Acta Numerica (2012), pp. 307–378
doi:10.1017/S0962492912000049
© Cambridge University Press, 2012
Printed in the United Kingdom

Mathematical cancer therapy planning in deep regional hyperthermia*

Peter Deuflhard[‡†], Anton Schiela[‡] and Martin Weiser[‡]
Zuse Institute Berlin,
Takustrasse 7,
D-14195 Berlin, Germany
E-mail: deuflhard@zib.de

This paper surveys the mathematics required for a typically challenging problem from computational medicine: cancer therapy planning in deep regional hyperthermia. In the course of many years of close cooperation with clinics, the medical problem has given rise to many subtle mathematical problems, some of which were unsolved when the project started. Efficiency of numerical algorithms, *i.e.*, computational speed and monitored reliability, plays a decisive role in the medical treatment. Off-the-shelf software had turned out to be insufficient to meet the requirements of medicine. Instead, new mathematical theory as well as new numerical algorithms had to be developed. In order to make our algorithms useful in the clinical environment, new visualization software, *i.e.*, a 'virtual lab', including three-dimensional geometry processing of individual virtual patients, had to be designed and implemented. Moreover, before the problems could be attacked by numerical algorithms, careful mathematical modelling had to be done. Finally, parameter identification and constrained optimization for the PDEs had to be newly analysed and realized over the individual patient's geometry. Our new techniques had an impact on the specificity of the treatment of individual patients and on the construction of an improved hyperthermia applicator.

* Colour online for monochrome figures available at journals.cambridge.org/anu.

† Institute of Mathematics, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany

‡ Supported by the DFG Research Center MATHEON 'Mathematics for key technologies', Berlin.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

CONTENTS

1 Introduction 308
2 Patient-specific virtual lab 311
3 Mathematical modelling 316
4 Efficient simulation 337
5 PDE-constrained optimization 349
6 Impact on medical technology 366
References 371

1. Introduction

The term *hyperthermia* in cancer medicine means a heat treatment of tumour tissue. In 1961, Scheid had already detected 'functional specialities in the microcirculation of carcinomas' under the influence of heating: see Scheid (1961). In an early attempt, Manfred von Ardenne (1907–1997) had experimented with what is today called *full-body* hyperthermia: he immersed cancer patients in a hot bath or exposed them to infrared radiation. However, his medical results were unclear: in rare cases, tumour growth was reduced, whereas in most cases there was no effect at all. Today such a procedure is only considered for cancer patients with widespread micro-metastases that cannot be detected by medical imaging. Instead, *regional* hyperthermia has evolved as an option. Here, tumours are only locally heated, while healthy tissue is generally saved. With increasing understanding about the generation and function of heat shock proteins, a lower temperature bound of roughly 42.5° appeared to be advisable. An upper temperature bound in the tumour should be 'not too high' to ensure that healthy tissue is kept close enough to the body temperature of 37°, and that the systemic body stress remains tolerable. At present, hyperthermia is applied via two typical modes.

- *Downstaging.* Hyperthermia treatment by itself can cause the tumour to have sharper boundaries, thus facilitating operative tumour extinction by the oncosurgeon – an effect that helps to reduce the probability of relapse.
- *Combination therapy.* In principle, one might think of a strategy where heat alone simply 'burns' the tumour, which would then be secreted by the kidneys as necrotic tissue. However, as long as healthy tissue cannot be guaranteed to be preserved unburnt in such a process, the therapeutic temperature level must remain moderate. In a temperature window between 41°C and 45°C, so-called *deep* regional hyperthermia (as opposed to surface hyperthermia, *e.g.*, for extremities) is usually combined with chemo- and radiotherapy: see, *e.g.*, Issels *et al.* (1987). In fact, testing

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

hyperthermia alone in a research situation, where other therapies such as chemo- or radiotherapy might help to cure the patient, is not ethically justified. This is why the actual cases open to our simulation studies were only deep-seated non-operable tumours.

As early as 1984, it had become clear that careful hyperthermia treatment planning required computer simulations: see, *e.g.*, Roemer and Cetas (1984). In the late 1980s, the physician (and physicist) Wust came to ZIB seeking advice about the fast numerical solution of a PDE called the bio-heat transfer equation, which turned out to be of Helmholtz type in the uncritical case. His interest was in two-dimensional computation, since the input from individual patients comprised two-dimensional cross-sections taken from a stack of images. Initially (see Wust *et al.* 1989, 1991), he was recommended to use the 1985 version of the two-dimensional elliptic solver PLTMG by Bank (1985). However, the two-dimensional calculations appeared to be in strong contradiction to the practical experience of medical radiologists: various patients complained about 'hot spots' quite distant from the tumour being heated, whereas the simulations did not exhibit any such phenomena. It was the mathematicians who suggested that patients should be regarded as three-dimensional! As a consequence, the new version of the three-dimensional elliptic solver Kaskade (Deuflhard, Leinen and Yserentant 1989, Bornemann, Erdmann and Kornhuber 1993, Götschel, Weiser and Schiela 2012) came into play. This started a close collaboration between the numerical analysis group at ZIB and the radiology department at Charité, a collaboration that continued over more than a decade: see, *e.g.*, Wust *et al.* (1999) and Sreenivasa *et al.* (2003). In the course of time, nearly all mathematical and computer science tools necessary to tackle the challenging hyperthermia problems were tested thoroughly. As a result, all pieces were redesigned, until finally the whole computational line appeared to be fast, robust, and reliable enough to be used in a clinical environment. For this purpose, integrated software, the *virtual lab* HyperPlan, was elaborated (with eventually about half a million lines of code). It included a selection of both visualization and numerical algorithms for the construction of individual three-dimensional *virtual patients*.

On the numerical side, it contained adaptive multilevel finite element methods for the parabolic bio-heat transfer (BHT) equation coupled with the time-harmonic heterogeneous Maxwell's equation, the latter in the then still underrated intermediate regime where the wavelength is of the size of the obstacles. For the BHT equation, Yserentant's hierarchical basis preconditioner for two dimensions (Yserentant 1986, Deuflhard *et al.* 1989), as well as Xu's BPX preconditioner for three dimensions (Xu 1989, Bramble, Pasciak and Xu 1990; in the variant due to Yserentant 1993 and Bornemann *et al.* 1993), appeared at just the right time. For the time-harmonic

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Maxwell's equation in the radiowave regime, the key turned out to be the use of (a) Whitney forms (Whitney 1957), the lowest-order three-dimensional finite edge elements (Nédélec 1980, Bossavit 1988, Beck 1993) to avoid spurious discrete solutions, and (b) the discrete Helmholtz decomposition due to Hiptmair (see Hiptmair 1996, Beck and Hiptmair 1999, Beck *et al.* 1999 and Hiptmair 2002*b*) within multigrid methods. In 2000, a first survey on this kind of work was given in Deuflhard (2000). These PDE solvers had been coupled to compute an optimal tuning of the antenna parameters. As a first step, this has been modelled by an unconstrained minimization problem, where the medically necessary heating constraints were taken into account via some penalty term.

Unfortunately, the DFG funding of the ZIB–Charité cooperation (within the collaborative research unit SFB 273) terminated at the end of 2002. Fortunately, the topic of regional hyperthermia could be further pursued within the DFG Research Center Matheon, now on a more mathematical basis. The above unconstrained minimization problem with penalty term was substituted by the medically more reasonable constrained minimization problem, where the heating constraints were modelled as strict inequalities. When this substitution was envisaged, the associated mathematical topic (adaptive multilevel methods for PDE-constrained optimization) was far from mature enough to be used with medical reliability. It is the purpose of the present survey to report on progress made since then in the context of past achievements.

Outline

The paper is organized as follows. In Section 2 we give the basics of the virtual lab HyperPlan. This software environment fixes all patient-specific details necessary to pose the hyperthermia problem: individual geometry, material coefficients, electromagnetic antennas at the applicator, and the water bolus. After that, in Section 3, the medical problem is transformed into a mathematical model in terms of partial differential equations (PDEs), a Helmholtz equation in the uncritical case for the distribution of heat in the body, and the time-harmonic Maxwell equation, which is structurally similar to a Helmholtz equation, in the critical case. The therapeutic goal is transformed into a variety of mathematical optimization problems of increasing difficulty, starting from a traditional functional with penalized temperature constraints and continuing to a medically preferable functional with PDE equality constraints and strict temperature inequality constraints. On this basis, Section 4 discusses efficient numerical techniques to solve the PDEs. Section 5 presents the recent theoretical analysis and algorithmic progress for PDE-constrained optimization. An interior point method is derived and worked out as a numerical path-following method in terms of a barrier parameter. As it turns out, the log-barrier function, well-known from finite-

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

dimensional optimization, needs to be replaced by a hierarchy of rational barrier functions in function space optimization. Finally, in Section 6, the impact of the mathematical developments over the years with respect to clinical application is surveyed.

2. Patient-specific virtual lab

The many years of common work between ZIB and Charité in the hyperthermia project taught much to both sides, and this has paid off in all later collaborations between mathematics and medicine. This is why we start with a description of the basic pattern (paradigm) and patient-specific three-dimensional grid generation.

2.1. Paradigm of virtual medicine

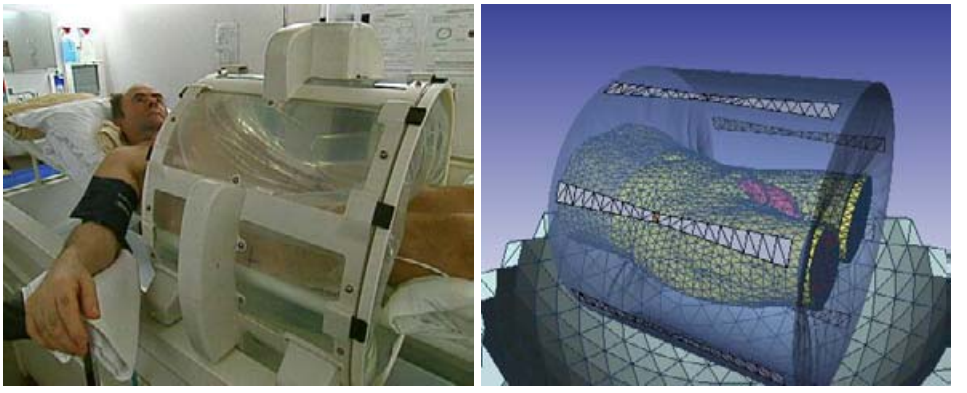
For the cooperation between medicine and mathematics the following three-step procedure has evolved (see Figure 2.1).

(I) Construct an individual *virtual patient*. The construction of a patient-specific geometrical three-dimensional model starts from a stack of two-dimensional images originating from CT, MRT or ultrasound, for example. In a first computational step, three-dimensional details of the patient's body need to be identified, a procedure called *segmentation*. In order to obtain reasonable *starting grids* for a multigrid PDE solution, surface triangulation of interfaces, coarsening of surface grids, and finally tetrahedral grid generation are successively performed.

(II) Model and simulate the treatment in a *virtual lab*. The modelling of the therapy in question consists of a description of the essential physiological and physical processes, as accurate as is needed for the medical problems to be tackled. Usually, the description consists of a set of partial differential equations (PDEs) and, in the best case, a functional to be minimized under constraints. In order to suggest an 'optimal' treatment planning, these PDEs need to be solved quickly and reliably. Moreover, in order to be of use as a basis for medical decision, any suggested therapy requires careful visualization. All this must be realized on a hardware and software platform that can be made available in a clinical environment.

(III) Transfer the simulation results *back to the clinical environment*. In some prephase, the model and its simulation will be validated against the clinical situation. Any discrepancies between model and reality need to be discussed and, if necessary, taken into account by remodelling. Once this adaptation has been done, the computationally suggested 'optimal' therapy is transferred back to patient-specific treatment in the clinical environment.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 2.1. Hyperthermia treatment planning. (a) Real patient in hospital.[1] (b) Virtual patient in the virtual lab.

Step (I) above is essentially independent of the particular medical problem, but only dependent on the body region treated: see Section 2.2 below. Step (II) depends on the medical problem in question, which in this case is deep regional hyperthermia: see Sections 3, 4.1 and 5. Step (III) will be illustrated by many examples of virtual patients inserted in the text throughout the paper.

2.2. Construction of geometrical three-dimensional model

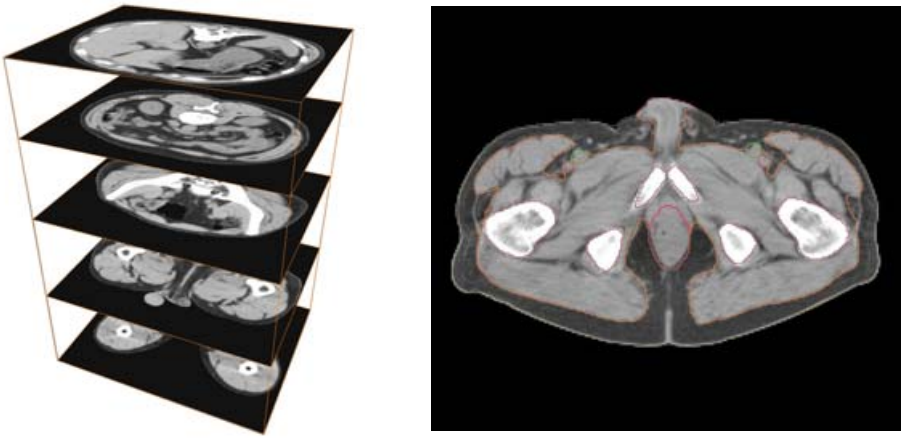
The above paradigm has been implemented in the virtual lab HyperPlan (Stalling *et al.* 1996, Deuflhard 2000), an integrated software system that has been especially developed to support Hyperthermia treatment Planning in a clinical environment. The implementation of roughly half a million lines of code made heavy use of C++ software tools such as OpenGL (graphic library) and OpenInventor (object-oriented class library) which are themselves of high algorithmic complexity.

HyperPlan combines the following features.

- Efficient three-dimensional grid generation techniques, which we will describe immediately below.
- Fast and reliable simulation techniques, *e.g.*, adaptive multilevel finite element methods: see Section 4.1 below.
- Efficient visualization techniques, which are beyond the scope of the present article (for reference see, *e.g.*, Hege and Stalling 1998 and Stalling 1998).

[1] The patient agreed to be shown recognizably within a scientific context.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 2.2. Medical imaging input. (a) Stack of two-dimensional computer tomograms (CTs). (b) Segmented raw image data.

Medical imaging input

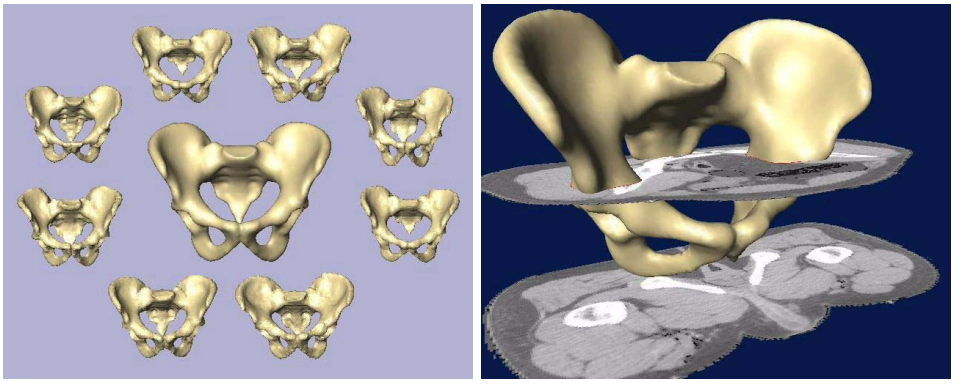
Patient-specific geometrical input usually comes from medical imaging techniques such as CT, MRT, ultrasound or mixtures thereof. We skip the case where different sources of input have to be matched (the so-called registration problem: see, *e.g.*, Zitov and Flusser 2003). As shown for CT in Figure 2.2, this input comes as a stack of two-dimensional images that merely give mass density information.[2] In view of the solution of PDEs on the specific patient data, we need more than just density information: see below under segmentation. Moreover, for the purpose of presenting numerical results to medical doctors, a three-dimensional grid generation with clear material identification needs to be realized. This is done in several steps, which are described next.

Segmentation

As can be seen from Figure 2.2, the original CT input contains only information about the mass density distribution in planar cross-sections of the body. What we need, however, to solve PDEs on an individual patient, is a clear correspondence of the various body parts with material properties such as electromagnetic or thermal constants. This task is called segmentation. From computer science, several segmentation techniques such as region growing (Adams and Bischof 1994) or watershed methods (Vincent and Soille 1991) are quite popular. They typically work on the stack of two-dimensional cross-sections and, in the best case, exploit some connection between neighbouring two-dimensional images. Such techniques

[2] More precisely, it gives information about the attentuation factor in tissue.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 2.3. Segmentation of a pelvic region by statistical shape analysis. (a) Representation of known patient data by their average and eight individual data (construction of essential coordinate frame). (b) Simultaneous adaptation of a shape model for the full stack of two-dimensional images of new patient data (representation in the frame of essential coordinates). From Lamecker, Seebass, Hege and Deuflhard (2004).

usually require a significant amount of user interaction, which is not desirable. Therefore, at ZIB, we followed a rather different approach called *statistical shape analysis* (Cootes, Hill, Taylor and Haslam 1994), which moves through three dimensions. In principle, this method performs the following steps.

(1) Start from a sample set of carefully segmented patient data, a time-consuming interactive job done by experts. For illustration, see the eight pelvic regions arranged in a circle in Figure 2.3(a).

(2) Compute an 'average' shape. See the centre shape in Figure 2.3(a). In order to compute a theoretically satisfactory average, one must solve the associated correspondence problem between two surfaces (see, *e.g.*, Lamecker 2008) wherein the way is paved towards a 'variational shape analysis'.

(3) Compute the differences 'individual–average' as fluctuations.

(4) Select 'essential coordinates' by selecting the fluctuations with the help of PCA (principal components analysis).

(5) Whenever new patient data are to be segmented, try to simply compute the essential coordinates for the new data: see Figure 2.3(b). If this works to sufficient accuracy, then exit. Otherwise enrich the patient database by these new data and go to step (1).

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

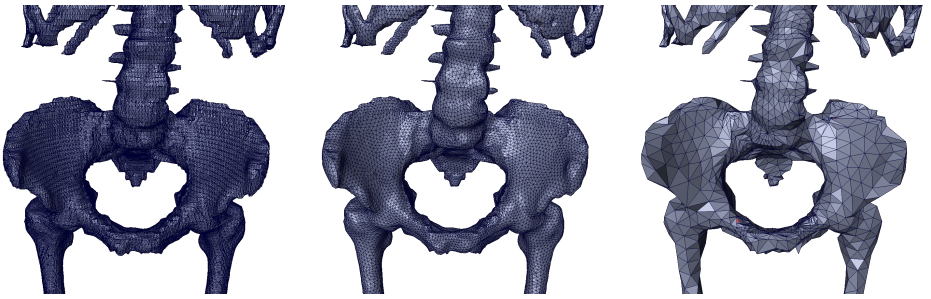


Figure 2.4. Grid coarsening in the pelvic region. From 478 000 nodes via 98 000 nodes to 11 000 nodes.

At the beginning of the above procedure, the number of essential coordinates will grow with all new patient data, but after a while it will get saturated, thus defining the 'essential dimension' of the patient data set at hand. Of course, in order to keep this dimension low, patient data should be clustered according to criteria such as sex, age, sports activities, biological group, *etc.* Figure 2.3 illustrates such a three-dimensional segmentation for some pelvic region. All in all, the techniques summarized here have reached a state of maturity that makes them equivalent in quality to a 'second human radiologist': see the award-winning paper by Kainmueller, Lange and Lamecker (2007), from an international contest on liver segmentation.

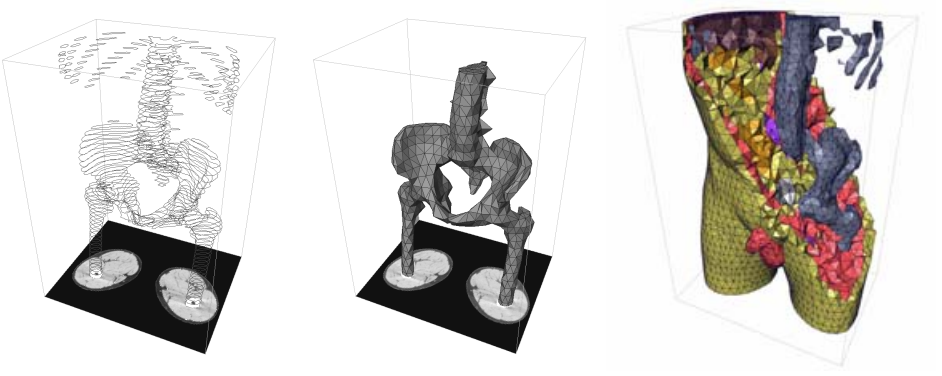
Surface triangulation coarsening

Typically, meshes generated after segmentation are much too fine: for an illustration see Figure 2.4(a). For the purpose of visualization, this large number can be handled, but in view of the numerical solution of PDEs on the virtual patients (*e.g.*, by multigrid methods), this number is far too large. For this reason, some precoarsening of the grids is needed. The associated iterative process is based on minimizing local approximation errors of coarser triangulated surfaces, which is closely related to local curvature: wherever the local curvature is 'low', the number of local nodes is reduced, whereas it is left unaltered where the local curvature is 'high' (see, *e.g.*, Heckbert and Garland (1997)). In Figure 2.4, the coarsening procedure (including mesh quality control) due to Zilske, Lamecker and Zachow (2008) is illustrated for a pelvic bone region. For multigrid methods to be discussed below in Section 4.1, this kind of grid then serves as the initial grid.

Tetrahedral grid generation

As a final step, the interface triangular grids are used as a starting point to construct spatial tetrahedral grids in the full volume. This is done by a special adaptation of the 'advancing front' algorithm of Löhner and Parikh (1988). A more recent improvement is the remeshing technique due to Zachow, Zilske and Hege (2007). This final step is illustrated in Figure 2.5(c).

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b) (c)

Figure 2.5. Stepwise construction of adaptive three-dimensional grid from a stack of cross-sectional images as shown in Figure 2.2. (a) Segmented surface and interface lines between different material regions. (b) Coarsened triangulation of surface and interfaces, as illustrated in Figure 2.4. (c) Generation of tetrahedral volume grid.

Remark 2.1. HyperPlan was the predecessor of the popular system Amira (Stalling, Westerhoff and Hege 2005), which applies to general systems in science and medicine.

3. Mathematical modelling

The purpose of this section is to describe the mathematical modelling necessary to translate the medical treatment planning problem into a mathematical formulation accessible to numerical methods. It covers the physics of heat generation by radio-wave emission in the antennas of the applicator, which is modelled by the time-harmonic Maxwell equation, and the transport of heat inside the human body, which is modelled by some bioheat transfer equation. Moreover, it includes the identification of unknown parameters in the mathematical models and several choices of quantitative formulations of the therapeutic goal, which leads to optimal control problems that are dealt with in Section 5.

Hyperthermia treatment of a real hospitalized patient is depicted in Figure 2.1(a). As the patient has quite a large rectal carcinoma, his abdomen is lying inside a so-called applicator, which contains eight antennas (pairwise coupled). The antennas emit radiowaves of frequency 100 MHz, which are absorbed in the human body, thus generating heat. The associated mathematical model, the time-harmonic Maxwell equation, is discussed in Section 3.1. Between the applicator and the body is a so-called water bolus to ease the entry of the radiowaves into the body and, at the same time,

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

to cool the skin – thus setting a temperature value at the body surface. The absorption of the electromagnetic radiation is different for different body compartments (essentially depending on the portion of water in the tissue). By virtue of blood flow as well as heat conduction, the heat is distributed in the body. This process is modelled mathematically by a parabolic PDE, the so-called bio-heat transfer equation: see Section 3.2. A solution of the coupled system of PDEs is a necessary prerequisite to compute an optimal tuning of the antenna parameters such that the tumour is heated (within a temperature window), but not healthy tissue. This leads to a formulation in terms of mathematical optimization problems, to be elaborated in Section 3.4.

3.1. Emission of radiowaves

From first principles, the radiowaves emitted by the applicator antennas are known to be governed by Maxwell's equations. Simplified models are available for the case when the wavelength is much larger or much smaller than the geometry of the objects. In the hyperthermia case, however, the wavelength in water (which is a good approximation for the material of the body) is about 30 cm, which is the order of magnitude of the human body (the 'obstacle'), *i.e.*, just in the intermediate regime. As a consequence, there is no way to avoid solving the full Maxwell equations for the heterogeneous medium 'water bolus–human body'.

Time-harmonic Maxwell equation: strong formulation

Fortunately, only a single angular frequency ω is emitted from the antennas, which means that we can make the time-harmonic ansatz

$$E(x,t) = \mathrm{Re}\,(E(x)\mathrm{e}^{-\mathrm{i}\omega t}), H(x,t) = \mathrm{Re}\,(H(x)\mathrm{e}^{-\mathrm{i}\omega t}),$$

for the electromagnetic fields $E, H \in \mathbb{C}^3$. Insertion into the time-dependent Maxwell equations for linear isotropic dielectric media,

$$\mathrm{div}(\mu H) = 0, \tag{3.1a}$$

$$\mathrm{div}(\epsilon E) = 0, \tag{3.1b}$$

$$\mathrm{curl}\, H = \sigma E + \epsilon \frac{\partial E}{\partial t}, \tag{3.1c}$$

$$\mathrm{curl}\, E = -\mu \frac{\partial H}{\partial t}, \tag{3.1d}$$

applying the curl operator to (3.1d), and inserting (3.1c) yields the so-called *double-curl equation* or *time-harmonic Maxwell equation*

$$\mathrm{curl}\, \frac{1}{\mu}\, \mathrm{curl}\, E = \epsilon' \omega^2 E \tag{3.2}$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

for the complex electrical field E. Here μ is the magnetic permeability and $\epsilon' = \epsilon + \mathrm{i}\sigma/\omega$ the complex dielectric constant related to Ohm's law $j = \sigma E$, with electric conductivity $\sigma \geq 0$ for any current density j and the generic dielectric constant $\epsilon > 0$. The material constants μ, ϵ, σ depend, as the name indicates, on the local materials (*e.g.*, muscle, fat, bone, *etc.*) and should therefore, more precisely, be written as $\mu(x), \epsilon(x), \sigma(x)$. They are allocated on the basis of the geometrical patient model: see Section 2.2 above.

For spatially constant material parameters (homogeneous material), the term on the left-hand side may be reformulated to lead to

$$-\Delta E + \nabla \underbrace{\operatorname{div} E}_{=0} = \mu\epsilon'\omega^2 E. \tag{3.3}$$

Here the Laplace operator Δ is applied componentwise, such that (3.3) represents three decoupled Helmholtz-type equations in the critical case. Otherwise, with material jumps, the three PDEs are coupled: see Section 4.1.

Boundary conditions

The computational domain of interest (see Figure 2.1(b)) consists of an artificially chosen sphere coating the water bolus, the antenna array, and the virtual patient model. Hence we have three types of boundary conditions to be imposed. First, the metallic antennas on the physical boundary of the applicator are regarded as perfect conductors, which implies *homogeneous* Dirichlet boundary conditions for the tangential components $E_t = (n \times E) \times n$ of the electric field E, *i.e.*,

$$E_t = 0 \quad \Leftrightarrow \quad n \times E = 0 \quad \text{on} \quad \Gamma_{\text{ant}}.$$

Here Γ_{ant} denotes the thin metallic antennas shown in Figure 3.1 and n denotes the outer unit normal of the surface Γ_{ant}. Second, the small gap between the trapezoidal antenna sheets is modelled by an artificial boundary Γ_{gap} on which the tangential component of E is prescribed, so as to yield the voltage of the power generator when integrated across the gap:

$$E_t = E_t^0 \quad \text{on} \quad \Gamma_{\text{gap}}.$$

Third, on the external sphere Γ_{ext}, the Sommerfeld radiation condition (also called the Silver–Müller condition in the electrical engineering community)

$$n \times \left(\frac{1}{\mu} \operatorname{curl} E\right) - n \times \left(\frac{\mathrm{i}\omega}{\mu c} n \times E\right) = 0 \quad \text{on} \quad \Gamma_{\text{ext}} \tag{3.4}$$

is prescribed as an *approximate transparent* boundary condition in order to avoid reflection at the artificial boundary Γ_{ext}. Here $c = 1/\sqrt{\epsilon\mu}$, the electromagnetic wave velocity in the dielectric material, has been inserted.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

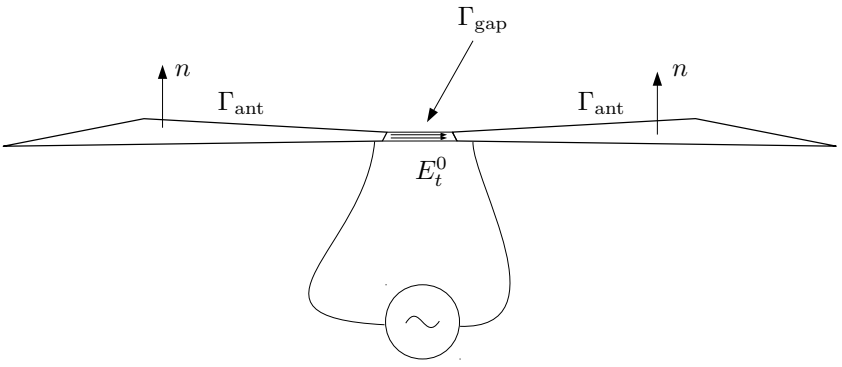


Figure 3.1. Antenna modelling with Dirichlet boundary conditions.

Time-harmonic Maxwell equation: weak formulation
The weak formulation starts from the double-curl equation (3.2) directly. For *lossy* materials (*i.e.*, for $\sigma > 0$), the material constant ϵ' is complex. As a consequence, the above PDE with boundary conditions represents a *non-self-adjoint* operator. Then, following Beck and Hiptmair (1999), we introduce the appropriate Hilbert space for the operator equation as

$$H_\Gamma(\text{curl};\Omega) := \{v \in L_2(\Omega)^3; \text{curl}\, v \in L_2(\Omega)^3,\ n \times v = 0 \text{ on } \Gamma_{\text{ant}} \cup \Gamma_{\text{gap}},\\ n \times v \in L_2(\Gamma_{\text{ext}})^3 \text{ on } \Gamma_{\text{ext}}\},$$

equipped with the scalar product

$$\langle v, w\rangle_{H(\text{curl})} := \langle \text{curl}\, v, \text{curl}\, w\rangle_{L_2(\Omega)} + \langle v, w\rangle_{L_2(\Omega\cup\Gamma\text{ext})},$$

and a sesquilinear form $a : H_\Gamma(\text{curl};\Omega) \times H_\Gamma(\text{curl};\Omega) \to \mathbb{C}$ (conjugate linear in the first component):

$$a(E,v) := \int_\Omega \left(\frac{1}{\mu}(\text{curl}\, E)^* \,\text{curl}\, v - \omega^2(\epsilon' E)^* v\right) \mathrm{d}x \\ - \int_{\Gamma_{\text{ext}}} \frac{\mathrm{i}\omega}{\mu c}(n \times E)^*(n \times v)\, \mathrm{d}s. \tag{3.5}$$

We end up with the following variational problem: find $E \in H_\Gamma(\text{curl};\Omega)$ such that

$$a(E,v) = f(v) \quad \forall\, v \in H_\Gamma(\text{curl};\Omega), \tag{3.6}$$

where $f : H_\Gamma(\text{curl};\Omega) \to \mathbb{C}$ is a linear form containing the inhomogeneity of the Dirichlet conditions.

Properties of (3.6)
In order to study the question of *uniqueness* of a solution, we follow the usual procedure: see Monk (2003, Theorem 4.12). Suppose there exists some $\bar{E}$ such that $a(\bar{E},\bar{E}) = 0$, which implies that both its real and its

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

imaginary part vanish. We assume homogeneous, possibly lossy material, *i.e.*, $\sigma \geq 0$. For the imaginary part we get

$$-\mathrm{Im}\, a(\bar{E}, \bar{E}) = \omega\left(\sigma \int_\Omega |\bar{E}|^2_{\mathbb{C}^3}\, \mathrm{d}x + \frac{1}{\mu c}\int_{\Gamma_{\mathrm{ext}}} |n \times \bar{E}|^2_{\mathbb{C}^3}\, \mathrm{d}s\right).$$

Both integrals are positive. For lossless material ($\sigma = 0$), the first term vanishes. The second one is positive, unless the external boundary integral vanishes, which would be the case for homogeneous Dirichlet or Neumann conditions. For the radiation condition (3.4), however, this is not the case. For lossy material ($\sigma > 0$), both terms are positive, which leads to a contradiction to the assumption that they should vanish. If a solution of (3.6) exists at all, it is unique whenever $\sigma \geq 0$.

Two further features of (3.5) deserve special attention: (a) the first term in the volume integral, representing the principal part of the PDE, has an 'ample' nullspace

$$H^0_\Gamma(\mathrm{curl}; \Omega) = \ker \mathrm{curl}$$

containing all gradient functions, in particular, functions of high frequency, (b) in this nullspace, the second term in the volume integral, representing the zero-order term in the PDE, is negative definite, since the factor $-\omega^2 \mathrm{Re}\, \epsilon' = -\omega^2 \epsilon$ is negative. Thus, the functional a is highly *indefinite*. The indefiniteness is the more marked the higher the angular frequency ω, a fact clearly given in hyperthermia (with 100 MHz).

On simply connected Lipschitz domains $H^0_\Gamma(\mathrm{curl}; \Omega)$ gives rise to a regular decomposition of $H_\Gamma(\mathrm{curl}; \Omega)$ (Hiptmair 2002*a*):

$$v = \phi + \mathrm{grad}\, p, \quad \phi \in H^1(\Omega)^3,\ p \in H^1(\Omega),\ \mathrm{grad}\, p \in H^0_\Gamma(\mathrm{curl}; \Omega), \tag{3.7}$$

such that

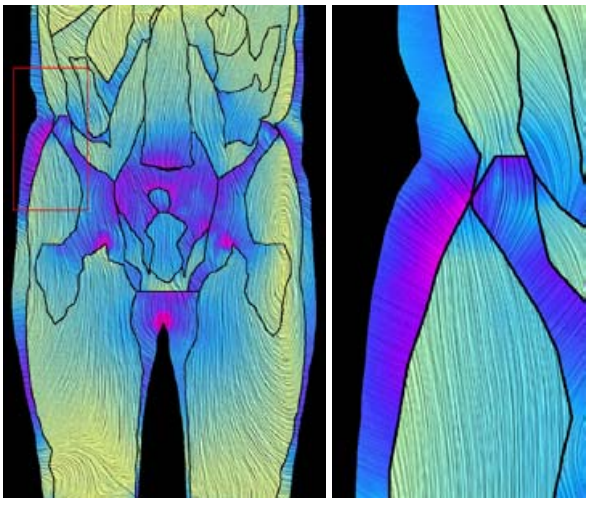
$$\|\phi\|_{H^1} + \|p\|_{H^1} \leq C \|v\|_{H(\mathrm{curl})}. \tag{3.8}$$

In this way, our problem is split into a vectorial H^1-part and a negative definite H^1-component for the potentials. This decomposition is a modification of the classical *Helmholtz* decomposition.

Interface conditions at material jumps

Both quantities μ and ϵ have jumps at material interfaces that have been determined by segmentation: see Section 2.2 above. At these interfaces special conditions must be observed, which in weak formulation are automatically satisfied (via the incorporated vanishing of the remainder terms stemming from an integration by parts). These conditions imply a kink in the electric field vector E at the boundaries; for an illustration in FastLIC representation (Hege and Stalling 1998) see Figure 3.2.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 3.2. FastLIC representation of electrical field lines (Hege and Stalling 1998). (a) Abdominal body part. (b) Zoom.

Superposition

The radiowaves (100 MHz) are emitted from k_E channels of pairwise coupled $2k_E$ antennas. Since the Maxwell equations are linear, superposition can be applied. Let $E_k(x)$ denote the electric field (the antenna profile) originating from channel k for $k = 1, \ldots, k_E$. Then the total electrical field can be written as a linear combination

$$E(x) = \sum_{k=1}^{k_E} \alpha_k E_k(x), \quad \alpha_k \in \mathbb{C}, \tag{3.9}$$

in terms of complex amplitudes $\alpha_1, \ldots \alpha_{k_E}$. They may be parametrized via

$$\alpha_k = a_k \ \exp(-\mathrm{i}\theta_k) \quad k = 1, \ldots, k_E, \tag{3.10}$$

with real amplitudes a_k and phase delays θ_k that can be controlled independently. Where appropriate, the k_E complex parameters α_k will be interpreted as $2k_E$ real parameters $\{\mathrm{Re}\,\alpha_k, \mathrm{Im}\,\alpha_k\}$.

Remark 3.1. Kremer and Louis (1990) treated the problem of approximating a given electrical field (obtained, *e.g.*, from optimization) via excitation through n antennas. For their theoretical analysis, they define a hyperthermia operator that maps arbitrary antenna excitations at a boundary circle of a cylinder to any electrical fields in that cylinder. This operator is shown to be compact in L_2. By means of singular value decomposition they can conclude that hot spots outside the tumour volume can only be suppressed for sufficiently large n.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

3.2. Heat distribution inside the human body

Compared to the *generation* of heat as discussed in the preceding section, the *distribution* of heat inside a specific patient's body is much harder in terms of modelling.

Heat absorption
Inside the body, the electromagnetic waves generate heat by absorption to be characterized via the *absorption rate density*

$$S = \frac{1}{2}\,\sigma |E|^2_{\mathbb{C}^3}, \tag{3.11}$$

where σ is the electric conductivity of the specific tissue. If one exploits the superposition (3.9), one ends up with the quadratic term

$$S(x) = \frac{1}{2}\sigma(x)|E(x)|^2_{\mathbb{C}^3} = \frac{1}{2}\sigma(x) \sum_{j,l=1}^{k_E} \alpha_j^* \alpha_l \, E_j^*(x)\, E_l(x), \tag{3.12}$$

where σ varies among different tissues. From the definition (3.12) it is apparent that simultaneous phase shifts of the controls α_k lead to a corresponding phase shift of the electrical field E and thus have no thermal impact at all. Therefore, one of the phases can be fixed *a priori*, *e.g.*, by setting $\mathrm{Im}\,(\alpha_{k_E}) = 0$. We will therefore think of α containing only $2k_E - 1$ real parameters, which will be important in Section 5.1 to get rid of a trivial local non-uniqueness of optimal solutions.

Bio-heat-transfer (BHT) equation: strong formulation
The standard heat transfer model used in our simulations has been the so-called bio-heat-transfer (BHT) equation over the domain $\Omega \subset \mathbb{R}^3$ covered by the patient's body:

$$\rho c \frac{\partial T}{\partial t} = \mathrm{div}\,(\kappa \,\mathrm{grad}\, T) - c_b w\,(T - T_{\mathrm{blood}}) + S \quad \text{in } \Omega, \tag{3.13}$$

where ρ is the density, c the specific heat capacity of tissue, κ the thermal conductivity, c_b the specific heat capacity of blood, w the mass flow rate of blood per unit volume of tissue, the so-called *perfusion*, T_{blood} the blood temperature in the unheated body (*e.g.*, 37°C), and S the absorption rate density as defined in (3.12). All material constants are understood to be local constants, different in different tissues. Already in 1948, this Helmholtz-type equation in the uncritical case was proposed by Pennes (1948) on the basis of theoretical consideration *and* experimental measurements – a rare scientific combination!

The boundary conditions are of Robin type:

$$-\kappa \frac{\partial T}{\partial n} = h(T - T_{\mathrm{bolus}}) \quad \text{on } \partial\Omega. \tag{3.14}$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

For the time-dependent case initial conditions

$$T(0,x) = T_a$$

are prescribed. In hyperthermia, the steady state is typically reached after about 15–20 minutes and maintained for 40–60 minutes. For this reason, we may well be content with solving the stationary elliptic problem rather than the time-dependent parabolic problem.

As with the other material coefficients, the discontinuity of the thermal conductivity here also yields interface conditions across the tissue interfaces. Again, these conditions are automatically satisfied if the PDE is solved in its weak form.

BHT equation: weak formulation

Here we interpret the stationary BHT equation (with notation as in (3.13)) as an operator equation:

$$\langle A(T) - B(\alpha), \varphi \rangle = 0 \quad \forall \varphi \in C^\infty(\Omega), \tag{3.15}$$

where A and B are defined by

$$\begin{aligned} \langle A(T), \varphi \rangle &:= \int_\Omega \kappa \, (\operatorname{grad} T)^{\mathrm{T}} \operatorname{grad} \varphi + c_b w \, (T - T_{\mathrm{blood}}) \varphi \, \mathrm{d}x \\ &\quad + \int_{\partial\Omega} h (T - T_{\mathrm{bolus}}) \varphi \, \mathrm{d}S, \\ \langle B(\alpha), \varphi \rangle &:= \int_\Omega \frac{\sigma}{2} |E(\alpha)|^2_{\mathbb{C}^3} \varphi \, \mathrm{d}x. \end{aligned} \tag{3.16}$$

The domain Ω consists of a number of subdomains Ω_i, corresponding to various types of tissue. All coefficients depend on the tissue type and are chosen piecewise constant on each subdomain associated with a tissue type. They may vary significantly from tissue to tissue. For a more detailed description of the parameters we refer to Deuflhard, Weiser and Seebass (2000).

Multiscale heat transport

Inside the patient's body, heat is distributed only by diffusion in tissue and by direct transport in blood vessels (with velocity $v(x)$, say), but not by any process described by the zero-order term in the BHT equation (3.13) above. Therefore, from first principles, a PDE of *diffusion–advection* type including the source term seems to be more appropriate:

$$\rho c \, \frac{\partial T}{\partial t} = \operatorname{div} (\kappa \operatorname{grad} T) - \rho c \, v^{\mathrm{T}} \operatorname{grad} T + S, \tag{3.17}$$

However, due to the multiscale structure of the velocity field v, both from the modelling aspect and from the point of view of computational complexity,

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

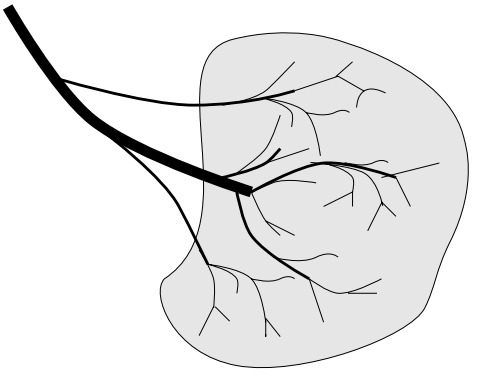


Figure 3.3. Three-scale pattern of heat distribution in the human body: large blood vessels (macroscale), medium-size blood vessels (mesoscale), capillaries (microscale). Homogenization of capillaries leads to a porous medium macroscale model.

such a PDE cannot be used to successfully describe the situation within an individual patient.

The multiscale vascular situation is schematically represented in Figure 3.3. Three scales can be distinguished.

(1) On the macroscale, large blood vessels should be treated as full three-dimensional objects (*e.g.*, as objects in a finite element setting; see also Lagendijk, Schellekens, Schipper and van der Linden 1984).

(2) On some mesoscales, smaller vessels could be modelled as one-dimensional objects (but with a finite width; see D'Angelo and Quarteroni 2007, Quarteroni, Tuveri and Veneziani 2000 and Massing 2007).

(3) On the microscale, detailed phenomena within the capillaries cannot be resolved, neither from modelling nor computational perspectives. That is why some kind of averaging has to take place, which results in a macroscale model for a porous medium region.

Wust *et al.* (1989) had suggested the above BHT equation as the averaged microscale model. It had been derived by a potential flow assumption in view of Darcy's law – known to arise from homogenization techniques in reservoir simulation. That assumption was certainly not self-explanatory in the present context. Therefore we turn to the mathematical *homogenization* approach (see, *e.g.*, Bensoussan, Lions and Papanicolau 1978), which was worked out by Hochmuth and Deuflhard (Deuflhard and Hochmuth 2004, Hochmuth and Deuflhard 2004) for the case of hyperthermia.

In the homogenization framework, we consider a periodic arrangement of microscopic cubic cells εY, periodic along the three spatial coordinate axes y_1, y_2, y_3, with insertions of blood εQ, the capillaries, assumed to have constant temperature T_{blood}: see Figure 3.4. Let Ω_ε denote the solid tissue

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

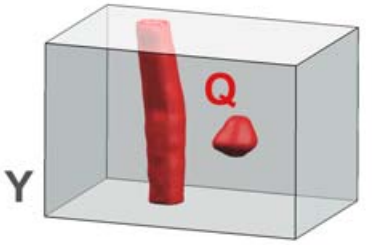


Figure 3.4. Normalized cells Y with inserted blood regions Q: non-isolated as well as isolated case. Surface of insertions $\partial Q \setminus \partial Y$.

volume without the capillaries. Therein heat transfer takes place only by diffusion. For simplicity, Deuflhard and Hochmuth (2004) scaled the thermal conductivity to $\kappa = 1$. As a consequence, the temperature T_ε is defined by the simple model equation

$$-\Delta T_\varepsilon = S_\varepsilon \quad \text{in } \Omega_\varepsilon,$$

with thermal source term $S_\varepsilon = \frac{1}{2}\sigma_\varepsilon |E_\varepsilon|^2$, exterior Dirichlet boundary conditions

$$T_\varepsilon = T_{\text{bas}} \quad \text{on } \partial\Omega_\varepsilon,$$

in terms of some basic temperature T_{bas}, and interior Robin boundary conditions

$$\frac{\partial T_\varepsilon}{\partial n} = \varepsilon\alpha(T^\varepsilon_{\text{blood}} - T_\varepsilon) \quad \text{on } \partial(\varepsilon Q).$$

Due to Deuflhard and Hochmuth (2004), the following asymptotic result can be shown to hold.

Theorem 3.2. Notation as just introduced. Let $|\cdot|$ denote the Lebesgue measure and define relative surfaces and volumes with and without blood by

$$\theta_{\text{surface}} := \frac{|\partial Q \setminus \partial Y|}{|Y \setminus Q|}, \quad \theta_{\text{volume}} := \frac{|Y \setminus Q|}{|Y|} < 1.$$

Assume that, in weak*-convergence, we have

$$S_\epsilon \rightharpoonup \theta_{\text{volume}} S, \quad T^\epsilon_{\text{blood}} \rightharpoonup \theta_{\text{volume}} T_{\text{blood}}.$$

Then, for $\varepsilon \to 0$, the temperature converges weakly as

$$T_\epsilon - T^\varepsilon_{\text{blood}} \rightharpoonup \theta_{\text{volume}}(T_0 - T_{\text{blood}}) \quad \text{in } L_2,$$

where the limit temperature T_0 satisfies the PDE

$$-\operatorname{div}(\mathcal{A} \operatorname{grad} T_0) + \alpha\theta_{\text{surface}}(T_0 - T_{\text{blood}}) = S \quad \text{in } \Omega$$

with Dirichlet boundary conditions

$$T_0 = T_{\text{bas}} \quad \text{on } \partial\Omega$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

and perturbed conductivity

$$\mathcal{A} = a_{ij}, \quad a_{ij} = \delta_{ij} - \frac{1}{|Y \setminus Q|} \int_{Y \setminus Q} \frac{\partial \chi^j}{\partial y_i} \, \mathrm{d}y,$$

where the functions $\chi^j,\ j = 1, 2, 3$ are understood to be solutions of

$$-\Delta \chi^j = 0 \quad \text{in } Y \setminus Q, \quad \frac{\partial (\chi^j - y_j)}{\partial n}, \quad \chi^j \quad Y\text{-periodic.} \tag{3.18}$$

The proof is a bit technical, mainly because of the occurrence of Robin boundary conditions (for $T^{\varepsilon}_{\text{blood}} = 0$ a proof was given earlier; see, *e.g.*, the book by Allaire, Damlamian and Hornung 1995, and references therein). It starts from the above diffusion–advection equation (3.17), assumed locally in a microscale periodic cell structure which is spatially averaged. Details are left to Deuflhard and Hochmuth (2004). Summarizing, the above theorem confirms the validity of the BHT equation in the microvascular regime – apart from some perturbation factors $\theta_{\text{surface}}, \theta_{\text{volume}} \approx 1$ that take care of the relative blood volume versus the full volume.

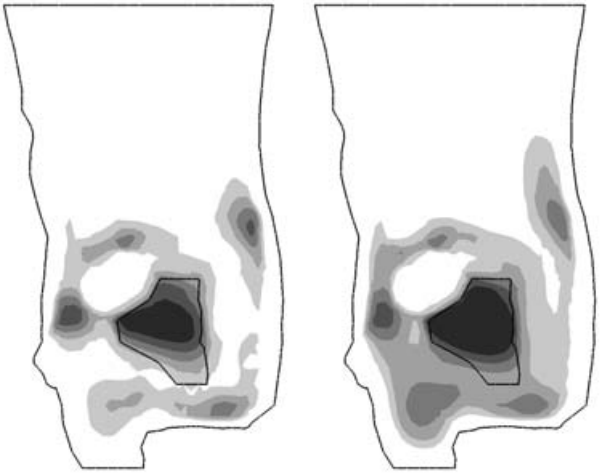
Remark 3.3. An approach aiming at directed advective heat transport in tumours has been developed by Gerber (2007), but not validated.

Temperature-dependent perfusion

In a more detailed investigation, temperature-induced vasodilation leads to an increased perfusion of heated tissue, and thus gives rise to a more realistic, *nonlinear* version of the BHT equation. Experiments (Song *et al.* 1984) have shown that the blood flow in normal tissues, *e.g.*, skin and muscle, increases significantly when heated up to 41–43°C, whereas in the tumour zone the blood flow may increase or decrease with temperature depending on the type of tumour. On this experimental basis, the perfusion $w = w(x, T)$ can be chosen monotonically increasing with respect to T in the muscle and fatty tissue, but monotonically decreasing in tumour tissue.

Simulations show significant qualitative differences between the temperature distributions predicted by the linear and nonlinear heat transfer model, as illustrated in Figure 3.5. Generally speaking, self-regulation of healthy tissue reflected by the nonlinear model reduces 'hot spots' caused by local maxima of the absorbed electromagnetic fields and hence allows the total energy deposition to rise. This is one reason for slightly better tumour heating (*c.* 0.5°C) predicted by the nonlinear model. An analogous result is reported in Tompkins *et al.* (1994*b*) for ferromagnetic thermoseed hyperthermia. An important finding is that the nonlinear model has an impact on the optimal treatment parameters as well. Maximal discrepancies turned out to be 22° for the phases θ_k and 0.22 for the relative amplitudes a_k; for notation see (3.10). For a more detailed discussion see Deuflhard *et al.* (2000) and Lang, Erdmann and Seebass (1999).

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 3.5. Optimized temperature distributions in a frontal section of the pelvic region. *Black lines*, body outline and tumour contour; *light grey to dark grey shading*, regions heated from 39°C to 43°C. (a) Linear model. (b) Nonlinear model.

Besides local vasodilation, systemic thermoregulation can play an important role, *e.g.*, via centrally controlled vasodilation of the skin and a corresponding shift of perfusion from the body core to the extremities, or an increased heart rate. A feedback mechanism observed rather recently is the so-called *steal effect*, where different regions compete for a joint blood supply. An increased perfusion in one region, *e.g.*, a heated muscle that actively reduces its peripheral resistance by vasodilation, will lead to a larger blood flow and hence a greater decrease of effective blood pressure along the vessels feeding both regions. In this case, another region with less capability for thermoregulation, *e.g.*, a tumour, will experience reduced perfusion and a smaller cooling effect. Such a non-local interaction may open up new possibilities for therapy optimization. It might lead to higher tumour temperatures if, instead of focusing the power on the tumour, one were to heat up muscles supplied by the same vessels in order to cut off the tumour from cooling perfusion.

3.3. Parameter identification

All of the above thermoregulation models contain a number of parameters, most of which are unknown. Such parameters must be determined either from direct measurements or from computations comparing indirect measurements with the applied model. The most relevant parameters are as follows.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



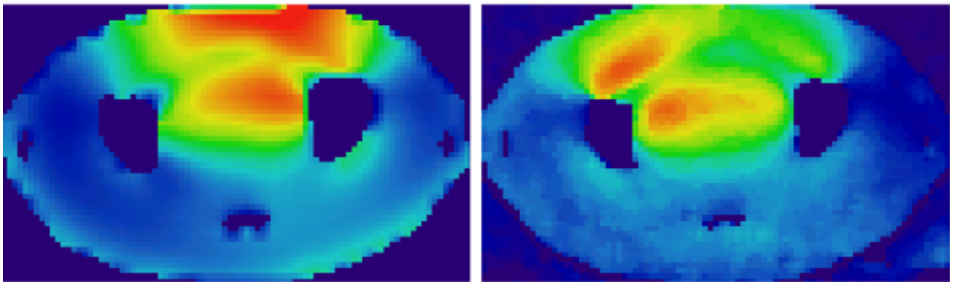
Figure 3.6. Hyperthermia applicator system in open MRT.

- *Material coefficients.* Thermal, electrical, and physiological properties of tissue depend on the specific patient under consideration. Several techniques are available to get hold of reasonable values. Parts of these data are obtained from measurements: see, *e.g.*, Wust *et al.* (1991, Table 1) for electromagnetic and bioheat data, or Seebass *et al.* (2000) with values from Paulsen, Geimer, Tang and Boyse (1999), where data for vessels, lungs, spleen were taken from Gabriel, Lau and Gabriel (1996). In Wust *et al.* (1999, Figure 1), only slight differences between different patients for the absorption rate density are reported. One of the most important parameters is the *perfusion* w, both due to its dominant effect on the temperature and its high inter-individual variability. It can be measured by MRI using a contrast agent, but this is not generally indicated.

- *Geometry.* Since therapy planning and, in particular, segmentation takes a significant amount of time, the patient's geometry is acquired some time before the therapy starts. Moreover, several therapy sessions are performed based on this single geometry. Positioning the patient in exactly the same location every time is a major practical difficulty. Even with the utmost care, spatial deviations of roughly 15 mm are unavoidable. Movements of the patient during therapy lead to additional offsets. Changes in body mass and intestine or bladder contents also affect the actual geometry.

- *Technical parameters.* Clinically available HF power generators are less stable in maintaining requested amplitudes and phase shifts than would be desirable. Moreover, part of the HF energy is reflected at contact resistances in coaxial cable connectors between the power generator and the applicator. These resistances depend on the mechanical force with which the connectors have been put together. While the transmission properties of the cable network can, in principle, be measured, this is not always possible in clinical practice.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 3.7. Cross-section of absorption rate density in simulation (a) and MR measurement (b) for a phantom (Weihrauch *et al.* 2007).

Computational identification

A validation of the models describing the electrical field and the heat transport requires *in vivo* measurements of relevant quantities, in particular the temperature. If sufficient data are acquired, the parameters entering the models can be identified, which allows improvement of the therapy. Three-dimensional distributed data sets can be obtained from magnetic resonance tomography (MR) measurements (Rieke and Pauly 2008) in a hybrid MR–hyperthermia applicator: see Figure 3.6. At a rather limited number of discrete times, temperature probes can be placed inside the patient's body. By MR thermometry based on positron resonance frequency shift (PRFS), a combination of tissue temperature increase $T - T_0$ and perfusion w can be measured. Due to lack of better understanding of the relevant processes, a simple measurement model of the form

$$m \approx a(T - T_0) + bw$$

has been used, where the coefficients a and b depend on the tissue type. In fatty tissue in particular, the perfusion influence vanishes almost completely ($b \approx 0$). Resolution, signal-to-noise ratio, and measurement time are mutually dependent. For useful thermometry data, a measurement takes at least 30 s with current MR scanners.

Adaptation of electric fields

As mentioned above, several model parameters are not exactly known in clinical practice. These modelling errors lead to predictions significantly deviating from measurements (see Figure 3.7).

A radical remedy would be to omit the numerical solution of Maxwell's equation completely and to identify the self-adjoint heating effect $M(x) \in \mathbb{C}^{k_E,k_E}$ of applicator control settings using only the k_m MR thermometry measurements (Kowalski, Behnia, Webb and Jin 2002). Assuming the correctness of the computed absorption rate density and the validity of the

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

BHT equation (3.13), Tikhonov regularization leads to the quadratic optimization problem

$$\min_{M\in L^2(\Omega;\mathbb{C}^{k_E,k_E})} \sum_{i=1}^{k_m} \|m_i - aT(t_i;M) - bw\|^2_{L_2(\Omega)} + \gamma\|M\|^2, \tag{3.19}$$

subject to the equality constraints (3.13) and (3.14) with source term $S(x) = \frac{1}{2}\alpha^* M(x)\alpha$. An implementation of this concept in clinical practice is prevented by the number and type of required measurements. Since, for each point or voxel $x \in \Omega$, the number of real parameters in $M(x)$ for an applicator with k_E channels is k_E^2, many measurements have to be taken for a complete identification (the Sigma-Eye applicator with $k_E = 12$ would require more than one hour just to calibrate the model). Moreover, in order to allow a reasonable identification, sufficiently different controls $\alpha \in \mathbb{C}^{k_E}$ have to be used, most of which are of questionable benefit for the patient. Motion artefacts at tissue boundaries prevent the identification of heating effects in a significant part of the domain.

The number of parameters to be identified can be reduced by considering the antenna profiles $\tilde{E} = (E_j)_{j=1,\dots,k_E}$ instead of the heating effect M. Since $M(x) = \tilde{E}(x)^*\tilde{E}(x)$, this leads to a nonlinear identification problem with $6k_E$ parameters per voxel. Because simultaneous phase shifts and simultaneous spatial rotation of field vectors have no thermal effect, a rank defect of at least three reduces the number of parameters further. Still, relying on models identified from scratch is impractical.

A promising approach is to start with antenna profiles obtained as the solutions $\tilde{E}_0$ of Maxwell's equation and adapt them by least-change updates so as to reproduce measurements (Weihrauch *et al.* 2007, Ranneberg *et al.* 2010). This allows us to start the therapy with a reasonably good model, to be improved further whenever new measurements are available. Without the need to sample the whole space of applicator controls, it is, moreover, possible to deliver the best therapy that is known at any time during the whole hyperthermia session.

Solving identification problems of the type (3.19) subject to the time-dependent BHT equation giving $T(t;M)$ in real time is a challenging task. Spatial decoupling by neglecting heat conduction in the identification leads to a set of smaller problems to be solved independently for every voxel, *e.g.*, by a Gauss–Newton method:

$$\min_{\tilde{E}\in\mathbb{C}^{3\times k_E}} \sum_{i=1}^{k_m} |m_i - aT(t_i;M) - bw|^2 + \gamma\|\tilde{E} - \tilde{E}_0\|^2. \tag{3.20}$$

Heat conduction can nevertheless be taken into account without solving a spatially coupled identification problem. Splitting the antenna profiles $\tilde{E}$ into a prediction part $\tilde{E}_0$ based on Maxwell's equations and a correction part

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

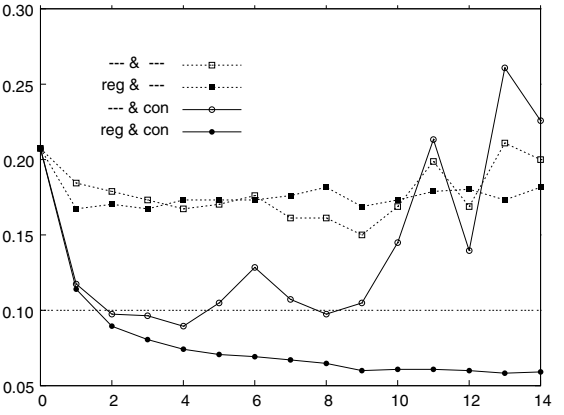


Figure 3.8. L_2-error of the predicted temperatures (in K) over the number of adaptation steps on a simulated therapy. The curves correspond to antenna profile adaptation with and without regularization (reg) as well as with and without taking heat conduction (con) into account. The straight dotted line at 0.1 is the L_2-norm of normally distributed noise applied to generation of artificial measurements from exact temperatures. From Ranneberg *et al.* (2010).

$\delta\tilde{E}$ to be identified, and the temperature $T = T_0 + \delta T$ into corresponding contributions, the time-dependent BHT equation (3.13) reads

$$\rho c \frac{\partial(T_0 + \delta T)}{\partial t} = \operatorname{div}(\kappa \operatorname{grad}(T_0 + \delta T)) - c_b w (T_0 + \delta T - T_{\text{blood}}) + \frac{\sigma}{2}\alpha^* \tilde{E}_0 \tilde{E}_0 \alpha + \frac{\sigma}{2}\alpha^* \big(2\tilde{E}_0^* + \delta\tilde{E}^*\big)\delta\tilde{E}\alpha.$$

Assuming $\delta\tilde{E}$ and δT to be comparatively small, we may neglect heat conduction in δT only, and obtain

$$\rho c \frac{\mathrm{d}\delta T}{\mathrm{d}t} = -c_b w \delta T + \frac{\sigma}{2}\alpha^* \big(2\tilde{E}_0^* + \delta\tilde{E}^*\big)\delta\tilde{E}\alpha$$

as the constraint to (3.20). The temperature T_0 satisfying (3.13) does not depend on $\delta\tilde{E}$, and can thus be computed before or during the MR measurement. This approach yields much better results than neglecting heat conduction completely: see Figure 3.8. For details we refer to Ranneberg *et al.* (2010).

One important point to notice is that, despite being underdetermined, the identification problem is severely ill-conditioned and hence leads to significant noise amplification whenever similar applicator controls α_i are used – a fact that had been long overlooked in the literature. Applied in a feedback loop to improve the applicator controls during therapy, the process will converge towards some optimum. Hence, very similar controls are bound to occur, and sufficient regularization is necessary: see Figure 3.8.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Identification of perfusion

Among the heat transfer model parameters perfusion is the most influential one. Hence, measurement or identification of *individual* perfusion values would be very valuable. Quite accurate values can be obtained from MR data using contrast agents (Lüdemann, Wust and Gellermann 2008). The application of contrast agents may incur an undesirable risk for the patient, which is why they are only used if absolutely necessary. Moreover, due to the required injection of the contrast agent, only very few of these measurements can be obtained, such that a continuous monitoring of thermoregulatory feedback is impossible. Consequently, such data are not generally available.

As a non-invasive alternative, the MR thermometry data m mentioned above can be considered. This leads to an identification problem for temperature and perfusion (Gänzler, Volkwein and Weiser 2006):

$$\min_{T,w} \|m - aT - bw\|^2_{L_2(\Omega)} + \gamma \|w - w_0(T)\|^2_H + b(w, w_0(T)),$$

subject to the equality constraints (3.13) and (3.14). Here, $w_0(T)$ is the *a priori* perfusion model. The norm

$$\|w\|^2_H := \|w\|^2_{L_2(\Omega)} + \sum_{i=1}^{\sharp\text{tissues}} \|\nabla w\|^2_{L_2(\Omega_i)}$$

for the regularization term is constructed based on the assumption that the perfusion depends smoothly on the temperature in every tissue region, but may jump across tissue boundaries. The barrier function b enforces non-negativity of the perfusion, but has to be skewed such that its minimizer is the reference perfusion $w_0(T)$, in order not to introduce a systematic bias.

Clinical application of that concept is prevented by the fact that there can be two different value pairs of temperature and perfusion which are compatible with both the BHT equation and the measurements, which is a consequence of the bilinearity of the BHT equation. Hence, while local consistency is fostered by regularization, global consistency cannot be guaranteed. In fact, numerical experiments indicate that the occurrence of both compatible values must be expected (Weiser 2009).

The consequence is that more data are required to provide the clinical staff with reliable identified temperature and perfusion values. Such data might come from additional measurements taken with MR sequences, which exhibit impact coefficients a and b different from PRFS.

3.4. Therapy planning problem

Our first step is to formulate the medical problem of therapy planning, which then has to be translated into mathematical terms.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Medical problem setting

Due to the production of so-called heat shock proteins, the temperature should preferably be above 42.5°C. Up to now the heating could not be safely restricted to tumour cells alone, not least since it is impossible to distinguish tumour cells reliably from healthy ones. Therefore, in order to be on the safe side for healthy tissue, local temperatures in tumours cannot be allowed to be too high. Thus the therapeutic aim is in general not to burn tumour cells, but to make them more sensitive to other treatments such as radiotherapy or chemotherapy (cytostatica). *In vitro*, experiments have confirmed an increased sensitivity of tumour cells by a factor of 7 over healthy ones, an effect also seen in a milder form *in vivo*: see, *e.g.*, Brizel *et al.* (1996). The physiological mechanism behind this effect is captured in the so-called oxygen enhancement ratio (OER). Consequently, the general aim can be specified as follows:

Control the antenna parameters such that tumour tissue is heated, while healthy and uncertain tissue is not heated above a prescribed threshold temperature.

After these preparations, we are now ready to translate the above medical problem into a mathematical one. Let T denote the temperature inside the body. Following the medical intention, an optimal temperature distribution should satisfy the following items.

- In the region V_{tumour} populated by tumour cells, the temperature should be sufficiently high to incur a therapeutic benefit. In view of the metabolism of heat shock proteins, a temperature level $T_t \geq 42.5°\text{C}$ should be maintained.
- In the region V_{healthy} populated by a significant portion of healthy tissue, the temperature should not exceed a tissue-type-dependent threshold $T_h \approx 42$–$44°\text{C}$, in order to avoid subcutaneous burns.

In addition, there is a region where both tumour and healthy cells are present, which can be treated in a special way (see Section 6.2).

Apart from the medical intention, technical restrictions are imposed by the applicator. The amplitude of each antenna is limited, as is the total power emitted. Moreover, the antenna phases are only reliable as long as the amplitudes are not too different.

These aims and restrictions, partially conflicting, have been modelled and formulated as mathematical optimization problems in a variety of ways, which we will present in the following.

Cost functionals

The first approaches to the quantification of therapeutic benefit to be maximized were formulated in terms of the absorption rate density (3.11) alone

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

(Paulsen *et al.* 1999, Kremer and Louis 1990, Köhler, Maass, Wust and Seebass 2001), based on the idea that high power deposition leads to high temperature. The associated cost functional to be minimized is

$$f(E) = -\frac{1}{2}\int_{V_{\text{tumour}}} \sigma |E(\alpha)|^2_{\mathbb{C}^3}\,\mathrm{d}x. \tag{3.21}$$

An advantage is that this functional can be easily minimized by solving a low-dimensional eigenvalue problem (Kremer and Louis 1990, Köhler *et al.* 2001), as long as no temperature constraints are prescribed.

For medical reasons, a temperature-based objective functional is preferable. Therefore, in Nikita, Maratos and Uzunoglu (1993) and Paulsen *et al.* (1999), such a functional has been suggested. The idea is to penalize 'cold spots' in the tumour below a therapeutic temperature $T_t \approx 43°C$, since the impact of hyperthermia on cancer cells drops quickly below this threshold. This leads to the functional

$$f(T) = \int_{V_{\text{tumour}}} (T_t - T)_+^2\,\mathrm{d}x \tag{3.22}$$

with $(\cdot)_+ = \max(\cdot, 0)$. This approach can be extended to the concept of thermal dose, with the aim of minimizing the fraction of surviving tumour cells. The thermal damage rate inflicted on cancer cells follows the *Arrhenius law* (Dewey 1994, Szasz and Vincze 2006), with a rate constant

$$k(T) = A\exp\left(\frac{\Delta E}{RT}\right).$$

Here, the frequency factor A and the activation energy ΔE depend on the tissue type, and R is the universal gas constant. An estimate s of the fraction of surviving cells is given by the simple differential equation

$$\dot{s}(t,T) = k(T)s(t,T), \quad s(0,T) = 1,$$

which gives rise to

$$s(t,T) = \exp(tk(T)). \tag{3.23}$$

Thus the objective functional to be minimized is the fraction of cancer cells surviving after a prescribed duration t of a therapy session:

$$f(T) = \int_{V_{\text{tumour}}} \exp\left(-tA\exp\left(\frac{\Delta E}{RT}\right)\right)\mathrm{d}x. \tag{3.24}$$

A schematic representation of such a function is given in Figure 3.9. The thermal isoeffect dose is an established quantity for assessing the therapeutic benefit of a treatment (Dewey 1994, Rosner, Clegg, Prescott and Dewhirst 1996). Treatment planning based on the tumour cell survival has been proposed for thermoseed placement (Tompkins *et al.* 1994*a*), but up to

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

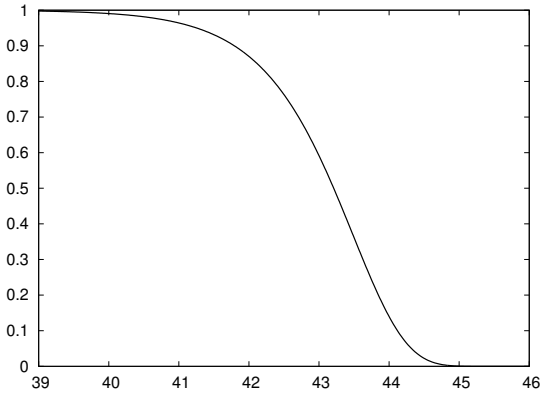


Figure 3.9. Dependence of cancer cell survival rate on tissue temperature according to (3.23).

now rather *ad hoc* cost functionals like (3.22) have been used in regional hyperthermia planning.

Temperature constraints

At the time when mathematical therapy planning started, a sufficient mathematical understanding of state constraints and an efficient algorithmic treatment were lacking. A common approach was therefore an external penalization of 'hot spots' arising in healthy tissue with temperatures above a tissue-dependent threshold $T_{\lim}$, occasionally augmented by an additional, weaker penalization of heating healthy tissue above a comfortable temperature $T_h < T_{\lim}$ (Erdmann, Lang and Seebass 1997). This leads to extended cost functionals

$$J(T) = f(T) + \int_{V_{\text{healthy}}} \left(\gamma_1 (T - T_h)_+^2 + \gamma_2 (T - T_{\lim})_+^2\right) \mathrm{d}x,$$

with fixed penalty parameters $0 \leq \gamma_1 \ll \gamma_2$. Due to external penalization, however, 'hot spots' cannot be completely prevented. In fact, they are bound to arise unless $\gamma_2 \to \infty$. An analysis of the asymptotic behaviour performed in Schiela and Hintermüller (2011) suggests that even under favourable circumstances the maximum constraint violation only decreases by a rate of $O(\gamma_2^{-2/5})$. We would like to point out that in hot spots with less regularity of the temperature distribution, in particular due to discontinuous temperature bound or singularities of the absorption rate density, the maximum violation decreases even more slowly.

A strict *prevention of hot spots* in healthy tissue will require the explicit satisfaction of state constraints of the kind

$$T(x, \alpha) \leq T_{\lim}, \quad x \in V_{\text{healthy}}. \tag{3.25}$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

One way to guarantee feasibility of the temperature distribution is to use a barrier method. In Section 5 below, we will work out results in terms of both theory and algorithms.

Control constraints

Additional constraints directly on the control α are imposed to model the limited power of the microwave applicator:

$$\begin{aligned} |\alpha_k|_{\mathbb{C}} &\leq \alpha_{\max}, \quad k = 1, \dots, k_E, \\ \sum_{k=1}^{k_E} |\alpha_k|_{\mathbb{C}}^2 &\leq \alpha_{\max}^{\text{total}}. \end{aligned} \tag{3.26}$$

If desired, stability of antenna phases can be ensured by confining the amplitude ratios to

$$\alpha_{\text{ratio}}^{-1} \leq \frac{|\alpha_k|_{\mathbb{C}}}{|\alpha_j|_{\mathbb{C}}} \leq \alpha_{\text{ratio}} \quad j, k = 1, \dots, k_E, \quad \text{with } \alpha_{\text{ratio}} > 1,$$

or by requesting all antennas to have the same amplitude:

$$|\alpha_j|_{\mathbb{C}} = |\alpha_k|_{\mathbb{C}}, \quad j, k = 1, \dots, k_E.$$

Optimal control problems

Putting all the pieces together, different optimization problems can be formulated. Omitting the control constraints, we state two variants that have been used for therapy planning.

Penalized problem:

$$\min_{\alpha \in \mathbb{C}^{k_E}, T \in H^1(\Omega)} \int_{V_{\text{tumour}}} (T_t - T)_+^2 \, \mathrm{d}x + \int_{V_{\text{healthy}}} \left(\gamma_1 (T - T_h)_+^2 + \gamma_2 (T - T_{\text{lim}})_+^2 \right) \mathrm{d}x \tag{3.27a}$$

subject to the BHT equation (3.15),

$$A(T) - B(\alpha) = 0. \tag{3.27b}$$

Constrained problem:

$$\min_{\alpha \in \mathbb{C}^{k_E}, T \in H^1(\Omega)} \int_{V_{\text{tumour}}} \exp\left(-tA \exp\left(\frac{\Delta E}{RT} \right) \right) \mathrm{d}x \tag{3.28a}$$

subject to the BHT equation (3.15) and temperature constraints,

$$A(T) - B(\alpha) = 0, \quad T \leq T_{\text{lim}}. \tag{3.28b}$$

This above set of goal functionals and constraints supplies the starting point for the actual computation of optimal therapies for individual patients.

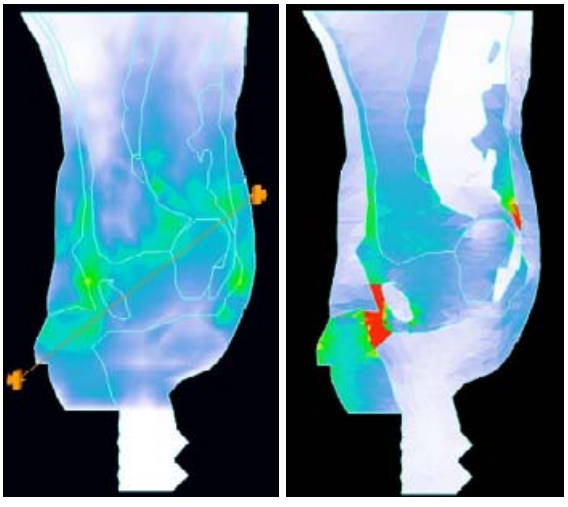
https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

4. Efficient simulation

In the numerical simulation of the whole hyperthermia procedure, *computational efficiency* is of utmost importance, since clinical decision-making requires numerical algorithms that run on a local PC *both fast and reliably*. The simulation involves solving both the time-harmonic Maxwell equation and the BHT equation. The latter, a Helmholtz equation in the non-critical case, can be readily solved by algorithms which are now standard textbook material: see, *e.g.*, Deuflhard and Weiser (2012). That is why we skip this part here. Rather, we concentrate on the two main sources of computational cost, the adaptive multilevel finite element methods for the time-harmonic Maxwell equation in Section 4.1 and their unidirectional coupling with the BHT equation in Section 4.2. An initial survey of the decisive role of these methods in hyperthermia treatment planning has been given in Deuflhard and Seebass (1999).

4.1. Adaptive FEM for the time-harmonic Maxwell equation

The first algorithmic attempts to solve the time-harmonic Maxwell equation were based on the rather popular FDTD method: see Sullivan (1990).



(a) (b)

Figure 4.1. Comparison of electrical field computations. (a) Popular FDTD method (Sullivan 1990) based on the strong formulation (without correct interface conditions). (b) Adaptive finite element method based on the weak formulation (with automatically correct interface conditions). See Section 4.1.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

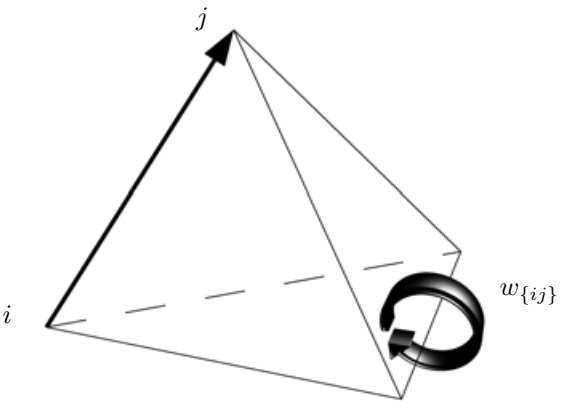


Figure 4.2. The tetrahedral edge element shape function $w_{\{ij\}}$ associated to edge $\{ij\}$ is a circular field rotating around the opposite edge.

These methods started directly from the voxel discretization, that is, the raw data, as input for numerical integration, certainly an intriguing feature. In addition, the methods appeared to be very fast; but, after testing them in a hyperthermia environment, it turned out that they were unable to treat the interface conditions correctly, which in an FEM would be automatically satisfied. A comparison of results for a special virtual patient is given in Figure 4.1.

The typical ingredients of an *adaptive multilevel* finite element method (FEM) are (a) an *a posteriori local error estimator*, (b) an *adaptive mesh refinement strategy*, and (c) a *fast solver* for the large linear systems arising. For Poisson-type PDEs, such techniques are well developed. For the time-harmonic Maxwell equation, however, a deeper understanding of the indefinite structure is necessary to reach optimal multigrid complexity.

Edge elements

We return to the weak formulation (3.5) in Section 3.1 above. In order to obtain a finite element discretization of the weak equation (3.6), we generate a *tetrahedral* triangulation $\mathcal{T}_h$ of the domain Ω and employ Nédélec's curl-conforming finite elements of first order (Nédélec 1980), also called *edge elements* (Bossavit 1988) or Whitney 1-forms. Let us denote the space of linear Lagrange elements by S_h and the nodal basis functions associated to vertex i by λ_i. Then the edge element basis function associated with the edge $e = \{ij\}$ is defined by

$$w_{\{ij\}} = \lambda_i \mathrm{grad}\, \lambda_j - \lambda_j \mathrm{grad}\, \lambda_i \tag{4.1}$$

(see Figure 4.2). The corresponding FE space is

$$ND_h := \mathrm{span}\{w_{\{ij\}}\}.$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

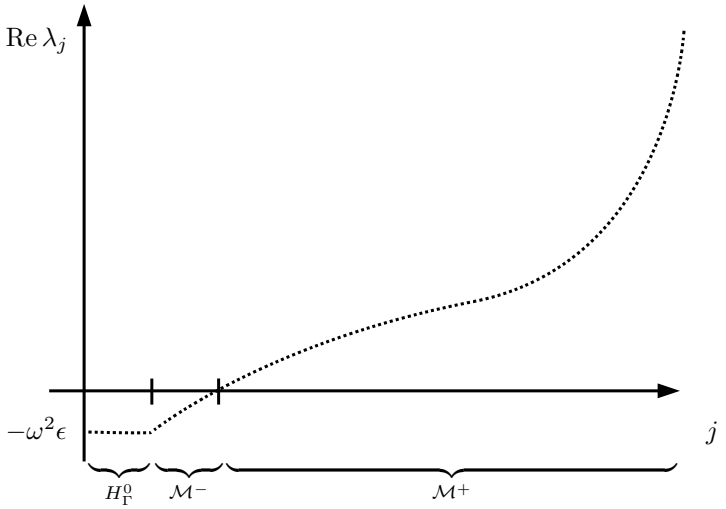


Figure 4.3. Eigenvalue structure for the time-harmonic Maxwell operator from (3.5).

Edge elements possess several distinct advantages compared to vectorial Lagrange elements. They guarantee the desired continuity of the tangential components of the electric field while preserving jumps of the normal components. In particular, concave metallic edges and internal boundaries between materials with jumps in the coefficients can be naturally incorporated in agreement with physical continuity relations. Moreover, they avoid the occurrence of spurious modes, whereas nodal basis functions may give rise to wrong solutions (see Bossavit 1990). The convergence behaviour of edge element discretizations of (3.6) is studied in Monk (1992). For the actual realization we note that for a vector field E the degree of freedom $E_e \in \mathbb{C}$ associated with each edge e in $\mathcal{T}_h$ is given by the path integral

$$E_e = \int_e E \cdot t_e \, \mathrm{d}s \tag{4.2}$$

along the edge e with direction vector t_e.

Spectral properties

Before we dive into algorithmic details, let us have a closer look at the spectral structure of the time-harmonic Maxwell operator from (3.5) that occurs in the weak equation (3.6). As illustrated in Figure 4.3, its spectrum divides into three distinct parts: (i) eigenvalues with negative real part originating from the nullspace $H^0_\Gamma(\mathrm{curl}, \Omega)$ of the curl-operator, (ii) eigenvalues with negative real part to be associated with some space $\mathcal{M}^-$ spanned by the corresponding eigenfunctions, and (iii) eigenvalues with positive real part, which go with some space $\mathcal{M}^+$.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Coarse mesh

Suppose we formally expand both sides of the weak equation (3.6) in terms of eigenfunctions of the Maxwell operator. Then the system would decouple in terms of these eigenfunctions. Those terms on the right-hand side with *smallest* coefficients will dominate the solution structure in the generic case – assuming, of course, that all eigenfunctions are present in the actual solution. Recall that, in the Laplace operator, the *smooth* eigenfunctions have the smallest eigenvalues and thus dominate the solution structure. The situation is very different for the indefinite time-harmonic Maxwell operator. To see this, a simple plane wave analysis has been given in Beck *et al.* (1999), which we repeat here. For simplicity, let the material be homogeneous, so that the coefficients μ, ϵ are constants. Let $E(x) := \exp(\mathrm{i}kx)e$ denote a plane spatial wave in a polarization direction e. Upon ignoring the boundary terms in A, eigenvalues $|k|^2 - \omega^2\mu\epsilon$ will come up. Obviously, wavenumbers k close to the critical wavenumber $k_C = \omega\sqrt{\mu\epsilon}$ give rise to the smallest eigenvalues of the time-harmonic operator A. The associated eigenfunctions are bound to dominate the solution generically. Let the critical wavelength be defined by

$$\lambda_C = \frac{2\pi}{k_C} = \frac{2\pi}{\omega\sqrt{\mu\epsilon}}.$$

Then an appropriate coarse mesh able to resolve solutions in the space $\mathcal{M}^-$ (see Figure 4.3) should obey a mesh size restriction of the kind

$$h \leq \frac{\lambda_C}{2}, \tag{4.3}$$

which is often called the *Nyquist condition*. It will again show up below in the context of multigrid methods. In passing we note that the above restriction depends, of course, on the material. In the hyperthermia setting, it is only restrictive in the water bolus, producing a nearly uniform mesh therein.

Large linear system solvers

Replacing $H_\Gamma(\mathrm{curl};\Omega)$ by ND_h in the variational formulation (3.6) leads to a large, very sparse linear equation system

$$Au = b \tag{4.4}$$

with complex symmetric and indefinite matrix A. The solution of (4.4) can be obtained by either direct or iterative solvers. *Direct* solvers have experienced significant progress in recent years (Davis 2006). Despite a still considerable memory requirement due to fill-in, they are very effective on coarse up to moderately fine meshes. It is only for very fine grids, or if low memory requirement is important, that iterative solvers have to be used.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Iterative solvers face two difficulties, as can be seen from the eigenvalue distribution of the operator A in Figure 4.3, which carries over to the matrix A. First, the large eigenvalues grow with decreasing mesh size h. The growing condition number of the matrix A renders both matrix decomposition methods and Krylov methods inefficient on fine meshes. Second, due to the eigenvalues with negative real part, usual positive smoothers such as Jacobi or Gauss–Seidel methods do not converge: the spectral radius of their iteration matrix $I - B^{-1}A$ exceeds 1, so the corresponding modes in $H^0_\Gamma \cup \mathcal{M}^-$ are amplified rather than reduced.

Adaptive mesh refinement

Due to the complex geometrical structure and strongly varying material properties of virtual patients, corner singularities and other local solution features are bound to arise. Hence we consider adaptive mesh refinement to be of crucial importance to ensure both efficiency of the solver and reliability of its computational results. The foundation of adaptive mesh refinement are *a posteriori error estimators*, which are most often based on localized operations. As we are dealing with equations of a partly hyperbolic nature in this context, it is clear that any local estimator will be unable to capture far-field contributions. Accordingly we prefer to use the name *error indicator* in the sense of Babuška and Aziz (1972).

Among the different flavours of error indicators, we consider hierarchical error indicators in the spirit of Deuflhard *et al.* (1989), which have been observed to yield very favourable results with efficiency quite close to 1 on simple test problems (Beck *et al.* 1999, Section 8.2). Quantitative reliability is, of course, of utmost importance in medical applications.

Conceptually, the ansatz space $V^L = ND_h$ is extended by a hierarchical surplus V^H to give a higher-order ansatz space $V^+ = V^L \oplus V^H$, in which (4.4) assumes the block structure

$$\begin{pmatrix} A^{LL} & A^{LH} \\ A^{HL} & A^{HH} \end{pmatrix} \begin{pmatrix} u^L \\ u^H \end{pmatrix} = \begin{pmatrix} b^L \\ b^H \end{pmatrix}. \tag{4.5}$$

For the mere purpose of error indication, (4.5) only needs to be solved approximately via a defect equation for the hierarchical components:

$$A^{HH}\, u^H = b^H - A^{HL}\, u^L. \tag{4.6}$$

Hierarchical extensions of the lowest-order edge elements have been described in Savage and Peterson (1996), where the basis functions are given in terms of Lagrangian nodal functions λ_i associated to the vertices i of the tetrahedral mesh. With each edge $\{ij\}$, one basis function

$$w^H_{\{ij\}} = \lambda_i \mathrm{grad}\, \lambda_j + \lambda_j \mathrm{grad}\, \lambda_i$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Table 4.1. Estimated discretization errors versus refinement level j with node numbers N_j. The 'coarse' mesh ($j = 0$) is taken from Figure 4.4.

j	N_j	η
0	128 365	10.1 %
1	373 084	4.96 %
2	1 085 269	2.54 %

and with each face $\{ijk\}$, two basis functions

$$w^{H,1}_{\{ijk\}} = \lambda_i\lambda_j \mathrm{grad}\,\lambda_k - \lambda_i\lambda_k \mathrm{grad}\,\lambda_j, \quad w^{H,2}_{\{ijk\}} = \lambda_i\lambda_j \mathrm{grad}\,\lambda_k - \lambda_j\lambda_k \mathrm{grad}\,\lambda_i$$

are associated. An approximate solution of (4.6) may be obtained via one block-Jacobi step only (!), where it is essential to keep the (2×2)-block-entries of both functions $w^{H,l}_{\{ijk\}}$, $l = 1, 2$, attached to each face of the triangulation (see Beck *et al.* 1999 for details). At first glance, this block-Jacobi sweep seems to be critical, as the matrix contains negative eigenvalues and the related eigenmodes may be amplified. However, as the modulus of all negative eigenvalues is comparatively small, these modes can be expected to give only minor contributions. Let $\tilde{u}^H$ denote the first Jacobi iterate of the exact u^H. For the definition of a *relative* discretization error estimator we set

$$\eta^2 = \frac{\|\tilde{u}^H\|^2_E}{\|u^L\|^2_E},$$

where, in the absence of a better idea, we employ the norm

$$\|v\|^2_E := \int_\Omega \left(\frac{1}{\mu} \mathrm{curl}\, v^* \ \mathrm{curl}\, v + \omega^2 \epsilon v^* \, v \right) \mathrm{d}x + \int_{\Gamma_{\mathrm{ext}}} \beta (n \times v^*)\,(n \times v)\, \mathrm{d}s. \tag{4.7}$$

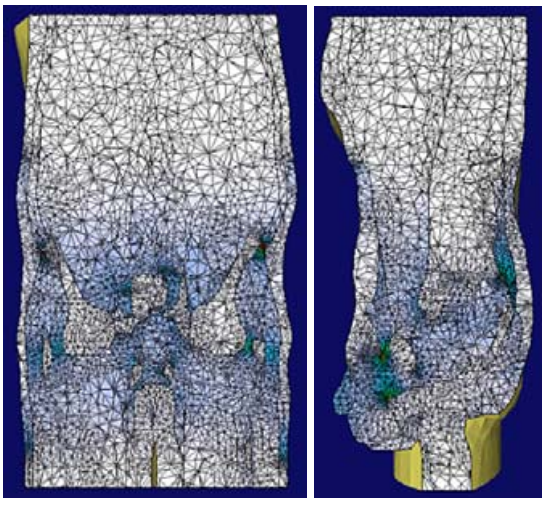
This error estimator is the key ingredient in an *adaptive refinement strategy* in the spirit of a technique suggested by Babuška and Rheinboldt (1978) for purely elliptic problems. As an illustration of an adaptive grid constructed along these lines, results for a virtual patient mesh are shown in Figure 4.4.

Table 4.1 gives a list of error indicators as a function of the refinement levels starting from the mesh of Figure 4.4.

Multilevel preconditioning

In the hyperthermia problem, a multigrid V-cycle preconditioner is used within a conjugate residual iterative method (CR): see Freund (1992). On

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 4.4. Adaptive mesh of a virtual patient consisting of about 128 000 nodes. Note that a uniform grid would have required an estimated 16 000 000 nodes. (a) Frontal cut. (b) Sagittal cut.

Table 4.2. Convergence history for multilevel solvers with standard Gauss–Seidel smoothing (Std) compared to multiplicative (M-H) and additive (A-H) versions using hybrid Gauss–Seidel smoothing on each level. The iteration is terminated if the ratio of the Euclidean norms of residuals and right-hand sides is below 10^{-5}. From Deuflhard and Seebass (1999).

Ref. depth	Nodes	No. of iterations			CPU (min)		
		Std	M-H	A-H	Std	M-H	A-H
0	128 365	4250	354	413	150	24	20
1	373 084	4832	265	277	800	76	60
2	1 085 269	> 10000	186	194	> 2000	215	160

the subspace $\mathcal{M}^+$, *standard multigrid* methods are effective in reducing the error – however, at the expense of amplifying error components in $H^0_\Gamma \cup \mathcal{M}^-$. This leads to a deterioration of the whole multigrid procedure in the present case: see Table 4.2 under columns Std. Therefore, the standard multigrid approach needs to be modified.

The general applicability of multigrid methods to indefinite problems has been investigated in Bramble, Kwak and Pasciak (1994). In order to capture the error part in $\mathcal{M}^-$, certain restrictions on the mesh size of the coarse grid have to be observed – unfortunately, given there in terms of unknown

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

theoretical quantities. Fortunately, both local mode analysis and numerical experience confirmed that it is enough to observe the above Nyquist condition (4.3). The coarse grid correction is then computed by a direct solver. Treating the part originating from $H^0_\Gamma(\mathrm{curl}, \Omega)$ is more delicate, since it contains functions with all spatial frequencies and therefore cannot be captured on the coarse mesh.

For Nédélec finite elements there exists a discrete counterpart of the representation of $H^0_\Gamma(\mathrm{curl}, \Omega)$ as a space of gradients of potentials. Thus, on simply connected domains each curl-free element of ND_h can be represented as the gradient of a scalar potential in the space S_h of nodal finite elements:

$$ND_h \cap H^0_\Gamma(\mathrm{curl}, \Omega) = \mathrm{grad}\, S_h, \tag{4.8}$$

Moreover, the stable decomposition (3.7) has a (perturbed) discrete counterpart, too. To this end, let $I^{\mathrm{curl}} : (S_h)^3 \to ND_h$ denote the edge interpolation operator, defined via (4.2). Then

$$v = \tilde{v} + I^{\mathrm{curl}}\phi + \mathrm{grad}\, p, \quad \phi \in (S_h)^3,\ p \in S_h,\ \tilde{v} \in ND_h \tag{4.9}$$

such that

$$\|h^{-1}\tilde{v}\|_{L_2} + \|\phi\|_{H^1} + \|p\|_{H^1} \leq C\|v\|_{H(\mathrm{curl})}. \tag{4.10}$$

This outstanding feature of Nédélec elements (see, *e.g.*, Beck and Hiptmair 1999, Hiptmair 2002*a*) is of crucial importance for the construction of efficient multigrid solvers for the linear systems arising, as we will illustrate for the simpler example of *positive definite problems*.

A multigrid method, applied to a linear system $Au = b$ on a space V, consists of a (non-direct) splitting

$$V = \sum V_i$$

and preconditioners B_i, defined on each V_i. Here we will choose V_i as one-dimensional spaces that come from a multilevel hierarchy and $B_i = A$. The convergence theory then depends on the stability condition

$$\inf_{v=\sum v_i} \sum \|v_i\|^2_A \leq K_1 \|v\|^2_A, \tag{4.11}$$

and the strengthened Cauchy–Schwarz inequality

$$\langle v_k, v_l \rangle_A \leq \gamma_{kl} \|v_k\|_A \|v_l\|_A \quad K_2 := \|\gamma\|_2, \tag{4.12}$$

which must be established for the given splitting (Xu 1992). In the additive case (*e.g.*, Jacobi-type smoothing) the product $K_1 K_2$ gives an estimate for the condition number of the preconditioned system. If $K_1 K_2$ admits an upper bound that is independent of h, we obtain an optimal preconditioner, *i.e.*, a bounded condition number. Similar results hold in the multiplicative case, *e.g.*, for Gauss–Seidel-type smoothing.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

To understand why standard multigrid smoothers are inefficient in an $H(\mathrm{curl})$-setting, consider $\|\cdot\|_A = \|\cdot\|_{H(\mathrm{curl})}$ and a multilevel splitting

$$ND_h = \sum \Phi_i,$$

where Φ_i are one-dimensional spaces, spanned by the edge element basis functions on different levels. Now let $v \in ND_h \cap H^0_\Gamma(\mathrm{curl}, \Omega)$ (recall that $\mathrm{curl}\, v = 0$), such that

$$v = \sum_{\phi_i \in \Phi_i} \phi_i.$$

The basis functions satisfy $\|\phi_i\|_{L_2} \le c h_i \|\phi_i\|_{H(\mathrm{curl})}$. Thus we obtain

$$\|v\|^2_{H(\mathrm{curl})} = \|v\|^2_{L_2} \le C \inf_{v=\sum \phi_i} \sum \|\phi_i\|^2_{L_2} \le C \inf_{v=\sum \phi_i} \sum h_i^2 \|\phi_i\|^2_{H(\mathrm{curl})}. \tag{4.13}$$

Whenever v bears high spatial frequencies, this implies

$$K_1 \ge C h^{-2},$$

and thus leads to mesh-dependent condition numbers and contraction rates.

Returning again to (4.8), we may introduce a potential $p \in S_h$, such that $v = \mathrm{grad}\, p$. Then, for a multilevel decomposition

$$S_h = \sum P_i,$$

where P_i are one-dimensional subspaces spanned by the nodal basis functions of S_h on different levels, standard H^1-multigrid theory yields

$$\inf_{p=\sum p_i} \sum \|p_i\|^2_{H^1} \le K_{1,H^1} |p|^2_{H^1} = K_{1,H^1} \|\mathrm{grad}\, p\|^2_{L_2} = K_{1,H^1} \|v\|^2_{H(\mathrm{curl})}. \tag{4.14}$$

For general $v \in ND_h$, we can apply the stable discrete decomposition (4.9), which finally yields a stable multigrid decomposition of the form

$$V = \sum \Phi_i + \sum \mathrm{grad}\, P_i, \tag{4.15}$$

where Φ_i and P_i are taken from a multilevel hierarchy of ND_h and S_h, respectively. Then a multigrid estimate similar to (4.14) for the rotational part of v, together with (4.14), yields an h-independent constant K_1. Note that the mere existence of such a splitting is sufficient: it need not be computed explicitly. More details can be found in Beck and Hiptmair (1999), Beck *et al.* (1997, 1999), Hiptmair (1999), Hiptmair and Xu (2007) and Xu, Chen and Nochetto (2009).

On the basis of the splitting (4.15), a hybrid smoothing procedure was constructed in Beck and Hiptmair (1999) and Beck *et al.* (1999). There, as a basic solver the conjugate residual (CR) method was used, which is similar to the well-known conjugate gradient algorithm, but adjusted to symmetric

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

indefinite systems. Its basic operations are Gauss–Seidel sweeps, both in the Nédélec space ND_h, coping with the elliptic part of A, and in the space S_h of potentials. Within this framework, efficient transfer operators between field representations in S_h and ND_h are essential.

If we represent a vector field $v \in ND_h \cap H^0_\Gamma(\mathrm{curl}, \Omega)$ by $v = \mathrm{grad}\, p$, then the representation in ND_h can be obtained easily from (4.2):

$$v_e = \int_{P_1}^{P_2} \mathrm{grad}\, p \cdot t \, \mathrm{d}s = p(P_2) - p(P_1).$$

Here P_1 and P_2 denote the positions of the endpoints of the edge e.

In the following, P_{S_h} denotes the transfer operator from the potential space S_h into the Nédélec space ND_h on $\mathcal{T}_h$. $P^*_{S_h}$ denotes the adjoint operator, defining the canonical restriction. Then the smoother A_S for the nullspace is constructed by the Galerkin product $A_S = P^*_{S_h} A P_{S_h}$. The hybrid smoothing algorithm for a given right-hand side vector $b \in ND_h$ will provide an updated solution vector $u \in ND_h$ via the following steps:

One **Gauss–Seidel** step for $Au = b$
$\tilde{b} \leftarrow b - Au$
$b_S \leftarrow P^*_{S_h} \tilde{b}$
$u_S \leftarrow 0$
One **Gauss–Seidel** step for $A_S u_S = b_S$
$u \leftarrow u + P_{S_h} u_S$

In order to be on safer theoretical grounds, we realize symmetric preconditioning, which means an immediately following step of the adjoint procedure. Note that in this approach both spaces are treated in a *multiplicative* fashion in the sense of domain decomposition methods.

As an alternative, we propose a symmetric *additive* version:

One **symmetric Gauss–Seidel** step for $Au = b$
$b_S \leftarrow P^*_{S_h} b$
$u_S \leftarrow 0$
One **symmetric Gauss–Seidel** step for $A_S u_S = b_S$
$u \leftarrow u + P_{S_h} u_S$

If we employ adaptive mesh refinement, thus creating a sequence of nested triangulations $\mathcal{T}_0 \subset \mathcal{T}_1 \subset \ldots \subset \mathcal{T}_h$, then the extension to a multilevel solver is quite straightforward. Taking into consideration that the associated sequence of Nédélec spaces is nested, *i.e.*, $ND_0 \subset ND_1 \subset \cdots \subset ND_h$, we may adopt the classical multigrid idea (Hackbusch 1985) by using canonical grid transfer operations between these spaces, but employing a hybrid smoother on each level.

Finally, in order to demonstrate the efficiency of this hybrid algorithm, we present comparative results for both the multiplicative and the additive

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

hybrid algorithm using the discrete Helmholtz decomposition in Table 4.2 above. As can be seen, the hybrid versions exhibit a superior performance and do not deteriorate with increasing refinement level. Note that we have no direct factorization available on the coarse grid, thus facing comparatively large iteration counts. A proof of optimal multigrid complexity of this type of algorithm can be found in Beck, Deuflhard, Hege and Stalling (1997) and Hiptmair (1996).

4.2. Coupling of the BHT and the Maxwell equations

In the covered temperature range of 30–50°C, the electromagnetic material parameters are essentially constant, such that the coupling of the time-harmonic Maxwell equation and the BHT equation is unidirectional. Thus, we can first solve the time-harmonic Maxwell equation and insert the results into the BHT equation. Since the Maxwell equation is linear, we can compute the antenna profiles E_k for each antenna channel $k = 1, \ldots, k_E$ and use superposition afterwards. For the insertion into the BHT equation, however, a distinction between the linear and the nonlinear case is necessary.

Linear BHT equation

As can be seen from equation (3.12), the superposition of the electric field E from k_E antenna profiles gives rise to a superposition of k_E^2 source modes to the source term S:

$$\sum_{j,l=1}^{k_E} \alpha_j^* \alpha_l E_j^* E_l = \sum_{j=1}^{k_E} |\alpha_j|_{\mathbb{C}}^2 |E_j|_{\mathbb{C}}^2 + 2 \sum_{j=2}^{k_E} \sum_{l=1}^{j-1} \big(\mathrm{Re}\,(\alpha_j^* \alpha_l) \mathrm{Re}\,(E_j^* E_l) - \mathrm{Im}\,(\alpha_j^* \alpha_l) \mathrm{Im}\,(E_j^* E_l) \big).$$

The source modes,

$$S_{jl} = \begin{cases} \mathrm{Re}\,(E_j^* E_l) & j \le l, \\ \mathrm{Im}\,(E_j^* E_l) & j > l, \end{cases}$$

correspond to temperature elevations T_{jl} above the basic tissue temperature T_{bas} for $\alpha = 0$ and can be computed in advance by solving the homogeneous BHT equation

$$\begin{aligned} -\mathrm{div}(\kappa \,\mathrm{grad}\, T_{jl}) &= -c_b w T_{jl} + \frac{1}{2} \sigma S_{jl} \quad \text{in } \Omega, \\ -\kappa \frac{\partial T_{jl}}{\partial n} &= h T_{jl} \qquad\qquad\qquad\quad \text{on } \partial\Omega. \end{aligned}$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

With coefficients $\gamma_{jl} = 2\,\mathrm{Re}\,(\alpha_j^* \alpha_l)$ for $j < l$, $\gamma_{jj} = |\alpha_j|_{\mathbb{C}}^2$, and $\gamma_{jl} = -2\,\mathrm{Im}\,(\alpha_j^* \alpha_l)$ for $j > l$, the temperature is obtained as

$$T(\alpha) = T_{\mathrm{bas}} + \sum_{j,l=1}^{k_E} \gamma_{jl} T_{jl}.$$

Let $\mathrm{cost}_{\mathrm{linBHT}}$ denote the computational cost for one solution of the BHT equation. Then the total computational cost sums up as

$$\mathrm{cost}_{\mathrm{total}} = k_E\,\mathrm{cost}_{\mathrm{Maxwell}} + (k_E^2 + 1)\,\mathrm{cost}_{\mathrm{linBHT}} + \mathrm{cost}_{\mathrm{Opt}}. \tag{4.16}$$

As an illustration, see Table 6.1 below. The cost for the optimization as a whole depends, of course, on the number of iterations.

Nonlinear BHT equation

In the case of a temperature-dependent perfusion $w(T)$, some kind of iteration is needed to take care of this nonlinearity. Deuflhard and Seebass (1999) have suggested the following fixed point iteration:

k_E **Maxwell solves:** E_k

let $\alpha^0 := 0$

for $m = 0, \ldots, n_{\mathrm{iter}}$:

one *nonlinear* BHT equation solve for $S(\alpha^m)$
supplies temperature T^m and perfusion $w^m := w(T^m)$

$k_E^2 + 1$ ***linear* BHT equation solves** for S_{jl} using w^m
supplies T_{bas}^m and temperature modes T_{jl}^m

optimization $\min_\alpha f(T_{\mathrm{bas}}^m + \sum_{j,l=1}^{k_E} \gamma_{jl}(\alpha) T_{jl}^m)$
supplies α^{m+1}

The idea behind this algorithm has been that the Maxwell solves are considerably more expensive than the BHT solves. The iteration converges linearly and quite fast at an observed contraction rate $\theta \approx 0.3$. So the total computational cost sums up as

$$\begin{aligned}\mathrm{cost}_{\mathrm{total}} = {} & k_E\,\mathrm{cost}_{\mathrm{Maxwell}} \\ & + n_{\mathrm{iter}}\,(\mathrm{cost}_{\mathrm{nonlinBHT}} + (k_E^2 + 1)\,\mathrm{cost}_{\mathrm{linBHT}} + \mathrm{cost}_{\mathrm{Opt}}).\end{aligned} \tag{4.17}$$

In typical computations, $n_{\mathrm{iter}} \approx 6$ optimization iterations are required. The nonlinear BHT equation has been solved by a Newton iteration in function space (Deuflhard *et al.* 2000), typically requiring no more than two steps due to the fast local convergence.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

5. PDE-constrained optimization

In Section 3.4, we had already worked out the medically preferable optimization problem with *strict temperature limits in healthy tissue*. When this aspect had come into the focus of interest, there were no efficient algorithms to tackle such problems. That is why thorough theoretical investigations then began: we present these in Section 5.1, and postpone algorithmic details to Section 5.2. In order to derive efficient *adaptive* algorithms, we follow a multistage computational paradigm. In the first stage, we develop the necessary numerical analysis background in a function space setting, which covers both optimality conditions and formal algorithms for the infinite-dimensional problem. In the second stage, we construct an iterative method, again in the infinite-dimensional setting, for the nonlinear problems arising. Only in the last stage, the innermost loop, where linear operator equations are to be solved, do we discretize the infinite-dimensional problem up to a prescribed error tolerance. This kind of scheme has emerged as the general basis for any adaptive PDE algorithm.

5.1. Theoretical background

The hyperthermia treatment planning problem can be formulated as an optimization problem *in function space*, subject to equality constraints and inequality constraints, in the following form:

$$\min f(T, \alpha), \tag{5.1a}$$

subject to

$$A(T) - B(\alpha) = 0, \tag{5.1b}$$

$$|\alpha_k|_{\mathbb{R}^2} \leq \alpha_{\max} \quad T \leq T_{\lim}. \tag{5.1c}$$

Here A and B represent the BHT equation in weak form (3.15). From Section 3.2 we have α as a set of $2k_E - 1$ *real* parameters.

Various cost functionals f and inequality constraints have been discussed in Section 3.4. The most challenging parts are the constraints on the temperature $T \leq T_{\lim}$. They can be considered as an infinite number of inequality constraints in function space. In PDE-constrained optimization and optimal control they are classified as pointwise *state constraints*, and have been subject to intense research in recent years; see, *e.g.*, Casas (1986), Meyer, Tröltzsch and Rösch (2006), Hintermüller and Kunisch (2006) and Deckelnick and Hinze (2007).

The following derivations are based on Schiela and Weiser (2010) and Schiela (2009). More details and proofs can be found therein.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Barrier regularization and strict feasibility

In *finite* dimensions, barrier regularizations give rise to the highly popular class of interior point methods (Terlaky 1996, Wright 1987). With $y = (y_1, \dots, y_n)$, the principal idea is to replace an inequality-constrained problem of the kind

$$\min_{y \in \mathbb{R}^n} f(y) \text{ subject to } y_i \geq 0 \quad i = 1 \dots n$$

by an unconstrained problem, depending on a parameter $\mu > 0$,

$$\min_{y \in \mathbb{R}^n} f(y) - \mu \sum_{i=1}^{n} \log(y_i),$$

where the functional f has been extended by a sum of *logarithmic barrier functions*. This new minimization problem with minimizer y^μ can then be solved approximately. By driving the barrier parameter to its limit $\mu = 0$ we can approximate the solution of the original problem. The logarithmic terms ensure strict feasibility $y_i^\mu > 0$ for all $\mu > 0$, so Newton's method can be applied to the corresponding first-order optimality conditions in a *feasible* neighbourhood of y^μ.

If we want to extend this idea to *infinite*-dimensional function spaces, *e.g.*, for the problem

$$\min_{y \in Y} f(y) \text{ subject to } y(x) \geq 0 \quad \forall x \in K,$$

then a straightforward idea might be to substitute the above finite sum by an integral. Following this approach, the associated barrier regularization would then read

$$\min_{y \in Y} f(y) - \mu \int_K \log(y(x)) \, \mathrm{d}x.$$

However, we have to take into account two additional issues.

(i) We have to achieve strict feasibility of y^μ in the sense that $y^\mu(x) \geq \varepsilon(\mu) > 0$ for $x \in K$ and for all $\mu > 0$.

(ii) The neighbourhood of y^μ that contains the Newton iterates has to be taken with respect to the $\| \cdot \|_\infty$-norm (or stronger). Any norm that is finite for a pointwise unbounded function yields neighbourhoods that are not contained in the feasible set.

Unfortunately, the logarithmic barrier functional guarantees strict feasibility of y^μ only *almost everywhere* in K, and possibly allows touch points where $y^\mu(x) = 0$ for some $\mu > 0$, even when y^μ is a smooth function. A simple example of such a case has been constructed in Schiela (2009). Of course, in such a case no feasible neighbourhood exists, even in the $\| \cdot \|_\infty$-norm.

A way out of this difficulty is shown by the following lemma.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Lemma 5.1. Let $K \subset \mathbb{R}^d$ be a compact set, satisfying a cone property. Let $y \geq 0$ be in the space $C^\beta(K)$ of Hölder-continuous functions, $1/y^m \in L_1(K)$, $0 < \beta \leq 1$. Assume that $m \geq d/\beta$. Then $1/y \in C(K)$.

Proof. Assume that without loss of generality $0 \in K$ and $y(0) = 0$. We will show that this contradicts the assumption $1/y^m \in L_1(K)$. By the Hölder-continuity of y we infer that $y(x) < cr^\beta$ for all $x \in B(0,r) \cap K$. Using the cone property of K, there is a cone $C_0 \subset K$, and we can compute for sufficiently small $R < 1$

$$\|y^{-m}\|_{L_1(K)} \geq \int_{C_0} y(x)^{-m}\,\mathrm{d}x \geq c\int_{[0,R]} r^{-\beta m} r^{d-1}\,\mathrm{d}r \geq c\int_{[0,R]} r^{-1}\,\mathrm{d}r = \infty,$$

which shows $1/y^m \notin L_1(K)$. Hence, by contradiction, $y > 0$ in K, which implies by compactness of K that there exists $\psi > 0$ with $y \geq \psi$ and thus $1/y \in C(K)$. □

Hence, in order to obtain strict feasibility of barrier minimizers in Hölder spaces, we may use a hierarchy of *higher-order barrier functions*

$$l_m(y) := \begin{cases} -\mu \log(y) & m = 1, \\ \dfrac{\mu^m}{(m-1)y^{m-1}} & m > 1, \end{cases}$$

and solve

$$\min_{y \in C^\beta(K)} f(y) + \int_K l_m(y(x))\,\mathrm{d}x.$$

Then, for sufficiently large m and Hölder-continuous y, the existence of a touch point $y(x) = 0$ would imply

$$\int_K l_m(y(x))\,\mathrm{d}x = \infty,$$

in contradiction to the optimality of y.

BHT equation as operator equation in function space

In order to study the weak form of the BHT equation (3.15) in a functional analytic setting, we have to impose some regularity assumptions. Our set of assumptions has to cover the case of discontinuous coefficients, which arise due to the different types of tissues in the human body.

The presence of state constraints forces us to impose an $\|\cdot\|_\infty$-topology on the space of temperature distributions. For this reason, we have to depart from the usual H^1-framework for elliptic equations, since $H^1(\Omega) \not\subset C(\overline{\Omega})$ for spatial dimension greater than one.

To proceed, let $q > 3$, and define its dual exponent q' via $1/q + 1/q' = 1$. Let $D_q \subset H^1(\Omega)$ be the set of all T, such that $A(T) \in W^{1,q'}(\Omega)^*$, *i.e.*,

$$\langle A(T), \varphi \rangle \leq M\|\varphi\|_{W^{1,q'}} \quad \forall \varphi \in C^\infty(\overline{\Omega}).$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

We have to achieve that the operator

$$A(\cdot) - B(\cdot) : D_q \times \mathbb{R}^{2k_E - 1} \to W^{1,q'}(\Omega)^*$$

is well defined, and D_q can be embedded into $C(\overline{\Omega})$. For a strictly monotone perfusion term $w(T,x)(T - T_{\text{blood}})$, general results from elliptic regularity theory (Haller-Dintelmann, Rehberg, Meyer and Schiela 2009) show that A has a continuous inverse $A^{-1} : W^{1,q'}(\Omega)^* \to C(\overline{\Omega})$, *i.e.*, $D_q \subset C(\overline{\Omega})$, and even $\|T\|_{C^\beta(\overline{\Omega})} \leq c\|A(T)\|_{W^{1,q'}(\Omega)^*}$ for some $\beta > 0$, where $C^\beta(\overline{\Omega})$ is the space of Hölder-continuous functions. Moreover, D_q only depends on the principal part of A, and is thus independent of T.

To ensure that B maps into $W^{1,q'}(\Omega)^*$, we have to impose a technical regularity assumption on the antenna profiles, which requires slightly more regularity than the usual one for Maxwell's equation, namely $E_k \in H(\text{curl}, \Omega)$. We assume that each E_k is contained in $L_q(\Omega, \mathbb{C}^3)$. Then $B(\alpha)$ is contained in $L_{q/2}(\Omega)$, which, in turn, is contained in $W^{1,q'}(\Omega)^*$ by the Sobolev embedding theorem. In this way singularities in the temperature distribution T can be excluded. Such regularity has been shown under certain smoothness assumptions on the tissue interfaces in Druet *et al.* (2011) and Druet (2007). In any case, practical experience shows that such singular temperature distributions do not occur in the hyperthermia setting. In this framework A and B are then twice continuously Fréchet-differentiable as long as the perfusion term is twice differentiable with respect to T. The derivative of A with respect to T,

$$A' : C(\overline{\Omega}) \supset D_q \to W^{1,q'}(\Omega)^*,$$

is not continuous, but it has a continuous inverse. Moreover, since the principal part of A is linear, $A'(T) - A'(\tilde{T})$ does not contain a differential operator. These facts allow us to apply results such as the inverse function theorem to A, and thus to establish continuous differentiability of the control-to-state mapping

$$\alpha \mapsto T(\alpha) := (A^{-1} \circ B)(\alpha).$$

Application of the barrier regularization to the pointwise state constraints

$$T(x) \leq T_{\lim}(x) \quad \forall\, x \in \Omega$$

immediately leads us to identify the above variable y with $T_{\lim}(x) - T(x)$. Consequently, we arrive at the following barrier minimization problem:

$$\begin{aligned} &\min f(T,\alpha) + \int_\Omega l_m(T_{\lim} - T)\,\mathrm{d}x, \\ &\text{subject to} \quad A(T) - B(\alpha) = 0. \end{aligned} \tag{5.2}$$

Our regularity considerations yield $T \in C^\beta(\Omega)$, and since $T_{\lim}$ is constant on each type of tissue, Lemma 5.1 applies separately on each tissue. With these

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

considerations a generalization of the Weierstrass theorem on existence of minimizers yields the following result (for details of the proof see Schiela and Weiser 2010).

Theorem 5.2. There exists $m < \infty$, such that for every $\mu > 0$ the barrier problem (5.2) has an optimal solution, which is strictly feasible with respect to the inequality constraints.

First- and second-order optimality conditions

For a strictly feasible barrier minimizer (T, α), first- and second-order optimality conditions were derived in Schiela and Weiser (2010). To this end, it is convenient to define the Lagrangian function of the problem:

$$L(T, \alpha, \lambda) := f(T, \alpha) + \int_\Omega l_m(T_{\lim} - T)\,\mathrm{d}x + \langle \lambda, A(T) - B(\alpha)\rangle.$$

Here, as $A(T) - B(\alpha) \in W^{1,q'}(\Omega)^*$, λ is a Lagrangian multiplier in $W^{1,q'}(\Omega)$.

Theorem 5.3. (KKT conditions) If (T, α) is a locally optimal strictly feasible solution of (5.2), then there exists a unique Lagrange multiplier $\lambda \in W^{1,q'}(\Omega)$, such that

$$0 = F(T, \alpha, \lambda) := \begin{cases} f_T(T, \alpha) + \displaystyle\int_\Omega l_m'(T_{\lim} - T)\,\mathrm{d}x + A'(T)^*\lambda, \\ f_\alpha(T, \alpha) - B'(\alpha)^*\lambda, \\ A(T) - B(\alpha). \end{cases} \tag{5.3}$$

Of course, F also depends on the parameter μ, since l_m' does so. Since $A'(T)^*$ is a differential operator, the first line of this system is another (linear elliptic) partial differential equation, the *adjoint equation*, and λ is called the *adjoint state*.

Our algorithm will be based on this version of the KKT conditions and can therefore be considered as a *primal* interior point method. Alternatively one could also introduce Lagrangian multipliers for the state constraints, which would lead to *primal–dual* interior point methods. These are known to be more efficient than their primal counterparts for *finite-dimensional* problems. However, the analysis in function space yields that these multipliers can only be represented as Borel measures (Casas 1986), which are very hard to treat numerically. Below, a specific pointwise damping step is presented that recovers some of the benefits of primal–dual methods.

Theorem 5.4. (second-order optimality conditions) Let (T, α, λ) be a solution of (5.3). Moreover, let

$$\mathcal{K} := \ker\left(A'(T), -B'(\alpha)\right) := \{(\delta T, \delta\alpha) : A'(T)\delta T - B'(\alpha)\delta\alpha = 0\}$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

denote the kernel of the linearized constraints, and let

$$H(T,\alpha,\lambda)(\delta T,\delta\alpha)^2 := L_{TT}(T,\alpha,\lambda)\delta T^2 + L_{\alpha\alpha}(T,\alpha,\lambda)\delta\alpha^2 \tag{5.4}$$

denote the Hessian of the Lagrangian.

(i) If (T,α) is a minimizer of (5.2), then $H(T,\alpha,\lambda)$ is positive semi-definite on $\mathcal{K}$.

(ii) $H(T,\alpha,\lambda)$ is positive definite on $\mathcal{K}$ if and only if (T,α) is a local minimizer of (5.2) and J satisfies a local quadratic growth condition. Then, for each $(r_1,r_2,r_3) \in H^1(\Omega)^* \times \mathbb{R}^{2k_E-1} \times W^{1,q'}(\Omega)^*$, the linear system

$$\begin{pmatrix} L_{TT}(T,\alpha,\lambda) & 0 & A'(T)^* \\ 0 & L_{\alpha\alpha}(T,\alpha,\lambda) & -B'(\alpha)^* \\ A'(T) & -B'(\alpha) & 0 \end{pmatrix} \begin{pmatrix} \delta T \\ \delta\alpha \\ \delta\lambda \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \tag{5.5}$$

has a unique solution $(\delta T,\delta\alpha,\delta\lambda) \in D_q \times \mathbb{R}^{2k_E-1} \times H^1(\Omega)$, depending continuously on (r_1,r_2,r_3).

Finally, in

$$L_{TT}(T,\alpha,\lambda)\delta T^2 = f_{TT}(T,\alpha)\delta T^2 + \int_\Omega l_m''(T_{\lim}-T)\delta T^2\,\mathrm{d}x + \langle\lambda, A''(T)\delta T^2\rangle \tag{5.6}$$

we encounter the second derivative l_m'' of the barrier term. This factor can be very large if T is close to its upper bound, and thus induces a strong scaling to be taken into account in solution algorithms.

Central path

The solvability of (5.5) in a neighbourhood of a locally unique solution according to Theorem 5.4 allows us to compute Newton steps for the solution of the KKT system (5.3). Here it is important to obtain $\delta T \in D_q \subset C(\overline{\Omega})$, instead of the weaker result $\delta T \in H^1(\Omega)$. This ensures that Newton's method stays in an L_∞-neighbourhood of the minimizer and yields feasible iterates.

Introducing the shorthand notation $z := (T,\alpha,\lambda)$, we can formulate the following result.

Corollary 5.5. If $H(z)$ is positive definite on $\ker(A',-B')(T,\alpha)$, Newton's method, applied to $F(z)$, converges locally superlinearly to a zero z^μ of F.

Proof. Since F' depends continuously on z, we may just use a qualitative local convergence result for Newton's method (see, *e.g.*, Ortega and Rheinboldt 1970, Theorem 10.2.2). □

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

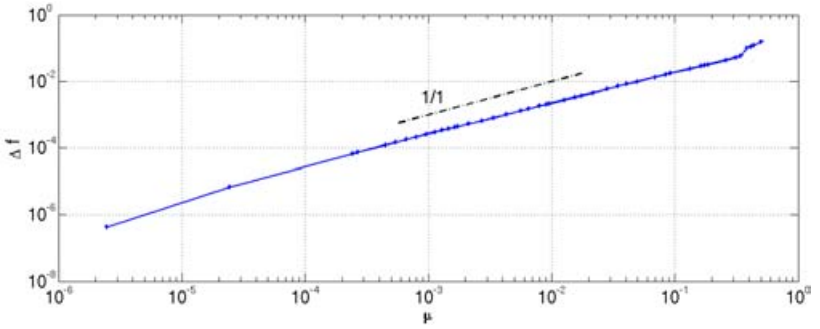


Figure 5.1. Experimental results on the rate of convergence of the function values (see (5.7)) for a hyperthermia planning problem.

In particular, the implicit function theorem yields local existence and differentiability of the so-called *central path* of barrier minimizers.

Corollary 5.6. If $H(z)$ is positive definite on $\ker(A', -B')(T, \alpha)$, then, locally, there is a differentiable path $\mu \mapsto z^\mu$ of local minimizers of the barrier problems, defined in some open interval $]\overline{\mu}, \underline{\mu}[\supset \mu$.

A convergence theory for the path of barrier minimizers has been established for the case of a linear partial differential equation as an equality constraint and a convex functional. In this case, estimates of the form

$$f(T^\mu, \alpha^\mu) - f(T_{\text{opt}}, \alpha_{\text{opt}}) \leq C\mu \tag{5.7}$$

have been established, which implies under a quadratic growth condition on α the estimate

$$\|T^\mu - T_{\text{opt}}\|_{L_\infty(\Omega)} + |\alpha^\mu - \alpha_{\text{opt}}|_{\mathbb{R}^{2k_E-1}} \leq C\sqrt{\mu}.$$

Furthermore, the qualitative result of Corollary 5.5 can be quantified in this case, giving a μ-dependent lower bound on the radius of convergence. Then a global convergence result can be established (Schiela 2008) for a barrier path-following method.

In hyperthermia, the equality constraints are nonlinear, and so our problem is generically non-convex. However, in a neighbourhood of the central path, under second-order sufficient optimality conditions, our barrier problems behave like convex ones, so we can expect these results to carry over. Numerical results give clear evidence for this conjecture, as can be seen in Figure 5.1.

From an algorithmic point of view, the second-order optimality conditions yield well definedness of a Newton path-following method, as presented in the following section, once an initial guess close to the central path is found.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

To find such a point, algorithms have to be used that are robust with respect to non-convexity. For our numerical experiments we employ a new algorithm (see the forthcoming report by Schiela and Weiser 2012) that is based on the idea of a composite step method (see, *e.g.*, Conn, Gould and Toint 2000) with a cubic regularization (Weiser, Deuflhard and Erdmann 2007) of the functional and an affine covariant treatment of the constraints.

5.2. Adaptive central path-following in function space

Section 5.1 laid the groundwork for a Newton path-following method with respect to a barrier parameter μ. The aim of this section is to describe the development of an efficient algorithm based on this principle. We base our algorithmic approach on an affine covariant adaptive path-following method, as described in Deuflhard (2006, Chapter 5). However, compared to the standard case, additional difficulties appear.

- Our problem is posed in function space, and thus Newton steps cannot be computed exactly, which means that an inexact Newton corrector has to be used.
- The radius of convergence of Newton's method tends to zero as $\mu \to 0$. Thus, the subproblems become more and more difficult to solve. This has to be taken into account in an adaptive update strategy for μ.

An algorithm that considers these issues will consist of three loops.

(i) In the outer path-following loop, the barrier parameter μ is reduced successively. This requires an adaptive strategy.

(ii) Each step of the outer loop invokes an inexact Newton method in function space. Here we have to design termination criteria, and criteria on the success of the Newton iteration. An estimate of the *local contraction quantity* $\Theta(z)$ is crucial in this respect.

(iii) In the inner loop each Newton step is computed up to a desired relative accuracy. Typical sources of errors are truncation errors of iterative solvers, but also discretization errors which arise during the approximate computation of Newton steps in function space by finite element discretizations. We have to devise error estimators that are suited to the particular problem structure.

Local convergence of an inexact Newton method

A local affine covariant convergence theory of inexact Newton methods lays the groundwork for our algorithmic considerations. A detailed discussion of this topic can be found in Deuflhard (2006). For the convenience of the reader we give a simplified presentation, based on the analysis of Schiela and Günther (2011).

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

The starting point is the following very simple local convergence result for Newton's method. It is a generalization of the affine covariant refined Newton–Mysovskikh theorem due to Deuflhard and Potra (1992), which can be applied to semi-smooth systems F as well. It is formulated exclusively in terms of quantities in the domain space of the problem, *i.e.*, affine covariant, and covers the inexact computation of Newton steps.

Theorem 5.7. Let R be a linear space, Z a normed space, $D \subset Z$ and $F : D \to R$. Assume there exists $z_* \in D$ with $F(z_*) = 0$. For given $z \in D$ assume that there exists an invertible linear mapping $F'(z)(\cdot) : Z \to R$ such that

$$\Theta(z) := \frac{\left\| F'(z)^{-1}\left[F'(z)(z - z_*) - (F(z) - F(z_*))\right]\right\|}{\|z - z_*\|} \tag{5.8}$$

is well defined. Assume that an inexact Newton step results in

$$z_+ := z - F'(z)^{-1}F(z) + e,$$

where the relative error $\gamma(z) := \|e\|/\|z - z_*\|$ is bounded by

$$\gamma(z) + \Theta(z) \le \beta < 1. \tag{5.9}$$

Then

$$\|z_+ - z_*\| \le \beta\|z - z_*\|. \tag{5.10}$$

Proof. We compute for one inexact Newton step

$$\begin{aligned}\|z - z_*\| &= \|z - F'(z)^{-1}F(z) + e - z_*\| \\ &\le \left\|F'(z)^{-1}\left[F'(z)(z - z_*) - (F(z) - F(z_*))\right]\right\| + \|e\|.\end{aligned}$$

Inserting the definition of $\Theta(z)$ and assumption (5.9), we obtain

$$\|z - z_*\| \le (\Theta(z) + \gamma(z))\|z - z_*\| \le \beta\|z - z_*\|. \qquad \square$$

Equations (5.9) and (5.10) show that the local convergence of Newton's method is governed by the local contraction rate Θ and the relative accuracy γ. As we will show, in an adaptive Newton path-following algorithm, both quantities can be controlled algorithmically.

To be able to react adaptively to the nonlinearity of the problem we have to provide computable estimates for Θ and γ. We will construct an estimate for γ in the paragraph below about adaptive refinement.

Next we derive a computable estimate for $\Theta(z)$. From (5.8) we observe that $\Theta(z)$ cannot be computed exactly unless z_* is known. However, with the help of a simplified Newton step,

$$\overline{\delta z} := -F'(z)^{-1}F(z_+),$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

we can replace the solution z_* in (5.8) by the new iterate z_+, and compute

$$[\Theta(z)] := \frac{\left\|F'(z)^{-1}\left[F'(z)(z-z_+) - (F(z)-F(z_+))\right]\right\|}{\|z-z_+\|} = \frac{\|\overline{\delta z}\|}{\|\delta z\|}. \quad (5.11)$$

In view of local quadratic convergence of Newton's method, we may define a parametrized model for $\Theta(z)$ as follows:

$$[\widetilde{\Theta}(z)] := \frac{[\omega]}{2}\|z-z_*\|.$$

The constant $[\omega]$ can be computed for given z via (5.11):

$$[\omega] := \frac{2[\Theta(z)]}{\|z-z_+\|} = \frac{2\|\overline{\delta z}\|}{\|\delta z\|^2}.$$

Then $[\omega]$ can be interpreted as an estimate from below for an affine covariant Lipschitz constant ω, as used in Deuflhard (2006). This gives an estimate for the radius of contraction, *i.e.*, the radius of the largest ball around x_*, where a contraction $\Theta(z) \le \Theta$ takes place:

$$r_\Theta \ge \frac{2\Theta}{\omega}. \quad (5.12)$$

Replacing ω by our estimate $[\omega]$, we get a quantitative estimate of the size of the region of convergence. Moreover, the remaining error via Newton's method satisfies the triangle inequality,

$$\|z_+ - z_*\| = \Theta(z)\|z-z_*\| \le \Theta(z)(\|z-z_+\| + \|z_+-z_*\|),$$

and thus, solving for $\|z_+-z_*\|$ and replacing $\Theta(z)$ by $[\Theta(z)]$, we obtain the estimate

$$[\|z_+-z_*\|] := \frac{[\Theta(z)]}{1-[\Theta(z)]}\|z-z_+\| \quad (5.13)$$

for the remaining error.

Adaptive reduction of barrier parameter

The geometric picture of a successful path-following method is sketched in Figure 5.2. For some prescribed desired contraction $\Theta_d < 1$ (in practice one chooses $\Theta_d \in [0.1, 0.5]$), each iterate z_j should remain in the region of convergence around a target solution z^{μ_j}, which lies on the central path, *i.e.*,

$$\|z_j - z^{\mu_j}\| \le r_{\Theta_d}(\mu_j). \quad (5.14)$$

We want to achieve this inductively by a suitable adaptive choice of μ_j. First of all, r_{Θ_d} can be estimated via (5.12), if we assume a relation between ω and μ. Numerical experiments, and estimates for related problem classes (Weiser, Gänzler and Schiela 2008), suggest the relation $\omega \sim \mu^{-1/2}$.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

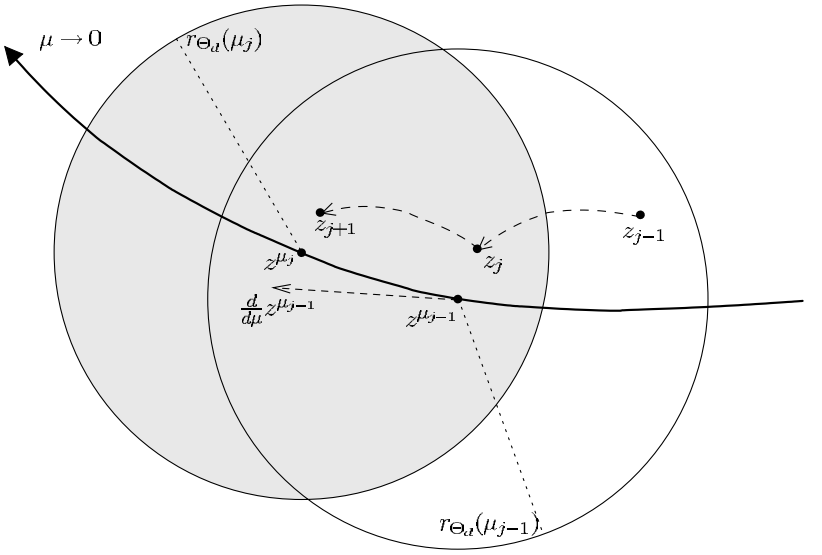


Figure 5.2. Sketch of a Newton path-following method. In each corrector step the iterates approach a target point on the central path and eventually enter the region of convergence of the next target.

To obtain a choice of μ_j that yields (5.14), we also have to provide an estimate for

$$\|z_j - z^{\mu_j}\| \leq \|z_j - z^{\mu_{j-1}}\| + \|z^{\mu_{j-1}} - z^{\mu_j}\|.$$

The first term of the right-hand side is the remaining error of the last Newton corrector and can be estimated via (5.13). The second term depends on the *sensitivity* of z^μ with respect to μ, and we have the estimate

$$\|z^{\mu_{j-1}} - z^{\mu_j}\| = \left\| \int_{\mu_j}^{\mu_{j-1}} \frac{\mathrm{d}}{\mathrm{d}\mu} z^\mu \, \mathrm{d}\mu \right\| \leq \int_{\mu_j}^{\mu_{j-1}} \underbrace{\left\| \frac{\mathrm{d}}{\mathrm{d}\mu} z^\mu \right\|}_{\eta(\mu)} \mathrm{d}\mu. \tag{5.15}$$

For finite-dimensional problems or for fixed discretizations, we may estimate this slope simply by finite differences in the iterates. If, as in our approach, adaptive grid refinement is performed during the corrector, finite differences are not reliable due to the presence of discretization errors. Then, at the cost of one system solve, an estimate $[\eta(\mu_{j-1})]$ can be computed via the formula

$$\frac{\partial}{\partial z} F(z^{\mu_{j-1}}) \frac{\mathrm{d}}{\mathrm{d}\mu} z^{\mu_{j-1}} + \frac{\partial}{\partial \mu} F(z^{\mu_{j-1}}),$$

where we have to replace $z^{\mu_{j-1}}$ by its approximation z_j. Then, using the

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

relation $\eta(\mu) \sim \mu^{-1/2}$, we may compute the estimate

$$[\|z^{\mu_{j-1}} - z^{\mu_j}\|] := \int_{\mu_j}^{\mu_{j-1}} [\eta(\mu_{j-1})]\sqrt{\mu_{j-1}/\mu}\,\mathrm{d}\mu = 2[\eta(\mu_{j-1})]\sqrt{\mu_{j-1}}(\sqrt{\mu_{j-1}} - \sqrt{\mu_j}).$$

To summarize, we may write down (5.14) as an implicit inequality in μ_j:

$$[\|z_j - z^{\mu_{j-1}}\|] + 2[\eta(\mu_{j-1})]\sqrt{\mu_{j-1}}(\sqrt{\mu_{j-1}} - \sqrt{\mu_j}) \le \frac{2\Theta_d}{[\omega(\mu_{j-1})]}\sqrt{\frac{\mu_{j-1}}{\mu_j}}, \quad (5.16)$$

and compute the smallest μ that satisfies this inequality as a prediction for μ_j. In the subsequent inexact Newton corrector the estimated error is reduced up to a tolerance $\mathrm{TOL}_{\mathrm{corr}}$, which is chosen in such a way that usually one corrector step is sufficient, so that we terminate if

$$[\|z_j - z^{\mu_{j-1}}\|] \le \mathrm{TOL}_{\mathrm{corr}} := \rho\frac{2\Theta_d}{[\omega(\mu_j)]}, \quad \rho \in [1, 1.2].$$

This rule suggests a *multiplicative* update strategy. Define $\sigma_j := \mu_j/\mu_{j-1}$. Then (5.16) can be written as an implicit inequality in σ_j:

$$[\|z_j - z^{\mu_{j-1}}\|][\omega(\mu_{j-1})] + 2[\eta(\mu_{j-1})][\omega(\mu_{j-1})]\mu_{j-1}(1 - \sqrt{\sigma_j}) \le \frac{2\Theta_d}{\sqrt{\sigma_j}}. \quad (5.17)$$

Under our model assumptions $[\omega] \sim \mu^{-1/2}$, $[\eta] \sim \mu^{-1/2}$, which imply $\mathrm{TOL}_{\mathrm{corr}} \sim \sqrt{\mu}$, we observe that μ_{j-1} cancels out in (5.17), and thus σ is constant. Of course our model assumptions only hold approximately in computational practice, but their use leads to an unbiased and strictly positive update $\mu_{j-1} \to \mu_j > 0$.

If it terminates successfully, we accept the result of the corrector as the new iterate z_j, and compute μ_{j+1} as the new homotopy parameter, just as before. In the case of unsuccessful termination (for example, if $[\Theta(z)] > 1$ for some iterate z), with the additional knowledge of $[\omega]$ gained in this correction step, a more conservative choice for μ_j can be computed and the corrector is restarted at z_{j-1}.

Local norms for barrier subproblems

Up to now our considerations have not included any special structure of our problem. One way to do this is to choose a norm specific to the problem. In the context of barrier methods the strong scaling property of the Hessian, as observed in (5.6), should be taken into account. Thus (for fixed μ) a suitable local norm can be defined as follows:

$$\|\delta z\|_z^2 := \|\delta T\|_{L_2(\Omega)}^2 + \int_\Omega l_m''(T_{\mathrm{lim}} - T)\delta T^2\,\mathrm{d}x + |\delta\alpha|_{\mathbb{R}^{2k_E-1}}^2 + \|\delta\lambda\|_{L_2(\Omega)}^2. \quad (5.18)$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Qualitatively speaking, the inclusion of the barrier function l_m into the norm has the effect that errors are given more weight in regions of Ω where T is close to its upper bound $T_{\lim}$. Thereby it is taken into account that the problem is more nonlinear and thus more sensitive to errors in these regions.

In practical computations, the multiplier term $\|\delta\lambda\|^2_{L_2(\Omega)}$ can be omitted. The reason is that λ can be reconstructed approximately from the primal variables T, α by suitable multiplier update rules (Conn, Gould and Toint 2000) and plays the role of an auxiliary variable.

In a neighbourhood of the central path, where positive definiteness of the Hessian H of the Lagrangian from (5.4) can be assumed on $\ker(A', -B')$, the generic L_2-norms of δT and $\delta\alpha$ in (5.18) can be substituted by the norm induced by H for the tangential components, which dominate during the homotopy due to the mild nonlinearity of the state equation. This leads to the affine conjugate norm

$$\|\delta z\|^2_{H(z)} := H(z)(\delta T, \delta\alpha)^2. \tag{5.19}$$

The normal components can also be taken into account if desired: see, *e.g.*, Volkwein and Weiser (2002).

As the local norm changes in every Newton step, contraction $\Theta < 1$ for a fixed norm need not imply contraction between different local norms. Whether this is guaranteed depends on the actual impact of z on $\|\cdot\|_z$. For details, we refer to Deuflhard (2006) and Volkwein and Weiser (2002).

Pointwise damping

Barrier methods rely on iterates that are feasible with respect to the inequality constraints. Since in the barrier context Newton's method approximates a rational function by a linear one, Newton steps tend to violate the constraints. On the other hand, the nonlinearity, introduced by the barrier functions, is given in a pointwise fashion and highly structured. This should be exploited.

In the following we propose a modification of Newton's method, which addresses these issues and may be considered as a pointwise damping strategy. The idea exploits the pointwise structure of the problem and guarantees strict feasibility of the iterates. In the whole discussion $\mu > 0$ is fixed.

For what follows, define the Lagrangian function of the unconstrained problem:

$$\tilde{L}(z) := f(T, \alpha) + \langle \lambda, A(T) - B(\alpha) \rangle.$$

Further, let $z_+ = (T_+, \alpha_+, \lambda_+)$ be the possibly infeasible next Newton iterate, which satisfies

$$F(z) + F'(z)(z_+ - z) = 0. \tag{5.20}$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

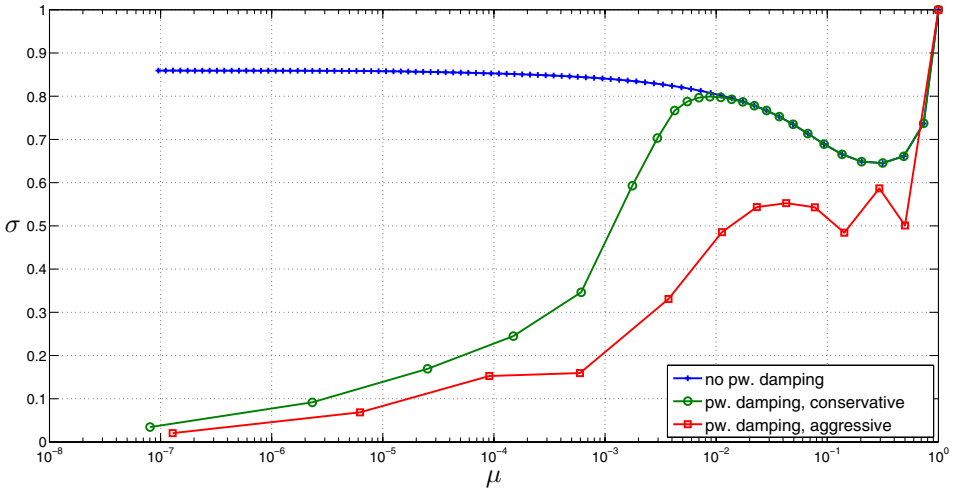


Figure 5.3. Comparison of path-following algorithm with and without pointwise damping. Horizontal axis: barrier parameter μ. Vertical axis: reduction factor $\sigma := \mu_+/\mu$. For pointwise damping one can observe $\sigma \to 0$ for $\mu \to 0$.

The main idea is to construct a modified *feasible* iterate $z_C = (T_C, \alpha_+, \lambda_+)$ that satisfies the first row of (5.3), *i.e.*,

$$\tilde{L}_T(z_C) + l'_m(T_{\lim} - T_C) + A'(T_C)^*\lambda_+ = 0. \tag{5.21}$$

This is an equation in $(H^1)^*$ which seems to be difficult to solve for T_C. However, if we subtract the first row of (5.20) from (5.21) we observe that the temperature-independent differential operators in $A'(\cdot)^*$ cancel out, and we obtain a *pointwise* equation in T_C that depends on z and z_+. If we neglect those higher-order terms that are independent of l_m, we end up with a pointwise equation of the form

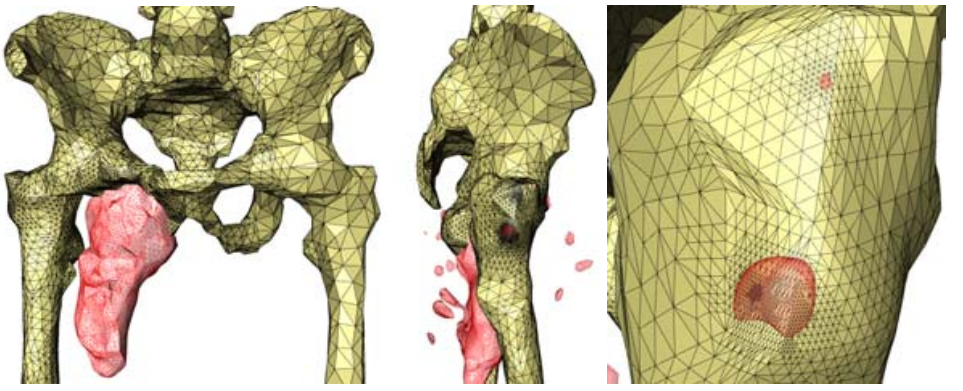
$$\tilde{L}_{TT}(z)T_C + l'_m(T_{\lim} - T_C) = r(z, z_+).$$

This equation can be solved for T_C at each node of the discretization. For scalar equations there are efficient and robust algorithms available (*e.g.*, Brent 1971), and the computational cost of performing this step is marginal compared to the overall computation time.

To obtain a pointwise damping step we use only those $T_C = T_C(x)$ for which $|T_C - T| \le |T_+ - T|$; these are exactly those for which $T_+ \ge T_C \ge T$ holds. Hence, we are able to compute a *strictly* feasible corrected iterate T_C from a possibly infeasible iterate T_+ in a *natural way*. It can be shown that

$$\lim_{\|T - T^\mu\|_\infty \to 0} \frac{\|T_C - T_+\|_\infty}{\|T - T_+\|_\infty} = 0.$$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b) (c)

Figure 5.4. Bone and tumour surfaces with adaptively refined grid. (a) Frontal view. (b) Sagittal view. (c) Zoom into upper femur with 43°C isosurface; adaptive refinement occurs near hot spots.

This implies that undamped steps are recovered asymptotically close to a strictly feasible solution. An estimate for the Newton contraction $\Theta(z)$ can be computed in analogy to the case described above.

Our pointwise damping strategy has a marked effect on the numerical performance of the path-following method. As Figure 5.3 indicates, an algorithm without pointwise damping shows relatively slow linear convergence behaviour for an academic test problem. In contrast, with pointwise damping the reduction of μ often takes place at a superlinear rate. This is particularly beneficial in the context of our hyperthermia problem.

Discretization and adaptive grid refinement

So far we have assumed that our algorithm can compute Newton steps in function space exactly. However, this is not realistic because the Newton steps have to be discretized for computation. Discretization of the temperature T and the Lagrange multiplier λ is done straightforwardly by linear finite elements, and the precomputed electric fields are discretized using edge elements of first order, as described in Section 4.

As an additional numerical difficulty the barrier integrals and their derivatives cannot be evaluated exactly. However, it can be shown that the use of the *trapezoidal rule* is sufficient to obtain satisfactory *a priori* error estimates (Hinze and Schiela 2011).

For *a posteriori* error estimation and adaptive grid refinement we rely once more on hierarchical error estimators, as presented in Section 4.1.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Discretizing (5.5) with an extended ansatz space, we obtain a block linear system of the form (4.5)

$$\begin{pmatrix} F'(z)^{LL} & F'(z)^{LH} \\ F'(z)^{HL} & F'(z)^{HH} \end{pmatrix} \begin{pmatrix} \delta z^L \\ \delta z^H \end{pmatrix} = - \begin{pmatrix} F(z)^L \\ F(z)^H \end{pmatrix}. \tag{5.22}$$

The only difference is that each block of this system is in turn a block matrix of the form (5.5). Finally, as in Section 4.1, we end up with the problem of solving the block linear system

$$F'(z)^{HH}\delta z^H = r^H := -(F(z)^H + F'(z)^{HL}\delta z^L). \tag{5.23}$$

Once δz^H is computed, one can compute cheaply an estimate in the affine conjugate norm defined in (5.19). If $\delta z^H = (\delta T^H, \delta\alpha^H, \delta\lambda^H)$, define $\delta\tilde{z}^H := (\delta T^H, \delta\alpha^H, -\delta\lambda^H)$. Then one can compute via (5.5) that

$$[\|e\|_{H(z)}]^2 := \langle \delta\tilde{z}^H, r^H\rangle = \langle \delta\tilde{z}^H, F'(z)^{HH}\delta z^H\rangle = \|\delta z^H\|^2_{H(z)},$$

because the off-diagonal blocks of $F'(z)^{HH}$ cancel out. Hence, we can estimate the error in the same norm that is used to measure the length of the Newton steps.

Now we can use these estimates to *control* the relative error

$$[\gamma(z)] := \frac{\langle \delta z^H, \tilde{r}^H\rangle^{1/2}}{\|\delta z^L\|_{H(z)}}$$

by adaptive mesh refinement. The refinement loop is stopped as soon as the estimate for $[\gamma(z)]$ drops below a given relative tolerance $\mathrm{TOL}_{\mathrm{disc}} \in [0.1, 0.5]$. A closely related adaptive refinement scheme results from goal-oriented error estimation for optimal control problems, considering the error in the cost functional (Becker, Kapp and Rannacher 2000, Weiser 2012).

In the context of Theorem 5.7 it can be seen that the corresponding inexact Newton iteration will converge asymptotically linearly in function space. A superlinear or quadratic convergence rate is not aimed for, since this usually leads to unnecessarily high accuracy and a corresponding dramatic increase of unknowns, impeding the efficiency of the algorithm.

Solution of linear systems

Finally, we have to solve discretized versions of the system (5.5) in order to compute the discretized Newton steps, and also to solve the defect equation (5.23).

This can be done by reducing it to a linear system in the control $\alpha \in \mathbb{R}^{2k_E-1}$ via a Schur complement approach. To obtain the reduced linear system, each of the discretized PDEs $A'(T)v = r$ and $A'(T)^*w = s$ has to be solved for $2k_E - 1$ right-hand sides. For a wide range of discretizations a sparse direct solver (*e.g.*, MUMPS: Amestoy *et al.* 2011) performs quite

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

efficiently. Alternatively, one can solve the PDEs iteratively up to a certain accuracy by a preconditioned conjugate gradient method and perform an outer iteration to obtain a solution of the total system.

In particular, when solving (5.23) approximately, the solution of the PDEs $A'(x)v = r$ and its adjoint can be found iteratively via a Jacobi preconditioned conjugate gradient method.

Putting together all the pieces described above, we now arrive at the following pseudo-code.

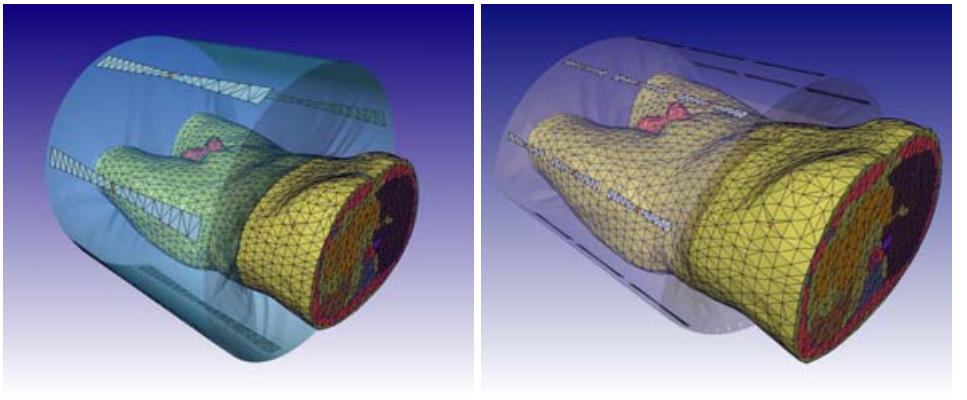
algorithm `AdaptivePathfollowing`(z_0, μ_0) (*Initial Guess*)

for $j = 1, \ldots$ (*Path-following Loop*)
  $(\widetilde{z}; \text{success}) \leftarrow$ `InexactNewtonCorrector`$(z_{j-1}; \mu_{j-1})$
  if success
    compute new $\mu_j < \mu_{j-1}$
    $z_j \leftarrow \widetilde{z}$
  else
    compute corrected $\mu_j > \mu_{j-1}$
    $z_j \leftarrow z_{j-1}$

subroutine `InexactNewtonCorrector`$(z; \mu)$:

do (*Newton Iteration Loop*)
  do (*Inexact Newton Step*)
    RefineMarkedElements
    $z_+ := z - F_h'(z)^{-1} F_h(z)$
    $[\|e\|_{H(z)}] :=$ ErrorEstimator:Discretization(z)
    MarkElements
  while $[\|e\|_{H(z)}] > \text{TOL}_{\text{disc}} \|z_+ - z\|$
  compute a pointwise damped iterate z_C
  compute $[\Theta(z)]$
  compute $[\|z_C - z^\mu\|_z]$
  success $= ([\|z_C - z^\mu\|_z] \leq \text{TOL}_{\text{corr}})$
  failure $= ([\Theta(z)] >$ RequiredContraction$)$
  $z \leftarrow z_C$
while not(success $\vee$ failure)
return$(z; \text{success})$

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 6.1. Applicator design. (a) Old applicator, Sigma-60, with 8 antennas ($k_E = 4$) and circular cross-section; see also Figure 2.1(b). (b) New applicator, Sigma-Eye, with 24 antennas ($k_E = 12$) and eye-shaped cross-section, designed on the basis of our simulations.

6. Impact on medical technology

The earlier versions of our virtual lab HyperPlan have already had significant impact on treatment via deep regional hyperthermia, part of which we will present here.

6.1. Redesign of applicator

During the past ten years the applicator most frequently used for regional hyperthermia had been the Sigma-60 applicator of BSD Medical Corporation, Salt Lake City, Utah:[3] see Figure 6.1(a). This applicator consists of 8 antennas (*i.e.*, $k_E = 4$ channels) arranged on a ring with a diameter of 60 cm, hence its name. The space between the antenna ring and the patient's body (see Figure 2.1(a) for an actual photo and Figure 6.1(a) for a computer image) is filled with a *water bolus* containing de-ionized water. In early 1988, based on the results of our first simulations, the new Sigma-Eye applicator (see Figure 6.1(b)), was introduced. It has 24 antennas (*i.e.*, $k_E = 12$ channels) arranged in three parallel rings, thus allowing for an additional power steering along the patient's axis. The much smaller bolus volume has an 'eye'-shaped cross-section, hence the name.

In order to illustrate the progress, let us document some comparative results. Table 6.1 lists the computation times for the whole simulation process (on a SUN UltraSparc), based on the linear BHT as heat transfer model.

[3] Now in Munich, Germany.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Table 6.1. Coarse grid computation times (SUN UltraSparc) for the whole simulation process with linear BHT equation. From Deuflhard and Seebass (1999).

	Sigma-60 $k_E = 4$	Sigma-Eye $k_E = 12$
Grid generation	15 min	
Field calculation	80 min	120 min
Temperature calculation	2 min	20 min
Optimization	6 sec	1 min

Table 6.2. Comparison of old applicator Sigma-60 and the new applicator Sigma-Eye for three patients with different tumour locations. From Deuflhard and Seebass (1999).

	Part of tumour volume heated to above 43°C	
	Sigma-60	Sigma-Eye
Distal (supra-anal) rectal carcinoma	17.5%	62.5%
Highly presacral rectal carcinoma	0.7%	18.4%
Cervical carcinoma at pelvic wall	24.8%	49.1%

Note that, essentially due to its smaller bolus volume, the field calculation time per channel of the Sigma-Eye is only about half that of the Sigma-60. Moreover, recall from the cost model (4.16) that the temperature calculation times roughly increase with k_E^2, whereas the Maxwell solves need $O(k_E)$.

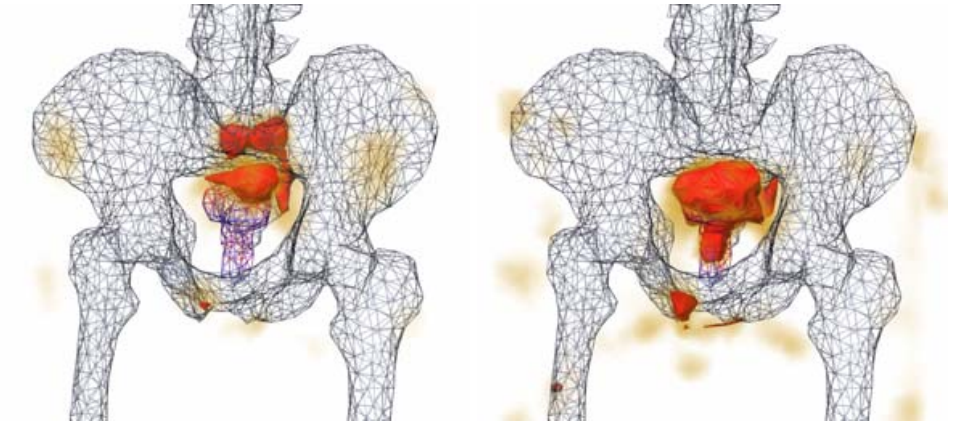
In Table 6.2, the progress in hyperthermia is documented at three patients with typical tumour locations. The new applicator has significantly improved tumour heating, keeping 'hot spots' in healthy tissue at a constant level or slightly reducing them – and not merely for these cases. Incidentally, the second location is the most difficult one, since the tumour is to a large extent enclosed by bony structures.

In Figure 6.2, we illustrate the progress in the resulting therapy with visualizations of the optimal solution in both cases.

Frequency study

Initial frequency studies were performed by Paulsen *et al.* (1999) for 100, 150, and 200 MHz, but only for power functionals such as (3.21); for their applicator test cases, they came up with an 'optimal' frequency of about

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

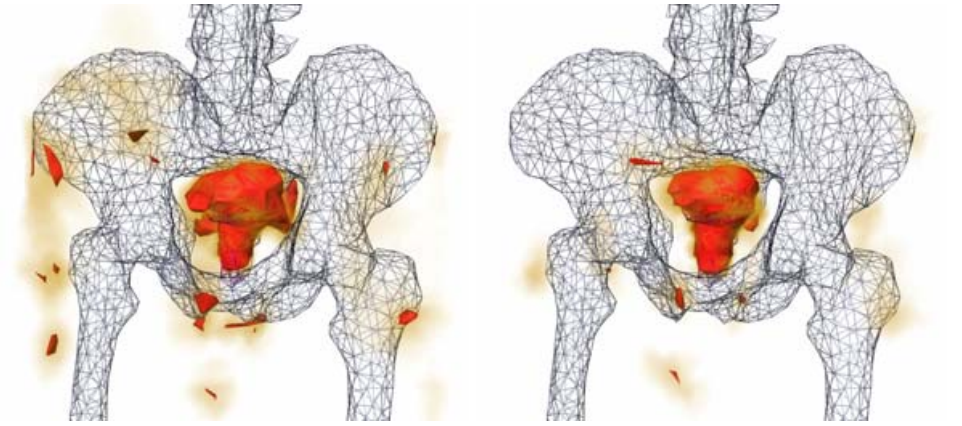
Figure 6.2. Optimal temperature distribution for a rectal carcinoma at a frequency of 100 MHz. (a) Old applicator (Sigma-60). (b) New applicator (Sigma-Eye).

150 MHz. In Figure 6.3 we show our results for the new applicator in Figure 6.2(b). For all simulations, in various patient cases, a frequency of around 150 MHz also appeared to be preferable.

6.2. Case study: femoral tumour

Finally, in order to give an impression of the computational impact of the new optimization techniques, we present a case study. The anonymous patient data refer to a femoral tumour: see Figure 6.4(a). Of the various objective functions from Section 3.4 we selected the Arrhenius functional (3.24) inside the tumour. Outside the tumour, in healthy tissue, we imposed the strict restriction $T \leq 44°\mathrm{C}$, and a mild quadratic penalty term, if tissue were heated above 41°C. To treat the uncertain tissue around the tumour, we surrounded the tumour with a region of width just one coarse mesh size: for this region we replaced the constraints with a strong quadratic penalty term, thus smoothing the temperature profiles to be computed. The $2k_E$ electrical fields were computed off-line once; they are a property of the given applicator. In order to satisfy the temperature constraints in healthy tissue, we applied the barrier algorithm from Section 5. The intermediate values of the barrier parameter μ are automatically determined from adaptive control, as described in Section 5.2. In the present example, the path-following method started at $\mu = 0.5$ and took the number of μ-steps documented in Table 6.3 down to $\mu = 10^{-6}$. The state shown in Figure 6.4(e) was reached after 7 μ-steps. The remaining 39 steps have been computed merely for

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press



(a) (b)

Figure 6.3. Optimal temperature distribution in new applicator for a rectal carcinoma at frequencies different from the one in Figure 6.2(b). (a) 150 Mhz. (b) 200 MHz.

reference purposes. In practice one would of course employ an automatic termination criterion.

Numerical central path-following

The computational homotopy chain for $\mu \to 0$ started at $\mu = 0.5$ and a conservative solution (no 43°C isothermal surface active): see again Figure 6.4(a). Upon examination of the other images in Figure 6.4, one observes that intermediate steps may well supply interesting planning options, depending on additional medical information to be taken into account, which opens up new possibilities for the medical staff in charge of the hyperthermia treatment. For large μ, the thermal dose is lower than for smaller values. For ever smaller μ, the appearance of hot spots becomes clearly visible. However, due to the barrier's regularization, these regions are still below the prescribed upper bounds of 44°C. Hence, some individual weighting of the occurrence of hot spots versus a better covering of the tumour by the temperature isosurface may be employed.

Mesh history

In contrast to the old penalty approach, the new approach via the PDE-constrained optimization solves the full nonlinear problem including adaptive mesh refinement within one algorithmic sweep. The geometry of the tissue regions is adjusted during mesh refinement. For practical reasons, the maximal number of degrees of freedom had to be restricted. Table 6.3 gives details of three refinement levels at different barrier parameters μ.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

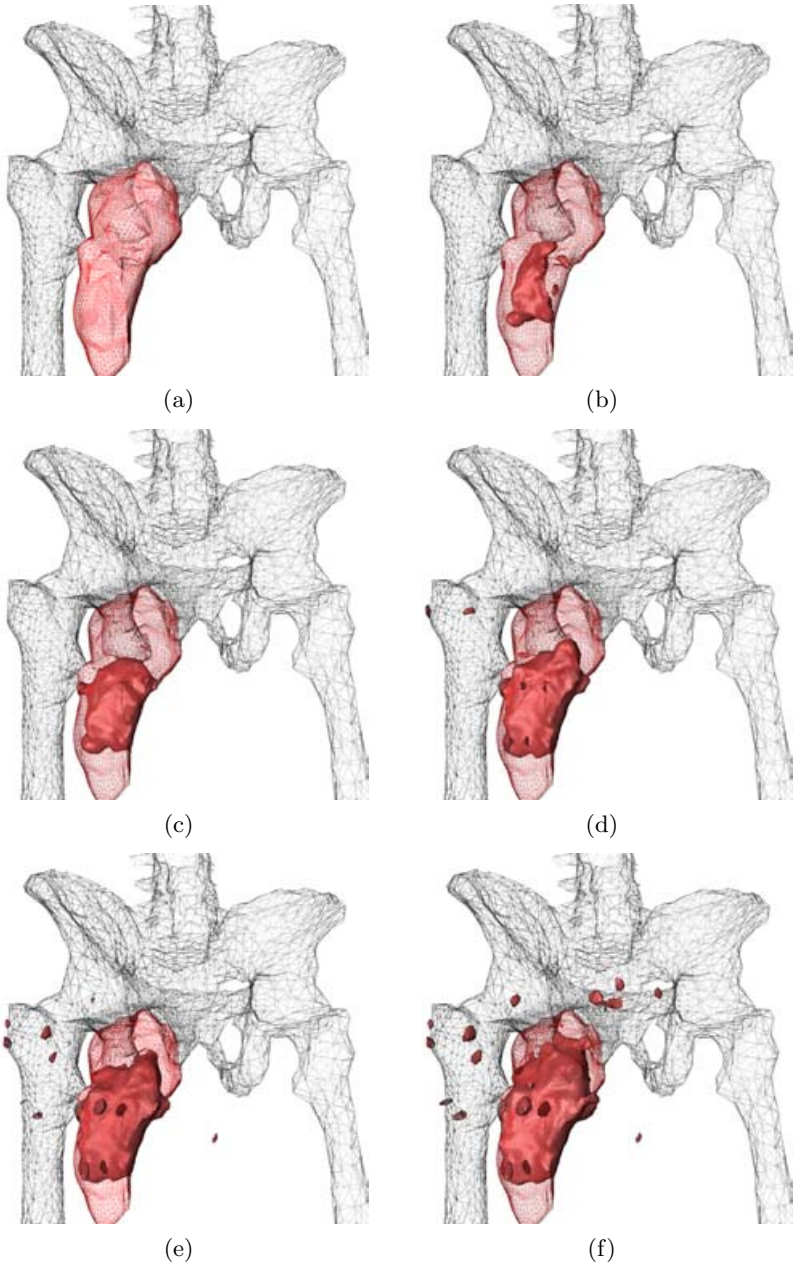


Figure 6.4. Femoral tumour: temperature distributions along the numerical central path. *Light grey grid*, bone; *medium grey grid*, tumour; *dark grey surface*, isothermal surface for 43°C. (a) Just bone and tumour (identical to $\mu = 0.5$), (b) $\mu = 0.44$, (c) $\mu = 0.42$, (d) $\mu = 0.38$, (e) $\mu = 0.26$, (f) $\mu = 10^{-6}$. Obviously, the difference between (f) and (e) is marginal.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

Table 6.3. Computational effort spent on the therapy planning problem for the femoral tumour. The first line represents a computation on the coarse grid.

Tetrahedra	Unknowns	μ-steps
74 457	28 691	39
617 178	222 972	46
3 658 357	1 275 057	55

Acknowledgements

The authors gratefully acknowledge many years of funding by the Deutsche Forschungsgemeinschaft: in the collaborative research unit SFB 273 (1996–2002), in a series of individual projects (WU 235/1-1, WU 235/1-2, WE 2937/2-2, WU 235/3-1) and finally within the DFG Research Center MATHEON (in the first two periods, 2002–2010). For the present paper, they wish to thank the ZIB members Stefan Zachow and Hans Lamecker for helpful discussions and support concerning the mathematical visualization, Lin Zschiedrich for intensive discussions on the electrodynamical part, and Lars Lubkoll for computational assistance. Finally, the authors wish to thank Charlie Elliott, who happened to be on our campus as an Alexander-von-Humboldt awardee, for his careful reading (which he called a 'superficial reading').

REFERENCES

R. Adams and L. Bischof (1994), 'Seeded region growing', *IEEE Trans. Pattern Analysis and Machine Intelligence* **16**, 641–647.

G. Allaire, A. Damlamian and U. Hornung (1995), Two-scale convergence on periodic surfaces and applications. In *Mathematical Modelling of Flow through Porous Media* (A. Bourgeat *et al.*, eds), World Scientific, chapter 38.

P. R. Amestoy, A. Buttari, I. S. Duff, A. Guermouche, J.-Y. L'Excellent and B. Uçar (2011), MUMPS. In *Encyclopedia of Parallel Computing* (D. Padua, ed.), Springer.

I. Babuška and A. Aziz (1972), Survey lectures on the mathematical foundations of the finite element method. In *Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations* (A. Aziz, ed.), Academic Press, pp. 1–359.

I. Babuška and W. Rheinboldt (1978), 'Error estimates for adaptive finite element computations', *SIAM J. Numer. Anal.* **15**, 736–754.

R. Bank (1985), PLTMG User's Guide. University of California, San Diego.

R. Beck (1993), Feldberechnung in dreidimensionalen Leitungsstrukturen der Mikroelektronik mittels p-adaptiver Finite-Elemente-Methoden. PhD thesis, Fachbereich Elektrotechnik, Technische Universität Berlin.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

R. Beck and R. Hiptmair (1999), 'Multilevel solution of the time-harmonic Maxwell's equations based on edge elements', *Int. J. Numer. Methods Engng* **45**, 901–920.

R. Beck, P. Deuflhard, H.-C. Hege and D. Stalling (1997), Numerical algorithms and visualization in medical treatment planning. In *Mathematical Visualization: Experiments, Simulations and Environments* (H.-C. Hege and K. Polthier, eds), Springer International.

R. Beck, P. Deuflhard, R. Hiptmair, R. Hoppe and B. Wohlmuth (1999), 'Adaptive multilevel methods for edge element discretizations of Maxwell's equations', *Surveys Math. Indust.* **8**, 271–312.

R. Becker, H. Kapp and R. Rannacher (2000), 'Adaptive finite element methods for optimal control of partial differential equations: Basic concepts', *SIAM J. Control Optim.* **39**, 113–132.

A. Bensoussan, J.-L. Lions and G. Papanicolau (1978), *Asymptotic Analysis for Periodic Structures*, Vol. 5 of *Studies in Mathematics and Applications*, North-Holland.

F. A. Bornemann, B. Erdmann and R. Kornhuber, R. (1993), 'Adaptive multilevel methods in three space dimensions', *Int. J. Numer. Meth. Engng* **36**, 3187–3203.

A. Bossavit (1988), 'Whitney forms: A class of finite elements for three-dimensional computation in electromagnetism', *Inst. Electr. Eng. Processing* A **135**, 493–500.

A. Bossavit (1990), 'Solving Maxwell's equation in a closed cavity and the question of spurious modes', *IEEE Trans. Magn.* **26**, 702–705.

J. Bramble, D. Kwak and J. Pasciak (1994), 'Uniform convergence of multigrid V-cycle iterations for indefinite and nonsymmetric problems', *SIAM J. Numer. Anal.* **31**, 1746–1763.

J. H. Bramble, J. E. Pasciak and J. Xu (1990), 'Parallel multilevel preconditioners', *Math. Comput.* **55**, 1–22.

R. P. Brent (1971), 'An algorithm with guaranteed convergence for finding a zero of a function', *Comput. J.* **14**, 422–425.

D. Brizel, S. Scully, J. Harrelson, L. Layfield *et al.* (1996), 'Radiation therapy and hyperthermia improve the oxygenation of human soft tissue sarcomas', *Cancer Res.* **56**, 5347–5350.

E. Casas (1986), 'Control of an elliptic problem with pointwise state constraints.', *SIAM J. Control Optim.* **24**, 1309–1318.

A. Conn, N. Gould and P. Toint (2000), *Trust-Region Methods*, SIAM.

T. Cootes, A. Hill, C. Taylor and J. Haslam (1994), 'Use of active shape models for locating structures in medical images', *Image and Vision Computing* **12**, 355–365.

C. D'Angelo and A. Quarteroni (2007), 'On the coupling of 1D and 3D diffusion–reaction equations: Application to tissue perfusion problems', *Math. Mod. Meth. Appl. Sci.* **18**, 1481–1504.

T. Davis (2006), *Direct Methods for Sparse Linear Systems*, SIAM.

K. Deckelnick and M. Hinze (2007), 'Convergence of a finite element approximation to a state-constrained elliptic control problem', *SIAM J. Numer. Anal.* **45**, 1937–1953.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

P. Deuflhard (2000), Differential equations in technology and medicine: Computational concepts, adaptive algorithms, and virtual labs. In *Computational Mathematics Driven by Industrial Problems* (R. Burkard, P. Deuflhard, A. Jameson, J.-L. Lions and G. Strang, eds), Springer, pp. 69–125.

P. Deuflhard (2006), *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, second edn, Springer.

P. Deuflhard and R. Hochmuth (2004), 'Multiscale analysis of thermoregulation in the human microvascular system', *Math. Meth. Appl. Sci.* **27**, 971–989.

P. Deuflhard and F. Potra (1992), 'Asymptotic mesh independence of Newton–Galerkin methods via a refined Mysovskii theorem', *SIAM J. Numer. Anal.* **29**, 1395–1412.

P. Deuflhard and M. Seebass (1999), Adaptive multilevel FEM as decisive tools in the clinical cancer therapy hyperthermia. In *Domain Decomposition Methods in Sciences and Engineering* (C.-H. Lai, P. E. Bjørstad, M. Cross and O. B. Widlund, eds), Domain Decomposition Press, Bergen, pp. 403–414.

P. Deuflhard and M. Weiser (2012), *Adaptive Numerical Solution of Partial Differential Equations*, de Gruyter textbook.

P. Deuflhard, P. Leinen and H. Yserentant (1989), 'Concepts of an adaptive hierarchical finite element code', *Impact Comp. Sci. Engng* **1**, 3–35.

P. Deuflhard, M. Weiser and M. Seebass (2000), 'A new nonlinear elliptic multilevel FEM applied to regional hyperthermia', *Comput. Vis. Sci.* **3**, 115–120.

W. Dewey (1994), 'Arrhenius relationships from the molecule and cell to the clinic', *Int. J. Hyperthermia* **10**, 457–483.

P.-É. Druet (2007), Higher integrability of the Lorentz force for weak solutions to Maxwell's equations in complex geometries. WIAS Preprint 1270, Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS).

P.-E. Druet, O. Klein, J. Sprekels, F. Tröltzsch and I. Yousept (2011), 'Optimal control of three-dimensional state-constrained induction heating problems with nonlocal radiation effects', *SIAM J. Control Optim.* **49**, 1707–1736.

B. Erdmann, J. Lang and M. Seebass (1997), Optimization of temperature distributions for regional hyperthermia based on a nonlinear heat transfer model. Technical report SC-97-59, Zuse Institute Berlin.

R. W. Freund (1992), 'Conjugate gradient-type methods for linear systems with complex symmetric coefficient matrices', *SIAM J. Sci. Statist. Comput.* **13**, 425–448.

S. Gabriel, R. Lau and C. Gabriel (1996), 'The dielectric properties of biological tissues III: Parametric models for the dielectric spectrum of tissues', *Phys. Med. Biol.* **41**, 2271–2293.

S. Gerber (2007), Perfusionsmodellierung in menschlichen Tumoren. Master's thesis, Freie Universität Berlin.

T. Gänzler, S. Volkwein and M. Weiser (2006), 'SQP methods for parameter identification problem arising in hyperthermia', *Optim. Meth. Software* **21**, 869–887.

S. Götschel, M. Weiser and A. Schiela (2012), Solving optimal control problems with the Kaskade 7 finite element toolbox. In *Advances in DUNE* (A. Dedner, B. Flemisch and R. Klöfkorn, eds), Springer, to appear 2012.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

W. Hackbusch (1985), *Multi-Grid Methods and Applications*, Vol. 4 of *Computational Mathematics*, Springer.

R. Haller-Dintelmann, J. Rehberg, C. Meyer and A. Schiela (2009), 'Hölder continuity and optimal control for nonsmooth elliptic problems', *Appl. Math. Optim.* **60**, 397–428.

P. Heckbert and M. Garland (1997), Survey of polygonal surface simplification algorithms. In *Siggraph 97 Course Notes*, Vol. 25, ACM Press.

H.-C. Hege and D. Stalling (1998), Fast LIC with piecewise polynomial filter kernels. In *Mathematical Visualization: Algorithms and Applications* (H.-C. Hege and K. Polthier, eds), Springer International.

M. Hintermüller and K. Kunisch (2006), 'Feasible and non-interior path-following in constrained minimization with low multiplier regularity', *SIAM J. Control Optim.* **45**, 1198–1221.

M. Hinze and A. Schiela (2011), 'Discretization of interior point methods for state constrained elliptic optimal control problems: Optimal error estimates and parameter adjustment', *Comput. Optim. Appl.* **48**, 581–600.

R. Hiptmair (1996), Multilevel preconditioning for mixed problems in three dimensions. Master's thesis, University of Augsburg, Mathematics Institute.

R. Hiptmair (1999), 'Multigrid method for Maxwell's equations', *SIAM J. Numer. Anal.* **36**, 204–225.

R. Hiptmair (2002*a*), Finite elements in computational electromagnetism. In *Acta Numerica*, Vol. 11, Cambridge University Press, pp. 237–341.

R. Hiptmair (2002*b*), Numerical methods for computational electromagnetism. Master's thesis, University of Augsburg, Mathematics Institute.

R. Hiptmair and J. Xu (2007), 'Nodal auxiliary space preconditioning in $H(\mathrm{curl})$ and $H(\mathrm{div})$ spaces', *SIAM J. Numer. Anal.* **45**, 2483–2509.

R. Hochmuth and P. Deuflhard (2004), 'Multiscale analysis for the bio-heat transfer equation: The nonisolated case', *Models Meth. Appl. Sci.* **14**, 1621–1634.

R. D. Issels, M. Wadepohl, K. Tiling, M. Müller, H. Sauer and W. Wilmanns (1987), 'Regional hyperthermia combined with systemic chemotherapy in advanced abdominal and pelvic tumors', *Recent Results Cancer Res.* **107**, 236–243.

D. Kainmueller, T. Lange and H. Lamecker (2007), Shape constrained automatic segmentation of the liver based on a heuristic intensity model. In *Proc. MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge*, pp. 109–116.

T. Köhler, P. Maass, P. Wust and M. Seebass (2001), 'A fast algorithm to find optimal controls of multiantenna applicators in regional hyperthermia', *Phys. Med. Biol.* **46**, 2503–2514.

M. E. Kowalski, B. Behnia, A. G. Webb and J.-M. Jin (2002), 'Optimization of electromagnetic phased-arrays for hyperthermia via magnetic resonance temperature estimation', *IEEE Trans. Biomed. Engng* **49**, 1229–1241.

J. Kremer and A. K. Louis (1990), 'On the mathematical foundations of hyperthermia therapy', *Math. Meth. Appl. Sci.* **13**, 467–479.

J. Lagendijk, M. Schellekens, J. Schipper and P. van der Linden (a984), 'A three-dimensional description of heating patterns in vascularized tissues during hyperthermia treatment', *Phys. Med. Biol.* **29**, 495–507.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

H. Lamecker (2008), Variational and statistical shape modeling for 3D geometry reconstruction. PhD thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin.

H. Lamecker, M. Seebass, H.-C. Hege and P. Deuflhard (2004), A 3D statistical shape model of the pelvic bone for segmentation. In *Medical Imaging 2004: Image Processing* (J. M. Fitzpatrick and M. Sonka, eds), Vol. 5370 of *Proceedings of SPIE*, pp. 1341–1351.

J. Lang, B. Erdmann and M. Seebass (1999), 'Impact of nonlinear heat transfer on temperature control in regional hyperthermia', *IEEE Trans. Biomed. Engrg* **46**, 1129–1138.

R. Löhner and P. Parikh (1988), 'Generation of three-dimensional unstructured grids by the advancing-front method', *Int. J. Numer. Methods Fluids* **8**, 1135–1149.

L. Lüdemann, P. Wust and J. Gellermann (2008), 'Perfusion measurement using DCE-MRI: Implications for hyperthermia', *Int. J. Hyperthermia* **24**, 91–96.

A. Massing (2007), 1D-reduktion thermal signifikanter Aderstränge in der Hyperthermie-Modellierung. Master's thesis, Freie Universität Berlin.

C. Meyer, F. Tröltzsch and A. Rösch (2006), 'Optimal control problems of PDEs with regularized pointwise state constraints', *COAP* **33**, 206–228.

P. Monk (1992), 'A finite element method for approximating the time-harmonic Maxwell equations', *Numer. Math.* **63**, 243–261.

P. Monk (2003), *Finite Element Methods for Maxwell's Equations*, Numerical Mathematics and Scientific Computation, Oxford Science Publications.

J. C. Nédélec (1980), 'Mixed finite elements in $\mathbb{R}^3$', *Numer. Math.* **35**, 315–341.

K. Nikita, N. Maratos and N. Uzunoglu (1993), 'Optimal steady-state temperature distribution for a phased array hyperthermia system', *IEEE Trans. Biomed. Engng* **40**, 1299–1306.

J. Ortega and W. Rheinboldt (1970), *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press.

K. Paulsen, S. Geimer, J. Tang and W. Boyse (1999), 'Optimization of pelvic heating rate distributions with electromagnetic phased arrays', *Int. J Hyperthermia* **15**, 157–186.

H. H. Pennes (1948), 'Analysis of tissue and arterial blood temperatures in the resting human forearm', *J. Applied Physiology* **1**, 93–122.

A. Quarteroni, M. Tuveri and A. Veneziani (2000), 'Computational vascular fluid dynamics: Problems, models, and methods', *Comp. Vis. Sci.* **2**, 163–197.

M. Ranneberg, M. Weiser, M. Weihrauch, V. Budach, J. Gellermann and P. Wust (2010), 'Regularized antenna profile adaptation in online hyperthermia treatment', *Medical Physics* **37**, 5382–5394.

V. Rieke and K.-B. Pauly (2008), 'MR thermometry', *J. Magn. Reson. Imaging* **27**, 376–390.

R. B. Roemer and T. C. Cetas (1984), 'Applications of bioheat transfer simulations in hyperthermia', *Cancer Research* (*Suppl.*) **44**, 4788–4798.

G. Rosner, S. Clegg, D. Prescott and M. Dewhirst (1996), 'Estimation of cell survival in tumours heated to nonuniform temperature distributions', *Int. J. Hyperthermia* **12**, 223–239.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

J. Savage and A. Peterson (1996), 'High-order vector finite elements for tetrahedral cells', *IEEE Trans Microwave Theory Techn.* **44**, 874–879.

P. Scheid (1961), 'Funktionelle Besonderheiten der Mikrozirkulation im Karzinom', *Bibliotheca Anatomica* **1**, 327–335.

A. Schiela (2008), An interior point method in function space for the efficient solution of state constrained optimal control problems. ZIB report 07-44, Zuse Institute Berlin.

A. Schiela (2009), 'Barrier methods for optimal control problems with state constraints', *SIAM J. Optim.* **20**, 1002–1031.

A. Schiela and A. Günther (2011), 'An interior point algorithm with inexact step computation in function space for state constrained optimal control', *Numer. Math.* **119**, 373–407.

A. Schiela and M. Hintermüller (2011), On the length of the primal–dual path in Moreau–Yosida-based path-following for state constrained optimal control: Analysis and numerics. ZIB report 11-37, Zuse Institute Berlin.

A. Schiela and M. Weiser (2010), Barrier methods for a control problem from hyperthermia treatment planning. In *Recent Advantages in Optimization and its Applications in Engineering*, Springer, pp. 419–428.

A. Schiela and M. Weiser (2012), An affine invariant composite-step method for non-convex optimal control problems. Technical report, Zuse Institute Berlin. In preparation.

M. Seebass, R. Beck, J. Gellermann, J. Nadobny and P. Wust (2000), Electromagnetic phased arrays for regional hyperthermia: Optimal frequency and antenna arrangement. Technical report, Zuse Institute Berlin.

C. Song, A. Lokshina, J. Rhee, M. Patten and S. Levitt (1984), 'Implication of blood flow in hyperthermic treatment of tumors', *IEEE Trans. Biomed. Engrg.* **31**, 9–16.

G. Sreenivasa, J. Gellermann, B. Rau, J. Nadobny, P. Schlag, P. Deuflhard, R. Felix and P. Wust (2003), 'Clinical use of the hyperthermia treatment planning system HyperPlan to predict effectiveness and toxity', *Int. J. Radiation Oncology Biol. Phys.* **55**, 407–419.

D. Stalling (1998), Fast texture-based algorithms for vector field visualization. PhD thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin.

D. Stalling, M. Seebass, H.-C. Hege, P. Wust, P. Deuflhard and R. Felix (1996), HyperPlan: An integrated system for treatment planning in regional hyperthermia. In *Congress on Hyperthermic Oncology*, Vol. II (C. Franconi, G. Arcangeli and R. Cavaliere, eds), University Tor Vergata, Rome, pp. 552–554.

D. Stalling, M. Westerhoff and H.-C. Hege (2005), Amira: A highly interactive system for visual data analysis. In *The Visualization Handbook* (C. Hansen and C. Johnson, eds), Elsevier, chapter 38, pp. 749–767.

D. M. Sullivan (1990), 'Three-dimensional computer simulation in deep regional hyperthermia using the FDTD method', *IEEE Trans. Microwave Theory and Techniques* **38**, 204–211.

A. Szasz and G. Vincze (2006), 'Dose concept of oncological hyperthermia: Heat-equation considering the cell destruction', *J. Cancer Research Therapeutics* **2**, 171–181.

T. Terlaky (1996), *Interior Point Methods of Mathematical Programming*, Springer.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

D. Tompkins, R. Vanderby, S. Klein, W. Beckman, R. Steeves and B. Paliwal (1994*a*), 'The use of generalized cell-survival data in a physiologically based objective function for hyperthermia treatment planning: A sensitivity study with a simple tissue model implanted with an array of ferromagnetic thermoseeds', *Int. J. Radiation Oncology Biol. Phys.* **30**, 929–943.

D. Tompkins, R. Vanderby, S. Klein, W. Beckman, R. Steeves, D. Frey and B. Paliwal (1994*b*), 'Temperature-dependent versus constant-rate blood perfusion modelling in ferromagnetic thermoseed hyperthermia: Results with a model of the human prostate', *Int. J. Hyperthermia*.

L. Vincent and P. Soille (1991), 'Watersheds in digital spaces: An efficient algorithm based on immersion simulations', *IEEE Trans. Pattern Analysis and Machine Intelligence* **13**, 583–598.

S. Volkwein and M. Weiser (2002), 'Affine invariant convergence analysis for inexact augmented Lagrangian–SQP methods', *SIAM J. Control Optim.* **41**, 875–899.

M. Weihrauch, P. Wust, M. Weiser, J. Nadobny, S. Eisenhardt, V. Budach and J. Gellermann (2007), 'Adaptation of antenna profiles for control of MR guided hyperthermia (HT) in a hybrid MR–HT system', *Medical Physics* **34**, 4717–4725.

M. Weiser (2009), 'Optimization and identification in regional hyperthermia', *Int. J. Appl. Electromagn. and Mech.* **30**, 265–275.

M. Weiser (2012), 'On goal-oriented adaptivity for elliptic optimal control problems', *Optim. Meth. Software*, to appear.

M. Weiser, P. Deuflhard and B. Erdmann (2007), 'Affine conjugate adaptive Newton methods for nonlinear elastomechanics', *Optim. Meth. Software* **22**, 413–431.

M. Weiser, T. Gänzler and A. Schiela (2008), 'A control reduced primal interior point method for PDE constrained optimization', *Comput. Optim. Appl.* **41**, 127–145.

H. Whitney (1957), *Geometric Integration Theory*, Princeton University Press.

S. Wright (1987), *Primal–Dual Interior-Point Methods*, SIAM.

P. Wust, J. Nadobny, R. Felix, P. Deuflhard, W. John and A. Louis (1989), 'Numerical approaches to treatment planning in deep RF-hyperthermia', *Strahlenther. Onkol.* **165**, 751–757.

P. Wust, J. Nadobny, R. Felix, P. Deuflhard, A. Louis and W. John (1991), 'Numerical approaches to treatment planning in deep RF-hyperthermia', *Int. J. Hyperthermia* **7**, 157–173.

P. Wust, J. Nadobny, M. Seebass, D. Stalling, J. Gellermann, H.-C. Hege, P. Deuflhard and R. Felix (1999), 'Influence of patient models and numerical methods on predicted power deposition patterns', *Int. J. Hyperthermia* **15**, 519–540.

J. Xu (1989), Theory of multilevel methods. Dissertation, Pennsylvania State University, University Park, USA.

J. Xu (1992), 'Iterative methods by space decomposition and subspace correction', *SIAM Review* **34**, 581–613.

J. Xu and L. Chen and R. H. Nochetto (2009), Optimal multilevel methods for $H(\mathrm{grad})$, $H(\mathrm{curl})$, and $H(\mathrm{div})$ systems on graded and unstructured grids. In *Multiscale, Nonlinear and Adaptive Approximation*, Springer, pp. 599–659.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press

H. Yserentant (1986), 'On the multi-level splitting of finite element spaces', *Numer. Math.* **49**, 379–412.

H. Yserentant (1993), Old and new convergence proofs for multigrid methods. In *Acta Numerica*, Vol. 2, Cambridge University Press, pp. 285–326.

S. Zachow, M. Zilske and H.-C. Hege (2007), 3D reconstruction of individual anatomy from medical image data: Segmentation and geometry processing. ZIB report 07-41, Zuse Institute Berlin.

M. Zilske, H. Lamecker and S. Zachow (2008), Adaptive remeshing of non-manifold surfaces. In *Eurographics 2008 Annex Conf. Proc.*, pp. 207–211.

B. Zitov and J. Flusser (2003), 'Image registration methods: A survey', *Image Vision Comput.* **21**, 977–1000.

https://doi.org/10.1017/S0962492912000049 Published online by Cambridge University Press