

Therapieplanung mit Numerischer Mathematik

Tim Kröger

Sommersemester 2007 (Version vom 19. Oktober 2007)

Inhaltsverzeichnis

1. Einleitung	4
2. Elektrostatik	6
2.1. Elektrisches Feld und Potenzial	6
2.2. Potenzialgleichung	9
2.3. Randbedingungen für die Potenzialgleichung	10
3. Finite Elemente für die Potenzialgleichung	13
3.1. Schwache Lösungen	13
3.2. Diskretisierung	18
3.3. Inhomogene Randbedingungen	23
4. Wärmeausbreitung im Gewebe	25
4.1. Wärmequelle durch elektrische Leistung	25
4.2. Skalierung der elektrischen Leistung	26
4.3. Wärmeleitung	28
5. Finite Elemente für die Wärmeleitungsgleichung	30
A. Klassifizierung von partiellen Differenzialgleichungen zweiter Ordnung	32
A.1. Klassifizierung	32
A.2. Typische Eigenschaften	33
A.3. Analytische Lösung der Prototypen	34

1. Einleitung

In dieser Veranstaltung beschäftigen wir uns mit der numerischen Simulation von Radiofrequenzablation. Radiofrequenzablation ist eine minimal invasive Therapiemethode zur Behandlung von Lebertumoren.

Wenn ein Patient eine Krebserkrankung hat – egal wo sich der primäre Tumor befindet –, gelangen Krebszellen ins Blut und werden durch den ganzen Körper transportiert. Besonders häufig setzen sich solche Krebszellen in der Leber ab, da diese das Blut filtert. Daraus entstehen dann Krebsmetastasen in der Leber. Somit entwickeln praktisch alle Krebspatienten recht bald auch Lebermetastasen; oft schon, bevor der Krebs diagnostiziert wurde. Daher ist es von großer Bedeutung, Verfahren zur Behandlung solcher Lebermetastasen zu entwickeln.

Eine operative Entfernung von Lebertumoren ist zur Zeit nur in etwa 25 % der Fälle erfolgreich anwendbar (zum Beispiel auf Grund ungünstiger Lage des Tumores). Außerdem ist eine solche Operation sehr aufwändig und teuer. Eine Chemotherapie alleine ist ebenfalls in der Regel nicht erfolgreich. Daher ist es von Bedeutung, nach alternativen, „minimal invasiven“ Verfahren zu suchen. Einige davon sind:

Radiofrequenzablation: Ein nadelförmiger Applikator wird von außen durch die Haut in die Leber eingestochen. Durch Anlegen eines elektrischen Stroms wird das Gewebe lokal erhitzt und der Tumor damit verkocht.

Laser-induzierte Thermotherapie: Es wird ebenfalls ein Applikator eingestochen, jedoch wird das Gewebe lokal dann durch Laserstrahlen zerstört.

Hyperthermie: Der Patient wird von außen mit elektromagnetischen Wellen bestrahlt, die im Tumor fokussieren und diesen dann wiederum durch die entstehende Hitze zerstören.

Kryotherapie: Es wird wiederum ein Applikator eingestochen, der diesmal jedoch stark gekühlt wird und durch die Kälte ebenfalls das Gewebe abtötet.

Das zerstörte Gewebe wird mit der Zeit vom Körper abgebaut und durch neues, intaktes Lebergewebe ersetzt. (Die Leber ist das einzige innere Organ des Menschen, das nachwächst.)

Wir beschränken uns hier auf die Radiofrequenzablation.

Bisher sind die Ärzte bei der Durchführung dieser Behandlung hauptsächlich auf ihre Erfahrung (bezüglich der Positionierung des Applikators, Einstellung des elektrischen Generators sowie Wahl der Ablationsdauer) angewiesen. Ziel einer numerischen Simulation ist eine Vorhersage des zerstörten Gewebegebiets in Abhängigkeit von diesen Parametern. Dies dient unter anderem folgenden Zwecken:

- Verbesserung der Planung eines Eingriffs;
- Einzelfallentscheidung für Radiofrequenzablation oder eine andere Therapieform;
- Training von Ärzten.

Die folgenden physikalischen und chemischen Effekte müssen bei der Simulation berücksichtigt werden:

- elektrisches Feld;
- Heizquelle durch Stromfluss;
- Wärmeleitung;
- Koagulation von Proteinen;
- Wärmetransport / Kühleffekt durch fließendes Blut;
- Sieden von Zellwasser;
- veränderliche Materialparameter.

Zu Beginn der Veranstaltung wird die Komplexität so gering wie möglich gehalten, so dass möglichst schnell ein „sinnvolles“ Gesamtverfahren entwickelt wird. Später werden wir nach und nach mehr Komplexitätsebenen hinzuzunehmen.

Dieses Skript umfasst jedoch nur einen Teil der Veranstaltung, nämlich denjenigen, der von Tim Kröger gehalten wurde. Die Teile der Veranstaltung, die von Tobias Preußer präsentiert wurden, sind nicht in diesem Skript enthalten. Außerdem wurden Teile der Veranstaltung in Form von studentischen Seminarvorträgen dargeboten; diese sind ebenfalls nicht in diesem Skript enthalten.

2. Elektrostatik

2.1. Elektrisches Feld und Potenzial

In Wirklichkeit wird am Applikator ein Wechselstrom mit einer Frequenz von etwa 500 kHz angelegt. (Daher resultiert auch die Bezeichnung „Radiofrequenzablation“.) In diesem Frequenzbereich herrscht physiologische Reizfreiheit, das heißt der Patient „bekommt keinen Stromschlag“. Für unsere Berechnungen gehen wir jedoch vereinfachend von einem Gleichstrom aus, so dass wir die Gesetze der Elektrostatik anwenden können. Dies ist gerechtfertigt, wenn zwei Bedingungen erfüllt sind:

- Zum Einen muss die Wellenlänge der elektromagnetischen Wellen groß gegen die Abmessungen der Elektrodenanordnungen sein. Die Wellenlänge bestimmt sich gemäß

$$\lambda = \frac{c}{f\sqrt{\epsilon_r}},$$

wobei $c = 3 \cdot 10^8$ m/s die Lichtgeschwindigkeit im Vakuum, $f = 5 \cdot 10^5$ s⁻¹ die Frequenz und ϵ_r die relative Dielektrizitätskonstante des untersuchten Mediums ist; sie liegt für Leber je nach Zustand im Bereich von etwa 1000 bis 5000 (einheitenlos). Die Wellenlängen liegen somit im Bereich von 8 bis 19 Metern; die Bedingung ist also erfüllt. Dies bedeutet, dass die zeitliche Änderung vernachlässigt werden kann; das elektromagnetische Feld ist *quasistationär*.

- Zum Anderen muss sichergestellt werden, dass das durch das sich verändernde elektrische Feld erzeugte Magnetfeld wiederum das elektrische Feld nicht nennenswert beeinflusst. Dies ist erfüllt, wenn die Größe

$$\delta = \sqrt{\frac{2}{\omega\sigma\mu}}$$

groß gegenüber dem Radius des durch den elektrischen Stromfluss beeinflussten Gewebebereichs ist. Dabei ist $\omega = 2\pi f$ die Kreisfrequenz, σ die elektrische Leitdichte und μ die magnetische Permeabilität des Gewebes. σ liegt zwischen 0.2 und 0.5 s³ A²/kg m³ und μ liegt nahe dem Wert im Vakuum von $\mu_0 = 1.256 \cdot 10^{-6}$ kg m/A² s². Somit liegt δ zwischen 1 und 2 Metern; die Bedingung ist also auch erfüllt. Somit kann das Magnetfeld komplett vernachlässigt werden; das Feld ist nicht nur quasistationär, sondern auch *quasistatisch*.

Der am Applikator angelegte elektrische Strom erzeugt eine Ladungsverteilung in den Elektroden und im Gewebe. Würde man einen geladenen Gegenstand in die Nähe des

Applikators bringen, so würde auf diesen eine Kraft \vec{F} wirken. Die Kraft \vec{F} hängt ab vom Ort \vec{x} und der Ladung q des Gegenstands. Sie ist proportional zur Ladung q , das heißt, es gilt die Beziehung

$$\vec{F}(q, \vec{x}) = q\vec{E}(\vec{x})$$

mit einem Vektorfeld \vec{E} , das als *elektrisches Feld* bezeichnet wird. Die SI-Einheiten sind:

$$[\vec{F}] = \text{kg m/s}^2 = \text{N},$$

$$[q] = \text{A s} = \text{C},$$

$$[\vec{E}] = \text{kg m/A s}^3.$$

Satz und Definition 2.1 *Es sei $G \subset \mathbb{R}^3$ ein sternförmiges Gebiet und $\vec{E} : G \rightarrow \mathbb{R}^3$ ein stetig differenzierbares Vektorfeld. Dann sind folgende Bedingungen äquivalent:*

1. *Es existiert eine zweimal stetig differenzierbare Funktion $\varphi : G \rightarrow \mathbb{R}$ mit $\vec{E} = -\nabla\varphi$.*
2. $\nabla \times \vec{E} = 0$.
3. $\oint \vec{E} \cdot d\vec{s} = 0$ über jeden geschlossenen Weg in G .

Sind diese Bedingungen erfüllt, so heißt \vec{E} ein konservatives Feld.

Beweis: Ist Bedingung 1 erfüllt, so folgt

$$\nabla \times \vec{E} = -\nabla \times \nabla\varphi = 0,$$

weil die Rotation eines Gradienten stets verschwindet.

Sei jetzt Bedingung 2 erfüllt und ein geschlossener Weg in G gegeben. Der geschlossene Weg kann als Rand ∂M einer glatten Mannigfaltigkeit $M \subset G$ dargestellt werden (indem man jeden Punkt der Kurve mit dem Zentrum \vec{z} des sternförmigen Gebiets G verbindet und dann in einer kleinen Kugel um z die Fläche glättet). Nach dem Satz von Stokes ist dann

$$\oint_{\partial M} \vec{E} \cdot d\vec{s} = \int_M (\nabla \times \vec{E}) \cdot \vec{n} dA = 0.$$

(Hierbei ist \vec{n} der Normalenvektor an die Fläche M , der in die Richtung zeigt, von der aus gesehen die Randkurve im Gegenuhrzeigersinn orientiert ist.)

Sei jetzt Bedingung 3 erfüllt. Wir definieren φ durch

$$\varphi(\vec{x}) = - \int_{\vec{z}}^{\vec{x}} \vec{E} \cdot d\vec{s}, \quad (2.1)$$

wobei \vec{z} das Zentrum des sternförmigen Gebiets G ist und das Integral über einen beliebigen Weg von \vec{z} nach \vec{x} erfolgt (zum Beispiel die Strecke von \vec{z} nach \vec{x}). Offensichtlich

ist der Wert des Integrals von der Wahl des Weges unabhängig. Jede Komponente von $\nabla\varphi$ hat die Form

$$\partial_j\varphi = \lim_{h \rightarrow 0} \frac{\varphi(\vec{x} + h\vec{e}_j) - \varphi(\vec{x})}{h},$$

wobei \vec{e}_j der j -te Einheitsvektor ist. Da die Wahl des Weges in (2.1) beliebig ist, kann man den Weg von \vec{z} nach $\vec{x} + h\vec{e}_j$ aus dem Weg von \vec{z} nach \vec{x} und der Strecke von \vec{x} nach $\vec{x} + h\vec{e}_j$ zusammensetzen. Für den gemeinsamen Wegteil fällt das Integral weg, und es bleibt:

$$\partial_j\varphi = - \lim_{h \rightarrow 0} \frac{1}{h} \int_{\vec{x}}^{\vec{x}+h\vec{e}_j} \vec{E} \cdot d\vec{s} = - \lim_{h \rightarrow 0} \frac{1}{h} \int_0^h \vec{E}(\vec{x} + \tau\vec{e}_j) \cdot \vec{e}_j d\tau.$$

Nach dem Mittelwertsatz existiert zu jedem h ein $\xi(h)$ mit $0 \leq \xi(h) \leq h$ so, dass

$$\int_0^h \vec{E}(\vec{x} + \tau\vec{e}_j) \cdot \vec{e}_j d\tau = h\vec{E}(\vec{x} + \xi(h)\vec{e}_j) \cdot \vec{e}_j$$

ist, also

$$\partial_j\varphi = - \lim_{h \rightarrow 0} \vec{E}(\vec{x} + \xi(h)\vec{e}_j) \cdot \vec{e}_j = -\vec{E}(\vec{x}) \cdot \vec{e}_j = -E_j(\vec{x}),$$

was zu zeigen war. □

Übung 2.2 Wenn man in der Voraussetzung den Begriff „sternförmig“ durch „zusammenhängend“ ersetzt, in welcher Beziehung stehen die drei Bedingungen dann? Welche Beweisteile gehen schief? Kann man ein Gegenbeispiel finden?

Um in unserem elektrischen Feld den geladenen Gegenstand vom Punkt \vec{x} zum Punkt \vec{y} zu bringen, ist die Energie

$$W = - \int_{\vec{x}}^{\vec{y}} \vec{F} \cdot d\vec{s} = -q \int_{\vec{x}}^{\vec{y}} \vec{E} \cdot d\vec{s}$$

aufzuwenden. Der Integrationsweg ist der Weg, entlang dem der Gegenstand bewegt wird. Ist $\vec{x} = \vec{y}$, so muss wegen des Prinzips der Energieerhaltung $W = 0$ sein, egal welcher Weg gewählt wurde. Somit gilt

$$\oint \vec{E} \cdot d\vec{s} = 0$$

entlang jedes geschlossenen Weges. Das elektrische Feld ist also konservativ, und es existiert eine Funktion φ mit

$$\vec{E} = -\nabla\varphi. \tag{2.2}$$

φ wird als *elektrisches Potenzial* bezeichnet. Die SI-Einheiten sind:

$$\begin{aligned} [W] &= \text{kg m}^2/\text{s}^2 = \text{J}, \\ [\varphi] &= \text{kg m}^2/\text{A s}^3 = \text{V}. \end{aligned}$$

Übung 2.3 Warum schreibt man wohl das Minus-Zeichen vor $\nabla\varphi$?

2.2. Potenzialgleichung

Es sei σ die *elektrische Leitdichte*. Wir müssen damit rechnen, dass σ nicht konstant ist, weil der Tumor beispielsweise andere elektrische Eigenschaften hat als gesundes Gewebe oder bereits koaguliertes Gewebe oder Blutgefäße und weil die elektrischen Eigenschaften sich auch mit der Temperatur ändern können. Außerdem ändert sich die elektrische Leitfähigkeit sehr stark, wenn das Wasser in den Zellen verdampft. Wir schreiben also $\sigma(\vec{x})$.

Die *elektrische Stromdichte* $\vec{J}(\vec{x})$ gibt an, wie viel Strom pro Zeiteinheit durch ein Flächenstück fließt (und in welche Richtung). Sie ist das Produkt des elektrischen Feldes und der elektrischen Leitfähigkeit:

$$\vec{J} = \sigma \vec{E}. \quad (2.3)$$

Die SI-Einheiten sind:

$$\begin{aligned} [\sigma] &= \text{s}^3 \text{A}^2 / \text{kg m}^3 = 1/\Omega \text{m}, \\ [\vec{J}] &= \text{A/m}^2. \end{aligned}$$

Betrachtet man ein Volumen $V \subset \mathbb{R}^3$ im Gewebe, das sich außerhalb der Elektroden befindet, so muss über den Rand genauso viel Strom hinein- wie herausfließen, da man im Gewebe keine Stromquelle oder -senke hat. (Die Elektroden werden als Stromquelle und -senke betrachtet, deswegen gilt dies dort nicht.) Man hat also

$$\int_{\partial V} \vec{J} \cdot \vec{n} dA = 0, \quad (2.4)$$

wobei \vec{n} die äußere Normale an das Volumen V ist. Mittels Satz von Gauß (partielle Integration) erhält man

$$\int_{\partial V} \vec{J} \cdot \vec{n} dA = \int_V \nabla \cdot \vec{J}(\vec{x}) d\vec{x}.$$

Ist $\nabla \cdot \vec{J}$ stetig, so folgt daraus, dass $\nabla \cdot \vec{J}$ außerhalb der Elektroden identisch verschwindet, denn wäre etwa $\nabla \cdot \vec{J}(\vec{y}) \neq 0$ für ein \vec{y} , so müsste $\nabla \cdot \vec{J}$ auf einer Umgebung V von \vec{y} ungleich null sein und einheitliches Vorzeichen haben, weswegen dann $\int_V \nabla \cdot \vec{J}(\vec{x}) d\vec{x}$ für dieses V nicht verschwinden könnte. Es gilt also:

$$0 = \nabla \cdot \vec{J}(\vec{x}) = \nabla \cdot (\sigma(\vec{x}) \vec{E}(\vec{x})) = -\nabla \cdot (\sigma(\vec{x}) \nabla \varphi(\vec{x})) \quad (2.5)$$

außerhalb der Elektroden. Dies ist eine elliptische partielle Differentialgleichung für φ . Bisher ist sie sogar linear.

Wir betrachten nun noch den Fall, dass σ und damit \vec{J} nur stückweise stetig ist. Der Einfachheit halber, sei angenommen, dass σ nur entlang einer glatten Fläche M springt. M unterteile eine Grundmenge Ω in zwei offene Teilmengen Ω_1 und Ω_2 . Es bezeichnen

$\sigma_1, \sigma_2, \varphi_1$ und φ_2 die Werte von σ und φ auf Ω_1 und Ω_2 . Wir nehmen an, dass all diese Funktionen stetig auf den Rand M fortsetzbar sind. Dann führt (2.4) zu

$$0 = \int_{\partial V} \vec{J} \cdot \vec{n} dA = \int_{\partial V \cap \Omega_1} \vec{J}_1 \cdot \vec{n}_1 dA + \int_{\partial V \cap \Omega_2} \vec{J}_2 \cdot \vec{n}_2 dA.$$

Nun ist $\partial V \cap \Omega_j$ ein Teil des Randes von $V \cap \Omega_j$; der andere Teil ist jeweils $V \cap M$. Definiert man auf M eine Normale \vec{n}_M , die in Richtung Ω_2 zeigt, so gilt¹:

$$\begin{aligned} \dots &= \int_{\partial(V \cap \Omega_1)} \vec{J}_1 \cdot \vec{n}_1 dA + \int_{\partial(V \cap \Omega_2)} \vec{J}_2 \cdot \vec{n}_2 dA + \int_{V \cap M} (\vec{J}_2 - \vec{J}_1) \cdot \vec{n}_M dA \\ &= \int_{V \cap \Omega_1} \nabla \cdot \vec{J}_1 d\vec{x} + \int_{V \cap \Omega_2} \nabla \cdot \vec{J}_2 d\vec{x} + \int_{V \cap M} (\vec{J}_2 - \vec{J}_1) \cdot \vec{n}_M dA \\ &= \int_{V \cap M} (\vec{J}_2 - \vec{J}_1) \cdot \vec{n}_M dA, \end{aligned}$$

weil $\nabla \cdot J$ ja im Inneren von Ω_1 und Ω_2 nach wie vor verschwindet. Da \vec{J}_1 und \vec{J}_2 entlang M jeweils stetig sind, muss (nach ähnlichen Argumenten wie oben) $(\vec{J}_2 - \vec{J}_1) \cdot \vec{n}_M$ überall auf M verschwinden. Einsetzen von (2.3) und (2.2) liefert die Bedingung

$$(\sigma_2 \nabla \varphi_2 - \sigma_1 \nabla \varphi_1) \cdot \vec{n}_M = 0, \quad (2.6)$$

die also entlang jeder Unstetigkeit von σ als das Äquivalent zur Differenzialgleichung (2.5) anzusehen ist. Wir werden das später bei der schwachen Formulierung wieder aufgreifen; dort wird es ganz natürlich ins Kalkül passen.

2.3. Randbedingungen für die Potenzialgleichung

Damit die Potenzialgleichung eine eindeutige Lösung hat, müssen wir sie noch mit Randwerten versehen. Dazu müssen wir uns zunächst überlegen, was unser Rechengebiet ist und wo es Ränder hat.

- Da man die numerischen Rechnungen nicht gut auf eine unbeschränkten Gebiet durchführen kann und da es außerdem anzunehmen ist, dass die Ablation nur die Region lokal um den Applikator beeinflusst, wählt man eine „Interessenregion“ (englische Bezeichnung: *Region of Interest*, abgekürzt *ROI*). Sie ist der Einfachheit halber quaderförmig und so groß, dass man annehmen kann, dass außerhalb keine für die Simulation relevanten Dinge passieren. Der Rand der Region of Interest ist somit auch Rand des Rechengebiets; wir nennen ihn *äußeren Rand*.
- Da (2.5) nur außerhalb der Elektroden gilt, müssen wir die Elektroden aus dem Rechengebiet ausschneiden. Somit wird der Rand der Elektroden auch zu einem

¹Die Punkte „ \dots “ am Anfang einer Gleichungskette bedeuten, dass die vorangegangene Gleichungskette (oder eine andere, wenn dies aus dem Zusammenhang hervorgeht) fortgesetzt wird.

Rand des Rechengebiets; wir haben also einen sogenannten *inneren Rand*. (Die Größen \vec{E} und φ existieren prinzipiell auch im Inneren der Elektroden, so dass die Sternförmigkeitsvoraussetzung in Satz 2.1 keine Probleme macht. Lediglich die Gleichung (2.5) gilt dort nicht.)

Wir beschäftigen uns zunächst mit den inneren Rändern. Es sei angenommen, dass die Elektroden eine unendlich große elektrische Leitfähigkeit haben. Dann ist auf ihrem Rand das elektrische Potenzial konstant. Für die numerische Berechnung setzen wir die Werte von φ auf den Elektroden auf 1 und -1 fest (wir setzen also Dirichlet-Randbedingungen an diesen Rändern). Das tatsächliche Potenzial unterscheidet sich dann von dem berechneten um eine multiplikative und eine additive Konstante. Die multiplikative Konstante wird später bestimmt; die additive spielt keine Rolle, da das Potenzial (als Stammfunktion des elektrischen Feldes) sowieso nur bis auf eine additive Konstante eindeutig ist. Ferner setzen wir das elektrische Potenzial im Inneren der Elektroden der Einfachheit halber konstant fort, weil der tatsächliche Wert dort für die Simulation nicht weiter relevant ist.

Die Bedingungen an den äußeren Rändern sind nicht so klar, da es sich dabei um künstliche Ränder handelt. Üblich sind folgende Möglichkeiten:

- Man kann annehmen, dass der Stromfluss über den Rand der ROI vernachlässigbar ist. In jedem Punkt des Randes ist dieser Stromfluss gegeben durch $\vec{J} \cdot \vec{n}$, wobei \vec{n} die äußere Normale an die ROI ist. Wegen $\vec{J} = -\sigma \nabla \varphi$ für das zu der Bedingung $\nabla \varphi \cdot \vec{n} = 0$ am Rand (sofern nicht $\sigma = 0$ ist, was bedeuten würde, dass das Gewebe ein elektrischer Isolator wäre). Diese Bedingungen werden als *Homogene Neumann-Randbedingungen* bezeichnet. Sie versagen allerdings bei der Verwendung sogenannter *monopolärer Applikatoren*, also Applikatoren mit nur einer Elektrode, weil hier der Stromkreis nur geschlossen wird, indem eine Neutralelektrode an einem entfernten Körperteil (zum Beispiel am Arm oder am Bein) des Patienten befestigt wird. Monopolare Applikatoren sind durchaus gebräuchlich, weil sie billiger sind. Die Berechnung mit Neumann-Randbedingungen am äußeren Rand führt dann zu einem konstanten Potenzial und somit zu einem verschwindenden elektrischen Feld. Damit kann man nichts anfangen.
- Homogene Dirichlet-Bedingungen: $\varphi = 0$.
- Robin-Randbedingungen:

$$\nabla \varphi \cdot \vec{n} = \frac{\vec{n} \cdot (\vec{s} - \vec{x})}{|\vec{s} - \vec{x}|^2} \varphi, \quad (2.7)$$

wobei \vec{s} der „Mittelpunkt der aktiven Zone“ ist (schwammiger Begriff). Die Idee dabei ist, dass sich das Potenzial in genügend großer Entfernung vom Applikator so verhält wie das von einer Punktladung (im Punkt \vec{s}) erzeugte Potenzial. Dieses kann man aber exakt berechnen:

$$\varphi_P = C \cdot \frac{1}{|\vec{s} - \vec{x}|}$$

mit einer Konstanten C , deren Wert in unserer Situation nicht ohne Weiteres bestimmbar ist. Man kann nachrechnen, dass φ_P unabhängig vom Wert von C die Gleichung (2.7) exakt erfüllt.

Wir entscheiden uns für die Robin-Randbedingungen und fassen zusammen: Das elektrische Potenzial wird modelliert durch das elliptische Randwertproblem

$$-\nabla \cdot (\sigma(\vec{x}) \nabla \varphi(\vec{x})) = 0 \quad \text{in } \Omega \setminus \overline{\Omega_{\pm}}, \quad (2.8a)$$

$$\varphi(\vec{x}) = 1 \quad \text{auf } \overline{\Omega_+}, \quad (2.8b)$$

$$\varphi(\vec{x}) = -1 \quad \text{auf } \overline{\Omega_-}, \quad (2.8c)$$

$$\nabla \varphi \cdot \vec{n} = \frac{\vec{n} \cdot (\vec{s} - \vec{x})}{|\vec{s} - \vec{x}|^2} \varphi \quad \text{auf } \Gamma. \quad (2.8d)$$

Hierbei haben wir folgende Bezeichnungen für bestimmten Teilmengen der \mathbb{R}^3 benutzt:

Ω : Region of Interest,

Ω_+ : positive Elektrode,

Ω_- : negative Elektrode,

Ω_{\pm} : $\Omega_+ \cup \Omega_-$,

Γ : $\partial\Omega$.

3. Finite Elemente für die Potenzialgleichung

Das elliptische Randwertproblem (2.8) lässt sich im Allgemeinen nicht analytisch lösen. Wir wollen sie numerisch mit der Methode der Finiten Elemente behandeln. Viele der fehlenden Beweise findet man in guten Lehrbüchern über Finite Elemente, zum Beispiel Braess.

Da wir in diesem Kapitel in ein anderes Gebiet der Mathematik eintauchen, benutzen wir zum Teil auch andere Bezeichnungen: Ω bezeichnet abweichend das Rechengebiet für die Potenzialgleichung, also die ROI *ohne* die Elektroden. Das gesuchte Potenzial nennen wir u . Außerdem verzichten wir auf Vektorpfeile über vektorwertigen Größen.

3.1. Schwache Lösungen

Ω ist also ein offenes, beschränktes Rechengebiet im \mathbb{R}^n (bei uns $n = 3$). Wir nehmen an, dass der Rand Γ stückweise glatt ist und sich in zwei Teile Γ_1 und Γ_2 unterteilt (wobei Γ_2 offen in Γ ist), auf denen Dirichlet- beziehungsweise Neumann-Bedingungen vorgegeben sind. Ferner nehmen wir an, dass alle Randbedingungen homogen sind. Dies ist ein Widerspruch zu (2.8b) und (2.8c); wir kommen später darauf zurück. Stattdessen nehmen wir an, dass die Gleichung selbst inhomogen ist. Wir haben also zu lösen:

$$-\nabla \cdot (\sigma(x)\nabla u(x)) = f(x) \quad \text{in } \Omega, \quad (3.1a)$$

$$u(x) = 0 \quad \text{auf } \Gamma_1, \quad (3.1b)$$

$$\nabla u(x) \cdot n(x) = 0 \quad \text{auf } \Gamma_2, \quad (3.1c)$$

mit einer gegebenen Funktion f .

Die Methode der Finiten Elemente basiert darauf, dass man die das Randwertproblem umformuliert in ein Variationsproblem. Dazu nehmen wir an, wir hätten eine Lösung u von (3.1). Wir multiplizieren alsdann (3.1) mit einer sogenannten „Testfunktion“ v und integrieren das Ergebnis über Ω . Wir erhalten:

$$0 = - \int_{\Omega} \nabla \cdot (\sigma(x)\nabla u(x))v(x) dx - \int_{\Omega} f(x)v(x) dx.$$

Im ersten Integral wenden wir partielle Integration (Satz von Gauß) an und erhalten:

$$0 = \int_{\Omega} \sigma \nabla u \cdot \nabla v dx - \int_{\Gamma_1} \sigma \nabla u \cdot n v ds(x) - \int_{\Gamma_2} \sigma \nabla u \cdot n v ds(x) - \int_{\Omega} f v dx.$$

3. Finite Elemente für die Potenzialgleichung

Das Γ_2 -Integral fällt weg, weil $\nabla u \cdot n = 0$ ist gemäß (3.1c). Wenn wir ferner annehmen, dass die Testfunktion v ebenfalls die Dirichlet-Randbedingung (3.1b) erfüllt, fällt das Γ_1 -Integral auch weg. Es gilt dann also

$$0 = \int_{\Omega} \sigma \nabla u \cdot \nabla v \, dx - \int_{\Omega} f v \, dx. \quad (3.2)$$

Satz 3.1 Seien $\sigma \in C(\overline{\Omega}) \cap C^1(\Omega)$ und $f \in C(\overline{\Omega})$. Sei

$$X = \{u \in C(\overline{\Omega}) \cap C^2(\Omega) : u|_{\Gamma_1} = 0\}$$

und $u \in X$. Dann sind äquivalent:

1. Es gilt (3.1).
2. Es gilt (3.2) für alle $v \in X$.

Beweis: Die Hin-Richtung haben wir gerade gezeigt. Für die Rück-Richtung sei angenommen, dass (3.1a) in einem Punkt $x_0 \in \Omega$ nicht erfüllt sei, etwa

$$-\nabla \cdot (\sigma(x_0) \nabla u(x_0)) > f(x_0).$$

Aus Stetigkeitsgründen gibt es eine Umgebung V von x_0 , deren Abschluss noch in Ω liegt und auf der überall $-\nabla \cdot (\sigma \nabla u) > f$ ist. Man findet dazu ein $v \in X$, das außerhalb von V verschwindet, in V nichtnegativ ist und $v(x_0) = 1$ hat. Setzt man dieses v in (3.2) ein und formt alles zurück um (die Randterme entfallen, weil v auf dem Rand verschwindet), so erhält man einen Widerspruch. Also gilt (3.1a). Ferner gilt (3.1b) nach Definition von X . Wäre jetzt $x_0 \in \Gamma_2$ ein Punkt, in dem (3.1c) nicht gilt. Dann findet man wieder eine geeignete von x_0 in Γ_2 , und so weiter wie eben. Beim Einsetzen in (3.2) und Zurückrechnen fällt diesmal die Differenzialgleichung weg, weil wir von dieser ja schon wissen, dass sie erfüllt ist. \square

Auf X definieren wir die Linearform l und die Bilinearform a gemäß

$$l(v) = \int_{\Omega} f v \, dx, \quad a(u, v) = \int_{\Omega} \sigma \nabla u \cdot \nabla v \, dx,$$

dann schreibt sich (3.2) formal so:

$$a(u, v) = l(v) \quad \text{für alle } v \in X.$$

Definition 3.2 Sei X ein normierter Raum.

1. Eine Linearform l auf X heißt stetig, falls eine Konstante $C < \infty$ existiert so, dass $|l(u)| \leq C \|u\|$ für alle $u \in X$.
2. Eine Bilinearform a auf X heißt stetig, falls eine Konstante $C < \infty$ existiert so, dass $|a(u, v)| \leq C \|u\| \|v\|$ für alle $u, v \in X$.

3. Eine Bilinearform a auf X heißt elliptisch, falls eine Konstante $\alpha > 0$ existiert so, dass $a(u, u) \geq \alpha \|u\|^2$ für alle $u \in X$.

Satz 3.3 (Lax–Milgram) Sei H ein Hilbertraum, l eine stetige Linearform auf H und a eine stetige, elliptische Bilinearform auf H . Dann existiert genau ein $u \in H$ so, dass $a(u, v) = l(v)$ ist für alle $v \in H$.

Man kann übrigens zeigen, dass an der Lösung u das Funktional $J(v) = a(v, v)/2 - l(v)$ sein eindeutiges Minimum annimmt, das ist für uns aber hier nicht von Bedeutung.

Für die Finiten Elemente spielt der Satz von Lax–Milgram eine große Rolle, weil seine Aussage sich, wie wir noch sehen werden, auf das diskretisierte Problem überträgt. Wir müssen uns von unserem oben eingeführten Raum X jedoch verabschieden und nach Lösungen in allgemeineren Räumen (sogenannten *schwachen Lösungen*) suchen, denn:

- Der Raum X lässt sich nicht derart normieren, dass alle Voraussetzungen des Satzes von Lax–Milgram erfüllt sind.
- Mit dem Raum X bekämen wir auch Probleme, wenn σ oder f nicht mehr genügend glatt wären. Man beachte, dass σ zum Beispiel in verschiedenen Gewebetypen unterschiedliche Werte haben könnte.
- Für Rechengebiete mit sogenannten „einspringenden Ecken“ ist selbst für glatte σ und f nicht garantiert, dass in X überhaupt eine Lösung existiert.

Wie führen jetzt Räume von Funktionen ein, die in einem schwächeren als dem herkömmlichen Sinne differenzierbar sind, um danach auf diesen Räumen den Satz von Lax–Milgram anzuwenden.

Definition 3.4 Sei $u \in L^2(\Omega)$ eine Funktion und $\alpha \in \mathbb{N}_0^n$ ein Multiindex. Man sagt, u besitzt eine schwache Ableitung $\partial^\alpha u \in L^2(\Omega)$, falls eine Funktion $u_\alpha \in L^2(\Omega)$ existiert so, dass

$$\int_{\Omega} u \partial^\alpha v \, dx = (-1)^{|\alpha|} \int_{\Omega} u_\alpha v \, dx$$

für alle $v \in C_0^\infty(\Omega)$. Dieses u_α bezeichnet man dann mit $\partial^\alpha u$.

Übung 3.5 Man zeige, dass u_α , wenn es existiert, eindeutig bestimmt ist. Ferner zeige man, dass im Falle dass $\partial^\alpha u$ im herkömmlichen Sinne existiert und in L^2 ist, die Definition mit der herkömmlichen Definition übereinstimmt.

Übung 3.6 Man zeige, dass die Funktion $u(x) = |x|$ auf dem Intervall $\Omega = (-1, 1)$ schwache erste Ableitung besitzt.

Definition 3.7 Sei $m \in \mathbb{N}_0$. Der Sobolev-Raum $H^m(\Omega)$ ist gegeben durch

$$H^m(\Omega) = \{u \in L^2(\Omega) : \partial^\alpha u \in L^2(\Omega) \text{ für alle } |\alpha| \leq m\}.$$

3. Finite Elemente für die Potenzialgleichung

Er wird wie folgt mit Skalarprodukt, Norm und Seminorm versehen:

$$(u, v)_{H^m} = \sum_{|\alpha| \leq m} \int_{\Omega} \partial^\alpha u \partial^\alpha v \, dx, \quad \|u\|_{H^m}^2 = \sum_{|\alpha| \leq m} \int_{\Omega} (\partial^\alpha u)^2 \, dx,$$

$$|u|_{H^m}^2 = \sum_{|\alpha|=m} \int_{\Omega} (\partial^\alpha u)^2 \, dx.$$

Satz 3.8 $H^m(\Omega)$ ist ein Hilbertraum.

Übung 3.9 Was ist $H^0(\Omega)$?

Um auf unser Problem den Satz von Lax–Milgram anzuwenden, wollen wir

$$H = \{u \in H^1(\Omega) : u|_{\Gamma_1} = 0\}$$

setzen. Da u als L^2 -Funktion aber nur fast überall eindeutig gegeben ist und Γ_1 im \mathbb{R}^n eine Nullmenge ist, ist nicht klar, ob $u|_{\Gamma_1}$ überhaupt erklärt ist. Dies ist aber in der Tat der Fall:

Satz 3.10 (Spursatz) Sei $\Omega \subset \mathbb{R}^n$ offen und beschränkt mit stückweise glattem Rand, und in jedem Punkt x des Randes sei die sogenannte Kegelbedingung erfüllt, das heißt es gibt einen Kegel mit Spitze in x , der ganz in Ω liegt. Dann ist für $u \in H^1(\Omega)$ die Restriktion $u|_{\Gamma}$ sinnvoll erklärt; es ist $u|_{\Gamma} \in L^2(\Gamma)$, und es gilt

$$\|u|_{\Gamma}\|_{L^2} \leq C \|u\|_{H^1},$$

mit einer Konstanten C , die nur von Ω abhängt.

Die oben angegebene Definition von H ist also sinnvoll. Ferner ist H als abgeschlossener Unterraum von $H^1(\Omega)$ vollständig. Wir wollen prüfen, ob die verbleibenden Voraussetzungen des Satzes von Lax–Milgram gültig sind. Dies sind die Stetigkeit von l sowie die Stetigkeit und Elliptizität von a .

Zur Stetigkeit von l :

$$|l(v)| = \left| \int_{\Omega} v f \, dx \right| \leq \|v\|_{L^2} \|f\|_{L^2} \leq \|v\|_{H^1} \|f\|_{L^2},$$

sofern zumindest $f \in L^2(\Omega)$ ist. (Das ist eine wesentlich schwächere Voraussetzung als das ursprünglich geforderte $f \in C(\overline{\Omega})$, siehe Satz 3.1.)

Zur Stetigkeit von a :

$$|a(u, v)| = \left| \int_{\Omega} \sigma \nabla u \cdot \nabla v \, dx \right| \leq \|\sigma\|_{L^\infty} \int_{\Omega} |\nabla u| |\nabla v| \, dx \leq \|\sigma\|_{L^\infty} \|\nabla u\|_{L^2} \|\nabla v\|_{L^2}$$

$$\leq \|\sigma\|_{L^\infty} \|u\|_{H^1} \|v\|_{H^1}.$$

Dafür muss also zumindest $\sigma \in L^\infty(\Omega)$ sein.

Für die Elliptizität von a nehmen wir zusätzlich noch an, dass σ von null weg beschränkt ist, also $\sigma \geq \sigma_0 > 0$. Dann gilt:

$$a(v, v) = \int_{\Omega} \sigma (\nabla v)^2 dx \geq \sigma_0 \int_{\Omega} (\nabla v)^2 dx = \sigma_0 |v|_{H^1}^2.$$

Das reicht so nicht aus, wir haben nur die Seminorm, nicht die Norm. Man kann jedoch mit der sogenannten Poincaré-Friedrichschen Ungleichung zeigen, dass die Elliptizität dennoch gegeben ist, sofern der Teilrand, auf dem Dirichlet-Bedingungen gegeben sind, (also Γ_1) positives $(n-1)$ -dimensionales Lebesgue-Maß hat (also nicht „fast nichts“ ist). Wir erinnern uns daran, dass Γ_1 bei uns die Oberfläche der Elektroden ist, die ja tatsächlich nicht „fast nichts“ sind.

Wir fassen zusammen:

Satz und Definition 3.11 *Es sei*

$$H = \{u \in H^1(\Omega) : u|_{\Gamma_1} = 0\},$$

und auf H seien eine Linearform l und eine Bilinearform a erklärt gemäß

$$l(v) = \int_{\Omega} f v dx, \quad a(u, v) = \int_{\Omega} \sigma \nabla u \cdot \nabla v dx. \quad (3.3)$$

Mit der H^1 -Norm ist H ein Hilbertraum, und alle Voraussetzungen des Satzes von Lax-Milgram (Satz 3.3) sind erfüllt. Somit existiert ein eindeutiges $u \in H$ so, dass $a(u, v) = l(v)$ für alle $v \in H$. Dieses u heißt schwache Lösung des Randwertproblems (3.1). Falls $u \in C(\overline{\Omega}) \cap C^2(\Omega)$ ist und außerdem $\sigma \in C(\overline{\Omega}) \cap C^1(\Omega)$ und $f \in C(\overline{\Omega})$, so ist u starke Lösung.

Wir wollen uns jetzt noch einmal um den Fall eines unstetigen σ kümmern, um zu sehen, dass die obige schwache Formulierung des Randwertproblems in dem Fall zu derselben Sprungbedingung wie in Abschnitt 2.2 (Gleichung (2.6)) führt. Es mögen analoge Bezeichnungen wie damals gelten (Ω_1 , Ω_2 , M und so weiter), und u sei eine schwache Lösung der Potenzialgleichung, die global stetig ist sowie C^2 in den beiden Teilgebieten Ω_1 und Ω_2 , und sie lasse sich beiden Seiten C^1 in die Trennfläche M fortsetzen. Mit denselben Argumenten wie im Beweis von Satz 3.1 sieht man, dass u im Inneren von Ω_1 und Ω_2 jeweils starke Lösung ist. Sei jetzt v eine Testfunktion aus $C_0^\infty(\Omega)$. Dann ist

$$\begin{aligned} 0 &= \int_{\Omega} \sigma \nabla u \cdot \nabla v dx - \int_{\Omega} f v dx = \int_{\Omega_1} (\sigma_1 \nabla u_1 \cdot \nabla v - f v) dx + \int_{\Omega_2} \dots dx \\ &= \int_{\Omega_1} (-\nabla \cdot (\sigma_1 \nabla u_1) v - f v) dx + \int_{\partial\Omega_1} \sigma_1 \nabla u_1 \cdot n_{\Omega_1} v ds(x) - \int_{\Omega_2} \dots dx + \int_{\partial\Omega_2} \dots ds(x). \end{aligned}$$

Die Volumenintegrale fallen weg, weil u jeweils starke Lösung ist. Von den Randintegralen bleibt nur der Teil über die Mannigfaltigkeit M über, weil am übrigen Rand v verschwindet. Mit $n_M = n_{\Omega_1} = -n_{\Omega_2}$ (wie zuvor) ist somit

$$0 = \int_M (\sigma_1 \nabla u_1 - \sigma_2 \nabla u_2) \cdot n_M v ds(x),$$

und weil v beliebig ist, gilt wieder (wie im Beweis von Satz 3.1) punktweise

$$(\sigma_1 \nabla u_1 - \sigma_2 \nabla u_2) \cdot n_M = 0$$

auf M , und das stimmt in der Tat überein mit (2.6). Dies zeigt, dass unsere schwache Formulierung die Modellsituation sogar *besser* beschreibt als die Differentialgleichung, weil sie gewisse Unstetigkeiten zulässt und dabei mit dem Modell konsistent bleibt.

3.2. Diskretisierung

Wir haben es also mit einer Situation wie im Satz von Lax–Milgram zu tun, das heißt, auf einem Hilbertraum H haben wir eine stetige Linearform l und eine stetige, elliptische Bilinearform a , und wir suchen das eindeutige $u \in H$, mit dem gilt

$$a(u, v) = l(v) \quad \text{für alle } v \in H. \quad (3.4)$$

Natürlich ist die exakte Lösung in der Regel immer noch nicht bestimmbar.¹ Die Grundidee der Methode der Finiten Elemente besteht nun in folgender Idee:

Es sei $V_h \subset H$ ein endlichdimensionaler Teilraum², und wir approximieren u durch $u_h \in V_h$ mit

$$a(u_h, v_h) = l(v_h) \quad \text{für alle } v_h \in V_h. \quad (3.5)$$

Dass ein solches u_h stets existiert und eindeutig bestimmt ist, liefert wiederum der Satz von Lax–Milgram, denn alle Voraussetzungen übertragen sich auf den Raum V_h (die Vollständigkeit folgt, weil V_h isomorph zu einem \mathbb{R}^n ist). Andererseits ist (3.5) äquivalent zu einem endlichdimensionalen linearen Gleichungssystem, das man mit geeigneten Verfahren lösen kann. Da wir ja nunmal wissen, dass eine eindeutige Lösung existiert, muss die Matrix dieses Gleichungssystems regulär sein.

Es bleiben zwei Fragen zu klären:

1. Gibt es Grund zu der Annahme, dass u_h tatsächlich eine gute Approximation an u ist?
2. Was ist eine günstige Wahl von V_h ?

Zur ersten Frage gibt es folgendes Resultat:

Lemma 3.12 (Céa) *In dieser Situation gilt*

$$\|u - u_h\|_H \leq \frac{C}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_H,$$

wobei C und α die Stetigkeits- und Elliptizitätskonstanten der Bilinearform a sind.

¹Der Beweis des Satzes von Lax–Milgram ist nicht konstruktiv; er nutzt die Vollständigkeit des Raumes aus und konstruiert die Lösung als Grenzwert einer Cauchy-Folge.

² h ist zunächst nur ein Symbol und wird später mit Leben gefüllt.

In Worten: Die von uns gewählte Approximation ist höchstens um einen festen Faktor (der nur vom Problem abhängt) schlechter als die beste Approximation, die in dem gewählten Raum V_h überhaupt möglich ist.

Beweis (des Lemmas): Für $v_h \in V_h$ gilt $a(u_h, v_h) = l(v_h)$, aber auch $a(u, v_h) = l(v_h)$ weil ja auch $v_h \in H$ ist. Durch Subtraktion ergibt sich

$$a(u - u_h, v_h) = 0 \quad \text{für alle } v_h \in V_h.$$

Für jedes $v_h \in V_h$ gilt nun

$$\alpha \|u - u_h\|_H^2 \leq a(u - u_h, u - u_h) = a(u - u_h, u - v_h) + a(u - u_h, v_h - u_h).$$

Da aber auch $v_h - u_h \in V_h$ ist, ist der zweite Summand nach der obigen Überlegung null, also

$$\dots = a(u - u_h, u - v_h) \leq C \|u - u_h\|_H \|u - v_h\|_H,$$

und die Behauptung folgt nach Division durch $\|u - u_h\|_H$. \square

Zwei Warnungen am Rande:

- Bei der Auswahl des Raumes V_h muss man sicher stellen, dass sich die exakte Lösung u in V_h genügend gut approximieren lässt; sonst nützt das Lemma nichts.
- Das Lemma von Céa gilt nur in der Norm des Raumes H ; das ist bei uns die Sobolev-Norm $\|\cdot\|_{H^1}$. Fehlerabschätzungen in „schöneren“ Normen bekommt man nur mit deutlich mehr Aufwand.

Im Rahmen dieser Veranstaltung wollen wir hier jedoch nicht weiter ins Detail gehen.

Wenden wir uns jetzt der Frage nach der Wahl des Raumes V_h zu. Für die Zwecke dieser Veranstaltung beschränken wir uns auf Räume V_h , die wie folgt entstehen: Wir unterteilen unser (quaderförmiges) Rechengebiet Ω in eine endliche Anzahl kongruenter, achsenparalleler Würfel mit Kantenlänge h . Sodann legen wir fest, dass V_h aus all jenen Funktionen besteht, die global stetig sind und auf jedem dieser Würfel trilinear, also eine Darstellung der Form

$$u(x) = a + b_1x_1 + b_2x_2 + b_3x_3 + c_{12}x_1x_2 + c_{13}x_1x_3 + c_{23}x_2x_3 + dx_1x_2x_3$$

(auf jedem Würfel) haben. Wir nennen unsere Unterteilung ein *Gitter* und die kleinen Würfel *Gitterzellen* oder auch *Elemente*³. Die Eckpunkte der Würfel heißen *Gitterknoten*.

Übung 3.13 Zeigen Sie, dass der so definierte Raum V_h tatsächlich ein Teilraum von H ist.

³Die Bezeichnung „Elemente“ ist üblich, aber problematisch, weil man so eigentlich die in Kürze eingeführten Basisfunktionen nennt.

Den Rest des Abschnitts verbringen wir mit der Frage, wie das lineare Gleichungssystem nun tatsächlich aussieht.

Der endlichdimensionale Raum V_h besitzt eine Basis; ihre Elemente mögen $\varphi_1, \dots, \varphi_N$ heißen (wobei N die Dimension ist). Unsere diskretisierte Lösung ist durch (3.5) charakterisiert, wobei man aber als Testfunktion v_h nur alle Basisfunktionen φ_j betrachten muss, weil der Rest dann aus der Linearität folgt. Ferner ist die Lösung u_h eine Linearkombination der Basisfunktionen:

$$u_h = \sum_{i=1}^N u_i \varphi_i.$$

Das Gleichungssystem lautet somit

$$\sum_{i=1}^N u_i a(\varphi_i, \varphi_j) = l(\varphi_j) \quad \text{für alle } j,$$

also

$$Lu = b,$$

wobei (mit Notationsmissbrauch)

$$L = \begin{pmatrix} a(\varphi_1, \varphi_1) & \cdots & a(\varphi_1, \varphi_N) \\ \vdots & & \vdots \\ a(\varphi_N, \varphi_1) & \cdots & a(\varphi_N, \varphi_N) \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}, \quad b = \begin{pmatrix} l(\varphi_1) \\ \vdots \\ l(\varphi_N) \end{pmatrix}$$

ist. Zu Erinnerung: In unserem Fall ist die Bilinearform a gegeben durch

$$a(u, v) = \int_{\Omega} \sigma \nabla u \cdot \nabla v \, dx$$

(siehe (3.3)) und somit *symmetrisch*, das heißt $a(u, v) = a(v, u)$, so dass wir feststellen, dass auch die Matrix L symmetrisch ist. Auch ist die Matrix positiv definit, denn ist v irgendein Vektor und $v_h = \sum_i v_i \varphi_i \in V_h$, so ist

$$v^t L v = \sum_{i=1}^N \sum_{j=1}^N v_i a(\varphi_i, \varphi_j) v_j = a(v_h, v_h) \geq \alpha \|v_h\|_H^2,$$

und weil auf dem endlichdimensionalen Raum V_h alle Normen äquivalent sind, ist $\|v_h\|_H \geq c|v|$ mit einer geeigneten Konstanten c .

Die tatsächliche Gestalt der Matrix L hängt von der Wahl der Basis ab. Es ist zweckmäßig, die Basis so zu wählen, dass möglichst viele Einträge der Matrix L verschwinden, die Matrix also *dünn besetzt* ist, weil viele Lösungsverfahren für lineare Gleichungssysteme dann sehr viel schneller sind. Man kann nachrechnen, dass eine Funktion $v_h \in V_h$ eindeutig bestimmt ist durch ihre Werte in allen Knoten. Wenn wir alle Knoten, auf

denen v_h nicht auf Grund der Dirichlet-Randbedingung verschwinden muss, von x_1 bis x_N durchnummerieren, so ist eine Basis $(\varphi_i)_{i=1,\dots,N}$ gegeben durch die Vorschrift

$$\varphi_i(x_j) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Für diesen Fall wollen wir jetzt die Einträge der Steifigkeitsmatrix berechnen, wobei wir der Einfachheit halber annehmen wollen, dass σ auf jeder Gitterzelle konstant ist. Es ist

$$L_{ij} = a(\varphi_i, \varphi_j) = \int_{\Omega} \sigma \nabla \varphi_i \cdot \nabla \varphi_j \, dx = \sum_T \int_T \sigma \nabla \varphi_i \cdot \nabla \varphi_j \, dx = \sum_T \sigma_T \int_T \nabla \varphi_i \cdot \nabla \varphi_j \, dx,$$

wobei T alle Gitterzellen durchläuft und σ_T der (konstante) Wert von σ auf T ist. Es genügt also, die lokalen Integrale $\int_T \nabla \varphi_i \cdot \nabla \varphi_j \, dx$ zu berechnen. Diese sind offenbar nur dann von null verschieden, wenn sowohl x_i als auch x_j Eckpunkte von T sind, da sonst φ_i oder φ_j auf T identisch verschwindet. Aus Symmetriegründen tauchen nur die folgenden vier Fälle auf:

- $x_i = x_j$ (also $i = j$);
- x_i und x_j liegen eine Kante entfernt;
- x_i und x_j liegen eine Flächendiagonale entfernt;
- x_i und x_j liegen eine Raumdiagonale entfernt.

Beispielhaft berechnen wir das Integral für den Fall, dass x_i und x_j eine Flächendiagonale voneinander entfernt liegen. Ohne Beschränkung der Allgemeinheit ist

$$T = (0, h) \times (0, h) \times (0, h),$$

und die Eckpunkte sind

$$x_i = (h, h, h)^t, \quad x_j = (0, 0, h)^t.$$

Dann haben die Basisfunktionen φ_i und φ_j auf T folgende Darstellung

$$\varphi_i(x) = \frac{x_1 x_2 x_3}{h^3}, \quad \varphi_j(x) = \frac{(h - x_1)(h - x_2)x_3}{h^3},$$

und die Gradienten sind

$$\nabla \varphi_i(x) = \frac{1}{h^3} \begin{pmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{pmatrix}, \quad \nabla \varphi_j(x) = \frac{1}{h^3} \begin{pmatrix} (x_2 - h)x_3 \\ (x_1 - h)x_3 \\ (x_1 - h)(x_2 - h) \end{pmatrix}.$$

Somit ergibt sich

$$\begin{aligned}
 & \int_T \nabla \varphi_i \cdot \nabla \varphi_j \, dx \\
 &= \frac{1}{h^6} \int_0^h \int_0^h \int_0^h \left(x_2(x_2 - h)x_3^2 + x_1(x_1 - h)x_3^2 + x_1(x_1 - h)x_2(x_2 - h) \right) dx_1 \, dx_2 \, dx_3 \\
 &= \frac{1}{h^6} \int_0^h \int_0^h \int_0^h \left(x_2^2 x_3^2 - h x_2 x_3^2 + x_1^2 x_3^2 - h x_1 x_3^2 + x_1^2 x_2^2 - h x_1 x_2^2 - h x_1^2 x_2 + h^2 x_1 x_2 \right) dx_1 \, dx_2 \, dx_3 \\
 &= \frac{1}{h^6} \int_0^h \int_0^h \left(h x_2^2 x_3^2 - h^2 x_2 x_3^2 - \frac{1}{6} h^3 x_3^2 - \frac{1}{6} h^3 x_2^2 + \frac{1}{6} h^4 x_2 \right) dx_2 \, dx_3 \\
 &= \frac{1}{h^6} \int_0^h \left(-\frac{1}{6} h^4 x_3^2 - \frac{1}{6} h^4 x_3^2 + \frac{1}{36} h^6 \right) dx_3 = \frac{1}{h^6} \int_0^h \left(-\frac{1}{3} h^4 x_3^2 + \frac{1}{36} h^6 \right) dx_3 \\
 &= \frac{1}{h^6} \left(-\frac{1}{9} h^7 + \frac{1}{36} h^7 \right) = -\frac{h}{12}.
 \end{aligned}$$

Die anderen Fälle lassen sich analog ausrechnen. Es ergeben sich insgesamt folgende Werte (in derselben Reihenfolge wie zuvor):

- $h/3$,
- 0 ,
- $-h/12$,
- $-h/12$.

(Für zwei Raumdimensionen hat man analog die drei Werte $2/3$, $-1/6$, $-1/3$; für eine Raumdimension hat man analog die zwei Werte $1/h$, $-1/h$. Man beachte, dass h in der $(d-2)$ -ten Potenz vorkommt, wobei d die Raumdimension ist.)

Aus diesen Werten setzt man die Steifigkeitsmatrix zusammen. Man beachte jedoch folgende Dinge:

- Die Terme sind gegebenenfalls noch mit Faktoren zu multiplizieren: Sind die Punkte x_i und x_j beispielsweise eine Flächendiagonale voneinander entfernt, so gibt es *zwei* Gitterzellen, die an diese Fläche grenzen. Für jede erhält man den Beitrag $-h/12$, also ist der Gesamteintrag L_{ij} in diesem Fall gleich $-h/6$.
- Die Faktoren, mit denen man multipliziert, ändern sich gegebenenfalls am Rand des Rechengebiets: Sind x_i und x_j eine Flächendiagonale voneinander entfernt und liegt diese Fläche am Rand des Rechengebiets (im Bereich der Neumann-Randbedingungen, also Γ_2), so gibt es wiederum nur *eine* Zelle, die an diese Fläche grenzt. Der Matrix-Eintrag L_{ij} ist in dem Falle also wieder nur $-h/12$.

- Falls σ nicht als auf jeder Gitterzelle konstant angenommen werden soll, sind die Integrale entsprechend aufwändiger zu berechnen. Gegebenenfalls verwendet man Quadraturformeln.
- Wenn die Zellen keine Würfel sondern Quader sind, hat man statt h drei Gitterweiten h_1 , h_2 und h_3 . In diesem Fall haben die obigen Terme kompliziertere Darstellungen; insbesondere ist der zweite Wert *nicht* mehr gleich null.

3.3. Inhomogene Randbedingungen

Wir verallgemeinern das Problem (3.1) jetzt, indem wir auf Γ_1 inhomogene Dirichlet-Bedingungen erlauben, also

$$-\nabla \cdot (\sigma(x)\nabla u(x)) = f(x) \quad \text{in } \Omega, \quad (3.6a)$$

$$u(x) = g(x) \quad \text{auf } \Gamma_1, \quad (3.6b)$$

$$\nabla u(x) \cdot n(x) = 0 \quad \text{auf } \Gamma_2. \quad (3.6c)$$

Ein üblicher Trick besteht darin, *irgendeine* Funktion $u_0 \in H^1(\Omega)$ zu bestimmen, die die Randbedingungen (3.6b) und (3.6c) erfüllt (jedoch nicht notwendig die Differenzialgleichung (3.6a)) und dann das sich ergebende Randwertproblem für die Differenz $\tilde{u} = u - u_0$ zu betrachten:

$$-\nabla \cdot (\sigma(x)\nabla \tilde{u}(x)) = f(x) + \nabla \cdot (\sigma(x)\nabla u_0(x)) \quad \text{in } \Omega, \quad (3.7a)$$

$$\tilde{u}(x) = 0 \quad \text{auf } \Gamma_1, \quad (3.7b)$$

$$\nabla \tilde{u}(x) \cdot n(x) = 0 \quad \text{auf } \Gamma_2. \quad (3.7c)$$

Man erhält also ein Problem der Form (3.1) mit geänderter rechter Seite. Dies löst man dann wie zuvor beschrieben und berechnet dann $u = \tilde{u} + u_0$.

Praktisch geht man insgesamt wie folgt vor:

1. Man überdeckt wie beschrieben das Rechengebiet mit einem Gitter.
2. Man stellt die Steifigkeitsmatrix und die rechte Seite auf, wobei man auch den Komponenten auf dem Dirichlet-Rand jeweils eine Komponente zugesteht (das heißt, den Dirichlet-Rand zunächst nicht berücksichtigt).
3. Jetzt wählt man den Vektor, der in den Komponenten der Dirichlet-Ränder die geforderten Randwerte annimmt und sonst null ist. Dies entspricht der Funktion u_0 . Diesen Vektor multipliziert man an die Matrix. Das Ergebnis subtrahiert man von der ursprünglichen rechten Seite.
4. Danach ersetzt man in der Steifigkeitsmatrix ersetzt in denjenigen Komponenten, die den Dirichlet-Rand-Knoten entsprechen, die Diagonalelemente durch Einsen und alle anderen Komponenten dieser Zeilen *und Spalten* durch Nullen. Ferner ersetzt man die entsprechenden Komponenten der rechten Seite durch Nullen. Somit

3. Finite Elemente für die Potenzialgleichung

hat man das Problem mit homogenen Randwerten korrekt diskretisiert, und die Matrix bleibt symmetrisch positiv definit. Die Nullen in den Spalten ändern nichts an der Lösung, weil die entsprechenden Komponenten der Lösung ja sowieso null sind.

5. Zur Lösung dieses Problems addiert man dann wiederum die Funktion u_0 hinzu.

4. Wärmeausbreitung im Gewebe

4.1. Wärmequelle durch elektrische Leistung

Mikroskopisch betrachtet besteht elektrischer Strom aus vielen kleinen Ladungsträgern. In einem Volumen V möge es n Ladungsträger geben; ihre Ladung sei q_i und ihre Geschwindigkeit \vec{v}_i (für $i = 1, \dots, n$). Der elektrische Strom ergibt sich aus dem Produkt der Ladungen und Geschwindigkeiten, aufsummiert für alle Ladungsträger. Makroskopisch betrachtet haben wir eine elektrische Stromdichte \vec{J} (vergleiche Kapitel 2), und der elektrische Strom in dem Volumen V ergibt sich aus dem Integral von \vec{J} über V . Da die Ladungsträger sehr klein und sehr viele sind, schreiben wir

$$\int_V \vec{J} d\vec{x} = \sum_{i=1}^n q_i \vec{v}_i.$$

Andererseits wirkt auf die Ladungsträger jeweils die Kraft $\vec{F}_i = q_i \vec{E}$ (wobei wir V als hinreichend klein annehmen, so dass E dort näherungsweise konstant ist). In der Zeit t führt dies zu der Arbeit (Energie) $W_i = \vec{F}_i \cdot \vec{v}_i t$. Im gesamten Volumen V verrichten die Ladungsträger also in der Zeit t die Arbeit

$$W = \sum_{i=1}^n W_i = \sum_{i=1}^n \vec{F}_i \cdot \vec{v}_i t = \sum_{i=1}^n q_i \vec{E} \cdot \vec{v}_i t = \vec{E} \cdot \sum_{i=1}^n q_i \vec{v}_i t = \vec{E} t \cdot \int_V \vec{J} d\vec{x}$$

Die Leistung ist Arbeit pro Zeit, ergibt sich die Leistung

$$P = \frac{W}{t} = \vec{E} \cdot \int_V \vec{J} d\vec{x} \quad (4.1)$$

und somit die *elektrische Leistungsdichte* (nämlich Leistung pro Volumen)

$$p = \vec{E} \cdot \vec{J} = \sigma |\nabla \varphi|^2. \quad (4.2)$$

Auf den Elektroden ist die Leistungsdichte null, weil \vec{E} verschwindet (weil φ konstant ist).

Die SI-Einheiten sind

$$\begin{aligned} [\vec{v}] &= \text{m/s}, \\ [t] &= \text{s}, \\ [W] &= \text{kg m}^2/\text{s}^2 = \text{N m} = \text{J}, \\ [P] &= \text{kg m}^2/\text{s}^3 = \text{J/s} = \text{W}, \\ [p] &= \text{kg/m s}^3 = \text{W/m}^3. \end{aligned}$$

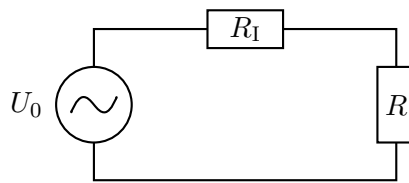


Abbildung 4.1.: Ersatzschaltbild für den elektrischen Generator und das Gewebe. U_0 ist die eingestellte elektrische Spannung, R_I der innere Widerstand des Generators und R der Gewebewiderstand.

Will man die Gesamtleistung P_Ω auf der Interessenregion Ω berechnen, so muss man p aufintegrieren. (Die Gleichung (4.1) gilt näherungsweise für kleine Volumen, auf denen \vec{E} als konstant angenommen werden kann):

$$P_\Omega = \int_\Omega \vec{E} \cdot \vec{J} d\vec{x} = \int_\Omega \sigma |\nabla\varphi|^2 d\vec{x} \quad (4.3)$$

(die zweite Gleichheit folgt aus den Formeln aus Kapitel 2).

4.2. Skalierung der elektrischen Leistung

Wie in Kapitel 2 erwähnt, haben wir die Randwerte ± 1 auf den Elektroden willkürlich gewählt und müssen φ (und damit p) noch korrekt skalieren. Am Generator stellt man eine Leistung P_{setup} ein, die jedoch nicht unbedingt der tatsächlich eingebrachten effektiven Leistung P_{eff} entspricht. Vielmehr hängt sie vom elektrischen Widerstand R , den der elektrische Strom auf dem Weg von der einen zur anderen Elektrode (normalerweise also im Lebergewebe) überwinden muss. Ist beispielsweise der Widerstand unendlich groß (etwa weil an den Elektroden gar nichts angeschlossen ist), so fließt überhaupt kein Strom, also ist auch die Leistung dann null. Ist der Widerstand null (zum Beispiel wenn man die beiden Elektroden durch einen Supraleiter verbindet), so gibt es einen Kurzschluss, und es wird eventuell im Generator warm, aber nicht am Applikator.

Wir gehen von der Annahme aus, dass sich der Generator bei fest gewählter Einstellung verhält wie eine konstante Spannungsquelle zusammen mit einem *inneren Widerstand*. Die Gesamtsituation wird dann durch das *Ersatzschaltbild* in Abbildung 4.1 wiedergegeben.

Der innere Widerstand ist am Typenschild des Generators angegeben. Der Gewebewiderstand ergibt sich aus den bekannten Beziehungen $R = U_{\text{el}}/I$ und $P_\Omega = U_{\text{el}}I$, denn wir wissen, dass die Leistung gleich P_Ω wäre, wenn unsere Annahme der Elektrodenpotenziale ± 1 richtig wäre, und das wäre eine Spannung von $U_{\text{el}} = 2\text{ V}$ (in Worten: zwei Volt) zwischen den beiden Elektroden. Somit gilt

$$R = \frac{2\text{ V}}{I} = \frac{2\text{ V}}{P_\Omega/2\text{ V}} = \frac{4\text{ V}^2}{P_\Omega}. \quad (4.4)$$

Der Gesamtwiderstand im Stromkreis des Ersatzschaltbildes ist somit $R + R_1$, und damit ist der Strom I gegeben durch

$$I = \frac{U_0}{R + R_1}.$$

Die elektrische Spannung, die tatsächlich zwischen den beiden Elektroden liegt, ergibt sich somit zu $RI = RU_0/(R + R_1)$, und damit ist die Leistung P_{eff} , die ins Gewebe eingebracht wird, gegeben durch

$$P_{\text{eff}} = \frac{RU_0^2}{(R + R_1)^2} \quad (4.5)$$

Hätte man also das elektrische Potenzial korrekt skaliert, so müsste sich die Leistungsdichte zu P_{eff} aufintegrieren. In Wirklichkeit integriert sie sich aber zu P_Ω auf (so war nämlich P_Ω definiert), also muss man die Leistungsdichte mit dem Faktor P_{eff}/P_Ω skalieren. (Das entspricht einer Skalierung des elektrischen Potenzials mit dem Faktor $\sqrt{P_{\text{eff}}/P_\Omega}$.) Die tatsächliche (skalierte) Leistungsdichte, die wir mit Q_{rf} bezeichnen wollen, ergibt sich also zu

$$Q_{\text{rf}}(\vec{x}) = \frac{P_{\text{eff}}}{P_\Omega} p(\vec{x}). \quad (4.6)$$

Es bleibt noch die Frage, wie groß U_0 ist. Zur Erinnerung: U_0 ist die konstante Spannung, die im Generator erzeugt wird, wenn der Regler, mit dem man vermeintlich die Leistung einzustellen meint, auf P_{setup} eingestellt ist. Die tatsächlich erzielte Leistung P_{eff} ergibt sich dann aus (4.5). Wir nehmen an, dass die Beschriftung des Reglers in der Weise mit der Realität zusammenhängt, als dass P_{setup} die *maximal erreichbare* Leistung ist, also der größtmögliche Wert, den (4.5) (als Funktion von R) liefern kann. Durch Nullsetzen der Ableitung stellt man fest, dass dies bei $R = R_1$ der Fall ist und somit

$$P_{\text{setup}} = \frac{U_0^2}{4R_1}$$

ist. Somit haben wir die U_0 -freie Formel

$$P_{\text{eff}} = \frac{4P_{\text{setup}}RR_1}{(R + R_1)^2} \quad (4.7)$$

gefunden.

Wir fassen zusammen, wie man die skalierte Leistungsdichte Q_{rf} bestimmt:

1. Man berechne in jedem Punkt die (unskalierte) elektrische Leistungsdichte $p(\vec{x})$ gemäß (4.2).
2. Man integriere sie auf zu P_Ω , siehe (4.3).
3. Man bestimme den Gewebewiderstand mittels (4.4).
4. Man bestimme die effektive Leistung P_{eff} gemäß (4.7).
5. Man skaliere die Leistungsdichte wie in (4.6).

4.3. Wärmeleitung

Wir bezeichnen mit $T(t, \vec{x})$ die Temperatur im Punkt \vec{x} zum Zeitpunkt t . Die thermische Energie zum Zeitpunkt t in einem Kontrollvolumen V ist gegeben durch

$$E_V = \int_V \rho(\vec{x}) e(T(t, \vec{x}), \vec{x}) d\vec{x},$$

wobei ρ die *Massendichte* des Gewebes (Masse pro Volumen) e die *massenspezifische thermische Energiedichte* (thermische Energie pro Masse) ist. Letztere hängt von hauptsächlich von der Temperatur ab (je wärmer, desto energiereicher), aber auch vom Ort (Gewebetyp). Für die zeitliche Änderung der Energie E_V lässt sich eine Bilanzgleichung formulieren: man kann sie zum einen durch Ableiten der obigen Integralformel bestimmen; zum anderen ergibt sie sich aber auch durch die über den Rand von V hinein- und herausfließende Energie sowie durch die Energiequelle in Form der Leistungsdichte Q_{rf} (siehe Abschnitt 4.2). Es gilt also

$$\int_V \rho(\vec{x}) \frac{d}{dt} e(T(t, \vec{x}), \vec{x}) d\vec{x} = - \int_{\partial V} \vec{j}(t, \vec{x}) \cdot \vec{n} ds(\vec{x}) + \int_V Q_{\text{rf}}(\vec{x}) d\vec{x}.$$

Dabei ist \vec{j} eben der Energiefluss. Die thermische Energie fließt stets in Richtung des steilsten Temperaturabstiegs, also gilt

$$\vec{j}(t, \vec{x}) = -\lambda(\vec{x}) \nabla T(t, \vec{x})$$

mit einem skalaren Faktor $\lambda(\vec{x})$, der als *Wärmeleitfähigkeit* bezeichnet wird. Ferner lässt sich die Zeitableitung der Energiedichte mit der Kettenregel ausrechnen:

$$\frac{d}{dt} e(T(t, \vec{x}), \vec{x}) = \partial_T e(T(t, \vec{x}), \vec{x}) \partial_t T(t, \vec{x}).$$

Hier kommt die Ableitung $\partial_T e$ der massenspezifischen thermischen Energiedichte nach der Temperatur vor; diese Größe heißt *massenspezifische Wärmekapazität* und wird mit $c(T(t, \vec{x}), \vec{x})$ bezeichnet. Die Abhängigkeit diese Größe von der Temperatur vernachlässigen wir zunächst. Somit ergibt sich die Bilanzgleichung

$$\begin{aligned} \int_V \rho(\vec{x}) c(\vec{x}) \partial_t T(t, \vec{x}) dx &= \int_{\partial V} \lambda(\vec{x}) \nabla T(t, \vec{x}) \cdot \vec{n} ds(\vec{x}) + \int_V Q_{\text{rf}}(\vec{x}) dx \\ &= \int_V (\nabla \cdot (\lambda(\vec{x}) \nabla T(t, \vec{x})) + Q_{\text{rf}}(\vec{x})) dx. \end{aligned}$$

Wie in Kapitel 2 ergibt sich unter der Annahme, dass alle Integranden stetig sind, die partielle Differenzialgleichung

$$\rho c \partial_t T - \nabla \cdot (\lambda \nabla T) = Q_{\text{rf}}. \quad (4.8)$$

Sie ist parabolisch und inhomogen.

Die SI-Einheiten sind:

$$\begin{aligned}
 [E_V] &= \text{kg m}^2/\text{s}^2 = \text{J}, \\
 [\rho] &= \text{kg}/\text{m}^3, \\
 [e] &= \text{m}^2/\text{s}^2 = \text{J}/\text{kg}, \\
 [\vec{j}] &= \text{kg}/\text{s}^3, \\
 [Q_{\text{rf}}] = [p] &= \text{kg}/\text{m s}^3 = \text{W}/\text{m}^3, \\
 [T] &= \text{K}, \\
 [\lambda] &= \text{kg m}/\text{K s}^3, \\
 [c] &= \text{m}^2/\text{K s}^2.
 \end{aligned}$$

Die Gleichung (4.8) muss noch in geeigneter Weise mit Rand- und Anfangsbedingungen versehen werden:

- In der Regel wird bei der Radiofrequenzablation mit gekühlten Applikatoren gearbeitet, das heißt, durch den gesamten Applikator fließt eine Kühlflüssigkeit. Damit soll der Effekt der Gewebeaustrocknung in Folge von Zellwasserverdampfung verringert werden. Um dies zu modellieren, führt man den Rand des Applikators als inneren Rand des Rechengebiets ein und setzt dort die Dirichlet-Bedingung $T = T_{\text{pr}}$ fest.
- Auf dem (künstlichen) äußeren Rand Γ_{out} kann man (analog zur Potenzialgleichung) näherungsweise annehmen, dass kein Wärmefluss über den Rand erfolgt, also $\vec{j} \cdot \vec{n} = 0$ ist, also $\nabla T \cdot \vec{n} = 0$. Man hat also wieder homogene Neumann-Randbedingungen. (Alternativ kann man annehmen, dass auf dem Rand die Temperatur konstant gleich der Körpertemperatur T_{body} (in Höhe von etwa 310 K), was dann wiederum zu Dirichlet-Randbedingungen führt.)
- Außerdem braucht man Anfangsbedingungen, also Bedingungen zum Zeitpunkt $t = 0$. Es ist sinnvoll, hierzu anzunehmen, dass die Temperatur zu Beginn der Behandlung überall konstant gleich der Körpertemperatur ist: $T(0, \vec{x}) = T_{\text{body}}$ für alle $x \in \Omega$.

Wir fassen zusammen: Die (zeitabhängige) Temperaturverteilung wird modelliert durch das parabolische Anfangs-Randwertproblem

$$\begin{aligned}
 \rho c \partial_t T - \nabla \cdot (\lambda \nabla T) &= Q_{\text{rf}}, & x \in \Omega \setminus \overline{\Omega_{\text{pr}}}, & t > 0, \\
 T &= T_{\text{body}}, & x \in \Omega \setminus \overline{\Omega_{\text{pr}}}, & t = 0, \\
 T &= T_{\text{pr}}, & x \in \overline{\Omega_{\text{pr}}}, & t > 0, \\
 \nabla T \cdot \vec{n} &= 0, & x \in \Gamma_{\text{out}}, & t > 0.
 \end{aligned}$$

5. Finite Elemente für die Wärmeleitungsgleichung

Wir haben also ein Problem zu lösen, das so etwa von folgender Form ist:

$$\mu(x)\partial_t u(t, x) - \nabla \cdot (\lambda(x)\nabla u(x)) = f(x) \quad \text{in } \Omega, \quad (5.1a)$$

$$u(t, x) = 0 \quad \text{auf } \Gamma_1, \quad (5.1b)$$

$$\nabla u(t, x) \cdot n(x) = 0 \quad \text{auf } \Gamma_2, \quad (5.1c)$$

$$u(0, x) = 0 \quad \text{in } \Omega, \quad (5.1d)$$

wobei die Bezeichnungen analog zu Kapitel 3 sind und auch die Transformation auf homogene Rand- und Anfangswerte wir dort durchgeführt werden kann.

Ein übliches Vorgehen zur numerischen Behandlung zeitabhängiger Problem besteht im *schrittweisen* Lösen der Differenzialgleichung. Man wählt also eine Zeitschrittweite $\tau > 0$ und berechnet Näherungslösungen $u^n(x)$ für $u(n\tau, x)$. Die Zeitableitung $\partial_t u(n\tau, x)$ wird dabei durch einen geeigneten Differenzenquotienten angenähert. Verschiedene Wahlmöglichkeiten und ihre Eigenschaften werden im zweiten Teil der Veranstaltung erläutert; wir beschränken uns hier auf die Wahl

$$\partial_t u(n\tau, x) \approx \frac{u^n(x) - u^{n-1}(x)}{\tau}.$$

Dies ist das sogenannte *implizite Euler-Verfahren*. Ist die Näherungslösung zum $(n-1)$ -ten Zeitschritt, also u^{n-1} bekannt, so lautet die Gleichung für u^n somit

$$\frac{\mu(x)}{\tau}(u^n(t, x) - u^{n-1}(t, x)) - \nabla \cdot (\lambda(x)\nabla u^n(x)) = f(x).$$

Wie zuvor multiplizieren wir mit einer Testfunktion v und wenden dann partielle Integration (im Ort) auf den elliptischen Teil an. Dies ergibt

$$\begin{aligned} 0 = & \int_{\Omega} \lambda \nabla u^n \cdot \nabla v - \int_{\Gamma_1} \lambda \nabla u^n \cdot n v \, ds(x) - \int_{\Gamma_2} \lambda \nabla u^n \cdot n v \, ds(x) \\ & + \int_{\Omega} \frac{\mu}{\tau} u^n v \, dx - \int_{\Omega} \left(\frac{\mu}{\tau} u^{n-1} + f \right) v \, dx. \end{aligned}$$

Die Randintegrale entfallen aus denselben Gründen wie vorher. Sortieren wir die restlichen Terme danach, ob u^n vorkommt oder nicht, so erhalten wir

$$\int_{\Omega} \lambda \nabla u^n \cdot \nabla v + \int_{\Omega} \frac{\mu}{\tau} u^n v \, dx = \int_{\Omega} \left(\frac{\mu}{\tau} u^{n-1} + f \right) v \, dx.$$

Tabelle 5.1.: Summanden, aus denen sich die Einträge der Steifigkeitsmatrix (S) beziehungsweise der Massematrix (M) zusammensetzen.

Entfernung von x_i und x_j	3D		2D		1D	
	S	M	S	M	S	M
$i = j$	$\frac{h}{3}$	$\frac{h^3}{27}$	$\frac{2}{3}$	$\frac{h^2}{9}$	$\frac{1}{h}$	$\frac{h}{3}$
Kante	0	$\frac{h^3}{54}$	$-\frac{1}{6}$	$\frac{h^2}{18}$	$-\frac{1}{h}$	$\frac{h}{6}$
Flächendiagonale	$-\frac{h}{12}$	$\frac{h^3}{108}$	$-\frac{1}{3}$	$\frac{h^2}{36}$		
Raumdiagonale	$-\frac{h}{12}$	$\frac{h^3}{216}$				

Die können wir wieder in die abstrakte Form $a(u, v) = l(v)$ bringen, wobei $u = u^n$ die gesuchte Näherungslösung im n -ten Zeitschritt ist und eben

$$a(u, v) = \int_{\Omega} \lambda \nabla u \cdot \nabla v + \int_{\Omega} \frac{\mu}{\tau} uv \, dx, \quad l(v) = \int_{\Omega} \left(\frac{\mu}{\tau} u^{n-1} + f \right) v \, dx.$$

Man überprüft leicht, dass die Voraussetzungen des Satzes von Lax–Milgram (Satz 3.3) hierfür erfüllt sind (mit analoger Definition des Hilbertraumes H wie in Kapitel 3). Die Diskretisierung verläuft ebenfalls analog wie in Kapitel 3. Die Einträge der Matrix des zu lösenden Gleichungssystems haben wieder die Form $a(\varphi_i, \varphi_j)$; wegen der zweigeteilten Form der Bilinearform a , bietet es sich jedoch an, die Matrix als $L + M$ zu schreiben, wobei die Einträge l_{ij} und m_{ij} durch

$$l_{ij} = \int_{\Omega} \lambda \nabla \varphi_i \cdot \nabla \varphi_j \, dx, \quad m_{ij} = \int_{\Omega} \frac{\mu}{\tau} \varphi_i \varphi_j \, dx$$

gegeben sind. L ist die schon bekannte Steifigkeitsmatrix. M hingegen ist bisher noch nicht aufgetaucht und wird *Massematrix* genannt. Die Massematrix ist ebenfalls symmetrisch und positiv definit, also auch die Summe $L + M$. Zur Bestimmung der Einträge m_{ij} der Massematrix kann man analog zur Steifigkeitsmatrix vorgehen: Ist μ auf jeder Gitterzelle konstant, so reduziert sich das Problem auf die Berechnung von

$$\int_T \varphi_i \varphi_j \, dx$$

für eine Gitterzelle T . Es treten dieselben Fälle auf wie bei der Steifigkeitsmatrix. Die Werte sind in Tabelle 5.1 aufgeführt; der Übersichtlichkeit halber sind auch die entsprechenden Werte der Steifigkeitsmatrix noch einmal mit aufgeführt.

A. Klassifizierung von partiellen Differenzialgleichungen zweiter Ordnung

In diesem Kapitel werden kurz die drei wichtigsten Klassen von partiellen Differenzialgleichungen zweiter Ordnung eingeführt und wichtige Eigenschaften ohne Beweis zusammengefasst. Die Prototypen der drei Klassen werden im jeweils einfachsten Fall analytisch gelöst, um zu demonstrieren, welche Rand- und Anfangsbedingungen jeweils zu wohlgestellten Problemen führen.

Das Kapitel ersetzt in keinem Fall den entsprechenden Abschnitt aus einer Vorlesung über partielle Differenzialgleichungen.

A.1. Klassifizierung

Es sei Ω ein Gebiet im \mathbb{R}^d . Auf Ω sei ein linearer Differenzialoperator L der Form

$$Lu(x) = - \sum_{i,j} a_{ij}(x) \partial_{ij} \partial_i u(x) + \sum_i b_i(x) u(x) + c(x) u(x)$$

gegeben. Wenn u zweimal stetig differenzierbar ist, kann angenommen werden, dass $a_{ij} = a_{ji}$ ist für alle i und j . Die Koeffizientenmatrix $A = (a_{ij})$ ist somit symmetrisch. Die Lineare Algebra lehrt, dass A reell diagonalisierbar ist. Sei ferner b der Vektor der b_i .

Definition A.1 *In dieser Situation heißt der Differenzialoperator L*

- elliptisch, falls alle Eigenwerte von A positiv sind (wenn als A positiv definit ist);
- hyperbolisch, falls ein Eigenwert von A negativ und alle anderen positiv sind;
- parabolisch, falls ein Eigenwert von A null ist und alle anderen positiv sind, sowie $b \cdot v \neq 0$ ist, wobei v der Eigenvektor von A zum Eigenwert 0 ist.

Die Bezeichnungen lehnen sich an die Kegelschnitte an. Man kann ausrechnen, dass die Lösungsmenge der zweidimensionalen algebraischen Gleichung $-x^t A x + b^t x + c = 0$ jeweils eine Ellipse, Hyperbel oder Parabel ist, wenn die (symmetrische) Matrix A die entsprechende Bedingung aus der Definition erfüllt.

Bei hyperbolischen und parabolischen Differenzialgleichungen ist eine Richtung im d -dimensionalen Raum vor den anderen ausgezeichnet, nämlich durch den Eigenvektor zum negativen beziehungsweise verschwindenden Eigenwert. Diese Differenzialgleichungen rühren in der Regel aus *zeitabhängigen* physikalischen Problemen her. Da man Zeit-

und Raumkoordinaten üblicherweise getrennt voneinander behandelt, ist der ausgezeichnete Eigenvektor dann meist der Einheitsvektor e_1 (wenn die Zeit als die erste Komponente betrachtet wird). Elliptische Gleichungen rühren hingegen in der Regel aus *stationären* Problemen her, also solchen, in denen keine Zeit vorkommt.

Prototypen für Differenzialgleichungen dieser Klassen sind:

- die Potenzialgleichung $-\Delta u = 0$ als elliptische;
- die Wellengleichung $\partial_{tt}u - \Delta u = 0$ als hyperbolische; und
- die Wärmeleitungsgleichung $\partial_t u - \Delta u = 0$ als parabolische Gleichung.

A.2. Typische Eigenschaften

Hyperbolische Gleichungen treten überall dort auf, wo Bewegungen im Spiel sind und alle Diffusionseffekte (wie zum Beispiel Reibung) vernachlässigt werden: Ausbreitung von elektromagnetischen Wellen; Ausbreitung von Schall- oder Wasserwellen; Fließen eines Gewässers; Luftströmung um Flugzeugtragflügel; schwingende Saite und so weiter.

Parabolische Gleichungen treten auf, sobald Diffusionseffekte betrachtet werden: Wärmeausbreitung; Vermischung von Flüssigkeiten oder Gasen; und so weiter. Wenn sich die „äußeren Bedingungen“ in der Zeit nicht ändern, nähert sich das System einem stationären Zustand an, in dem die Zeitableitungen immer kleiner werden. Da durch Weglassen der Zeitabhängigkeit aus einer parabolischen Gleichung eine elliptische wird, sind diese eng verwandt. So treten in der Tat elliptische Gleichungen über all dort auf, wo davon ausgegangen werden kann, dass sich durch entsprechende Diffusionseffekte bereits der stationäre Zustand eingestellt hat.

Alle diese Gleichungen müssen mit Anfangs- und / oder Randwerten versehen werden, um wohlgestellte Probleme zu ergeben. Welche Anfangs- und Randbedingungen passend sind, hängt jedoch vom Typ ab:

- Elliptische Probleme sind wohlgestellt, wenn sie auf einem beschränkten Gebiet $\Omega \subset \mathbb{R}^d$ betrachtet werden und man auf dem gesamten Rand $\partial\Omega$ gewisse Bedingungen vorgibt. Dabei hat man die Wahl zwischen verschiedenen möglichen Randbedingungen: Man kann beispielsweise Funktionswerte auf dem Rand vorschreiben oder Ableitungen in Normalenrichtung oder Kombinationen aus diesen Dingen. Auf unbeschränkten Gebieten kann man sie nur mit speziellen Zusatzbedingungen betrachten.
- Parabolische Gleichungen sind wohlgestellt, wenn sie auf einem Gebiet der Form $\Omega \times (0, T)$ im Orts-Zeit-Raum betrachtet werden, wobei Ω beschränkt ist. Sie brauchen Anfangsbedingungen auf $\Omega \times \{0\}$ sowie Randbedingungen auf $\partial\Omega \times (0, T)$. Bei den Randbedingungen hat man eine ähnliche Wahlmöglichkeit wie bei elliptischen Problemen; für die Anfangsbedingungen muss man Funktionswerte vorschreiben. Dies gilt, wenn das Vorzeichen vor der Zeitableitung entgegengesetzt zu den Vorzeichen vor den zweiten Ortsableitungen ist (wie im Prototyp $\partial_t u - \Delta u = 0$).

Ein zusätzliches Fordern von „Endbedingungen“ auf $\Omega \times \{T\}$ macht die Gleichung unlösbar. Auch kann man die Anfangsbedingungen nicht durch Endbedingungen ersetzen, das heißt man kann parabolische Gleichungen nicht rückwärts in der Zeit rechnen. Der Lösungswert in einem gegebenen Punkt (x, t) im Zeit-Orts-Raum hängt von allen Anfangswerten auf ganz Ω sowie von allen Randwerten bis zum Zeitpunkt t ab; das heißt, Informationen breiten sich mit unendlicher Geschwindigkeit aus.

Ist die Vorzeichensituation andersherum (was normalerweise nicht vorkommt), so *muss* man Endbedingungen statt Anfangsbedingungen setzen.

- Hyperbolische Gleichungen brauchen ebenfalls ein Gebiet der Form $\Omega \times (0, T)$ im Orts-Zeit-Raum, hierbei kann Ω aber auch unbeschränkt sein. Man braucht wiederum Anfangsbedingungen auf $\Omega \times \{0\}$ sowie Randbedingungen auf $\partial\Omega \times (0, T)$. Die Randbedingungen sind wieder von vergleichbarer Art wie bei den anderen Typen, jedoch als Anfangswerte muss man Funktionswerte *und* Ableitungen in Zeitrichtung vorschreiben. Die Anfangswerte können wahlweise durch Endbedingungen ersetzt werden, das heißt man kann hyperbolische Gleichungen auch rückwärts in der Zeit rechnen. Der Lösungswert an einem gegebenen Punkt (t, x) im Zeit-Orts-Raum hängt nur von einer beschränkten Menge im Ortsraum zum Zeitpunkt 0 ab, und der Durchmesser dieser Menge ist von der Größenordnung $O(t)$. (Falls diese beschränkte Menge nicht ganz in Ω liegt, hängt der Wert zu (x, t) entsprechend noch von einigen der Randwerte ab.) Informationen breiten sich somit nur mit endlicher Geschwindigkeit aus. In der Akustik ist diese Geschwindigkeit beispielsweise die Schallgeschwindigkeit.

A.3. Analytische Lösung der Prototypen

Wir wollen nun die drei prototypische Gleichungen in einem einfachen Rahmen exakt lösen, um die Intuition für die Eigenschaften zu verbessern. Wir betrachten dafür das Rechengebiet $(0, 1) \times (0, 1)$ für (x, t) , wobei wir der Einheitlichkeit halber auch im elliptischen Fall t schreiben, obwohl hier die Bezeichnung y angebrachter wäre. Ferner setzen wir zu $x = 0$ und $x = 1$ die Randwerte $u = 0$, was für alle Gleichungen passend ist. Nun machen wir den *Separationsansatz*

$$u(x, t) = v(x)w(t).$$

Diesen Ansatz setzen wir in die drei Gleichungen

$$-\partial_{xx}u - \partial_{tt}u = 0, \quad -\partial_{xx}u + \partial_{tt}u = 0, \quad -\partial_{xx}u + \partial_t u = 0$$

ein und erhalten

$$-v''w - vw'' = 0, \quad -v''w + vw'' = 0, \quad -v''w + vw' = 0,$$

also

$$\frac{v''}{v} = -\frac{w''}{w}, \quad \frac{v''}{v} = \frac{w''}{w}, \quad \frac{v''}{v} = \frac{w'}{w}.$$

In allen Fällen hängt jeweils die linke Seite nur von x und die rechte nur von t ab, also müssen beide Seiten konstant sein. Es muss also

$$v''(x) = cv(x)$$

sein mit irgendeiner Konstanten c und außerdem

$$v(0) = v(1) = 0$$

wegen der Randbedingung. Dieses Problem hat die abzählbar unendlich vielen Lösungen

$$v(x) = \sin(n\pi x), \quad n \in \mathbb{N},$$

wobei die Konstante c dann den Wert $-n^2\pi^2$ hat. Für w haben wir nun jeweils die Gleichungen

$$w''(t) = n^2\pi^2 w(t), \quad w''(t) = -n^2\pi^2 w(t), \quad w'(t) = -n^2\pi^2 w(t)$$

zu lösen. Der Lösungsraum ist zweidimensional (im letzten Fall eindimensional), und eine Basis ist jeweils gegeben durch

$$\begin{aligned} w(t) &= \exp(n\pi t), & w(t) &= \sin(n\pi t), & w(t) &= \exp(-n^2\pi^2 t). \\ w(t) &= \exp(-n\pi t), & w(t) &= \cos(n\pi t), & & \end{aligned}$$

Betrachten wir nun zuerst das dritte Beispiel (parabolische Gleichung). Die Lösung hat die allgemeine Darstellung

$$u(x, t) = \sum_{n=1}^{\infty} a_n \sin(n\pi x) \exp(-n^2\pi^2 t), \quad (\text{A.1})$$

wobei die Koeffizienten a_n so gewählt sein müssen, dass die Reihe in geeigneter Weise konvergiert. Hat man die Anfangsbedingung

$$u(x, 0) = u_0(x) \quad (\text{A.2})$$

mit einer stetigen Funktion u_0 , die in den Punkten 0 und 1 verschwindet, so ergibt das die Bedingung

$$u_0(x) = \sum_{n=1}^{\infty} a_n \sin(n\pi x)$$

an die Koeffizienten a_n . Da sich jede stetige Funktion eindeutig in eine gleichmäßig konvergente Fourierreihe entwickeln lässt, sind die a_n dadurch eindeutig gegeben. Die

cos-Terme der Fourierreihe entfallen, da u_0 bei 0 und 1 verschwindet. Die Lösungsdarstellung (A.1) muss mit diesen Koeffizienten ebenfalls gleichmäßig konvergieren, da die Faktoren $\exp(-n^2\pi^2t)$ durch 1 beschränkt sind. Würde man (A.2) durch entsprechende Endbedingungen

$$u(x, 1) = u_1(x) \tag{A.3}$$

ersetzen, so könnte man die Funktion u_1 natürlich ebenfalls in eine Fourierreihe entwickeln. Die Koeffizienten wären durch die Bedingung

$$u_1(x) = \sum_{n=1}^{\infty} a_n \sin(n\pi x) \exp(-n^2\pi^2)$$

ebenfalls eindeutig gegeben. Beim Einsetzen in (A.1) träten dann aber Faktoren der Form $\exp(n^2\pi^2(1-t))$ auf, die für jedes $t < 1$ mit $n \rightarrow \infty$ unbeschränkt sind, so dass die Darstellung (A.1) nicht mehr konvergiert. Dies zeigt, dass man die Anfangsbedingungen nicht durch Endbedingungen ersetzen kann.

Betrachten wir jetzt das zweite der drei Beispiele, also die hyperbolische Gleichung. Hier ergibt sich für u die Darstellung

$$u(x, t) = \sum_{n=1}^{\infty} a_n \sin(n\pi x) \sin(n\pi t) + \sum_{n=1}^{\infty} b_n \sin(n\pi x) \cos(n\pi t). \tag{A.4}$$

Durch Anfangsbedingungen der Form (A.2) ergibt die Bedingung

$$u_0(x) = \sum_{n=1}^{\infty} b_n \sin(n\pi x),$$

woraus durch Fourierentwicklung die Koeffizienten b_n eindeutig bestimmt sind und der entsprechende Teil der Lösungsdarstellung (A.4) wiederum gleichmäßig konvergiert. Über die Koeffizienten a_n ist aber noch nichts bekannt. Setzt man zusätzlich eine Endbedingung wie in (A.3), so nützt das auch nichts, da hierin die a_n ebenfalls wegfallen. Ergänzt man die Anfangsbedingung stattdessen durch eine zusätzliche Bedingung der Form

$$\partial_t u(x, 0) = \bar{u}_0(x),$$

so erhält man die Bedingung

$$\bar{u}_0(x) = \sum_{n=1}^{\infty} a_n \sin(n\pi x),$$

woraus sich dann die a_n eindeutig bestimmen lassen. Beide Reihen in der Darstellung (A.4) konvergieren somit gleichmäßig. Ersetzt man beide Anfangsbedingungen durch Endbedingungen, so ändert sich (bis auf ein paar Vorzeichen) nichts.

Betrachten wir jetzt das erste der drei Beispiele, die elliptische Gleichung. Hier ergibt sich die allgemeine Lösungsdarstellung

$$u(x, t) = \sum_{n=1}^{\infty} a_n \sin(n\pi x) \exp(n\pi t) + \sum_{n=1}^{\infty} b_n \sin(n\pi x) \exp(-n\pi t). \quad (\text{A.5})$$

Mit den Anfangswerten

$$u(x, 0) = u_0(x), \quad \partial_t u(x, 0) = \bar{u}_0(x)$$

würde man die Koeffizienten a_n und b_n eindeutig bestimmen, aber die a_n -Reihe würde für $t > 0$ nicht konvergieren. Nimmt man stattdessen Endwerte

$$u(x, 1) = u_1(x), \quad \partial_t u(x, 1) = \bar{u}_1(x),$$

so konvergiert entsprechend der b_n -Teil nicht. Durch die Kombination

$$u(x, 0) = u_0(x), \quad u(x, 1) = u_1(x),$$

also dem einfachen Vorschreiben von Funktionswerten an *allen* Rändern, ergibt sich jedoch eine konvergente Lösungsdarstellung. Genauso gut funktioniert es mit Kombinationen wie

$$u(x, 0) = u_0(x), \quad \partial_t u(x, 1) = \bar{u}_1(x).$$

Wichtig ist für elliptische Gleichungen also nur, dass an *jedem* Rand Bedingungen vorschrieben sind.