DIPLOMARBEIT

Comparison of Different Optimization and Regularization Methods for the Solution of Inverse Problems

Andrea Mehsner

Institut für Grundlagen und Theorie der Elektrotechnik
Technische Universität Graz
8010 Graz, Inffeldgasse 18

Begutachter: Univ.-Ass. Dipl.-Ing. Dr. techn. Alice Reinbacher-Köstinger
Betreuer: A.o. Univ.-Prof. Dipl.-Ing. Dr. techn. Christian Magele

Graz, im Februar 2013

Abstract

The solution of inverse problems is a frequent task in science and in technical applications. The goal is to determine the cause of a problem from measurable effects. Examples for such problems can be found in biomedical applications, where non invasive techniques are crucial or when monitoring technical processes inside a pipeline. Such problems are called tomography problems. To compare and assess different methods for solving inverse problems the model of a resistor network with linear resistances is investigated. Voltages at certain accessible points of the network are to be determined, when a given current is injected into the network. The values of different adjacent resistors, which are diverse to the rest of the resistances (called the background) are to be reconstructed. Such networks are used to simulate different parts of the human body in biomedical applications. A major advantage of this forward problem is, that the determination of the accessible voltages from the resistances and the currents injected is straight forward and can be done with less computational effort. The simulated voltages are then opposed to the "measured" ones. The unavoidable differences between those sets of values, called the residuals, can be used to formulate an optimization problem, namely a least squares problem. To solve that type of problem deterministic methods will be employed only. Unfortunately, the resulting optimization problem can be classified as "ill posed", which is quite frequently the case when dealing with inverse problems. This feature makes it all but impossible to solve the resulting system of equations directly. It is essential to stabilize and regularize the system of equations. This is preferably done by adding a new term, the penalty term to the optimization problem. Methods for regularization can be divided into two groups, those which rely on the L2 norm, as in the case of Tikhonov regularization and those, which rely on the L1 norm, represented by the method of Total Variation. Irrespectively of the applied norm, a proper regularization parameter has to be determined. Different approaches to do so are discussed and assessed. To be as close as possible to real world applications both data without and with noise were used in the least squares approach, in order to evaluate the different methods applied and to be able to make somewhat general statements.

Kurzfassung

Die Lösung von inversen Problemen ist eine häufige Aufgabenstellung in der Technik. Dabei handelt es sich um die Ermittlung der Ursache einer Aufgabenstellung aus deren Wirkung, da nur diese messbar zugänglich ist. Als Beispiel sind hierzu nichtinvasive Messmethoden in der biomedizinischen Technik oder in der Prozesstechnik anzuführen. Diese Probleme werden als Tomografie Probleme bezeichnet.
Zum Vergleich mehrerer Methoden wurde als Beispiel das mathematische Modell eines Widerstandsnetzwerks mit linearen Elementen herangezogen. Dabei sollen durch Messung von Spannungen an bestimmten Punkten am Rand des Netzwerks die Widerstandswerte des Netzwerks bestimmt werden. Der dazu notwendige Strom kann auch nur an diesen bestimmten Punkten am Rand des Netzwerks eingespeist werden. Die Widerstandswerte wurden zu Gruppen gleichwertig angenommen, sodass sich das Problem als Objekte auf einem Hintergrund beschreiben lässt. Ähnliche Widerstandsmodelle werden auch für biomedizinische Zwecke als sogenannte Phantome verwendet, um Messverfahren für den menschlichen Körper zu entwickeln.
Diese Aufgabenstellung wurde gewählt, da es sich dabei um einfach zu berechnendes Vorwärtsproblem handelt. Das Vorwärtsproblem beschreibt die Berechnung der Wirkung aus der gegebenen Ursache des Problems, die dann der gemessenen Wirkung gegenübergestellt wird. Dadurch ergibt sich ein Optimierungsproblem (Minimierungsproblem), das auch als Least-Squares-Problem bezeichnet wird. Zur Lösung dieses Problems wurden in diesem Fall nur deterministische Methoden herangezogen.
Es handelt sich dabei in der gegebenen Aufgabenstellung auch um ein schlecht gestelltes unterbestimmtes Gleichungssystem, das eine direkte Lösung nicht möglich macht. Das Problem muss zusätzlich stabilisiert oder regularisiert werden. Dabei wird zum Optimierungsproblem ein Strafterm addiert. Die Regularisierungsverfahren können aufgrund der Norm dieses Strafterms in zwei Gruppen eingeteilt werden. Die L2-Norm wird von der Tikhonov Regularisierungsmethode und die L1-Norm von der Methode der Totalen Variation verwendet. Zusätzlich werden die möglichen Parameterbestimmungsverfahren und Unterschiede erläutert.
Die Rekonstruktionen der gegebenen Widerstandsverteilungen erfolgten ohne zusätzlichem Messrauschen und mit angenommenem Rauschen, das in praktischen Problemen nicht zu vermeiden ist. Die Unterschiede zwischen den angewendeten Verfahren werden durch Darstellungen und Fehleranalyse gezeigt.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Ort	Datum	Unterschrift

Danksagung

An erster Stelle möchte ich mich bei meinem Betreuern Alice Reinbacher-Köstinger und Christian Magele für ihre unermüdliche Unterstützung und die hilfreichen Diskussionen bedanken.
Ein besonderes Dankeschön gilt auch meiner Familie und meinen Mitarbeitern, die mich während der ganzen Zeit so großartig unterstützt haben.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Inverse problems often arise in science and in technical applications. In such the quantity, which can be called the *effect* of something can be measured, while the corresponding *cause* of the measurements is not accessible or measurable. (The opposite of that is the forward or direct problem where the *effect* is calculated from the known *cause*.) For example, it is quite often desirable to have a "look" inside a pipe, a vessel or the human body with noninvasive techniques. This kind of problem is termed a tomography problem. Then perturbations are made from the exterior of the object to be examined and measurements are also just possible on the surface or at the boundary, as it is called. With that measurements the obtained internal structure should be determined in the best possible way for the given application. Thereof arise some additional problems like the computational effort of the reconstruction process. There also exist problems where the distribution of the sought values is not time invariant, for example when investigating a breathing human body. But also optimization problems looking for certain geometrical parameters are similarly solved. Another example is the earthquake location problem.
There are many different inverse problems in different applications but in this work emphasis is put on a simple circuit problem, a network of ohmic resistors. For that setup the forward problem is easy to be calculated and therefore it is a good choice for comparing different methods for solving inverse problems. This problem belongs to the group of electrical sensing techniques which are called electrical impedance tomography (EIT) problems. Basically EIT can be divided into three groups depending on the sought or dominant value. The three groups are the Electro-Magnetic Inductance Tomography (EMT), the Electrical Capacitance Tomography (ECT), and finally the Electrical Resistance Tomography (ERT). The problem in this thesis belongs to the last group, of course, since the sought values are the internal linear resistor values. Resistor circuits are often used as a simulation board, for example in biomedical applications for the cross section of a human torso, because the true values are easy to be determined and show a high long time stability. With the help of this problem basic deterministic reconstruction methods and their characteristics are described and analyzed.

1.2 Aim of This Thesis

In this thesis the main features when solving a linear inverse problem are discussed. Different methods to perform this task are presented. Regularization is crucial in this respect, so the most important methods are introduced. The obtained results are then applied to a nonlinear circuit problem, see figure 1.1 on the following page.

The main task is to determine the resistance of every resistor in the network as accurately as

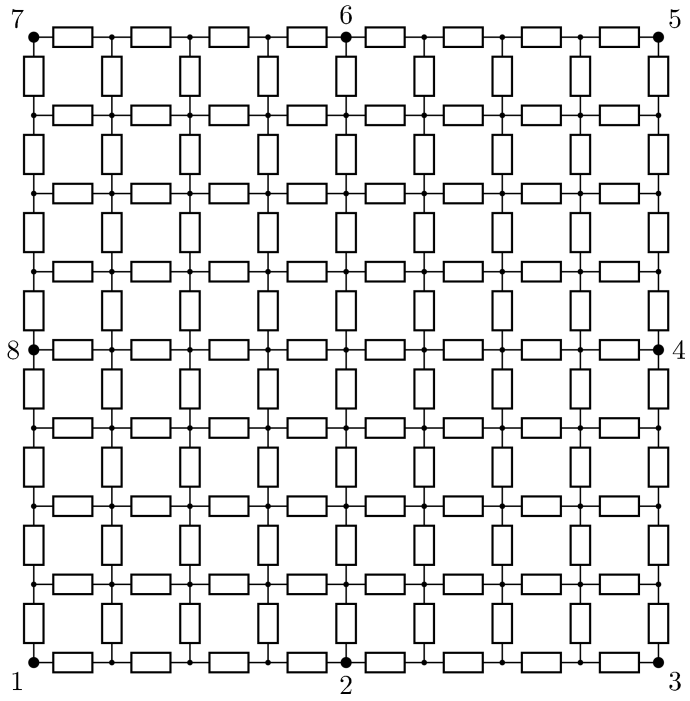


Figure 1.1: Network of resistors with 8 electrodes

possible only from measuring some voltages at the boundary of the network. Therefore, some nodes at the boundary are appointed as "electrodes", where currents can be injected and voltages can be measured. The other nodes at the boundary and in the interior are not accessible. Alternately, two active electrodes are defined, one for currents injecting and one reference electrode. Therefore, a *discrete inverse problem* has to be solved, because starting from exterior measurement values (which are the *effect*) one calculates the *cause* of the measurement values, which are the resistor values. So one has a discrete number of values. It will be shown, that this problem is a nonlinear one and has to be solved iteratively. The solution strategies for a linear problem are also discussed, because nonlinear problems are typically solved by successive linearization.

1.3 Thesis Outline

- In the **first chapter** the general problem is discussed, the forward problem is determined and on the basis of a small network the structure of the problem is analyzed. Also the main characteristics of inverse problems are reviewed. Additionally the structure of the measurements and the sensitivity of these measurements are presented.
- The **second chapter** contains the Optimization and Regularization theory. With the help of an one-dimensional linear problem the main characteristics are discussed. The implemented algorithms are presented and the line search procedure used is shown.
- In the **third chapter** the common calculations of the regularization parameter (iteratively and non iteratively) are described.

- In the **fourth chapter** the determination of the Regularization Matrix of the Tikhonov Functional is discussed and different forms of the matrix are presented.
- In the **fifth chapter** the Total Variation method is applied to the present problem and the differences to the Tikhonov Functional are shown.
- In the **sixth chapter** different reconstruction methods are applied to the circuit problem and the results are discussed. MATLAB [2] is used for all calculations in this work.

1.4 Problem Description

1.4.1 Forward Problem

The forward problem is analyzed by defining the graph of the given circuit made of resistors and sources, which consists of branches and nodes. The branches are the line segments containing the resistors (and sources) and the nodes are the terminal points of the branches.
Therefore, the general branch b_k of a linear time invariant network is considered (see figure 1.2) [3, 4]. All initial conditions are assumed to be included in the independent sources. Then the

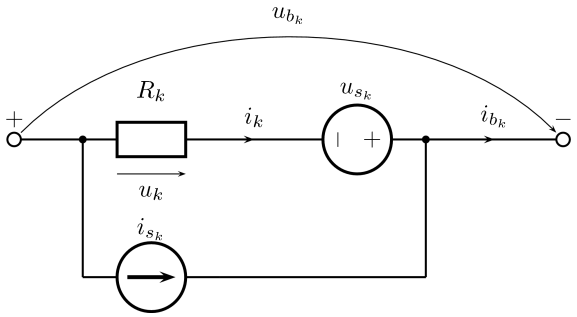


Figure 1.2: General branch b_k

branch current i_{b_k} is given as the sum of the current source i_{s_k} and the element current i_k. Similarly, the branch voltage u_{b_k} is the algebraic sum of the voltage source u_{s_k} and the voltage drop across the element u_k.

$$i_{b_k} = i_k + i_{s_k} \ , \ k = 1, 2, \ldots, B$$
$$u_{b_k} = u_k - u_{s_k} \ , \ k = 1, 2, \ldots, B$$

for all B branches present in the network. The currents and voltages can be written in vector form:

$$\begin{aligned}
\mathbf{i}_b &= (i_{b_1}, i_{b_2}, \ldots, i_{b_B})^T \\
\mathbf{i}_s &= (i_{s_1}, i_{s_2}, \ldots, i_{s_B})^T \\
\mathbf{i} &= (i_1, i_2, \ldots, i_B)^T \\
\mathbf{u}_b &= (u_{b_1}, u_{b_2}, \ldots, u_{b_B})^T \\
\mathbf{u}_s &= (u_{s_1}, u_{s_2}, \ldots, u_{s_B})^T \\
\mathbf{u} &= (u_1, u_2, \ldots, u_B)^T
\end{aligned}$$

Then one gets

$$\mathbf{i}_b = \mathbf{i} + \mathbf{i}_s \tag{1.1}$$

$$\mathbf{u}_b = \mathbf{u} - \mathbf{u}_s \tag{1.2}$$

Each element voltage u_k is related to the corresponding element current i_k through the equation

$$u_k = R_k i_k \tag{1.3}$$

since in branch b_k there is only a resistor with resistance R_k in the present case. This equation can also be converted to

$$i_k = \frac{1}{R_k} u_k = Y_k u_k$$

where $Y_k = 1/R_k$ is the conductance of the resistor. For further considerations the equation is needed in matrix form

$$\mathbf{i} = \mathrm{Y}\ \mathbf{u} \tag{1.4}$$

where Y is the element admittance matrix, a diagonal matrix consisting the conductances of the respective branches. To obtain the set of equations corresponding to the forward problem the nodal analysis method [4] is used.
To derive this method Kirchhoff's Current Law (KCL) is needed [3]: For any lumped electrical network the algebraic sum of the currents entering or leaving a node is equal zero. This can be arranged in matrix form for all branches

$$\mathrm{A_a}\mathbf{i}_b = \mathbf{0} \tag{1.5}$$

where $\mathrm{A_a}$ is said to be the augmented incidence matrix. It consists of the elements a_{jk},which are defined according to the direction of the directed branches b_k. (Every branch is getting a distinct direction according to the assumed direction of the current flow.)

$a_{jk} = +1$ when b_k in incident to the node n_j and is directed away from it

$a_{jk} = -1$ when b_k is incident to n_j and is directed toward it

$a_{jk} = 0$ when b_k is not incident to n_j

Not all of the equations resulting from equation (1.5) on the preceding page are linearly independent, because one row can always be represented through the others. By deleting one equation, the set of nodal equations becomes linearly independent. This can be written in matrix form

$$\mathrm{A}\mathbf{i}_b = \mathbf{0} \tag{1.6}$$

where A is the incidence matrix. It is obtained by deleting one row of the augmented incidence matrix A_a. The node corresponding to the deleted row is called the reference node. Additionally the nodal voltages u_{n_j} are defined: These are the voltages from the corresponding node to the reference node. The vector of the nodal voltages $\mathbf{u}_n$ is defined as

$$\mathbf{u}_n = (u_{n_1}, u_{n_2}, \ldots, u_{n_N})^T$$

and the relation between the nodal voltages and the branch voltages is given as

$$\mathbf{u}_b = \mathrm{A}^T \mathbf{u}_n.$$

To obtain the equations without the branch voltages, these are replaced in equation (1.2) on the previous page

$$\mathbf{u} - \mathbf{u}_s = \mathrm{A}^T \mathbf{u}_n \tag{1.7}$$

Similarly the branch currents vector $\mathbf{i}_b$ is replaced in equation (1.6) with equation (1.1) on the previous page:

$$\mathrm{A}\ \mathbf{i}_b = \mathrm{A}(\mathbf{i} + \mathbf{i}_s) = \mathbf{0} \quad \rightarrow \quad \mathrm{A}\ \mathbf{i} = -\mathrm{A}\ \mathbf{i}_s,$$

Then the element currents are replaced through equation (1.4) on the preceding page to obtain

$$\mathrm{A}\ \mathrm{Y}\mathbf{u} = -\mathrm{A}\ \mathbf{i}_s$$

and the element voltages are replaced with equation (1.7)

$$\mathrm{A}\ \mathrm{Y}\ \mathrm{A}^T \mathbf{u}_n = \mathrm{Y}_n\ \mathbf{u}_n = -\mathrm{A}\ \mathrm{Y}\mathbf{u}_s - \mathrm{A}\ \mathbf{i}_s$$

where Y_n is called the nodal admittance matrix. In the presented case only a current source is present, hence the term with the voltage sources is neglected.

$$\mathrm{A}\ \mathrm{Y}\ \mathrm{A}^T \mathbf{u}_n = \mathrm{Y}_n \mathbf{u}_n = -\mathrm{A}\mathbf{i}_s$$

Finally the nodal voltage vector is obtained with the equation

$$\mathbf{u}_n = -\mathrm{Y}_n^{-1} \mathrm{A}\mathbf{i}_s \tag{1.8}$$

which is calculated in MATLAB [2]. But only few nodal voltages (according to the measurable voltages) are needed. The result is the vector of nodal voltages $\mathbf{u}(\mathbf{r})$, which can be compared with the measurement voltages. Since there are no real measurements available, the vector of node voltages is set equal to a precalculated vector biased with noise. The noise is assumed to be

normally distributed with zero mean and standard deviation σ. To analyze the structure of the present problem the following small network is considered.

1.4.2 Small Network

The determination of the resistances from measured voltages in a small network with only 4 nodes (figure 1.3) is analyzed and the structure of the forward problem is shown. On the basis of

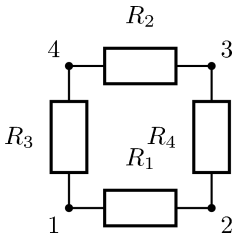


Figure 1.3: Small network of resistors

equation (1.8) on the previous page the problem is formulated. First the reference and injection nodes have to be defined. Node 1 is selected to be the reference node and node 2 to be the injection node with the injection current i_s, see figure 1.4. Also the nodal voltages are defined in the figure.

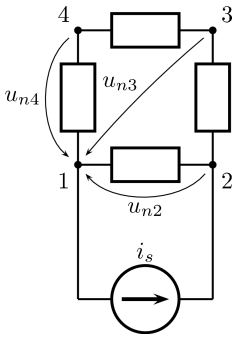


Figure 1.4: Network of resistors with source current and nodal voltages

To calculate them, first the directed graph has to be defined (see figure 1.5). With this graph the

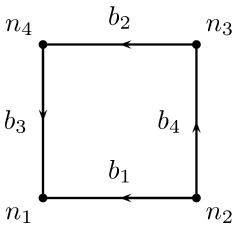


Figure 1.5: Directed graph of the network

augmented incidence matrix A_a can be determined as

$$\mathrm{A}_a = \begin{pmatrix} -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & -1 & 1 & 0 \end{pmatrix}$$

The first row of A_a is deleted because this row corresponds to node 1, which is defined as the reference node. According to equation (1.8) on page 5 the node admittance matrix Y_n is calculated

$$\begin{aligned} \mathrm{Y}_n &= \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{R_1} & 0 & 0 & 0 \\ 0 & \frac{1}{R_2} & 0 & 0 \\ 0 & 0 & \frac{1}{R_3} & 0 \\ 0 & 0 & 0 & \frac{1}{R_1} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{R_1} + \frac{1}{R_4} & -\frac{1}{R_4} & 0 \\ -\frac{1}{R_4} & \frac{1}{R_2} + \frac{1}{R_4} & -\frac{1}{R_2} \\ 0 & -\frac{1}{R_2} & \frac{1}{R_2} + \frac{1}{R_3} \end{pmatrix}. \end{aligned}$$

Then the complete system of equations can be written as

$$\begin{pmatrix} \frac{1}{R_1} + \frac{1}{R_4} & -\frac{1}{R_4} & 0 \\ -\frac{1}{R_4} & \frac{1}{R_2} + \frac{1}{R_4} & -\frac{1}{R_2} \\ 0 & -\frac{1}{R_2} & \frac{1}{R_2} + \frac{1}{R_3} \end{pmatrix} \begin{pmatrix} u_{n_2} \\ u_{n_3} \\ u_{n_4} \end{pmatrix} = \begin{pmatrix} i_s \\ 0 \\ 0 \end{pmatrix}$$

Now the nodal voltages have to be calculated so that the later used structure $\mathbf{u}(\mathbf{r}) = \mathbf{u}_n$ can be analyzed. The following solution is obtained:

$$\begin{aligned} u_{n_2} &= \frac{i_s R_1 \left(R_2 + R_3 + R_4\right)}{R_1 + R_2 + R_3 + R_4} \\ u_{n_3} &= \frac{i_s R_1 \left(R_2 + R_3\right)}{R_1 + R_2 + R_3 + R_4} \\ u_{n_4} &= \frac{i_s R_1 R_3}{R_1 + R_2 + R_3 + R_4} \end{aligned} \tag{1.9}$$

This result can be written as

$$\mathbf{u}_n = \mathbf{u}(\mathbf{r}, \mathbf{i}_s) \tag{1.10}$$

where $\mathbf{r}$ is the vector of resistances and is defined as $\mathbf{r} = (R_1, R_2, R_3, R_4)^T$.
For further considerations the general characteristics of linear problems are given [5]:

- Principle of superposition: $\mathrm{G}(\mathbf{x}_1 + \mathbf{x}_2) = \mathrm{G}(\mathbf{x}_1) + \mathrm{G}(\mathbf{x}_2)$ and
- Principle of homogeneity or scaling: $\mathrm{G}(\alpha\, \mathbf{x}) = \alpha\, \mathrm{G}(\mathbf{x})$ where α is a scalar

The vector $\mathbf{x}$ is equivalent to the vector of resistances $\mathbf{r}$ and the matrix G would be equivalent to a matrix U, which is independent from $\mathbf{r}$, but this representation is not possible for the equation (1.10). It can be seen that the principle of scaling is valid, but the principle of superposition is violated, hence the forward problem is a nonlinear problem. The dependency of the nodal voltage

vector $\mathbf{u}_n$ on the source or injection current i_s is of course a linear relation (Ohm's law), but the remaining entire resistance is a relation of all resistances. Of course, if all resistors have the same resistance the problem is linear again. It should be marked that the resistors are linear elements, but the structure of the elements in equation (1.9) on the preceding page leads to a nonlinear system.

1.4.3 Structure of the Network

The present forward problem according to the network in figure 1.1 on page 2 can also be described as

$$\mathbf{u}(\mathbf{r}, \mathbf{i}_s) = \mathbf{u},$$

where $\mathbf{u}(\mathbf{r}, \mathbf{i}_s)$ is the system, dependent on the known source current vector $\mathbf{i}_s$ and the vector of the unknown resistances $\mathbf{r}$, and $\mathbf{u}$ are the resulting voltages at the boundary electrodes. For future computations $\mathbf{u}(\mathbf{r}, \mathbf{i}_s)$ is only described as $\mathbf{u}(\mathbf{r})$ and $\mathbf{i}_s$ is neglected, because the source current i_s is a given constant value.

$$\mathbf{u}(\mathbf{r}) = \mathbf{u}.$$

Now the structure of the present problem is analyzed:
There are $m = 9$ rows of nodes and $p = 9$ columns of nodes, hence the number of nodes is $n = m \cdot p = 81$. The nodes, where the voltages can be measured, are defined as the electrodes. They are positioned at the corners and in the middle of the edges and they are named from 1 to 8 counterclockwise (see figure 1.1 on page 2). Thus there are $e = 8$ electrodes with access for measurements. The number of unknown resistances is calculated by $b = p{\cdot}(m{-}1){+}m{\cdot}(p{-}1) = 144$. One electrode is defined as reference electrode, hence there exist $n - 1 = 80$ nodal voltages but only $e - 1 = 7$ are measurable. With the circular method $i = 8$ successive measurements are made, hence the whole number of measured voltages is given by $v = (e - 1) \cdot i = 56$. Because of the reciprocity of the problem only one half is linearly independent [6]. This can be displayed by calculation of the system matrix $\mathrm{J}^T\mathrm{J}$ (where J is denoted as the Jacobian matrix) which has always the rank of $(e{-}1){\cdot}i/2 = 28$. There is always noise present in practical applications. For this reason it is usual to use all measurements for the reconstruction [6]. Hence, for the solution of the inverse problem there are 144 unknown resistors and 28 linearly independent voltage measurements. This is a highly underdetermined or rank deficient problem, therefore, another configuration with 12 electrodes is used (see figure 1.6 on the next page). Then the rank of the matrix $\mathrm{J}^T\mathrm{J}$ equals $(e - 1) \cdot i/2 = (12 - 1) \cdot 12/2 = 66$. Compared to the configuration with 8 electrodes more information about the resistances is available but it is still a rank deficient problem.

1.4.4 Determination of the Jacobian Matrix

The Jacobian matrix J is an important matrix of the given system and is defined as the first derivatives of the model function.

$$\mathrm{J} = \frac{\partial \mathbf{u}(\mathbf{r})}{\partial \mathbf{r}} = [\nabla u_1, \nabla u_2, \ldots, \nabla u_v]^T$$

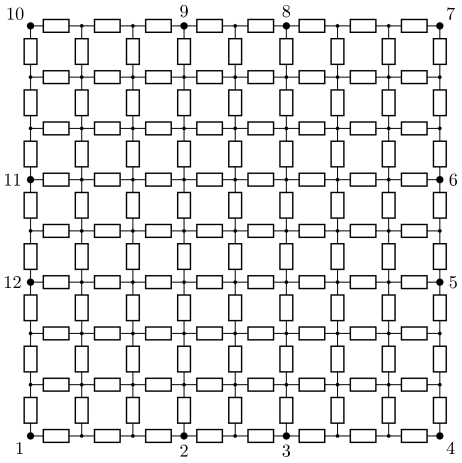


Figure 1.6: Configuration with 12 electrodes

and is needed for solving the inverse problem. It can be interpreted as the mapping between the distribution of the voltages at the boundary and the distribution of the resistor values [7]. But also it is a norm of the sensitivity of the problem, hence it is called the sensitivity matrix. In the given case it is not possible to compute the first derivatives of the model function analytically, therefore the Jacobian matrix J is approximated using finite difference operators repeatedly. Every resistance R_i is shifted separately by a small value $\delta(\delta = 10^{-4})$, and the forward problem is computed to get the shifted nodal voltages $\widetilde{\mathbf{u}}$ for each shifted resistor. The computation of each coefficient of the Jacobian matrix can be described as

$$\mathrm{J}_{i,j} = \frac{\partial u_j}{\partial r_i} \cong \frac{\widetilde{u}_{i,j} - u_j}{\delta} \quad \text{with } i = 1 \dots b,\ j = 1 \dots v,$$

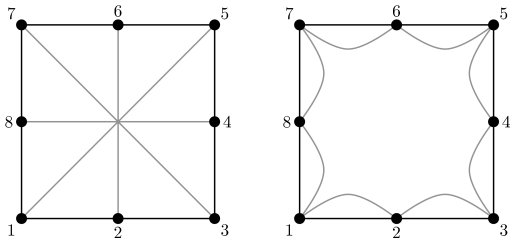
where b denotes the number of resistors and v denotes the number of voltages along the boundary. Here one can see why it is also termed sensitivity matrix. It displays the sensitivity of the boundary voltages according to small perturbations of the resistances. The computation of the Jacobian matrix requires $b = 144$ solutions of the forward problem, which is no problem for the given network problem but for more time consuming forward problems other computation methods have to be used (see for example [8]).

1.4.5 Current Pattern and Sensitivity Analysis

In principle all combinations of reference and injection node are allowed, but the two main patterns (combinations of reference and injection electrode) shall be analyzed here:

- the circular method: the reference electrode and the injection electrode are abreast each other.
- the opposite method: the reference electrode and the injection electrode lie opposite to each other.

For the given problem, 8 fixed electrodes (one at each vertex and one in the middle of each edge) are used. Figure 1.7 on the following page shows the two schematics, the circular method on



(a) opposite method (b) circular method

Figure 1.7: Schematic of the current pattern

the right and the opposite method on the left. To analyze the two patterns, 10 resistors of the network are selected (see figure 1.8) and one after another is perturbed with a small constant and the forward problem is computed for each pattern to display the change of the outside values according to a change of an inside value. Hence, the sensitivity computation equals the Jacobian matrix computation for the selected resistors.

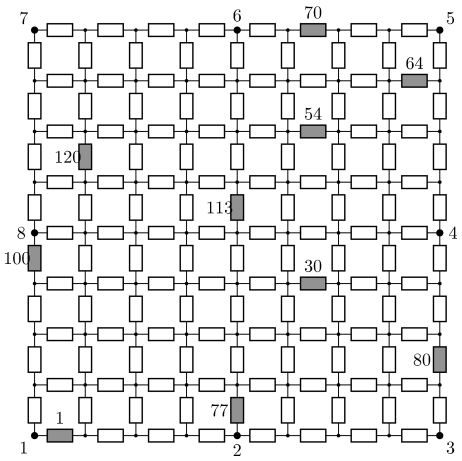


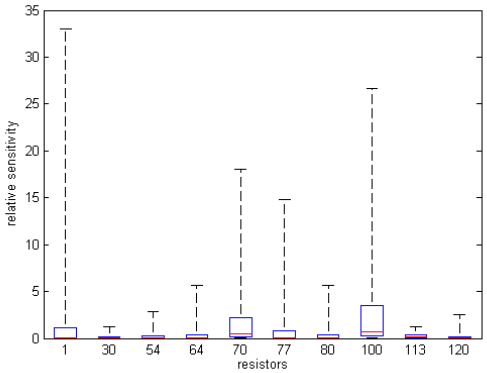
Figure 1.8: Selected resistors for sensitivity analysis

The difference is, that only the size of the change of the voltage values and not its direction is needed. Therefore, the absolute values are used. Because of the dependency of the sensitivity from the present resistor values, this computation is made for four different distributions (The unit of the resistor values (Ω) is not specified.):

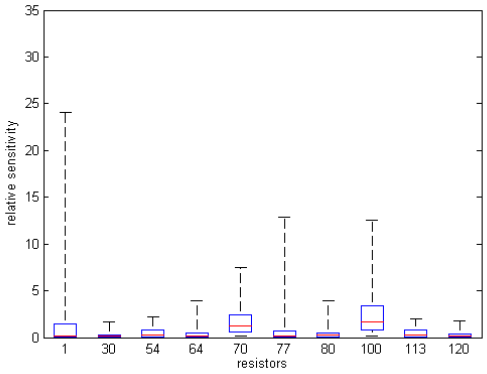
- homogeneous distribution, all resistor values are $R = 3$
- circuit 1: two different objects: one small object with the values $R = 3$ and one big object with the values $R = 5$ on a background with $R = 1$
- circuit 2: two different objects: one small object with the values $R = 5$ and one big object with the values $R = 3$ on a background with $R = 1$
- circuit 3: one big object with the values $R = 5$ on a background with $R = 1$

The first distribution equals to the starting distribution for the reconstruction processes in chapter 6 and the other three distributions equal to the circuits in chapter 6. In order to compare the

computed values, the values of each distribution are correlated to the mean sensitivity value of the distribution. The resulting values for every resistor are plotted in figure 1.9 It can be seen,



(a) circular method



(b) opposite method

Figure 1.9: Relative sensitivity values for the selected resistors. The edges of the box are the 25th and 75th percentiles, the red line marks the median, the black line marks the complete range.

that the sensitivity for resistors at the boundary is higher than the one for resistors inside the network. Especially next to the electrodes there are very high maximum values and the mean range of the values is also bigger than elsewhere. Comparing the two figures for the different patterns the maximum values for the boundary resistors are a bit smaller in the opposite case than in the circular case, but the sensitivity for the inside values are about as worse as in the circular case. For the calculations in chapter 6 only the circular method is used, because there are only insignificant differences to be expected in the reconstruction results.
For other practical situations where the computation time is a main problem it may be useful to choose fewer combinations, because with growing amount of combinations the computation time grows. Since the accuracy of the results is also proportional to the number of combinations, a good trade off has to be found.

1.4.6 Inverse Problems

The forward problem is defined as the problem if the system parameters and the input values are given and the output values are sought.
In opposite to the forward problem the inverse problem can be defined as

- the problem, if the input and output values are given and the system parameters are sought,

or as

- the problem if the output values and the system parameter are given and the input values are sought.

For the present problem the input and output values are given and the system parameters, the resistor values are sought.
An important issue of inverse problems is, whether the problem is well-posed or ill-posed. Well-posedness implies that the solution can easily be determined by inversion, since there exists a well-defined, continuous inverse operator [9]. In the beginning of the 20th century Hadamard defined the conditions of well-posed problems:

- The solution exists
- and the solution is unique
- and the solution depends continuously on the data.

For all admissible data, if one of the conditions is violated the problem is named ill-posed [10, 5]. The third item can additionally be described such a way, that the solution is stable with respect to perturbations in the data. It is violated if for example an arbitrarily small perturbation of the data causes an arbitrarily large perturbation of the solution.

2 General Optimization and Regularization Methods

2.1 Introduction

The solution of an inverse problem is often connected to the solution of an optimization problem. Therefore, the basic knowledge about optimization methods is crucial and will be described in the first section. Furthermore, some Regularization techniques are needed as well for ill-posed problems. This will be discussed in the second section. Most of the regularization methods can be expressed as or combined with an optimization method like in the presented algorithms in the third section.
So first of all optimization methods are introduced. Starting from the the result of the forward problem, which has the form

$$\mathbf{u}(\mathbf{r}) = \mathbf{u}, \tag{2.1}$$

where $\mathbf{u}(\mathbf{r})$ are the simulated voltages derived from the forward model and $\mathbf{u}$ is the resulting boundary voltage vector. Now one could expect that the boundary voltages $\mathbf{u}$ are equal the corresponding measured voltages $\mathbf{u}_r$ at the boundary, but in real world problems the measurement values are never exactly equal to the solution of the forward model either because of noise or because of the approximation of the the model or some environmental influences. Therefore, equation (2.1) will not exist in this form (see the Hadamard conditions). To ensure that a solution exists in every case a least squares minimization problem has to be formulated minimizing the difference of the two voltage vectors, which is named the residuals $\overline{\mathbf{r}} = \mathbf{u}(\mathbf{r}) - \mathbf{u}_r$. Hence the minimization functional $\psi(\mathbf{r})$ can be developed as

$$\min_{\mathbf{r}} \psi(\mathbf{r})$$

with

$$\psi(\mathbf{r}) = ||\mathbf{u}(\mathbf{r}) - \mathbf{u}_r||_2^2. \tag{2.2}$$

This is a general optimization problem one has to solve as accurately as possible. Equation (2.2) can be written as the scalar product of $\overline{\mathbf{r}}^T$ and $\overline{\mathbf{r}}$.

$$\psi = \overline{\mathbf{r}}^T \overline{\mathbf{r}}.$$

In case of a linear model $\mathrm{K}\mathbf{f} = \mathbf{d}$, which will be used to introduce different approaches, one gets

$$\psi(\mathbf{f}) = ||\mathrm{K}\mathbf{f} - \mathbf{d}||_2^2, \tag{2.3}$$

where K is the model matrix, $\mathbf{f}$ are the sought discrete function values and $\mathbf{d}$ are the discrete data values (measurements). The notation is selected according to the linear example which will be introduced now. According to [9, p. 1] the Fredholm first kind integral equation of convolution type in one space dimension is used to explain linear methods in the next sections. This equation occurs in two-dimensional optical imaging, but here a one dimensional version only is used.

$$g(x) = \int_0^1 k(x - x^*)\, f(x^*) dx^*, \quad 0 < x < 1,$$

There f represents the light source intensity as a function of spatial position, g represents the image intensity, and k is called the kernel and characterizes the blurring effects that occur during the image formation. Here the Gaussian kernel is used because it models the long-time average effects of atmospheric turbulence of light propagation [9]. The one-dimensional form is:

$$k(x - x^*) = c\, e^{\frac{-(x-x^*)^2}{2\gamma^2}}$$

where γ and c are positive parameters and they are set to $\gamma = 0.025$ and $c = 1/(\gamma\sqrt{2\pi})$. The resulting equation becomes

$$g(x) = \int_0^1 c e^{\frac{-(x-x^*)^2}{2\gamma^2}} f(x^*) dx^*, \quad 0 < x < 1. \tag{2.4}$$

This equation defines the forward problem, because for a given function f and a typically smooth kernel function k the function g can be calculated. Contrary the inverse problem is defined with the given kernel k and the given blurred image g, while the source f has to be calculated.
The equation equation (2.4) needs to be discretized, which is done by using collocation in the independent variable x and quadrature in x^*. The discrete linear system $\mathrm{K}\mathbf{f} = \mathbf{d}$ is generated with the entries of K, which, if midpoint quadrature is applied, are defined as [9]

$$\mathrm{K}(i,j) = \frac{1}{n} c e^{-\frac{((i-j)\frac{1}{n})^2}{2\gamma^2}}, \quad 1 \le i, j \le n,$$

where n is the number of discrete data points. A certain amount of data points is necessary to obtain an accurate quadrature approximation, hence n must be relatively large. But the matrix K becomes more ill-conditioned with increasing data points, and therefore errors in $\mathbf{d}$ are amplified during the reconstruction process to obtain $\mathbf{f}$. To show this effect, the constructed blurred data $\mathbf{g}$ are additionally perturbed with noise (leading to the data vector $\mathbf{d}$) to take into account errors which are always present in practical applications. The true distribution $\mathbf{f}$ and the data points, which are used to create the data vectors $\mathbf{g}$ and $\mathbf{d}$ are shown in figure 2.1 on the following page, the blurred data $\mathbf{g}$ and the noisy data $\mathbf{d}$ are displayed in figure 2.2 on the next page. The data vector $\mathbf{d}$ is used to reconstruct the distribution $\mathbf{f}$. The straightforward approach is the solution of $\mathbf{f} = \mathrm{K}^{-1}\mathbf{d}$. This fails, due to the amplification of the present errors in the data, as can be seen in figure 2.3 on page 16.

If no noise is present, the reconstruction of the source is successful, although the limited resolution due to the 80 data points is visible, see figure 2.4 on page 16.

Now some standard deterministic methods for solving this optimization problem are described.

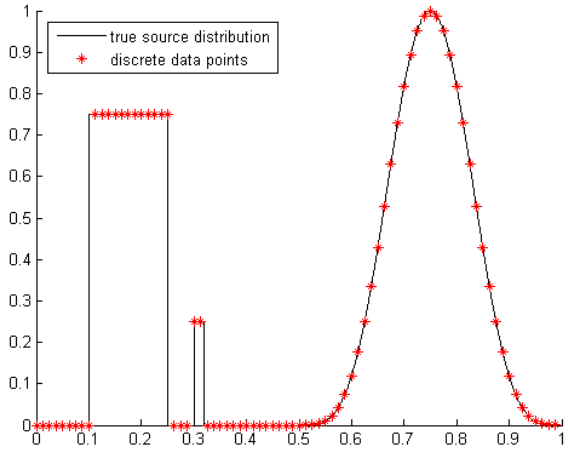


Figure 2.1: The linear problem: The true source distribution **f** (black line) and the 80 data points which are used to create the data vectors

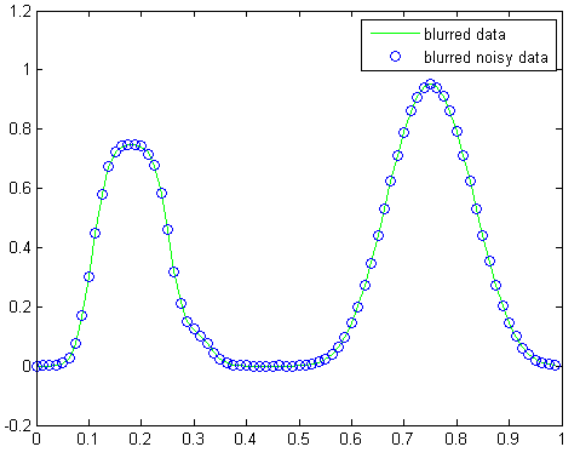


Figure 2.2: The blurred image **g** (green line) and noisy data **d** (blue circles)

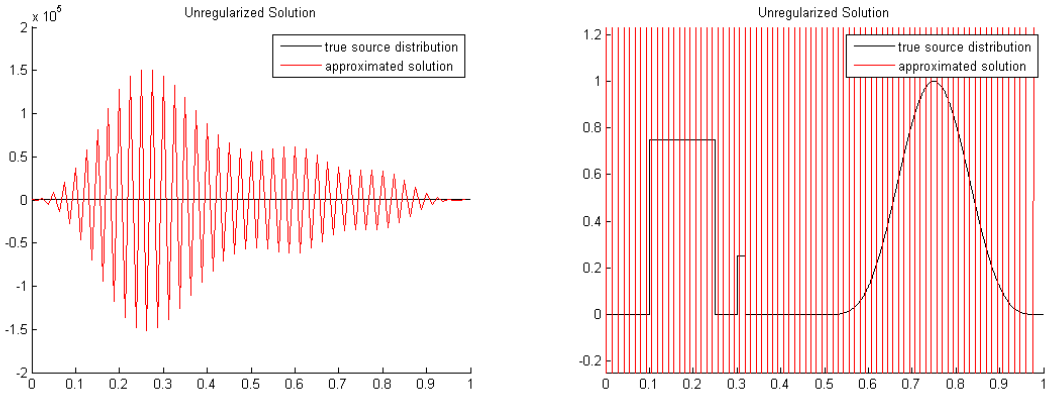


Figure 2.3: The true distribution $\mathbf{f}$ (is not visible on the left due to the dimension of the y-axis) and the simple inverse solution obtained by $\mathbf{f} = \mathrm{K}^{-1}\mathbf{d}$.
On the right a detailed view including the true distribution is shown.

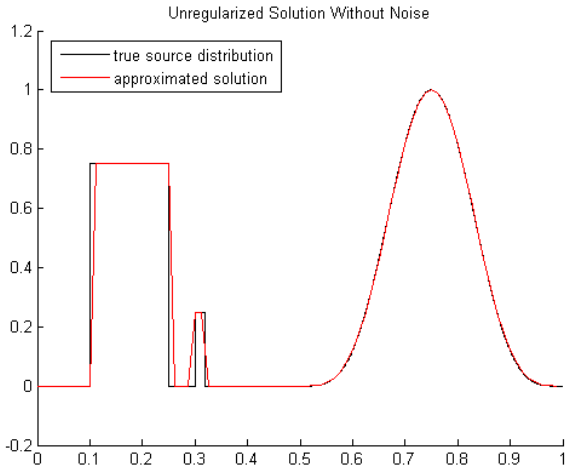


Figure 2.4: The true distribution $\mathbf{f}$ and the simple inverse solution obtained by $\mathbf{f} = \mathrm{K}^{-1}\mathbf{d}$, if no noise is present

- The Steepest Descent Method, where the search direction for the next parameter values is chosen as the negative gradient of the functional,

- The Standard Newton Method, using first and second order information,

- The Gauss Newton Method, which is a special form of the Newton Method and

- The Levenberg-Marquardt Method, a trust region method (contrary to step length methods) and can be seen as something between the steepest descent and the Newton method.

There are other practical, often used methods like the Conjugate Gradient Method, which avoids the direct calculation of the Hessian, but approximates this matrix during the iteration process (see [11, 9] for example). Because of the simple forward problem one does not need such approximations. There are other special Newton-type methods, the Quasi-Newton method for example, which are also not described here. They are shown in [11, 12] for example.
The result of the Newton method and the steepest descent method is a search direction. This direction is a decent direction, but a full step in this direction does not ensure to arrive the minimum value. Therefore the minimum value has to be searched along this direction. This essential algorithm, which is part of most optimization methods is named *line search procedure* and is described in the last subsection. Only the Levenberg-Marquardt method doesn't need a line search because the trust region restricts the values and the corresponding variable has to be adjusted appropriately.
This basic optimization methods are needed to solve inverse problems, but additional measures have to be taken. It only ensures, that a solution exists [13]. The remaining two points of Hadamard's conditions, that the solution is unique and stable, are not ensured when just minimizing the functional. Therefore, the optimization problem has to be modified. In other words the problem needs some regularization. Regularization means, that additional information or assumptions of the system are added to get a useful solution.
The regularization methods, which are treated in detail, are:

- Truncated Singular Value Decomposition
- Standard Tikhonov Regularization Method
- Landweber Iteration Method
- Total Variation Method

The last section forms the basis of the implemented algorithms, which are often combinations of optimization and regularization methods.

- Regularized Gauss-Newton Algorithm
- Occam's Inversion
- Regularization with additional box constraints, a special "Active Set" method

2.2 Standard Optimization Methods

2.2.1 Steepest Descent Method

The steepest descent method is a simple first order iterative method (without computation of the Hessian matrix). To compute the search direction, the negative gradient is used, hence only first derivatives are necessary. In the linear case, the functional which should be minimized is given by

$$\psi(\mathbf{f}) = \frac{1}{2}||\mathrm{K}\mathbf{f} - \mathbf{d}||_2^2$$

The factor 1/2 is just added to avoid the factor 2 in the gradient which is given by

$$\mathbf{g} = \nabla\psi(\mathbf{f}) = \mathrm{K}^T(\mathrm{K}\mathbf{f} - \mathbf{d}).$$

The gradient vector of a scalar function defines the direction and the magnitude of maximum increase in the functions values at a given point [14, p. 102]. Hence the negative gradient represents the steepest descent of the function. Whenever the gradient is not equal to zero, the search direction $\mathbf{s}$ is automatically a descent direction because of $\mathbf{g}^T\mathbf{s} = -||\mathbf{g}||_2^2$ [14, p. 102] with

$$\mathbf{s} = -\mathbf{g} = -\nabla\psi(\mathbf{f})$$

in a given point. The update step to obtain the next iterate of $\mathbf{f}$ ($\mathbf{f}_{k+1}$) is given by

$$\mathbf{f}_{k+1} = \mathbf{f}_k + \gamma\mathbf{s}_k,$$

where the step size γ is obtained by a line search algorithm in order to minimize the functional ψ along the search direction $\mathbf{s}_k$. This can be stated as

$$\min_{\gamma>0} \ \psi(\mathbf{f}_k + \gamma\mathbf{s}_k)$$

In every iteration k the gradient $\mathbf{g}_k$ and the step size γ_k have to be calculated.
The following simple algorithm is given by [9, p. 35]

$k = 0$
$\mathbf{f}_0 =$ *inital guess*
begin loop
  $\mathbf{s}_k = -\nabla\psi(\mathbf{f}_k)$ compute the negative gradient
  $\gamma_k = \min_{\gamma>0} \psi(\mathbf{f}_k + \gamma\mathbf{s}_k)$ line search
  $\mathbf{f}_{k+1} = \mathbf{f}_k + \gamma_k\mathbf{s}_k$ update
  $k = k + 1$
end loop

The steepest descent method is a standard numerical optimization tool, but there are some disadvantages about it [14, p. 102]: Usually a large number of iterations is needed, because due the orthogonality of the successive gradients ($\mathbf{g}_k^T \mathbf{g}_{k+1} = 0$), a zigzag path is generated. Additionally the convergence rate of an quadratic functional is linear, but can be very slow in more general, non quadratic cases, especially if the Hessian is ill-conditioned [9, p. 31].

2.2.2 Standard Newton Method

Newton-based methods are quite often referred to when using regularization methods. So a short description of Newton's method is given here. Newton's method is an iterative method for linear and nonlinear problems and is constructed in the following way [11]. Starting from a quadratic model derived from Taylor series expansion of $\psi(\mathbf{x})$ about $\mathbf{x}_k$, which can be written as

$$\psi(\mathbf{x}_k + \boldsymbol{\delta}) \approx \widetilde{\psi}_k(\boldsymbol{\delta}) = \psi(\mathbf{x}_k) + \mathbf{g}_k^T \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}^T \mathrm{H}_k \boldsymbol{\delta}$$

where $\boldsymbol{\delta} = \mathbf{x} - \mathbf{x}_k$ and $\widetilde{\psi}_k$ is the quadratic approximation of $\psi(\mathbf{x}_k + \boldsymbol{\delta})$. The two other variables denote the gradient $\mathbf{g}_k$ and the Hessian matrix H_k which are known. The next iterate is defined as $\mathbf{x}_{k+1} = \mathbf{x}_k + \boldsymbol{\delta}$, where the correction $\boldsymbol{\delta}$ minimizes $\widetilde{\psi}_k(\boldsymbol{\delta})$. To get a unique minimizer the Hessian H_k must be positive definite. The minimizer $\boldsymbol{\delta}$ is calculated by setting the first derivative of $\widetilde{\psi}_k(\boldsymbol{\delta}_k)$ to zero. So Newton's method for the step k can be written

- solve $\mathrm{H}_k \boldsymbol{\delta} = -\mathbf{g}_k$ for $\boldsymbol{\delta} = \boldsymbol{\delta}_k$,
- set $\mathbf{x}_{k+1} = \mathbf{x}_k + \boldsymbol{\delta}_k$

The correction $\boldsymbol{\delta}_k$ is named search direction or Newton direction [15, 11] $\mathbf{s}_k$. The classical Newton method uses a unit step with $\gamma = 1$, but it can be improved by the line search [14, p. 106]. With additional line search, which determines the step size γ the update step follows

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \gamma_k \mathbf{s}_k.$$

This describes the fundamental idea of the standard procedure for the Newton method with line search. Now the linear case with the simple optimization functional is considered:

$$\psi(\mathbf{f}) = \frac{1}{2} ||\mathrm{K}\mathbf{f} - \mathbf{d}||_2^2$$

the gradient $\mathbf{g}_k = \nabla \psi$ is derived as

$$\mathbf{g}_k = \mathrm{K}^T(\mathrm{K}\mathbf{f}_k - \mathbf{d})$$

and the constant Hessian matrix $\mathrm{H} = \nabla^2 \psi$

$$\mathrm{H} = \mathrm{K}^T \mathrm{K}$$

Hence the solution for the next iteration step is

$$\mathbf{f}_{k+1} = \mathbf{f}_k - \gamma(\mathrm{K}^T \mathrm{K})^{-1} \mathrm{K}^T(\mathrm{K}\mathbf{f}_k - \mathbf{d})$$

which is equal to the optimal point in the "single step method" derived by setting the gradient of the functional $\mathbf{g} = \mathrm{K}^T(\mathrm{K}\mathbf{f} - \mathbf{d}) = \mathbf{0}$ with $\gamma = 1$

$$\mathbf{f}_{opt} = (\mathrm{K}^T\mathrm{K})^{-1}(\mathrm{K}^T\mathbf{d})$$

If the model $\mathrm{K}\mathbf{f} - \mathbf{d}$ is accurate and the Hessian matrix is positive definite, then only one step is necessary to find the minimum of the quadratic model from every starting point [15, 14]. For not accurate models the iterative Newton method is preferred.
There is another modification of Newton's method, an iterative method where the solution is computed without calculation of the real Hessian matrix. This is called the Quasi-Newton method. Only first order derivatives (the gradients) are computed, the Hessian matrix is approximated with the help of the previous iteration and the first order derivatives ("update formulas") [14, p. 107].

2.2.3 Gauss-Newton Method

The Gauss-Newton method is a modified Newton method for nonlinear least squares [14]. The functional of the network is used to derive the iterative Gauss-Newton scheme [13]:

$$\psi(\mathbf{r}) = ||\mathbf{u}(\mathbf{r}) - \mathbf{u}_r||_2^2 \tag{2.5}$$

and the Taylor series approximation of this functional is

$$\widetilde{\psi}(\mathbf{r}) = \psi(\mathbf{r}_k) + \left(\frac{\partial\psi}{\partial\mathbf{r}}(\mathbf{r}_k)\right)(\mathbf{r} - \mathbf{r}_k) + \frac{1}{2}(\mathbf{r} - \mathbf{r}_k)^T\left(\frac{\partial^2\psi}{\partial\mathbf{r}^2}(\mathbf{r}_k)\right)(\mathbf{r} - \mathbf{r}_k) \tag{2.6}$$

where the first order derivative is defined as the gradient $\mathbf{g}_k = \mathbf{g}(\mathbf{r}_k)$

$$\mathbf{g}_k = \frac{\partial\psi}{\partial\mathbf{r}}(\mathbf{r}_k)$$

and the second order derivative is defined as the Hessian matrix $\mathrm{H}_k = \mathrm{H}(\mathbf{r}_k)$

$$\mathrm{H}_k = \frac{\partial^2\psi}{\partial\mathbf{r}^2}(\mathbf{r}_k)$$

The equation (2.6) at the point $\mathbf{r} = \mathbf{r}_{k+1}$ is given as

$$\widetilde{\psi}(\mathbf{r}_{k+1}) = \psi(\mathbf{r}_k) + \mathbf{g}_k(\mathbf{r}_{k+1} - \mathbf{r}_k) + \frac{1}{2}(\mathbf{r}_{k+1} - \mathbf{r}_k)^T\mathrm{H}_k(\mathbf{r}_{k+1} - \mathbf{r}_k)$$

To get the optimal value, the minimum of the functional above, the first order derivative is set equal zero :

$$\frac{\partial\widetilde{\psi}}{\partial\mathbf{r}}(\mathbf{r}_{k+1}) = \mathbf{g}_k + \mathrm{H}_k(\mathbf{r}_{k+1} - \mathbf{r}_k) = \mathbf{0}$$

Under assumption that the Hessian matrix is invertible it follows

$$(\mathbf{r}_{k+1} - \mathbf{r}_k) = -\mathrm{H}_k^{-1}\mathbf{g}_k$$

Now the gradient is derived from equation (2.5) on the previous page

$$\begin{aligned}\mathbf{g}(\mathbf{r}) &= \frac{\partial \psi}{\partial \mathbf{r}} \\ &= \frac{\partial}{\partial \mathbf{r}} \left[(\mathbf{u}(\mathbf{r}) - \mathbf{u}_r)^T (\mathbf{u}(\mathbf{r}) - \mathbf{u}_r) \right] \\ &= 2 \left(\frac{\partial \mathbf{u}(\mathbf{r})}{\partial \mathbf{r}} \right)^T (\mathbf{u}(\mathbf{r}) - \mathbf{u}_r) \\ &= 2\mathrm{J}(\mathbf{r})^T (\mathbf{u}(\mathbf{r}) - \mathbf{u}_r)\end{aligned}$$

because the Jacobian matrix is $\mathrm{J}(\mathbf{r}) = \frac{\partial \mathbf{u}(\mathbf{r})}{\partial \mathbf{r}}$. The Hessian can be determined as follows

$$\mathrm{H}(\mathbf{r}) = \frac{\partial^2 \psi}{\partial \mathbf{r}^2} \tag{2.7}$$

$$= 2 \left(\frac{\partial \mathbf{u}}{\partial \mathbf{r}} \right)^T \left(\frac{\partial \mathbf{u}}{\partial \mathbf{r}} \right) + \underbrace{2 \sum_k \frac{\partial^2 \mathbf{u}_k(\mathbf{r})}{\partial \mathbf{r}^2} (\mathbf{u}_k(\mathbf{r}) - \mathbf{u}_{r\,k})}_{negligible} \tag{2.8}$$

$$\approx 2\mathrm{J}(\mathbf{r})^T \mathrm{J}(\mathbf{r}) \tag{2.9}$$

The second term in equation (2.8) is negligible in the neighborhood of the solution, because the residuals are small there [14, p. 173]. This is denoted the Gauss-Newton-Hessian approximation. So the gradient $\mathbf{g}_k = \mathbf{g}(\mathbf{r}_k)$ and the Hessian matrix $\mathrm{H}_k = \mathrm{H}(\mathbf{r}_k)$ at iteration k are given as

$$\begin{aligned}\mathbf{g}_k &= 2\mathrm{J}_k^T (\mathbf{u}(\mathbf{r}_k) - \mathbf{u}_r) \\ \mathrm{H}_k &= 2\mathrm{J}_k^T \mathrm{J}_k\end{aligned}$$

where $\mathrm{J}_k = \mathrm{J}(\mathbf{r}_k)$ is the Jacobian matrix at iteration k. The updating correction can be written in the following compact form :

$$\boldsymbol{\delta}\mathbf{r}_{k+1} = -\mathrm{H}_k^{-1}\mathbf{g}_k$$

with $\boldsymbol{\delta}\mathbf{r}_{k+1} = \mathbf{r}_{k+1} - \mathbf{r}_k$.

2.2.4 Levenberg-Marquardt Method

In case the Hessian matrix H_k is not positive definite, a diagonal matrix (in the standard case a weighted identity matrix) is added to ensure that the "modified Hessian matrix" becomes positive definite [14, p. 107] . This is known as the Levenberg-Marquardt method and can be written as

$$(\mathrm{H}_k + \nu \mathrm{I})\mathbf{s}_k = -\mathbf{g}_k.$$

The value ν is chosen in such a way that the modified Hessian matrix $(\mathrm{H}_k + \nu \mathrm{I})$ becomes positive definite and yields a good search direction [14, p. 175]. The exceptional is that the eigenvectors remain unchanged and the eigenvalues are augmented by ν and are positive if ν is positive. This also improves the condition number of the Hessian matrix. This modification can also seen in a different way: For $\nu = 0$ the result is the Newton direction and for $\nu \to \infty$ it becomes the descent direction. So the Levenberg-Marquardt direction lies somewhere in between the Newton

and the steepest descent method [11]. Additionally shall be stated, that the Levenberg-Marquardt method is not a regularization method like the standard Tikhonov method. There is no additional information included like with the Tikhonov penalty term when using the identity matrix. The additional identity matrix has the only purpose to obtain a positive definite Hessian matrix and to ensure stability.
This method is also nicely applicable to the nonlinear least squares case (Gauss-Newton), when the Hessian matrix is approximated. Generally speaking, the unconstrained optimization methods can be splitted into two main classes: *step-length methods* and *trust-region methods* [12, p. 113]. It can be shown that the Levenberg-Marquardt method is a trust region method, because it is related to the problem

$$\begin{aligned} &\min_{\mathbf{s}} \ \mathbf{g}_k^T\mathbf{s} + \frac{1}{2}\mathbf{s}^T\mathrm{H}_k\mathbf{s} \\ &\text{s.t. } ||\mathbf{s}||_2 \leq \Delta \end{aligned}$$

where Δ is the trust region [12, p. 114]. In [12, p. 114] this relation is shown for the two cases if either $\nu = 0$ and $||\mathbf{s}||_2 \leq \Delta$, or $\nu \geq 0$ and $||\mathbf{s}||_2 = \Delta$. In the first case the trust region Δ is large enough so the result equals the general Newton direction $\mathbf{s}$. Otherwise a restriction is active and the trust region bound holds and ν is set to a positive value.

2.2.5 Line Search Procedure

The line search procedure is necessary in Newton based regularization methods for enhanced convergence. The solution of a Newton based regularization method is a search direction, without specification where the minimum lies. Therefore the line search procedure is needed to determine the distance to the minimum value along this search direction. Basically the line search consists of two distinct phases [15, 11]:

- bracketing phase: find an interval $[a, b]$ where the minimum is expected
- sectioning phase: divide $[a, b]$ in sub-intervals $[a^j, b^j]$, which become smaller and smaller, until a point is located as the minimum

For practical usage the standard line search procedure is too much time consuming. There are several forward problem solution and especially also Jacobian matrices have to be computed. Because one line search is needed at each iteration and since the real problem is approximated at each iteration, a reduced and inexact line search procedure is used, which is described here. [16, 17, 18] It uses only the quadratic interpolation for the residual function without a bracketing phase, there are no loops included and no additional Jacobian matrix is necessary. Only one forward problem is calculated for a full Newton step. This simple approximation is only possible because it is a very good approximation of the real function present in the network problem. First a full Newton step is done with

$$\mathbf{r}_{new} = \mathbf{r} + \delta\mathbf{r}$$

and the new voltages $\mathbf{u}_{new}$ are calculated. For each individual function of the residual $\bar{\mathbf{r}}$ a quadratic interpolation is made.

$$\bar{\mathbf{r}}(\gamma) = (\bar{\mathbf{r}}_1 - \bar{\mathbf{r}}'_0 - \bar{\mathbf{r}}_0) \cdot \gamma^2 + \bar{\mathbf{r}}'_0 \cdot \gamma + \bar{\mathbf{r}}_0.$$

The derivative of $\bar{\mathbf{r}}_0$ with respect to γ is calculated as follows:

$$\bar{\mathbf{r}}'_0 = \frac{\mathrm{d}\bar{\mathbf{r}}_0(\mathbf{r})}{\mathrm{d}\gamma} = \frac{\mathrm{d}\bar{\mathbf{r}}_0(\mathbf{r})}{\mathrm{d}\mathbf{r}} \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}\gamma} = \mathrm{J} \cdot \delta\mathbf{r}$$

because of

$$\mathbf{r}(\gamma) = \mathbf{r}(0) + \gamma\delta\mathbf{r}$$

and

$$\mathrm{J} = \frac{\mathrm{d}\bar{\mathbf{r}}_0(\mathbf{r})}{\mathrm{d}\mathbf{r}}$$

So the quadratic interpolation can be seen as

$$\bar{\mathbf{r}}(\gamma) = \mathbf{a}\,\gamma^2 + \mathbf{b}\,\gamma + \mathbf{c}$$

with

$$\mathbf{c} = \bar{\mathbf{r}}_0 = \begin{bmatrix} \mathbf{u}(\mathbf{r}) - \mathbf{u}_r \\ \sqrt{\alpha}\ \mathrm{L}\ \mathbf{r} \end{bmatrix}$$

$$\mathbf{b} = \bar{\mathbf{r}}'_0 = \begin{bmatrix} \mathrm{J}\ \delta\mathbf{r} \\ \sqrt{\alpha}\ \mathrm{L}\ \delta\mathbf{r} \end{bmatrix}$$

$$\mathbf{a} = \bar{\mathbf{r}}_1 - \bar{\mathbf{r}}'_0 - \bar{\mathbf{r}}_0 = \begin{bmatrix} \mathbf{u}_{new} - \mathbf{u}_r - \mathrm{J}\ \delta\mathbf{r} - (\mathbf{u} - \mathbf{u}_r) \\ \mathbf{0} \end{bmatrix}$$

The complete functional f is calculated from

$$\begin{aligned} f(\gamma) &= \mathbf{r}(\gamma)^T\mathbf{r}(\gamma) \\ &= (\mathbf{a}\gamma^2 + \mathbf{b}\gamma + \mathbf{c})^T(\mathbf{a}\gamma^2 + \mathbf{b}\gamma + \mathbf{c}) \\ &= \mathbf{a}^T\mathbf{a}\gamma^4 + 2\mathbf{a}^T\mathbf{b}\gamma^3 + (2\mathbf{a}^T\mathbf{c} + \mathbf{b}^T\mathbf{b})\gamma^2 + 2\mathbf{b}^T\mathbf{c}\gamma + \mathbf{c}^T\mathbf{c} \end{aligned}$$

where $\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}$ for $\mathbf{x}, \mathbf{y} = \mathbf{a}, \mathbf{b}, \mathbf{c}$. The step size can be determined by setting the first derivative of f to zero.

$$\begin{aligned} \frac{\mathrm{d}f(\gamma)}{\mathrm{d}\gamma} &= f'(\gamma) \\ &= 4\mathbf{a}^T\mathbf{a}\gamma^3 + 6\mathbf{a}^T\mathbf{b}\gamma^2 + 2(2\mathbf{a}^T\mathbf{c} + \mathbf{b}^T\mathbf{b})\gamma + 2\mathbf{b}^T\mathbf{c} \\ &= \mathbf{0} \end{aligned}$$

To determine the possible values of γ which solve that function, the Matlab function *roots()* [2] is used. The largest positive and real value of the roots-vector inside the range $0 \leq \gamma \leq 1$ is chosen as the maximum γ value. The maximum step size is one Newton step even if the minimum γ value is bigger than one. Figure 2.5 on the next page shows the schematic diagram for the line search procedure.

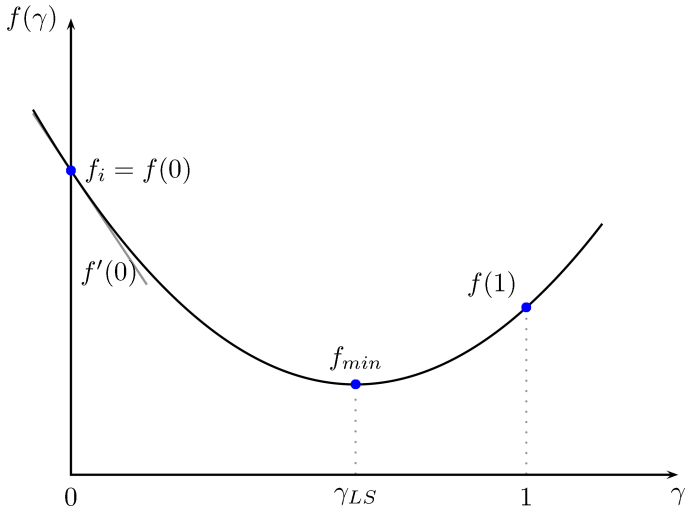


Figure 2.5: Schematic diagram for the line search

2.3 Regularization Methods

2.3.1 Truncated Singular Value Decomposition (TSVD)

One of the standard tools for solving linear problems is the Truncated Singular Value Decomposition (TSVD). To understand this method, it is necessary to have a look at the Singular Value Decomposition (SVD) of a matrix briefly.

Singular Value Decomposition (SVD)

To explain the Singular Value Decomposition of a matrix, consider a rectangular matrix K $\in R^{m\times n}$ with the rank(K) = k [1, 19, 14]. The SVD of the matrix K has the form

$$\mathrm{K} = \mathrm{U}\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}\mathrm{V}^T = \mathrm{USV}^T,$$

where

- U is a $m \times m$ orthogonal matrix $\mathrm{U} = (\mathbf{u}_1, ..., \mathbf{u}_m) \in \mathbb{R}^{m\times m}(\mathrm{U}^T = \mathrm{U}^{-1})$, where $\mathbf{u}_i$ are the left singular vectors,
- V is a $n \times n$ orthogonal matrix $\mathrm{V} = (\mathbf{v}_1, ..., \mathbf{v}_n) \in \mathbb{R}^{n\times n}$ $(\mathrm{V}^T = \mathrm{V}^{-1})$, where $\mathbf{v}_i$ are the right singular vectors and
- Σ is a $k \times k$ diagonal matrix with the diagonal entries $\sigma = (\sigma_1, ..., \sigma_k)$ with $\sigma_k > 0$.

The diagonal entries of the matrix S are $\sigma = (\sigma_1, ...\sigma_p)$ with $p = \min(m, n)$. This values are called the **singular values** of the matrix K. All singular values are non negative and can be arranged in decreasing order so that

$$\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_p \geq 0$$

An important value in this context is the condition number of the matrix K

$$\text{cond}(\text{K}) = \frac{\sigma_1}{\sigma_k}.$$

It is the ratio between the biggest and the smallest nonzero singular value of the matrix K. The matrix K is named ill-conditioned if the condition number is large, where the specification of "large" depends on the particular problem [9, p. 31]. The inverse solution obtained by the SVD equals the unregularized solution in figure 2.3 on page 16, see figure 2.6

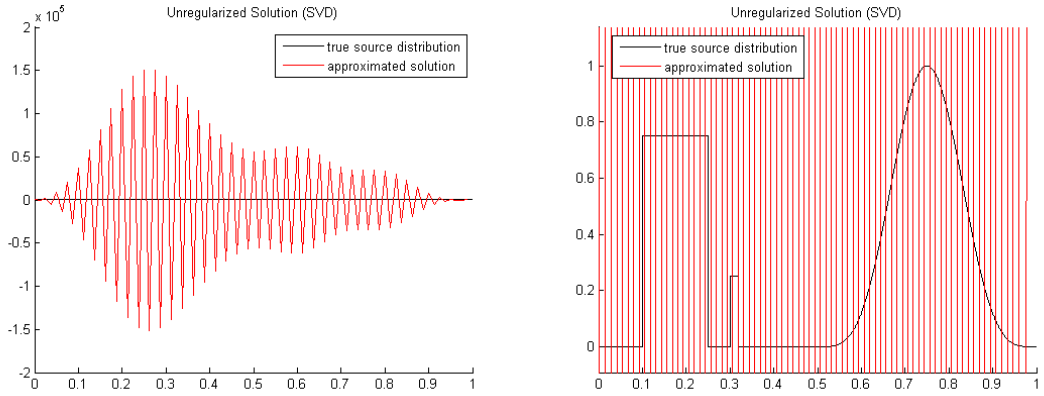


Figure 2.6: The true distribution $\mathbf{f}$ (is not visible on the left due to the dimension of the y-axis) and the inverse solution obtained by the SVD.
On the right a detailed view with the true distribution is shown.

TSVD

For further consideration the SVD is used to solve linear least squares problems of the form

$$\min_{\mathbf{f}} \|\text{K}\mathbf{f} - \mathbf{d}\|_2^2$$

The SVD of the matrix $\text{K} \in \mathbb{R}^{m \times n}$ can be seen as

$$\text{K} = \text{USV}^T = \sum_{i=1}^{k} \mathbf{u}_i \sigma_i \mathbf{v}_i^T$$

with $\text{rank}(\text{K}) = k$.
If $\mathbf{d} \in \mathbb{R}^m$ then the solution $\mathbf{f}^*$ is unique and can be defined by

$$\mathbf{f}^* = \sum_{i=1}^{k} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\sigma_i} \right) \mathbf{v}_i$$

Now the matrix K^+ is defined, which is called the pseudo-inverse or Moore-Penrose generalized inverse of K.

$$\text{K}^+ = \text{V}\Sigma^+\text{U}^T$$

where $\Sigma^+ = \text{diag}(\sigma_1^{-1}, ..., \sigma_k^{-1}, 0, ...0) \in \mathbb{R}^{n \times n}$ then the solution is

$$\mathbf{f}^* = \mathrm{K}^+ \mathbf{d}$$

The pseudo-inverse can additionally be formulated by

$$\mathrm{K}^+ = (\mathrm{K}^T \mathrm{K})^{-1} \mathrm{K}^T.$$

If $\text{rank}(\mathrm{K}) = m = n$, so the matrix is square and has full rank, then $\mathrm{K}^+ = \mathrm{K}^{-1}$. In this case the singular values are all nonzero. Additionally the squared singular values of K are the eigenvalues of $\mathrm{K}^T\mathrm{K}$.
According to Hansen [1] two characteristics are often found in connection with discrete ill-posed problems.

- The singular values σ_i decay gradually to zero with no particular gap in the spectrum.
- The left and right singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ tend to have more and more changes in signs in their elements for increasing index i.

The singular values of the present matrix K are plotted in figure 2.7. They decay without a

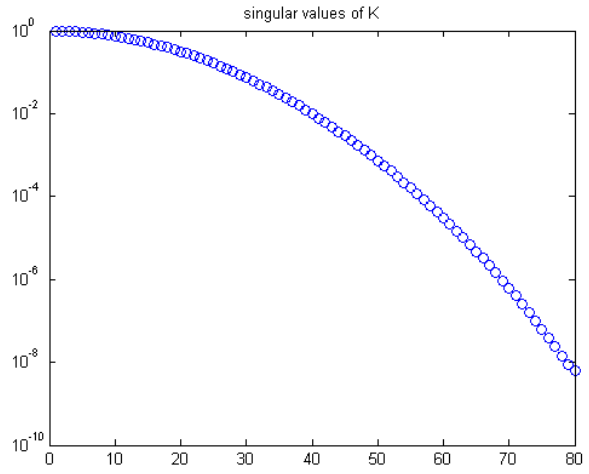


Figure 2.7: The singular values of the matrix K

particular gap from $s_1 = 0.9971$ to $s_{80} = 6.2032 \cdot 10^{-9}$. The condition number of the matrix K is quite large: $\text{cond}(\mathrm{K}) = 1.6074 \cdot 10^8$. It is shown in figure 2.8 on the next page that for decreasing singular value σ_i the elements of the singular vectors show more oscillations. For the solution $\mathbf{f}^*$ it can be expected, that for bigger index i it will get more and more oscillations, since the values of $\mathbf{d}$ are amplified dramatically by the very small but nonzero σ_i. (The high-frequency contents in d are amplified.)
If there is a notable discontinuity in the spectrum of singular values the problem is well-posed and no regularization is needed. As a logical consequence of this, only the first r singular values are used and the rest is neglected (set to zero), which is called Truncated Singular Value decomposition

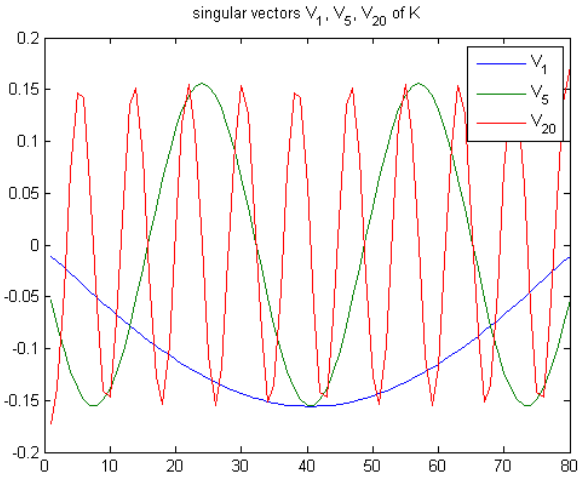


Figure 2.8: The singular vectors $\mathbf{v}_1$, $\mathbf{v}_5$, $\mathbf{v}_{20}$, of the matrix K

(TSVD). So the solution $\mathbf{f}_{TSVD}$ is given by

$$\mathbf{f}_{TSVD} = \sum_{i=1}^{r} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\sigma_i} \right) \mathbf{v}_i.$$

The choice of r depends on the particular problem. In practice not r but the threshold α with the condition $\sigma_i^2 > \alpha$ is defined. All singular values which violate the condition are set to zero. This leads to the **filter function**

$$w_\alpha(\sigma^2) = \begin{cases} 1 & \text{if } \sigma^2 > \alpha \\ 0 & \text{if } \sigma^2 \leq \alpha \end{cases} \tag{2.10}$$

The approximated solution $\mathbf{f}_\alpha$ can then be written in the form

$$\mathbf{f}_\alpha = \sum_{\sigma_i^2 > \alpha} \sigma_i^{-1} (\mathbf{u}_i^T \mathbf{d}) \mathbf{v}_i = \sum_i w_\alpha(\sigma_i^2) \sigma_i^{-1} (\mathbf{u}_i^T \mathbf{d}) \mathbf{v}_i. \tag{2.11}$$

The solution for a given threshold $\alpha = 0.001$ is plotted in figure 2.9 on the following page. The SVD is often used to calculate the inverse of the Hessian matrix of the given functional. To handle functionals with an additional penalty term like in the Tikhonov functional the matrix pair (K, L) (see equation (2.12) on the next page) it is necessary to extend or generalize the SVD. This modified SVD is known as generalized SVD or GSVD, see [1].

2.3.2 Tikhonov Regularization (Standard Method)

The Tikhonov Regularization method is a frequently used regularization method, so it is also applied to the network of resistors. First of all, the linear problem is discussed and the similarities in between the TSVD and the Tikhonov Regularization method are shown. Sometimes the Tikhonov Regularization is also denoted as Tikhonov-Phillips Regularization method [19, p. 87]. Tikhonov originally used the penalty term $\|\mathbf{f}\|^2$ which is equivalent to a regularization matrix $\mathrm{L} = \mathrm{I}$ (with I the identity matrix), hence this constellation is also known as standard Tikhonov Regularization

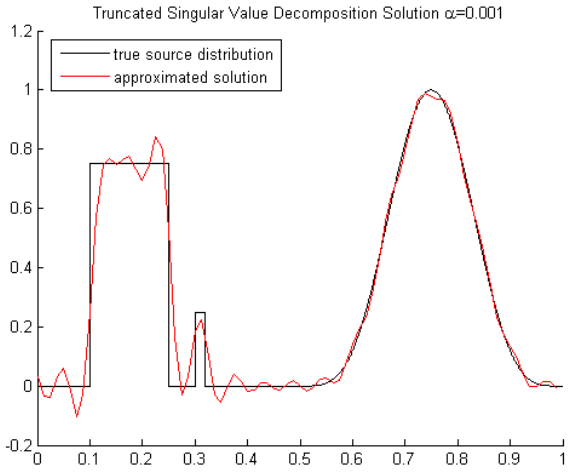


Figure 2.9: Solution obtained with TSVD with $\alpha = 0.001$

or zeroth order Tikhonov Regularization [5]. Other choices for the regularization matrix L are shown in chapter 4. Nevertheless the linear case is considered here and the functional is given by

$$\min_{\mathbf{f}}(||\mathrm{K}\mathbf{f} - \mathbf{d}||_2^2 + \alpha||\mathrm{L}\mathbf{f}||_2^2), \tag{2.12}$$

where $\mathrm{K}\mathbf{f}$ is the linear model of the system, $\mathbf{d}$ is the vector of the measurement values and α is the regularization parameter (see next chapter). Additionally the above defined linear functional in equation (2.12) can be interpreted as a constrained least squares problem in two different ways [14].

On the one hand an ill-posed problem is considered and some prior information of the form $||\mathrm{L}\mathbf{f}||_2$ is incorporated. One specifies that the L_2 norm of the sought values is smaller than a prescribed value ϵ.

$$\min_{\mathbf{f}} ||\mathrm{K}\mathbf{f} - \mathbf{d}||_2 \text{ subject to } ||\mathrm{L}\mathbf{f}||_2 \leq \epsilon$$

This is known as the method of quasi solution relative to L.

On the other hand some a-priori knowledge about the behavior (for example: a smoothness behavior) of the solution is known and the least squares problem is given by

$$\min_{\mathbf{f}} ||\mathrm{L}\mathbf{f}||_2 \text{ subject to } ||\mathrm{K}\mathbf{f} - \mathbf{d}||_2 \leq \delta$$

Now the standard Tikhonov functional with $\mathrm{L} = \mathrm{I}$ is analyzed and the similarity compared to the TSVD is shown. Equation (2.12) leads to the least squares problem

$$\min_{\mathbf{f}}(||\mathrm{K}\mathbf{f} - \mathbf{d}||_2^2 + \alpha||\mathbf{f}||_2^2)$$

The solution can be written as

$$\mathbf{f}_\alpha = \mathrm{K}_\alpha^I \mathbf{d}$$

where the matrix $\mathrm{K}_\alpha^I \in \mathbb{R}^{n\times n}$ is a '*regularized inverse*' which is defined by

$$\mathrm{K}_\alpha^I = \mathrm{V}\Sigma_\alpha^I \mathrm{U}^T$$

with

$$\Sigma_\alpha^I = \mathrm{diag}\left[\frac{\sigma_1}{\sigma_1^2+\alpha}, \ldots, \frac{\sigma_n}{\sigma_n^2+\alpha}\right] \in \mathbb{R}^{n\times m}$$

This solution also can be written as

$$\mathbf{f}_\alpha = \sum_{i=1}^{n} \frac{\sigma_i(\mathbf{u}_i^T\mathbf{d})}{\sigma_i^2+\alpha}\mathbf{v}_i = (\mathrm{K}^T\mathrm{K}+\alpha\mathrm{I})^{-1}\mathrm{K}^T\mathbf{d}.$$

This leads to the Tikhonov filter function similar to the TSVD filter function in equation (2.10) on page 27 and the according solution in equation (2.11) on page 27

$$w_\alpha(\sigma^2) = \frac{\sigma^2}{\sigma^2+\alpha} \tag{2.13}$$

In figure 2.10 the similarity of the linear Tikhonov regularization to the TSVD is shown. The TSVD

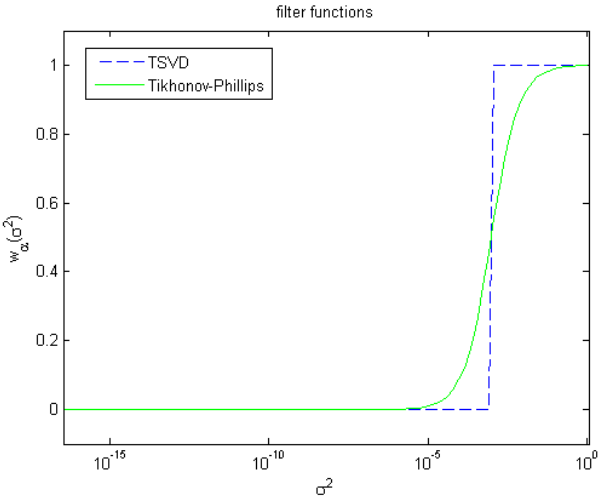


Figure 2.10: Similarity of the TSVD and Tikhonov filter functions $w_\alpha(\sigma^2)$

cuts off the smaller singular values, the Tikhonov method has a smooth but similar transition if the regularization parameter α is chosen appropriately. Therefore, the result of the chosen linear problem (which is shown in figure 2.11 on the next page) is similar to the TSVD, even though the oscillations are slightly smaller.

The influence of the Tikhonov filter function could be described that for relatively large values of σ_i^2 the solution is only slightly modified, but for small values, which have strong influence in amplifying the data errors, the solution is damped [19, p. 91]. In opposite to the TSVD the terms are not omitted, but replaced with "incorrect" terms which do not corrupt the solution.
Now the problem is treated as a least squares problem and the standard single-step method is shown. This is possible for linear problems only, if the model is of high accuracy. On the basis of

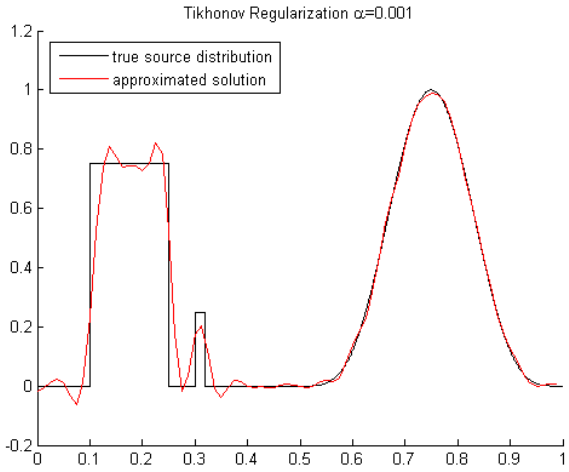


Figure 2.11: Single step Tikhonov Regularization according to the TSVD method with $\alpha = 0.001$

this linear problem it is shown in figure 2.2 on page 15, that the simple solution $\mathbf{f} = \mathrm{K}^{-1}\mathbf{d}$ is not satisfactory. Consider the linear optimization functional above, then the minimization problem is

$$\psi(\mathbf{f}) = ||\mathrm{K}\,\mathbf{f} - \mathbf{d}||_2^2 + ||\mathrm{L}\mathbf{f}||_2^2$$

or

$$\psi(\mathbf{f}) = (\mathrm{K}\,\mathbf{f} - \mathbf{d})^T(\mathrm{K}\,\mathbf{f} - \mathbf{d}) + \alpha(\mathrm{L}\mathbf{f})^T(\mathrm{L}\mathbf{f})$$

To determine the optimal parameter, set the first derivation to zero.

$$2\mathrm{K}^T(\mathrm{K}\,\mathbf{f} - \mathbf{d}) + 2\alpha\mathrm{L}^T\mathrm{L}\mathbf{f} = 0$$

and it follows

$$(\mathrm{K}^T\mathrm{K} + \mathrm{L}^T\mathrm{L})\mathbf{f} = \mathrm{K}^T\mathbf{d}$$

and the optimal solution is

$$\mathbf{f} = (\mathrm{K}^T\mathrm{K} + \mathrm{L}^T\mathrm{L})^{-1}(\mathrm{K}^T\mathbf{d}).$$

Additionally it should be marked, that the penalty term with the Identity matrix $\mathrm{L} = \mathrm{I}$ forces the minimization of the parameters of $\mathbf{f}$, which could lead to bad results if this assumption is not correct.

2.3.3 Landweber Iteration Method

The Landweber iteration method is similar to the steepest descent method, but the γ value is set to a fixed constant with $0 < \gamma < 1/||\mathrm{K}||^2$ [9, p. 10]. It has the updating function for every iteration k

$$\mathbf{f}_{k+1} = \mathbf{f}_k - \gamma\mathrm{K}^T(\mathrm{K}\mathbf{f}_k - \mathbf{d}). \tag{2.14}$$

In [19, p. 107] and [9, p. 10] it is shown, that the Landweber iteration can also be interpreted as the filter function

$$w_k(\sigma^2) = 1 - (1 - \gamma\sigma^2)^k. \tag{2.15}$$

[9, p. 10] The filter function depends on the number of iterations k, hence for few iterations the solution is a very smooth approximation of the real function, on the other hand if the number of iterations becomes too large, the solution becomes highly oscillatory. The iteration counter plays the role of the regularization parameter, since $\alpha = 1/k$ [19, p. 106-7]. The method tends to need a high number of iterations to arrive at a useful solution, therefore the practical usage is limited. In [10, p. 154] the linear Landweber iteration is described as a steepest descent method without line search, equivalent with $\gamma = 1$. The plots in figure 2.12 show the results after 5, 50, 500 and 5000 iterations. In case of a nonlinear $\mathbf{k}(\mathbf{f})$, the updating formula becomes

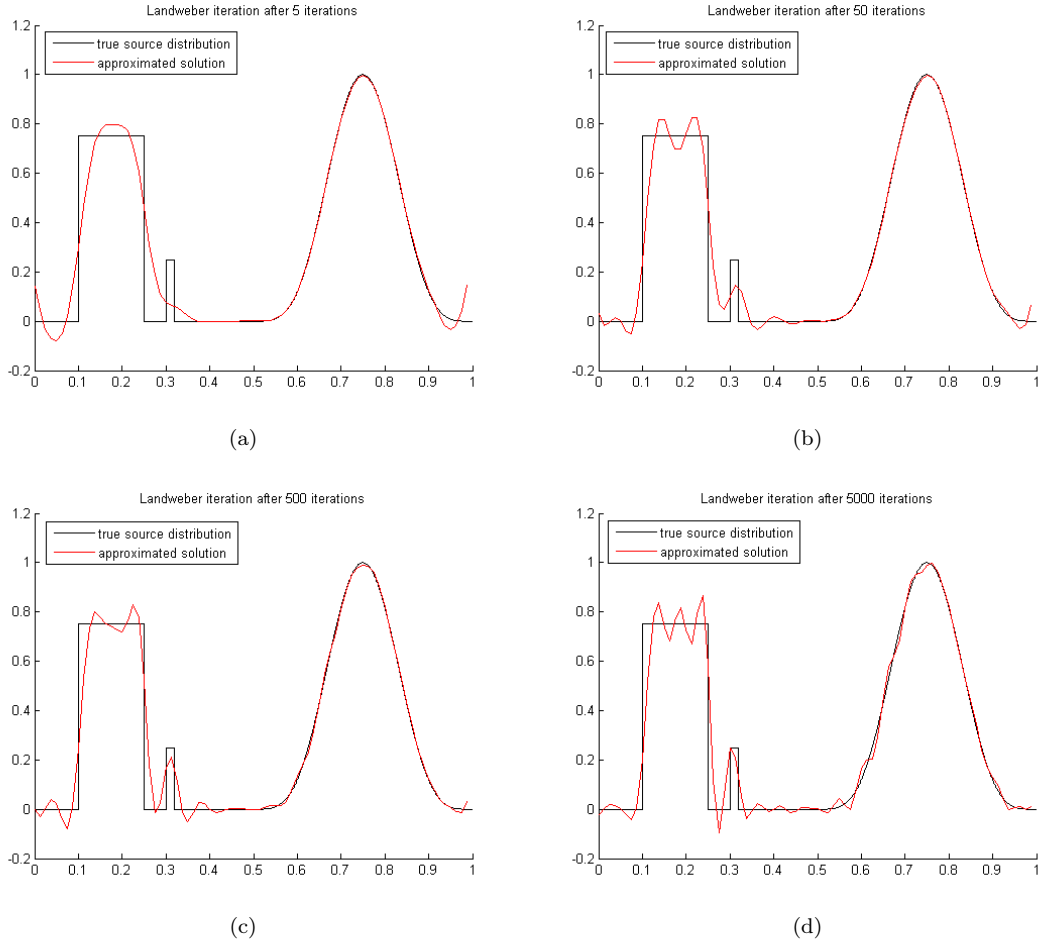


Figure 2.12: Landweber iteration method: solution after a different number of iterations

$$\mathbf{f}_{k+1} = \mathbf{f}_k - \mathbf{k}'(\mathbf{f}_k)(\mathbf{k}(\mathbf{f}_k) - \mathbf{d}) \tag{2.16}$$

where $\mathbf{k}'(\mathbf{f})$ is the Frechet-derivative of $\mathbf{k}(\mathbf{f})$ [10].

2.3.4 Total Variation Method

The Total Variation method is in general is a regularization method with a different selection of the penalty term. In the Tikhonov penalty term the L_2-Norm is used, the penalty term of the Total Variation (TV) Method can be seen as L_1-norm [9]. With this choice 'blocky' images can be reconstructed with good results. 'Blocky' images means nearly constant areas with jump discontinuities or steep changes of the values between them. Unfortunately, the method it is more difficult to implement. First consider the following linear least squares minimization functional similar as above:

$$\min_{\mathbf{f}}(||\mathrm{K}\mathbf{f} - \mathbf{d}||_2^2 + \alpha \mathrm{TV}(\mathbf{f})) \tag{2.17}$$

where $\mathrm{TV}(\mathbf{f})$ is in general

$$\mathrm{TV}(\mathbf{f}) = \int_\Omega |\nabla \mathbf{f}|_\Omega \, d\Omega. \tag{2.18}$$

As example the discrete one dimensional TV functional is [9, p. 9]

$$\begin{aligned}\mathrm{TV}(\mathbf{f}) &= \sum_{k=1}^{n-1} |f_{k+1} - f_k| \\ &= \sum_{k=1}^{n-1} \left| \frac{f_{k+1} - f_k}{\Delta x} \right| \Delta x.\end{aligned}$$

This method reveals some problems, which are discussed later in chapter 5. As motivation: for the given linear problem a solution can be computed with the lagged diffusivity fixed point method for the Total Variation penalized least squares [9], see figure 2.13 and see chapter 5. The main characteristic of the TV-functional is that blocky structures are nicely reconstructed, while continuous functions (like the sin-function) in the example is reconstructed with almost blocky elements. The computed fixed point method is developed in chapter 5 and [9, p. 136].

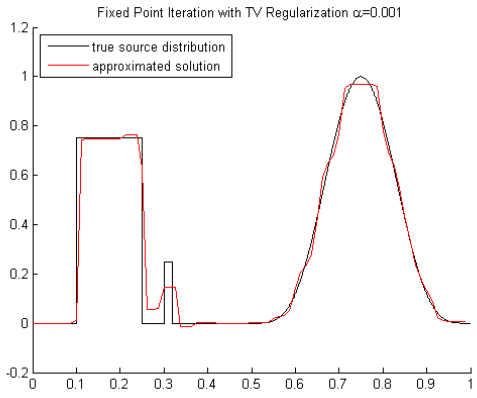


Figure 2.13: Fixed Point method for the Total Variation functional $\alpha = 0.001$ with 50 iterations (because the TV-method needs to be iterative for linear functions)

2.4 Implemented Algorithms with Tikhonov Regularization

2.4.1 Regularized Gauss-Newton Method

The derivation of the iterative Gauss-Newton scheme [13]

$$\psi(\mathbf{r}) = ||\mathbf{u}(\mathbf{r}) - \mathbf{u}_r||_2^2 + \alpha||\mathrm{L}\,\mathbf{r}||_2^2$$

and via the Taylor series approximation (see equation (2.6) on page 20) the gradient $\mathbf{g}(\mathbf{r})$ and the Hessian matrix $\mathrm{H}(\mathbf{r})$ can be determined as

$$\begin{aligned}\mathbf{g}(\mathbf{r}) &= \frac{\partial \psi}{\partial \mathbf{r}} \\ &= \frac{\partial}{\partial \mathbf{r}}\left[(\mathbf{u}(\mathbf{r}) - \mathbf{u}_r)^T(\mathbf{u}(\mathbf{r}) - \mathbf{u}_r) + \alpha\,(\mathrm{L}\,\mathbf{r})^T(\mathrm{L}\,\mathbf{r})\right] \\ &= 2\mathrm{J}(\mathbf{r})^T(\mathbf{u}(\mathbf{r}) - \mathbf{u}_r) + 2\alpha\mathrm{L}^T(\mathrm{L}\,\mathbf{r})\end{aligned}$$

$$\begin{aligned}\mathrm{H}(\mathbf{r}) &= \frac{\partial^2 \psi}{\partial \mathbf{r}^2} \\ &= 2\left(\frac{\partial \mathbf{u}}{\partial \mathbf{r}}\right)^T\left(\frac{\partial \mathbf{u}}{\partial \mathbf{r}}\right) + \underbrace{2\sum_k \frac{\partial^2 \mathbf{u}_k}{\partial \mathbf{r}^2}(\mathbf{u}_k(\mathbf{r}) - \mathbf{u}_{r\,k})}_{negligible} + 2\alpha\mathrm{L}^T\mathrm{L} \\ &\approx 2\mathrm{J}(\mathbf{r})^T\mathrm{J}(\mathbf{r}) + 2\alpha\mathrm{L}^T\mathrm{L}\end{aligned}$$

with the Jacobian matrix $\mathrm{J} = \frac{\partial \mathbf{u}}{\partial \mathbf{r}}(\mathbf{r})$. The second order term is negligible in the neighborhood of the solution. This is denoted as Gauss-Newton Hessian matrix approximation.
Hence the gradient $\mathbf{g}_k = \mathbf{g}(\mathbf{r}_k)$ at iteration k is

$$\mathbf{g}_k = \frac{\partial \psi}{\partial \mathbf{r}}(\mathbf{r}_k) = 2\mathrm{J}_k^T(\mathbf{u}(\mathbf{r}_k) - \mathbf{u}_r) + 2\alpha\mathrm{L}^T(\mathrm{L}\,\mathbf{r}_k) \tag{2.19}$$

and the Hessian matrix $\mathrm{H}_k = \mathrm{H}(\mathbf{r}_k)$ at iteration k is

$$\mathrm{H}_k = \frac{\partial^2 \psi}{\partial \mathbf{r}^2}(\mathbf{r}_k) = 2\mathrm{J}_k^T\mathrm{J}_k + 2\alpha\mathrm{L}^T\mathrm{L} \tag{2.20}$$

The solution of every iteration k is the difference $\boldsymbol{\delta}\mathbf{r}$ to calculate the resistor values for the next step.

$$\boldsymbol{\delta}\mathbf{r}_{k+1} = -\mathrm{H}_k^{-1}\mathbf{g}_k$$

with $\boldsymbol{\delta}\mathbf{r}_{k+1} = \mathbf{r}_{k+1} - \mathbf{r}_k$, where $\mathbf{u}_r$ is the vector of measured voltages or the data vector, $\mathbf{u}(\mathbf{r})$ is the vector of calculated voltages, α is the regularization parameter, $\mathbf{r}$ is the vector of current resistor values and L is the regularization matrix. The first term which is named the residual norm or objective functional contains only of the measurement values and simulated values of the voltages. With this term only it is not possible to get a useful and stable solution, because of the ill-posedness of the problem. Therefore the second term, the penalty term is added. This term expands the problem with additional information. An update step of the resistor values $\mathbf{r}$

for iteration k is given by

$$\mathbf{r}_{k+1} = \mathbf{r}_k + \gamma\,\boldsymbol{\delta}\mathbf{r}$$

with

$$\boldsymbol{\delta}\mathbf{r} = -(\mathrm{J}^T\mathrm{J} + \alpha\mathrm{L}^T\mathrm{L})^{-1}(\mathrm{J}^T(\mathbf{u}(\mathbf{r}) - \mathbf{u}_r) + \alpha\mathrm{L}^T\mathrm{L}\,\mathbf{r})$$

and γ is the step size calculated by the line search procedure, which was shown earlier in this chapter. As maximum a full Newton step is made hence the maximal γ value is $\gamma = 1$.

2.4.2 Occam's Inversion

Occam's Inversion [5] is an iterative algorithm which uses the discrepancy principle (see next chapter) to chose the best α for every iteration. It is similar to the Gauss-Newton algorithm, but uses only a Taylor series approximation for the model $\mathbf{u}(\mathbf{r})$ without a line search procedure. The basic idea is a linear approximation of the model in every step. In the network problem the real δ-value from the discrepancy principle can be computed because the real resistances are known. The method approximates $\mathbf{u}(\mathbf{r})$ with a Taylor series approximation:

$$\begin{aligned}\mathbf{u}(\mathbf{r}_{k+1}) &= \mathbf{u}(\mathbf{r}_k) + \frac{\partial\mathbf{u}(\mathbf{r}_k)}{\partial\mathbf{r}}^T(\mathbf{r}_{k+1} - \mathbf{r}_k)\\ &= \mathbf{u}(\mathbf{r}_k) + \mathrm{J}_k^T\boldsymbol{\delta}\mathbf{r}_{k+1}\end{aligned}$$

For the iteration $k+1$ the functional is given by

$$\psi_{k+1} = \|\mathbf{u}(\mathbf{r}_{k+1}) - \mathbf{u}_r\|_2^2 + \alpha\,\|\mathrm{L}\,\mathbf{r}_{k+1}\|_2^2$$

with the approximation of $\mathbf{u}(\mathbf{r}_{k+1})$ and a new data vector $\widehat{\mathbf{u}}_r = \mathbf{u}_r - \mathbf{u}(\mathbf{r}_k) + \mathrm{J}(\mathbf{r}_k)\mathbf{r}_k$ and $\mathbf{r}_{k+1} = \mathbf{r}_k + \boldsymbol{\delta}\mathbf{r}$ it can be written

$$\psi_{k+1} = \|\mathrm{J}(\mathbf{r}_k)\mathbf{r}_{k+1} - \widehat{\mathbf{u}}_r\|_2^2 + \alpha\,\|\mathrm{L}\,\mathbf{r}_{k+1}\|_2^2$$

The gradient can be computed as follows and is set to zero to determine the optimal parameters

$$\frac{\partial\psi}{\partial\mathbf{r}_{k+1}} = 2\mathrm{J}(\mathbf{r}_k)^T(\mathrm{J}(\mathbf{r}_k)\mathbf{r}_{k+1} - \widehat{\mathbf{u}}_r) + \alpha 2\mathrm{L}^T\mathrm{L}\mathbf{r}_{k+1} = \mathbf{0}$$

with $\mathrm{J}_k = \mathrm{J}(\mathbf{r}_k)$ the optimal next resistor values are determined as

$$\mathbf{r}_{k+1} = (\mathrm{J}_k^T\mathrm{J}_k + \alpha\mathrm{L}^T\mathrm{L})^{-1}\mathrm{J}_k^T\widehat{\mathbf{u}}_r \tag{2.21}$$

The algorithm adjusts the regularization parameter α every step according to the best factor determined by the discrepancy principle [5, p. 245]. Equation (2.21) is equal with the Gauss Newton update function without a line search procedure ($\gamma = 1$).

2.4.3 Active Set Method with Box Constraints

Additional information can also be implemented if upper and lower bounds of the sought values are known [11]. This bounds can be implemented as simple box constraints. There are two

constraints for every resistor and because the resistor can only be at one bound at the same time, there can only be one constraint active. This is computed according to the algorithm of [11] with considerations from [12]. The optimization functional $\psi(\mathbf{r})$ is given by

$$\min_{\mathbf{r}} \ \psi(\mathbf{r}) = \min \|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|_2^2 + \alpha \|\mathrm{L}\mathbf{r}\|_2^2 \qquad (2.22)$$
$$\text{s.t. } \mathbf{lb} \leq \mathbf{r} \leq \mathbf{ub}$$

with the Taylor series expansion of the Gauss Newton method the active set problem becomes

$$\min \left(\boldsymbol{\delta}\mathbf{r}^T \mathbf{g}_k + \frac{1}{2} \boldsymbol{\delta}\mathbf{r}^T \mathrm{H}_k \boldsymbol{\delta}\mathbf{r} \right)$$
$$\text{s.t. } \boldsymbol{\delta}\mathbf{lb} \leq \boldsymbol{\delta}\mathbf{r} \leq \boldsymbol{\delta}\mathbf{ub}$$

where $\boldsymbol{\delta}\mathbf{lb} = \mathbf{lb} - \mathbf{r}_k$, $\boldsymbol{\delta}\mathbf{r} = \mathbf{r} - \mathbf{r}_k$ and $\boldsymbol{\delta}\mathbf{ub} = \mathbf{ub} - \mathbf{r}_k$. The gradient $\mathbf{g}_k$ and the Hessian matrix H_k where defined with the Gauss-Newton approximation (see equation (2.20) and equation (2.19) on page 33)

$$\mathbf{g}_k = 2\mathrm{J}_k^T(\mathbf{u}(\mathbf{r}_k) - \mathbf{u}_\mathbf{r}) + 2\alpha \mathrm{L}^T \mathrm{L}\mathbf{r}_k$$
$$\mathrm{H}_k = 2\mathrm{J}_k^T \mathrm{J}_k + 2\alpha \mathrm{L}^T \mathrm{L}$$

The general linear constraints are represented as $\mathrm{A}\ \boldsymbol{\delta}\mathbf{r} \geq \mathbf{b}$. In this particular case the matrix A consists of positive or negative rows of the identity matrix dependent on the bound (**lb** or **ub**), and the vector **b** consists of the relative negative upper and positive lower bounds. The given constraints are inequality constraints. They become active if they reach the bound and then they are treated like equality constraints. Hence the active constraints are equality constraints, the other inactive constraints are ignored.

$$\min \left(\boldsymbol{\delta}\mathbf{r}^T \mathbf{g}_k + \frac{1}{2} \boldsymbol{\delta}\mathbf{r}^T \mathrm{H}_k \boldsymbol{\delta}\mathbf{r} \right) \qquad (2.23)$$
$$\text{s.t. } \mathbf{a}_j^T \boldsymbol{\delta}\mathbf{r} = 0 \quad \text{for } j \in \text{active set}$$

where $\mathbf{a}_j^T$ are the rows of A_t which denotes the active constraints. It is important for the algorithm to start inside the feasible region (no inequality constraint is violated). To calculate the solution the null space of A needs to be defined in such a way that $\mathrm{A}^T\mathrm{Z} = 0$ (Z is the null space of A^T). In this special case with boundary constraints only, the matrix A_t contains the rows of the identity matrix which belong to the active constraints and the matrix Z_t contains the rows of the identity which belong to the inactive constraints. Because of two constraints for every resistor, if one constraint of a resistor is active the opposite constraint of this resistor can be removed (not active and not inactive). In each iteration the bounded variables (active variables) are removed from the equations (elimination method) and the optimal solution is obtained from the system of inactive constraints. This system is denoted as the reduced optimization system with the substitution

$$\boldsymbol{\delta}\mathbf{r} = \mathrm{Z}_t \mathbf{y}. \qquad (2.24)$$

Inserting equation (2.24) into equation (2.23) leads to

$$\min \left(\mathbf{y}^T \mathrm{Z}_t^T \mathbf{g}_k + \frac{1}{2} \mathbf{y}^T \mathrm{Z}_t^T \mathrm{H}_k \mathrm{Z}_t \mathbf{y} \right)$$

with the reduced Hessian matrix H_r defined as

$$\mathrm{H}_r = \mathrm{Z}_t^T \mathrm{H}_k \mathrm{Z}_t$$

and the reduced gradient $\mathbf{g}_r$ is

$$\mathbf{g}_r = \mathrm{Z}_t^T \mathbf{g}_k.$$

Hence the solution of the reduced system can be written as

$$\mathbf{y} = -\mathrm{H}_r^{-1} \mathbf{g}_r = - \left(\mathrm{Z}_t^T \mathrm{H}_k \mathrm{Z}_t \right)^{-1} \mathrm{Z}_t^T \mathbf{g}_k$$

To ensure that no constraint is violated, the step size needs to be limited

$$\gamma_{max} = \min_{\mathbf{a}_j^T \mathbf{s}_k < 0} \frac{\mathbf{b}_j - \mathbf{a}_j^T \, \boldsymbol{\delta} \mathbf{r}_k}{\mathbf{a}_j^T \mathbf{s}_k} \text{ for } j : j \notin \text{ active set}$$

with the index of the inactive constraints j and the iteration count k. If some computed $\boldsymbol{\delta}\mathbf{r}$-values violate the constraints, the value which causes the maximum step size γ_{max} for the next step $\mathbf{r}_{k+1} = \mathbf{r}_k + \gamma_{max} \boldsymbol{\delta} \mathbf{r}$ is set active and the solution is computed again. If no constraint is violated, it is checked if some constraints become inactive. Therefore the first order estimates of the Lagrange multipliers at the solution are needed and calculated by solving the equation system

$$\mathbf{g} = \mathrm{A} \boldsymbol{\lambda}$$

In the case with box constraints, for the solution on the lower bound $\lambda_j = g_i$ and for the solution at the upper bound $\lambda_j = -g_i$. The constraint denoted with the smallest $\lambda_j < 0$ is removed from the active set. Every main iteration (computation of the forward solution) only 3 active set iterations are made, since the system is an approximation only (Gauss-Newton Hessian matrix approximation).
Additionally a line search is made but the maximum step size γ_{max} is given by the equation above or the full Newton step ($\gamma = 1$). Therefore if the boundary is reached multiple times, the convergence rate is lower because of the restricted step size. If the boundary is set too wide, there is no difference in convergence compared to the unconstrained case.

3 Calculation of the Regularization Parameter

3.1 Introduction

Again the following structure of the linear minimization functional (standard Tikhonov functional) is considered.

$$\min_{\mathbf{f}} \left\{ \underbrace{\|\mathrm{K}\mathbf{f} - \mathbf{d}\|_2^2}_{\text{residual norm}} + \alpha \quad \underbrace{\|\mathbf{f}\|_2^2}_{\text{solution norm}} \right\}, \tag{3.1}$$

where the penalty term consists of a regularization matrix that is only the Identity matrix (therefore the name *solution norm*). The regularization parameter α guides the influence of the penalty term in the functional and is controlling the degree of regularization [5, p. 93]. It is an important quantity which controls the properties of the regularized solution [1, p. 10]. As seen in chapter 2 in the example of the TSVD and the Tikhonov Regularization, α plays the role of the truncation threshold of the filter function. There exist basically three standard methods to determine the regularization parameter :

- discrepancy principle
- generalized cross validation method
- L-curve method (Hansen)

One often used criterion is the discrepancy principle, also known as Morozov criterion. The idea behind is that one cannot expect more accuracy in the approximate solution than the one present in the data, hence it is possible to define a boundary value δ for the residual norm.
The generalized cross validation (GCV) method and especially the L-curve method are methods which need a lot of iterations to arrive a proper regularization parameter. So they need a considerable amount of computation time. Especially for the given nonlinear least squares problem resulting from the network of resistors

$$\min_{\mathbf{r}} \left\{ \underbrace{\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|_2^2}_{\text{residual norm}} + \alpha \quad \underbrace{\|\mathrm{L}\mathbf{r}\|_2^2}_{\text{penalty term}} \right\} \tag{3.2}$$

the idea comes up to use in each iteration the best regularization parameter. This means, that one have to compute the best parameter in each iteration, which in practice will not be applicable. The computation of the forward problems is the most time consuming part of the inverse problems. There is an iterative method using the discrepancy principle named Occam's inversion [5] which

is described in the previous chapter. Therefore the forward problem has to be computed several times per iteration (8-12) to get the best parameter of the computed ones, which need not to be the *best* parameter for the problem. If the computation of the forward problem is extremely time consuming even this is not possible. Hence there are other methods which should also be mentioned:

- adaptive method (EMT)
- simple methods of diminishing α

The adaptive method uses the weighted ratio of the residual norm and the penalty term to adjust the regularization parameter in each step. The condition number of the Hessian matrix is used to get an initial parameter and the weighting factor. This method was empirically found and is probably not applicable for every problem. It is described in more detail later.
The idea behind the simple diminishing of the regularization parameter is to get first a rough version of the solution and then refining the solution by putting more emphasis on the residual term for minimization. In the case that a smoothing L-matrix is used and the diminishing regularization factor should reduce the smoothing later in the iteration process. A popular adjustment formula is

$$\alpha_k = \alpha_0 q^{k-1},$$

where α_0 is the properly chosen initial value and q is a constant value $0 < q < 1$. This equals the notation $\alpha_{k+1} = \alpha_k \cdot q$. Another and very simple choice of the regularization parameter is given by [10, p. 287]

$$\alpha_k = 2^{-k}.$$

Of course the number of iterations has to be limited appropriately, in dependence of the present errors (noise and model error).

3.2 Discrepancy Principle

The discrepancy principle [14, 67] is also known as the Morozov criterion. It requires a knowledge of a upper noise limit δ and can be formulated as the functional

$$h(\alpha; \mathbf{d}) = ||\mathrm{K}\mathbf{f}_\alpha - \mathbf{d}||_2$$

is the residual norm for the approximate solution $\mathbf{f}_\alpha$. If the following condition for the data holds

$$||\widehat{\mathbf{d}} - \mathbf{d}||_2 \leq \delta < ||\mathbf{d}||_2$$

where $\widehat{\mathbf{d}}$ is the real data vector without noise, then the approximate regularization parameter $\alpha(\delta)$ is determined from

$$\delta = h(\alpha; \mathbf{d}).$$

This gives the final equation for the network of resistors

$$\left\|\mathbf{u}(\mathbf{r}_\alpha^\delta) - \mathbf{u}_r^\delta\right\|_2 = \delta. \tag{3.3}$$

Therefrom the optimal regularization parameter α_δ according to δ is computed [9, 106]. The computation of the regularization parameter does not need to be very exact, because typically the residual norm changes only little with small changes in α. The main problem of the discrepancy principle is, that exact details of the included noise are necessary, white noise is assumed and the variance is needed [9, p. 126]. The convergence of the discrepancy principle can be shown in the sense that if the noise level δ goes to zero the solution goes to the true solution [10, 5]. Vogel in [9, p. 126] derived that the discrepancy principle for the stochastic setting is order optimal. Engl, Grever [20] show, that the L-curve can be used to calculate the optimal regularization parameter via the Morozov criterion. In the results in chapter 6 the δ bound, derived from the computed forward problem with the correct solution and the data vector which is perturbed with noise, is plotted in the L-curve.

3.3 Generalized Cross Validation Method (GCV)

The Generalized Cross Validation Method [14, p. 68] is based on statistical assumptions about the data error (according to Wahba) and defined as the following minimization problem where the optimal α can be determined:

$$\min_{\alpha>0} |||\mathrm{K}\mathbf{f}_\alpha - \mathbf{d}||_2^2 - \mathbf{m}\theta\sigma^2|$$

with $\mathbf{d}$ being the discrete data perturbed with additive noise and this noise is assumed to have zero mean and uncorrelated with the variance σ^2 and a fudge factor $0.7 \leq \theta \leq 0.95$. This function is based on the stochastic smoothing principle of Wahba and Wold. Hansen [1] defines the cross validation function as

$$q(\alpha) = \frac{||\mathrm{K}\mathbf{f}_\alpha - \mathbf{d}||_2^2}{(\mathrm{trace}(\mathrm{I} - \mathrm{K}\mathrm{K}_\alpha^I))^2}, \tag{3.4}$$

where K is the matrix from $\mathrm{K}\mathbf{f} = \mathbf{d}$ and K_α^I is the matrix which produces the regularized solution in $\mathbf{f}_\alpha = \mathrm{K}_\alpha^I\mathbf{d}$, so its the "regularized inverse". The chosen regularization parameter minimizes the function $q(\alpha)$. The derivation of this method is shown in [5, p. 116] and also the ordinary or "leave-one-out" cross validation is defined. The principle is, that the model can be obtained by omitting one data point out of the fitting process. The model will predict the missing data value. The regularization parameter is selected by minimization of the predictive errors of all data points. The number of problems equals the number of data points therefore the method of the Generalized Cross Validation is used to speed up this computation [5, p. 115]. In contrast to the discrepancy principle there is no knowledge necessary about the variance of the noise for the computation and as in the discrepancy principle white noise is assumed [9, p. 126].

3.4 L-Curve Method

The L-curve [1] is a popular graphical tool to determine the regularization parameter, where no information about the errors of the problem is necessary. In principal the L-curve displays the compromise between minimizing the two quantities of the regularization problem, the norm of the

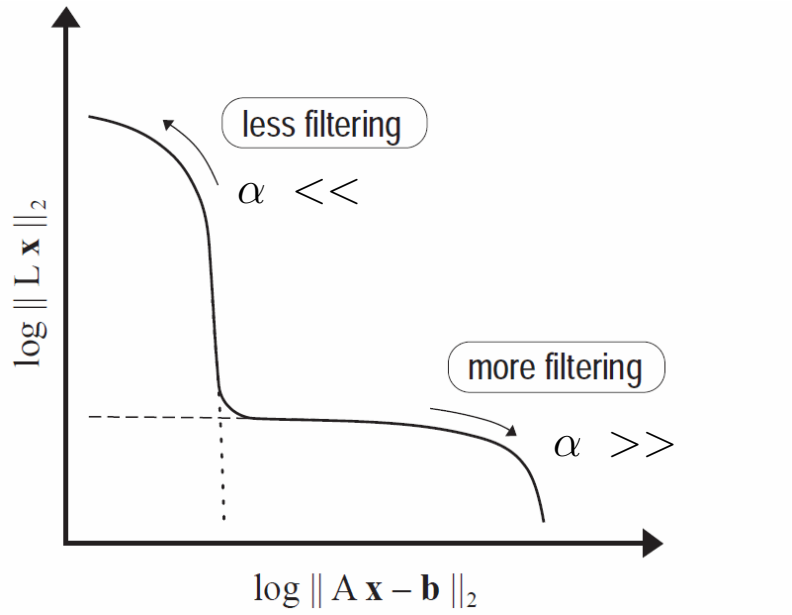


Figure 3.1: L-curve from Hansen [1], where $\log||\mathbf{Ax}-\mathbf{b}||_2$ equals $\log||\mathrm{K}\mathbf{f}-\mathbf{d}||_2$ and $\log||\mathrm{L}\mathbf{x}||_2$ equals $\log||\mathrm{L}\,\mathbf{f}||_2$

residuals and the norm of the regularization or penalty term, see figure 3.1.

$$\psi = \underbrace{||\mathbf{u}(\mathbf{r})-\mathbf{u}_r||_2^2}_{\text{residual norm}} + \alpha \quad \underbrace{||\mathrm{L}\,\mathbf{r}||_2^2}_{\text{penalty term}} \tag{3.5}$$

By detecting the corner of the L-curve, the optimal regularization parameter can be found [1], because the solution is computed with a good balance of the regularization parameter. For some problems no corner can be detected, so the L-curve method fails. The corner can be defined in different ways [10, 110]. In the definition of Hansen [1] the corner of the L-curve is the point of maximum curvature with respect to α. Another definition [10, 110] says that the L-curve is concave in the neighborhood of the corner and the tangent of the corner has to have the slope k with $k = -1$. Unfortunately both definitions are not suitable for the network of resistors. Most of the computed L-curves have no standard L-form and more than one corner. The regions around the corners can be defined as convex with a negative slope. So in each region the point with maximal curvature is defined as the corner. The corner in which the error between simulated and measured data is the smallest is chosen to determine α. The L-curve of the linear problem, on the other hand, has a nice L-shape (see figure 3.2 on the next page).

It has been shown [21] that the L-curve criterion works nicely, if the discrete Picard condition is fulfilled. The discrete Picard condition insures stability in the solution and arises from the computation of the SVD solution of the inverse problem. Considering the linear problem $||\mathrm{K}\mathbf{f}-\mathbf{d}||_2$ and assuming that K is a $n \times n$ Matrix with full rank, then the SVD solution can be written

$$\mathbf{f}^* = \mathrm{V}\mathrm{S}^{-1}\mathrm{U}^T\mathbf{d} = \sum_{j=1}^{n} \frac{\mathbf{u}_j^T\mathbf{d}}{\sigma_j}\mathbf{v}_j$$

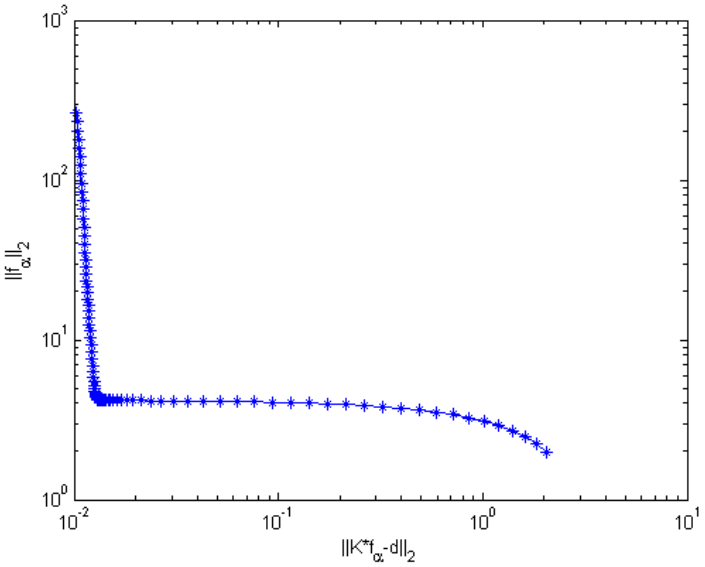


Figure 3.2: L-curve of the linear problem

where $\mathbf{u}_j$ and $\mathbf{v}_j$ are the singular vectors corresponding to σ_j.

The discrete Picard condition is satisfied if the values of $|\mathbf{u}_j^T\mathbf{d}|$ decay to zero more quickly than the singular values σ_j [5, p. 67]. The Picard plot of linear problem is shown in figure 3.3 on the following page and can be used to analyze the discrete Picard condition. Between value number 35 to 40 can be seen that the singular values continue to decay while the values of $|\mathrm{U}^T\mathbf{d}|$ remain almost constant. In other words the values of $|\mathrm{U}^T\mathbf{d}|$ reach the noise floor [5, p. 99]. Thus the ratio $|\mathbf{u}_j^T\mathbf{d}|/\sigma_j$ increases, which implies that the singular values after this point should be damped (Tikhonov) or omitted (TSVD), because it cannot be expected to get useful information from this singular values [5, p. 99].

According to Hansen [1] the functionality of the L-curve is described in the case that the unperturbed data vector $\overline{\mathbf{d}}$ satisfies the discrete Picard condition. Therefore the Fourier coefficients $|\mathrm{U}^T\overline{\mathbf{d}}|$ in average decay faster to zero than the singular values σ. The Picard plot without noise is shown in figure 3.4 on the next page.

The horizontal part of the L-curve corresponds to solutions where the regularization error dominates and the solution becomes very smooth. On the other hand, the vertical part contains solutions which are dominated by perturbation error due to the division by the small singular values. Because its impossible to reach a point below the L-curve in the Tikhonov case (because of the continuous L-curve in opposite to the TSVD), the optimality of the L-curve method is defined that for a given residual norm there does not exist a solution with a smaller semi norm than the Tikhonov solution and therefore the Tikhonov regularization is optimal [1].

Additionally the computed MATLAB [2] function *comp_lcurve.m* is explained here. It consists of 5 parts:

- Part 1: Drawing the L-curve with about 50 or less points (α values: if there are too little points no corner can be detected, but the computation time is raising with the number of

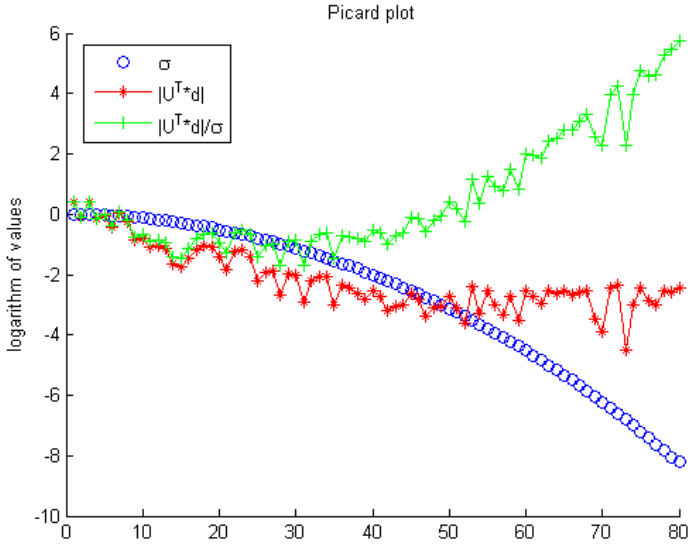


Figure 3.3: Picard plot for the linear problem

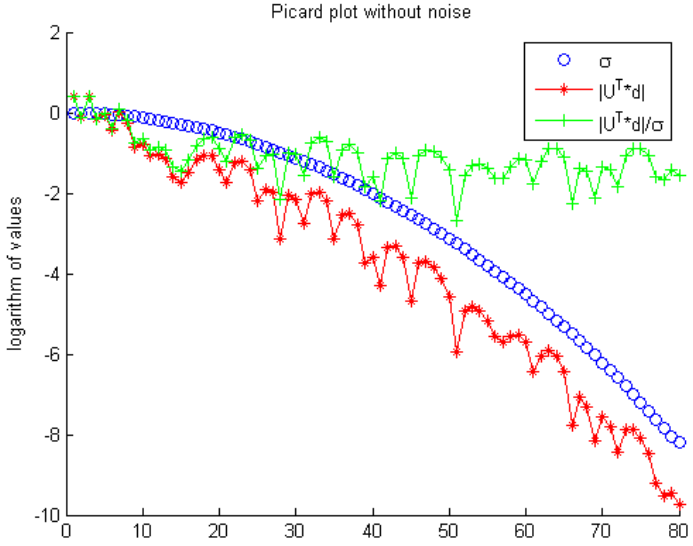


Figure 3.4: Picard plot for the linear problem without noise in the data vector

calculated points. The number of points and the range of the values depends on the particular problem.

- Part 2: Determining the corner regions. A possible corner consists of points which are adjacent together, concave and / or with negative slope.

- Part 3: Refinement of corner regions: To increase the accuracy of the L-curve and therefore the accuracy of the regularization parameter, the curve can be refined with additional points placed in the possible corner regions(s). The factor of refinement can be chosen between 1 and 10, where 1 means no refinement. This parameter is multiplied with the number of points containing every corner region and results in the new number of points in the corner region. This part isn't used in the results in chapter 6, because the accuracy for the given problem is sufficient without it.

- Part 4: Calculation of the corner of every region. The best point of each corner region is defined as the point with the maximum curvature [1]. The spline toolbox is used to obtain a function for every corner.

- Part 5: Determining the best corner, if there are more than one. The best corner is chosen as the corner with the smallest residual norm.

The major characteristic of the L-curve method is that no error information of the problem is needed even no information of the error in the measurements. It can fail if there is no corner detectable or the values of α doesn't match the values of other methods. In [21] is shown that the functionality of the L-curve method fails if the discrete Picard condition is not satisfied. The non convergence of the L-curve for the TSVD and the standard Tikhonov regularization method is also shown in [9, p. 124-126]. And in [10] it is shown that no scheme which contains noisy data with no information about the noise level is able to converge, even if the noise level goes to zero. Thus if useful information of the existing errors in the problem are known, an other method might be more useful [9].

3.5 Adaptive Method

If the L-matrix is the discrete Laplace operator (see chapter 4), it includes a smoothing assumption for the solution, which is often not a correct assumption as in the case of the resistor network. Therefore the regularization parameter is adjusted adaptively after each iteration to get a better solution [22]. This method chooses the initial regularization parameter in dependence of the condition number of the Hessian matrix $\mathrm{H}_k = (\mathrm{J}_k^T \mathrm{J}_k + \alpha \mathrm{L}^T \mathrm{L})$. The condition number of a Matrix is defined as the ratio between the biggest and the smallest singular value of the matrix (see chapter 2 SVD). Then the regularization parameter is chosen in dependence of the initial parameter and

the resulting norms.
Hence, the determination of the regularization parameter can be described in two steps:

- First an estimate of the regularization parameter via the calculation of the condition number of the Hessian matrix. The following minimization problem is solved for the initial value of the regularization parameter.

$$\alpha_0 = \arg\min_{\alpha} \left\{ \left\| \text{cond}\left(\mathrm{J}^T\mathrm{J} + \alpha \mathrm{L}^T\mathrm{L}\right) - c \right\|_2^2 \right\}$$

  where cond(A) denotes the condition number of A and c denotes a empirically found constant value.

- The second step contains an iterative adaptation of α. The idea behind is to use ratio between the residual norm and the regularization term. The residual norm gets smaller after each iteration and also the regularization term should get smaller (see the L-curve method). The following adaption strategy is used.

$$\alpha_k = c(\alpha_0) \frac{\|\mathbf{u}(\mathbf{r}_k) - \mathbf{u}_r\|_2^2}{\|\mathrm{L}\,\mathbf{r}_k\|_2^2}$$

  where $c(\alpha_0)$ is a weighting factor in dependence of α_0. For the constructed problem it works very well, but it is possible that this applies not for every problem. But if this succeeds for the particular problem the regularization parameter is chosen completely automatically.
  In our case, it is difficult to find a appropriate weighting factor for the network problem because the combinations of different resistances is not bounded due to prior information.

4 Tikhonov Regularization: Different Methods to Determine the Regularization Matrix

4.1 Introduction

There exist many different approaches (see [23]) to include prior information in the regularization matrix. The standard Tikhonov functional contains the identity matrix as L-matrix, which is also called zeroth order Tikhonov regularization. For the linear case the first and second order Tikhonov regularization also can be easily explained. The linear Tikhonov regularization functional has the form

$$\|\mathrm{K}\mathbf{f} - \mathbf{d}\|_2^2 + \alpha \|\mathrm{L}\mathbf{f}\|_2^2$$

and the L-matrix of the first order Tikhonov regularization is denoted as

$$\mathrm{L}_1 = \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{bmatrix}$$

This $\mathrm{L}_1\mathbf{f}$ term is the finite-difference approximation proportional to the first order derivative, which will favor solutions that are relatively "flat" [5, p. 103] or slowly changing [24]. The L-matrix of the second order Tikhonov regularization in given by

$$\mathrm{L}_2 = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -2 & 1 \end{bmatrix}$$

This results in the approximation of the second order derivative, which will favor "smooth" solutions. For higher dimensional problems the second order Tikhonov regularization is often implemented using the finite difference approximation to the Laplacian operator [5, p. 104], which is constructed in the next section. The choice of the identity matrix for the resistor network forces the minimization of every resistor value and in our case it doesn't really assist the search for the real solution. Hence for the general case, where no special a priori information is available, the

Laplacian operator is preferred.
Another point of view of the function of the Tikhonov regularization method is that it pulls the solution towards the null space $N(\mathrm{L})$ of the regularization matrix L [23], because the resulting distributions (slowly changing, smooth) form the basis of the null spaces of the respective matrices (L_1 and L_2). Hence if prior information about the true distribution is available, then the regularization matrix can be constructed to guide the solution towards the known distribution. From this another problems arises if the prior information is incorrect. The features of these this special methods, which are

- the Basis Constraint Method
- the Subspace Regularization Method
- the Diagonal Weighting of $\mathrm{J}^T\mathrm{J}$
- the Minimum Mean Variation Method

are described in the following sections. There are other statistical methods to determine the regularization matrix for example see [7].
An additional remark:
When replacing $||\mathrm{L}\mathbf{r}||$ with $||\mathrm{L}(\mathbf{r}-\mathbf{r}^*)||$,then $\mathbf{r}^*$ can represent prior information, which is not represented in the null space of L [23]. The values of $\mathbf{r}^*$ are often the initial parameters of the regularization [13, 23].

4.2 Discrete Laplacian Operator

In this case the regularization matrix L is a discrete Laplacian operator (approximated with finite differences). If the element $\mathrm{L}(i,j)$ $(= R_j)$ is a neighbor of the main element $\mathrm{L}(i,i)$ $(= R_i)$, then $\mathrm{L}(i,j) = -1$ or zero otherwise. For the used definition of the neighbor elements see figure 4.1. The main elements are given by $\mathrm{L}(i,i) = -\sum_j \mathrm{L}(i,j)$ for $i \neq j$. If the main element is horizontal, the neighboring elements are the vertical ones next to this main element. For vertical main elements it is similar. This choice of the regularization matrix incorporates a smoothness assumption about the variation of the resistor values.

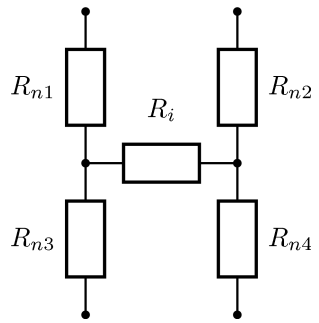


Figure 4.1: Definition of the neighbors

4.2.1 Some Different Definitions of Element-Neighbors

Other definitions of neighbors are also possible, see figure 4.2 for example. These alternative

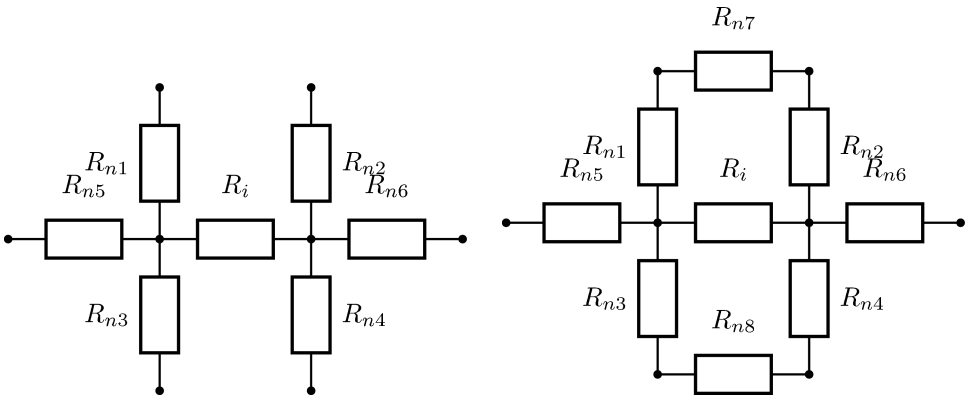


Figure 4.2: Other possible definitions of the neighbors

definitions of the neighbor elements take more neighbors into account than the first definition, hence they lead to a bigger smoothing of the regularized solution. It is also possible to weight the neighbors differently, so that the neighbors further away are less weighted than the adjacent ones. It is essential, that the value assigned to the "main" resistor is the negative sum of the values assigned to the neighboring resistors, so that the sum of each row of L equals zero.

4.3 Basis Constraint Method

The idea behind the basis constraint method is described here [23, 25] starting from the minimization functional

$$\min_{\mathbf{r}}\{||\mathbf{u}_r - \mathbf{u}(\mathbf{r})||^2 + \alpha||\mathrm{L}\mathbf{r}||^2\}.$$

First a set of preselected basis functions $\mathbf{w}_m$ is used, to set up an approximation of the sought distribution.

$$\mathbf{r} = \sum_{m=1}^{M} c_m \mathbf{w}_m(x) \quad , \quad c_m \in \mathbb{R},$$

where c_m are weights and $\mathbf{w}_m$ are the possible distribution vectors or basis functions. The number of different preselected basis functions M should be small (for example $3-5$). The construction of these basis function is possible with known structures or conductivities. A priori knowledge, like results of earlier measurements (like MRI) is also used. This kind of set of distributions is then also known as a learning set. The Subspace $\mathrm{S}_w = \mathrm{span}(\mathbf{w}_m)$ is assumed to be the space, in which vicinity the true solution is assumed to lie. A set of possible sought distributions (vectors) $\mathbf{r}_n$ is set up, with $n = 1, ..., N$, next to which the true distribution is assumed to be (learning set). The real distribution is next to one or next to a linear combination of these vectors. This set can be seen as an prior model. It is very inefficient to stabilize if N is very large, so a low-dimensional subspace is used (with predefined M elements). The covariance matrix C is calculated with

$$\mathrm{C} = N^{-1}\hat{\mathrm{R}}\hat{\mathrm{R}}^T,$$

where the matrix $\hat{\mathrm{R}} = [\mathbf{r}_1 ... \mathbf{r}_N]$. To approximate the smaller set with M elements, one calculate the M largest eigenvalues and their orthonormal eigenvectors $\mathbf{w}_m$ of C. This is a good compromise between a good approximation (large M) and good regularization (small M). The disadvantage of this method is, that when the prior information is incompatible with the reality, one gets misleading results, so the estimates exhibits the structures of the prior information although they are not existent. In medical detection of some unimplemented features it will fail [23].

4.4 Subspace Regularization Method

This method uses the same idea of the basis constraint method, but uses no strict linear combination of these basis functions [23]. It leads the approximation towards the Subspace $\mathrm{S}_w = span(\mathbf{w}_m)$. This is done by forming a new regularization matrix with the null space $N(\mathrm{L}) = \mathrm{S}_w$. the orthogonal projector onto S_w is WW^T where W is the matrix with the orthonormal eigenvectors $\mathbf{w}_m$ as its columns. The regularization matrix is constructed as follows

$$\mathrm{L} = \mathrm{I} - \mathrm{WW}^T,$$

with I the identity matrix. The regularization matrix is constructed in that way because of the relation

$$\mathrm{WW}^T\mathbf{r} \approx \mathbf{r}$$

which is valid for such vectors that comply approximately with the prior model. Now the whole matrix construction for such compatible vectors is considered

$$(\mathrm{I} - \mathrm{WW}^T)\mathbf{r} = \mathbf{r} - \mathrm{WW}^T\mathbf{r} \approx \mathbf{0},$$

therefore such vectors are favored in the reconstruction. This choice of the regularization matrix is the same as to use the distance between the subspace S_w and the solution $\mathbf{r}$ as penalty.

4.5 Diagonal Weighting of $\mathrm{J}^T\mathrm{J}$

In this section a diagonal weighting of $\mathrm{J}^T\mathrm{J}$ is used as regularization matrix L [23].

$$\mathrm{L} = \mathrm{diag}(\mathrm{J}^T\mathrm{J})$$

so that

$$\boldsymbol{\delta}\mathbf{r}_k = (\mathrm{J}^T\mathrm{J} + \alpha\ \mathrm{diag}(\mathrm{J}^T\mathrm{J}))^{-1}[\mathrm{J}^T(\mathbf{u}_r - \mathbf{u}(\mathbf{r}_k))]$$

is. The diagonal matrix can be thought here as an approximation of the part of the second derivative.

4.6 Minimum Mean Variation Method

If the cross section is subdivided in parts with different length of the edges, the regularization matrix can be weighted with the length of the edges to ensure different strength of the regularization

[23]. First the regularization matrix L is chosen as a discrete first order difference operator (for row i: 1 and -1 are set for the pixels on the common edge) , that the part $||\mathrm{L}\mathbf{r}||$ gives an approximation for the mean variation of $\mathbf{r}$. Additionally each row has been weighted with the length of the edge d_i of the edge i. The diagonal matrix D consists of the vector $\mathbf{d}$ in the diagonals. The final form can be written as

$$\boldsymbol{\delta}\mathbf{r}_k = (\mathrm{J}^T\mathrm{J} + \alpha\mathrm{L}^T\mathrm{D}^T\mathrm{DL})^{-1}[\mathrm{J}^T(\mathbf{u}_r - \mathbf{u}(\mathbf{r}_k)) - \alpha\mathrm{L}^T\mathrm{D}^T\mathrm{DL}\mathbf{r}_k].$$

4.7 Modifications of Standard Regularization Matrices due to Spatial Prior Information

According to Borsic in [13, p. 78] and [26] where the Tikhonov Regularization with a special regularization matrix is used, the main idea of the modification which is named Anisotropic Regularization is described here. Assuming that in many practical applications blocky structures are present, which means areas with constant values and steep variations between them, the smoothing effect of the Laplacian operator used as regularization matrix does not lead to optimal reconstructions because it doesn't allow steep variations. In many cases (for example the human body) the position of the areas is known but the size is only roughly known or is varying. Therefore, the whole area is separated in areas with expected small variations and in areas with expected steep variations. In the regions where small variations are expected, the standard Laplacian operator is used. In the regions with expected steep variations (the direction of the variation in usually known) the Laplacian operator values are relaxed only along the direction of the steep variation, allowing faster transitions in this direction while keeping the necessary smoothness tangentially [26]. It is also ensured that the sum of each row of L is zero hence the value of the center element is set to the negative sum of all appearing elements.

Another similar approach is made in [24], where the first order difference operator is used as regularization matrix L and modified to include prior information. The locations of the sharp edges are assumed a priori. The corresponding rows to this sharp edges are removed of the L matrix. With this modification resistance changes (according to the network problem) are allowed to be arbitrarily and not forced to be small over this edge as elsewhere. If the assumption is incorrect and the regularization parameter is well adjusted, the results are almost as good as the reconstruction without prior assumptions.

Also worth mentioning is the approach to reduce the degrees of freedom with the aid of groups. In [27] the idea is presented. The problem is assumed as a two phase-field, where only two different parameter values are present. The aim is the improvement of the spatial resolution of the reconstruction. Starting from a Gauss Newton based reconstruction with some iterations the resulting parameters are partitioned into three groups. The groups are defined by two thresholds μ_{TH} and μ_{TL}, where μ_{TL} is the lower threshold and all parameter values below this threshold are assigned to the background group BG. A similar procedure is done for the values which are higher than the upper threshold μ_{TH}, they are assigned to the target group TG. The remaining values,

which are between the two thresholds are assigned to the unadjusted group UG. All parameter in the background and target group are reduced to one sought parameter each. Doing that, the degrees of freedom are reduced and after some iterations the grouping is repeated. In [27] it is shown that the number of unknowns is reduced successively and thus the ill-posedness of the inverse problem is diminished.

5 Total Variation Method

5.1 Introduction

In chapter 2 the basics of the Total Variation theory were introduced. Now a detailed implementation is shown [9]. When "blocky" structures are assumed in the true object the Total variation method is a good choice for the penalty term. "Blocky" structure means piecewise constant values with steep changes between them. The complete minimization functional of linear form can be defined as

$$\min_{\mathbf{f}}(||\mathrm{K}\mathbf{f}-\mathbf{d}||_2^2+\alpha\mathrm{TV}(\mathbf{f})) \tag{5.1}$$

where $\mathrm{TV}(\mathbf{f})$ is in general

$$\mathrm{TV}(\mathbf{f})=\int_\Omega |\nabla\mathbf{f}|_\Omega \mathrm{d}\Omega.$$

This formula can also be written in the range of 0 to 1 without loss of generality, for the one dimensional case as

$$\mathrm{TV}_{1D}(f)=\int_0^1 \left|\frac{\mathrm{d}f}{\mathrm{d}x}\right| \mathrm{d}x. \tag{5.2}$$

and equivalent in the two dimensional case

$$\mathrm{TV}_{2D}(f)=\int_0^1\int_0^1 |\nabla f|\mathrm{d}x\ \mathrm{d}y,$$

where $\nabla f=(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y})$ denotes the gradient and $|(a,b)|=\sqrt{a^2+b^2}$ denotes the Euclidean norm [9, p. 129]. These formulas are true for smooth functions, this means for infinitely differentiable functions. There exists an important problem when implementing numerical methods: the L_1 norm is not differentiable at the origin. So the often used approximation of $|\mathbf{x}|$ like $\sqrt{|\mathbf{x}|^2+\beta^2}$ is used to handle that problem and this yields to the following formulas:
for the one dimensional case:

$$\mathrm{TV}_{1D}(f)=\int_0^1 \sqrt{\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)^2+\beta^2}\ \mathrm{d}x$$

and equivalent for the two dimensional case:

$$\mathrm{TV}_{2D}(f)=\int_0^1\int_0^1 \sqrt{\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)^2+\left(\frac{\mathrm{d}f}{\mathrm{d}y}\right)^2+\beta^2}\ \mathrm{d}x\ \mathrm{d}y.$$

where β is a small positive parameter. This approximation has the deviation of β at the origin and converges outside the origin to the original function [9, 17, 5].
In a geometrical interpretation of f the $\mathrm{TV}(f)$ can be seen as the lateral surface area of the graph

of f [9, p. 130].
The main difference between the first order Tikhonov and the Total Variation regularization is the grade of the used norm of the penalty term. The Tikhonov Regularization uses the L_2 norm independent of the used L-matrix, the Total Variation regularization uses the L_1 norm with the first order differences Operator (denoted as the matrix D) like in the first order Tikhonov Regularization.

$$\min_{\mathbf{f}}(||\mathrm{K}\mathbf{f} - \mathbf{d}||_2^2 + \alpha||\mathrm{D}\mathbf{f}||_1) \tag{5.3}$$

This general model has to be adopted numerically to the given discrete problem, which leads to some problems. This is made for 1D and 2D problems as well as for the network problem. Additionally the implemented and some other algorithms for the linear and nonlinear case are shown, but first of all the difference between L_1 and L_2 norm is demonstrated.

5.2 Difference Between L_1 and L_2 Norm

The main difference between the L_1 and L_2 norm should be explained with a simple piecewise linear function (see figure 5.1) [17, p. 34]. This penalty term applied to the given function $f(x)$

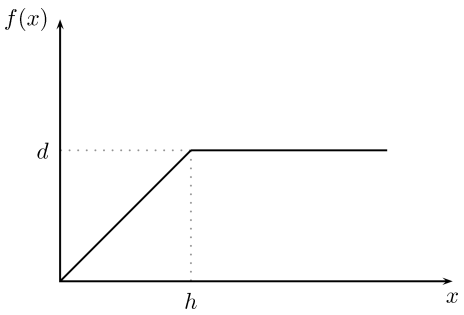


Figure 5.1: Piecewise linear function to explain the behavior of different norms

in the range of $(0, h)$ gives

$$\mathrm{TV}(f) = \int_0^h |f'(x)|dx.$$

To analyze this we need the general notation of norms with grade p applied to the given problem $f(x) = \frac{d}{h}x$

$$\|f'(x)\|_p^p = \int_0^h |f'(x)|^p dx = \int_0^h \left(\frac{d}{h}\right)^p dx = \left(\frac{d}{h}\right)^p h = \frac{d^p}{h^{p-1}}.$$

The norm is then given by

$$\|f'(x)\|_p = \sqrt[p]{\frac{d^p}{h^{p-1}}} = \frac{d}{h^{\frac{p-1}{p}}}.$$

The result for the L_1 and L_2 norm follows to

$$\|f'(x)\|_1 = d$$
$$\|f'(x)\|_2 = \frac{d}{\sqrt{h}}$$

It can be seen that both norms penalizes the transitions in f (because of d) but the L_2 norm penalizes fast transitions (with small h) more than slow transitions (with big h), in contrast to the L_1 norm. There is no difference between fast and slow transitions. Therefore the L_2 norm penalizes the abrupt changes in f and the L_1 norm has no effect on the gradient of $f(x)$ because it is no function of h.

5.3 1D-Discretization of the Functional

The 1D-TV-functional given in equation (5.2) on page 51 has to be discretized for implementation. Hence the function $f(x)$ is supposed to be a smooth function with [9, p. 131]

$$\mathbf{f} = (f_0, ..., f_n) \quad \text{with} \quad f_i \approx f(x_i), x_i = i\Delta x, \Delta x = \frac{1}{n}$$

and the approximation of the derivative is defined as

$$\mathrm{D}_i\mathbf{f} = \frac{f_i - f_{i-1}}{\Delta x} \quad i = 1, ..., n.$$

So the vector D_i has the size $(n+1) \times 1$ and is nonzero on the position i and $i-1$ and can be seen as one row of the matrix D in equation (5.3) on the preceding page.

$$\mathrm{D}_i = [0, ..., 0, -\frac{1}{\Delta x}, \frac{1}{\Delta x}, 0, ..., 0]$$

Hence the discrete penalty term arises to

$$\mathrm{TV}_d(\mathbf{f}) = \frac{1}{2}\sum_{i=1}^{n} \phi((\mathrm{D}_i\mathbf{f})^2)\Delta x$$

The function ϕ is a smooth function and is integrated to handle the non-differentiability at the origin with the property

$$\phi'(t) > 0 \quad \text{whenever} \quad t > 0.$$

A common approximation for $\phi(t)$ is

$$\phi(t) = 2\sqrt{t + \beta^2}.$$

The variable t is defined as $t = (\mathrm{D}_i\mathbf{f})^2$. The derivatives of $\phi(t)$ results in

$$\phi'(t) = \frac{1}{\sqrt{t + \beta^2}} \tag{5.4}$$

$$\phi''(t) = -\frac{1}{2}\left(t + \beta^2\right)^{-\frac{3}{2}} \tag{5.5}$$

There are another possible approximations of $\phi'(t)$ shown in [9, p. 131] and in [5, p. 177].
Now the optimal solution of the linear optimization functional has to be computed. This derivation is taken from [5, p. 176] because its easier to understand than in [9, p. 133] while the result is the same with the assumption $\Delta x = 1$ which can always be absorbed by the regularization parameter.

Starting with the functional (see equation (5.3) on page 52)

$$\min_{\mathbf{f}}(||\mathrm{K}\mathbf{f}-\mathbf{d}||_2^2+\alpha||\mathrm{D}\mathbf{f}||_1) \tag{5.6}$$

and with the substitution $\mathbf{y}=\mathrm{D}\mathbf{f}$, getting $||\mathrm{D}\mathbf{f}||_1=\sum_{j=1}^{m}|y_j|$ with m number of elements of $\mathbf{y}$ the gradient is obtained by

$$\mathbf{g}(\mathbf{f})=2\mathrm{K}^T(\mathrm{K}\mathbf{f}-\mathbf{d})+\alpha\sum_{j=1}^{m}\nabla|y_j|$$

For nonzero y_j (what is ensured above)

$$\nabla|y_j|=\frac{\partial|y_j|}{\partial f_i}=\mathrm{D}_{j,i}\ \mathrm{sgn}(y_j)=\mathrm{D}_{j,i}\frac{y_j}{|y_j|}$$

Now the diagonal matrix $\mathrm{W}(\mathbf{f})$ can be defined. $\mathrm{W}(\mathbf{f})$ has the elements

$$\mathrm{W}_{j,j}(\mathbf{f})=\frac{1}{|y_j|}=\frac{1}{\sqrt{(\mathrm{D}_j\mathbf{f})^2+\beta^2}}.$$

Then the gradient of the penalty functional is given as

$$\nabla(||\mathrm{D}\mathbf{f}||_1)=\mathrm{D}^T\mathrm{W}(\mathbf{f})\mathrm{D}\mathbf{f}$$

The complete gradient is given by

$$\mathbf{g}(\mathbf{f})=2\mathrm{K}^T(\mathrm{K}\mathbf{f}-\mathbf{d})+\alpha\mathrm{D}^T\mathrm{W}(\mathbf{f})\mathrm{D}\mathbf{f}$$

To obtain the optimal solution the gradient is set to zero

$$(2\mathrm{K}^T\mathrm{K}+\alpha\mathrm{D}^T\mathrm{W}(\mathbf{f})\mathrm{D})\mathbf{f}=2\mathrm{K}^T\mathbf{d}$$

The matrix $\mathrm{W}(\mathbf{f})$ equals the matrix $\mathrm{diag}(\phi'((\mathrm{D}_j\mathbf{f})^2))$ [9, p. 132] and depends on $\mathrm{D}\mathbf{f}$. It is an nonlinear system of equations. Because $\mathrm{W}(\mathbf{f})$ depends on the sought solution, the iteration process has to be started with $\mathbf{f}_0$ and some iterations are needed to gain convergence. So opposite to the Tikhonov Regularization the Total Variation method is not a single step solution. Additionally the Hessian matrix can be calculated with [9, p. 133]

$$\mathrm{H}(\mathbf{f})=2\mathrm{K}^T\mathrm{K}+\alpha\mathrm{D}^T\mathrm{W}(\mathbf{f})\mathrm{D}+\alpha\mathrm{D}^T\mathrm{W}'(\mathbf{f})\mathrm{D}$$

with the definition $\mathrm{W}'(\mathbf{f})=\mathrm{diag}(2(\mathrm{D}\mathbf{f})^2\phi''((\mathrm{D}_j\mathbf{f})^2))$ and ϕ'' equivalent to the approximation in equation (5.5) on the preceding page. In the following algorithms the following notations according to [9] are used

$$\mathrm{L}(\mathbf{f})=\mathrm{D}^T\mathrm{W}(\mathbf{f})\mathrm{D}$$
$$\mathrm{L}'(\mathbf{f})\mathbf{f}=\mathrm{D}^T\mathrm{W}'(\mathbf{f})\mathrm{D}$$

To analyze the penalty term, the entries of the diagonal matrices W and W′ are plotted. For the approximated Hessian matrix only the first term W is used. One entry of this diagonal matrix is denoted as $y_1(x)$ and for the complete Hessian matrix both terms $\mathrm{W} + \mathrm{W}'$ are used. One entry of this matrix is denoted as $y_2(x)$ where x corresponds to $\mathrm{D}_j\mathbf{f}$, the difference of the neighboring elements.

$$y_1(x) = \frac{1}{\sqrt{x^2 + \beta^2}}$$
$$y_2(x) = \frac{\beta^2}{(x^2 + \beta^2)^{\frac{3}{2}}}$$

Both terms are plotted with $\beta = 0.1$ in figure 5.2. It can be seen that the term for the approximation of the Hessian matrix converges much slower to zero than the term of the complete Hessian matrix but the maximum value which is $1/\beta$ is the same.

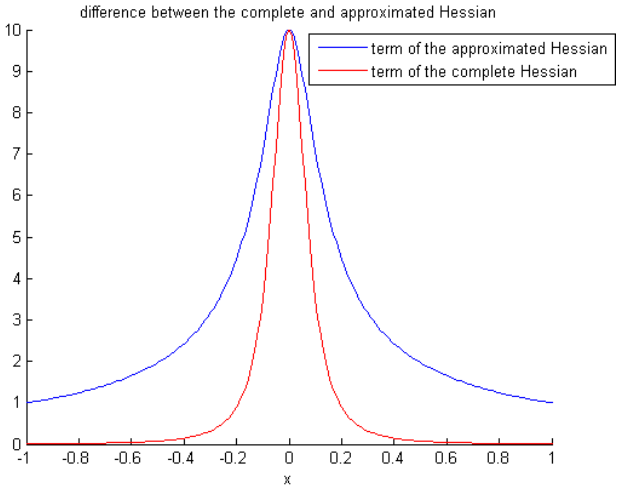


Figure 5.2: Constructed entries of the diagonal matrices W and $W+W'$ versus x which corresponds to $\mathrm{D}_i\mathbf{f}$ with $\beta = 0.1$

The algorithm with the solution in the second chapter plotted figure 2.13 on page 32 can be written as

$k = 0$

$\mathbf{f}_0 =$ initial guess

begin loop

$\mathrm{H}_k = 2\mathrm{K}^T\mathrm{K} + \alpha\mathrm{L}(\mathbf{f}_k)$ approximated Hessian matrix

$\mathbf{f}_{k+1} = \mathrm{H}_k^{-1}(2\mathrm{K}^T\mathbf{d})$ update step derived by setting the gradient $\mathbf{g}_k = \mathbf{0}$

end loop repeat until the solution converges

Additionally some other linear algorithms and the corresponding solutions are discussed briefly now.

5.4 Some Linear Algorithms

5.4.1 Steepest Descent Algorithm

The simple algorithm from the steepest descent method, where the search direction is the negative gradient, is modified for Total Variation-Penalized Least Squares [9, p. 135]

$k = 0$

$\mathbf{f}_0 =$ initial guess

begin loop

$\mathbf{g}_k = 2\mathrm{K}^T(\mathrm{K}\mathbf{f}_k - \mathbf{d}) + \alpha \mathrm{L}(\mathbf{f}_k)\mathbf{f}_k$ calculation of the gradient

$\gamma_k = \arg\min_{\gamma>0} \psi(\mathbf{f}_k - \gamma \mathbf{g}_k)$ linesearch

$\mathbf{f}_{k+1} = \mathbf{f}_k - \gamma \mathbf{g}_k$ update

$k = k + 1$

end loop

In figure 5.3 the result for the steepest descent with TV algorithm for 50 iterations is plotted. The selected number of iterations is small for the steepest descent algorithm, hence the result is

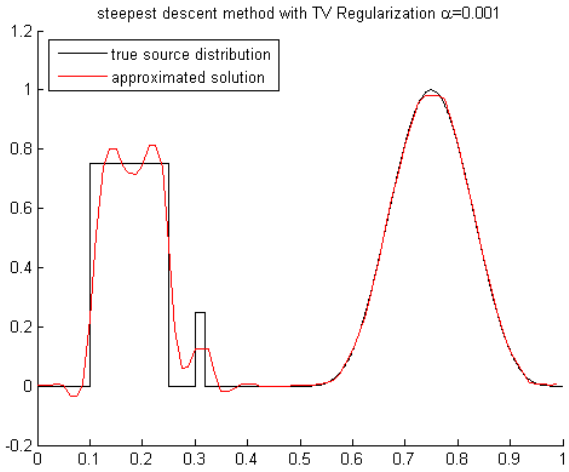


Figure 5.3: Total variation steepest descent method with $\alpha = 0.001$ after 50 iterations

smooth. With 500 iterations the result is similar to figure 2.13 on page 32.

5.4.2 Newton Method

The next algorithm is the modified Newton algorithm, where the full Hessian matrix is used.

$k = 0$

$\mathbf{f}_0 =$ initial guess

begin loop

$\mathbf{g}_k = 2\mathrm{K}^T(\mathrm{K}\mathbf{f}_k - \mathbf{d}) + \alpha \mathrm{L}(\mathbf{f}_k)\mathbf{f}_k$ calculation of the gradient

$\mathrm{H}_k = 2\mathrm{K}^T\mathrm{K} + \alpha[\mathrm{L}(\mathbf{f}_k) + \mathrm{L}'(\mathbf{f}_k)\mathbf{f}_k]$ Hessian matrix

$\mathbf{s}_k = -\mathrm{H}_k^{-1}\mathbf{g}_k$ calculation of one Newton step, search direction

$\gamma_k = \arg\min_{k>0} \psi(\mathbf{f}_k + \gamma \mathbf{s}_k)$ linesearch

$\mathbf{f}_{k+1} = \mathbf{f}_k + \gamma \mathbf{s}_k$ update

$k = k + 1$

end loop

The complete Hessian matrix is more difficult to be controlled than the approximated Hessian matrix. A possible method to overcome this problem is the successive decrease of β every iteration. Additionally a well working line search procedure is needed. The result, equivalent to the Quasi-Newton approach is shown in figure 5.4.

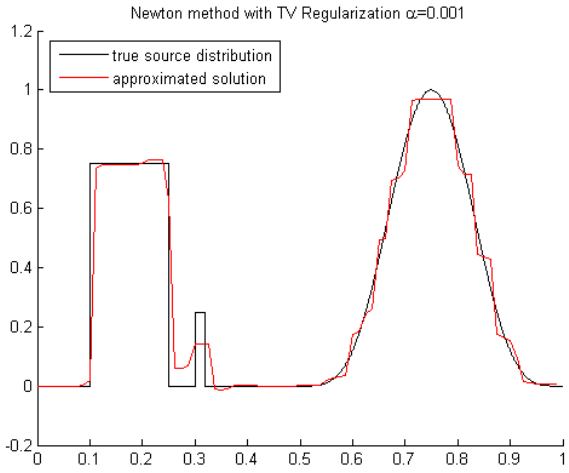


Figure 5.4: Total variation Newton method with $\alpha = 0.001$ and $\beta = 0.001$

5.4.3 Lagged Diffusivity Fixed Point Method

In general this is the same as the algorithm of the solution in the second chapter, the similarity is shown starting from the equation for the iteration step k [9]

$$\mathbf{f}_{k+1} = [2\mathrm{K}^T\mathrm{K} + \alpha \mathrm{L}(\mathbf{f}_k)]^{-1} 2\mathrm{K}^T\mathbf{d} = \mathbf{f}_k - [2\mathrm{K}^T\mathrm{K} + \alpha \mathrm{L}(\mathbf{f}_k)]^{-1}[2\mathrm{K}^T(\mathrm{K}\mathbf{f} - \mathbf{d}) + \alpha \mathrm{L}(\mathbf{f}_k)\mathbf{f}_k].$$

The first definition is the equation of the fixed point iteration like derived above, the second is the Quasi-Newton form, both are equivalent. The Quasi-Newton form can be derived by a simplified

Hessian without the $\mathrm{L}(\mathbf{f})'\mathbf{f}$-Term, where the result equals the optimal solution derived above. This results to

$$[2\mathrm{K}^T\mathrm{K} + \alpha\mathrm{L}(\mathbf{f}_k)]\Delta\mathbf{f} = 2\mathrm{K}^T\mathbf{d} \quad \text{with} \quad \Delta\mathbf{f} = \mathbf{f}_{k+1} - \mathbf{f}_k$$

The following algorithm is based on the Quasi Newton form. This form is not so sensitive to roundoff errors than the fixed point form [9].

$k = 0$

$\mathbf{f}_0 =$ initial guess

begin loop

$\mathbf{g}_k = 2\mathrm{K}^T(\mathrm{K}\mathbf{f}_k - \mathbf{d}) + \alpha\mathrm{L}(\mathbf{f}_k)\mathbf{f}_k$ calculation of the gradient

$\mathrm{H}_k = 2\mathrm{K}^T\mathrm{K} + \alpha\mathrm{L}(\mathbf{f}_k)$ Hessian matrix

$\mathbf{s}_k = -\mathrm{H}_k^{-1}\mathbf{g}_k$ calculation of one Newton step, search direction

$\mathbf{f}_{k+1} = \mathbf{f}_k + \mathbf{s}_k$ update

$k = k + 1$

end loop

The term $\mathrm{L}(\mathbf{f})$ is termed the discrete diffusion operator. Therefrom the name "lagged diffusivity" is derived. If $\mathrm{K}^T\mathrm{K}$ is positive definite, it is proven that the algorithm converges globally without a line search and a linear rate of convergence can be expected [9, p. 136]

5.5 Implemented Gauss-Newton Method with Total Variation

According to the linear case above,the variable t is defined as $t_j = \mathrm{D}_j\mathbf{r}$ where D_j is a row of the D-Matrix.

$$t_j = \mathrm{D}_j\mathbf{r} = \frac{r_k - r_j}{\Delta r} \quad for \quad j = 1 \ldots b, \quad k = 1 \ldots \text{ neighbors of } j$$

where b is the number of resistors and D_j the discrete derivative approximation. And Δr is set to $\Delta r = 1$ for simplicity. The D-Matrix is first constructed from the L-Matrix, for every neighboring resistor of the current resistor a row D_i is constructed. As neighbor definition the L-Matrix is used (see figure 4.1 on page 46), so for each "main" resistor 4 neighbors are defined and 4 rows in the D-Matrix are constructed. So the D-Matrix would have the size 512×144. The D-Matrix is also given when taking the horizontal neighbors into account only, because every difference between two neighboring resistors appears. Then the D-Matrix has the size 256×144. Now it is needed to determine the gradient and the Hessian of the new penalty term. Although its a planar problem it's not a real 2D-problem with x and y variable, because the sought vector is $\mathbf{r}$, hence it is computed as 1D-problem with the above constructed D-Matrix. Stability problems occurred with the use of the $\mathrm{L}'(\mathbf{r})\mathbf{r}$ term in Hessian matrix (full Newton step), hence only the quasi-Newton steps with the penalty term of Hessian matrix $\mathrm{L}(\mathbf{r})$ is used, which leads to the following derivation of the Gauss Newton method with the Total Variation Functional.

The following minimization problem is discussed

$$\min_{\mathbf{r}} \psi(\mathbf{r})$$

with

$$\psi = \underbrace{\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|_2^2}_{\text{residual norm}} + \alpha \underbrace{\|\mathrm{D}\ \mathbf{r}\|_1}_{\text{TV penalty term}} \tag{5.7}$$

Equivalent to the Gauss-Newton approximation the Taylor-series approximation is applied and the gradient of the functional is computed

$$\mathbf{g}_k = 2\mathrm{J}_r(\mathbf{u}(\mathbf{r}_k) - \mathbf{u}_r) + \alpha \nabla \|\mathrm{D}\ \mathbf{r}_k\|_1$$

The term $\frac{\partial u}{\partial \mathbf{r}}$ denotes the Jacobian matrix J and the derivative of the TV term can be written as

$$\nabla \|\mathrm{D}\ \mathbf{r}_k\|_1 = \mathrm{D}^T \mathrm{W}(\mathbf{r}_k)\mathrm{D}\ \mathbf{r}_k$$

where the matrix W is $\mathrm{W} = \mathrm{diag}\left(\phi'((\mathrm{D}_j\mathbf{r})^2)\right)$ as previous defined. The Hessian matrix is given by

$$\mathrm{H}_k = 2\mathrm{J}_k^T\mathrm{J}_k + \alpha \mathrm{D}^T\mathrm{W}(\mathbf{r}_k)\mathrm{D} + \underbrace{2\frac{\partial^2 \mathbf{u}}{\partial \mathbf{r}^2}(\mathbf{u}(\mathbf{r}_k - \mathbf{u}_r) + \alpha \mathrm{D}^T \frac{\partial \mathrm{W}}{\partial \mathbf{r}}\mathrm{D}\mathbf{r}_k}_{\text{2nd order term is neglected}}$$

Equivalent to the standard Gauss-Newton case the second order terms are neglected.

$$\delta\mathbf{r}_k = (\mathbf{r}_{k+1} - \mathbf{r}_k) = -\mathrm{H}_k^{-1}\mathbf{g} = -(2\mathrm{J}_k^T\mathrm{J}_k + \alpha \mathrm{D}^T\mathrm{W}(\mathbf{r}_k)\mathrm{D})^{-1}(2\mathrm{J}_k^T(\mathbf{u}(\mathbf{r}_k) - \mathbf{u}_r) + \alpha \mathrm{D}^T\mathrm{W}(\mathbf{r}_k)\mathrm{D}\mathbf{r}_k)$$

In the linear case the Total Variation algorithm is an iterative algorithm because of the dependency of the penalty functional of the sought values. In the (Gauss-) Newton case the Total Variation term is only dependent on the actual constant resistor values $\mathbf{r}_k$, hence there are no additional iterations necessary. The following Gauss-Newton-based algorithm is used:

$k = 0$

$\mathbf{r}_0 =$ initial guess

begin loop

  $\mathrm{L}(\mathbf{r}_k) = \mathrm{D}^T\mathrm{W}(\mathbf{r}_k)\mathrm{D}$

  $\mathbf{g}_k = 2\mathrm{J}^T(\mathbf{u}(\mathbf{r}_k) - \mathbf{u}) + \alpha \mathrm{L}(\mathbf{r}_k)\mathbf{r}_k$ calculation of the gradient

  $\mathrm{H} = 2\mathrm{J}^T\mathrm{J} + \alpha \mathrm{L}(\mathbf{r}_k)$ approximated Hessian matrix

  $\mathbf{s}_k = -\mathrm{H}^{-1}\mathbf{g}_k$ calculation of one Newton step, search direction

  $\gamma_k = \arg\min_{\gamma>0} \psi(\mathbf{f}_k + \gamma \mathbf{s}_k)$ linesearch

  $\mathbf{r}_{k+1} = \mathbf{r}_k + \gamma \mathbf{s}_k$ update

  $k = k + 1$

end loop

6 Results and Discussion

6.1 Introduction

In this chapter the implemented regularization methods are briefly described and results are presented. First the circuits used are shown. After that possible errors in an practical problem are described and the error approximation used is defined. Finally the measurement values are described.

6.1.1 Used Circuits

Three different circuits are selected to show reconstruction results. They are consisting of continuous areas with homogeneous resistances. The distribution can be seen as objects against a background. The unit of the resistances (Ω) is not specified. The background consists of resistors with resistance 1. The first circuit contains a small object with value 3 and a big object with value 5, see figure 6.1. On the basis of this network the implemented algorithms are applied and the results are compared. The second circuit consists also of two objects similar to the first one but with interchanged resistances, see figure 6.2 on the following page. Additionally a third circuit (see figure 6.3 on the next page) is used. It consists of one big object in the center. For every circuit error calculations are made. Because of the rank deficiency of the problem, even if no

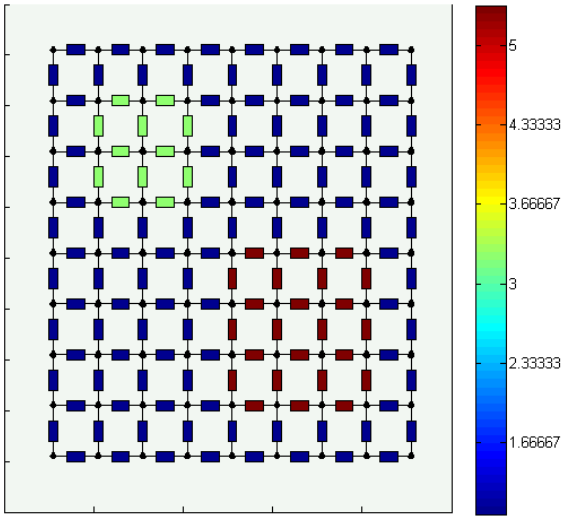


Figure 6.1: Circuit 1: Two different objects diagonally arranged

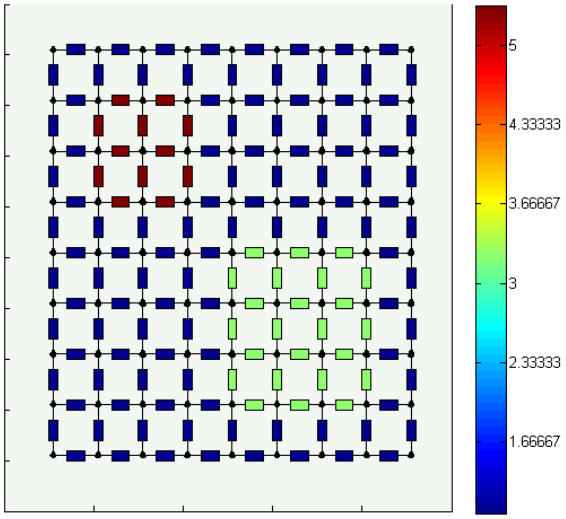


Figure 6.2: Circuit 2: Two different objects diagonally arranged with interchanged values

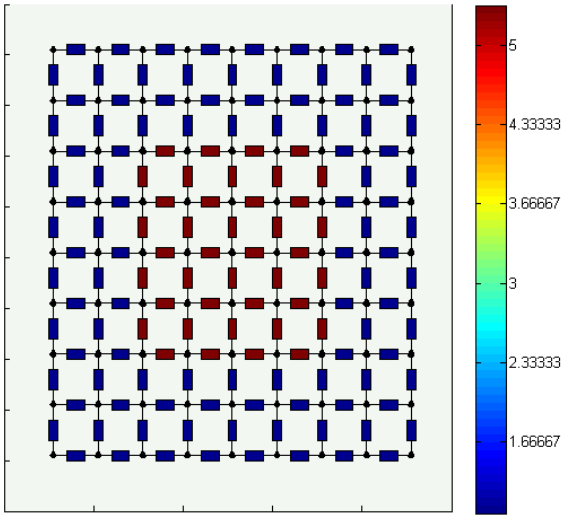


Figure 6.3: Circuit 3: One big object in the center

noise is present the true distribution cannot be reconstructed easily. The missing equations have to be replaced with information from the regularization term. Therefore, reconstructions are shown with and without noise.

6.1.2 Errors

Errors can occur in measuring the values $\mathbf{u}_r$ and in computation of the optimal solution $\mathbf{r}$.
The measurement errors should be as small as possible, but they are always present because of limited accuracy of the measuring instruments, which can be seen as instrument noise. There is always some noise present in practical applications, which is common for every electrode at the boundary. It can be seen as noise induced from the measurement environment, for example [28].
Additionally errors (due to the computation of the solution) can occur in the solution because of following reasons [12].

- Rounding errors: The cause is the representation of numbers on the computer. The numbers are typically floating point numbers with bounded length and usually written as a signed fraction times a power of 10. Hence, only a finite set of numbers can be represented, so the real number must be rounded.
- Cancellation errors: They occur during arithmetic operations caused by the bounded storage space for the numbers.

Additionally it should stated, that the ill-posedness or ill-conditioning of problems has nothing to do with the floating point-operations. It is a property of the problem itself.
Due to the simple forward problem of the present network problem, the errors occurring in the computation are assumed to be very small and they are negligible. The measurement errors which are always present in practical applications are simulated. They are assumed as additive Gaussian noise with zero mean and a standard deviation of 1% of each boundary voltage. Additionally the reconstruction is made without noise to visualize the best possible results. After that the different methods are applied to the different circuits, and the solutions are discussed on the basis of the relative mean error ϵ_{mean}, the relative maximum error $\epsilon_{r,max}$ and the absolute maximum error $\epsilon_{a,max}$. They are defined as follows [29, 30, 31]

$$\begin{aligned} \epsilon_{mean} &= \sqrt{\frac{\sum_i (R_{i\ recon} - R_{i\ true})^2}{\sum_i (R_{i\ true})^2}} \ \times \ 100\% \\ \epsilon_{r,max} &= \max \frac{|R_{i\ recon} - R_{i\ true}|}{|R_{i\ true}|} \ \times \ 100\% \\ \epsilon_{a,max} &= \max |R_{i\ recon} - R_{i\ true}|. \end{aligned} \tag{6.1}$$

The same errors are calculated for the voltage values $\mathbf{u}_r$ for comparison with the calculated resistances. For the resulting solution $\mathbf{r}$ the number of iterations k needed, the resulting residual norm and the maximum and minimum resistances are displayed as well. The values of the regularization parameter computed with the different methods are compared.

6.1.3 Measurement Voltages with Noise

Here the used measurement voltages with noise are shown for every circuit. The values are calculated with the same noise vector for comparison (for 8 and 12 electrodes separately). The noise vector is weighted with the corresponding values to construct a noise equal to about 1 percent of the value for each value. The correct and the perturbed values of the voltage vector $\mathbf{u}_r$ for the 8 electrode arrangement are shown in figure 6.4, figure 6.5 and figure 6.6. Similarly, the values for the 12 electrode arrangement are shown in figure 6.7, figure 6.8 and figure 6.9. Additionally the maximum and minimum values and the errors are shown in table 6.1 for the 8 electrode arrangement and in table 6.2 on page 65 for the 12 electrode arrangement.

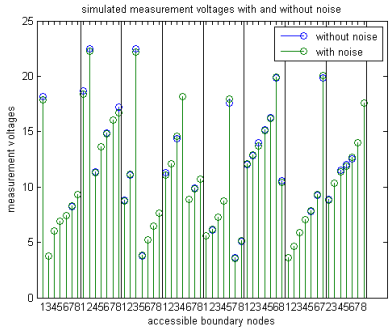


Figure 6.4: Measurement values for circuit 1 with 8 electrodes

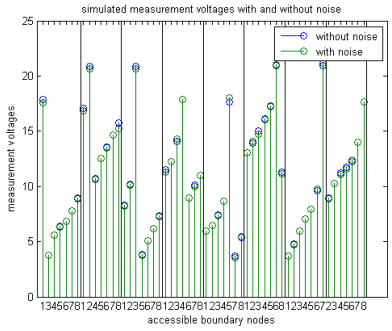


Figure 6.5: Measurement values for circuit 2 with 8 electrodes

circuit	U_{max} mV	U_{min} mV	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ mV
1	22.2348	3.5293	1.1009	3.0292	0.5211
2	21.1449	3.5745	1.0847	3.0292	0.4771
3	19.0328	3.2792	1.0733	3.0292	0.4192

Table 6.1: Measurement values for the circuits with 8 electrodes

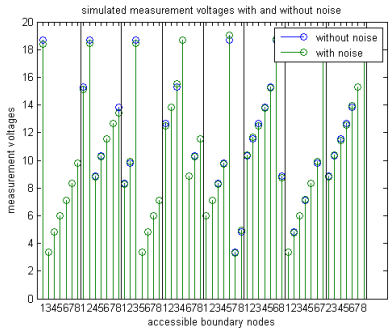


Figure 6.6: Measurement values for circuit 3 with 8 electrodes

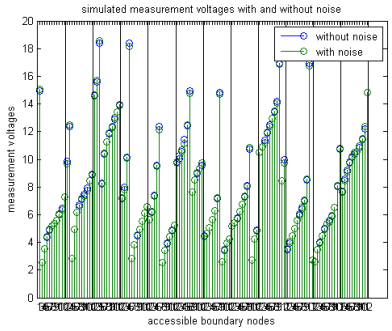


Figure 6.7: Measurement values for circuit 1 with 12 electrodes

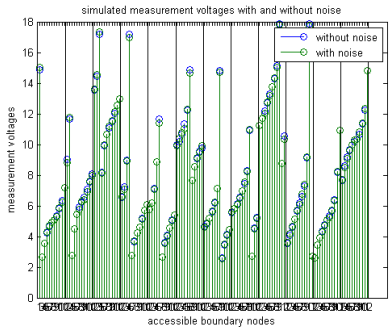


Figure 6.8: Measurement values for circuit 2 with 12 electrodes

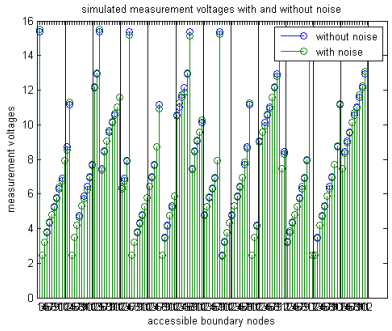


Figure 6.9: Measurement values for circuit 3 with 12 electrodes

network	U_{max} mV	U_{min} mV	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ mV
1	18.5576	2.5321	0.9773	2.4969	0.2853
2	17.8204	2.5503	0.9660	2.4969	0.2828
3	15.4994	2.4028	0.9839	2.4969	0.3036

Table 6.2: Measurement values for the circuits with 12 electrodes

6.2 Circuit 1

6.2.1 True Distribution

In figure 6.10 the true distribution of resistances is displayed. It is shown in two representations. The left figure shows the circuit of resistors and the color of each resistor shows its resistance. The right figure shows a three dimensional plot of the resistances. Due to the structure of the network, not every row consists of the same number of resistors. There are rows with 8 or 9 resistors and the position of the values should be comparable to the real configuration. Therefore, the values in between are interpolated. The result is a net of 17 x 17 values, which can be displayed in three dimensions.

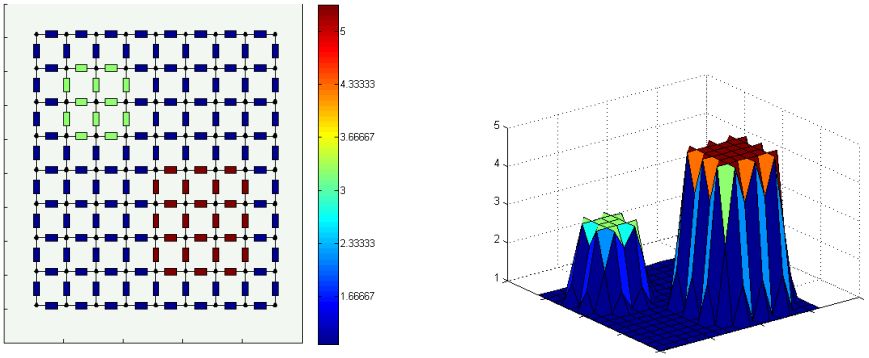


Figure 6.10: Circuit 1, true distribution

6.2.2 Tikhonov Regularization Method

Results Without Noise

First reconstructions with the second order Tikhonov Regularization method are shown. The L-matrix is selected as the discrete Laplace operator with the neighboring element definitions shown in figure 4.1 on page 46. There is no measurement noise added to the vector $\mathbf{u}_r$. The reconstructions are computed for both electrode arrangements (8 and 12 electrodes). Since there is no noise present, the stopping criteria is a very small value of the residual norm. If the residual norm reaches the value 10^{-8} the iteration is stopped. Additionally a maximum number of 50 iterations is specified. The regularization parameter is diminished every iteration with $\alpha_{k+1} = \alpha_k/10$ and the starting value is selected with $\alpha_0 = 0.1$ and positive definiteness is enforced in every iteration ($\alpha = 2\alpha$ until positive definiteness is given). The used L-matrix is the discrete Laplace operator (see chapter 4), which incorporates a smoothness assumption. In figure 6.11 on the following page the result with 8 electrodes is shown. The solution is very smooth (because of the assumption of the L-matrix), but the two objects are located. Figure 6.12 on the next page shows the result for the 12 electrode model. There the algorithm tries to approximate the true distribution. The results are shown in table 6.3 on page 68. There e is the number of electrodes, k is the iteration count, $||\mathbf{u}(\mathbf{r}) - \mathbf{u}_r||_2^2$ is the residual norm after k iterations, R_{max} and R_{min} are

the maximum and minimum resistances after k iterations and α_k is the regularization parameter at iteration k. The remaining values ϵ_{mean}, $\epsilon_{r,max}$ and $\epsilon_{a,max}$ are the errors which are defined in equation (6.1) on page 62. The reconstruction of the 12 electrode model reaches the maximum number of iterations. Therefore, the reconstruction is repeated with 80 iterations and the results are shown in figure 6.13 on the following page and table 6.4 on the next page. The residual norm has not changed significantly from iteration 51 to 80. Since the smoothness assumption is not correct for this problem, the algorithm is not able to reach the stopping criterion. In case of the 8 electrode model, there is much less information of the true distribution available. Therefore, the stopping criterion is reached with a smooth solution. In table 6.3 can be seen that the relative maximum error for 12 electrodes is higher than for 8 electrode, because of oscillations in the background values. Also R_{min} is smaller than in the result with 8 electrodes.

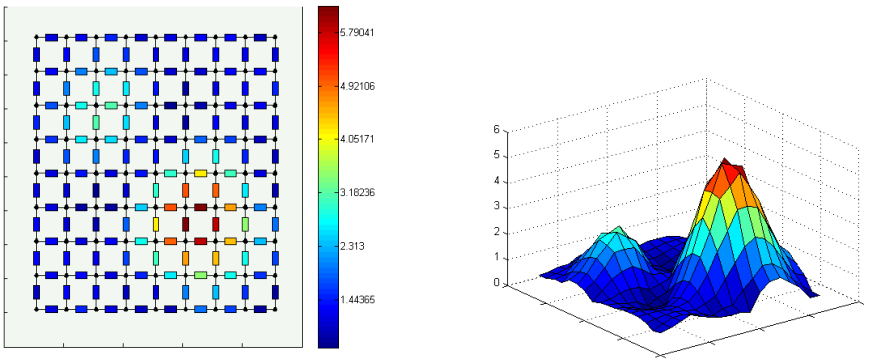


Figure 6.11: 2nd order Tikhonov regularization without noise, circuit 1 with 8 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

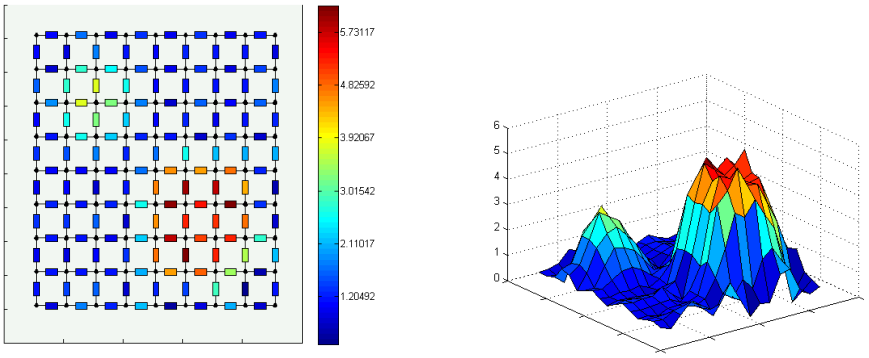


Figure 6.12: 2nd order Tikhonov regularization without noise, circuit 1 with 12 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

Additionally one reconstruction with the zeroth-order Tikhonov Regularization (where the Regularization matrix is the Identity matrix) is shown without adding noise (see figure 6.14 on page 69). The resulting values are shown in table 6.5 on page 69. The result is much worse than the result with the smoothness assumption, since the used penalty term tries to push the solution $\mathbf{r}$ towards

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	25	6.0236e-9	5.7904	0.5743	32.4304	157.9615(1)	2.5782(5)	1.3744e-14
12	50	6.4363e-4	5.7312	0.2997	22.5447	165.7045(1)	1.8287(5)	1.3292e-14

Table 6.3: Tikhonov Regularization without noise, circuit 1

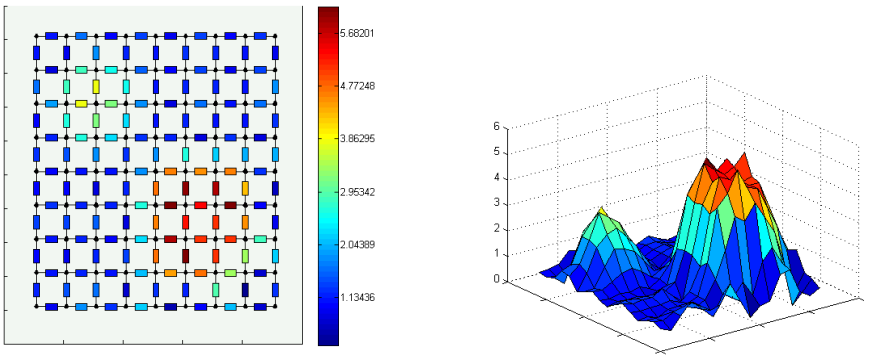


Figure 6.13: Reconstruction without noise, circuit 1 with 12 electrodes: 2nd order Tikhonov regularization with regularization parameter adjusted with $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$ after 80 iterations

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
12	80	3.1906e-4	5.6820	0.2248	23.0700	169.2961(1)	1.9378(5)	1.6850e-14

Table 6.4: Tikhonov reconstruction without noise, circuit 1 with 80 iterations

zero.

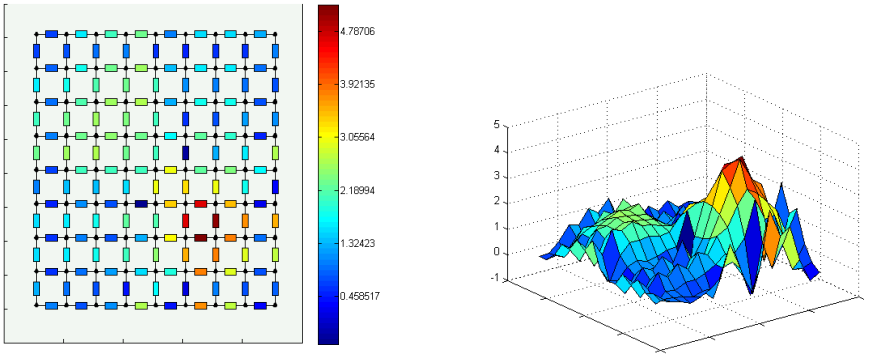


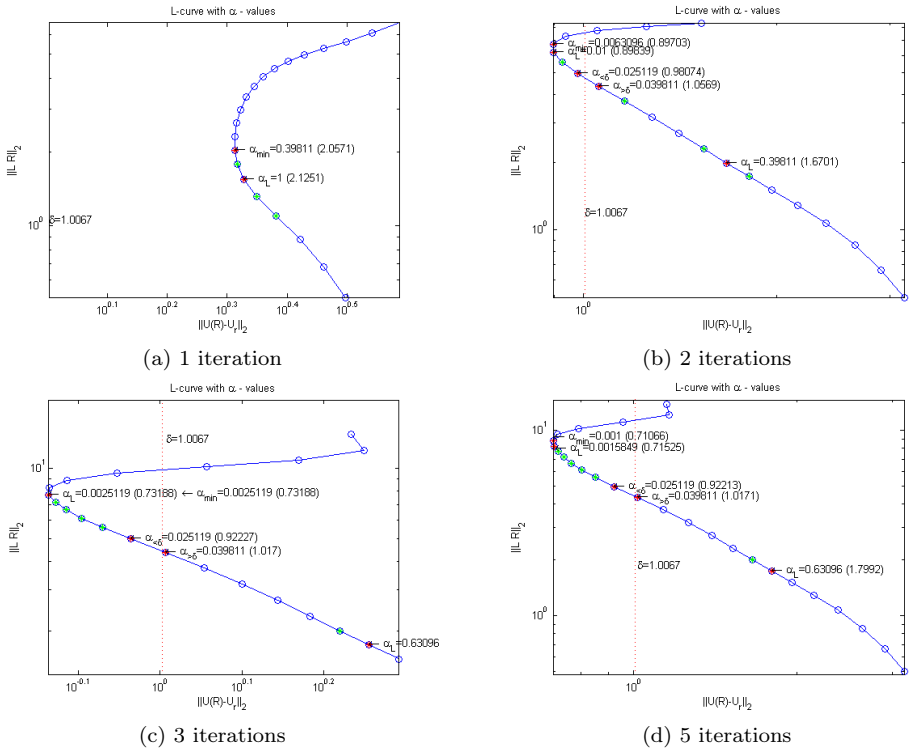
Figure 6.14: Reconstruction without noise, circuit 1 with 8 electrodes: zeroth order Tikhonov regularization with the parameter adjusted with $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	32	9.3377e-9	4.7871	-0.4072	47.5091	234.6539(1)	3.4051(5)	1.8447e-13

Table 6.5: Tikhonov reconstruction without noise, circuit 1 with the identity matrix

Determination of the Regularization Parameter with the L-curve

The measurement values of the 8 electrode model are perturbed with noise (see figure 6.4 on page 63 and table 6.1 on page 63). As above the L-matrix is the discrete Laplace operator. The regularization parameter α is obtained by the L-curve method for different number of iterations. To calculate the L-curve several regularization parameter are chosen. With each parameter the reconstruction is performed. The resulting terms, the residual norm (x-axis) and the penalty term (y-axis) are plotted for 1, 2, 3 and 5 iterations, see figure 6.15 on the following page. Additionally the δ parameter from the discrepancy principle (which is a measure of the noise level) is displayed in the L-curve. There it can be seen that the L-curve parameter is chosen independently from the noise level. The α values obtained from the L-curve are displayed in table 6.6 on the next page for the different number of iterations. Each of this values is used as a fixed regularization parameter for the reconstructions. The stopping criteria is the δ parameter and 2 additional iterations are performed after reaching this threshold δ. With the first parameter the stopping criterion cannot be reached because it is too large. The solution after 50 iterations is very smooth (see figure 6.16 on the following page), therefore this parameter cannot be used as a fixed regularization parameter but only as a starting value for iterative methods. The other values are much smaller and hence, the solutions are much better (see figure 6.17, figure 6.18 and figure 6.19). The corresponding results are shown in table 6.7 on page 72. The error values of the reconstruction with the third value are the best, but there occur negative resistances. This is possible, because it is not prevented by an additional constraint. So the algorithm doesn't know that this values are impossible. Therefore box-constraints are applied next.



(a) 1 iteration

(b) 2 iterations

(c) 3 iterations

(d) 5 iterations

Figure 6.15: L-curve of circuit 1 for the 8 electrode model and different number of iterations for each α value

iterations	α_L	residual norm mV^2	α_{min}	residual norm mV^2
1	1	2.1251	0.39811	2.0571
2	0.01	0.89839	0.00631	0.89703
3	0.0025	0.73188	0.0025	0.73188
5	0.0015849	0.71525	0.001	0.71066

Table 6.6: Regularization parameter obtained by the L-curve for the different number of iterations

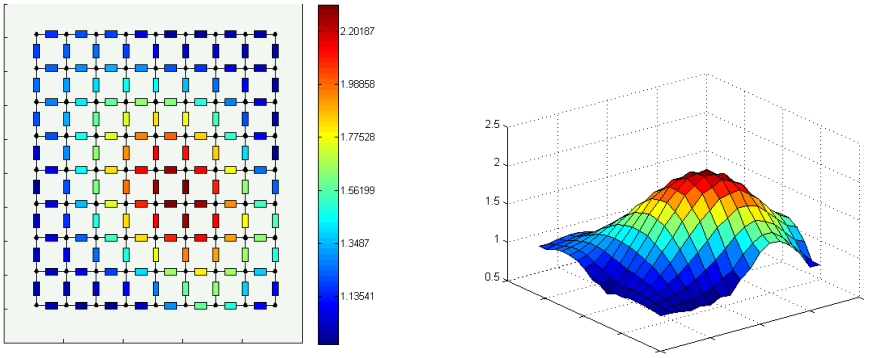


Figure 6.16: Result obtained with the L-curve parameter after 1 iteration

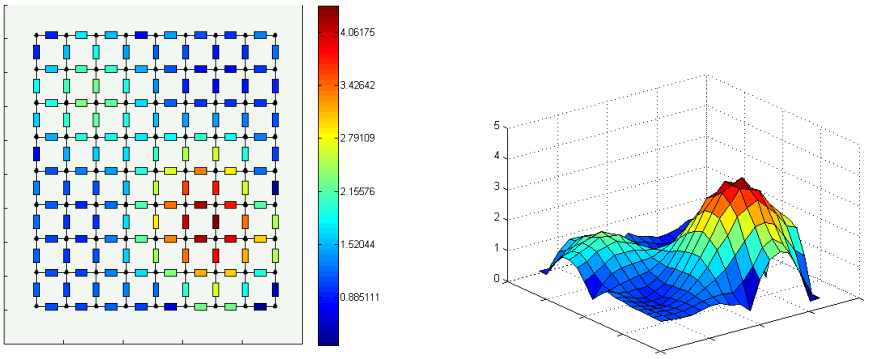


Figure 6.17: Result obtained with the L-curve parameter after 2 iterations

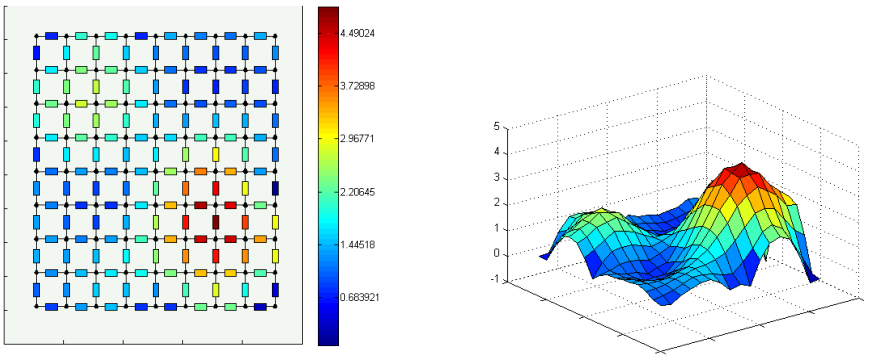


Figure 6.18: Result obtained with the L-curve parameter after 3 iterations

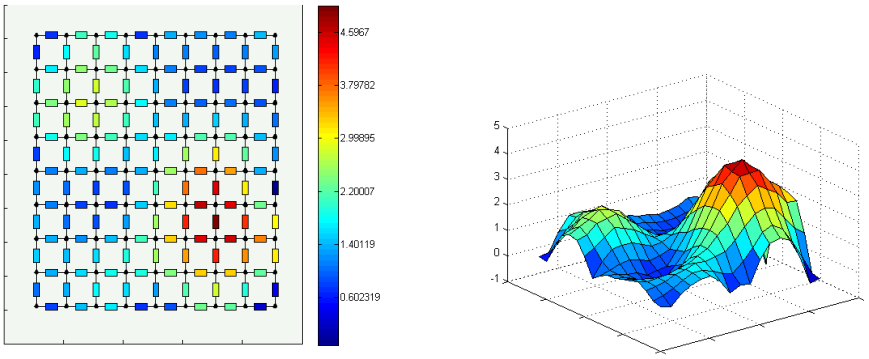


Figure 6.19: Result obtained with the L-curve parameter after 5 iterations

i	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
1	50	3.8404	2.2019	0.9221	58.5164	103.6687(1)	3.4347(5)	1
2	4	0.6423	4.0617	0.2498	41.3768	178.0676(1)	2.8848(5)	0.0100
3	5	0.5277	4.4902	-0.0773	40.4919	221.6456(1)	2.9947(5)	0.0025
5	5	0.5116	4.5967	-0.1966	40.8390	234.7525(1)	3.1071(5)	0.0016

Table 6.7: Tikhonov reconstruction of circuit 1 with fixed α-values obtained by the L-curve, 8 electrode model

Active Set Method

The box-constraints are applied as active set algorithm to the same configuration with 8 electrodes. The fixed regularization parameter obtained by the L-curve after 3 iterations is used. The upper and lower bound are selected exactly according to the true distribution. The result is shown in figure 6.20 and in table 6.8. Both bounds are reached and much more iterations are needed than in the unconstrained case. The result is considerably better than in the unconstrained case.

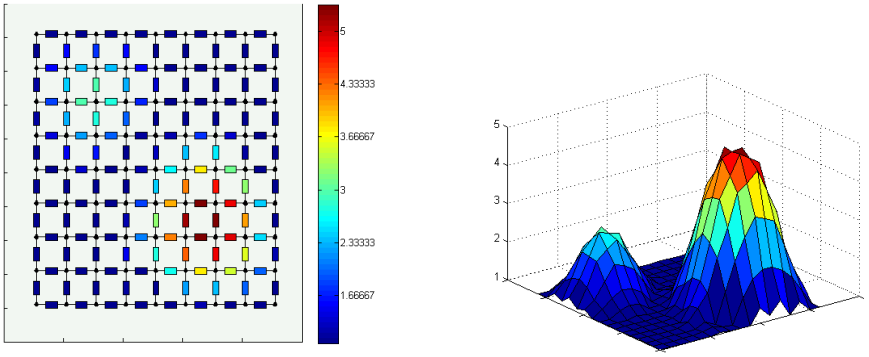


Figure 6.20: Active set result with the L-curve parameter after 3 iterations, 8 electrode model

i	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
3	26	0.7018	5	1	30.8644	141.8425(1)	2.7587(5)	0.0025

Table 6.8: Active set method, results with the L-curve parameter after 3 iterations, 8 electrode model

Simple Diminishing of α

Due to the L-curve parameter shown in table 6.6 on page 70, where the parameter are getting smaller for growing iteration count, the regularization parameter is now diminished after each iteration. Similarly to the reconstruction without noise the regularization parameter is selected with $\alpha_{k+1} = \alpha_k/2$ starting with the value $\alpha_0 = 0.1$. The result is shown in figure 6.21 on the next page for the 8 electrode model and in figure 6.22 on the following page for the 12 electrode model (see also table 6.9 on page 74). They are smoothed distributions similar to the distributions

obtained by the L-curve parameters. The result with 12 electrodes is slightly better than the result with 8 electrodes.

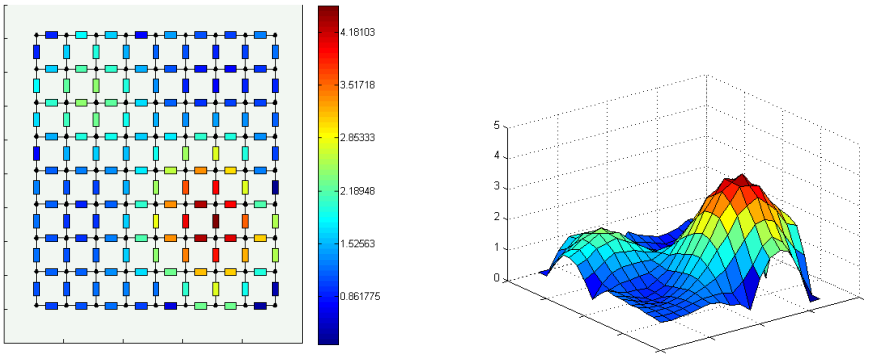


Figure 6.21: Results obtained with Tikhonov regularization and α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$, circuit 1 with 8 electrodes

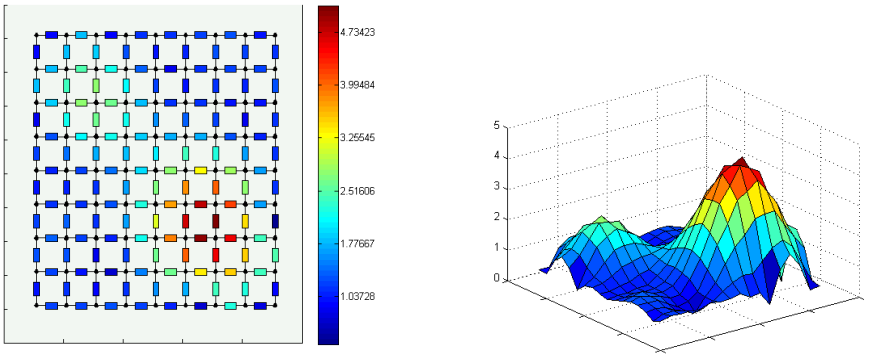


Figure 6.22: Results obtained with Tikhonov regularization and α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$, circuit 1 with 12 electrodes

GCV Regularization Parameter

Additionally the reconstruction is shown with the regularization parameter obtained by GCV for each iteration. The results are shown in figure 6.23 on the following page for the 8 electrode model and in figure 6.24 on page 75 for the 12 electrode model. The results are very similar to the results obtained by diminishing α each iteration. Each iteration the GCV-functional is calculated for 9 points which span one decade above and one decade below the actual α. (The matrix in the GCV-functional is shown for linear problems, but according to Vogel [9] it could be used for nonlinear problems, by using the Frechet-Derivative. The matrix is the same as used in Occam's inversion, it's the update-matrix which results by linearization of the system at each iteration). Therefore, a initial value is needed. It is chosen $\alpha_0 = 0.1$. The stopping criterion is identical with the one above defined.

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	5	0.6076	4.1810	0.1979	40.9486	187.7858(1)	2.8855(5)	0.0063
12	6	0.6797	4.7342	0.2979	35.5585	123.7251(1)	2.6085(5)	0.0031

Table 6.9: Results for circuit 1 obtained by Tikhonov regularization, α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$

Additionally should be stated, that the computational effort is much higher with the GCV method than with the previous method (diminishing α), but no additional information is needed, but as in the previous method the factor has to be predefined with a proper value.

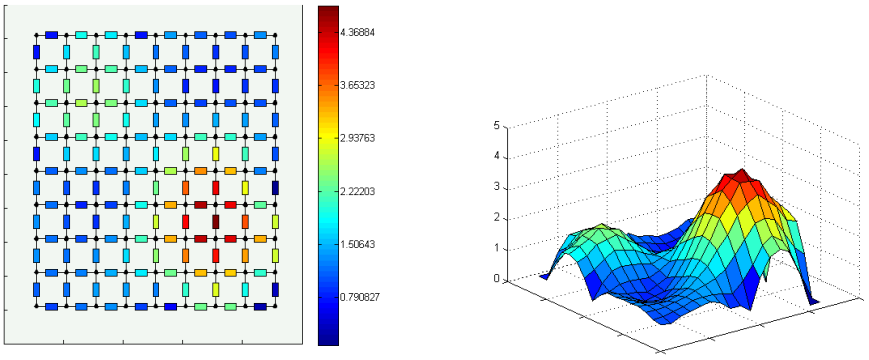


Figure 6.23: Results obtained with Tikhonov regularization, α calculated by the GCV method, circuit 1 with 8 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	6	0.5564	4.3688	0.0752	40.4389	205.3699(1)	2.8770(5)	0.0040
12	6	0.6580	4.8279	0.2641	35.0028	128.4550(1)	2.5571(5)	0.0025

Table 6.10: Results for circuit 1 obtained by Tikhonov regularization, α calculated by the GCV method, $\alpha_0 = 0.1$

Occam's Inversion

Here the regularization parameter is obtained by the discrepancy principle after each iteration. The δ value is chosen correctly. The same regularization matrix as above is used. It is the same value which is used for the stopping criterion. The number of iterations is extended, to ensure that the values converge to a solution. The results are shown in figure 6.25 on the following page for the 8 electrode model and in figure 6.26 on the next page for the 12 electrode model. Additionally the values are shown in table 6.11 on page 76. The distributions are smoother than the solutions before and the mean error is also worse, but they would be similar if removing the two additional iterations below the δ threshold.

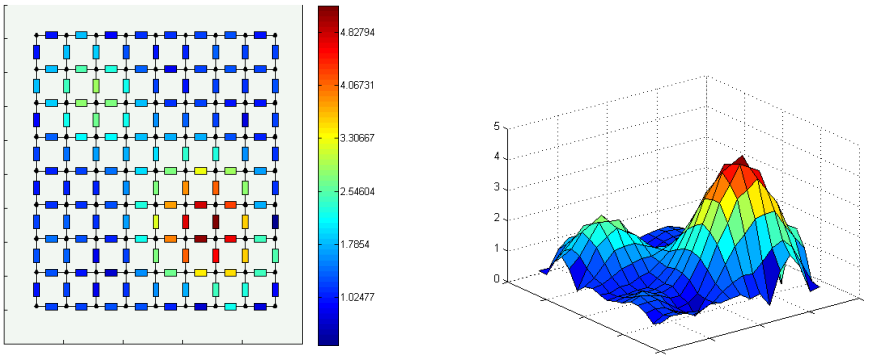


Figure 6.24: Results obtained with Tikhonov regularization, α calculated by the GCV method, circuit 1 with 12 electrodes

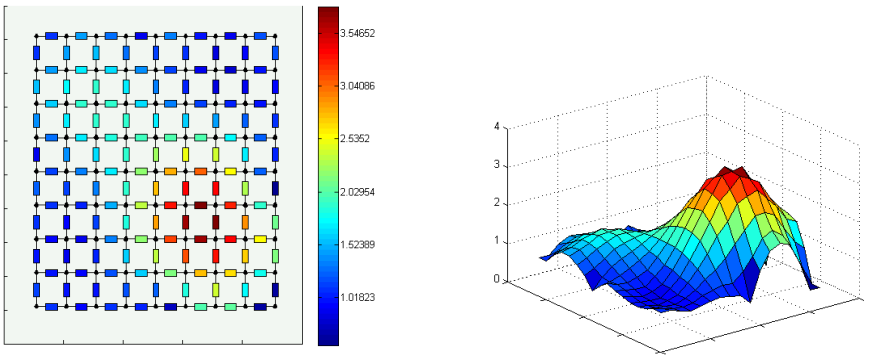


Figure 6.25: Results obtained with Occam's Inversion, α calculated by the discrepancy principle, circuit 1 with 8 electrodes

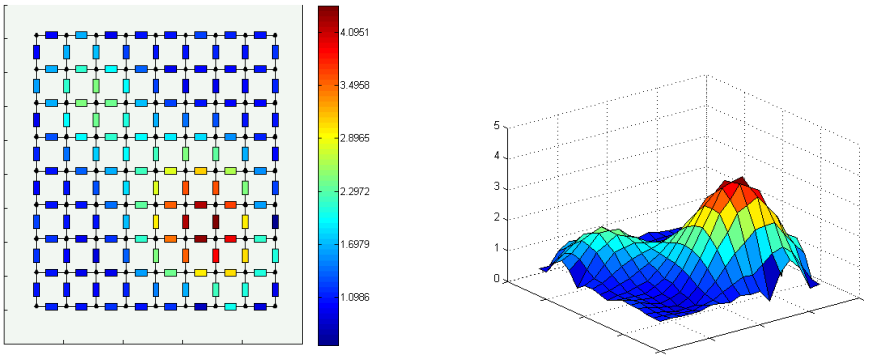


Figure 6.26: Results obtained with Occam's Inversion, α calculated by the discrepancy principle, circuit 1 with 12 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	12	0.9344	3.5465	0.5126	44.7863	138.0341(1)	2.9856(5)	0.0316
12	13	0.9654	4.0951	0.4993	39.5321	127.7857(1)	2.8063(5)	0.0200

Table 6.11: Results obtained for circuit 1 with Occam's Inversion, α calculated by the discrepancy principle

6.2.3 Total Variation Regularization

Results without Noise

The Total Variation method is implemented according the algorithm derived in chapter 5. Similarly to the Tikhonov case, first reconstruction with the measurement values without noise is performed. The regularization parameter α is also selected equally. The reconstruction with 8 electrodes is shown in figure 6.27. The corresponding values are shown in table 6.12. There the mean error is slightly worse than the error in the comparable Tikhonov reconstruction. This is is not the case with 12 electrodes (see figure 6.28 on the next page). There all reconstruction values are much better than in the Tikhonov result. The two objects are well reconstructed, also the values, especially the maximum value are well reconstructed. The maximum number of iterations is not reached. In comparison to the Tikhonov results, the TV regularization is able to reconstruct the true image better because abrupt edges are allowed.

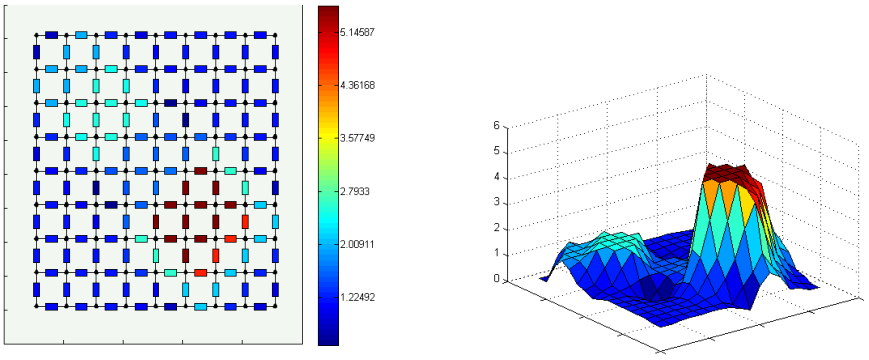


Figure 6.27: Total Variation regularization without noise, circuit 1 with 8 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	37	9.1225e-9	5.1459	0.4407	35.0090	141.2171(1)	3.4972(5)	9.4447e-16
12	43	7.0460e-9	5.0121	0.6597	10.4770	34.1380(1)	1.7069(5)	9.9035e-16

Table 6.12: Total Variation regularization without noise, circuit 1, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

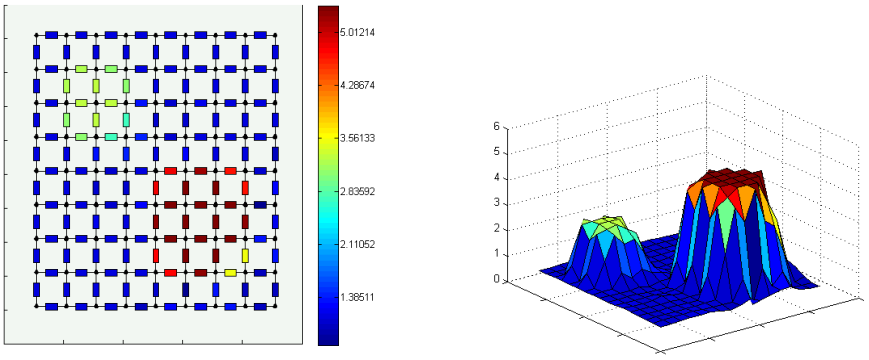


Figure 6.28: Total Variation regularization without noise, circuit 1 with 12 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

Simple Diminishing of α

The used measurement values are perturbed with noise. The regularization parameter is selected equally to the Tikhonov reconstruction with a diminishing α ($\alpha = \alpha/2$) and the starting value $\alpha_0 = 0.1$. The results are shown in figure 6.29 for the 8 electrode model and in figure 6.30 on the next page for the 12 electrode model. The positions of the two objects cannot be determined very well. The maximum values are much lower than in the Tikhonov results. Again a negative minimum resistance occurs. In comparison with the Tikhonov results the TV results with noise present are worse (see table 6.13 on the following page).

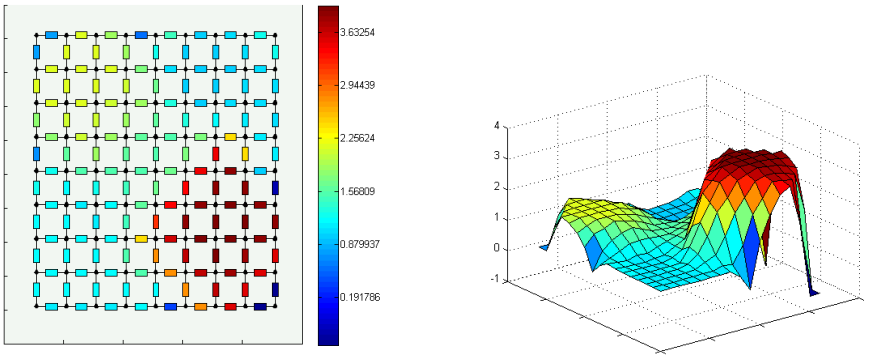


Figure 6.29: Results obtained with TV regularization, α calculated by $\alpha = \alpha/2$, $\alpha_0 = 0.1$, circuit 1 with 8 electrodes

GCV Regularization Parameter

The TV regularization is shown with the GCV-Parameter. The results are shown in figure 6.31 on the next page for the 8 electrode model and in figure 6.32 on page 79 for the 12 electrode model. The results are similar to the results of the corresponding diminishing α method. Again they are worse than the results in the Tikhonov reconstruction (see table 6.14).

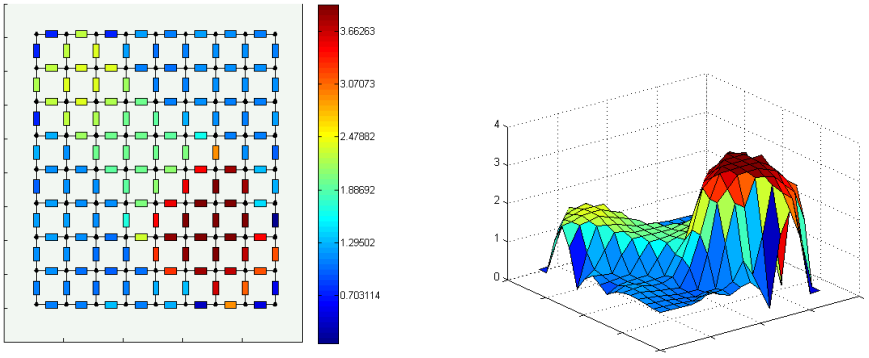


Figure 6.30: Results obtained with TV regularization, α calculated by $\alpha = \alpha/2$, $\alpha_0 = 0.1$, circuit 1 with 12 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	7	0.5639	3.6325	-0.4964	46.8929	259.1235(1)	3.5014(5)	0.0016
12	7	0.6825	3.6626	0.1112	39.7020	239.2542(1)	3.1083(5)	0.0016

Table 6.13: Resulting values for circuit 1 obtained by TV regularization, α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$

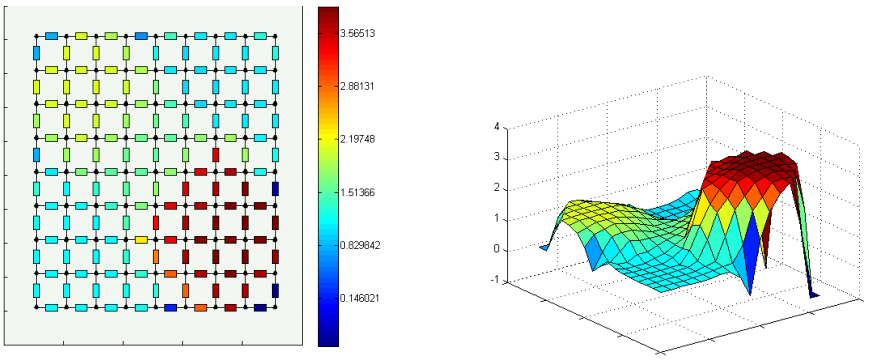


Figure 6.31: Results obtained with TV regularization, α calculated by the GCV method, $\alpha_0 = 0.1$, circuit 1 with 8 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	7	0.5704	3.5651	-0.5378	47.4969	256.4827(1)	3.4410(5)	0.0032
12	8	0.6843	3.6029	0.1413	40.4548	234.3709(1)	3.1296(5)	0.0020

Table 6.14: Resulting values for circuit 1 obtained by TV regularization, α calculated by the GCV method, $\alpha_0 = 0.1$

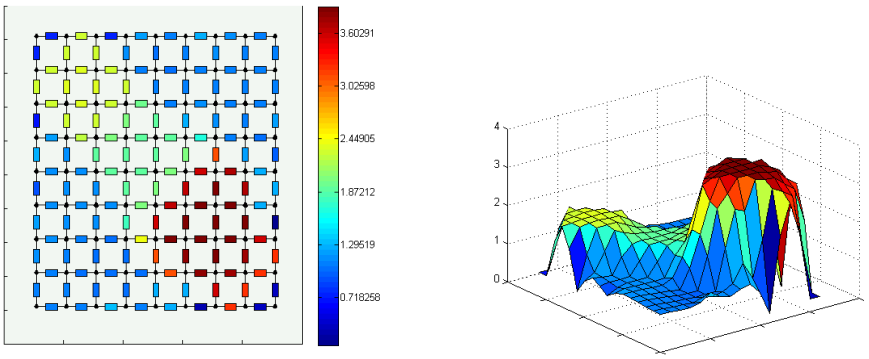


Figure 6.32: Results obtained with TV regularization, α calculated by the GCV method, $\alpha_0 = 0.1$, circuit 1 with 12 electrodes

6.2.4 Comparison of the Results for Circuit 1

The following tables show the results for the first circuit in comparison. The results are assembled for the 8 and 12 electrode model and the Tikhonov and Total Variation regularization method. Following acronyms are used for the tables:

- **WN**: Reconstruction without noise in the measurement values.
- **LC2**: The regularization parameter is selected with the L-curve method, the parameter obtained after 2 iterations is used.
- **LC3**: The regularization parameter is selected with the L-curve method, the parameter obtained after 3 iterations is used.
- **AS3**: The active set method is used. The regularization parameter is selected with the L-curve method, the parameter obtained after 3 iterations is used.
- **SDα**: The regularization parameter is obtained by simple diminishing of α each iteration.
- **GCV**: The regularization parameter is obtained by the GCV method after each iteration.
- **OI**: Occam's Inversion is applied. The regularization parameter is obtained by the discrepancy principle after each iteration.

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
WN	25	6.0236e-9	5.7904	0.5743	32.4304	157.9615(1)	2.5782(5)	1.3744e-14
LC2	4	0.6423	4.0617	0.2498	41.3768	178.0676(1)	2.8848(5)	0.0100
LC3	5	0.5277	4.4902	-0.0773	40.4919	221.6456(1)	2.9947(5)	0.0025
AS3	26	0.7018	5	1	30.8644	141.8425(1)	2.7587(5)	0.0025
SDα	5	0.6076	4.1810	0.1979	40.9486	187.7858(1)	2.8855(5)	0.0063
GCV	6	0.5564	4.3688	0.0752	40.4389	205.3699(1)	2.8770(5)	0.0040
OI	12	0.9344	3.5465	0.5126	44.7863	138.0341(1)	2.9856(5)	0.0316

Table 6.15: Reconstruction of circuit 1 with 8 electrodes, Tikhonov based regularization methods

The results obtained for the 8 electrode model and the Tikhonov regularization method are compared in table 6.15. It can be seen, that the active set method achieves the best mean error value and relative maximum error value, better than in the reconstruction without noise. This is possible because of the correct additional prior information about the true distribution in form of box constraints.

Comparing the errors of the TV regularization in table 6.16 with the corresponding errors for

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
WN	37	9.1225e-9	5.1459	0.4407	35.0090	141.2171(1)	3.4972(5)	9.4447e-16
SDα	7	0.5639	3.6325	-0.4964	46.8929	259.1235(1)	3.5014(5)	0.0016
GCV	7	0.5704	3.5651	-0.5378	47.4969	256.4827(1)	3.4410(5)	0.0032

Table 6.16: Reconstruction of circuit 1 with 8 electrodes, TV regularization method

the Tikhonov regularization in table 6.15 on the previous page the Tikhonov regularization outperforms the TV regularization, which is not obvious from the beginning.

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
WN	50	6.4363e-4	5.7312	0.2997	22.5447	165.7045(1)	1.8287(5)	1.3292e-14
SDα	6	0.6797	4.7342	0.2979	35.5585	123.7251(1)	2.6085(5)	0.0031
GCV	6	0.6580	4.8279	0.2641	35.0028	128.4550(1)	2.5571(5)	0.0025
OI	13	0.9654	4.0951	0.4993	39.5321	127.7857(1)	2.8063(5)	0.0200

Table 6.17: Reconstruction of circuit 1 with 12 electrodes, Tikhonov based regularization methods

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
WN	43	7.0460e-9	5.0121	0.6597	10.4770	34.1380(1)	1.7069(5)	9.9035e-16
SDα	7	0.6825	3.6626	0.1112	39.7020	239.2542(1)	3.1083(5)	0.0016
GCV	8	0.6843	3.6029	0.1413	40.4548	234.3709(1)	3.1296(5)	0.0020

Table 6.18: Reconstruction of circuit 1 with 12 electrodes, TV regularization method

Comparing the results for the 12 electrode model with the Tikhonov regularization (see table 6.17) and the TV regularization (see table 6.18), the TV regularization is better only in the case if no noise is present in the measurement values. For the reconstructions with noise the Tikhonov regularization method seems to be the better choice.
But it should be stated, that the used regularization parameter are not chosen on the basis of an optimality analysis. With other configurations or other initial parameters the result can differ. Nevertheless the mean error values differ not much in the comparable results, there are no unexplainable differences.

6.3 Circuit 2

6.3.1 True Distribution

The true distribution, which should be obtained, is shown in figure 6.33. It is similar to the first one, but the resistances of the objects are interchanged. Which should lead to no problems on first sight.

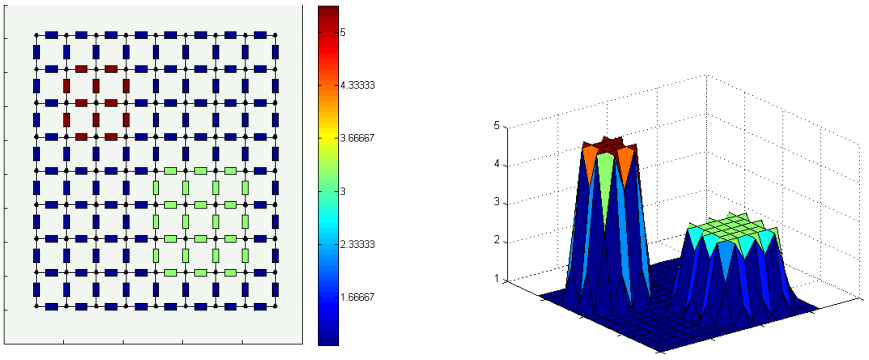


Figure 6.33: Circuit 2, true distribution

6.3.2 Tikhonov Regularization

Results Without Noise

Similarly to circuit 1, the results without noise in the measurements are obtained. The result with

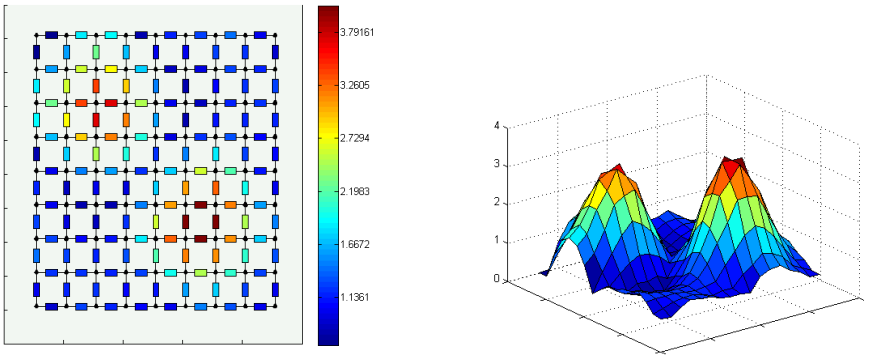


Figure 6.34: Tikhonov regularization without noise, circuit 2 with 8 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

8 electrodes is shown in figure 6.34. It differs extremely from the true distribution. The biggest errors are in the object with smaller values in the true distribution. The size of the two objects is nearly the same. Looking at the results in table 6.19 on the following page shows a very small residual norm. There must be a duality between big objects with small values and small objects

with big values.
The result with 12 electrodes is shown in figure 6.35. There the position of objects is correctly found, nevertheless the smoothness assumption is not a correct assumption and the norm is not reached (similarly to circuit 1), the result after 50 iterations is shown. Because of the failure of reconstruction with 8 electrodes, the reconstructions with noise are only shown with 12 electrodes.

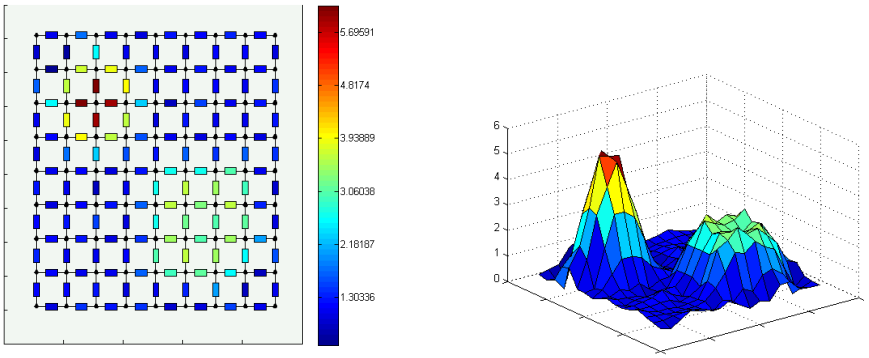


Figure 6.35: Tikhonov regularization without noise, circuit 2 with 12 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

elec	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	15	2.7420e-10	3.7916	0.6050	37.4924	130.6354(1)	2.5878(5)	3.2000e-14
12	50	2.4366e-5	5.6959	0.4249	23.8618	138.6688(1)	1.6212(5)	1.3292e-14

Table 6.19: Tikhonov regularization without noise, circuit 2, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

Results With Noise

The reconstructions for the 12 electrode model and the parameter selected with the simple diminishing of α and the GCV method are compared. Both reconstructions are very smooth and also similar in values. The GCV result is slightly better in the mean error, the maximum value and the absolute maximum error. The true maximum value is not well detected by both methods.

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
SDα	5	0.6942	3.3388	0.6052	38.1349	113.5978(1)	2.5473(5)	0.0063
GCV	6	0.6346	3.5831	0.5522	36.6462	119.5880(1)	2.4227(5)	0.0025

Table 6.20: Results obtained with Tikhonov regularization, circuit 1 with 12 electrodes, α calculated by $\alpha_{k+1} = \alpha_k/2$ and the GCV method, $\alpha_0 = 0.1$

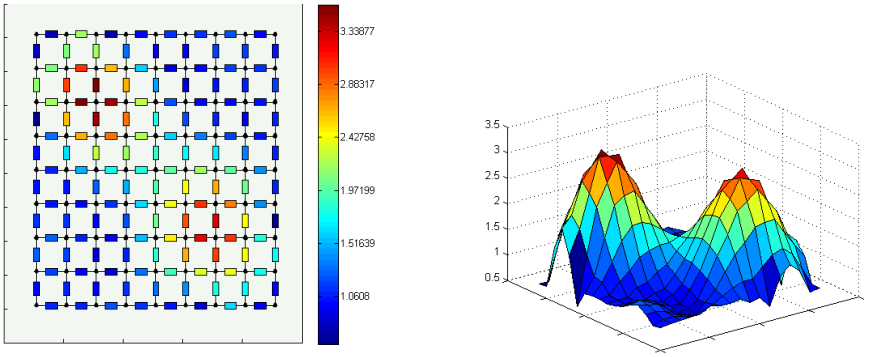


Figure 6.36: Results obtained with Tikhonov regularization, α calculated by $\alpha_{k+1} = \alpha_k/2$ and $\alpha_0 = 0.1$, circuit 2 with 12 electrodes

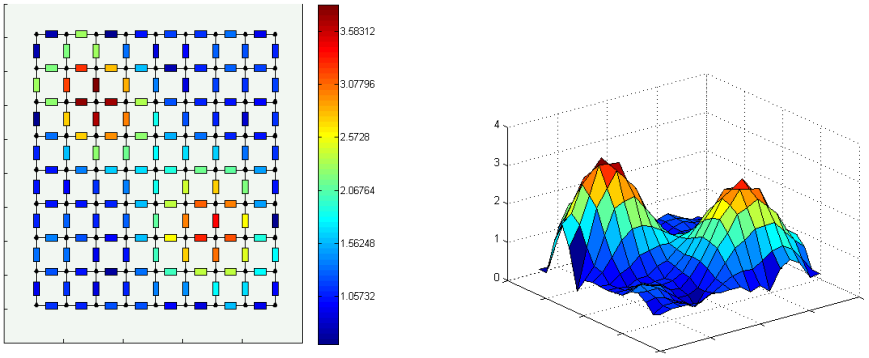


Figure 6.37: Results obtained with Tikhonov regularization, α calculated by the GCV method, $\alpha_0 = 0.1$, circuit 2 with 12 electrodes

6.3.3 Total Variation Regularization

Results Without Noise

The Total Variation Reconstruction for the 8 electrode model is shown in figure 6.38. It suffers from the same problem as in the Tikhonov regularization method. The reconstruction fails, the larger values are in the big object which has smaller values in the true distribution, and these values are too large. On the contrary, reconstruction with 12 electrodes is working almost perfectly. The corresponding values are shown in table 6.21. Similarly to the Tikhonov regularization method the reconstructions with noise are only shown with 12 electrodes.

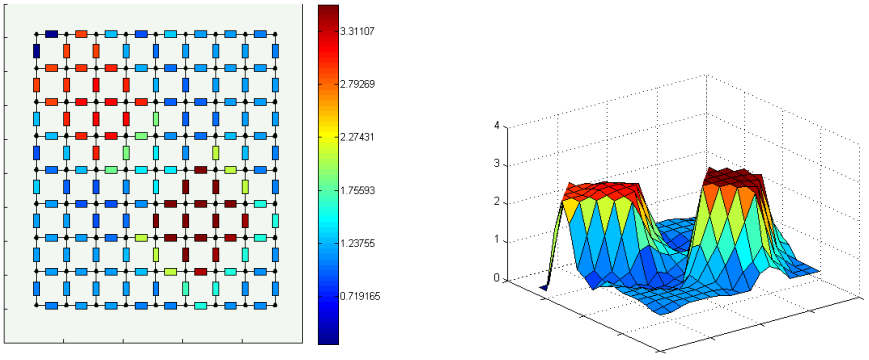


Figure 6.38: TV regularization without noise, circuit 2 with 8 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

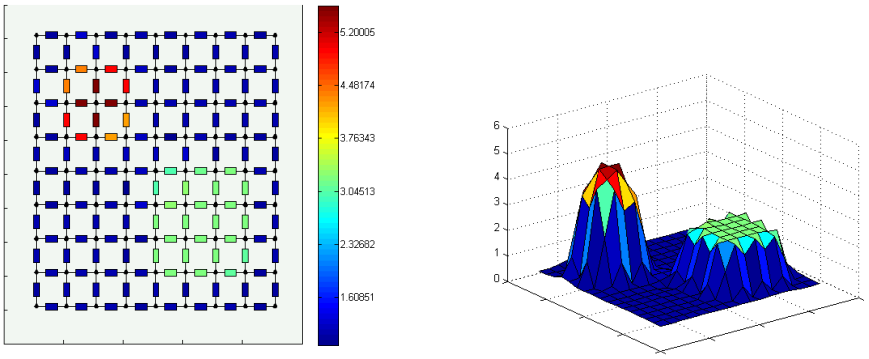


Figure 6.39: TV regularization without noise, circuit 2 with 12 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

elec	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
8	32	8.5463e-9	3.3111	0.2008	40.8422	180.2822(1)	2.2539(5)	2.8823e-15
12	41	4.1438e-9	5.2000	0.8902	8.6880	20.3373(5)	1.0169(5)	3.0949e-15

Table 6.21: TV regularization without noise, circuit 2, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

Results With Noise

The reconstruction values with noise present in the measurement values are shown with the GCV-Parameter and the simple diminishing of α (see figure 6.40 and figure 6.41). The simple diminishing of α leads to slightly better results than the GCV method, but they are both worse than the comparable Tikhonov solutions. Similarly to the Tikhonov results the maximum value is not well detected. The resulting values are shown in table 6.22 on the following page.

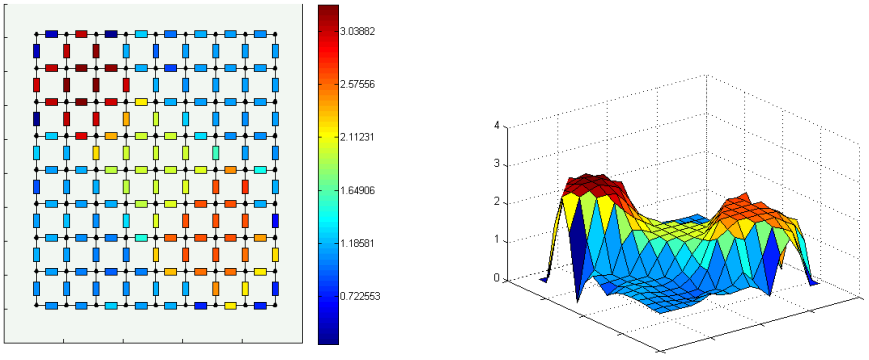


Figure 6.40: Results obtained with TV regularization, α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$, circuit 2 with 12 electrodes

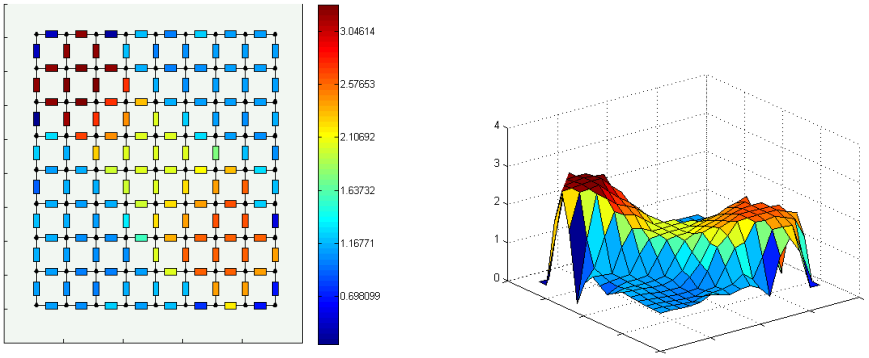


Figure 6.41: Results obtained with TV regularization, α calculated by the GCV method, $\alpha_0 = 0.1$, circuit 2 with 12 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
SDα	7	0.6596	3.0388	0.2593	42.7105	198.5031(1)	2.7910(5)	0.0016
GCV	7	0.6618	3.0461	0.2285	44.5333	198.6251(1)	2.6927(5)	0.0020

Table 6.22: Results obtained with TV regularization, circuit 2 with 12 electrodes, α calculated by $\alpha_{k+1} = \alpha_k/2$ and the GCV method, $\alpha_0 = 0.1$

6.4 Circuit 3

6.4.1 True Distribution

The true distribution of circuit 3 is shown in figure 6.42. It has only one big object in the center.

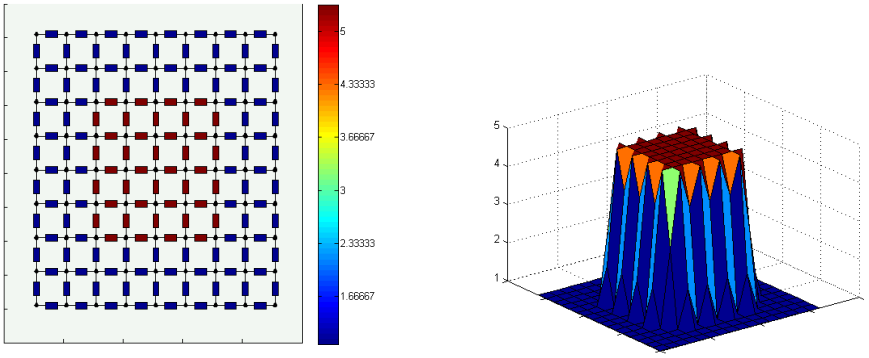


Figure 6.42: Circuit 3, true distribution

6.4.2 Tikhonov Regularization Applied to the 8 Electrode Model

The reconstruction of the 8 electrode model without noise present is shown in figure 6.43. It is a very smooth result. The maximum values are much higher than the real values, due to the smoothness assumption. The reconstruction with 12 electrodes, looks much better, but the smoothness assumption prevents the correct reconstruction. In this case an additional averaging of high and low values could give a good result. In figure 6.44 on the next page the result with noise present is shown, where the diminishing α regularization parameter is used. In figure 6.45 on the following page the corresponding result with the GCV regularization parameter is shown. Both results are similar and very smooth. The resulting values are shown in table 6.23 on the next page.

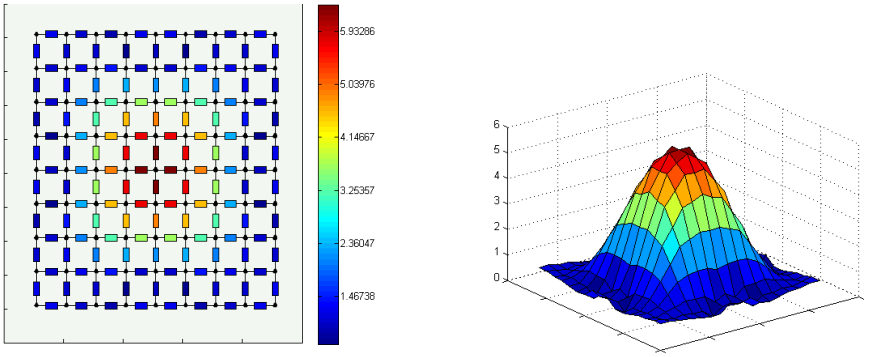


Figure 6.43: Tikhonov regularization without noise, circuit 3 with 8 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

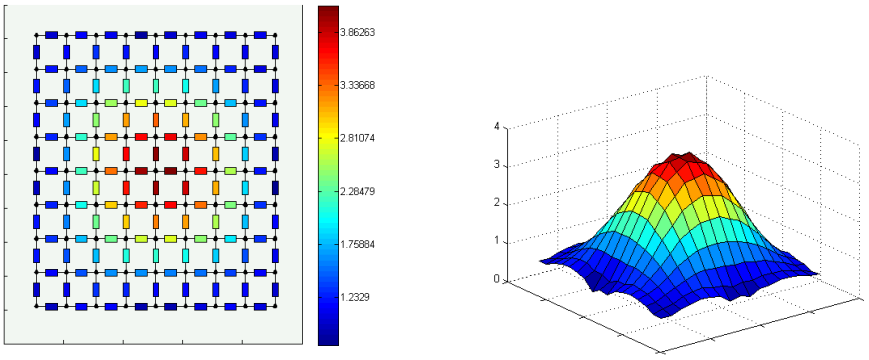


Figure 6.44: Results obtained with Tikhonov regularization, α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$, circuit 3 with 8 electrodes

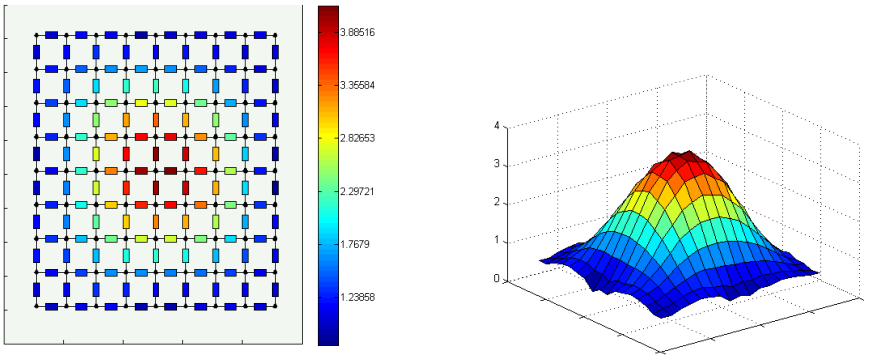


Figure 6.45: Results obtained with Tikhonov regularization, α calculated by the GCV method, $\alpha_0 = 0.1$, circuit 3 with 8 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
WN	24	3.2642e-9	5.9329	0.5743	27.6488	107.1360(1)	2.0066(5)	1.7180e-14
SDα	4	0.5140	3.8626	0.7070	42.4409	139.3554(1)	2.8085(5)	0.0125
GCV	4	0.5138	3.8852	0.7093	42.3148	141.1383(1)	2.8127(5)	0.0100

Table 6.23: Tikhonov regularization, circuit 3 with 8 electrodes

6.4.3 Tikhonov Regularization Applied to the 12 Electrode Model

The results obtained by the Tikhonov regularization method and without noise present are shown in figure 6.46. The true distribution is well obtained, but the smoothness assumption prevents the exact reconstruction. Similarly to circuit 1 and 2 the stopping criterion is not reached (see table 6.24 on the next page). Additionally the reconstructions are performed with noise present.

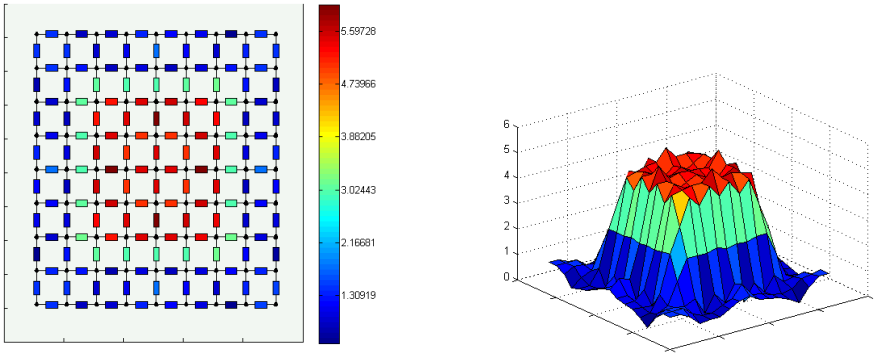


Figure 6.46: Tikhonov regularization without noise, circuit 2 with 12 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

The results are shown in figure 6.47 for the diminishing α and in figure 6.48 on the next page for the GCV parameter. The results are also shown in table 6.24 on the following page. They are slightly better for the GCV parameter, especially the maximum resistance is much higher, but both distributions are very smooth.

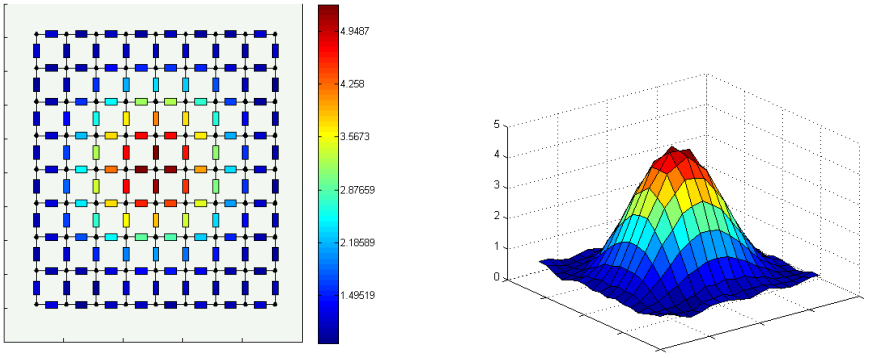


Figure 6.47: Results obtained with Tikhonov regularization, α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$, circuit 3 with 12 electrodes

6.4.4 Total Variation Regularization Applied to the 8 Electrode Model

Also the TV regularization method is applied to the 8 electrode model. The result without noise present is shown in figure 6.49 on the next page. There the shape of the object is not reconstructed correctly, but the maximum resistance is very close to the true value. Also the reconstructions

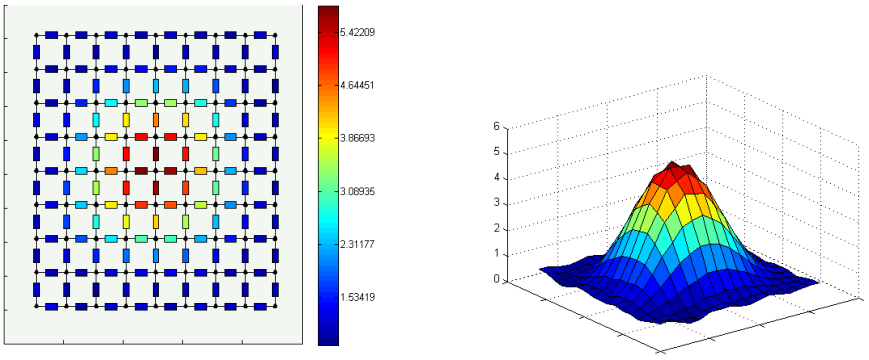


Figure 6.48: Results obtained with Tikhonov regularization, α calculated by the GCV method, $\alpha_0 = 0.1$, circuit 3 with 12 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
WN	50	1.3929e-4	5.5973	0.4516	25.6929	190.1507(1)	1.9015(1)	6.6461e-15
SDα	5	0.7358	4.9487	0.8045	35.5617	134.7627(1)	2.8957(5)	0.0063
GCV	6	0.6761	5.4221	0.7566	33.3848	128.6227(1)	2.8957(5)	0.0032

Table 6.24: Tikhonov regularization, circuit 3 with 12 electrodes

with noise present are shown similar to the Tikhonov reconstrution (see figure 6.50 on the following page and figure 6.51 on the next page). The corresponding values are shown in table 6.25 on the following page. The result without noise present is better than the corresponding result obtained by the Tikhonov regularization. Both reconstructions with noise present are similar, but they are worse than the values obtained with the Tikhonov regularization method.

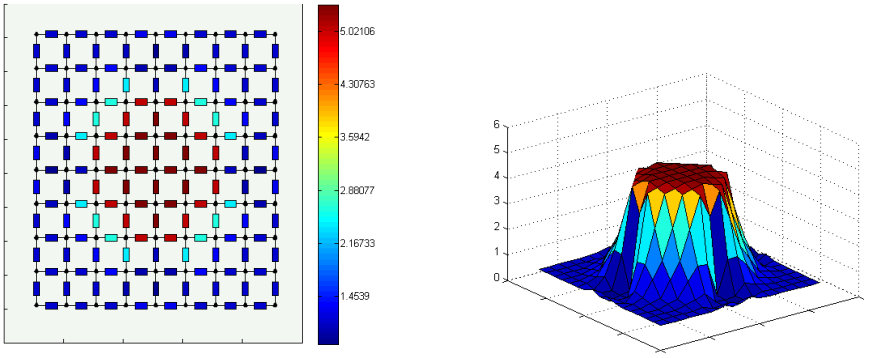


Figure 6.49: TV regularization without noise, circuit 3 with 8 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

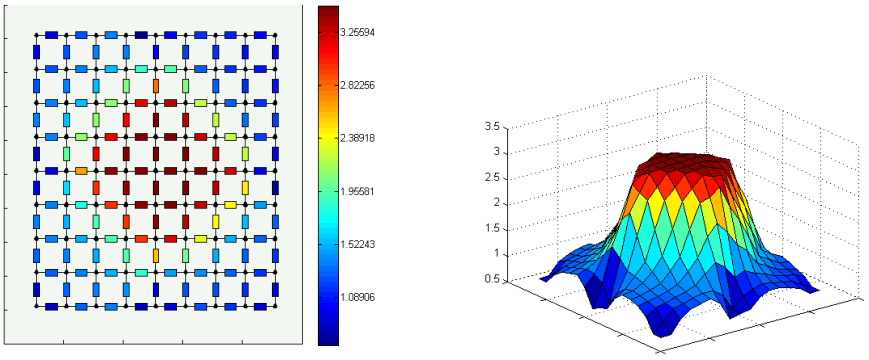


Figure 6.50: Results obtained with TV regularization, α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$, circuit 3 with 8 electrodes

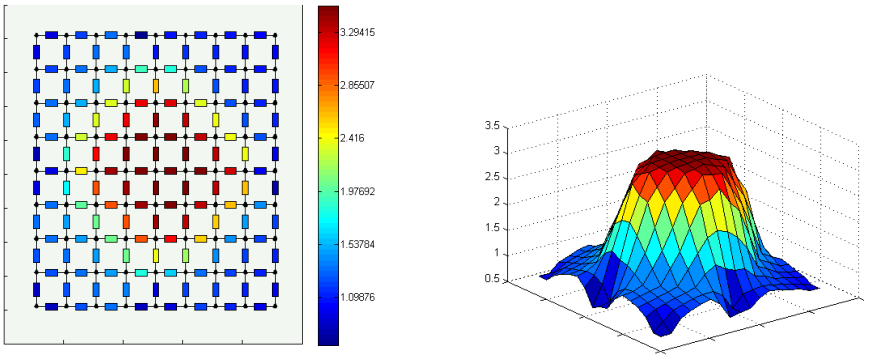


Figure 6.51: Results obtained with TV regularization, α calculated by the GCV method, $\alpha_0 = 0.1$, circuit 3 with 8 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
WN	50	1.1513e-8	5.0211	0.7405	24.6335	126.8274(1)	2.5625(5)	3.3231e-15
SDα	6	0.5167	3.2559	0.6557	44.0817	205.4313(1)	3.1775(5)	0.0031
GCV	5	0.5061	3.2942	0.6597	43.3115	208.7100(1)	3.0857(5)	0.0040

Table 6.25: TV regularization, circuit 3 with 8 electrodes

6.4.5 Total Variation Regularization Applied to the 12 Electrode Model

The obtained reconstructions with the TV method applied to the 12 electrode model are shown. The result without noise present (see figure 6.52) is exactly the true distribution. Additionally the reconstructions with noise present are shown in figure 6.53 and figure 6.54 on the next page. They are similar again, but the GCV parameter result is slightly better. The results are shown in table 6.26 on the following page. In this case the mean errors are better than errors of the corresponding Tikhonov results.

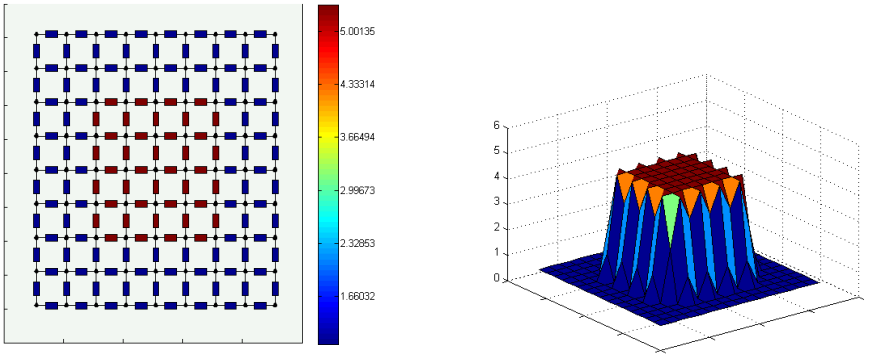


Figure 6.52: TV regularization without noise, circuit 3 with 12 electrodes, $\alpha_{k+1} = \alpha_k/10$ and $\alpha_0 = 0.1$

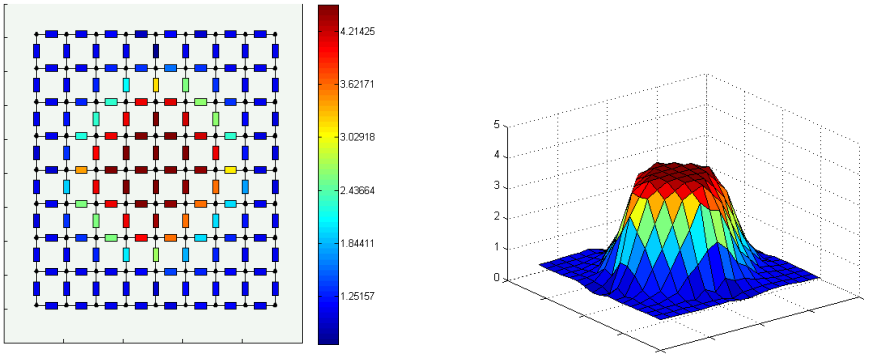


Figure 6.53: Results obtained with TV regularization, α calculated by $\alpha_{k+1} = \alpha_k/2$, $\alpha_0 = 0.1$, circuit 3 with 12 electrodes

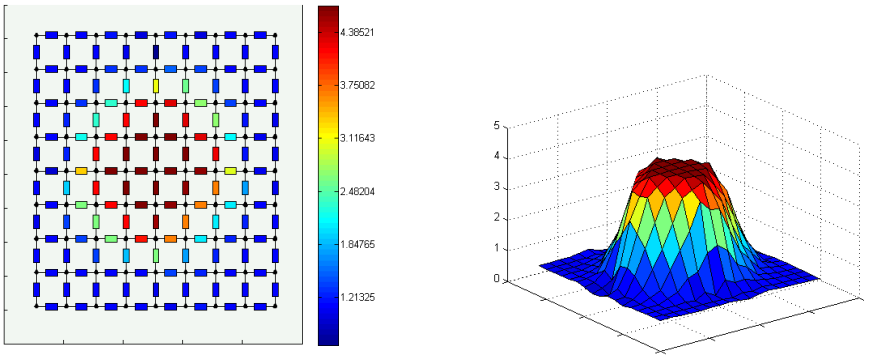


Figure 6.54: Results obtained with TV regularization, α calculated by the GCV method, $\alpha_0 = 0.1$, circuit 3 with 12 electrodes

e	k	$\|\|\mathbf{u}(\mathbf{r}) - \mathbf{u}_r\|\|_2^2$ mV^2	R_{max} Ω	R_{min} Ω	ϵ_{mean} %	$\epsilon_{r,max}$ %	$\epsilon_{a,max}$ Ω	α_k
WN	15	1.4736e-9	5.0013	0.9921	0.1973	0.7884(1)	0.0222(5)	1.0000e-15
SDα	7	0.7243	4.2142	0.6590	33.8517	221.6411(1)	3.1533(5)	0.0016
GCV	6	0.6796	4.3852	0.5789	32.2658	209.0032(1)	3.1137(5)	0.0032

Table 6.26: TV regularization, circuit 3 with 12 electrodes

7 Conclusion

In this thesis different methods to solve inverse problems were introduced and their main features were discussed. These methods were applied to a linear and a nonlinear problem. It was shown that reconstruction is much more complex in the nonlinear case. In the linear case just a few iterations (sometimes even a single iteration) are required, while nonlinear problems are successively linearized and hence iteratively solved. It was shown that there is not "the procedure" to solve each individual problem. On the contrary, one has to find a more or less tailor-made algorithm for the problem at hand. When doing so, several criteria have to be taken into account. It makes a big difference, how time consuming the underlying forward problem is or whether the Hessian matrix can be calculated with reasonable effort or is better approximated in some way or another. A main feature of each inverse algorithm is the a kind of regularization, which is essential in the process of reconstruction and the way, the regularization parameter is computed. It can be done iteratively or by fixing this parameter after a suitable number of iterations. In any case the type of regularization has to be selected carefully in order to represent the distribution of the sought data (material values etc.) as correctly as possible. Furthermore, it makes a big difference whether the results are to be obtained in real time or not. A big issue is the distribution of the measured data (which are to be simulated correctly by the underlying forward problem) and the signal to noise ratio. It has to be investigated how sensitive the selected algorithms reacts to variable signal to noise ratios. A vital question, which has to answered before starting the reconstruction process, is the number of measured data, the amount of information, which has to be supplied in order to get meaningful results. It was shown with the help of the second circuit, that a number of 8 measuring electrodes does not give feasible results at all (even without any noise), while grading up to 12 electrodes improves the results substantially, even though the resulting system of equations is still massively under determined. Having the problems investigated in mind it can be concluded that reconstructing continuous areas of different features with a sufficiently high number of measured data and a sufficiently low signal to noise ratio the Total Variation methods is a good choice for the regularization term. Nevertheless, if the position of some continuous area is more important than its exact dimensions, Tikhonov regularization is also a promising method, even more so, if some a priori knowledge about the magnitude of the values of the continuous is available.

Bibliography

[1] P. C. Hansen. Regularization tools. Technical report, Department of Mathematical Modelling, Technical University of Denmark, 2800 Lyngby, Denmark, June 1992, Last revision September 2001.

[2] The MathWorks, Inc. www.mathworks.com. *MATLAB Software, Version R2009b.*

[3] B. Peikari. *Fundamentals of network analysis and synthesis.* Prentice-Hall, 1974.

[4] J. W. Nilsson and S. A. Riedel. *Electric Circuits.* Pearson Prentice Hall, 2005.

[5] R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter Estimation and Inverse Problems.* Academic Press (Elsevier), 2013.

[6] L M Heikkinen, M Vauhkonen, T Savolainen, and J P Kaipio. Modelling of internal structures and electrodes in electrical process tomography. *Meas. Sci. Technol.*, 12:1012–1019, 2001.

[7] B. Brandstätter. Fast reconstruction of beams in intense proton accelerators. *IEEE Transactions on Magnetics*, 38 NO 2:1113–1116, 2002.

[8] H. Yan, F.Q. Shao, and S.Wang. Fast calculation of sensitivity distributions in capacitance tomography sensors. *Electronic Letters*, 34 No 20:1936–1937, 1998.

[9] C. R. Vogel. *Computational Methods for Inverse Problems.* SIAM Society for Industrial and Applied Mathematics, 2002.

[10] A. Neubauer H. W. Engl, M. Hanke. *Regularization of Inverse Problems.* Kluwer Academic Publishers, 2000.

[11] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, 1990.

[12] M. H. Wright P. E. Gill, W. Murray. *Practical Optimization.* Academic Press, 1981.

[13] A. Borsic. *Regularization Methods for Imaging from Electrical Measurements.* PhD thesis, Oxford Brookes University, 2002.

[14] P.Neittaanmäki, M.Rudnicki, and A.Savini. *Inverse Problems and Optimal Design in Electricity and Magnetism.* Oxford University Press, 1996.

[15] U. Baumgartner, T. Ebner, and C. Magele. *Optimization in Electrical Engineering, Skriptum der Vorlesung Numerische Optimierungsverfahren.* Technical University of Graz, Institute for Fundamentals and Theory in Electrical Engineering, 2006.

[16] B. Brandstätter. *Unterlagen und Skriptum Vorlesung und Rechenübung Inverse Probleme.* 2006.

[17] A. Missner. 3d-magnetic induction tomography: Single-step and iterative image reconstructions with different regularization methods. Master's thesis, Graz University of Technology, 2006.

[18] M. Al-Baali and R. Fletcher. An efficient line search for nonlinear least squares. *Journal of Optimization Theory and Applications*, 48 No. 3:359–377, 1986.

[19] A.K. Louis. *Inverse und schlecht gestellte Probleme.* Teubner Studienbücher Mathematik, 1989.

[20] H. W. Engl and W. Grever. Using the l-curve for determining optimal regularization parameters. *Numer. Math.*, 69:25–31, 1994.

[21] C.R. Vogel. Non-convergence of the l-curve regularization parameter. *Inverse Problems*, 12:535–547, 1996.

[22] D. Watzenig, B. Brandstätter, and G. Holler. Adaptive regularization parameter adjustment for reconstruction problems. *IEEE Transactions on Magnetics*, 40. NO. 2:1116–1119, 2004.

[23] M. Vauhkonen, D. Vadasz, P. A. Karjalainen, E. Somersalo, and J. P. Kaipio. Tikhonov regularization and prior information in electrical impedance tomography. *IEEE Transactions on Medical Imaging*, 17 NO. 2:285–293, 1998.

[24] V. Kolehmainen, M. Vauhkonen, P.A. Karjalainen, and J.P. Kaipio. Spatial inhomogeneity and regularization in eit. *Proceedings - 19th International Conference - IEEE/EMBS*, 19:449–452, 1997.

[25] M Vauhkonen, J P Kaipio, E Somersalo, and P A Karjalainen. Electrical impedance tomography with basis constraints. *Inverse Problems*, 13:523–530, 1997.

[26] A.Borsic, W. R. B. Lionheart, and C. N. McLeod. Generation of anisotropic-smoothness regularization filters for eit. *IEEE Transactions on Medical Imaging*, 21 No. 6:579–587, 2002.

[27] K. Y. Kim, B.S.Kim, M. C. Kim, S. Kim, Y. J. Lee, H.J. Jeon, B. Y. Choi, and M. Vauhkonen. Electrical impedance imaging of two-phase fields with an adaptive mesh grouping scheme. *IEEE Transactions on Magnetics*, 40 No 2:1124–1127, 2004.

[28] L.M. Heikkinen, M. Vaukhonen, T. Savolainen, K. Leinonen, and J.P.Kaipio. Electrical process tomography with known internal structures and resistivities. Technical report, University of Kuopio, Department of Applied Physics, 2000.

[29] M. Glidewell and K. T. Ng. Anatomically constrained electrical impedance tomography for anisotropic bodies via a two-step approach. *IEEE Transactions on medical Imaging*, 14, NO 3:498–503, 1995.

[30] E. Kreyszig. *Advanced Mathematics.* John Wiley & Sons, 2006.

[31] J. Li. *Multifrequente Impedanztomographie zur Darstellung der elektrischen Impedanzverteilung im menschlichen Thorax.* PhD thesis, Universität Stuttgart, 2000.

[32] G. Holler, D. Watzenig, and B. Brandstätter. A fast gauss-newton based ect algorithm with automatic adjustment of the regularization parameter. *3rd World Congress on Industrial Process Tomography, Banff, Canada*, 3:415–420.

[33] G. Rodriguez and D. Theis. An algorithm for estimating the optimal regularization parameter by the l-curve. *Rendiconti di Matematica*, 25:69–84, 2005.

[34] J.P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo. Construction of nonstandard smoothness priors. Technical report, University of Kuopio Department of Applied Physics, 1998.

[35] P. Hua, E.J. Woo, J. G. Webster, and W. J. Tompkins. Iterative reconstruction methods using regularization and optimal current patterns in electrical impedance tomography. *IEEE Transactions on Medical Imaging*, 10 No. 4:621–628, 1991.

[36] U. Baysal and B. M.Eyüboglu. Use of a priori information in estimating tissue resistivities-a simulation study. *Phys. Med. Biol.*, 43:3589–3606, 1998.

[37] U. Baysal and B. M. Eyüboglu. Use of a priori information in estimating tissue resistivities-application to measured data. *Phys. Med. Biol.*, 44:1677–1689, 1999.