmetalen in staat relatief grote rekken en spanningen te genereren. Er bestaat hierom een grote interesse voor het gebruik van geheugenmetalen in een scala van (micro-)actuatietoepassingen. Echter, het ontwerpen van een goed functionerende geheugenmetaalconstructie wordt bemoeilijkt door twee factoren. Ten eerste is het materiaalgedrag van geheugenmetaal niet eenvoudig te voorspellen, en ten tweede dienen bij het ontwerpen veelal de electrische, thermische en mechanische eigenschappen van een constructie tegelijkertijd beschouwd te worden. Dit onderzoek heeft daarom tot doel te onderzoeken op welke wijze en in hoeverre ontwerpoptimalisatiemethoden het ontwerpen met geheugenmetaal kunnen vergemakkelijken.

De toepasbaarheid en effectivititeit van verscheidene ontwerpoptimalisatiemethoden voor geheugenmetaalconstructies is onderzocht door deze gebruiken in representatieve toepassingen. Dit proefschrift richt zich hierbij met name op opkomende toepassingen in de gezondheidszorg, waar de populariteit van minimaal invasieve behandelingen een grote vraag naar geminiaturiseerde instrumenten met uitgebreide functionaliteit tot gevolg heeft. Geheugenmetaal is een aantrekkelijke optie voor het realiseren van de actuatie in dergelijke instrumenten. De nadruk in dit onderzoek ligt op het geheugenmetaalgedrag ten gevolge van de

zogenaamde austeniet/R-fase transformatie in NiTi legeringen. Deze transformatie vindt plaats binnen een relatief klein temperatuurbereik, heeft een geringe hysterese en een uitstekende cyclische stabiliteit. Dit zijn zeer geschikte kenmerken voor actuatiedoeleinden in het beoogde toepassingsgebied. Praktische geheugenmetaaltoepassingen die in dit proefschrift met name beschouwd worden zijn geminiaturiseerde grijpinstrumenten en bestuurbare catheters.

Om een breed toepasbare en veelzijdige ontwerpmethodologie te ontwikkelen, worden eindige-elementenmodellen gebruikt om het functioneren van geheugenmetaalconstructies te evalueren. In dit proefschrift wordt een niet-lineair quasi-statisch model gebruikt, dat bestaat uit een sequentieel gekoppelde electrische, thermische en mechanische analyse. Een drie-dimensionaal materiaalmodel is ontwikkeld dat het gedrag van NiTi beschrijft ten gevolge van de austeniet/R-fase transformatie. Een empirische benadering is toegepast, in combinatie met de fundamentele eigenschappen van de onderliggende fasetransformatie, om zo de essentie van het materiaalgedrag in een model te vatten. Het beperkte temperatuurbereik dat de toepassingen vereisen en de geringe hysterese maken belangrijke vereenvoudigingen mogelijk. Dit leidt tot een pad-onafhankelijk model, dat bij uitstek geschikt is voor ontwerpoptimalisatie. Het ontwikkelde materiaalmodel is toegepast in de analyse van een nieuw concept van een bestuurbare catheter met geïntegreerde actuatiegedeeltes van geheugenmetaal.

Vervolgens worden vormoptimalisatiestudies van een geminiaturiseerd grijpinstrument van geheugenmetaal behandeld. Als eerste wordt een directe optimalisatie beschouwd, waarin geen ontwerpafgeleiden worden gebruikt. De uitslag van de bekken van de grijper wordt gemaximaliseerd, met inachtneming van restricties betreffende de werktemperaturen en de maximale rek. Een Meerpunts-Benaderings-Methode wordt gebruikt voor het uitvoeren van de optimalisatie. Deze methode maakt gebruik van het construeren van analytische benaderingen van het modelgedrag, hetgeen de rekenkosten reduceert. In combinatie met parallelle berekeningen om de totale duur van de procedure te verminderen, wordt aangetoond dat deze methode een effectieve manier biedt om geoptimaliseerde grijperontwerpen te genereren.

Hetzelfde miniatuur grijpinstrument wordt tevens gebruikt in een optimalisatiestudie waarin ook de invloed van onzekerheden wordt beschouwd. Doorgaans zijn in de beginfase van het ontwerpproces onvoldoende betrouwbare statistische gegevens beschikbaar om onzekerheden te kunnen karakteriseren. Hierom worden in deze studie onzekerheden slechts gedefinieerd door de grenzen waarbinnen de waarden van de onzekere parameters kunnen variëren. Het ontwerpoptimalisatieprobleem is zodanig opgesteld dat het tracht de verwachte prestaties te maximaliseren, terwijl de meest ongunstigste combinatie van onzekerheden nog steeds een toelaatbaar ontwerp geeft. Als optimalisatieprocedure is een aangepaste anti-optimalisatie benadering gebruikt, gebaseerd op een Meerpunts-Benaderings-Methode. Hierbij wordt afgewisseld tussen hoofd- en anti-optimalisatiecycli, wat de totale rekentijd sterk reduceert. Tevens zijn ook hier parallelle rekenmethoden toegepast, waarmee deze niet-deterministische vormoptimalisatie van een geheu-

genmetaalconstructie succesvol afgerond kon worden binnen een beperkte tijdsduur.

Naarmate optimalisatieproblemen meer ontwerpvariabelen bevatten en modellen meer rekentijd vergen, wordt het aantrekkelijker om, in plaats van directe, gradiënt-gebaseerde methoden te gebruiken. De benodigde gradiëntinformatie wordt berekend middels gevoeligheidsanalyse met het gebruikte model. Beschouwd zijn globale eindige differenties, globale eindige differenties in combinatie met snelle heranalyse, semi-analytische en verfijnde semi-analytische directe differentiatiemethoden. De semi-analytische formulering vereist speciale aandacht voor het impliciete karakter van het materiaalmodel en de koppeling tussen het electrische, thermische en mechanische probleem. Deze studie laat zien dat alle beschouwde methoden resulteren in ontwerpafgeleiden van vergelijkbare nauwkeurigheid. Verder blijken semi-analytische benaderingen efficiënter dan de eindige differentie methode in combinatie met snelle heranalyse. Echter, de implementatie van de laatstgenoemde methode is aanzienlijk eenvoudiger. De relatief lage rekenkosten van de gevoeligheidsanalyse voor de beschouwde klasse van geheugenmetaalmodellen zijn overigens voor een groot deel te danken aan het padonafhankelijke karakter van het gebruikte materiaalmodel.

De beschikbaarheid van ontwerpafgeleiden maakt gradiëntgebaseerde vormoptimalisatie van geheugenmetaalconstructies mogelijk. De effectiviteit van deze benadering wordt aangetoond middels optimalisatiestudies van een geminiaturiseerd grijpinstrument van geheugenmetaal, alsmede van de eerder beschreven bestuurbare catheter. Een bijzonderheid van het model van deze catheter is, dat het niet-lineaire beperkende voorwaarden tussen de vrijheidsgraden van meerdere punten bevat. Deze worden gebruikt om de symmetrie- en periodiciteitscondities op te leggen. In de gebruikte oplossingsmethode worden deze voorwaarden behandeld volgens de zogenaamde augmented Lagrangian benadering. In deze situatie hebben globale eindige differenties in combinatie met snelle heranalyse de voorkeur voor de gevoeligheidsanalyse. Een correctie is geïntroduceerd om de invloed van een restfout in de vervulling van de nevenvoorwaarden, die resteert na het oplossen van het nominale probleem, in rekening te brengen. In combinatie met strengere convergentievoorwaarden in de geperturbeerde gevallen, leidt deze correctie tot een duidelijke verbetering van de nauwkeurigheid van de ontwerpafgeleiden. Gemodificeerde Newton-iteraties worden gebruikt in de heranalyse, op basis van de gedecomponeerde nominale systeemmatrix, hetgeen de benodigde rekentijd verder reduceert. Sequentieel Kwardratisch Programmeren en de Methode van Verschuivende Asymptoten zijn toegepast als optimalisatiealgorithmes. Deze gradiëntgebaseerde optimalisatieprocedures blijken in alle gevallen veel minder functie-evaluaties te vereisen dan de eerder gebruikte directe methode, en zijn tevens in staat het optimale ontwerp preciezer te bepalen.

Tot slot wordt ook topologie-optimalisatie van thermische actuatoren van geheugenmetaal beschouwd in dit proefschrift. In topologie-optimalisatie wordt de geometrie van het uiteindelijke ontwerp volledig vrijgelaten, en slechts het ontwerpdomein en de randvoorwaarden worden gedefiniëerd. In deze studie wordt

een nieuwe topologische parameterisatie toegepast, die is gebaseerd op het veranderen van de connectiviteit tussen eindige-elementen. Gebruikelijke topologische parameterisaties zijn daarentegen gebaseerd op het veranderen van materiaaleigenschappen van elementen. Deze Element Connectiviteits Parameterisatie gaat drie problemen die gebruikelijke benaderingen ondervinden uit de weg. Ten eerste is het niet nodig een arbitraire keuze te maken voor de interpolatie van materiaaleigenschappen. Daarnaast is deze aanpak ook ongevoelig voor numerieke instabiliteiten door buitensporige vervorming van elementen met lage stijfheid. Ten derde speelt het materiaalmodel geen rol in de gevoeligheidsanalyse, hetgeen de gevoeligheidsanalyse aanzienlijk vereenvoudigd. Met behulp van de Methode van Verschuivende Asymptoten konden innovatieve geheugenmetaalconstructies worden gegenereerd voor willekeurige belastingsgevallen. Tevens bleek de ontwikkelde topologie-optimalisatieprocedure in staat om grote prestatieverbeteringen te realiseren bij het aanpassen van een reeds gegeven ontwerp.

De modellerings- en ontwerpbenaderingen die in dit proefschrift worden beschreven vormen tezamen een verzameling van veelzijdige en effectieve ontwerpprocedures voor geheugenmetaalconstructies. Het is de verwachting dat ontwerpers voor een breed scala van specifieke geheugenmetaaltoepassingen baat zullen hebben bij de ontwikkelde systematische procedures. Dit geldt in het bijzonder voor de ontwikkeling van toekomstige geheugenmetaalconstructies met complexe geometrieën, zoals bijvoorbeeld geavanceerde en geminiaturiseerde medische instrumenten. Daarbij biedt de Element Connectiviteits Parameterisatie, zoals die toegepast is bij de topologie-optimalisatie van thermische actuatoren van geheugenmetaal, ook perspectieven voor topologie-optimalisatie waarin ander niet-lineair materiaalgedrag een rol speelt. Tenslotte behaalde het nieuwe ontwerp voor een bestuurbare catheter, zoals dat in dit onderzoek is voorgesteld, goede resultaten na ontwerpoptimalisatie. Het zou hierom uitstekend kunnen dienen als een basis voor toekomstige prototypes.

List of Publications

Publications in refereed journals and conference proceedings, related to this dissertation:

- Langelaar, M. and Van Keulen, F. (2003). A simple R-phase transformation model for engineering purposes. In *Proceedings of the European Symposium On Martensitic Transformations and Shape-Memory, ESOMAT 2003, August 17–22, Cirencester, England.*

- Gurav, S., Langelaar, M., Goosen, J., and Van Keulen, F. (2003). Different approaches to deal with Bounded-But-Unknown uncertainties: Application to MEMS. In *Proceedings of the 5^{th} World Conference on Structural and Multidisciplinary Optimization, 19–23 May, Venice, Italy.*

- Gurav, S., Langelaar, M., Goosen, J., and Van Keulen, F. (2003). Bounded-but-unknown uncertainty optimization of micro-electro-mechanical-systems. In *Proceedings of the Second M.I.T. Conference on Computational Fluid and Solid Mechanics, June 17–20, M.I.T. Cambridge, MA, U.S.A.*

- Langelaar, M. and Van Keulen, F. (2004). A simple R-phase transformation model for engineering purposes. *Materials Science and Engineering: A - Structural Materials: Properties, Microstructure and Processing*, 378(1–2):507–512.

- Langelaar, M. and Van Keulen, F. (2004). Modeling of a shape memory alloy active catheter. In *Proceedings of the 12^{th} AIAA/ASME/AHS Adaptive Structures Conference, Palm Springs, CA, U.S.A.*

- Langelaar, M. and Van Keulen, F. (2004). Design optimization of shape memory alloy structures. In *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, August 30–September 1, Albany, NY, U.S.A.*

- Langelaar, M., Yoon, G.H., Kim, Y.Y. and Van Keulen, F. (2005). Topology Optimization of Shape Memory Alloy Actuators using Element Connectivity Parameterization. In *Proceedings of the 6th World Congress on Structural and Multidisciplinary Optimization, May 30–June 3, Rio de Janeiro, Brazil.*

- Langelaar, M., Yoon, G.H., Gurav, S., Kim, Y.Y. and Van Keulen, F. (2005). Modeling and Design of Shape Memory Alloy Actuators. In *Proceedings of the 6th IEEE EuroSimE Conference, April 18–20, Berlin, Germany*, invited keynote contribution.

- Langelaar, M., Yoon, G.H., Gurav, S., Kim, Y.Y. and Van Keulen, F. (2005). Analysis and design techniques for shape memory alloy microactuators for space applications. In *Proceedings of the 5th ESA Round Table on Micro/Nano Technologies for Space, 3–5 October, Noordwijk, The Netherlands.*

- Langelaar, M. (2005). Designing Memory Metal Devices. Design optimization of shape memory alloy structures – targeted at medical and microsystem applications. In *Dingeldein, N., editor, DIMES White Papers 2005, pages 41–52, DIMES, Delft.*

- Gurav, S., Langelaar, M., Goosen, J., and Van Keulen, F. (2005). Uncertainty-based design optimization of shape memory alloy microgripper using combined cycle-based alternating anti-optimization and nested parallel computing. In *Proceedings of the 6th World Congress on Structural and Multidisciplinary Optimization, May 30–June 3, Rio de Janeiro, Brazil.*

- Zeoli, M., Van Keulen, F. and Langelaar, M. (2005). Fast Reanalysis of Geometrically Nonlinear Problems After Shape Modifications. In *Proceedings of the 6th World Congress on Structural and Multidisciplinary Optimization, May 30–June 3, Rio de Janeiro, Brazil.*

- Langelaar, M. and Van Keulen, F. (2006). Sensitivity analysis of shape memory alloy shells. In *Proceedings of the 3rd European Conference on Computational Mechanics, 5–8 June, Lisbon, Portugal.*

- Langelaar, M. and Van Keulen, F. (2006). Sensitivity analysis and optimization of a shape memory alloy gripper. In *Proceedings of the 8th International Conference on Computational Structures Technology, 12–15 September, Las Palmas de Gran Canaria, Spain.*

- Langelaar, M. and Van Keulen, F. Modeling of shape memory alloy shells for design optimization. *Computers and Structures*, in review.
- Langelaar, M. and Van Keulen, F. Sensitivity analysis of shape memory alloy shells. *Computers and Structures*, in review.
- Langelaar, M. and Van Keulen, F. Sensitivity analysis and optimization of a shape memory alloy gripper. *Computers and Structures*, in review.
- Langelaar, M., Yoon, G.H., Kim, Y.Y. and Van Keulen, F. Topology optimization of shape memory alloy thermal actuators using element connectivity parameterization. *Computer Methods in Applied Mechanics and Engineering*, in review.
- Gurav, S.P., Langelaar, M. and Van Keulen, F. (2006). Cycle-based alternating anti-optimization combined with nested parallel computing: application to shape memory alloy microgripper. *Computers and Structures*, in review.
- Yoon, G.H., Kim, Y.Y., Langelaar, M. and Van Keulen, F. Theoretical aspect of the internal element connectivity parameterization approach for topology optimization. In preparation, to be submitted to *International Journal of Numerical Methods in Engineering*.
- Langelaar, M. and Van Keulen, F. (2007). Gradient-based design optimization of shape memory alloy active catheters. In *Proceedings of the 15th AIAA/ASME/AHS Adaptive Structures Conference, April 23–26, Waikiki, Hawaii, U.S.A.*, accepted.
- Langelaar, M. and Van Keulen, F. (2007). Design optimization of shape memory alloy active structures using the R-phase transformation. In *Proceedings of the 14th International SPIE Symposium on Smart Structures and Materials, 18–22 March, San Diego, CA, U.S.A.*, accepted.

Acknowledgements

First of all, I thank José, for all the support, encouragement and inspiration she gave me over the past few years, and for helping me balance my professional and personal life. I also thank my parents, family and friends for all their interest and support. Furthermore, I would like to thank my supervisor Fred van Keulen, for his commitment, enthusiasm and advice, and for giving me the freedom to develop my own ideas. Also each time when the final date of my contract was approaching, he managed to find a source of funding that allowed me to continue my research, for which I am very grateful. I also thank all my collegues from the Structural Optimization and Computational Mechanics group in Delft: Jan, Hans, Marianne, Javad, Jacqueline, Marten Jan, Koen, Andriy, Gerard, Teun, Sham, Peterjan, Chiara, Bert, Gih-Keong and Caspar. They all together created an enjoyable and stimulating working environment.

A special thanks goes to Sham Gurav, with whom I collaborated on including uncertainties in shape memory alloy design optimization and on the development of parallel computing scripts, which have been of great use for many of the results in this thesis. I also especially would like to mention Jan Booij, for solving numerous computer problems, providing a well-maintained working environment, tolerating my occasional computer pranks and not even complaining when due to my use of an obsolete operating system the entire network of the group was hacked. Furthermore, I thank Niels Bakker and Paul Breedveld from the Biomedical Engineering department, for sharing their expertise on minimally invasive surgery, and Vincent Henneken and Warner Venstra for sharing their insight and experiences regarding microsystem design and development.

Furthermore, I would like to say *감사합니다*[*] to Professor Yoon Young Kim of Seoul National University, for his great hospitality and for creating the ideal conditions for a fruitful collaboration. Also his guidance and advice during the regular progress meetings was invaluable. An equally strong *감사합니다* goes out to Gil Ho Yoon, who taught me all the secrets of topology optimization and element connectivity parameterization. With his knowledge and ideas, he contributed significantly to the realization of the topology optimization procedure discussed in this thesis. Next to this, I also feel indebted to Professor Yoon Young Kim and Gil Ho Yoon for introducing me to a wide range of Korean customs, food and drinks, and in this regard I also should mention Yongkeun Park and Young-Soo Joung for being excellent company.

I furthermore thank Dr. Andrzej Ziolkowski of the Polish Institute of Fundamental Technological Research, for his discussions on modeling of shape memory alloys, for sharing his knowledge on the R-phase transformation with me, and for sending me the relevant literature. I also want to acknowledge Karsten Svanberg for the use of his Method of Moving Asymptotes, Sivan Toledo for the use of his fast direct sparse linear solver library TAUCS, and Leo Breebaart for the use of his LaTeX thesis template. Finally, I also would like to acknowledge the Delft Center of Mechatronics and Microsystems, which has financially supported a considerable part of this research.

Matthijs Langelaar
Delft, October 2006

[*] "Kamsa hamnida", meaning "Thank you", in Korean.

About the author

February 12, 1976	Born in Apeldoorn, the Netherlands.
1988 – 1994	High school, CLA Gymnasium in Apeldoorn. Graduated cum laude.
1994 – 1999	BSc. and MSc. study in Mechanical Engineering at University of Twente, the Netherlands. Graduated cum laude, on a graduation project carried out at the Product Application Center of Corus, IJmuiden, the Netherlands, supervised by Prof.dr.ir. J. Huétink and Dr.ir. H. Vegter. Title of MSc-thesis: *Verification of the Vegter Yield Criterion*. This thesis was awarded the Dutch Central Institute of Industrial Development (Centraal Instituut Voor Industrie Ontwikkeling) prize in November 2000, by the Royal Holland Society of Sciences and Humanities (Koninklijke Hollandsche Maatschappij der Wetenschappen).
2000 – 2001	Researcher at the Control Design Engineering group of the Institute of Robotics and Mechatronics. This institute is a division of DLR, the German Aerospace Center, located in Oberpfaffenhofen, Germany. Participant in European Union funded project "REALSIM", aimed at real-time simulation for the design of multi-physics systems.

2001 – present	Ph.D. candidate at the Structural Optimization and Computational Mechanics group, which is part of the Precision and Microsystem Engineering Department of the Faculty of Mechanical, Materials and Maritime Engineering of Delft University of Technology, Delft, the Netherlands.
November 2004 – January 2005	Visiting scholar at the Integrated Design and Analysis of Structures Laboratory of Prof. Y.Y. Kim, at Seoul National University, Seoul, Korea. Collaborated with Dr. G.H. Yoon on the development of topology optimization of shape memory alloy structures.

From January 2007, Matthijs Langelaar is appointed as a post-doctoral researcher at the Precision and Microsystem Engineering Department, and will work on a project aimed at topology optimization of microcomponents including reliability aspects, which is part of the national MicroNed research program.