

Kann Mathematik der elektrischen Impedanztomographie zum Durchbruch verhelfen?

Martin Brühl

Martin Hanke-Bourgeois

5. Januar 2001

1 Tomographie und Mathematik

Die Computertomographie ist aus dem heutigen medizinischen Alltag nicht mehr wegzudenken. Aber inwiefern ist schon bekannt, daß dieser Fortschritt ohne mathematische Grundlagenforschung an den Universitäten undenkbar gewesen wäre? Beispielsweise beruht die 1979 mit einem Nobelpreis für Medizin (A. M. Cormack und G. N. Hounsfield) ausgezeichnete Röntgentomographie im wesentlichen auf einer Inversionsformel von Radon aus dem Jahr 1917 für die nach ihm benannte Integraltransformation. Nicht nur dieses Paradebeispiel widerlegt das immer wieder zitierte Vorurteil, Mathematiker lebten in dem Elfenbeinturm ihrer Wissenschaft, und ihre Ergebnisse seien für die Gesellschaft ohne erkennbaren Nutzen. Ein solches Vorurteil erklärt sich aber vielleicht dadurch, daß zwischen den theoretischen Resultaten und der praktischen Umsetzung oft viel Zeit verstreicht — im obigen Fall gar mehr als ein halbes Jahrhundert.

In dieser Übersicht beschreiben wir unlängst erzielte Fortschritte bei dem Rekonstruktionsproblem der *elektrischen Impedanztomographie*. Es ist zu hoffen, daß diese Resultate in mittelbarer Zukunft dazu führen können, die elektrische Impedanztomographie als Diagnoseverfahren in der medizinischen Praxis zu etablieren.

2 Klassische Tomographieverfahren

Allen tomographischen Methoden ist das Ziel gemeinsam, aus Messungen auf dem Rand oder außerhalb des Körpers auf ein Bild des Körperinneren zu schließen. Dies wird dadurch erreicht, daß man unterschiedliche Gewebestrukturen mit Hilfe ihrer spezifischen biophysikalischen Eigenschaften hervortreten läßt, etwa durch Bestrahlung des Körpers oder mit Hilfe von elektromagnetischen Feldern. Die jeweiligen physikalischen Prinzipien der einzelnen Tomographieverfahren führen zu unterschiedlichen mathematischen Modellen und diese erfordern dann naturgemäß auch ganz spezielle Rekonstruktionsalgorithmen.

Beginnen wir mit der verbreitetsten Tomographiemethode, der schon erwähnten *Röntgentomographie*, die oft mit dem allgemeineren Begriff ‚Computertomographie‘ gleichgesetzt wird. Bei diesem Verfahren wird eine Röntgenstrahlenquelle um den Patienten bewegt. Die Röntgenstrahlung durchdringt den Körper geradlinig, verliert aber durch Absorption proportional zur lokalen Massendichte und zur Weglänge an Intensität, welche auf der ge-

genüberliegenden Seite mittels Detektoren gemessen wird. Aus dieser Information wird dann durch Inversion der bereits zitierten Radontransformation ein Bild der Massendichte im Körperinneren rekonstruiert. Ein Nachteil dieser Technologie ist die gesundheitliche Belastung, die eine fortdauernde Beobachtung über längere Zeitspannen hinweg ausschließt.

Die *Kernspintomographie* (MRI oder NMR) erzeugt ein Bild der magnetischen Resonanz der Atomkerne, wenn diese starken Magnetfeldern ausgesetzt werden. Hauptnachteil ist hier die sehr aufwendige und langwierige Datenerfassung.

Bei der *Ultraschalltomographie* werden hochfrequente Schallimpulse ausgesendet, die sich in den verschiedenen Gewebarten unterschiedlich schnell ausbreiten und an Übergängen zwischen zwei Medien gebrochen und reflektiert werden. Aus der Messung der Laufzeit und der Intensität der gestreuten Schallwellen kann auf die akustische Dichte im Innern zurückgeschlossen werden, in der Mathematik spricht man hier von einem inversen Streuproblem. Für den Patienten ist diese Methode frei von gesundheitlichen Belastungen; allerdings kann das Verfahren nicht zur Bildgebung von Körperregionen eingesetzt werden, welche durch Knochen oder Luft gegen die Ultraschallwellen abgeschirmt sind, also etwa das Gehirn oder innere Organe im Brustkorb.

Im Unterschied zu den bislang genannten Methoden, die völlig ohne Eingriff in den Körper auskommen, werden bei der *Emissionstomographie* radioaktive Präparate in den Körper injiziert. Die Intensität der anschließend austretenden Strahlung gibt Aufschluß über die örtliche Verteilung des Präparats. Auf diese Weise läßt sich bestimmen, wo sich die Substanzen im Körper ablagern. Als Vertreter dieser Verfahren seien SPECT (Single Photon Emission Computerized Tomography) und PET (Positron Emission Tomography) angeführt.

3 Impedanztomographie

Seit den frühen achtziger Jahren werden auch tomographische Verfahren entwickelt, die die elektrischen Leitfähigkeitskoeffizienten im Innern des Körpers bestimmen. Dabei wird ausgenutzt, daß die verschiedenen Organe, Muskeln, Knochen, Fettgewebe, eingeatmete Luft, usw. unterschiedlich hohen elektrischen Widerstand leisten.

Während der Untersuchung wird dem Patienten durch einen auf der Haut aufliegenden Elektrodengürtel (mit typischerweise 32 oder 64 Elektroden) eine Folge von Strom-

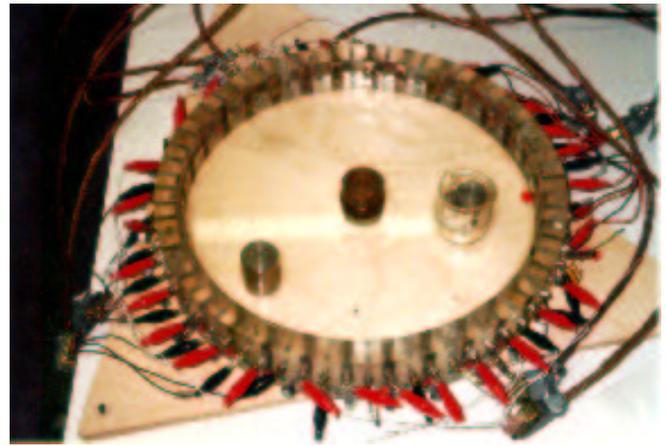
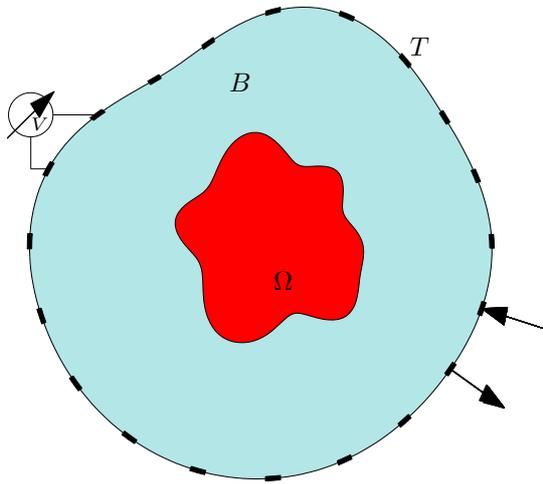


Abbildung 1: Elektrische Impedanztomographie

mustern eingespeist, die für den menschlichen Körper keinerlei Belastung darstellen. Neben diesem offensichtlichen Vorteil — etwa gegenüber der Röntgentomographie — lassen auch die weitaus geringeren Untersuchungskosten die Impedanztomographie attraktiv erscheinen.

Die linke Figur in Abb. 1 skizziert eine gängige Versuchsanordnung für den Meßvorgang; die Elektroden sind hierin schwarz markiert. Die Pfeile zeigen an, daß in diesem Experiment durch eine ausgewählte Elektrode Strom in den Körper hinein- und durch die benachbarte Elektrode wieder hinausfließt. Zwischen allen anderen Elektroden wird währenddessen die Spannung abgegriffen. Rechts im Bild ist ein Versuchsaufbau, der an der Oxford Brookes University für Testzwecke verwendet wird. Bei diesem Tank sind Meßelektroden (rot) und stromführende Elektroden (schwarz) zueinander versetzt. Dadurch ist es möglich, von allen Seiten Strom in den Körper einzuspeisen und gleichzeitig an *allen* Meßelektroden Spannungen abzugreifen. In dem Tank befinden sich ein Elektrolyt mit homogener Leitfähigkeit, zwei Metallzylinder (höhere Leitfähigkeit) sowie ein kaum leitendes Becherglas. Auf dieses Beispiel kommen wir später zurück.

Jede Randstromvorgabe erzeugt im Patienten ein elektrisches Feld und somit Potentialdifferenzen zwischen den aufliegenden Elektroden. Seien J , σ und E das *Stromfeld*, der ortsabhängige *Leitfähigkeitskoeffizient* und das *elektrische Feld* im Innern des Körpers, den wir im folgenden, wie in Abb. 1 mit B bezeichnen. Das *Ohmsche Gesetz*

$$J = \sigma E \quad (1)$$

und die sogenannte *Kontinuitätsgleichung*

$$\operatorname{div} J = 0 \quad (2)$$

bilden die physikalischen Grundlagen der Impedanztomographie. Die Kontinuitätsgleichung spiegelt das physikalische Erhaltungsprinzip wider, nach dem sich in jeder Zelle des Körpers die Summe der ein- und ausfließenden Ströme zu Null summieren muß.

Das elektrische Feld ist ein Gradientenfeld, d.h. es existiert eine skalare Funktion u , das *Potential*, mit

$$E = -\operatorname{grad} u. \quad (3)$$

Die gemessene Spannung zwischen zwei Elektroden auf der Haut ist gerade die Differenz der entsprechenden Werte des Potentials.

Eine Kombination der Modellgleichungen (1), (2) und (3) ergibt die partielle Differentialgleichung

$$\operatorname{div}(\sigma \operatorname{grad} u) = 0 \quad \text{in } B \quad (4)$$

für u . Aus der Theorie der partiellen Differentialgleichungen ist bekannt, daß die Gleichung (4) eine eindeutige Lösung u besitzt, sofern zusätzlich noch *entweder* der Stromfluß¹

$$\nu \cdot J = f \quad (5)$$

oder die Spannung

$$u = g \quad (6)$$

auf der Körperoberfläche T (dem Rand des Körpers B) vorgegeben wird; ν bezeichnet hierbei den äußeren Normalenvektor auf dem Rand, $\nu \cdot J$ also die senkrecht zum Rand nach außen fließende Stromstärke. In der Mathematik spricht man im ersten Fall (5) auch von *Neumannrandwerten* und im zweiten Fall (6) von *Dirichletrandwerten* für die Differentialgleichung (4).

In unserem Fall liegen jedoch sowohl die angelegten Randströme als auch Randspannungen als Meßdaten vor. Dieser Informationsüberschuß soll genutzt werden, um den noch unbekannt Parameter σ in der Differentialgleichung (4) (den gesuchten Leitfähigkeitskoeffizienten) zu rekonstruieren.

Es ist allerdings intuitiv einleuchtend, daß diese Meßdaten allein noch nicht ausreichen, da der Informationsüberschuß nur aus der Kenntnis einer Funktion (etwa

¹Bei Vorgabe des Randstroms ist zu beachten, daß sich die Summe aller Ströme zu Null addiert und Eindeutigkeit der Lösung nur durch Erdung erreicht wird.

der Spannung u) auf dem Rand T besteht, während die Leitfähigkeit im gesamten Körper B gesucht wird. Daher wird das Experiment mit verschiedenen Randstromvorgaben wiederholt, also wie eingangs erwähnt eine Serie von N Randströmen f_n , $n = 1, \dots, N$, dem Körper eingespeist; die jeweils resultierenden Spannungen bezeichnen wir entsprechend mit g_n .

Bevor wir auf numerische Rekonstruktionsverfahren eingehen werden, soll zumindest ganz kurz die Frage angeschnitten werden, inwieweit der Leitfähigkeitskoeffizient überhaupt aus den vorhandenen Daten rekonstruierbar ist. Generell könnte man sich ja vorstellen, daß zwei völlig verschiedene Leitfähigkeitskoeffizienten für alle denkbaren Randstromvorgaben gleiche Spannungswerte auf dem Rand liefern.

Die Beantwortung dieser Frage ist ein extrem schwieriges mathematisches Problem, welches erst unlängst befriedigend gelöst werden konnte. Demnach kann (zumindest für praxisrelevante Leitfähigkeitsverteilungen) davon ausgegangen werden, daß im idealisierten Fall mit unendlich vielen unabhängigen Randstromvorgaben die Randspannungen die Werte des Leitfähigkeitskoeffizienten eindeutig festlegen.

Allerdings war bei der obigen Problemstellung die Rede von *unendlich vielen* Randstromvorgaben, während in der Praxis nur eine begrenzte Zahl von Messungen vorgenommen werden kann. Insofern können die zitierten theoretischen Ergebnisse allenfalls zur Rechtfertigung dieser neuen Tomographiemethode herangezogen werden; darüber hinaus sind sie ohne größere praktische Bedeutung.

4 Das Problem ist schlechtgestellt

Bei fast allen zum Einsatz kommenden Rekonstruktionsverfahren für die elektrische Impedanztomographie wird die Untersuchung des Patienten für gewisse Leitfähigkeitsvorgaben $\tilde{\sigma}$ auf dem Computer simuliert, d.h. die Differentialgleichung (4) wird für die eingespeisten Randstromvorgaben numerisch gelöst. Die auf diese Weise berechneten Randspannungen \tilde{g}_n werden dann mit den gemessenen Randspannungen g_n verglichen, etwa durch Auswertung der Abweichung im quadratischen Mittel,

$$\Phi(\tilde{\sigma}) = \sum_{n=1}^N \int_T |g_n - \tilde{g}_n|^2, \quad (7)$$

dem sogenannten Datenfit. In diesem Funktional werden also für jedes einzelne Experiment (für jedes n) die Abweichungen der berechneten von den gemessenen Spannungswerten über den gesamten Rand aufintegriert und die so erhaltenen einzelnen Fehler summiert. $\Phi(\tilde{\sigma})$ ist demzufolge genau dann nahe bei Null, wenn die gemessenen und die simulierten Randspannungen für *jedes* Experiment *überall* auf dem Rand des Körpers gut übereinstimmen. Da die Randspannungen den Leitfähigkeitskoeffizienten eindeutig festlegen, wird man dann erwarten, daß $\tilde{\sigma}$ eine gute Approximation an die tatsächlich vorliegende Leitfähigkeitsverteilung ist. Nach geeigneter Visualisierung von $\tilde{\sigma}$ kann das so erzeugte Bild vom Arzt entsprechend interpretiert werden.

Die Praxis ist aufgrund unvermeidbarer Meßungenauigkeiten in den vorliegenden Spannungswerten leider erheblich komplizierter. Hinzu kommen subtile Phänomene an den Meßelektroden, die nur schwer in das mathematische Modell integriert werden können: So treten an der Grenzschicht zwischen der Haut und dem Metall der Elektroden zusätzliche Kontaktwiderstände auf, die zu wesentlich komplizierteren Randbedingungen als (5) und (6) führen und nicht vernachlässigt werden können, ohne Einbußen in der Rekonstruktionsqualität in Kauf zu nehmen.

Diese verschiedenen Faktoren haben zur Folge, daß selbst für den exakten Leitfähigkeitskoeffizienten $\tilde{\sigma} = \sigma$ das Fehlerfunktional (7) in der Regel nicht den Wert Null, sondern nur einen kleinen Wert δ ergibt, den wir im folgenden *Rauschniveau* nennen wollen. Es hat sich daher eingebürgert, zwei Leitfähigkeiten als *nicht unterscheidbar* einzustufen, wenn die Abweichung im quadratischen Mittel unterhalb des Rauschniveaus liegt. Unglücklicherweise können zwei Leitfähigkeiten grundverschieden und trotzdem anhand der gemessenen Daten nicht unterscheidbar sein. Besonders eklatante Beispiele lassen sich konstruieren, wenn in die Mitte eines Körpers ein elektrischer Isolator eingepflanzt wird. In diesem Fall ist auf der Haut so gut wie kein Unterschied in den Spannungsdifferenzen meßbar, da fast der gesamte Strom erst gar nicht so tief in den Körper eindringt und somit ohnehin an dem Isolator vorbeifließt.

Dies ist in Abb. 2 illustriert: Links sieht man dort die Äquipotentiallinien für eine homogene Leitfähigkeit und eine Stromvorgabe wie in Abb. 1, bei der an einer Elektrode Strom eingespeist wird und an einer benachbarten Elektrode Strom abgezogen wird. Die rechte Abbildung zeigt das entsprechende Bild, wenn das Innere des kreisförmigen Körpers einen zusätzlichen Isolator enthält (in der Abbildung als schwarzes Hindernis eingezeichnet). Der Strom fließt jeweils senkrecht zu den gestrichelt eingezeichneten Äquipotentiallinien, vgl. (1) und (3). Folglich wird die überwiegende Mehrzahl der Stromlinien (durchgezogene Linien) von dem Hindernis kaum beeinflusst. Entsprechend ähnlich sind die resultierenden Spannungen auf dem Kreisrand T : Abb. 3 zeigt die relativen Werte des Potentials u auf dem Kreisrand, aufgetragen über dem zugehörigen Winkel; die gestrichelte Kurve, die fast auf der durchgezogenen Kurve liegt, gehört zu dem Modell mit Isolator.

Glücklicherweise ergeben sich etwas größere Unterschiede in den Meßdaten, wenn der Randstrom durch alle zur Verfügung stehenden Elektroden *gleichzeitig* eingespeist wird. Für diesen Fall läßt sich quantitativ abschätzen, daß bei einem Rauschniveau von etwa 0.1% der Meßdaten zumindest noch Objekte von dem Durchmesser eines Fingers prinzipiell zu erkennen sind.

Wir wollen die oben erwähnte Tatsache, daß grundverschiedene Leitfähigkeiten annähernd gleiche Meßdaten liefern können, noch etwas eingehender beleuchten. Dieser Umstand hat nämlich umgekehrt zur Folge, daß unser *inverse Problem* der Rekonstruktion der Leitfähigkeit aus eben diesen Meßdaten extrem sensibel gegenüber Datenstörungen ist: Mathematiker nennen Probleme dieser Art *schlechtgestellt*.

Dieser etwas merkwürdige Begriff wurde erstmals 1923

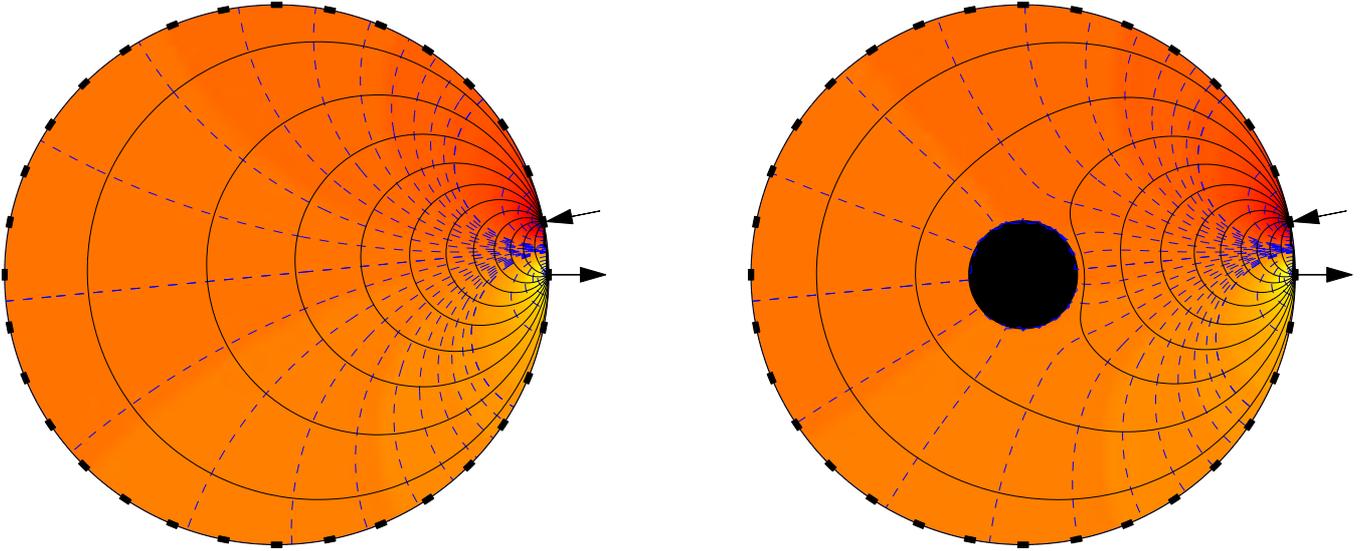


Abbildung 2: Stromlinien mit und ohne Isolator

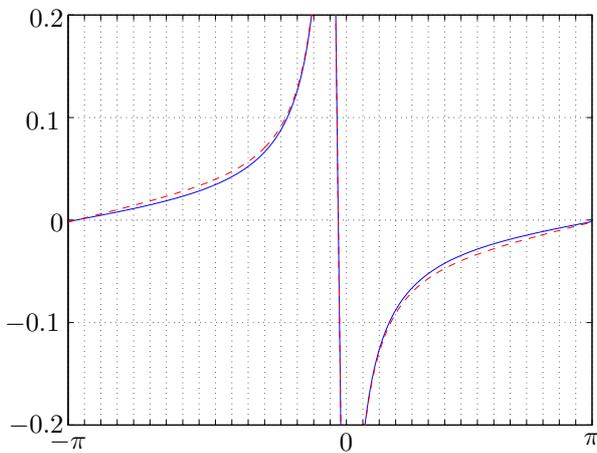


Abbildung 3: Randspannungen mit und ohne Isolator

von Hadamard verwendet, der nur solche mathematischen Problemstellungen als physikalisch sinnvoll ansah, die eine eindeutige und stetig von den Daten abhängende Lösung besitzen. Letztere Forderung soll sicherstellen, daß selbst zu verrauschten Meßdaten, wie sie in der Realität fast immer vorliegen, noch brauchbare Näherungslösungen gehören, die bei Verbesserung der Datenqualität, also Verminderung des Rauschniveaus, die exakte Lösung zunehmend besser approximieren.

Anders als noch von Hadamard vermutet, sind mittlerweile zahlreiche Fragestellungen aus Anwendungen herausgewachsen, die auf schlechtgestellte Probleme führen; beispielsweise kann man hierzu fast ausnahmslos die zuvor aufgeführten Tomographiemethoden zählen.

Deutlich wird dies etwa im Fall der elektrischen Impedanztomographie anhand des obigen Isolatorbeispiels. Für die folgende quantitative Analyse dieses Beispiels nehmen wir an, der kreisförmige Isolator im Zentrum des Einheitskreises aus Abb. 2 habe den Radius ρ und es lägen

Randstromvorgaben in der Form trigonometrischer Muster vor, etwa $f_n = \sin n\theta$, $n = 1, 2, \dots$. Hierzu errechnen sich die zugehörige Randspannungen

$$\tilde{g}_n = \frac{1}{n} \frac{1 + \rho^{2n}}{1 - \rho^{2n}} \sin n\theta.$$

Die Randspannungen haben also wiederum trigonometrische Gestalt, wobei die Vorfaktoren von der Frequenz n und dem Radius ρ des Isolators abhängen. Vergleicht man dies mit den entsprechenden Werten $g_n = \frac{1}{n} \sin n\theta$ zur homogenen Leitfähigkeit $\sigma = 1$, dann ergibt sich die Differenz zu

$$\tilde{g}_n - g_n = \frac{2}{n} \frac{\rho^{2n}}{1 - \rho^{2n}} \sin n\theta. \quad (8)$$

An dem Zähler erkennt man, daß die Messungen umso schwieriger zu unterscheiden sind, je hochfrequenter die angelegten Randstromvorgaben sind, also je größer n ist, und diese Unterscheidbarkeit nimmt sogar exponentiell ab. Für $\rho = 0.1$ bewegt sich diese Differenz spätestens ab $n = 8$ im Bereich der Rechengenauigkeit heutiger Computer, was bedeutet, daß Spannungsdifferenzen dann nicht mehr von Rundungsfehlern getrennt werden können. Dagegen dürfte bereits für $n = 2$ die übliche Meßgenauigkeit unterschritten sein. Dies alles sind Indizien für die Schlechtgestellttheit des Problems.

Um den Parameter ρ zu rekonstruieren, genügt theoretisch eine einzige Messung der Form (8). Nehmen wir jetzt an, daß um ε gestörte Meßdaten in der speziellen Form

$$\left(\frac{2}{n} \frac{\rho^{2n}}{1 - \rho^{2n}} + \varepsilon \right) \sin n\theta$$

vorliegen, dann ergibt sich daraus für kleine ε als Rekonstruktion

$$\tilde{\rho} \approx \rho \left(1 + \left(\frac{1 - \rho^{2n}}{2} \right)^2 \frac{\varepsilon}{\rho^{2n}} \right),$$

d.h. ein infinitesimaler Datenfehler wird im wesentlichen um den Faktor $1/\rho^{2n}$ verstärkt. Dies deutet darauf hin,

daß kleinere Objekte (ρ klein) wesentlich schwieriger zu rekonstruieren sind und die Instabilität zunimmt, je hochfrequenter der Randstrom ist (n groß), auf dem die Rekonstruktion basiert.

Damit Rekonstruktionsverfahren aus diesem Grund nicht schon von vornherein zum Scheitern verurteilt sind, muß der Schlechtgestellttheit schon bei der Konstruktion des Verfahrens Rechnung getragen werden. Daher sind in der Vergangenheit sogenannte *Regularisierungsverfahren* entwickelt worden, die zusätzlich für Stabilität der Näherungslösungen sorgen sollen.

Das am weitesten verbreitete Regularisierungsverfahren ist die *Tikhonov-Regularisierung*; in unserem Fall würde man hierbei das Fehlerfunktional (7) durch das *Tikhonov-Funktional*

$$\Phi_\alpha(\tilde{\sigma}) = \sum_{n=1}^N \int_T |g_n - \tilde{g}_n|^2 + \alpha \int_B |\tilde{\sigma}|^2 \quad (9)$$

mit einem Parameter $\alpha > 0$ ersetzen. Der neu hinzugekommene Term bestraft große Ausschläge der Leitfähigkeit; daher handelt es sich bei der (9) minimierenden Leitfähigkeit $\tilde{\sigma}_\alpha$ um eine „gedämpfte“ Version der exakten Lösung. Die Dämpfung läßt sich über den Parameter α steuern; dieser sollte in Abhängigkeit von der Größe des Rauschniveaus δ gewählt werden: je größer δ ist, umso größer muß auch α gewählt werden, um damit den Einfluß des Strafterms im Funktional (9) zu erhöhen.

Die optimale Einstellung des Parameters α ist eine Kunst für sich und Gegenstand zahlreicher Veröffentlichungen in diesem Gebiet. Eine sehr einleuchtende Strategie zur Wahl von α ist das sogenannte *Diskrepanzprinzip*: Dabei versucht man α so einzustellen, daß der Datenfit (7) gerade die Größenordnung des Rauschniveaus erreicht. Die so resultierende Approximation $\tilde{\sigma}$ kann auch durch ein naheliegendes Optimierungsproblem beschrieben werden: Unter allen Leitfähigkeitskoeffizienten, deren Datenfit im Bereich des Rauschniveaus liegt, ist $\tilde{\sigma}$ am „kleinsten“, d.h.

$$\int_B |\tilde{\sigma}|^2 \longrightarrow \min$$

unter der Nebenbedingung $\Phi(\tilde{\sigma}) \leq \delta$.

Während sich diese Vorgehensweise bei vielen linearen Problemen bewährt hat, ist allerdings bei dem vorliegenden nichtlinearen Problem theoretisch die Existenz eines solchen Minimums und damit einer entsprechenden Näherungslösung $\tilde{\sigma}$ nicht unbedingt gewährleistet. In der Praxis stellt sich dieser Vorbehalt jedoch als irrelevant heraus.

5 Iterative Ausgleichverfahren

Da die Potentiale samt ihrer Randwerte nichtlinear von der Leitfähigkeit abhängen, handelt es sich bei der Minimierung von (7) oder (9) um ein *nichtlineares Ausgleichproblem*, das nicht explizit gelöst werden kann und deshalb numerisch mit iterativen Methoden angegangen werden muß.

Hierfür bieten sich Standardverfahren aus der nichtlinearen Optimierung an, sogenannte *Gradientenverfahren*,

bei denen in jedem Iterationsschritt zunächst eine geeignete Suchrichtung bestimmt wird, entlang der anschließend, ausgehend von der vorherigen Näherung das Fehlerfunktional (7) minimiert wird. Diese Projektion auf eindimensionale Minimierungsprobleme birgt meist einen so großen Informationsverlust in sich, daß in der Folge die Konvergenz des Verfahrens sehr langsam wird und entsprechend viele Iterationsschritte benötigt werden.

Schnellere Konvergenz erzielt man im allgemeinen mit *newtonartigen* Verfahren. Die zugrundeliegende Idee besteht hier darin, das nichtlineare Problem durch eine lokale Linearisierung zu ersetzen. Die Lösung des linearisierten Problems ist dann sehr viel einfacher zu bewältigen und bildet einen Schritt des zugehörigen Iterationsverfahrens. Zur Berechnung dieser Linearisierung muß die Ableitung des Fehlerfunktionals (7), $\partial\Phi/\partial\tilde{\sigma}$, bestimmt werden, d.h. die Variation der Potentiale u bei infinitesimaler Änderung der Leitfähigkeit. Die Ableitungen $\partial u/\partial\tilde{\sigma}$ können wieder über Randwertprobleme mit Differentialgleichungen der Form (4) ausgedrückt werden, deren Lösung aus numerischer Sicht für den hohen Rechenaufwand dieser Newtonverfahren verantwortlich ist.

Eine Möglichkeit zur weiteren Verringerung des Rechenaufwands besteht darin, die linearisierten Probleme lediglich näherungsweise zu lösen, etwa wiederum mit einem iterativen Verfahren. Man bezeichnet dies dann als *inexaktes Newtonverfahren*. Aber noch ein weiteres Argument spricht hier dafür, die linearen Teilprobleme nicht mit hoher Genauigkeit zu lösen. Denn als Konsequenz aus der angesprochenen Schlechtgestellttheit des Problems werden die iterierten Näherungslösungen umso instabiler, je länger man iteriert. Der vorzeitige Abbruch der Iteration wirkt hier also regularisierend, ähnlich wie die Stabilisierung durch das oben erwähnte Tikhonov-Funktional.

Hier leistet das bereits angeführte Diskrepanzprinzip wieder eine wichtige Hilfestellung, nämlich als Stopkriterium: Demnach wird die Newtoniteration abgebrochen, sobald bei der Minimierung des Fehlerfunktionals (7), also des Datenfits, das Rauschniveau unterschritten wird. Die Effektivität dieses Kriteriums läßt sich für eine Reihe von Modellbeispielen explizit nachweisen. Für die elektrische Impedanztomographie demonstriert das numerische Beispiel in Abb. 4 mit den Daten des Modellexperiments aus Abb. 1 die Wirksamkeit dieser Stopregel: Dort werden die Rekonstruktionen nach 7 bzw. 30 Iterationen gezeigt, wobei hier bei der Minimierung von (7) im siebten Iterationsschritt das Rauschniveau erstmalig unterschritten und danach immer weiter verkleinert wird. Die vermeintlich genauere Rekonstruktion nach 30 Iterationen enthält jedoch keinerlei verwertbare Informationen: Diese Näherung hat sich, salopp gesprochen, von den Meßfehlern in den Daten in die Irre führen lassen. Andererseits enthält die Rekonstruktion nach sieben Iterationen wesentliche Merkmale der Vorgabe aus Abb. 1: in Rottönen dargestellt erkennt man die höhere Leitfähigkeit der beiden Metallteile und im dunkelblauen Farbbereich die niedrigere Leitfähigkeit des Becherglases.

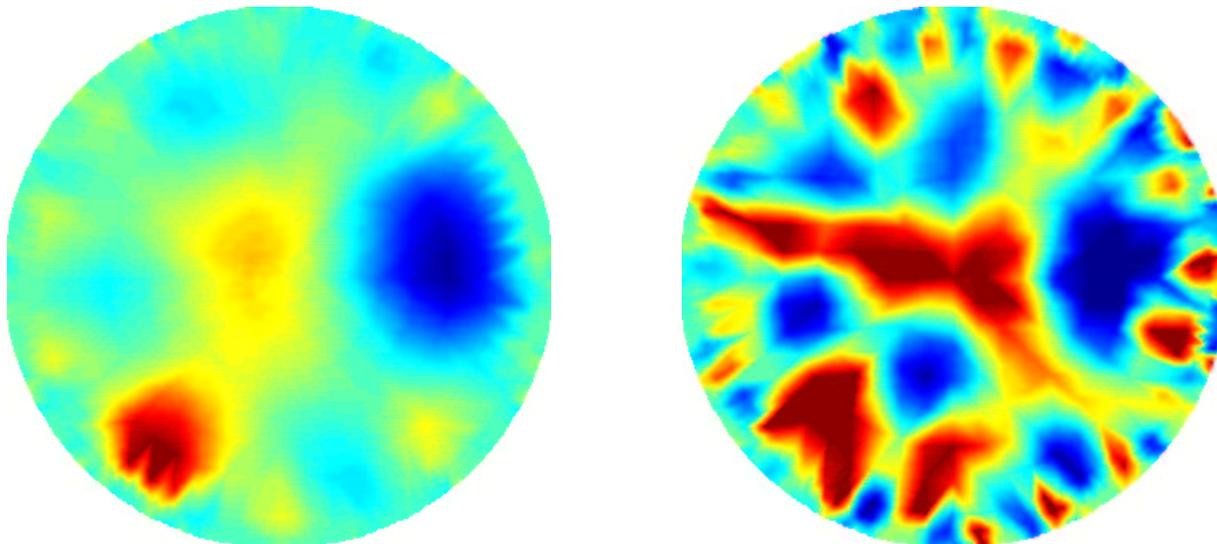


Abbildung 4: Rekonstruktion mit und ohne rechtzeitigem Abbruch

6 Ein neuartiges Verfahren

In vielen Fällen kommt es gar nicht darauf an, die Leitfähigkeit in allen Einzelheiten zu rekonstruieren, beispielsweise wenn man lediglich an Abweichungen von einem Referenzzustand interessiert ist. Bei dem Modellproblem aus Abb. 1 könnte man sich vorstellen, daß das Interesse vorrangig der Position der drei Objekte beziehungsweise ihrer Anzahl überhaupt gilt. Speziell für diese Situation haben wir ein Verfahren entwickelt, mit dem es möglich ist, Einschlüsse in einer homogenen Hintergrundleitfähigkeit mit sprunghaft erhöhten oder erniedrigten Leitfähigkeitswerten zu rekonstruieren.

Bei diesem Verfahren vergleichen wir die gemessenen Spannungswerte mit denjenigen Spannungen, die im Fall eines homogenen Körpers auf der Haut resultieren würden. Letztere können wir natürlich nicht durch Messungen erhalten, sondern müssen sie im Computer simulieren, indem wir die Lösung der Differentialgleichung (4) mit der Referenzleitfähigkeit $\sigma = 1$ durch ein Verfahren der numerischen Mathematik approximieren. Im folgenden fixieren wir vorübergehend den Randstrom f und bezeichnen dann mit u_σ das Lösungspotential für die tatsächlich vorhandene Leitfähigkeit σ und mit u das simulierte Potential für die Referenzleitfähigkeit σ , jeweils bei gleichem Randstrom f .

Dabei ist von entscheidender Bedeutung, daß σ und die Referenzleitfähigkeit σ außerhalb der Menge Ω der vorliegenden Einschlüsse übereinstimmen. In diesem Teilgebiet zwischen der Oberfläche des Körpers und den eingeschlossenen Organen lösen also sowohl u_σ wie auch u die Differentialgleichung

$$\operatorname{div}(\operatorname{grad} u) = 0;$$

Lösungen dieser speziellen Differentialgleichung nennt man auch *harmonische Funktionen*. Die Differenz $w = u_\sigma - u$ ist dann ebenfalls eine in diesem äußeren Teilgebiet harmonische Funktion. Aufgrund der identischen Randstromvorgabe f für die beiden Potentiale u_σ und u

genügt w überdies der homogenen Randbedingung

$$\nu \cdot \operatorname{grad} w = 0 \quad \text{auf } T. \quad (10)$$

Harmonische Funktionen sind besonders „glatte Funktionen“, insbesondere sind sie beliebig oft differenzierbar. Stärkere Oszillationen oder ähnlich „unglatte“ Phänomene können allenfalls in der Nähe des Randes ihres Definitionsbereichs auftreten, in unserem Fall also in der Nähe der Einschlüsse und in der Nähe von T , dem äußeren Rand. Da nun aber die Funktion w auf T die „glatte Nebenbedingung“ (10) erfüllt, können Oszillationen von w sogar nur in der Nähe der Einschlüsse auftreten: Je weiter man sich von der Einschlusmenge Ω und ihrem Rand entfernt, umso glatter wird w .

Dies bildet die Grundlage unseres Algorithmus. Aufgrund der gemessenen Spannungswerte und der berechneten Simulationen für die Referenzleitfähigkeit kennen wir für jeden eingespeisten Randstrom $f = f_n$, $n = 1, \dots, N$, die Werte der zugehörigen Spannungsdifferenzen $w = w_n$ auf dem äußeren Rand T . Die Glattheit dieser Funktionen w_n hängt nach dem eben gesagtten hauptsächlich von dem Abstand der Einschlüsse Ω von der Körperoberfläche ab. Die Gesamtheit der Funktionen w_n kodiert in gewisser Weise die Information über diese Abstände.

Um diese Information zu dekodieren bedienen wir uns eines Tricks. Wir simulieren in einem Punkt $z \in B$ ein elektrisches Dipolpotential D_z im Inneren des ansonsten in der Simulation elektrisch isolierten Körpers B mit homogener Referenzleitfähigkeit. Dies entspricht dem physikalischen Modell, durch diesen vorgegebenen Punkt z einen festen Strom fließen zu lassen. Auf diese Weise können wir gewissermaßen den Körper von innen „elektrisch ausleuchten“. Aufgrund der Isolierung des Körpers ist das resultierende Dipolpotential $w = D_z$ ebenfalls eine harmonische Funktion mit der Nebenbedingung (10), allerdings geht die Harmonizität im Punkt $x = z$ verloren, da dort das Dipolpotential eine Polstelle aufweist. Mit anderen Worten: Genau dann, wenn der Punkt z im Einschluß liegt, ist das Dipolpotential im gesamten äußeren

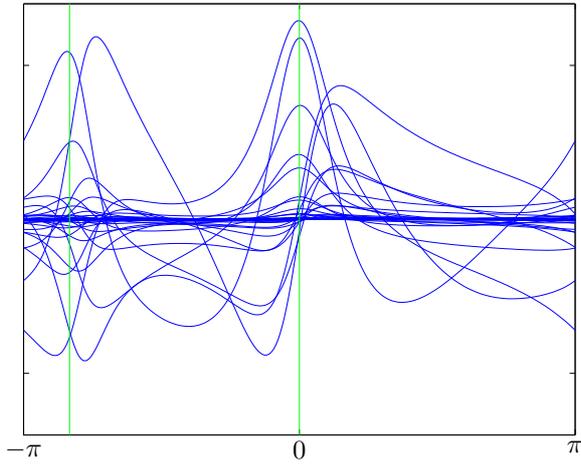


Abbildung 5: Die Funktionen w_n auf dem Rand T

Teilgebiet harmonisch. Und genau diese Eigenschaft teilt D_z dann mit den zuvor betrachteten Potentialdifferenzen w_n .

Wir können jedoch die Dipolpotentiale D_z nicht unmittelbar in Beziehung mit diesen Funktionen w_n setzen, da uns, wie schon gesagt, von letzteren nur ihre Randwerte auf T bekannt sind. Wir müssen also mit den Randwerten $h_z = D_z|_T$ vorliebnehmen. Hier kommt uns aber eine weitere Eigenschaft harmonischer Funktionen zugute, nämlich, daß sie im gesamten äußeren Teilgebiet schon durch ihre Dirichlet- und Neumannrandwerte auf T , also ihre Randspannungen und ihre (in unserem Fall verschwindenden) Randströme eindeutig festgelegt sind. Dies erlaubt es uns, durch einen geeigneten Vergleich der Funktion h_z mit den Randwerten der Funktionen w_n herauszufinden, ob z im äußeren Teilgebiet, dem Harmonizitätsgebiet der $\{w_n\}$ enthalten ist. Grob gesprochen gehört ein Testpunkt z genau dann zu Ω , wenn die Randfunktion h_z des zugehörigen Dipolpotentials in jedem Punkt des Randes T mindestens so glatt ist wie alle Funktionen w_n , $n = 1, \dots, N$.

An dieser Stelle muß jedoch betont werden, daß diese skizzenhafte Erläuterung der Grundidee unseres Verfahrens allenfalls eine stark überzeichnete Abstraktion des tatsächlichen Algorithmus darstellt. Bewußt haben wir den Begriff der Glattheit auch zu Beginn dieser Darstellung in Anführungsstriche gesetzt. Die hier gemeinte Glattheit ist mathematisch nur schwer zu fassen. Auch anschaulich ist es unmöglich, anhand eines optischen Vergleiches zu entscheiden, für welche Punkte $z \in B$ die Randfunktion h_z „mindestens so glatt“ ist wie die Gesamtheit der vorgegebenen Funktionen w_n .

Zur Illustration sei auf Abb. 5 verwiesen, in der alle Randwerte der Funktionen w_n für das Beispiel aus Abb. 1 mit trigonometrischen Randstrommustern übereinander gezeichnet sind. Man erkennt, daß im Bereich der eingezeichneten grünen Markierungen die Funktionen w_n besonders starke Ausschläge aufweisen. Die grünen Markierungen entsprechen den 3 Uhr- und 8 Uhr-Positionen eines Zifferblatts (wie auch in Abb. 6 eingezeichnet), was auf die Lage der zwei äußeren Objekte im Modelltank hinweist.

Mehr kann durch bloßen Augenschein aus dieser Abbildung aber beim besten Willen nicht geschlossen werden.

Die Mathematik erlaubt hingegen mehr: Unser mathematisch exakter „Glattheitstest“ läuft darauf hinaus, die Funktion h_z in optimaler Weise durch eine Linearkombination der Referenzfunktionen w_n zu approximieren,

$$h_z \approx \sum_{n=1}^N \alpha_n w_n, \quad (11)$$

mit reellen Entwicklungskoeffizienten α_n , $n = 1, \dots, N$, um danach die Größe dieser Entwicklungskoeffizienten zu analysieren. Wenn der Punkt z nicht zu der Einschlußmenge gehört, dann weist h_z eine charakteristische Unglattheit auf, die in den Funktionen w_n nicht vorhanden ist. Folglich ergeben sich in diesem Fall verhältnismäßig große Entwicklungskoeffizienten α_n , um diese fehlende Glattheit approximieren zu können. Gehört der Testpunkt z hingegen zu der Einschlußmenge Ω , dann kann h_z relativ einfach, das heißt mit kleineren Entwicklungskoeffizienten in der Form (11) dargestellt werden.

Die zulässige Größe der Entwicklungskoeffizienten für einen Punkt $z \in \Omega$ läßt sich mathematisch exakt beschreiben und auf diese Weise ergibt sich letztendlich eine überraschend einfache Charakterisierung der Einschlußmenge Ω . Für die praktische Durchführung ist dabei hilfreich, daß die Testfunktion h_z mit Hilfe der vorhandenen Daten leicht explizit berechnet werden kann, und zwar ohne erst das gesamte Dipolpotential D_z im Innern des homogenen Körpers ausrechnen zu müssen. Somit haben wir einen einfachen Test zur Verfügung, mit dem wir prinzipiell für jeden beliebigen Punkt $z \in B$ feststellen können, ob er zu der Einschlußmenge Ω gehört oder nicht. In unserer Implementierung können wir daher sehr effizient für alle Punkte eines feinen Rasters testen, ob sie in der Einschlußmenge liegen.

Das resultierende Verfahren ist bei weitem effizienter als die in Abschnitt 5 beschriebenen iterativen Verfahren. Zum Vergleich: Die Berechnung der optimalen Rekonstruktion aus Abb. 4 dauerte mehr als 10 mal so lange wie die in Abb. 6 gezeigte Rekonstruktion mit der neuen Methode. Dabei wurden für das numerische Beispiel in Abb. 6 die drei Inhomogenitäten aus dem Beispieltank aus Abb. 1 nachgebildet. Die (simulierten) Originaleinschlüsse sind durch gestrichelte Linien markiert und die Rekonstruktionen grau hinterlegt. Man beachte, wie genau die drei Objekte gefunden werden; andererseits ist es mit diesen Ergebnissen nicht möglich, den Unterschied zwischen besser leitenden Metallkörpern und dem schlechter leitenden Becherglas zu erkennen.

Unvermeidlich sind natürlich auch bei dieser Vorgehensweise — wie bei jedem anderen numerischen Verfahren für die Impedanztomographie — die Unsicherheiten, die sich aus den Diskretisierungsfehlern bei der numerischen Simulation sowie den allgegenwärtigen Meßfehlern in den Spannungswerten ergeben. Letztendlich kann kein noch so fundiertes Verfahren darüber hinwegtäuschen, daß das Inversionsproblem der elektrischen Impedanztomographie schlecht gestellt ist.

Wie unser Verfahren auf diese praktischen Einschränkungen reagiert, müssen die zukünftigen Untersu-

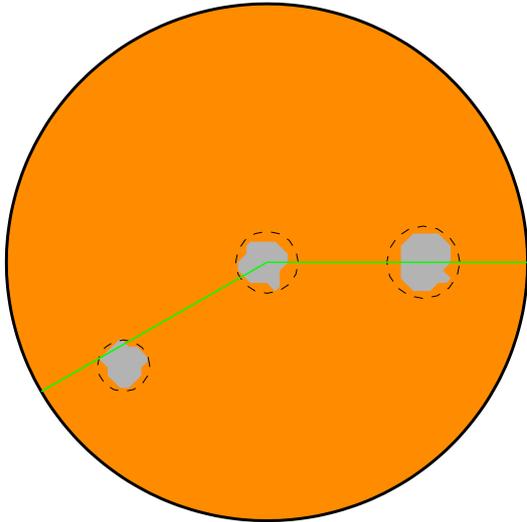


Abbildung 6: Rekonstruktion mit der neuen Methode

chungen erweisen. Derzeit passen wir unser Programm in einem ersten Schritt an die realen Daten mit den zugehörigen Elektrodenmodellen an.

7 Zusammenfassung und Ausblick

Das von uns neu entwickelte Verfahren zielt ausschließlich darauf ab, Unstetigkeiten in der Leitfähigkeitsverteilung aufzuspüren. Es eignet sich daher besonders für die in der Praxis häufig anzutreffenden Situation, daß verschiedene Teilgebiete durch deutlich abweichende Leitfähigkeiten hervortreten (wie z.B. die Organe im menschlichen Körper). Unsere Methode ist aufgrund des geringeren Rechenaufwands zudem deutlich schneller als die bislang verwendeten iterativen Verfahren.

Mit dem gleichen Verfahren können prinzipiell auch anisotrope Einschlüsse bestimmt werden, bei denen im Ohmschen Gesetz (1) ein Leitfähigkeitstensor vorliegt, wenn also die elektrische Leitfähigkeit richtungsabhängig ist wie z.B. in Muskelfasern. Bei einer solchen Vorgabe würden iterative Methoden vollständig versagen.

Abschließend sei noch hinzugefügt, daß die möglichen Anwendungen der Impedanztomographie weit über die Medizin hinausreichen; beispielhaft seien nichtdestruktive Prüfmethode, die Überwachung von Mehrphasenströmungen im Chemieingenieurwesen sowie die Steuerung von industriellen Fertigungsprozessen in der Nahrungsmittelindustrie angeführt.