

Inversion Methods and Resolution Analysis for the 2D/3D Reconstruction of Resistivity Structures from DC Measurements

Von der Fakultät für Geowissenschaften, Geotechnik und Bergbau
der Technischen Universität Bergakademie Freiberg

genehmigte

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium

Dr. rer. nat.

vorgelegt

von Diplom-Geophysiker Thomas Günther

geboren am 03. Februar 1975 in Marienberg

Gutachter Prof. Dr. Klaus Spitzer, Bergakademie Freiberg
 Prof. a.D. Dr. Peter Weidelt, TU Braunschweig
 Dr. Maxwell A. Meju, Lancaster University

Tag der Verleihung: 03. Dezember 2004

Abstract

The presented thesis deals with the multi-dimensional reconstruction of the earth's conductivity distribution based on DC resistivity data. This task represents a non-linear and ill-posed minimization problem with many degrees of freedom. In this work, techniques for regularization and controlling of this problem are depicted and classified. Particularly, it is concentrated on explicit regularization types, which impose constraints onto the model. The system of equations as resulting from the application of the Gauss-Newton minimization can be solved efficiently. Furthermore, it is shown how the regularization strength can be controlled.

The method of non-linear resolution analysis plays a central role in the thesis. It represents a powerful tool to estimate the quality of inversion results. Furthermore, the derived resolution measures provide the basis for the optimization of experimental design concerning information content and efficiency.

Methods of error estimation, forward modeling and the calculation of the Jacobian matrix for DC resistivity data are developed. Procedures for appropriate parameterization and inversion control are pointed out by studies of synthetic models. Different inversion and regularization methods are examined in detail. A linearized study is used to compare different data sets considering their efficiency. Moreover, a triple-grid-technique for the incorporation of topography into three-dimensional inversion is presented.

Finally the inversion methods are applied to field data. The depicted optimization strategies are realized in practice, which increases the economic relevance of three-dimensional data acquisition. The structure of the subsurface is imaged in detail for several applications in the fields of cavity detection, archaeology and the investigation of ground falls. The resolution analysis is successfully established to appraise the obtained results.

Kurzfassung

In der vorliegenden Arbeit wird die mehrdimensionale Rekonstruktion der Leitfähigkeitsverteilung im Untergrund ausgehend von geoelektrischen Messungen verfolgt. Das stellt in der Regel ein nichtlineares, schlecht-gestelltes Minimierungsproblem mit vielen Freiheitsgraden dar. Es werden Verfahren vorgestellt und klassifiziert, die dieses Problem regularisieren und steuern können. Im besonderen konzentriert sich die Arbeit auf explizite Regularisierungsverfahren, in denen man zusätzliche Forderungen an das Modell stellt. Es wird gezeigt, wie das bei der Anwendung des Gauss-Newton-Verfahrens entstehende Gleichungssystem effizient gelöst und die Stärke der Regularisierung optimiert werden kann.

Ausgehend von der Minimierungsstrategie wird die Methode der nichtlinearen Auflösungsanalyse herausgearbeitet. Mit Hilfe dieser kann zum einen die Qualität erhaltener Ergebnisse eingeschätzt werden. Zum anderen liefern die abgeleiteten Auflösungsmaße die Basis zur Optimierung des Experimentaldesign hinsichtlich des Informationsgehaltes.

Für den Fall der Geoelektrik werden Verfahren zur Fehlerschätzung, zur Simulation von Messungen mit Finiten Differenzen und zur Berechnung der benötigten Jacobi-Matrix weiterentwickelt. Anhand von synthetischen Studien werden Verfahrensweisen zur Parametrisierung und Steuerung des inversen Problems herausgearbeitet. Verschiedene Regularisierungs- und Parametrisierungsverfahren werden auf den Prüfstand gestellt. Eine linearisierte Studie vergleicht dabei verschiedene Datensätze hinsichtlich ihrer Effizienz. Weiterhin wird an einem Drei-Gitter-Verfahren gezeigt, wie die dreidimensionale Inversion eine vorhandene Topographie einbinden kann.

Schließlich werden die Inversionsverfahren auf Felddaten angewandt. Dabei können die herausgearbeiteten Optimierungsstrategien unter praktischen Bedingungen umgesetzt werden, wodurch dreidimensionale Messungen ökonomische Relevanz erhalten. Für verschiedene Anwendungen in der Hohlräumeuche, Archäologie und bei der Untersuchung von Erdfällen können detaillierte Rückschlüsse auf die Struktur des Untergrundes geschlossen werden. Die Auflösungsanalyse wird dabei zur kritischen Bewertung der Ergebnisse erfolgreich eingesetzt.

Contents

Contents	I
Used Symbols	VI
1 Introduction	1
1.1 DC Resistivity Methods	1
1.2 Modeling, Inversion and Resolution	2
1.3 Guideline of the Thesis	3
2 Non-linear Inversion and Resolution	7
2.1 Inversion Methods	8
2.1.1 Inversion Methodology	8
2.1.2 The Steepest Descent Method	10
2.1.3 The Nonlinear Conjugate Gradients Method	11
2.1.4 Newton Type Methods	12
2.2 Regularization	14
2.2.1 Regularization Types	14
2.2.2 Explicit Regularization	17
2.2.3 Implementation	18
2.2.4 Choosing the Regularization Parameter	21
2.3 Equation Solvers for Inverse Problems	25
2.3.1 Fixed Regularization Strength Solvers	26
2.3.2 Multiple Regularization Strength Solvers	29
2.4 SVD and GSVD	31
2.4.1 The Singular Value Decomposition	31
2.4.2 The Generalized SVD	34
2.5 Resolution Analysis	36
2.5.1 Resolution of Linear Problems	36
2.5.2 Resolution of Non-linear Problems	37
2.5.3 Interpretation of Resolution Measures	39
2.5.4 Resolution in Terms of SVD/GSVD	40
2.5.5 Resolution Approximation	41

2.6	Conclusions	41
3	Inversion of DC Resistivity Data	43
3.1	Fundamentals of DC Resistivity	43
3.2	Errors of DC Measurements	48
3.2.1	Geometrical Errors	48
3.2.2	Voltage Errors	51
3.2.3	Numerical Implementation	51
3.3	Parameterization and Regularization	53
3.3.1	The Parameter Function	53
3.3.2	Geometrical Parameter Arrangement	54
3.4	DC Forward Calculation	57
3.4.1	Finite Difference Discretization	57
3.4.2	Singularity Removal Technique	61
3.4.3	Accuracy and Speed	62
3.4.4	Finite Element Modeling	68
3.4.5	Forward Approximations	69
3.4.6	2D Finite Difference Modeling	69
3.5	Calculation of Sensitivities	70
3.5.1	Derivation of DC Sensitivities	71
3.5.2	Sensitivities for Inhomogeneous Conductivities	72
3.5.3	Secondary Potential Sensitivity Calculation	74
3.5.4	Sensitivities for Logarithms	75
3.5.5	Example	76
3.6	Conclusions	77
4	Inversion and Resolution Studies	79
4.1	Two-dimensional Inversion Studies	80
4.1.1	Numerical Setup	80
4.1.2	Inversion Methods and Resolution Analysis	81
4.1.3	Comparative Linearized Resolution Study	87
4.1.4	Consequences of Electrode Errors	94
4.2	Three-dimensional Inversion Studies	95
4.2.1	The 11x11 Electrodes Grid	96
4.2.2	The Parallel Profiles	99
4.3	3D-Inversion Including Topography	101
4.3.1	Motivation	101
4.3.2	The Synthetic Model	102
4.3.3	The Three Grids of Inversion	102

4.3.4	Inversion Method	104
4.3.5	Inversion Results	106
4.4	Conclusions	108
5	Application to field data	111
5.1	The Rothsönberg Gallery	111
5.1.1	2D-Inversion and the Global L-curve	111
5.1.2	Optimization of Profile Data	114
5.1.3	3D-Inversion	117
5.2	Ground Fall Problems on Salt-waste Dumps	122
5.2.1	Motivation	122
5.2.2	Two-dimensional Investigation	123
5.2.3	Three-dimensional Investigation	124
5.3	Archaeological Investigation at the Walbeck Castle	127
6	Discussion and Conclusions	131
	List of Figures	133
	List of Tables	137
	Bibliography	139
	Index	147

Used Symbols

	General notations
a	scalar value
\mathbf{x}	numerical vector (assembled scalars)
x_i	i^{th} component of \mathbf{x}
\mathbf{A}	numerical matrix (assembled vectors)
\mathbf{A}^\dagger	generalized inverse of \mathbf{A}
A_{ij}	element of i^{th} row and j^{th} column
\vec{F}	vectorial property of space
$\text{diag}(\mathbf{x})$	diagonal matrix built up of the elements of \mathbf{x}
	Inversion symbols
N	number of data
$\mathbf{d} \in \mathbb{R}^N$	data vector
$\phi_i(\vec{r})$	model basis function
M	number of model parameter
m_i	single model parameter
$\mathbf{m} \in \mathbb{R}^M$	model vector assembling the model parameters
\mathbf{m}^k	model vector of the k^{th} iteration step
$\mathbf{f}(\mathbf{m}) \in \mathbb{R}^N$	forward response of \mathbf{m}
$\mathbf{S} \in \mathbb{R}^{N \times M}$	sensitivity matrix
$\Delta \mathbf{m}$	model update vector
$\Delta \mathbf{d}$	data discrepancy vector
Φ	functional to be minimized
$\nabla_{\mathbf{m}}$	gradient with respect to the model parameters
γ	negative gradient
\mathbf{H}	Hessian matrix
Φ_d	data functional
$\mathbf{D} \in \mathbb{R}^{N \times N}$	data weighting matrix
ϵ	data error
Φ_m	model functional
$\mathbf{C} \in \mathbb{R}^{p \times M}$	model constraint matrix
\mathbf{m}^0	starting (or a-priori) model
λ	regularization / Lagrangian parameter
$\mathbf{C}_d \in \mathbb{R}^{N \times N}$	data covariance matrix
$\mathbf{C}_m \in \mathbb{R}^{M \times M}$	(a-priori) model covariance matrix
$\mathbf{P} \in \mathbb{R}^{p \times M}$	parameter mapping matrix
$\hat{\mathbf{S}} = \mathbf{D}\mathbf{S}$	error-weighted sensitivity matrix

$\Delta \hat{\mathbf{d}}$	error-weighted discrepancy vector
	SVD/GSVD symbols
$\mathbf{U} \in \mathbb{R}^{N \times N}$	data eigenvector matrix
$\mathbf{V} \in \mathbb{R}^{M \times M}$	model eigenvector matrix
$\mathbf{\Lambda} \in \mathbb{R}^{N \times M}$	singular value matrix
s_i	singular value
f_i	filter factor
$\mathbf{W} \in \mathbb{R}^{M \times M}$	constraint eigenvector matrix
σ_i, μ_i	generalized singular value pair
γ_i	generalized singular value
	Resolution symbols
\mathbf{n}	noise vector
$\mathbf{R}^M \in \mathbb{R}^{M \times M}$	model resolution matrix
$\mathbf{R}^D \in \mathbb{R}^{N \times N}$	data information matrix
$Cov(\mathbf{m})$	Covariance of the vector \mathbf{m}
IC	Information content
RD	Resolution degree
IE	Information efficiency
MCM	Model covariance matrix
	DC symbols
U	voltage
I	electric current
ρ_a	apparent resistivity
k	configuration (or geometry) factor
σ	electric conductivity
ρ	electric resistivity
\vec{j}	current density
\vec{E}	electric field
φ	electric potential
L_σ	partial differential operator of the continuity equation
$\delta(\vec{r})$	Dirac function
\overline{AM}	distance between electrodes A and M
x_i, y_i, z_k	discrete FD grid node positions
\mathbf{K}	(FD) coupling matrix
φ_p, φ_s	primary and secondary potential
$K(\eta)/E(\eta)$	elliptic integrals of first/second order
k_y	spatial wavenumber with respect to y
$\tilde{\varphi}$	potential in wavenumber domain
E	number of electrodes

1 Introduction

1.1 DC Resistivity Methods

Geophysical methods make use of different physical fields interacting with the existing materials. The task is to find a concept (or model) of the earth's structures on the basis of their physical properties. Since a direct investigation of the physical parameters is rarely possible, in most cases the interpretation of geophysical measurements represents an inverse problem.

Amongst geophysical methods direct current (DC) resistivity measurements have been used for many years to investigate the ground. In analogy with the impedance tomography in medical imaging, the measured voltages caused by injected currents reveal information about the earth's resistivity structure. Typical applications arise in the hydrogeological and environmental fields. DC resistivity measurements are used for engineering or archaeological problems as well.

For many years DC sounding and profiling techniques have been applied, which allow for the reconstruction of one- or two-dimensional resistivity models, respectively. Whereas conventional measurements were obtained by much effort for every single data, in the early 1990s multi-electrode systems were developed. They allow for employing arbitrary electrode combinations of a pre-installed electrode array. Nowadays, many single data can be obtained shortly making the DC resistivity profiling method one of the standard investigation techniques for near-surface tasks.

The earth often shows three-dimensional characteristics, which can limit a 2D interpretation. As multi-electrode systems were developing rapidly, e.g., by the use of multi-channel recorders, it became more and more interesting to carry out measurements allowing for a three-dimensional reconstruction. Such "3D data" can be acquired either by several 2D profiles or by the use of an electrode grid, whose perpendicular lines and diagonals can be used (Loke and Barker, 1996b). Moreover, the rapid advancements of personal computers allow for the application of multi-dimensional inversion algorithms.

Many environmental tasks involve polarizable materials, which can be investigated by measuring the induced polarization (IP). With the development of spectral IP devices many questions, particularly in hydro-geological fields, could be answered (Kemna

et al., 2002).

In the recent years, more and more borehole-to-surface and cross-borehole measurements have been carried out (LaBreque and Ward, 1990; Daily and Owen, 1991; Kemna *et al.*, 2004). Since boreholes exist in many areas, subsurface electrodes were included in the data acquisition. They help to improve the quality of resolved structures at depth (Zhou and Greenhalgh, 2002; Sugimoto, 1999).

1.2 Modeling, Inversion and Resolution

An essential part of every inversion scheme is the numerical simulation of measurements for a given parameter distribution. This forward procedure is generally represented by the solution of partial differential equations.

Besides the calculation for closed bodies by means of boundary integral methods the first numerical solutions for arbitrary two-dimensional resistivities were presented by Mufti (1976) and Dey and Morrison (1979b) using finite difference techniques. The main difficulty to overcome is the inclusion of the three-dimensionality of the source into the 2D model solved by a Fourier transform of the partial differential equation with respect to y and a calculation in the wavenumber domain.

It was not until the early 1980s, that approaches for the two-dimensional inversion were developed, e.g. the works of Inman (1975) and Tripp *et al.* (1984) have to be mentioned. The paper of Loke and Barker (1996b) was the basis of the commercially available resistivity inversion program RES2DINV, which has been proved in practice. The foundations for the forward calculation based on finite differences were given by Dey and Morrison (1979a), Spitzer (1995) and other authors. The introduction of improved boundary conditions by Zhang *et al.* (1995) and the singularity removal technique by Lowry *et al.* (1989) improved the quality of the modeling results significantly. Speed and accuracy of different discretization schemes and equation solvers have been investigated by Spitzer and Wurmstich (1995). As a result of the rapid advancement of computers, it is now possible to carry out accurate computations for large models with high resistivity contrasts.

The first 3D inversion of pole-pole data by Park and Van (1991) was followed by other Newton type schemes, e.g. Li and Oldenburg (1999). Alternatively, algorithms based on non-linear conjugate gradients methods were successfully used (Ellis and Oldenburg, 1994; Zhang *et al.*, 1995). The inversion schemes for induced polarization reach from single-step inversions based on a DC resistivity inversion (Beard *et al.*, 1996) to complex algorithms depicted by Kemna *et al.* (2000) or Oldenburg and Li (1994).

Most inversion programs assume a flat surface of the earth. The incorporation of

topography for 2D-models can be accomplished using a Schwarz-Christoffel transformation (Tong and Yang, 1990). If the topography has to be considered for 3D-problems, the forward calculation cannot be solved realistically by finite differences. In this case, finite element techniques provide suitable solutions (Sasaki, 1994; Sugimoto, 1999).

Parallel to the development of inversion routines questions about the resolution properties arose. First, they were assessed by modeling Narayan (1992); Sasaki (1992); Dahlin and Loke (1998, e.g.). The depth of investigation index presented by Oldenburg and Li (1999) yielded the first numerical resolution measure for DC/IP data. An intuitive comprehension of resolution has been obtained by sensitivity studies, see e.g. Friedel (1997), Spitzer (1998) and Dietrich (1999). Following the uncertainty analysis for linear problems (Menke, 1989) resolution matrices were defined for non-linear problems (Meju, 1994a). They have been investigated in detail for DC resistivity (Friedel, 2000) as well as electromagnetic data (Alumbaugh and Newman, 2000).

The quality of the inversion results can be appraised directly from the resolution matrices (Friedel, 2003). Moreover, the resolution analysis provides the basis for the optimization of experimental design (Maurer *et al.*, 2000), being particularly interesting for multi-electrode DC measurements as recently stated by Stummer *et al.* (2004). However, the linearized scheme of Friedel (2003) is restricted to the special case of the truncated SVD inversion and shall be generalized in this thesis.

The central objectives of the thesis are methods for modeling, inversion and resolution analysis of DC resistivity data. The challenges for three-dimensional resistivity inversion are

- arbitrary electrode positions and configurations.
- incorporation of data errors (measured or estimated).
- implementation of problem-dependent inversion and regularization schemes.
- active control of regularization method and strength.
- efficient and fast inverse problem solution.
- fast and accurate forward calculation.
- estimation of quality and uncertainty of model.

1.3 Guideline of the Thesis

The thesis is composed of four main chapters as outlined in the following.

Non-linear Inversion and Resolution

The second chapter works as a basis providing insight into the inversion process in an abstract way without touching the special problems of DC data. It starts with an overview about non-linear inversion techniques, which iteratively minimize the discrepancy between observed data and model response.

Since multi-dimensional inverse problems prove to be ill-posed, a stabilizing mechanism, denoted by regularization, has often to be introduced. A classification of regularization methods is presented and the determination of the regularization strength is depicted.

The application of the Gauss-Newton method results in a linear least squares system. It is shown, how these normal equations can be solved efficiently using conjugate gradient techniques. In particular, equation solvers for multiple regularization parameters are presented that allow for optimized regularization.

By the singular value decomposition (SVD) a comprehension of the mathematical processes in solving linear least squares problems can be obtained. It provides model and data eigenvectors of the problems, which can be used for distinct regularization methods. The generalized SVD (GSVD) helps to solve problems of arbitrary model constraints.

Following linear resolution, the resolution analysis is formulated for non-linear inversion with explicit regularization. It is shown, how the SVD and GSVD can be used to calculate resolution matrices. Starting from that, resolution measures are derived that appraise the quality of inversion results and the information content of data sets.

Inversion of DC Resistivity Data

This chapter provides the prerequisites for a successful application of the minimization techniques to DC resistivity data. To control the inversion process, it is important to estimate the errors of the observed data. The investigation of different noise sources under practical considerations allows for the construction of data weighting matrices used in the inversion. Methods for appropriate parameterization of the subsurface are given.

A central part of every inversion is the forward calculation predicting the response of the model. A finite difference operator in connection with the singularity removal technique is presented. To appraise the accuracy of the forward response, the reciprocity principle is used.

The Jacobian or sensitivity matrix contains the partial derivatives of the model responses with respect to the model parameters, which is necessary in every iteration step to solve the linearized subproblem. It is demonstrated, how DC sensitivities can be obtained for homogeneous and inhomogeneous models. For this purpose, the re-

sults of the singularity removal procedure are used to calculate the Jacobian without additional effort.

Inversion and Resolution Studies

The depicted inversion and regularization methods are compared considering profile data produced by a synthetic model. This involves the determination of appropriate regularization and line search parameters. Furthermore, the resolution of different model cells are investigated.

A linearized study is presented to derive comparable resolution measures. It is shown how different inversion methods affect resolution. Moreover, different data sets are investigated concerning information content and efficiency. Thus, strategies for the optimization of data sets are set up.

The experience for profiles data is used to design three-dimensional data acquisition. Two typical electrode layouts are considered to optimize 3D experimental design. Different parameterization techniques are compared.

Moreover, an approach is presented that involves Finite Element calculations on tetrahedral grids to incorporate topography into 3D DC inversion.

Application to Field Data

In the last chapter the inversion routines are applied to field data. Problems of cavity detection, ground falls on salt-waste dump and archaeology are successfully investigated. The results of the experimental optimization are proved in practice. Furthermore, the resolution analysis is used to appraise the inversion results.

Finally, the main results are summarized and discussed to point out future objectives.

2 Non-linear Inversion and Resolution

In applied geophysics a concept of the earth has to be found based on a limited number of measurements. In most cases this represents a non-linear inverse problem. Solutions are found by the iterative minimization of the misfit between the data and the model's response. In section 2.1 the reader is introduced to the methodology of non-linear minimization. In particular, gradient techniques and Newton type methods are described.

Many multi-dimensional inverse parameter problems are known to be non-unique to solve and numerically instable, depicted by the term ill-posed. Hence, the inverse problem has to be regularized by appropriate techniques. Section 2.2 describes, how inversion can be controlled by various methods. A classification of regularization techniques is given. One successful way is to introduce a penalty functional, which is weighted by regularization parameter. It is shown, how the strength of the regularization can be determined appropriately.

In every iteration step, a linear sub-problem representing a large-scale system of equations has to be solved. In Section 2.3 it is presented, how equation solvers on the basis of conjugate gradient methods can be adapted to the specific formulation. The resulting algorithms prove to be very efficient. Furthermore, they can be extended for the creation of a set of models used to search for optimum regularization.

The singular value decomposition (SVD) is a well-known numerical tool helping to understand inverse problems. By generalization it can be applied to the construction of generalized inverse matrices for all explicit regularization schemes as depicted in section 2.4.

Once a result is found by non-linear inversion, methods have to be applied to provide information on the reliability of the model. Resolution analysis as known from linear inverse problems is formulated for non-linear inversion in section 2.5. The resolution matrices are used to derive measures for the information content of the data set. They can be obtained by the SVD/GSVD or by solving a linear inverse problem.

2.1 Inversion Methods

2.1.1 Inversion Methodology

Assume a set of N measurements affected by the physical property p of the subsurface. The set of all data is represented by the data vector $\mathbf{d} = (d_1, d_2, \dots, d_N)^T$. We are then interested in finding a spatial parameter distribution $p(\vec{r})$ that explains our data to a certain degree. As we know, data are contaminated with noise. Thus, we try to fit that part of the data, which is generated by parameter variations. The function $p(\vec{r})$ is discretized into a limited number M of model parameters m_i serving as weighting coefficients for basis functions $\phi_i(\vec{r})$ such that

$$p(\vec{r}) = \sum_{i=1}^M m_i \phi_i(\vec{r}) \quad . \quad (2.1)$$

Arranging the model parameters yields the model vector¹

$$\mathbf{m} = (m_1, m_2, \dots, m_M)^T \quad .$$

The basis functions ϕ_i have to be selected such that the reality can be described with satisfying accuracy. For that reason it seems necessary to introduce many basis functions. In contrast, it is important to limit the degree of freedom. The ϕ_i can be spectral functions, Chebychev polynomials as well as problem inherent functions (Oldenburg *et al.*, 1993). Another way is to subdivide the region of interest into sub-domains Ω_i and to define the basis functions being

$$\phi_i(\vec{r}) = 1 \text{ for } \vec{r} \in \Omega_i \text{ and } 0 \text{ elsewhere.}$$

The sub-domains Ω_i are convex bodies such as hexahedrons, either defining a grid or not, or tetrahedrons. Central objective of inversion is to find a model \mathbf{m} , whose response² $\mathbf{f}(\mathbf{m})$ fits the data vector \mathbf{d} . For linear problems there exists a linear relation between \mathbf{m} and \mathbf{f} , which can be expressed by a matrix-vector multiplication

$$\mathbf{f}(\mathbf{m}) = \mathbf{Fm} \quad .$$

Examples for linear relations in geophysics are gravimetry, magnetics, or vertical seismic profiling. For non-linear problems the forward operation depends on the model \mathbf{m} itself, which holds for all methods that are based on Maxwell's equations. Beginning from a starting model \mathbf{m}^0 , an iterative process is applied to update the model until

¹The model vector is often identified with the term "the model".

²The (forward) response are the data, which would have been measured assuming a model.

data fit or convergence is achieved. In each iteration step k a new model is calculated by adding a model update $\Delta \mathbf{m}^k$

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \Delta \mathbf{m}^k \quad . \quad (2.2)$$

A Taylor approximation of first order yields

$$\mathbf{f}(\mathbf{m}^k + \Delta \mathbf{m}^k) = \mathbf{f}(\mathbf{m}^k) + \frac{\partial \mathbf{f}(\mathbf{m}^k)}{\partial \mathbf{m}} \Delta \mathbf{m}^k + \dots \approx \mathbf{f}(\mathbf{m}^k) + \mathbf{S} \Delta \mathbf{m}^k \quad ,$$

where the partial derivative of the model response with respect to the model parameters is called Jacobian or sensitivity matrix $\mathbf{S} \in \mathbb{R}^{N \times M}$ with the elements

$$S_{ij}(\mathbf{m}^k) = \frac{\partial f_i(\mathbf{m}^k)}{\partial m_j} \quad . \quad (2.3)$$

Setting the response of the new model $\mathbf{f}(\mathbf{m} + \Delta \mathbf{m})$ equal to the data \mathbf{d} we obtain the non-quadratic equation

$$\mathbf{S} \Delta \mathbf{m} = \mathbf{d} - \mathbf{f}(\mathbf{m}) \quad , \quad (2.4)$$

which has to be solved in some sense to minimize the residual vector $\mathbf{d} - \mathbf{f}(\mathbf{m})$. Often, for each data point d_i an error ϵ_i is known or can be estimated, which is used for weighting the residual. Using an L_p -norm of the weighted residual, a data functional Φ_d to be minimized is defined by

$$\Phi_d(\mathbf{m}) = \sum_{i=1}^N \left| \frac{d_i - f_i(\mathbf{m})}{\epsilon_i} \right|^p = \|\mathbf{D}(\mathbf{d} - \mathbf{f}(\mathbf{m}))\|_p^p \quad (2.5)$$

$$\text{with } \mathbf{D} = \text{diag} \left(\frac{1}{\epsilon_i} \right) \quad .$$

Generally, different values for p can be used (Farquharson and Oldenburg, 1998) corresponding to the expected noise characteristics. The L_1 -norm is particularly advantageous, if the noise has a long-tailed distribution. Since it is less sensitive to outliers in the data, an L_1 minimization procedure is often called "robust inversion" (Claerbout and Muir, 1973).

In the following, we assume the noise to be Gaussian corresponding to the use of the L_2 -norm measure. The mean value of the data functional $\chi^2 = \Phi_d/N$ is often considered for linking to the statistical nature. A value of $\chi^2 = 1$ implies the data being fitted within their errors. Thus, the functional norm can be written as

$$\Phi_d = [\mathbf{D}(\mathbf{d} - \mathbf{f}(\mathbf{m}))]^T [\mathbf{D}(\mathbf{d} - \mathbf{f}(\mathbf{m}))] = (\mathbf{d} - \mathbf{f}(\mathbf{m}))^T \mathbf{D}^T \mathbf{D} (\mathbf{d} - \mathbf{f}(\mathbf{m})) \quad . \quad (2.6)$$

The product $\mathbf{D}^T \mathbf{D}$ is the inverse of the data covariance matrix \mathbf{C}_d as used by Tarantola (1978) for the case of uncorrelated errors with standard deviations ϵ_i and variances ϵ_i^2 . If correlations between the individual errors are present, the covariance matrix \mathbf{C}_d does not remain diagonal. A profound insight into the nature of the error sources is necessary to find out, whether systematical data correlations exist. Examples of correlated errors in geophysics are:

- The individual responses as function of frequency in magnetotelluric measurements are disturbed by local conductivities producing static shift, which acts as systematic source of error, if not considered in inversion.
- If the electrode positions of a multi-electrode system are considered to show uncertainties, all measurements using an individual electrode will show systematic variations.

As we will see later, the functional to be minimized can be expanded by other terms yielding a combined Φ . At first, we set $\Phi = \Phi_d$.

2.1.2 The Steepest Descent Method

It is obvious, that one must seek the minimum of Φ in the direction of the steepest descent of Φ . The model update is a proportion of the negative gradient of Φ , $\gamma^k = -\nabla_{\mathbf{m}} \Phi$, the partial derivative vector $\nabla_{\mathbf{m}}$ reads

$$\nabla_{\mathbf{m}} = \left(\frac{\partial}{\partial m_1}, \frac{\partial}{\partial m_2}, \dots, \frac{\partial}{\partial m_M} \right)^T .$$

A step length τ of the descent step has to be estimated such that $\Phi(\mathbf{m}^k + \tau \gamma^k)$ is minimized. A procedure, which tries to search for an optimum solution along the line defined by varying τ is referred to as line search.

Algorithm 1 Steepest descent method

```

k = 0
while  $\Phi(\mathbf{m}^k) > \Phi^*$  do
   $\gamma^k = -\nabla_{\mathbf{m}} \Phi(\mathbf{m}^k)$  {find negative gradient}
  find  $\tau^k$  that minimizes  $\Phi(\mathbf{m}^k + \tau^k \gamma^k)$  {line search}
   $\mathbf{m}^{k+1} = \mathbf{m}^k + \tau^k \gamma^k$  {model update}
  k = k + 1
end while

```

In every iteration the forward response and the functional's gradient have to be calculated. The latter can be achieved by the explicit use of the Jacobian matrix or by

additional forward calculations. The convergence rate for this method is very slow for ill-posed problems. This results in many forward calculations, which are generally very time-consuming. Hence, this method is very simple to implement, but rarely of practical use.

2.1.3 The Nonlinear Conjugate Gradients Method

The technique of the conjugate gradients was developed by Hestenes and Stiefel (1952) to solve a linear system of equations $\mathbf{Ax} = \mathbf{b}$ for sparse matrices \mathbf{A} . The underlying principle is to find a set of orthogonal directions and to compute accompanying weights in a manner that every search direction is used only once. As a result, convergence is generally fast compared to steepest descent equation solvers. Since the method is based on an iterative minimization of the functional $1/2\mathbf{x}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{b}$, this technique can also be exploited for non-linear minimization (Shewchuk, 1994; Vogel, 2002), as successfully used for the inversion of magnetotelluric (Mackie and Madden, 1993; Rodi and Mackie, 2001) and DC resistivity (Zhang *et al.*, 1995; Ellis and Oldenburg, 1994) data.

Algorithm 2 Nonlinear conjugate gradients (NLCG)

```

 $k = 0$ 
 $\mathbf{g}^0 = \nabla_{\mathbf{m}}\Phi(\mathbf{m}^0)$ 
 $\gamma^0 = -\mathbf{g}^0$ 
 $\delta^0 = \|\mathbf{g}^0\|^2$ 
while  $\|\gamma^k\| > \text{tol}$  do
  find  $\tau$  that minimizes  $\Phi(\mathbf{m}^k + \tau\gamma^k)$ 
   $\mathbf{m}^{k+1} = \mathbf{m}^k + \tau\gamma^k$ 
   $\mathbf{g}^{k+1} = \nabla_{\mathbf{m}}\Phi(\mathbf{m}^{k+1})$ 
   $\delta^{k+1} = \|\mathbf{g}^{k+1}\|^2$ 
   $\beta^{k+1} = \delta^{k+1}/\delta^k$ 
   $\gamma^{k+1} = -\mathbf{g}^{k+1} + \beta^{k+1}\gamma^k$ 
   $k = k + 1$ 
end while

```

The gradient \mathbf{g}_d of the functional Φ_d can be easily computed from the sensitivity matrix using the chain rule

$$\mathbf{g}_d^k = 2\nabla_{\mathbf{m}}\Phi_d(\mathbf{m}^k) = \mathbf{S}^T\mathbf{D}^T\mathbf{D}(\mathbf{f}(\mathbf{m}^k) - \mathbf{d}) \quad .$$

In each iteration k the model response $\mathbf{f}(\mathbf{m}^k)$ has to be calculated as well as the gradient \mathbf{g}^k . In some cases it is possible to calculate the product of the sensitivity matrix and

an arbitrary vector without explicitly forming \mathbf{S} economizing memory. One example using this method is the 3D DC inversion of Zhang *et al.* (1995).

2.1.4 Newton Type Methods

To derive Newton's method, we start with a second order Taylor series for the functional Φ of an updated model $\mathbf{m} + \Delta\mathbf{m}$

$$\Phi(\mathbf{m} + \Delta\mathbf{m}) \approx \tilde{\Phi}(\mathbf{m} + \Delta\mathbf{m}) \quad \text{with} \quad (2.7)$$

$$\tilde{\Phi}(\mathbf{m} + \Delta\mathbf{m}) = \Phi(\mathbf{m}) + (\nabla_{\mathbf{m}}\Phi(\mathbf{m}))^T \Delta\mathbf{m} + \frac{1}{2}\Delta\mathbf{m}^T (\nabla_{\mathbf{m}}^2\Phi(\mathbf{m}))^T \Delta\mathbf{m} + \dots \quad (2.8)$$

where the second derivative $\nabla_{\mathbf{m}}^2$ is the Hessian matrix with the elements

$$(\nabla_{\mathbf{m}}^2\Phi)_{ij} = (\nabla_{\mathbf{m}}\nabla_{\mathbf{m}}^T\Phi)_{ij} = \frac{\partial^2\Phi}{\partial m_i \partial m_j} \quad .$$

The second order approximation $\tilde{\Phi}$ is minimized by setting its first partial derivative with respect to $\Delta\mathbf{m}$ to zero³, which results in

$$\nabla_{\mathbf{m}}\Phi(\mathbf{m}) + \nabla_{\mathbf{m}}^2\Phi(\mathbf{m})\Delta\mathbf{m} = 0 \quad .$$

Hence, the model update $\Delta\mathbf{m}^k$ is sought by solving the equation

$$(\nabla_{\mathbf{m}}^2\Phi) \Delta\mathbf{m}^k = -\nabla_{\mathbf{m}}\Phi \quad . \quad (2.9)$$

Equation (2.9) represents a linear subproblem to be solved in every iteration step. Applying an additional line search to Newton's method results in algorithm 3. For the functional Φ_d as defined by (2.6) holds

$$\begin{aligned} \mathbf{H}_d^k &= \nabla_{\mathbf{m}}\nabla_{\mathbf{m}}^T\Phi_d = \nabla_{\mathbf{m}}(2\mathbf{S}^T\mathbf{D}^T\mathbf{D}(\mathbf{f}(\mathbf{m}^k) - \mathbf{d})) \\ &= 2\mathbf{S}^T\mathbf{D}^T\mathbf{D}\mathbf{S} + 2(\nabla_{\mathbf{m}}^T\mathbf{S}^T)\mathbf{D}^T\mathbf{D}(\mathbf{f}(\mathbf{m}^k) - \mathbf{d}) \quad . \end{aligned} \quad (2.10)$$

The second term at the right hand side is generally difficult to compute. It can be neglected, if the problem is not strongly non-linear ($\nabla_{\mathbf{m}}^T\mathbf{S}^T$ is small).

Gauss-Newton and quasi-Newton methods

The method using the Hessian approximation $\mathbf{H}_d^k = 2\mathbf{S}^T\mathbf{D}^T\mathbf{D}\mathbf{S}$ is called Gauss-Newton method. Besides the easier computation it has the advantage that the Hessian approximation is positive semidefinite, which guarantees the Gauss-Newton step to be

³The necessary condition for a minimum is the positivity of the Hessian matrix, which can be shown easily for all subsequent functionals.

Algorithm 3 Newton's method

```

 $k = 0$ 
choose appropriate  $\mathbf{m}^0$ 
calculate  $\Phi(\mathbf{m}^0)$ 
while  $\Phi(\mathbf{m}^k) > \Phi^*$  do
  calculate  $\gamma^k = -\nabla_{\mathbf{m}}\Phi(\mathbf{m}^k)$  {negative gradient}
  calculate or update  $\mathbf{H}^k = \nabla_{\mathbf{m}}^2\Phi(\mathbf{m}^k)$  {Hessian}
   $\Delta\mathbf{m}^k = \mathbf{H}^{k-1}\gamma^k$  {solve linear subproblem}
  find  $\tau^k$  minimizing  $\Phi(\mathbf{m}^k + \tau\Delta\mathbf{m}^k)$  {line search}
   $\mathbf{m}^{k+1} = \mathbf{m}^k + \tau\Delta\mathbf{m}^k$  {model update}
   $k = k + 1$ 
  calculate  $\Phi(\mathbf{m}^k)$  {forward calculation}
end while

```

a descent direction. However, the quadratic convergence of Newton's method is lost for strong non-linearity. With $\Delta\mathbf{d}^k = \mathbf{d} - \mathbf{f}(\mathbf{m}^k)$ the linear subproblem (2.9) can be written as

$$((\mathbf{DS})^T(\mathbf{DS})) \Delta\mathbf{m}^k = (\mathbf{DS})^T\mathbf{D}\Delta\mathbf{d}^k \quad (2.11)$$

and interpreted as least squares solution of $\mathbf{DS}\Delta\mathbf{m}^k = \mathbf{D}\Delta\mathbf{d}^k$.

Due to the second order of the Taylor approximation Newton's method has quadratic convergence. Thus, the number of iterations will be small, if the starting model is in the neighborhood of the minimum. If there are several local minima, the solution is uniquely defined by the starting model. The quest for a global minimum can be achieved by starting from a set of different starting models \mathbf{m}^0 . Other globalization schemes are trust region techniques or homotopy methods (Jegen *et al.*, 2001; Vogel, 2002). However, for most problems there is no need for such methods.

The crucial point of Newton's method is to recalculate the Hessian matrix in every iteration, which can be very time consuming. For some problems it can be obtained as a byproduct of the forward calculation routine. An alternative is to update the Hessian by previous gradients, which results in a secant method, also referred to as quasi-Newton method. The well-known BFGS method developed by Broyden, Fletcher, Goldfarb and Shanno uses the approximation

$$\mathbf{H}^{k+1} = \mathbf{H}^k - \frac{\mathbf{H}^k\Delta\mathbf{m}^k\Delta\mathbf{m}^{kT}\mathbf{H}^k}{\Delta\mathbf{m}^{kT}\mathbf{H}^k\Delta\mathbf{m}^k} + \frac{\mathbf{y}^k\mathbf{y}^{kT}}{\Delta\mathbf{m}^{kT}\mathbf{y}^k}, \quad (2.12)$$

where $\mathbf{y} = \nabla_{\mathbf{m}}\Phi(\mathbf{m}^{k+1}) - \nabla_{\mathbf{m}}\Phi(\mathbf{m}^k)$.

A variant of this method is used by Loke and Barker (1996b) in the form of (Broyden, 1972)

$$\mathbf{H}^{k+1} = \mathbf{H}^k + \frac{[\mathbf{f}(\mathbf{m}^{k+1}) - \mathbf{f}(\mathbf{m}^k) - \mathbf{H}^k \Delta \mathbf{m}^k] \Delta \mathbf{m}^{kT}}{\Delta \mathbf{m}^{kT} \Delta \mathbf{m}^k} .$$

Loke and Dahlin (2002) showed that for most 2D DC resistivity problems this method is sufficient.

For strongly non-linear problems the line search procedure becomes inevitable. Generally the optimum τ lies between 0 and 1. The more linear the problem is, the closer the minimum approaches $\tau = 1$. For implementation of line search methods see section 4.1.

2.2 Regularization

2.2.1 Regularization Types

Multidimensional inversion problems generally have more free model parameters than data. In many cases the model domain shows both well resolved and poorly resolved model parameters. Consequently, there exist a lot of models fitting the data well. The use of the above described minimization scheme would probably lead to highly oscillating models with huge parameter contrasts. Therefore, it becomes necessary to eliminate the ambiguity of the problem. Such techniques are denoted with the term regularization. It denotes the procedure of removing the singularity of the left hand side matrix in the normal equations (2.11).

Implicit vs. explicit methods

One type of regularization involves approximate solutions to the basis equation

$$\mathbf{D} \mathbf{S} \Delta \mathbf{m}^k = \mathbf{D} \Delta \mathbf{d}^k$$

or the corresponding normal equations without explicitly demanding any specific model characteristics. Krylov subspace methods have the characteristics, that smooth model components converge faster than non-smooth model parts (Hansen, 1992). By stopping the iteration after k_{max} steps a regularization can be introduced. Small values of k_{max} correspond to smooth solutions, whereas large values result in more structured solutions. We can exploit this property either by the NLCG as mentioned above or in the form of the TLS (truncated least squares) method, which solves the Gauss-Newton subproblem incompletely on the basis of conjugate gradients (see section 2.3). Another

possibility is the use of reconstruction techniques derived from tomography problems such as SIRT (simultaneous iterative reconstruction technique).

In contrast, regularization can be applied directly to control the model in an explicit way. An example is the truncated SVD (TSVD) solution that applies a generalized inverse using a limited number of degrees of freedom, see section 2.4. An alternative is the (explicit) introduction of an additional model functional Φ_m to be minimized, which is a function of the model update or the model itself. Explicit regularization can be interpreted as numerical stabilization of the problem, the additional term is often called "stabilizer". Actually, the stabilizing term ensures the uniqueness of the solution, because the unregularized normal matrix is generally singular. For all explicit methods a generalized inverse \mathbf{A}^\dagger is found such that $\Delta\mathbf{m}^k = \mathbf{A}^\dagger\Delta\mathbf{d}^k$.

Local vs. global constraints

Constraining the model update $\Delta\mathbf{m}^k$ in every iteration step is referred to as local regularization. Each subproblem is treated as independent linear inverse problem to be regularized. This is the case for the truncated SVD inversion (TSVD, see section 2.4) as well as for the truncated least squares (TLS) and iterative reconstruction techniques like SIRT. The model functional is considered to have the form of an L_2 -norm

$$\Phi_m = \|\mathbf{c}(\Delta\mathbf{m})\|_2^2 \quad .$$

The constraint function $\mathbf{c}(\mathbf{m})$ represents some expected behavior of the model update. Probably the first and simplest function is $\mathbf{c} = \Delta\mathbf{m}$ used by the ridge regression method (Marquardt, 1963), also denoted as Marquardt-Levenberg method. The corresponding functional denotes the quadratic length of the update vector, which is kept small to prevent oscillations in the model. From this idea the term "damping" originates, which is often used in connection with locally constrained inversion (Loke and Barker, 1996b).

On the contrary to local regularization the term global regularization is used to constrain the model itself. Then, the model functional

$$\Phi_m = \|\mathbf{c}(\mathbf{m})\|_2^2$$

remains invariant during inversion process. For example, it is used to keep the model close to a model concept \mathbf{m}^0 derived from a-priori information. Other approaches are techniques, that prefer models that are "simple" in a certain sense based on Occam's⁴ principle (or Occam's razor):

⁴William of Occam was a monk living in Scotland's 14th century.

One should not increase, beyond what is necessary, the number of entities required to explain anything. (William of Occam)

Thus, of all possible explanations for a set of observations, the simplest has to be chosen. But what is simple? Simplicity can certainly be interpreted subjectively. For example, many authors identify Occam’s principle with smooth model inversion (Constable *et al.*, 1987; Beard *et al.*, 1996). Besides the minimization of parameter gradients in the model, the number of occurring parameter jumps can be treated as function to be minimized. Such approaches are referred to as total variation (Vogel, 2002) or focused (Portniaguine and Zhdanov, 1999) inversion and introduce a non-linear model functional. Other realizations for model simplicity can be defined by probability distributions or entropy (Press *et al.*, 1992).

Actually, every explicit global regularization is following Occam’s principle⁵ defining its own interpretation of ”simplicity”. To the contrary Occam’s principle is restricted to global regularizing methods, even though the desired model property (e.g. smoothness) can also be applied to the model update in a local regularization scheme. However, even if all model updates are relatively smooth, it is not evident that the final model has to be smooth as well.

The gradient methods Steepest Descent and NLCG have effects onto the model itself and can thus be denoted as globally regularizing. Table 2.1 provides an overview of the discussed inversion/regularization approaches in terms of explicit/implicit and local/global techniques.

	implicit/indirect	explicit/direct
local	TLS, SIRT	TSVD, Ridge Regression
global	NLCG, Steepest descent	Occam-Type, focused inv.

Table 2.1: Classification of regularization schemes regarding the effect on model or model update and explicit or implicit formulation

From the information point of view, regularization supplements the information provided by the data. The additional information is either obtained by a-priori knowledge or generated by expectations to the investigated structures. Amongst all models fitting the data equivalently, the one with minimum constraint Φ_m is favored.

⁵Unfortunately, the term Occam inversion is often used to denote smoothness constrained schemes, which is just one realization of simplicity.

2.2.2 Explicit Regularization

To weight data and model functional, a regularization parameter λ is introduced to construct the total functional to be minimized

$$\Phi = \Phi_d + \lambda\Phi_m \rightarrow \min \quad . \quad (2.13)$$

If the data variances are known, the data functional should converge to the corresponding Φ_d^* . Thus, the problem can be formulated as constraint minimization of Φ_m under the condition $\Phi_d = \Phi_d^*$. The introduction of a Lagrangian parameter μ yields

$$\bar{\Phi} = \Phi_m + \mu|\Phi_d - \Phi_d^*| \rightarrow \min \quad ,$$

whose solution is identical to those of equation (2.13) for $\mu = 1/\lambda$. For historical reasons the first formulation is used. The gradient of Φ is the weighted sum of the gradients of Φ_d and Φ_m , which also holds for the Hessian. Hence, Newton's method applied to Φ reads

$$(\nabla_{\mathbf{m}}^2\Phi_d + \lambda\nabla_{\mathbf{m}}^2\Phi_m) \Delta\mathbf{m} = -\nabla_{\mathbf{m}}\Phi_d - \lambda\nabla_{\mathbf{m}}\Phi_m \quad . \quad (2.14)$$

In terms of matrix inversion it is reasonable to write Φ_m as squared norm of a product of a constraint matrix \mathbf{C} and the difference between the model \mathbf{m} and a reference model \mathbf{m}^0

$$\Phi_m = \|\mathbf{C}(\mathbf{m} - \mathbf{m}^0)\|_2^2 = (\mathbf{m} - \mathbf{m}^0)^T \mathbf{C}^T \mathbf{C} (\mathbf{m} - \mathbf{m}^0) \quad , \quad (2.15)$$

which possesses the gradient

$$\nabla_{\mathbf{m}}\Phi_m = 2\mathbf{C}^T \mathbf{C} (\mathbf{m} - \mathbf{m}^0)$$

and the Hessian

$$\nabla_{\mathbf{m}}^2\Phi_m = 2\mathbf{C}^T \mathbf{C} \quad .$$

It is possible to introduce the additional gradient into gradient methods like NLCG. However, as those include intrinsic regularization by stopping the iteration before convergence, the advantages of the NLCG algorithm are destroyed.

Considering the Gauss-Newton method, equation (2.11) is expanded by Φ_m 's gradient and Hessian, yielding

$$(\mathbf{S}^T \mathbf{D}^T \mathbf{D} \mathbf{S} + \lambda \mathbf{C}^T \mathbf{C}) \Delta\mathbf{m}^k = \mathbf{S}^T \mathbf{D}^T \mathbf{D} (\mathbf{d} - \mathbf{f}(\mathbf{m}^k)) - \lambda \mathbf{C}^T \mathbf{C} (\mathbf{m}^k - \mathbf{m}^0) \quad . \quad (2.16)$$

Equation (2.16) represents the regularized normal equations. It is a special case for the statistical model estimator of Tarantola (1978)

$$(\mathbf{S}^T \mathbf{C}_d^{-1} \mathbf{S} + \lambda \mathbf{C}_m^{-1}) \Delta\mathbf{m}^k = \mathbf{S}^T \mathbf{C}_d^{-1} \Delta\mathbf{d}^k - \lambda \mathbf{C}_m^{-1} (\mathbf{m}^k - \mathbf{m}^0) \quad . \quad (2.17)$$

In analogy to $\mathbf{D}^T\mathbf{D} = \mathbf{C}_d^{-1}$, the matrix $\mathbf{C}^T\mathbf{C}$ can be interpreted as the inverse of the a-priori model covariance matrix \mathbf{C}_m . For reasons of clarity we continue using the data weighting matrix \mathbf{D} and the constraint matrix \mathbf{C} , which has advantages in the implementation of equation solvers and in the calculation of resolution properties. Nevertheless, one has to keep in mind the role of the matrices $\mathbf{C}^T\mathbf{C}$ and $\mathbf{D}^T\mathbf{D}$ as a-priori covariances for model and data, respectively.

The application of a local regularization scheme minimizing

$$\Phi = \|\mathbf{D}(\mathbf{S}\Delta\mathbf{m}^k - \Delta\mathbf{d}^k)\|_2^2 + \lambda \|\mathbf{C}\Delta\mathbf{m}^k\|_2^2$$

with respect to $\Delta\mathbf{m}^k$ in every iteration step leads to

$$(\mathbf{S}^T\mathbf{D}^T\mathbf{D}\mathbf{S} + \lambda\mathbf{C}^T\mathbf{C}) \Delta\mathbf{m}^k = \mathbf{S}^T\mathbf{D}^T\mathbf{D}\Delta\mathbf{d}^k \quad , \quad (2.18)$$

which differs from equation (2.16) merely by the missing term $-\lambda\mathbf{C}^T\mathbf{C}(\mathbf{m} - \mathbf{m}^0)$ at the right hand side. Hence, local constraints can be implemented as global constraints by setting $\mathbf{m}^0 = \mathbf{m}^k$ in every iteration.

The first iteration model \mathbf{m}^1 is identical for both schemes. It points out, that the main model structures appear in the first inversion step. In the further course of inversion the results differ. Assuming, λ is large enough to dominate the inverse subproblem, the solution for $\Delta\mathbf{m}$ tends to approximate $\mathbf{m}^0 - \mathbf{m}^k$. Hence, \mathbf{m}^{k+1} is biased towards the reference model \mathbf{m}^0 (Meju, 1994a). In the course of iterations trade-off between distance to \mathbf{m}^0 and data fit is established. This can be advantageous for convergence reasons as well as disadvantageous, e.g. if \mathbf{m}^0 is far from reality and the data cannot be fitted appropriately. See resolution analysis in section 2.5 for details.

2.2.3 Implementation

There is a variety of regularization schemes in inverse theory differing in the use of the constraint matrix \mathbf{C} and the reference model \mathbf{m}^0 . For an overview the reader is referred to Zhang *et al.* (1996) and Vogel (2002).

Zeroth-order regularization

The simplest method to be implemented is the application of the identity matrix \mathbf{I} used by the ridge regression method of Marquardt (1963) as local stabilizer. The global constraints $\Phi_m = \|\mathbf{m} - \mathbf{m}^0\|_2^2$ are successfully used in cases, where some a-priori model \mathbf{m}^0 is known from other investigations or geological concepts. However, using a more or less regular model parameterization, it turns out that well resolved model cells in the neighborhood of the sensors are over-accentuated while badly resolved regions, e.g. in deep layers, show much too less structure.

One way to circumvent this naturally occurring effect is to create a parameterization that takes into account the physical resolution properties. However, resolution is not known before inversion, so the realization becomes a trial-and-error procedure. In practice, an idea of geometrical resolution can often be derived by experience.

Another disadvantage of using the identity matrix as stabilizer is the fact, that the model update vector is implicitly expected to possess a Gaussian distribution with zero mean. For many cases, for example large bodies of constant resistivity, this assumption leads to unreasonable artifacts. However, the method is suitable for small parameter inversions or in the absence of geometrical relations, e.g. for the subspace methods of Oldenburg *et al.* (1993) or Siripunvaraporn and Egbert (2000).

Another possibility is the introduction of a weighting function for the individual model parameter. The matrix \mathbf{C} is then a diagonal matrix with the weighting function on the main diagonal and zeros elsewhere. The product of $\sqrt{\lambda}\mathbf{C}$ can also be interpreted as diagonal Lagrangian parameter matrix.

Generally, the weighting function obtains large values for well resolved and small values for badly resolved parameters. This function can be an approximation to the diagonal of the resolution matrix. However, changing the constraints produces a changed resolution matrix resulting in an iterative constraint balancing. An interesting approach of changing the weights during inversion is the active constraint balancing method of Yi *et al.* (2003). In several cases it proves reasonable to make use of the coverage or cumulative sensitivity. In analogy to linear tomography problems, it is the sum of all (absolute values of the) sensitivities for a given model parameter

$$\text{cov}_j = \sum_{i=1}^N |S_{ij}| \quad .$$

Since the coverage changes during the iterations as the sensitivity does, it seems reasonable to use the one of the starting model. This can be interpreted as an inverse a-priori model covariance matrix.

Smoothness constraints

Simplicity is often associated with smooth models motivated by the limited physical resolution. For geometrically arranged model parameters it is easily possible to calculate model gradients. The constraint matrix \mathbf{C} then represents a discrete derivation operator δ^n of order n , \mathbf{m}^0 can be neglected for a homogeneous starting model. Discrete first order derivation for a one-dimensional model is achieved by finite differences $\delta m_{i+1/2} = m_{i+1} - m_i$, which may either be divided by the distance $\delta x_{i+1/2}$ between m_i

and m_{i+1} or not. For the latter case the matrix \mathbf{C} reads

$$\mathbf{C}_{1st} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix} . \quad (2.19)$$

Note that for the actual differential operator every row i has to be multiplied with the factor $1/(x_{i+1}^m - x_i^m)$, if the x^m denote representative cell locations, e.g. the midpoint. Following equation (2.19), the gradient matrices for all directions can be assembled, denoted with \mathbf{C}_x , \mathbf{C}_y , \mathbf{C}_z . By introduction of weights $\alpha_x, \alpha_y, \alpha_z$ for the various orientations the total regularization matrix can be calculated (Li and Oldenburg, 1999)

$$\mathbf{C}^T \mathbf{C} = \alpha_x \mathbf{C}_x^T \mathbf{C}_x + \alpha_y \mathbf{C}_y^T \mathbf{C}_y + \alpha_z \mathbf{C}_z^T \mathbf{C}_z . \quad (2.20)$$

The weights α_ν can be used to enforce lateral or vertical model changes, which is useful in layered media or for discriminating vertical boundaries of geological units. Note that the matrix \mathbf{C} can also be assembled as

$$\mathbf{C} = \begin{pmatrix} \sqrt{\alpha_x} \mathbf{C}_x \\ \sqrt{\alpha_y} \mathbf{C}_y \\ \sqrt{\alpha_z} \mathbf{C}_z \end{pmatrix} .$$

The latter formulation is advantageous for computing resolution matrices using the GSVD (see section 2.4).

A discrete second order derivative with respect to x in one dimension is given by

$$\left(\frac{\delta^2 m}{\delta x^2} \right)_i = \frac{2}{x_{i+1}^m - x_{i-1}^m} \left(\frac{m_{i+1} - m_{ik}}{x_{i+1}^m - x_i^m} - \frac{m_i - m_{i-1}}{x_i^m - x_{i-1}^m} \right) .$$

It can be applied to all present directions, leading to an equation system that links neighboring model cells. All coefficients for the cells are assembled in a matrix \mathbf{C} . The element C_{mn} represents the contribution of the n^{th} cell to the smoothness at the m^{th} cell. At the boundaries of the model assumptions have to be made for the outward neighbors⁶. This corresponds to the boundary conditions for the discretization of partial differential equations. The use of homogeneous Neumann conditions, e.g., works if the model parameter were continued outside the model boundaries. As the roughness is required to be kept small, this can sometimes lead to artificial structures

⁶If the smoothness is only assembled for the inner cells, the cells at the model edges are unconstrained, which leads to artifacts in the result.

near the boundaries. One can avoid this by formulating derivatives at one boundary only with respect to the other directions. However, it turns out that models tend to show over-accentuated structures at the boundary cell line.

It has to be noted that the use of Neumann conditions at all boundaries leads to singular matrices, as known from the solution of elliptic boundary value problems. Although the regularity of the subproblem's left hand side matrix is still ensured, the necessary iteration numbers in the Krylov methods increase significantly. Since the smoothness is applied to the model changes, it shows reasonable to use homogeneous Dirichlet boundary conditions for the second order smoothness, which proves to yield acceptable results in most cases.

It can easily be seen, that for equidistant grid models ($\Delta x = \Delta y = \Delta z$) the second order smoothness matrix equals the quadratic first order smoothness

$$\mathbf{C}_{2nd} = \mathbf{C}_{1st}^T \mathbf{C}_{1st} \quad .$$

Assuming $\Delta x = \Delta y = \Delta z = 1$, the matrix \mathbf{C} constitutes of the constant value of -6 at the main diagonals and +1 at the off-diagonals representing the 6 respective neighboring cells.

Mixed constraints

Different constraints can be mixed using matrix algebra

- Addition $\mathbf{C} = \mathbf{C}_1 + \mathbf{C}_2$, e.g. smoothness constraints and step-length damping to make \mathbf{C} regular.
- Multiplication $\mathbf{C} = \mathbf{C}_1 \mathbf{C}_2$, e.g. to weight the smoothness for different model regions.

2.2.4 Choosing the Regularization Parameter

The regularization or Lagrangian parameter λ arises for explicit regularization schemes in equation (2.13). It weights the model constraints against the data misfit. Small values of λ will produce a highly structured model with huge parameter contrasts, explaining the data well. To the contrary, large λ values result in "simple" models with poor data fit. For implicit regularization schemes the number of iteration steps k_{max} (for TLS and NLCG) or the number of used model vectors r_{TSVD} (for TSVD) work as regularization parameter. Small k_{max}/r_{TSVD} correspond to a large λ and vice versa.⁷

⁷In the following, the explicit regularization parameter λ is adopted. However, the considerations can easily be transferred to implicit methods.

Generally, a trade-off between data fit and model constraints has to be sought for both local and global regularization. In many cases, reasonable values for λ can be derived regarding the physical units of data and model or from experience (Loke and Barker, 1996b). However, it is worth thinking about how λ can be optimized so as not to lose valuable information, because the regularization parameter significantly influences the quality of the model.

In the following, methods for choosing the Lagrangian parameter are presented. For an overview the reader is referred to Vogel (2002) and Kilmer and O'Leary (2001). Farquharson and Oldenburg (2004) compared selection methods for the 1D-inversion of electromagnetic loop-loop data.

Discrepancy principle For known data covariances the inverse problem can be formulated as minimization of Φ_m with the equality constraint

$$\Phi_d = \sum_{i=1}^N \left| \frac{d_i - f_i(\mathbf{m})}{\epsilon_i} \right|^2 = \Phi_d^* \quad . \quad (2.21)$$

The value of λ then has to be chosen to obtain the target value $\Phi_d^* = N$. This can be done by searching the null point of an n-point approximation of the function $\Phi_d(\lambda) - N$, e.g. by a bisection scheme.

In non-linear inversion the target data misfit cannot be reached until a few iterations are applied. One solution to this problem is to define local target values for every iteration, e.g. a logarithmically spaced series down-going from $\Phi_d(\mathbf{m}^0)$ to N , which requires the knowledge of the final iteration number.

L-curve methods The plot of the data functional Φ_d and the model functional Φ_m for a varying λ is called L-curve due to its typical L-shaped form (Hansen and O'Leary, 1993). Figure 2.1 shows the L-curve of a typical linear ill-posed problem in log-log scale. As pointed out above, a low data functional Φ_d corresponds to high model functional Φ_m and vice versa. At each point a better data fit is penalized by decreasing model simplicity, whereas a "simpler" model results in poor data fit. In the center of the plot one can often observe a characteristic "corner" predicting an appropriate λ .

One very simple suggestion for the optimum is the point, where the curvature of the L-curve is maximized. The curvature c of the parametric function $(\psi_d(\lambda), \psi_m(\lambda))$ can be calculated by

$$c(\psi_d, \psi_m) = \frac{\ddot{\psi}_m \dot{\psi}_d - \dot{\psi}_m \ddot{\psi}_d}{(\dot{\psi}_d^2 + \dot{\psi}_m^2)^{3/2}} \quad , \quad (2.22)$$

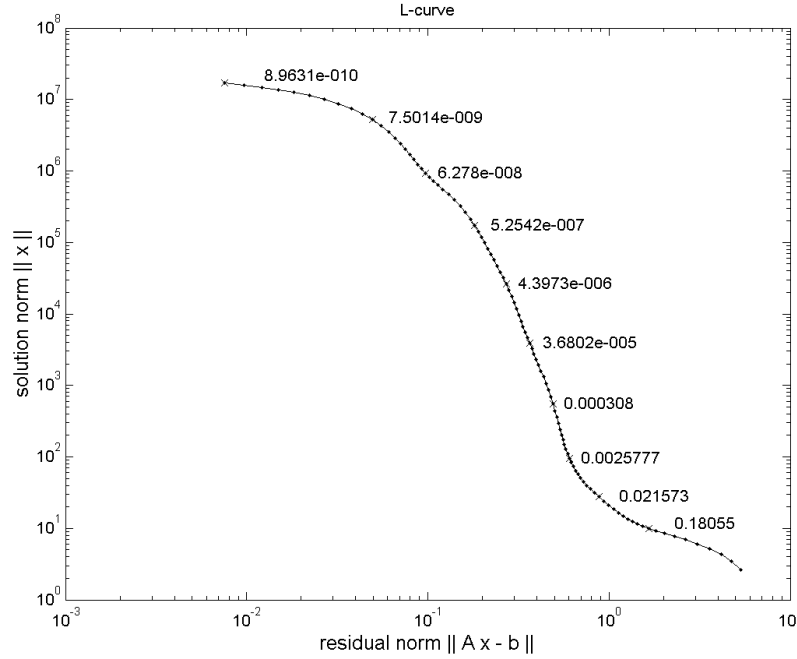


Figure 2.1: Typical L–curve as arising in ill-posed problems for the damped least squares solution of $\mathbf{Ax} = \mathbf{b}$

where the single and double dots represent the first and second order derivatives with respect to λ , respectively.

In case of local regularization ψ_m is the norm of the model update constraint $\|\mathbf{C}\Delta\mathbf{m}^k\|$, whereas ψ_d represents the residual norm of the linear subproblem $\|\mathbf{DS}\Delta\mathbf{m}^k - \mathbf{D}\Delta\mathbf{d}^k\|$. For a global regularizing scheme we use

$$\psi_m = \|\mathbf{C}(\mathbf{m}^k + \Delta\mathbf{m}^k - \mathbf{m}^0)\| \quad \text{and} \quad \psi_d = \|\mathbf{D}(\mathbf{f}(\mathbf{m}^k + \Delta\mathbf{m}^k) - \mathbf{d})\| \quad .$$

Figure 2.2 shows the L–curve and its curvature for a reasonably limited range of λ . The maximum curvature and thus the optimum λ is clearly defined. In order to economize the time-consuming forward calculations for all models $\mathbf{m}^k + \Delta\mathbf{m}^k(\lambda)$ the linearization

$$\mathbf{f}(\mathbf{m}^k + \Delta\mathbf{m}^k) \approx \mathbf{f}(\mathbf{m}^k) + \mathbf{S}\Delta\mathbf{m}^k$$

is used for the approximation of the L–curve. There are other ideas of detecting the "corner" as pointed out by Oraintara *et al.* (1999) obtaining similar results.

Generalized cross validation (GCV) The idea of the generalized cross validation is based on the philosophy, that if any row of the unregularized normal equations is

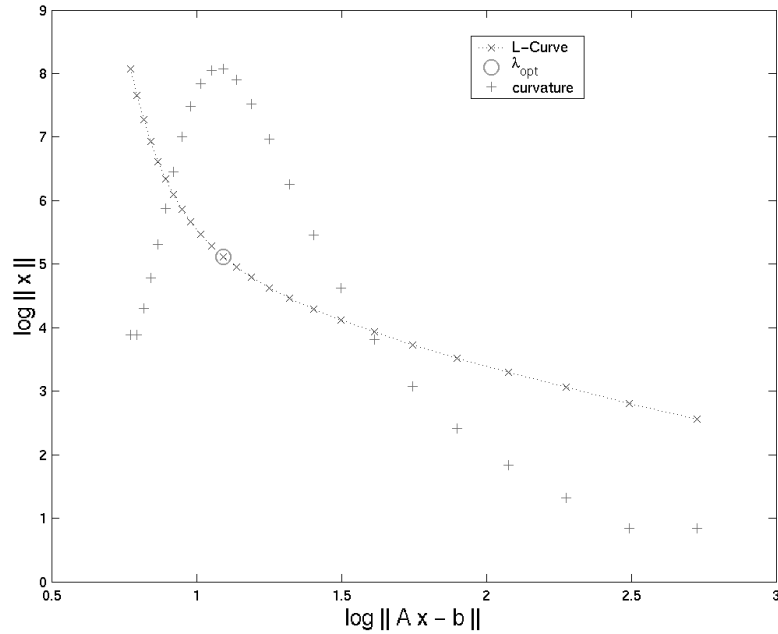


Figure 2.2: L-curve and its curvature for a reasonable interval of λ , the optimum regularization parameter is denoted by a circle

left out, the regularized solution will fit the data as well, independent of λ . For the least squares solution of $\mathbf{Ax} = \mathbf{b}$ the GCV function reads (Hansen, 1992)

$$g(\lambda) = \frac{\|\mathbf{Ax}_\lambda - \mathbf{b}\|_2^2}{\text{trace}(\mathbf{I} - \mathbf{AA}_\lambda^\dagger)^2} \quad (2.23)$$

The minimum point of the GCV function determines the optimized regularization. For the computation of the main diagonal of $\mathbf{AA}_\lambda^\dagger$ an SVD or bidiagonalization scheme is required. However, for the estimation of λ only a few iterations have to be carried out. For details of the GCV method see Vogel (2002) or Farquharson and Oldenburg (2004).

It was pointed out by Vogel (1996), that the L-curve method is not guaranteed to achieve plausible results and the GCV should be the method of choice. Nevertheless, by limiting λ to a reasonable range for most geophysical examples plausible models can be obtained by the maximum curvature method (Li and Oldenburg, 1999).

Changing λ during inversion

A well-discussed question in non-linear problems is that of changing the regularization parameter during inversion. Generally, the underlying minimization procedure (with

the use of globally operating constraints) treats λ as parameter, which merely depends on the number of data and their errors, the model discretization and the used constraints. All of these aspects are determined before the inversion, which pleads for constant regularization.

Different from global schemes, the use of local constraints leads to an independent linear sub-problem in every iteration. For each of them an appropriate λ can be determined, e.g., by the L-curve criterion. Many authors using local regularization schemes as Marquardt (1963), Loke and Barker (1996b) and Kemna (2000) discuss the use of decreasing λ beginning from a large starting value down to a minimum value. Also, Farquharson and Oldenburg (2004) applying a global smoothness constrained inversion use a cooling type schedule of decreasing λ . However, there are practical reasons such as the limited accuracy of the forward routines, which can entail interpretation failures. To prevent overshooting in the early iterations, a line search parameter can be applied to ensure convergence. Generally, the use of larger λ yields similar, but smoothed, structures with less magnitude, which represents an easy-to-control alternative to the line search procedure. However, the resolution analysis shows that the model is predicted by the final λ , which has to be chosen appropriately. The ultimate criterion is, whether the target value $\Phi_d^- N$ corresponding to $\chi^2 = 1$ is reached.

2.3 Equation Solvers for Inverse Problems

In every iteration step k , the linearized subproblem (2.16) to be solved reads

$$(\mathbf{S}^T \mathbf{D}^T \mathbf{D} \mathbf{S} + \lambda \mathbf{C}^T \mathbf{C}) \Delta \mathbf{m}^k = \mathbf{S}^T \mathbf{D}^T \mathbf{D} (\mathbf{d} - \mathbf{f}(\mathbf{m}^k)) \{-\lambda \mathbf{C}^T \mathbf{C} (\mathbf{m}^k - \mathbf{m}^0)\} \quad .$$

Note that for local regularization schemes the term within $\{ \dots \}$ vanishes. The Jacobian matrix $\mathbf{S} \in \mathbb{R}^{M \times N}$ is a full matrix, whereas the matrices $\mathbf{D} \in \mathbb{R}^{N \times N}$ and $\mathbf{C} \in \mathbb{R}^{M \times M}$ are generally sparse. The vectors have the dimensions $\mathbf{m}^* \in \mathbb{R}^M$ and $\mathbf{d}, \mathbf{f} \in \mathbb{R}^N$. The solution $\Delta \mathbf{m}^k$ depends on the matrices \mathbf{S} , \mathbf{D} , \mathbf{C} and the vectors $\Delta \mathbf{d}^k = \mathbf{d} - \mathbf{f}(\mathbf{m}^k)$ and $\delta \mathbf{m}^k = \mathbf{m}^k - \mathbf{m}^0$. The equation can be interpreted as solution of $\mathbf{A} \mathbf{x} = \mathbf{D} \mathbf{S} \Delta \mathbf{m}^k = \mathbf{D} \Delta \mathbf{d}^k = \mathbf{b}$ in a C-weighted least squares sense. In the following equation solvers are presented solving $\mathbf{A} \mathbf{x} = \mathbf{b}$ for \mathbf{x} in least squares senses.

For small-scale systems the normal equations can be solved by matrix inversion of the left hand side matrix, which is always possible for $\lambda > 0$, by appropriate methods like Gaussian elimination or QR decomposition. In multidimensional inversion, the number of model parameter and data are quite large, which prohibits the use of direct inversion from both computer time and memory usage point of view. Hence, an approximate solution is sought using iterative methods. To save computer memory, it is required that

- matrices of the form $\mathbf{A}^T \mathbf{A}$ are never formed explicitly and
- the transposed matrices are used without allocating additional memory.

2.3.1 Fixed Regularization Strength Solvers

The conjugate gradient method derived by Hestenes and Stiefel (1952) is widely used for iteratively solving large-scale systems of equations $\mathbf{Ax} = \mathbf{b}$. Since in every iteration only one matrix vector product has to be calculated, it is primarily used for sparse \mathbf{A} as arising in the discretization of partial differential equations. However, conjugate gradients are not restricted to sparse systems and can also be applied to the normal equations (Shewchuk, 1994). Although the numerical effort is increased for full matrices as the Jacobian represents, conjugate gradients techniques prove to be very efficient⁸.

Unregularized normal equations

Applying the conjugate gradient method to the normal equations

$$\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$$

yields Algorithm 4, called CGLS⁹ (Press *et al.*, 1992). Two matrix vector products are

Algorithm 4 Conjugate Gradients, applied to least squares (CGLS)

```

 $k = 0$ 
 $\mathbf{p}_0 = \mathbf{r}_0 = \mathbf{A}^T (\mathbf{b} - \mathbf{Ax}_0)$ 
while  $\|r_k\| > acc \|r_0\|$  do
   $\alpha_{k+1} = \frac{\|\mathbf{r}_k\|^2}{\|\mathbf{A}^T \mathbf{Ap}_k\|^2}$  {A-conjugation}
   $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_{k+1} \mathbf{p}_k$  {updating solution}
   $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_{k+1} \mathbf{A}^T \mathbf{Ap}_k$  {updating the residual}
   $\beta_{k+1} = \frac{\|\mathbf{r}_{k+1}\|^2}{\|\mathbf{r}_k\|^2}$ 
   $\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_{k+1} \mathbf{p}_k$ 
   $k = k + 1$ 
end while

```

required in every iteration. At first, a vector $\mathbf{q}_k = \mathbf{Ap}_k$ is calculated. Then, \mathbf{q}_k has to be multiplied with the transpose of \mathbf{A} . In order to avoid the storage of \mathbf{A}^T , the product is transformed such that $\mathbf{A}^T \mathbf{q}_k = (\mathbf{q}_k^T \mathbf{A})^T$.

⁸It has to be noted, that CG-based equation solvers are to be distinguished from the NLCC minimization method (see section 2.1), even though based on the same principles.

⁹The abbreviation stands for conjugate gradients least squares solver.

Note that all existent vectors and scalars can be overwritten it every iteration and thus no additional memory allocation is necessary. The only exception is to save the scalar $\|\mathbf{r}_k\|^2$, which is needed for the computation of β_{k+1} , before the residual vector \mathbf{r} is updated. The accuracy acc defines the stopping criterion with regard to the initial residual $\|r_0\|$.

As noted before, low-frequency components of the solution tend to converge faster than high-frequency parts in Krylov subspace methods. This can be used for an implicit regularization algorithm called truncated least squares (TLS). Doing so, the stopping criterion is replaced by ($k \leq k_{max}$), where k_{max} is treated as regularization parameter.

Regularized normal equations

Assume a Marquardt type of regularization resulting in the damped normal equations

$$(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})\mathbf{x} = \mathbf{A}^T \mathbf{b} \quad .$$

The additional term affects both the gradient directions \mathbf{p}_k and the coefficients α_k . Let \mathbf{z} be the residual of the basis equation $\mathbf{z} = \mathbf{b} - \mathbf{A}\mathbf{x}$. As in Algorithm 4, \mathbf{r} denotes the residual of the equation to be solved

$$\mathbf{r} = \mathbf{A}^T (\mathbf{b} - \mathbf{A}\mathbf{x}) - \lambda \mathbf{I}\mathbf{x} = \mathbf{A}^T \mathbf{z} - \lambda \mathbf{x} \quad .$$

Result of the changes yields Algorithm 5, called CGLSI.

Algorithm 5 Conjugate Gradients, applied to damped least squares (CGLSI)

```

k = 0
z0 = b - Ax0
p0 = r0 = ATz - λx0
while ||rk|| > acc ||r0|| do
  qk+1 = Apk
  αk+1 =  $\frac{\|\mathbf{r}_k\|^2}{\|\mathbf{q}_k\|^2 + \lambda \|\mathbf{p}_k\|^2}$ 
  xk+1 = xk + αk+1pk
  zk+1 = zk - αk+1qk+1
  rk+1 = Azk+1 - λxk+1
  βk+1 =  $\frac{\|\mathbf{r}_{k+1}\|^2}{\|\mathbf{r}_k\|^2}$ 
  pk+1 = rk+1 + βk+1pk
  k = k + 1
end while

```

If the constraint matrix \mathbf{C} differs from the identity matrix, a different computation of the residual \mathbf{r} and the coefficient α is needed. Furthermore, the data weighting matrix

can easily be introduced by additional matrix-vector multiplications, which are very fast for sparse \mathbf{D} . The solution to

$$(\mathbf{A}^T \mathbf{D}^T \mathbf{D} \mathbf{A} + \lambda \mathbf{C}^T \mathbf{C}) \mathbf{x} = \mathbf{A}^T \mathbf{D}^T \mathbf{D} \mathbf{b} \{ -\lambda \mathbf{C}^T \mathbf{C} \delta \mathbf{m}^k \}$$

is described in Algorithm 6, called CGLSCD.

Algorithm 6 Conjugate Gradients, weighted least squares (CGLSCD)

```

k = 0
z0 = D(b - Ax0)
p0 = r0 = ATDTz{-λCTCx0}
while ||rk|| > acc ||r0|| do
  qk+1 = DA pk
  αk+1 =  $\frac{\|\mathbf{r}_k\|^2}{\mathbf{q}_k^T \mathbf{q}_k + \lambda \mathbf{p}_k^T \mathbf{C}^T \mathbf{C} \mathbf{p}_k}$ 
  xk+1 = xk + αk+1 pk
  zk+1 = zk - αk+1 qk+1
  rk+1 = DA zk+1 {-λCTCxk+1}
  βk+1 =  $\frac{\|\mathbf{r}_{k+1}\|^2}{\|\mathbf{r}_k\|^2}$ 
  pk+1 = rk+1 + βk+1 pk
  k = k + 1
end while

```

It has practical benefits to apply the solution to the model update via a transformation matrix \mathbf{P} such that $\Delta \mathbf{m}^k = \mathbf{P} \mathbf{x}$. For the simplest case \mathbf{P} equals the identity matrix ($\mathbf{x} = \Delta \mathbf{m}^k$). \mathbf{P} can be used to change the model parameter to be inverted without rearranging the sensitivity matrix in the following ways:

- By deleting rows of \mathbf{P} certain model parameters can be excluded from being changed. This can be the case for fixing parameter values or for cells with low sensitivities.
- If the model geometry is described by a finite difference grid, usually badly resolved small cells are produced far from the sensors. Then, by adding rows of \mathbf{P} , cells can be combined to obtain model parameters with better resolution.
- The matrix \mathbf{P} can also be used as weighting function. An example may be a-priori knowledge from well logging data, the weights decrease with the distance to the borehole.

Consequently, the calculation of the residual vector \mathbf{r}_{k+1} and the coefficient α_{k+1} changes to

$$\mathbf{r}_{k+1} = \mathbf{PDAz}_{k+1} \{-\lambda \mathbf{P}^\dagger \mathbf{C}^T \mathbf{C} \mathbf{x}_{k+1}\} \text{ and}$$

$$\alpha_{k+1} = \frac{\|\mathbf{r}_k\|^2}{\mathbf{q}_k^T \mathbf{q}_k + \lambda \mathbf{P}^\dagger \mathbf{p}_k^T \mathbf{P}^T \mathbf{C}^T \mathbf{C} \mathbf{P} \mathbf{p}_k} .$$

The matrix \mathbf{P}^\dagger is the pseudo-inverse of \mathbf{P} , which can be easily obtained for the above described variants of row deletions and combination. The algorithm is then denoted by the term CGLSCDP, which stands for **c**onjugate **g**radients **l**east **s**quares solver using **c**onstraints, **d**ata weighting and **p**arametric mapping.

When solving a system of equations using conjugate gradients the question arises, can the use of preconditioners minimize the computational effort? Since incomplete decompositions of LU or Cholesky style are computationally prohibitive due to memory limitations, preconditioners without additional storage, e.g. of Jacobi type, remain. They can be easily installed computing the individual main diagonal elements of $\mathbf{S}^T \mathbf{S}$.

2.3.2 Multiple Regularization Strength Solvers

If the optimum value of the regularization parameter λ is not known - which is the case for most problems at least in the first iteration - it has to be estimated by appropriate methods, e.g. the L-curve criterion. For the construction of the L-curve, solutions for many λ_i have to be obtained. Since reasonable values extend over a range of several decades, a lot of computational work has to be done if the results are obtained independently.

All the solutions are expected to show similar model structures as indicated by the data. Models corresponding to large regularization parameters are usually of lower magnitude and smoother than results obtained with small λ 's. Considering the regularizing properties of the conjugate gradients leads directly to the idea of subsequently using the results of large λ as starting vector for the next smaller one. Beginning with the largest regularization parameter the numerical effort can be significantly reduced. An important advantage of this procedure lies in the fact, that the L-curve criterion can be applied after each run and once the maximum curvature is found, the computation for smaller λ can be omitted.

The left hand side matrices differ for various λ_i merely by multiples of the quadratic constraint matrix $\mathbf{C}^T \mathbf{C}$, which is the identity matrix for the ridge regression method. The main computational effort resides in the multiplication with the full matrix \mathbf{A} , which is the same for all "shifted" systems corresponding to different λ_i . This idea of a parallel computation of all systems was pointed out by Frommer and Maass (1999). As each left hand side matrix consists of a common and a separate part, the different gradients consist of common and separated parts. The solution, Algorithm 7 is based on an algorithm presented by Saad (1996), which is a combination of conjugate gra-

dients and Lanczos bidiagonalization methods yielding identical results compared to Algorithm 5.

Algorithm 7 CG/Lanczos least squares solver (CGLAN)

```

k = 0
r0 = b - Ax
β0 = ||r0||, v0 = r0/β0, p0 = r0
v-1 = 0, σ0 = β0, ω-1 = 0, γ-1 = 1
while ||σk|| > acc ||r0|| do
  q = Av
  δk = ||Axk||2 {start of Lanczos part}
  ṽk+1 = ATAv - δkvk - βkvk-1
  βk+1 = ||ṽk+1||
  vk+1 =  $\frac{\tilde{v}_{k+1}}{\beta_{k+1}}$ 
  αk+1 =  $\frac{\alpha_k}{(\delta_k + \lambda)\alpha_k - \omega_{k-1}}$  {end of Lanczos part}
  ωk+1 = (βk+1αk+1)2 {start of CG part}
  σk+1 = -βk+1αk+1σk
  xk+1 = xk + αk+1pk
  rk+1 = σk+1vk+1
  pk+1 = rk+1 + ωk+1pk {end of CG part}
end while

```

Note that the Lanczos part represents the solution of the unregularized normal equation, which is independent of λ . It provides that part of the solution, which is common to all systems and which includes the main computational effort, two matrix-vector multiplications per iteration.

The CG part of the algorithm depends on the concrete value for λ and has to be calculated separately for each system. Also, convergence has to be investigated separately and the CG part is applied to the systems that did not converge yet. Since that part consists of vector-vector and scalar-vector operations the main computation time is determined by the Lanczos part. Since the condition number generally decreases while increasing the stabilization effect, the number of iterations used for convergence is defined by the smallest λ . Thus, the minimum λ has to be chosen carefully to save computer time. As a result, a set of solutions corresponding to various λ is obtained with the effort of one single CGLS algorithm. However, the method cannot be generalized to arbitrary constraints up to now.

2.4 SVD and GSVD

2.4.1 The Singular Value Decomposition

The singular value decomposition is widely used as a powerful tool for obtaining insight into least squares problems of the form $\mathbf{Ax} = \mathbf{b}$ for non-quadratic matrices \mathbf{A} . It can be easily derived by eigenvalue analysis of the matrices $\mathbf{A}^T\mathbf{A}$ and \mathbf{AA}^T . For detailed information see, e.g., Menke (1989) or Golub and van Loan (1996).

At first, let us rewrite equations (2.16) and (2.18) symbolically

$$\Delta\mathbf{m} = \hat{\mathbf{S}}^\dagger \mathbf{D} \Delta\mathbf{d} \{-\mathbf{C}^\dagger \mathbf{C}(\mathbf{m}^k - \mathbf{m}^0)\} \quad \text{with} \quad (2.24)$$

$$\hat{\mathbf{S}}^\dagger = (\hat{\mathbf{S}}^T \hat{\mathbf{S}} + \lambda \mathbf{C}^T \mathbf{C})^{-1} \hat{\mathbf{S}}^T \quad \text{and} \quad \mathbf{C}^\dagger = \lambda (\hat{\mathbf{S}}^T \hat{\mathbf{S}} + \lambda \mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T$$

The introduction of the generalized inverse matrices $\hat{\mathbf{S}}^\dagger$ and \mathbf{C}^\dagger appears reasonable due to the symmetry between the error-weighted sensitivity $\hat{\mathbf{S}} = \mathbf{DS}$ and the constraint matrix \mathbf{C} . Note that $\hat{\mathbf{S}}^\dagger \hat{\mathbf{S}} + \mathbf{C}^\dagger \mathbf{C} = \mathbf{I}$.

For every matrix $\mathbf{A} \in \mathbb{R}^{N \times M}$ two orthonormal matrices $\mathbf{U} \in \mathbb{R}^{N \times N}$ and $\mathbf{V} \in \mathbb{R}^{M \times M}$ can be found transforming \mathbf{A} into a diagonal matrix $\mathbf{\Lambda}$

$$\mathbf{U}^T \mathbf{A} \mathbf{V} = \mathbf{\Lambda} \quad . \quad (2.25)$$

The matrix $\mathbf{V} = (\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_M)$ contains the model eigenvectors \mathbf{v}_i , whereas the data eigenvectors \mathbf{u}_i are assembled in the matrix $\mathbf{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_N)$. The vectors \mathbf{v}_i are mutually orthogonal and of length 1. The same holds for the \mathbf{u}_i such that

$$\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I}_N \quad \text{and} \quad \mathbf{V}^T \mathbf{V} = \mathbf{V} \mathbf{V}^T = \mathbf{I}_M \quad .$$

The matrix $\mathbf{\Lambda} \in \mathbb{R}^{N \times M}$ has the same size as \mathbf{A} and contains the singular values s_i as the main diagonal entries

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{with} \quad \mathbf{\Lambda}_r = \begin{pmatrix} s_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_r \end{pmatrix} \quad .$$

The s_i are usually ordered decreasingly and serve as weights for the corresponding data and model vectors. By the number r of nonzero singular values (the rank of \mathbf{A}) the null spaces for the model \mathbf{V}_0 and the data \mathbf{U}_0 can be separated

$$\mathbf{V} = (\mathbf{V}_r \ \mathbf{V}_0) \quad \text{and} \quad \mathbf{U} = (\mathbf{U}_r \ \mathbf{U}_0) \quad ,$$

where \mathbf{V}_r and \mathbf{U}_r are the first r columns of the model and data matrices, respectively. Since \mathbf{U} and \mathbf{V} are invertible, it can be found that

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{V}_r^T \quad . \quad (2.26)$$

Note that the null spaces have no contribution to \mathbf{A} . Since all matrices are easily invertible, a natural inverse¹⁰ of \mathbf{A} can be found

$$\mathbf{A}^\dagger = \mathbf{V}_r \mathbf{\Lambda}_r^{-1} \mathbf{U}_r^T \quad , \quad (2.27)$$

which is known as Moore-Penrose inverse or pseudo-inverse. The multiplication by \mathbf{A}^\dagger as solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ is then defined by

$$\mathbf{x} = \mathbf{A}^\dagger \mathbf{b} = \sum_{i=1}^r \frac{\mathbf{u}_i^T \mathbf{b}}{s_i} \mathbf{v}_i \quad (2.28)$$

and turns out to be the least squares solution. The weighting coefficients for the individual model vectors are determined dividing the dot product of \mathbf{b} and the corresponding data vector \mathbf{u}_i by the singular value s_i .

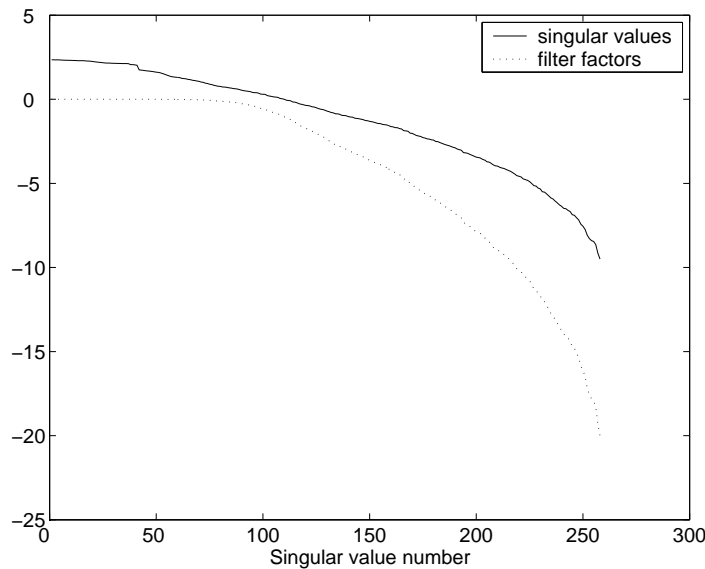


Figure 2.3: Singular values and filter factors for a typical ill-posed problem, note that the filter factors decrease faster

Figure 2.3 shows the singular value spectrum as typically arising for ill-posed problems. The spread of several decades between the largest and smallest singular value is an indication for the ill-posedness. Taking a closer look at the model vectors it can be observed that \mathbf{v}_i for small i show relatively simple characteristics, whereas model vectors connected to small singular values tend to show high-frequency oscillations (Friedel, 2000). If the vector \mathbf{b} in equation (2.28) is contaminated with noise, the projections onto the data vectors do not vanish. Due to the division by s_i the model vectors for small singular values are amplified resulting in a strongly structured model. Usually

¹⁰It is sometimes called "the" generalized inverse, even though many generalized inverses can be defined, as we will see.

this model fits the data (including noise) well, but lacks plausibility. Hence, a regularization procedure has to be introduced. The TSVD inverse as implicit regularization limits the summation up to an arbitrary pseudo-rank r_{TSVD} , from that on the singular values are treated as zero. A small r_{TSVD} value corresponds to large regularization and vice versa. In practice, a first choice of r_{TSVD} is obtained by assuming a noise level ϵ and choosing the s_i closest to ϵs_1 . In small parameter inversion the singular value spectrum is investigated for a gap or corner to detect r_{TSVD} .

Another way to reduce the effect of near-zero singular values from exploding is to introduce damping using filter factors f_i

$$\mathbf{A}^\dagger = \mathbf{V}_r \text{diag} \left(\frac{f_i}{s_i} \right) \mathbf{U}_r^T \quad . \quad (2.29)$$

Using an explicit regularization scheme as depicted in section 2.2 the filter factors for the solution of

$$\mathbf{x} = \mathbf{A}^\dagger \mathbf{b} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}$$

can be calculated by (Hansen, 1992)

$$f_i = \frac{s_i^2}{s_i^2 + \lambda} \quad . \quad (2.30)$$

Thus, the damped least squares solution reads

$$\mathbf{x} = \mathbf{A}^\dagger \mathbf{b} = \sum_{i=1}^r \frac{s_i \mathbf{u}_i^T \mathbf{b}}{s_i^2 + \lambda} \mathbf{v}_i \quad .$$

The ratio $f_i/s_i = s_i/(s_i^2 + \lambda)$ works as transform function for the individual model vectors. Note that the TSVD inversion can also be formulated by equation (2.29) using the filter factors

$$f_i = 1 \quad \forall \quad i \leq r_{TSVD} \quad \text{and} \quad f_i = 0 \quad \forall \quad i > r_{TSVD} \quad .$$

Hence, the transform function for TSVD is $1/s_i \forall i$ and 0 elsewhere. Figure 2.4 presents the filter factors and transform functions for TSVD and Tikhonov regularization. The two λ and r_{TSVD} are chosen such that the results are comparable. For the Tikhonov regularization the transform function is maximized at $i \approx 90 \dots 100$. Larger or smaller values for λ shift the maximum towards lower or higher singular value numbers, respectively. In contrast, the largest contributions of the TSVD inverse are provided by the model vectors directly below r_{TSVD} , resulting in a more structured model.

The SVD can be used to calculate the model update for TSVD as well as Marquardt-type regularization using $\mathbf{A} = \mathbf{D}\mathbf{S} = \hat{\mathbf{S}}$. Once the decomposition has been carried out,

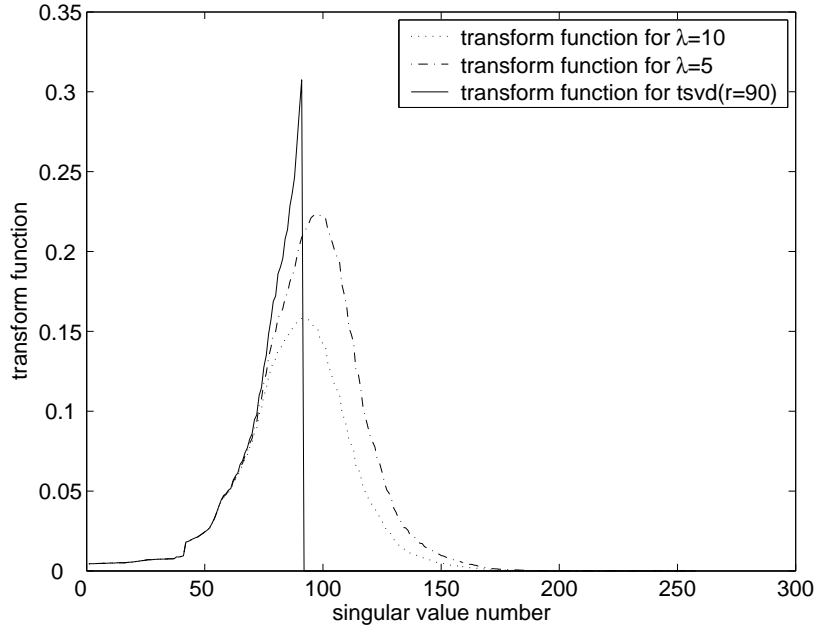


Figure 2.4: Transform function of Tikhonov ($\lambda = 10$, $\lambda = 5$) and TSVD ($r_{TSVD} = 90$) regularization for the sample problem of Figure 2.3

solutions for arbitrary λ can be obtained with less computational effort. This proves to be particularly efficient if the regularization parameter is completely unknown and the SVD provides data and model functional to choose λ by appropriate methods. When the model constraints are formulated using $\mathbf{C} \neq \mathbf{I}$ the generalized inverse can not be represented by SVD filter factors. If \mathbf{C} is easily invertible, the problem can be transferred to standard form using $\mathbf{A} = \hat{\mathbf{S}}\mathbf{C}^{-1}$.

2.4.2 The Generalized SVD

Another possibility for solving constrained least squares problems is the application of the generalized singular value decomposition introduced by van Loan (1976). For $\mathbf{A} \in \mathbb{R}^{N \times M}$ and $\mathbf{B} \in \mathbb{R}^{P \times M}$, there exist orthogonal matrices $\mathbf{U} \in \mathbb{R}^{N \times N}$, $\mathbf{W} \in \mathbb{R}^{M \times M}$ and an invertible matrix $\mathbf{V} \in \mathbb{R}^{M \times M}$ diagonalizing \mathbf{A} and \mathbf{B} such that

$$\mathbf{U}^T \mathbf{A} \mathbf{V} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{with } \mathbf{\Sigma} = \text{diag}(\sigma_i) \text{ and}$$

$$\mathbf{W}^T \mathbf{B} \mathbf{V} = \mathbf{M} = \text{diag}(\mu_i) \quad .$$

The paired singular values for \mathbf{A} and \mathbf{B} are denoted by σ_i and μ_i , respectively, and equal $\mu_i^2 + \sigma_i^2 = 1 \forall i$. The GSVD is computed by QR decomposition of the composite matrix of \mathbf{A} and \mathbf{B} and a subsequent CS decomposition, which involves a large numerical effort compared to the SVD.

The solution to the problem

$$(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{B}^T \mathbf{B}) \mathbf{x} = \mathbf{A}^T \mathbf{b}$$

can be described by means of the GSVD (Golub and van Loan, 1996)

$$\mathbf{x} = \sum_{i=1}^P \frac{\sigma_i \mathbf{u}_i^T \mathbf{b}}{\sigma_i^2 + \lambda \mu_i^2} \mathbf{v}_i + \sum_{i=P+1}^M \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i \quad . \quad (2.31)$$

For a quadratic stabilizer \mathbf{B} holds $P = M$ and the latter term disappears. Hence, equation (2.31) can be written

$$\mathbf{x} = \sum_{i=1}^M \frac{\sigma_i \mathbf{u}_i^T \mathbf{b}}{\sigma_i^2 + \lambda \mu_i^2} \mathbf{v}_i = \sum_{i=1}^M \frac{1}{\sigma_i} \frac{\gamma_i^2 \mathbf{u}_i^T \mathbf{b}}{\gamma_i^2 + \lambda} \mathbf{v}_i = \mathbf{A}^\dagger \mathbf{b} \quad , \quad (2.32)$$

where $\gamma_i = \sigma_i / \mu_i$ are the generalized singular values. Using $\mathbf{A} = \hat{\mathbf{S}}$ and $\mathbf{B} = \mathbf{C}$, the generalized inverse of $\hat{\mathbf{S}}$ can be formed as follows

$$\hat{\mathbf{S}}^\dagger = \mathbf{V}_r \text{diag} \left(\frac{f_i}{\sigma_i} \right) \mathbf{U}_r^T \quad , \quad (2.33)$$

which is identical to (2.29) except that σ_i instead of s_i is used. The ratio f_i / σ_i serves as transfer function for the model vectors. Note that the filter factors are defined the same way as for Tikhonov regularization replacing the singular values s_i by the generalized singular values γ_i

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \mu_i^2 \lambda} = \frac{\gamma_i^2}{\gamma_i^2 + \lambda} \quad .$$

Using $\mathbf{A} = \mathbf{C}$ and $\mathbf{B} = \hat{\mathbf{S}}$ and interchanging the corresponding matrices and singular vectors, it can be found that

$$\mathbf{C}^\dagger = \lambda \mathbf{V}_r \text{diag} \left(\frac{g_i}{\mu_i} \right) \mathbf{W}_r^T \quad , \text{ where}$$

$$g_i = \frac{1/\gamma_i^2}{1/\gamma_i^2 + 1/\lambda} = \frac{\lambda}{\lambda + \gamma_i^2} = 1 - f_i \quad , \text{ resulting in}$$

$$\mathbf{C}^\dagger = \lambda \mathbf{V}_r \text{diag} \left(\frac{1 - f_i}{\gamma_i} \right) \mathbf{W}_r^T \quad .$$

2.5 Resolution Analysis

Product of the inversion of data is a model, a simplified concept of the reality, which is rarely true. We have clearly in mind, that it is one of many possible models fitting the data within errors ¹¹. With this in mind, two important questions arise:

1. How well can we trust the model parameter values and geometry of the detected structures?
2. Which part of the model is determined by the data, and what model characteristics are a product of the constraints?

The first question is discussed widely, but mostly in terms of linear inversion. Oldenburg and Li (1999) try to find answers to the second question by a practical approach defining the depth of interest index. Recently a few approaches describing non-linear resolution in geophysics were published (Alumbaugh and Newman, 2000; Friedel, 2003; Stummer *et al.*, 2004).

2.5.1 Resolution of Linear Problems

For linear inversion theory, resolution has been described in detail. The reader may be referred to the books of Menke (1989) and Meju (1994b). Assume the data \mathbf{d} are equated by a product of the kernel matrix \mathbf{F} with the true model \mathbf{m}^{true} and noise vector \mathbf{n}

$$\mathbf{d} = \mathbf{F}\mathbf{m}^{true} + \mathbf{n} \quad . \quad (2.34)$$

The application of an explicit inversion scheme including regularization defines a generalized inverse \mathbf{F}^\dagger , which is used to find a model estimate

$$\mathbf{m}^{est} = \mathbf{F}^\dagger \mathbf{d} \quad . \quad (2.35)$$

Replacing \mathbf{d} from equation (2.34) we obtain

$$\mathbf{m}^{est} = \mathbf{F}^\dagger \mathbf{F} \mathbf{m}^{true} + \mathbf{F}^\dagger \mathbf{n} = \mathbf{R}^M \mathbf{m}^{true} + \mathbf{F}^\dagger \mathbf{n} \quad (2.36)$$

The matrix $\mathbf{R}^M = \mathbf{F}^\dagger \mathbf{F}$ combining forward and inverse mapping is referred to as resolution matrix. It serves as a kernel function transferring the reality into our model estimate, which is connected to the concept of Backus and Gilbert (1968). Ideally, \mathbf{R}^M would be the identity matrix, which corresponds to perfect resolution. However,

¹¹Actually, it is the one that minimizes our model constraints.

if we try to find a generalized inverse \mathbf{F}^\dagger resulting in a resolution matrix close to the identity, the solution is generally dominated by noise effects. Hence, regularization can be understood as a trade-off between the two terms on the right hand side of equation (2.36): Representation of the true model by $\mathbf{R}^M \mathbf{m}^{true}$ vs. effects of noise $\mathbf{F}^\dagger \mathbf{n}$.

2.5.2 Resolution of Non-linear Problems

In non-linear inversion the data vector is considered as the forward response \mathbf{f} of the true model \mathbf{m}^{true} , which is contaminated by a noise vector \mathbf{n}

$$\mathbf{d} = \mathbf{f}(\mathbf{m}^{true}) + \mathbf{n} \quad . \quad (2.37)$$

Assuming that in the k^{th} iteration the model \mathbf{m}^k is already close to the true model, a linearized Taylor expansion of \mathbf{f} at the point \mathbf{m}^k yields (Friedel, 2003)

$$\mathbf{d} = \mathbf{f}(\mathbf{m}^{true}) + \mathbf{n} = \mathbf{f}(\mathbf{m}^k) + \mathbf{S}(\mathbf{m}^{true} - \mathbf{m}^k) + \mathbf{n} \quad . \quad (2.38)$$

If explicit local regularization is considered, the iteration scheme reads

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \hat{\mathbf{S}}^\dagger \mathbf{D}(\mathbf{d} - \mathbf{f}(\mathbf{m}^k)) \quad . \quad (2.39)$$

Replacing $\mathbf{d} - \mathbf{f}(\mathbf{m}^k)$ from equation (2.38) and setting $\mathbf{m}^{est} = \mathbf{m}^{k+1}$ we obtain in coincidence with Friedel (2003)

$$\begin{aligned} \mathbf{m}^{est} &= \mathbf{m}^k + \hat{\mathbf{S}}^\dagger \mathbf{D} \mathbf{S} (\mathbf{m}^{true} - \mathbf{m}^k) + \hat{\mathbf{S}}^\dagger \mathbf{D} \mathbf{n} \\ &= \hat{\mathbf{S}}^\dagger \hat{\mathbf{S}} \mathbf{m}^{true} + (\mathbf{I} - \hat{\mathbf{S}}^\dagger \hat{\mathbf{S}}) \mathbf{m}^k + \hat{\mathbf{S}}^\dagger \mathbf{D} \mathbf{n} \\ &= \mathbf{R}^M \mathbf{m}^{true} + (\mathbf{I} - \mathbf{R}^M) \mathbf{m}^k + \hat{\mathbf{S}}^\dagger \mathbf{D} \mathbf{n} \quad . \end{aligned} \quad (2.40)$$

In analogy to linear inversion the resolution matrix \mathbf{R}^M serves as transformation from reality. Again, a term representing the influence of noise occurs. The term $\mathbf{I} - \mathbf{R}^M$ can be interpreted as a complement to \mathbf{R}^M . It points out, that in regions of missing resolution (where $R_{ii}^M \ll 1$) the model stays unchanged from \mathbf{m}^k . In fact, the resolution is linear regarding the model update $\Delta \mathbf{m}$. However, the effect of the starting model is not clear.

On the contrary, an explicit global regularization scheme including the minimization of $\|\mathbf{C}(\mathbf{m} - \mathbf{m}^0)\|$ throughout the inversion process has the iteration scheme

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \hat{\mathbf{S}}^\dagger \mathbf{D}(\mathbf{d} - \mathbf{f}(\mathbf{m}^k)) - \mathbf{C}^\dagger \mathbf{C}(\mathbf{m}^k - \mathbf{m}^0) \quad . \quad (2.41)$$

Replacing $\mathbf{d} - \mathbf{f}(\mathbf{m}^k)$ from equation (2.38) and setting $\mathbf{m}^{est} = \mathbf{m}^{k+1}$ yields

$$\begin{aligned} \mathbf{m}^{est} &= \mathbf{m}^k + \hat{\mathbf{S}}^\dagger \mathbf{S} \mathbf{D} (\mathbf{m}^{true} - \mathbf{m}^k) - \mathbf{C}^\dagger \mathbf{C} (\mathbf{m}^k - \mathbf{m}^0) + \hat{\mathbf{S}}^\dagger \mathbf{D} \mathbf{n} \\ &= \mathbf{m}^k + \hat{\mathbf{S}}^\dagger \hat{\mathbf{S}} \mathbf{m}^{true} - (\hat{\mathbf{S}}^\dagger \hat{\mathbf{S}} + \mathbf{C}^\dagger \mathbf{C}) \mathbf{m}^k + \mathbf{C}^\dagger \mathbf{C} \mathbf{m}^0 + \hat{\mathbf{S}}^\dagger \mathbf{D} \mathbf{n} \\ &= \mathbf{R}^M \mathbf{m}^{true} + (\mathbf{I} - \mathbf{R}^M) \mathbf{m}^0 + \hat{\mathbf{S}}^\dagger \mathbf{D} \mathbf{n} \quad . \end{aligned} \quad (2.42)$$

Equation (2.42) is identical to the resolution for local regularization except that \mathbf{m}^k is replaced by \mathbf{m}^0 . The difference is quite remarkable: It shows, that the model estimate is filled up with the starting model (and not the model of the preceding iteration) at regions of missing resolution. The matrix $\mathbf{I} - \mathbf{R}^M$ can be interpreted as constraint resolution matrix and describes, how \mathbf{m}^0 is mapped into reality.

The resolution equation (2.42) is directly linked to the idea of Oldenburg and Li (1999) defining the depth-of-ivestigation index. The DOI index reveals, to what degree the model parameters are determined by the starting model. The identical inversion scheme is applied twice using two different starting models \mathbf{m}_A^0 and \mathbf{m}_B^0 , e.g. two differing half-spaces. The first and the last term of equation (2.42) are expected to be invariant with respect to \mathbf{m}^0 . Hence, the difference reads $\mathbf{m}_A - \mathbf{m}_B = (\mathbf{I} - \mathbf{R}^M)(\mathbf{m}_A^0 - \mathbf{m}_B^0)$, which is small for good resolution ($\mathbf{R}^M \approx \mathbf{I}$) and large for poor resolution, as the DOI index depicts.

The result of a global regularization scheme is considerably independent of the way the iteration took, whereas all iteration stages strongly affect the final model in local regularization. From the resolution point of inversion the author favors the global minimization to keep control on how the model is determined. Particularly if a-priori information is available, the model is restricted to lie in the neighborhood of \mathbf{m}^0 .

In opposition to equation (2.42), combining forward and inverse matrix yields the data importance matrix \mathbf{R}^D

$$\begin{aligned} \mathbf{f}(\mathbf{m}^{k+1}) &= \mathbf{f}(\mathbf{m}^k) + \mathbf{S} (\mathbf{S}^\dagger \mathbf{D} (\mathbf{f}(\mathbf{m}^{true}) + \mathbf{n} - \mathbf{f}(\mathbf{m}^k)) - \mathbf{C}^\dagger \mathbf{C} (\mathbf{m}^k - \mathbf{m}^0)) \\ \mathbf{f}(\mathbf{m}^{est}) &= \mathbf{R}^D \mathbf{f}(\mathbf{m}^{true}) + \mathbf{R}^D \mathbf{n} + (\mathbf{I} - \mathbf{R}^D) \mathbf{f}(\mathbf{m}^0) \quad \text{with} \quad \mathbf{R}^D = \mathbf{S} \mathbf{S}^\dagger \quad . \end{aligned} \quad (2.43)$$

Besides the resolution matrix many authors (Meju, 1994a; Alumbaugh and Newman, 2000) investigate the (a-posteriori) model covariance matrix MCM^{12} . Using the rule of variance propagation $\text{Cov}(\mathbf{A}\mathbf{x} + \mathbf{b}) = \mathbf{A} \text{Cov}(\mathbf{x}) \mathbf{A}^T$ the MCM can be derived directly from equation (2.42). Since the true model has no variance and the noise has the variance of the data

$$\text{Cov}(\mathbf{n}) = \text{Cov}(\mathbf{d}) = \mathbf{D}^{-2} \quad ,$$

¹²To distinguish the MCM from \mathbf{C}_m , the terms "a-priori" and "a-posteriori" depict the use as constraint matrix and resolution measure, respectively.

the model covariance matrix reads

$$\begin{aligned} MCM &= (\mathbf{I} - \mathbf{R}^M)\text{Cov}(\mathbf{m}^0)(\mathbf{I} - \mathbf{R}^M)^T + \hat{\mathbf{S}}^\dagger \mathbf{D} \text{Cov}(\mathbf{d})(\hat{\mathbf{S}}^\dagger \mathbf{D})^T \\ &= (\mathbf{I} - \mathbf{R}^M)\text{Cov}(\mathbf{m}^0)(\mathbf{I} - \mathbf{R}^M)^T + \hat{\mathbf{S}}^\dagger (\hat{\mathbf{S}}^\dagger)^T \quad . \end{aligned} \quad (2.44)$$

It consists of two parts: The first one results from the variance of the starting model, which is not known and therefore often omitted. The second part originates from the error propagation throughout the inversion process.

Regarding Table 2.1, resolution matrices can be obtained for all regularization types except implicit global schemes, e.g. the NLCG method. The assumption for the validity of resolution matrices is that the model is "close to the true model". However, in best case we know the estimated model's forward response to be close to the one of the true model. Thus, the resolution is linked to the estimated model and only plausible models have interpretable resolution properties.

2.5.3 Interpretation of Resolution Measures

The individual columns of \mathbf{R}^M can be plotted as model vector displaying how anomalies in the corresponding model cells are imaged by the combined process of measurement and inversion. The element R_{ij}^M reveals, how much of the anomaly in the j^{th} model cell is transferred into the i^{th} model cell. Consequently, the diagonal element R_{ii}^M states, how much of the information is saved in the model estimate and can be interpreted as resolvability of m_i . Adding up all the diagonal elements we obtain a total information of the inverse process, referred to as information content

$$IC = \sum_{i=1}^M \mathbf{R}_{ii}^M \quad . \quad (2.45)$$

In contrast, the matrix $\mathbf{R}^C = \mathbf{I} - \mathbf{R}^M$ describes the transformation of the starting model and can be interpreted like \mathbf{R}^M . The sum of its main diagonal elements contains the missing information $M - IC$. By relating the IC to the number of model parameters M the resolution degree $RD = IC/M$ shows to what degree the model is resolved by the data.

The individual columns of \mathbf{R}^D can be viewed as data and show whether the corresponding data point can be interpreted independent of each other within data errors. Off-diagonal elements of \mathbf{R}^D show data correlations, while the main diagonal elements \mathbf{R}_{ii}^D depict the independence, or importance, of the data point. The sum of the \mathbf{R}_{ii}^D is then the total data importance, which is identical to the information content

$$\sum_{i=1}^N \mathbf{R}_{ii}^D = \sum_{i=1}^M \mathbf{R}_{ii}^M = IC \quad .$$

Dividing the information content by the number of data N we obtain the mean importance or information efficiency

$$IE = \frac{IC}{N} = \text{mean}(\mathbf{R}_{ii}^D) \quad . \quad (2.46)$$

Following the ideas of Friedel (2000, 2003), a resolution radius for each model cell can be determined. Assuming a piecewise constant cell resolution, r_i is the radius of a sphere at the midpoint of the i^{th} cell having a resolution of 1. For a three-dimensional hexahedron of dimensions Δx_i , Δy_i and Δz_i the resolution radius is

$$r_i = \sqrt[3]{\frac{3\Delta x_i \Delta y_i \Delta z_i}{4\pi R_{ii}^M}} \quad .$$

The model covariance matrix has the same dimensions as \mathbf{R}^M . Its columns can be displayed as co-variations of the model with the corresponding cell. It reveals the statistical nature of the inversion process. The essential information of the MCM is contained in the main diagonal elements MCM_{ii} , whose square-roots can be interpreted as uncertainties of the corresponding model parameters. This is valuable if the model parameter values are to be used for petrophysical purposes such as the calculation of porosity or water saturation. Then, the uncertainty of those parameters is known by error propagation as well.

In summary, both resolution matrices give information about the geometrical reliability of the model. The MCM is used for parameter uncertainties, whereas R^M and R^D provide information about the universe of the inverse process with respect to model parameters and data.

2.5.4 Resolution in Terms of SVD/GSVD

Applying equations (2.26) and (2.29) yields for ordinary SVD

$$\begin{aligned} \mathbf{R}^M &= \mathbf{V}_r \text{diag}\left(\frac{f_i}{s_i}\right) \mathbf{U}_r^T \mathbf{U}_r \text{diag}(s_i) \mathbf{V}_r^T \\ &= \mathbf{V}_r \text{diag}(f_i) \mathbf{V}_r^T \end{aligned} \quad (2.47)$$

and for the generalized GSVD holds with (2.33)

$$\mathbf{R}^M = \mathbf{V} \text{diag}((f_i \quad \mathbf{0})) \mathbf{V}^{-1} \quad . \quad (2.48)$$

For both SVD and GSVD holds

$$\mathbf{R}^D = \mathbf{U}_r \text{diag}(f_i) \mathbf{U}_r^T \quad . \quad (2.49)$$

It can easily be shown that the information content can be calculated by

$$IC = \sum_{i=1} R_{ii}^M = \sum_{i=1} f_i$$

without forming the resolution matrix explicitly. Note that, following the last equation, for TSVD inversion the information content equals the number of used singular vectors r_{TSVD} .

2.5.5 Resolution Approximation

The SVD is a procedure that consumes time and memory. The speed decreases with second order of the matrix size. If the effort is prohibitive or impractical, it becomes reasonable to approximate \mathbf{R}^M as used by Alumbaugh and Newman (2000). The equation $\mathbf{R}^M = \hat{\mathbf{S}}^\dagger \hat{\mathbf{S}}$ can be rearranged using $\hat{\mathbf{S}}^\dagger = (\hat{\mathbf{S}}^T \hat{\mathbf{S}} + \lambda \mathbf{C}^T \mathbf{C})^{-1} \hat{\mathbf{S}}^T$ resulting in

$$(\mathbf{S}^T \mathbf{D}^T \mathbf{D} \mathbf{S} + \lambda \mathbf{C}^T \mathbf{C}) \mathbf{R}^M = \mathbf{S}^T \mathbf{D}^T \mathbf{D} \mathbf{S} \quad . \quad (2.50)$$

Equation (2.50) represents an inverse subproblem for every column of the resolution matrix. It can be solved approximately, e.g. by the CGLS family routines depicted in section 2.3. As starting vectors the individual columns of the identity matrix can be used. Possibly, it is worth the numerical effort to obtain an efficient preconditioner matrix, which can reduce the number of iterations and thus the computation time. Alternatively, an LSQR algorithm based on bidiagonalization techniques can be used for the solution of equation (2.50) as pointed out by Berryman (2000).

To sum it up, the decomposition is quite fast for small-scale problems with up to 1000 model parameter and data, which is the case for most two-dimensional inverse problems. For large-scale problems (with more than 1000 data and unknown) such as three-dimensional inversion generates, it becomes more and more reasonable to apply an approximate calculation of resolution.

2.6 Conclusions

In this chapter, methods for non-linear, multidimensional inversion in form of functional minimization were presented. The non-linear conjugate gradients (NLCG) method represents a memory saving routine, if the gradients can be calculated by forward modeling. To the contrary, the Gauss-Newton method makes use of the full Jacobian matrix and is expected to show quadratic convergence.

To obtain acceptable solutions, regularization methods have to be applied. Besides implicitly working equation solvers, techniques were presented introducing additional

constraints onto the model update (local scheme) or the model itself (global scheme). The model constraints are weighted by a regularization parameter, which has to be determined appropriately, e.g., by the investigation of the L-curve. Note that the distinction into global and local schemes is essential for resolution properties and the treatment of λ , even though the underlying equations differ only slightly.

In the following, algorithms were presented solving the inverse sub-problem efficiently from computing time and memory point of view. Conjugate gradient techniques could be adapted to the weighted least squares problem. Special variants were developed for the parallel solution of systems with multiple regularization strengths.

The singular value decomposition (SVD) is a valuable tool for gaining insight into inversion problems as well as for the construction of resolution matrices of TSVD and Tikhonov inversion. It has been shown, how this method can be generalized for arbitrary model constraints using the GSVD (generalized SVD).

The quality of inversion results can be appraised by the adaption of non-linear resolution matrices as known from linear inversion. It can be seen, how the inversion model is constructed of the real model structures, the introduced model constraints and noise artifacts. Thus, regularization as a trade-off between data fit and model simplicity can also be interpreted as compromising good reality mapping at limited noise artifacts. It has to be proved if the defined information content is an appropriate measure for the quality of inversion results.

3 Inversion of DC Resistivity Data

This chapter deals with the peculiarities of the inversion of DC resistivity data. At first, a brief introduction into DC resistivity measurements is given in section 3.1. An essential task of inversion is the investigation of the measurement errors used for weighting the data. Section 3.2 points out how the individual error sources can affect the inversion scheme. The noise character of typical multi-electrode measurements is estimated.

The first procedure in every inversion scheme is an appropriate parameterization of the ground into independent model cells. Section 3.3 shows some specific aspects of the parameterization in DC resistivity inversion.

In every iteration the model response has to be calculated by appropriate ways, generally using finite different (FD) and finite element (FE) techniques. In section 3.4 it is shown, how FD methods can be used for efficient and accurate forward calculation in flat earth problems.

The Jacobian matrix assembling the sensitivities plays a central role in inversion. Sensitivities represent the partial derivatives of the forward responses with respect to the model parameters. They have to be determined in every iteration step. In section 3.5 methods for the sensitivity approximation are presented.

3.1 Fundamentals of DC Resistivity

The current density \vec{j} is related to the electric field \vec{E} by the conductivity σ denoted by Ohm's law

$$\vec{j} = \sigma \vec{E} \quad . \quad (3.1)$$

In general, the material parameter σ connecting the two vectors is a tensor. Derivations from multiples of the identity matrix can be physically interpreted as anisotropy. Anisotropy often occurs in nature, because many rocks have preferential directions defined by the original pressure conditions for metamorphic rocks or sedimentary formations. Although there exist anisotropic forward calculations, anisotropy is rarely incorporated into inversion algorithms (Pain *et al.*, 2003). One reason is the increased number of degrees of freedom, which has to be compensated by more measurements

or regularizing conditions in the inversion process. In the following, the conductivity is considered to be a scalar. Its reciprocal value, the resistivity $\rho = 1/\sigma$, is often used for the petrophysical description of rocks, while conductivity is often related to hydrological parameters as porosity or salinity.

From Maxwell's equations for the stationary case follows the electric field being the negative gradient of the electrical potential φ

$$\vec{E} = -\nabla\varphi \quad \Rightarrow \quad \vec{j} = -\sigma\nabla\varphi \quad . \quad (3.2)$$

The continuity equation demands the divergence of \vec{j} to vanish for all source free regions

$$\nabla \cdot \vec{j} = 0 \quad .$$

For a point-like current source of current I at the position \vec{r}_s its divergence $I\delta(\vec{r} - \vec{r}_s)$ has to be considered, which leads to the governing equation

$$L_\sigma\varphi = \nabla \cdot (\sigma\nabla\varphi) = -I\delta(\vec{r} - \vec{r}_s) \quad . \quad (3.3)$$

L_σ represents an elliptic partial differential operator of Poisson type¹. Numerical solutions of elliptic PDE's have been studied widely. Generally, their treatment is easy compared to parabolic or hyperbolic differential operators. The main problem for the solution of equation (3.3) is created by the infinite source term, which has to be included into the discrete problem.

Due to the linearity of L_σ the potential of an electrode combination can be obtained by superposition of the individual potentials. Hence, the forward task is considered to involve single current electrodes in the following. Since multi-electrode measurements generally use each electrodes more than two times, the computation of mono-poles saves time compared to the modeling of dipoles.

For a homogeneous half-space of conductivity σ_0 the potential caused by a point source at $\vec{r}_s = (x; y; 0)$ is

$$\varphi(\vec{r}) = \frac{I}{2\pi\sigma_0|\vec{r} - \vec{r}_s|} \quad . \quad (3.4)$$

For a subsurface source the effect of the earth's surface has to be considered. The solution can then be superposed by the source $\vec{r}_s = (x_s; y_s; z_s)$ and a mirror source $\vec{r}_s' = (x_s; y_s; -z_s)$, yielding

$$\varphi = \frac{I}{4\pi\sigma_0} \left(\frac{1}{|\vec{r} - \vec{r}_s|} + \frac{1}{|\vec{r} - \vec{r}_s'|} \right) \quad . \quad (3.5)$$

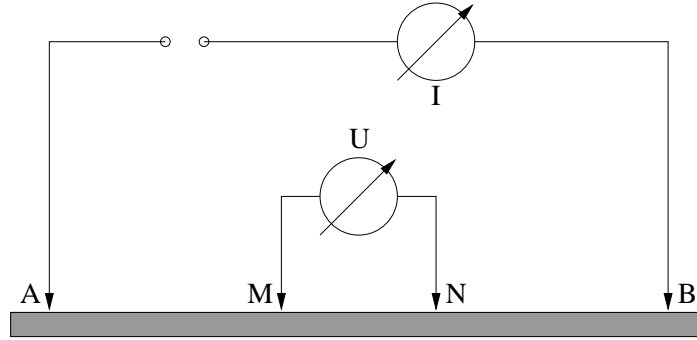


Figure 3.1: Sketch of a four-point DC measurement using surface electrodes

Note that equation (3.5) is identical to (3.4) for $z_s = 0$.

A single measurement is carried out by injecting a current I using two current electrodes A and B. Since generally transfer resistances of unknown magnitude occur at the electrodes, the voltage U is measured between two separate electrodes M and N. Figure 3.1 shows such a four-point configuration. The ratio of the voltage U and the current I is multiplied by a configuration factor k yielding the apparent resistivity

$$\rho_a = k \frac{U}{I} \quad . \quad (3.6)$$

The geometry factor k is determined such that $\rho_a = \rho$ for a homogeneous half-space of resistivity ρ using Neumann's equation, which for surface electrodes is

$$k = \frac{2\pi}{\frac{1}{AM} - \frac{1}{AN} - \frac{1}{BM} + \frac{1}{BN}} \quad . \quad (3.7)$$

For the general case of subsurface electrodes, the positions of the mirror current electrodes A' and B' have to be considered, yielding

$$k = \frac{4\pi}{\frac{1}{AM} + \frac{1}{A'M} - \frac{1}{AN} - \frac{1}{A'N} - \frac{1}{BM} - \frac{1}{B'M} + \frac{1}{BN} + \frac{1}{B'N}} \quad . \quad (3.8)$$

The apparent resistivity ρ_a represents the resistivity of the equivalent homogeneous half-space. This normalization is advantageous, e.g., ρ_a is ensured to be positive. Furthermore, data and model having identical physical units allows for easier estimation of regularization parameters, which would otherwise possess a physical unit. Note that equations (3.7) and (3.8) are valid for a flat surface. If an undulated topography is present, the configuration factor is unknown and can only be assessed by numerical modeling using homogeneous resistivity.

¹If σ is not constant, the equation is actually not a Poisson equation of the form $\Delta u = f$. However, theory and solution methods are comparable.

Actually, in most cases no direct current is injected. DC devices use low-frequency alternating current or rectangular pulses. For some cases, e.g. electronic conductors and electrolytic fluids in porous rocks, it is observed that the measured voltage does not follow the time function of the injected current. Due to polarization effects a phase shift between I and U occurs, denoted by the term induced polarization (IP). The theory of IP effects was described by Seigel (1959) using complex resistivities; for most rocks the behavior follows the model of Cole and Cole (1941). The reconstruction of the chargeabilities in the ground can be achieved by the application of complex algorithms. However, low contrast solutions are often sufficient, because the imaginary parts of the resistivity are generally small.

A single apparent resistivity represents a weighted mean of the earth's resistivities. Several independent data have to be acquired to obtain information about structures within the earth. This is done in two ways:

Mapping The entire electrode arrangement is moved along a profile to map out horizontal, mostly near-surface, variations.

Sounding In many areas the geological units are more or less layered structures. They can be investigated by increasing the electrode spread, which results in a deeper current penetration. Thus, information about deeper layers are obtained by resistivity sounding. For many decades resistivity sounding were carried out exclusively. This method is still practical for the investigation of - mostly hydrogeological - large-scale problems in sedimentary basins, where the layer thicknesses show small lateral variations.

There exist a variety of inversion programs, which produce layered resistivity models fitting the measured sounding curve. Many of them are based on TSVD and Marquardt type inversion schemes (Inman, 1975). Generally, the inversion parameters are the resistivities and thicknesses of a limited number of model layers. However, there are also approaches with fixed geometry and varying resistivities, which allow for smoothness constraints (Constable *et al.*, 1987). Due to the small number of parameters those inversions can be carried out without large numerical effort. Therefore, resolution and equivalence of the inverse problem are well studied.

If the earth's structure shows both horizontal and vertical variations, mapping and sounding have to be combined to obtain a resistivity structure as a function of x and z . Thus, the basic configuration is moved along the profile and expanded successively.

Classification of 2D DC data

Most measurements on multi-electrode instruments are carried out using common configurations. Table 3.1 provides a classification of configurations. Three space symbols between current (c) and potential (p) electrodes are introduced: While "." denotes a single electrode distance a , "-" show enlarging distances of $n \cdot a$ with the separation factor n .

Wenner- α	c-p-p-c	Pole-pole	c-p
Wenner- β	c-c-p-p	Pole-dipole	c-p.p
Wenner- γ	c-p-c-p	Pole-dipole (inc)	c-p:p
Schlumberger	c-p.p-c	Dipole-dipole	c.c-p.p
Schlumberger (inc)	c-p:p-c	Dipole-dipole (inc)	c:c-p:p

Table 3.1: Classification of 2D multi-electrode data sets using current (c) and potential (p) electrodes. "." denotes electrode distance a , "-" stands for separated distance $n \cdot a$ and ":" denotes increased dipole lengths

Note that big separation numbers n produce large configuration factors which can lead to erroneous measurements. As a consequence, an enlargement of the dipole lengths for Schlumberger, pole-dipole and dipole-dipole configurations seems reasonable, denoted with ":". Furthermore, it is recommended to shift the electrodes by the dipole length ":" instead of the electrode distance "." along the profile. Thus, a reduced number of data is gained. However, this can be compensated by applying a larger maximum separation factor. The data points are assigned to a representative x-location, e.g. the midpoint, and the separation factor n . Since large n are associated with increased investigation depths, it seems reasonable to plot the data in a pseudo-section with n going down. Remember, that a pseudo-section is just a graphical representation of the measured data. However, experienced geophysicists are able to retrieve a concept of the earth's structure without inverting the data.

For a numerical interpretation a model has to be found, which is able to explain the data. A simple method for data interpretation is an alternation of model change and forward calculation. To invert the data obtained by a profile measurement the model is considered to be composed of hexahedrons of infinite extension perpendicular to the profile line. Applying the concepts of non-linear inversion, an automated model finding process is created. To run the inversion, we need (a) a concept of the data errors, (b) an accurate forward modeling routine and (c) the sensitivities of the model response.

3.2 Errors of DC Measurements

The maximum relative error of ρ_a defined by equation (3.6) can be predicted as the sum of the relative errors of the configuration factor k , the voltage U and the current I (δa means variation of a)

$$\left| \frac{\delta \rho_a}{\rho_a} \right| \leq \left| \frac{\delta k}{k} \right| + \left| \frac{\delta U}{U} \right| + \left| \frac{\delta I}{I} \right| . \quad (3.9)$$

Since the current is kept constant and the stacking is usually applied to the impedance $Z = U/I$, the current is considered to have no variation. Then, the error consists of two parts, a geometrical error due to the variation of k and a device part due to uncertainties in U , which are investigated in the following.

3.2.1 Geometrical Errors

Single measurement

Geometrical errors can be represented by changes in the configuration factor k causing changes in the apparent resistivities

$$\left| \frac{\delta \rho_a}{\rho_a} \right|_k = \left| \frac{\delta k}{k} \right| .$$

Such changes occur for an undulated topography or for varying electrode positions. Topographical errors are systematic and can only be assessed by forward calculation.

The configuration factor for surface measurements is defined as

$$k = \frac{2\pi}{k} \quad \text{with} \quad \bar{k} = \frac{1}{AM} - \frac{1}{AN} - \frac{1}{BM} + \frac{1}{BN} . \quad (3.10)$$

The partial derivative of k with respect to the electrode position x_e can be written as

$$\frac{\partial k}{\partial x_e} = \frac{\partial k}{\partial \bar{k}} \cdot \frac{\partial \bar{k}}{\partial x_e} = -\frac{2\pi}{\bar{k}^2} \cdot \frac{\partial \bar{k}}{\partial x_e} = -\frac{k^2}{2\pi} \frac{\partial \bar{k}}{\partial x_e} . \quad (3.11)$$

\bar{k} is of the form $\frac{1}{x_1-x_2} + \dots$, so $\frac{\partial \bar{k}}{\partial x_{1/2}} = -\frac{1}{(x_1-x_2)^2} \cdot \pm 1$ so the minus sign in equation (3.11) vanishes. Note that the inner derivative of ± 1 has to be considered.

At first, consider a β -configuration with the electrode arrangement A-B-M-N. Its configuration factor reads

$$k = \frac{2\pi}{\frac{1}{x_M-x_A} - \frac{1}{x_N-x_A} - \frac{1}{x_M-x_B} + \frac{1}{x_N-x_B}} .$$

Then for the partial derivatives with respect to the x_e holds

$$\begin{aligned} \frac{\partial k}{\partial x_A} &= \frac{k}{2\pi} \left(-\frac{1}{AM^2} + \frac{1}{AN^2} \right) \quad , \quad \frac{\partial k}{\partial x_B} = \frac{k}{2\pi} \left(\frac{1}{BM^2} - \frac{1}{BN^2} \right) \\ \frac{\partial k}{\partial x_M} &= \frac{k}{2\pi} \left(\frac{1}{AM^2} - \frac{1}{BM^2} \right) \quad \text{and} \quad \frac{\partial k}{\partial x_N} = \frac{k}{2\pi} \left(-\frac{1}{AN^2} + \frac{1}{BN^2} \right) \quad . \end{aligned} \quad (3.12)$$

All other configurations can be expressed using factors α_i

$$k = \frac{2\pi}{\frac{\alpha_{AM}}{x_M - x_A} - \frac{\alpha_{AN}}{x_N - x_A} - \frac{\alpha_{BM}}{x_M - x_B} + \frac{\alpha_{BN}}{x_N - x_B}} \quad ,$$

where, e.g., α_{BM} is 1 for $x_M > x_B$, -1 for $x_M < x_B$ and 0 if one of the electrode is not used. This results in the partial derivatives

$$\begin{aligned} \frac{\partial k}{\partial x_A} &= \frac{k}{2\pi} \left(-\frac{\alpha_{AM}}{AM^2} + \frac{\alpha_{AN}}{AN^2} \right) \quad , \quad \frac{\partial k}{\partial x_B} = \frac{k}{2\pi} \left(\frac{\alpha_{BM}}{BM^2} - \frac{\alpha_{BN}}{BN^2} \right) \\ \frac{\partial k}{\partial x_M} &= \frac{k}{2\pi} \left(\frac{\alpha_{AM}}{AM^2} - \frac{\alpha_{BM}}{BM^2} \right) \quad \text{and} \quad \frac{\partial k}{\partial x_N} = \frac{k}{2\pi} \left(-\frac{\alpha_{AN}}{AN^2} + \frac{\alpha_{BN}}{BN^2} \right) \quad . \end{aligned} \quad (3.13)$$

A maximum relative error can be obtained by multiplying the sum of the absolute values with the electrode variance δx and dividing it by the configuration factor

$$\left| \frac{\delta k}{k} \right|_{max} = \left| \frac{\delta x}{k} \right| \left(\left| \frac{\partial k}{\partial x_A} \right| + \left| \frac{\partial k}{\partial x_B} \right| + \left| \frac{\partial k}{\partial x_M} \right| + \left| \frac{\partial k}{\partial x_N} \right| \right) \quad . \quad (3.14)$$

Figure 3.2 shows the maximum geometrical error for several widely used configurations. The equally spaced electrodes of spacing $a = 1$ m are assumed to have a positioning variance of $\delta x = 0.01$ m.

For all configurations the errors decrease with the separation factor. Note that the errors are weighted with the ratio $\delta x/a$. Thus, a spacing of $a = 10$ m with $\delta x = 10$ cm yields identical errors. However, in practice the relative positioning error decreases with spacing.

For all configuration types the largest errors occur for small separations. Pole-pole and Wenner configurations show the smallest errors. The relatively large error level of dipole-dipole and pole-dipole data (which has identical errors compared to Schlumberger configuration) can be significantly reduced by increasing the dipole length for large separations.

Geometrical errors for multi-electrode data

When the measurements are carried out using a multi-electrode system of E fixed electrodes, the errors of the N data do not remain independent of each other if electrode

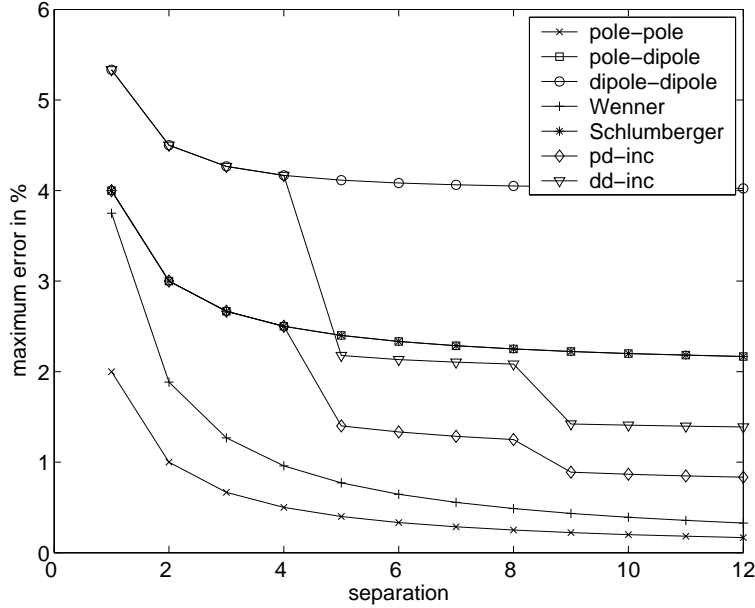


Figure 3.2: Relative geometric errors of several configurations with electrode distance 1 m, $\delta x = 1$ cm, dipole lengths for dd-inc and pd-inc are increased every 4th separation

variations are considered. Covariances occur, in the way that a misplacement of a single electrode has systematical effects onto all measurements using it. Therefore we expect a correlation amongst the data, represented by the covariance matrix \mathbf{C}_d . It replaces the uncorrelated data variance matrix $(\mathbf{D}^T \mathbf{D})^{-1}$ used in the minimization of the data functional Φ_d such that

$$\Phi_d = [\mathbf{d} - f(\mathbf{m})]^T \mathbf{C}_d^{-1} [\mathbf{d} - f(\mathbf{m})] \quad .$$

For the covariance of the elements of the vector $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$ with respect to the variation of the vector \mathbf{x} holds $\text{Cov}(\mathbf{y}) = \mathbf{A}\text{Cov}(\mathbf{x})\mathbf{A}^T$. The vector \mathbf{y} contains the apparent resistivities, \mathbf{x} are the electrode positions. A linearized Taylor series yields

$$\mathbf{y}(\mathbf{x}) = \mathbf{y}(\hat{\mathbf{x}}) + \frac{\partial y_i}{\partial x_j}(\hat{\mathbf{x}}) \cdot (\mathbf{x} - \hat{\mathbf{x}}) + \dots \quad ,$$

where $\hat{\mathbf{x}}$ are the expected electrode positions. Higher order terms can be neglected for small deviations. The matrix $\mathbf{A} \in \mathbb{R}^{N \times E}$ containing the partial derivatives is constructed as described above for every combination of data number and electrode number. Assuming the electrode positions to have identical standard deviations δx corresponding to the covariance $\text{Cov}(\mathbf{x}) = \delta x^2 \mathbf{I}$, the data covariance matrix due to geometrical errors reads

$$\mathbf{C}_d = \mathbf{A}^T \mathbf{A} \cdot \delta x^2 \quad . \quad (3.15)$$

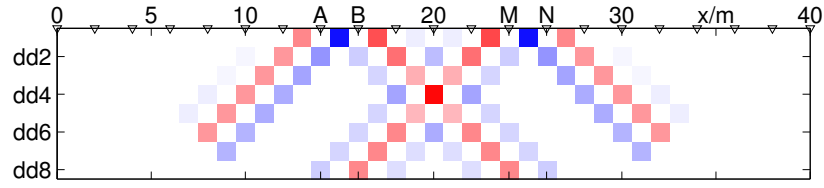


Figure 3.3: Data covariance structure for one configuration out of a dipole-dipole data set

Figure 3.3 exemplifies the covariance of all data with respect to an individual one, which is the corresponding column of \mathbf{C}_d . It can be seen, how all data using one of the four electrodes show co-variations. They are positive or negative according to the sensitivity structure.

3.2.2 Voltage Errors

Voltage errors are the relative deviations due to uncertainty of the measured voltage

$$\left| \frac{\delta \rho_a}{\rho_a} \right|_U = \left| \frac{\delta U}{U} \right| = \frac{\delta U |k|}{\rho_a I} .$$

δU is a small voltage, which cannot be distinguished from zero. Generally, it represents either the minimum accuracy of the measuring device, which is defined by the limited precision of the analog digital converter or voltages occurring in the device while measuring. Also, irrigation currents in the earth can produce a varying voltage level and disturb the measuring accuracy. Statistically varying voltages can be significantly reduced using stacking procedures. Thus, if irrigation currents are involved in the noise estimation, the effective stacking number has to be considered.

Figure 3.4 shows, using the same data sets as in Figure 3.2, the maximum error resulting from a limited voltage accuracy of $500 \mu\text{V}$, which is a reasonable value. The driving current is assumed to be 100 mA . The errors increase with the separation factor for all configurations. Dipole-dipole and pole-dipole configurations are sensible to voltage errors due to their larger configuration factors. However, the errors can be significantly reduced by dipole enlargement.

3.2.3 Numerical Implementation

The total error is considered to be a superposition of a device error and a geometrical error. The device error is supposed to have Gaussian statistics and can be measured by stacking impulses and computing the standard deviation of the measured impedance. If errors cannot be measured they have to be estimated by assuming the minimum volt-

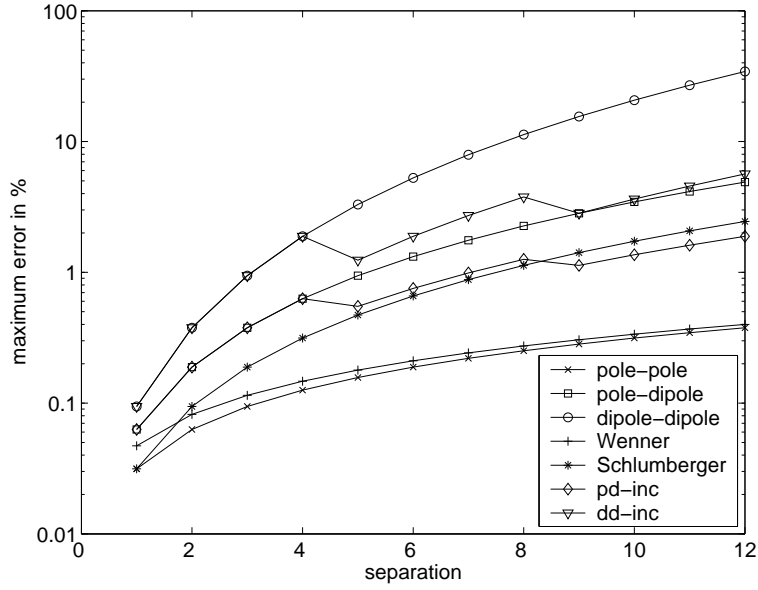


Figure 3.4: Relative errors due to voltage accuracy of several configurations with electrode distance 1 m, $\delta U = 500 \mu\text{V}$, $I = 100 \text{ mA}$, $\rho_a = 100 \Omega \text{ m}$, dipoles for dd-inc and pd-inc are increased every 4th separation

age, usual values are several $100 \mu\text{V}$. Also, the geometrical error needs to be estimated, which can be done in two ways:

1. The geometrical error is treated as uncorrelated maximum error or as correlated errors due to a Gaussian variation of the electrode positions. Hence we can work with the diagonal data weighting matrix \mathbf{D} as introduced in section 2.1. The relative variance reads

$$\frac{\delta\rho_a}{\rho_a} = p + \frac{\delta U|k|}{\rho_a I} + \frac{\delta k}{|k|}(\delta x) \quad . \quad (3.16)$$

The fixed relative error p can be used as an alternative to the geometrical error.

2. The data covariance matrix is the superposition of the squared statistical errors and the geometrical error matrix \mathbf{A}

$$\mathbf{C}_d = \mathbf{A}^T \mathbf{A} \cdot \Delta x^2 + \text{diag}(e_i^2) \quad . \quad (3.17)$$

In the algorithms described in section 2.3 the matrix $\mathbf{D}^T \mathbf{D}$ has to be replaced by \mathbf{C}_d^{-1} . Since \mathbf{C}_d has the dimension of $N \times N$, a matrix inversion is prohibitive regarding memory usage. Hence, in every iteration of the equation solver, the solution to the equation

$$\mathbf{C}_d \mathbf{x} = (\mathbf{A}^T \mathbf{A} \Delta x^2 + \mathbf{E}) \mathbf{x} = \mathbf{b}$$

has to be sought. The matrix $\mathbf{A} \in \mathbb{R}^{N \times E}$ is sparse having at most $4N$ nonzero entries, $\mathbf{E} = \text{diag}(e_i)$ is a sparse diagonal matrix. Since the subproblem is well-conditioned, the latter equation can easily be solved approximately using conjugate gradient techniques with minimum time and memory usage.

3.3 Parameterization and Regularization

3.3.1 The Parameter Function

The model parameter m_i may represent the resistivity ρ_i or the conductivity σ_i . By the inversion of resistivity or conductivity, resistive or conductive bodies are enhanced, respectively. This can be of advantage, but also leads to artifacts. To get rid of the choice, often m_i is chosen to represent the logarithm of the resistivity or conductivity. For details on logarithmic parameters see Tarantola (2001). It ensures the resistivities remaining always positive. Positivity has not to be claimed by regularization methods, because it represents a hard constraint.

Accordingly, the data are considered as logarithms of the apparent resistivity values. Consequently, the model update is also a logarithmic update. For the update of the resistivity values ρ_i in each iteration step by $m_i^{new} = m_i + \Delta m_i$ follows that

$$\rho_i^{new} = \rho_i \cdot \exp(\Delta m_i) \quad .$$

The idea of logarithmic non-negativity constraints can be expanded to claim the resistivity to be larger than a lower resistivity limit ρ_l using $m_i = \log(\rho_i - \rho_l)$, which results in the update formula

$$\rho_i^{new} = \rho_l + (\rho_i - \rho_l) \cdot \exp(\Delta m_i) \quad .$$

This method is referred to as logarithmic barrier constraint technique (Li and Oldenburg, 2003). It can also be used to limit ρ_i to an upper resistivity bound ρ_u by setting $m_i = \log(\rho_u - \rho_i)$. Both limits can be combined using the model parameter

$$m_i = \log \left(\frac{\rho_i - \rho_l}{\rho_u - \rho_i} \right) \quad ,$$

the accompanying update step is formulated as

$$\rho_i^{new} = \frac{\rho_u(\rho_i - \rho_l) \exp \Delta m_i + \rho_l(\rho_u - \rho_i)}{(\rho_i - \rho_l) \exp \Delta m_i + \rho_u - \rho_i} \quad .$$

Note that for small contrasts the inversion of conductivity, resistivity or the logarithm yields similar results. For large contrasts an improved convergence is observed for

logarithms. Therefore, in all subsequent examples the parameters are the logarithms of resistivity. In some cases, e.g. if the resistivity of a conductive basement is known, the lower bound technique can be successfully applied.

3.3.2 Geometrical Parameter Arrangement

The decomposition of the earth into sub-domains creates the discrete inverse problem. It has to satisfy the following conditions:

1. The sub-domains have to be small enough and the discretized region has to be large enough to ensure the existence of models fitting the data within error.
2. The number of model parameters (sub-domains) must not be too large due to the ill-posedness and the limitation of computer resources.
3. It must be possible to define appropriate regularizing functions.
4. The discrete model has to be transformed into a finite difference or finite element grid for an accurate forward calculation.

If there is no topography present, an FD-like grid is the matter of choice, because the FD forward calculation is quite fast and regularization methods like smoothness constraints can make use of the regular arrangement of the model cells. At first, the model domain has to be bounded. Typically, its boundaries are placed several (1 up to 5) electrode distances outside the bounding electrodes.

The layer discretization has to be chosen according to the depth penetration of the electrode configurations. An estimation of representative depth is an important question for all geophysical methods. Investigations for direct current methods were made early by Roy and Apparao (1971) and Barker (1989), who investigated penetrations of single configurations. However, for the general case of mixed or uncommon electrode arrangements an alternative method has to be found.

The one-dimensional sensitivity distribution helps to find a reasonable parameterization as follows. It can easily be shown, that for a pole-pole measurement with the distance $\overline{AM} = a$ the layer sensitivities in a homogeneous half-space is

$$\int_{z_1}^{z_2} \frac{z}{\sqrt{a^2 + z^2}^3} dz = \frac{a}{\sqrt{a^2 + z^2}} \Big|_{z_1}^{z_2}, \quad (3.18)$$

where z_1 and z_2 are the upper and lower layer boundaries, respectively. For four-electrode configurations the sensitivities can be obtained by superposition.

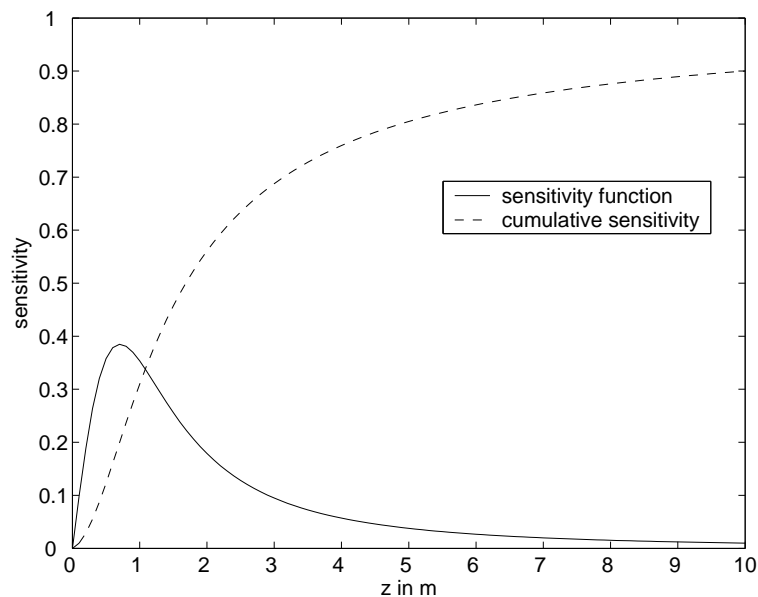


Figure 3.5: One-dimensional sensitivity distribution of a dipole-dipole measurement using $a=2$ m and $n=4$

Figure 3.5 shows the 1D-sensitivity of a dipole-dipole configuration. The layer sensitivity function has a clear maximum at $z=0.8$ m. The cumulative sensitivity, which equals (3.18) for $z_1=0$, begins to saturate at about 5–8 m. Since the integral reaches 1 only for infinite depth, the value of 0.9 is used for a representative penetration depth. Figure 3.6 displays the penetration depth per electrode spacing as a function of the separation factor for the above mentioned data sets. The pole-pole configuration has the largest penetration depth, whereas a dipole-dipole configuration penetrates only shallow. For a set of configurations the layer sensitivities can be added up to obtain a total investigation depth. The in-between layers are determined the way that either the sensitivities of the individual layers are equal or that the layer thicknesses increase with a constant factor.

The calculation of layer sensitivities leads to the idea of generating a layered starting model using a linearized inversion scheme. It has the advantage of calculating the forward responses semi-analytically (Li and Spitzer, 2002).

Besides the layers, horizontal grid spacings have to be defined. Note that the model has not necessarily to equal a finite difference grid itself. However, it must match the lines of a fine grid. From that two model parameter types arise denoted with the terms Grid and Para.

The Grid model is a regular finite difference grid. Grid spacings are the typical electrode distances Δx_e , Δy_e or their minimum for both values. In small-scale and shallow cases of well-covered ground a grid of $\Delta x_e/2$ can be used successfully.

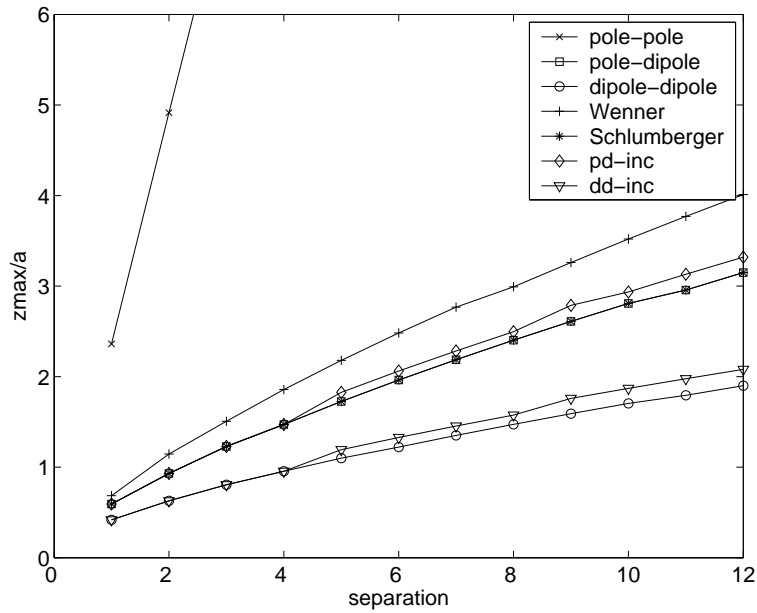


Figure 3.6: Maximum z (90% sensitivity) per electrode spacing a for several configurations as a function of separation factor n , dipoles for dd-inc and pd-inc are increased every 4th n

The Para model has element sizes varying from layer to layer. Generally, the elements increase with depth to consider the decreasing resolution. The edge lengths are multiples of a unit length, recommended is the minimum of $\Delta x_e/2$ and $\Delta y_e/2$.

Figure 3.7 shows how the model types may look like for a 2D discretization.

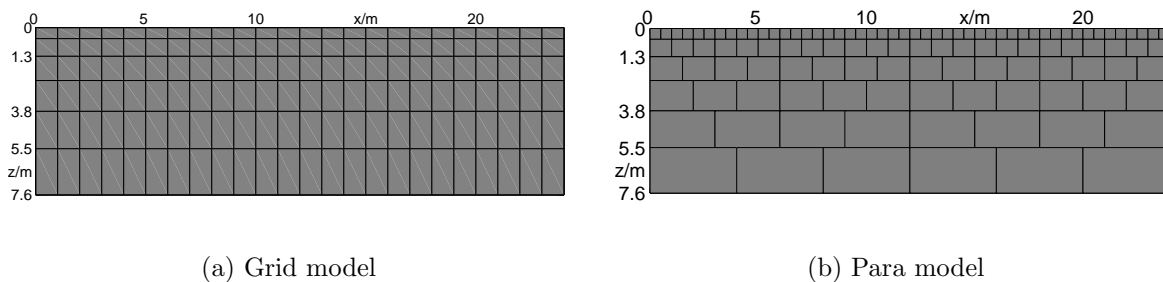


Figure 3.7: Sample two-dimensional Grid and Para type models. The Grid model contains $24 \times 6 = 144$ regular cells, the para model $48 + 24 + 16 + 12 + 8 + 6 = 114$ cells, whose lateral extent increases with depth

3.4 DC Forward Calculation

The term forward calculation denotes the simulation of synthetic data for a given model parameterization. For a point source of current I at the position \vec{r}_s the governing Poisson equation (3.3) reads

$$\nabla \cdot (\sigma \nabla \varphi) = \nabla \sigma \cdot \nabla \varphi + \sigma \nabla^2 \varphi = -I \delta(\vec{r} - \vec{r}_s) \quad , \quad (3.19)$$

which has to be solved for the potential $\varphi(\vec{r})$ by appropriate methods. Note that a non-unique solution for the domain Ω requires the specification of boundary conditions at its boundary Γ .

The continuity equation (3.19) is defined if the potential is twice differentiable ($\varphi \in C^2$) and the conductivity once ($\sigma \in C^1$) which holds within the sub-domains Ω of constant σ_i . At the inner boundaries of two sub-domains Ω_m and Ω_n the continuity equation yields $\sigma_m \nabla \varphi_m = \sigma_n \nabla \varphi_n$, which has to be satisfied. A common way is the solution of (3.19) in an integrated sense.

For bodies of closed geometry boundary integral methods can be used, which are seldom of interest in inverse problems. With the rapid development of computers in the last decades finite difference (FD) and finite element (FE) techniques were applied for the solution of differential equations in many branches. In the following is shown how FD calculations can be efficiently used. As a fundamental principle of numerical simulation holds, that the accuracy of the results increases with increasing numerical effort in form of a finer discretization. Main attention is paid to find a trade-off between reasonable accuracy within limited computing time.

3.4.1 Finite Difference Discretization

The basis of the finite difference technique is the construction of a discrete model in form of a hexahedral grid² with nodes at the cell corners. The existing partial derivatives are replaced by finite differences. An overview of finite difference modeling techniques and discretization schemes for DC problems was given by Spitzer (1999).

The three dimensions of the modeling domain are subdivided into a grid by the node positions x_i ($i \in 1 \dots i_{max}$), y_j ($j \in 1 \dots j_{max}$) and z_k ($k \in 1 \dots k_{max}$). There exist a lot of discretization schemes differing in the location of conductivities and partial derivatives (Dey and Morrison, 1979a; Zhang *et al.*, 1995; Spitzer, 1995). For a comparison see Spitzer and Wurmstich (1995). In the following, the discretization technique of Dey and Morrison (1979a) is applied. It is identical to the scheme of Brewitt-Taylor

²The finite difference grid does not necessarily correspond to the inversion parameter structure. However, a transformation, usually by refinement and prolongation, has to be found.

and Weaver (1976) used for 2D induction problems, which was applied to the 3D DC problem by Spitzer (1995).

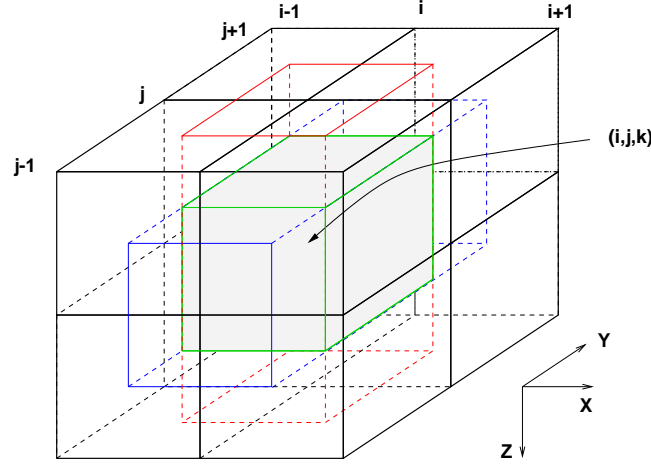


Figure 3.8: Section of the FD grid for the grid node ijk , the dashed line denotes the definition of the average conductivity at the grid nodes

Figure 3.8 shows a section of a finite difference grid. The elementary domain Ω_{ijk} with the conductivity σ_{ijk} is bounded by the grid nodes (i,j,k) , $(i+1,j,k)$, $(i,j+1,k)$, $(i+1,j+1,k)$, $(i,j,k+1)$, $(i+1,j,k+1)$, $(i,j+1,k+1)$ and $(i+1,j+1,k+1)$.

By integration of equation (3.19) on the model cell Ω_{ijk} we obtain

$$\iiint_{\Omega_{ijk}} \nabla \cdot (\sigma \nabla \varphi) d^3 \vec{r} = - \iiint_{\Omega_{ijk}} I \cdot \delta^3(\vec{r} - \vec{r}_s) d^3 \vec{r} = -I(\vec{r}) \quad .$$

By the application of Gauss' theorem the volume integral is transformed into a surface integral

$$\iiint_{\Omega_{ijk}} \nabla \cdot (\sigma \nabla \varphi) d^3 \vec{r} = \iint_{\Gamma_{ijk}} \sigma \frac{\partial \varphi}{\partial \eta} d\Gamma = -I(\vec{r}) \quad , \quad (3.20)$$

where Γ_{ijk} denotes the enclosing surface of Ω_{ijk} and η is its normal vector.

The potential gradient $\frac{\partial \varphi}{\partial \eta}$ is approximated using central differences on every edge. Thus the integral is replaced by a sum over the 6 faces. Figure 3.9 shows the conductivities at the edge in $+k$ -direction.

The conductivity σ is the weighted mean of the four adjacent conductivity cells, thus the integral at the bottom face is

$$\iint_{\Delta \Gamma_{i,j,k+1/2}^k} \sigma \frac{\partial \varphi}{\partial z} dx dy = \frac{\varphi_{i,j,k+1} - \varphi_{i,j,k}}{\Delta z_k} \left(\sigma_{i-1,j-1,k} \frac{\Delta x_{i-1} \Delta y_{j-1}}{4} + \right.$$

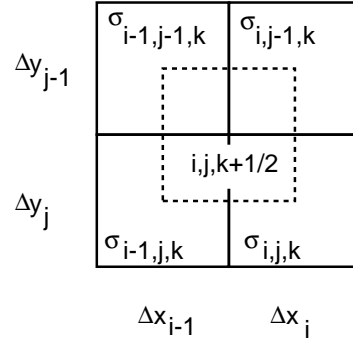


Figure 3.9: Plan view of the line between the grid nodes (i,j,k) and $(i,j,k+1)$, the dashed line represents the equivalent area with the weighted conductivity

$$+ \sigma_{i,j-1,k} \frac{\Delta x_i \Delta y_{j-1}}{4} + \sigma_{i-1,j,k} \frac{\Delta x_{i-1} \Delta y_j}{4} + \sigma_{i,j,k} \frac{\Delta x_i \Delta y_j}{4} \quad . \quad (3.21)$$

Replacing the partial derivatives with respect to all directions yields

$$C_{top} \cdot \varphi_{i,j,k-1} + C_{bottom} \cdot \varphi_{i,j,k+1} + C_{left} \cdot \varphi_{i-1,j,k} + C_{right} \cdot \varphi_{i+1,j,k} + \\ C_{front} \cdot \varphi_{i,j-1,k} + C_{back} \cdot \varphi_{i,j+1,k} + C_{self} \cdot \varphi_{i,j,k} = -I(x_s, y_s, z_s) \quad , \quad (3.22)$$

which represents the discrete partial differential equation at the node (i,j,k) . C_{left} , C_{right} , C_{back} , C_{front} , C_{top} and C_{bottom} are the coupling coefficients in -x, +x, -y, +y, -z and +z, respectively. They depend on the surrounding conductivities and grid spacings and can be interpreted as conductances between the adjacent nodes. For example the coupling coefficient into bottom direction is

$$C_{bottom} = -\frac{1}{\Delta z_k} \left(\sigma_{i-1,j,k} \frac{\Delta x_{i-1} \Delta y_j}{4} + \sigma_{i,j,k} \frac{\Delta x_i \Delta y_j}{4} + \right. \\ \left. \sigma_{i-1,j-1,k} \frac{\Delta x_{i-1} \Delta y_{j-1}}{4} + \sigma_{i,j,k} \frac{\Delta x_i \Delta y_j}{4} \right) \quad .$$

The self coupling coefficient C_{self} is the negative sum of the six coefficients

$$C_{self} = -(C_{top} + C_{bottom} + C_{left} + C_{right} + C_{front} + C_{back}) \quad . \quad (3.23)$$

Boundary conditions

At the boundary of the modeling domain the neighbors in the outward directions are missing. Therefore the behavior of the potential field has to be defined by boundary conditions. Generally, there exist three different types:

Neumann boundary conditions fix the potential derivative with respect to the outward direction. This is essential for the earth's surface, where the current flow

perpendicular to the surface is zero. They can be applied by introducing an additional conductivity-free layer above the surface, which sets C_{top} to zero, before the summation for C_{self} is done.

Dirichlet boundary conditions are used to fix the potential values. They can be calculated analytically for a homogeneous or layered half-space. Since the potentials are not known, usually the outward potentials are set to zero. To obtain an improved accuracy in the modeling domain, layers with prolonged grid spacings are added to the parameter model boundaries.

Mixed boundary conditions relate the potential and its derivative by assuming the potential characteristics. Dey and Morrison (1979a) presented boundary conditions for a single current electrode at the origin

$$\frac{\partial \varphi}{\partial \eta} + \frac{\varphi}{|\vec{r}|} \cos \theta = 0 \quad , \quad (3.24)$$

where θ is the angle between the position vector \vec{r} and the outward vector $\vec{\eta}$. Using the dot product we obtain, e.g. for the boundary at x,

$$\frac{\partial \varphi}{\partial \eta} + \frac{x}{r^2} \varphi = 0 \quad ,$$

which can be discretized and introduced into the system of equations easily. Zhang *et al.* (1995) presented improved mixed boundary conditions for multi-pole sources. Since for all subsequent simulations mono-poles are used, they are not of interest. Note that the coupling matrix has to be re-assembled for every individual electrode respecting their position. However, it is sufficient to formulate the boundary conditions using a representative source location in the middle of the electrode array, because the modeling boundaries are generally far from the electrodes.

Using equation (3.22) for all nodes yields the discrete differential equation represented by a system of equations

$$\mathbf{K} \cdot \boldsymbol{\varphi} = \mathbf{b} \quad , \quad (3.25)$$

which has to be solved for the vector $\boldsymbol{\varphi}$ containing the potentials for all existing nodes. The coupling matrix \mathbf{K} works as discretized differential operator L_σ . It is a sparse matrix of band-structure. The coupling coefficients fill the individual main and off-side diagonals. \mathbf{K} is guaranteed to be positive semi-definite, regular³ and thus has a unique

³The matrix becomes singular, if Neumann conditions are applied to all boundaries. Hence, at least at one point Dirichlet or mixed conditions have to be installed.

solution for all \mathbf{b} . The source vector \mathbf{b} represents a discrete Dirac function. All elements are zero except for the node where the current electrode is placed. All electrodes have to match the finite difference grid, which can lead to huge node numbers for irregular electrode positions. Since the potential follows a behavior proportional to $1/r$, it can hardly be discretized by a piecewise linear function. Hence, large discretization errors occur near the electrodes, which can only be diminished by a very fine grid.

3.4.2 Singularity Removal Technique

Lowry *et al.* (1989) introduced the singularity removal technique for DC modeling. However, secondary field simulations can be applied to other numerical modeling problems with complicated source fields (Günther, 2000). The approach is to consider the potential as sum of two parts

$$\varphi = \varphi_p + \varphi_s \quad . \quad (3.26)$$

The primary⁴ potential φ_p represents the known potential for a well-defined parameter structure. It satisfies the continuity equation for the conductivity distribution σ_0

$$L_{\sigma_0}\varphi_p = \nabla \cdot (\sigma_0 \nabla \varphi_p) = -I\delta(\vec{r} - \vec{r}_s) \quad . \quad (3.27)$$

The potential for the conductivity structure σ_0 must be known. Generally the homogeneous half-space is used. An alternative is a layered medium or a vertical contact, whose potentials can be derived following Li and Spitzer (2002). It is further possible to use the conductivities of the last model iterate.

The total potential satisfies the continuity equation (3.3) for σ

$$\nabla \cdot (\sigma \nabla (\varphi_p + \varphi_s)) = \nabla \cdot (\sigma \nabla \varphi_p) + \nabla \cdot (\sigma \nabla \varphi_s) = -I\delta(\vec{r} - \vec{r}_s) \quad . \quad (3.28)$$

By subtracting equation (3.27) from (3.28) follows that

$$\begin{aligned} \nabla \cdot (\sigma \nabla \varphi_s) &= \nabla \cdot ((\sigma_0 - \sigma) \nabla \varphi_p) \\ L_{\sigma}\varphi_s &= L_{\sigma_0 - \sigma}\varphi_p \quad . \end{aligned} \quad (3.29)$$

Sources of the secondary potential appear at regions of conductivity deviation and are strong near the electrodes, where $\nabla \varphi_p$ is large. For an anomalous, extensive body its boundaries bear positive and negative secondary sources working as a capacitance. The secondary potential shows no singularity. It is generally smoother and easier to approximate.

⁴The primary field is also referred to as "normal" or "singular", whereas the corresponding terms of the secondary field are "anomalous" and "regular".

Since the right side is defined by the potential instead of the discrete Dirac function, the singularity removal technique allows for grid independent electrode positioning. However, if the electrodes are placed off the grid nodes, the corresponding potentials have to be obtained by interpolation, which can introduce additional errors. Generally, a tri-linear interpolation of the secondary potential yields reasonable results.

Equation (3.29) is discretized replacing the operators by the coupling matrices

$$\mathbf{K}_\sigma \cdot \varphi_s = \mathbf{K}_{\sigma_0 - \sigma} \cdot \varphi_p \quad . \quad (3.30)$$

Thus we have additionally to assemble the coupling matrices for the conductivity deviation $\sigma_0 - \sigma$. It is obvious that the singularity is only removed if σ_0 equals the source conductivity as suggested by Zhao and Yedlin (1996). Unfortunately, secondary sources may arise at the boundaries leading to errors. The apparent resistivities can directly be calculated using

$$\rho_a = \rho_0 + \rho_a^s = \rho_0 + k \frac{\Delta \varphi^s}{I} \quad ,$$

where $\rho_0 = 1/\sigma_0$. Generally, the various sources are located at different conductivities σ_i . Then it proves useful to assemble a matrix \mathbf{K}_1 for $\sigma = 1$ and use the linearity, which yields for (3.30)

$$\mathbf{K}_\sigma \varphi_s = (\sigma_i \mathbf{K}_1 - \mathbf{K}_\sigma) \varphi_p = \sigma_i \mathbf{K}_1 \varphi_p - \mathbf{K}_\sigma \varphi_p \quad . \quad (3.31)$$

Let ρ_A and ρ_B be the conductivities at the source electrodes A and B. The apparent resistivity for a four-point measurement can be calculated by

$$\rho_a = \frac{k}{2\pi} \left(\frac{\rho_A}{AM} - \frac{\rho_A}{AN} - \frac{\rho_B}{BM} + \frac{\rho_B}{BN} \right) + \rho_a^s \quad . \quad (3.32)$$

3.4.3 Accuracy and Speed

Regular grids have to be refined highly to solve the total potential accurately. Since this results in large grids and huge computing time, the following computations are carried out using the singularity removal technique. The system of equations to be solved is represented by the quadratic, positively defined matrix \mathbf{A} and a left hand side matrix \mathbf{B} containing the secondary source vectors \mathbf{b}_i for all electrode positions. A broad overview of accuracy and speed for DC schemes has been given by Spitzer and Wurmstich (1995). However, the investigations merely apply single injection points and the effect of grid refinement was not considered.

There exists a variety of equation solvers. At first, direct methods as Gaussian elimination and QR-decomposition are to be mentioned. Although they may be interesting in

case of many right-side vectors, their use is limited to small-scale problems. Since the coupling matrix is sparse, iterative equation solvers are well-suited. Amongst them, Krylov subspace methods prove to converge fast. The conjugate gradient technique presented by Hestenes and Stiefel (1952) shows superior results (Spitzer and Wurmstich, 1995). Generally the satisfaction of appropriate boundary conditions demands boundaries far from the model region. This is usually accomplished by adding prolonging cells. The prolonging factor reaches from 1.3 to 5, the corresponding number of cells might equal 3 up to 20.

Irregular grids result in slow convergence indicated by large condition numbers. To accelerate the convergence, preconditioning is often successfully applied. The idea is to find an easy-to-invert matrix \mathbf{M} to solve the equivalent system of equations

$$\mathbf{M}^{-1}\mathbf{A}\mathbf{x} = \mathbf{M}^{-1}\mathbf{b} \quad ,$$

with a reduced condition number $\text{cond}(\mathbf{M}^{-1}\mathbf{A}) \ll \text{cond}(\mathbf{A})$. The easiest symmetric preconditioner is called Jacobi because it uses the main diagonal elements for scaling the system of equations. Asymmetric preconditioners use two matrices \mathbf{R}_1 and \mathbf{R}_2 instead of \mathbf{M}

$$\mathbf{R}_1^{-1}\mathbf{A}\mathbf{R}_2^{-1}\mathbf{x} = \mathbf{R}_1^{-1}\mathbf{R}_2^{-1}\mathbf{b} \quad . \quad (3.33)$$

Since triangular matrices can be inverted easily, they are typical. Due to the symmetry of \mathbf{A} , often only one $\mathbf{R} = \mathbf{R}_1 = \mathbf{R}_2^T$ is used. For example, the symmetric successive over-relaxation (SSOR) preconditioner can be written as

$$\mathbf{R} = \frac{1}{\sqrt{2-\omega}}\mathbf{D}^{-1}(\mathbf{D} + \mathbf{U}) \quad , \quad (3.34)$$

where \mathbf{D} is the diagonal matrix of \mathbf{A} , \mathbf{U} its upper right triangle matrix and ω is referred to as relaxation factor.

Another preconditioner is obtained by incomplete Cholesky factorization of $\mathbf{A} \approx \mathbf{R}^T\mathbf{R}$ (Wu *et al.*, 2003). In an iterative process \mathbf{R} is changed until the norms of \mathbf{A} and $\mathbf{R}^T\mathbf{R}$ differ by a tolerance *tol* (Saad, 1996). An alternative variant uses the same sparsity structure as the triangular matrix of \mathbf{A} , but its tolerance is limited.

Forward calculation errors

The errors arising in the forward calculation may be divided into two types:

Discretization errors Even if the model cells possess equal conductivities, errors are made using the finite difference approximations of the partial derivatives. Such errors occur at points, where the potential behavior departs from linear form,

particularly near the electrodes. Other discretization errors are introduced at point of changing finite distances, where the midpoint of the central differences departs from the grid node.

Conductivity contrast errors At the inner model boundaries arise sharp conductivity contrasts of infinite gradients, which are discretized by finite differences.

Generally, the error types can hardly be separated. Both types can be diminished using a finer discretization. However, a local refinement affects the entire modeling domain and increases the numerical work significantly. A comparison to analytical solutions is available only for few cases: a layered earth, a vertical contact and a sphere in homogeneous background. For models arising in inversion a measure of accuracy has to be found.

The principle of reciprocity states, that the measured impedance is invariant to interchanging current and potential electrodes (Friedel, 2000). Since for both realizations errors occur, a quality control can be derived comparing the real and reciprocal simulation yielding the reciprocity

$$rec = 2 \frac{\mathbf{f}_{forward} - \mathbf{f}_{reverse}}{\mathbf{f}_{forward} + \mathbf{f}_{reverse}} . \quad (3.35)$$

$\mathbf{f}_{forward}$ is the simulation result as the configuration predicts and for $\mathbf{f}_{reverse}$ current and potential electrodes are changed.

A first accuracy test applies pole-pole measurements over a conductive dike of infinite strike length. Within a homogeneous half-space of $100 \Omega \text{ m}$ the dike with a resistivity of $10 \Omega \text{ m}$ is situated between $x = 20 \text{ m}$ and 25 m . The source is located at the origin, potentials are measured at the line defined by $y = 0 \text{ m}$ and $z = 0 \text{ m}$. The model itself is spaced equidistantly with $\Delta x = \Delta y = \Delta z = 5 \text{ m}$ from $x = -40..110 \text{ m}$, $y = -50..50 \text{ m}$ and $z = 0..60 \text{ m}$. Four cells with thickness enlarged by factor 4 are added to the sub-surface boundaries to satisfy the boundary conditions.

Figure 3.10 shows the numerical result compared to the analytical solution and the solution error in comparison with the reciprocity. Note that the plotted result is the geometrical mean of forward and reciprocal simulation. For most regions, the analytical solution can be approximated well within one percent deviation. Larger deviations occur near the conductivity contrast, whereas in regions of constant conductivity both accuracy and reciprocity error are low. Hence, the reciprocity suites well for estimating the accuracy of numerical simulations. However, near the right boundary errors appear to both solutions that do not affect the reciprocity.

The reciprocity as a measure of accuracy can be displayed in form of pseudo-sections to delineate areas, which have to be refined. In the following, the maximum absolute

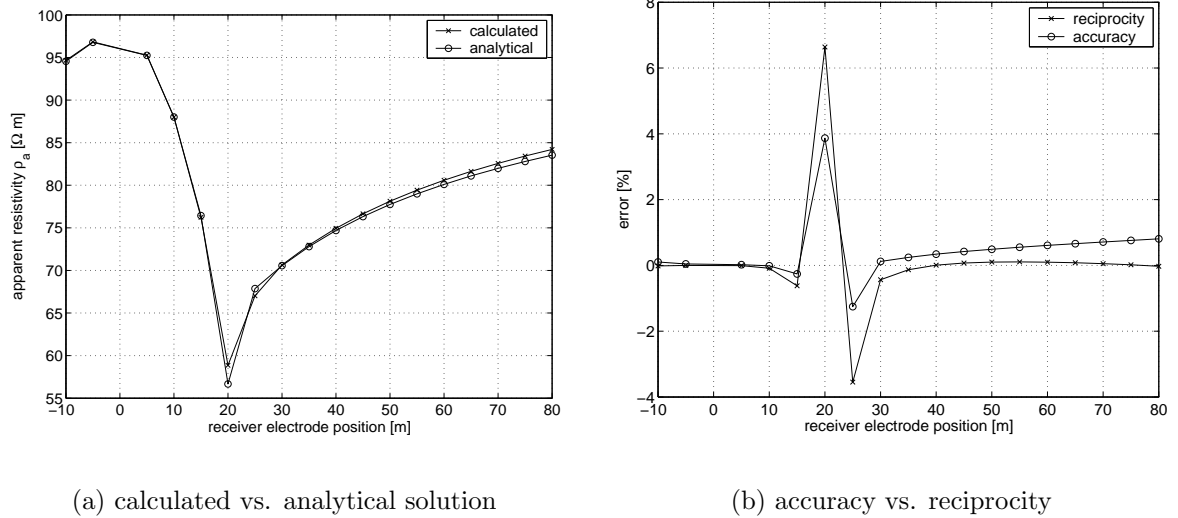


Figure 3.10: Comparison of results for the conductive dike, (a) apparent resistivity, (b) accuracy and reciprocity

value and the standard deviation of all reciprocity values are used for the estimation of simulation errors. For the above depicted example values of 6.74 % and 1.83 % were found, respectively.

Several investigations are carried out to determine the effect of refinement techniques and the resistivity contrasts on the reciprocity. Figure 3.11 displays the used model geometry. On a grid of 11×11 electrodes dipole-dipole measurements were simulated for both x- and y-direction. The model is discretized using horizontal grid spacings of $\Delta x = \Delta y = 1$ m. Six layers with thicknesses from 0.4 to 1.0 m define the vertical boundaries, the maximum depth of 4 m considers the total depth of investigation.

In a homogeneous half-space of $100 \Omega \cdot m$ several bodies of varying form and depth are located. The anomalous resistivities are defined by the contrast factor fak .

Accuracy considerations

The system of equations $\mathbf{Ax} = \mathbf{b}$ is solved via the method of conjugate gradients. The iteration process is stopped once the relative residual norm

$$\frac{\|\mathbf{Ax} - \mathbf{b}\|}{\|\mathbf{b}\|}$$

falls below an accuracy acc . To save computing time, it is necessary to find reasonable values for acc .

Discretization errors can be diminished by a node refinement. The refinement has to be applied globally and results in much larger systems of equation. For this example,

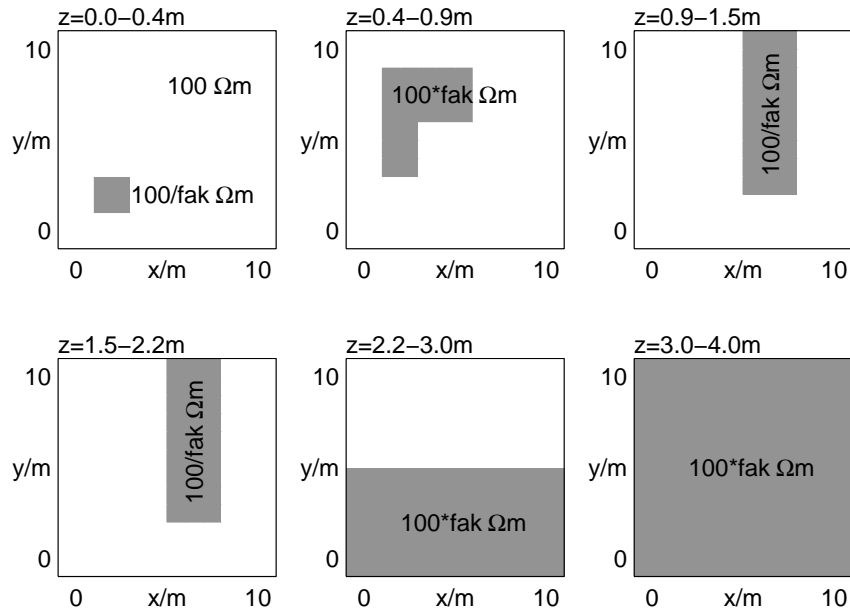


Figure 3.11: Synthetic model used for the accuracy investigations

one (refinement factor 2) or two (refinement factor 3) additional nodes are placed at the horizontal grid lines of the individual cells.

Accuracy <i>acc</i>	Unrefined		Refined by 2		Refined by 3	
	max(rec)	std(rec)	max(rec)	std(rec)	max(rec)	std(rec)
1e-1	77.6	14.7	380	29.8	108	22.3
1e-2	55.2	8.14	58.0	8.82	63.3	12.1
1e-3	36.8	5.49	28.9	4.10	20.1	3.17
1e-4	36.8	5.46	28.7	4.07	19.8	3.02
1e-5	36.8	5.45	28.7	4.07	19.8	3.02
1e-6	36.8	5.45	28.7	4.07	19.8	3.02

Table 3.2: Reciprocity (in %) as a function of refinement and solving accuracy, contrast factor $fak=5$

Table 3.2 shows both maximum and standard deviation of the reciprocity values for varying solver accuracy and refinement factors, a contrast factor of $fak=5$ is used. The largest errors of 20-40% arise at the near-surface variations, whereas the standard deviations range at several percent. The reciprocity values are significantly reduced by grid refinement. Regarding the solving accuracy it is clearly visible that values below $1e-3$ to $1e-4$ do not lead to more accurate results and thus $1e-4$ works as a reliable value used in subsequent simulations.

It is obvious, that the accuracy of the forward calculations strongly depends on the existing resistivity contrasts. In the following, the contrast is varied from 1:10 to 10:1 as denoted in Table 3.3. Note that only small errors arise for a contrast near 1 and that

resistivity contrast	1:10	1:5	1:2	4:5	5:4	2:1	5:1	10:1
maximum error	52.4	17.79	5.0	1.5	1.5	5.3	19.8	39.9
standard deviation	5.11	2.23	0.87	0.29	0.31	1.04	3.02	5.54

Table 3.3: Reciprocity (in %) as a function of the resistivity contrast between the half-space and the anomalous bodies, computations use triple refinement and $acc=1e-4$

they increase with extremal contrast factors. Consequently, the refinement has to be carried out depending on the resistivity contrasts. However, for FD grids a refinement is always global and strongly increases memory and computer time usage.

Speed of forward calculation

In the following is investigated, how the use of different preconditioning matrices affects computing time.

PC	ω or tol	nonzeros	pc time (s)	iter. time (s)	sum/mean iter.
Jacobi		8125	0.04	1384	20932/173
SSOR	1.4	31225	0.1	575	5434/44.9
$\omega =$	1.6	31225	0.1	562	5260/43.5
	1.8	31225	0.1	545	5115/42.3
CholInc	'0'	31225	2.23	83.3	2307/19.1
$tol=$	0.1	22459	0.15	93.6	2808/23.2
	0.01	48770	0.44	50.8	1274/10.5
	1e-3	145434	1.58	36.8	524/4.3
	1e-4	430993	6.51	40.7	279/2.3
	1e-5	1126411	20.0	56.0	159/1.3

Table 3.4: Comparison of different preconditioners, '0' denotes the use of the same sparsity structure

Table 3.4 shows the number of nonzero elements, the needed iteration numbers and the time consumed for different preconditioners. Although the Jacobi preconditioner allocates less memory it needs the largest number of iterations and computer time. Both are significantly reduced by the SSOR preconditioner. The optimal value of the relaxation factor is close to $\omega = 1.8$.

The incomplete Cholesky type preconditioners are superior to SSOR and Jacobi type from both iteration number and computer time point of view. With decreasing Cholesky tolerance tol the iteration numbers are significantly reduced. Although the iteration numbers are still decreasing, the computing time begins to increase for tolerances lower than $1e-3$. The reason is the larger number of non-zero elements, which increases the time for each matrix-vector multiplication. Moreover, the time used for calculating the preconditioning matrix rises significantly. Therefore a value of $tol=1e-3$ is used for subsequent simulations.

It has to be noted, that the incomplete Cholesky with same sparsity structure needs more memory and computer time than the one with tolerance 0.1. Thus, small tolerance values are to be applied instead if the computer memory is strongly limited. However, the memory needed by the forward calculation is much smaller than for storing the sensitivity matrix. For inversion with NLCG without explicit sensitivity calculation Cholesky preconditioners with small tolerances can be the method of choice for large-scale problems.

3.4.4 Finite Element Modeling

Finite Element (FE) methods have been used early to calculate electromagnetic and electric fields (Coggon, 1971; Pridmore *et al.*, 1980). Many authors like Li and Spitzer (2002) and Pain *et al.* (2003) use hexahedral elements. However, the main advantage of finite elements is the incorporation of topography (Fox *et al.*, 1980), which cannot be accomplished using hexahedrons. Tetrahedrons allow for the simulation of arbitrary geometry. Bing and Greenhalgh (2001) presented finite element simulations based on regularly arranged tetrahedral elements. Until now, only few approaches using unstructured tetrahedral grids have been published (Sugimoto, 1999).

The above denoted problems at inner boundaries are circumvented by the weak formulation. The differential equation $L_\sigma\varphi = f$ is solved in an inner product sense

$$\iiint_{\Omega} v L_\sigma\varphi d\Omega = \iiint_{\Omega} v f d\Omega \quad (3.36)$$

for certain test functions v .

Equation (3.36) can be solved for discrete functions v_h using a variational integral (Galerkin method) or by the method of weighted residuals. The potential is then constructed by form functions and the solution is obtained at the node positions of the finite elements.

As for finite differences, the singularity removal technique can be applied, for details see R ucker (2003). The primary potentials can be calculated analytically for flat-earth

problems. If topography is present, they can be obtained by forward calculation using a homogeneous conductivity. However, the total potential requires a much finer grid to ensure accurate results.

3.4.5 Forward Approximations

The sensitivity matrix as relative and absolute weighting function can be used to accomplish forward calculation. Thus, the model response vector can be approximated using the formula

$$\mathbf{f}(\mathbf{m}) \approx \mathbf{f}(\mathbf{m}^0) + \mathbf{S}(\mathbf{m} - \mathbf{m}^0) \quad .$$

However, it turns out that small errors in the sensitivity values lead to large errors in the forward approximation. The main problem is, that the sensitivity of an updated model is generally not known. Without update the mechanism represents a pure linear problem, which can be solved in one inversion step. Even if sensitivity updates like Broyden's method are used, they are under-linear and result in the known slow convergence of back-projection forward calculation algorithms.

3.4.6 2D Finite Difference Modeling

The conductivity is considered to be of infinite extent in one (let's say y -) direction. Hence, the partial derivatives with respect to y vanish as well as the coupling coefficients to the back and the front direction. The potentials are sought at the discretized x - z plane resulting in smaller computational effort. However, using the source terms as depicted would implicate current sources of infinite extent. The three-dimensionality of the point source is transformed into wavenumber domain using the Fourier-Cosine-transform of the potential

$$\tilde{\varphi}(x, k_y, z) = 2 \int_0^{\infty} \varphi(x, y, z) \cos(k_y y) dy \quad , \quad (3.37)$$

where $\tilde{\varphi}$ denotes the transformed potential and k_y is the wavenumber with respect to y . The corresponding transformation of equation (3.3) yields

$$\frac{\partial}{\partial x} \left(\sigma \frac{\partial \tilde{\varphi}}{\partial x} \right) + \frac{\partial}{\partial z} \left(\sigma \frac{\partial \tilde{\varphi}}{\partial z} \right) - k_y^2 \sigma \tilde{\varphi} = -I \delta(x - x_s) \delta(z - z_s) \quad . \quad (3.38)$$

The boundary conditions have to be transformed adequately. Equation (3.38) can then be solved using finite differences (Dey and Morrison, 1979b) or finite elements (Kemna,

2000). The solutions $\tilde{\varphi}$ for a series of k_y are then transformed back to spatial domain

$$\varphi(x, 0, z) = \frac{1}{\pi} \int_0^{\infty} \tilde{\varphi}(x, k_y, z) dk_y \quad .$$

Kemna (2000) and Rücker (1999) suggest a numerical integration using the Gauss-Legendre method for small and Gauss-Laguerre for large wavenumbers.

3.5 Calculation of Sensitivities

The sensitivity term arises inevitable in every non-linear inversion process. Even a trial-and-error inversion using subsequent modeling and forward calculations has to consider the sensitivity distribution. A single sensitivity value denotes the change of forward response f_i as a function of the model \mathbf{m} with respect to a change of the model parameter m_j

$$S_{ij}(\mathbf{m}) = \frac{\partial f_i(\mathbf{m})}{\partial m_j} \quad . \quad (3.39)$$

The term sensitivity is associated with an assumption of a model \mathbf{m} , which is never the truth⁵. Since it is a function of the model, we can never speak about "the sensitivity", even though the term is often used for the one of the homogeneous half-space or an a-priori model.

Arranging the sensitivities of all forward responses with respect to all model parameter in a matrix yields the Jacobian (or sensitivity) matrix

$$\mathbf{S} = \begin{pmatrix} \frac{\partial f_1}{\partial m_1} & \cdots & \frac{\partial f_1}{\partial m_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial m_1} & \cdots & \frac{\partial f_N}{\partial m_M} \end{pmatrix} \quad , \quad (3.40)$$

used in non-linear inversion for updating the model vector.

The sensitivities of all existing model parameters plotted in model space is referred to as sensitivity distribution. It provides a clear view into the physical processes and thus represents a powerful tool for the comprehension and the interpretation of measurements. In the last decade, numerous papers on sensitivities of electromagnetic measurements were published (Gomez-Trevino, 1987; Spies and Habashy, 1995; Dorn *et al.*, 2002). Also, for DC resistivity data the sensitivity have been investigated (Friedel, 1997; Dietrich, 1999; Spitzer, 1998).

⁵As we do not know the actual model producing our data, we can not speak of the "sensitivity of the data".

The sensitivity distribution of one single data primarily denotes regions, which are sensitive to the measurement. Following travel-time tomography, the sum of the sensitivities of all measurements yields a cumulative sensitivity or coverage

$$\text{cov}_j = \sum_{i=1}^N S_{ij} \quad .$$

The comparison of sensitivity distributions for the individual data provides a concept of how the model parameters can be distinguished from each other. When designing data sets, a trade-off between the abilities of detecting and distinguishing has to be sought.

3.5.1 Derivation of DC Sensitivities

Using the reciprocity principle of electrical impedance tomography it has been shown by Geselowitz (1971), how the impedance Z changes due to a change of the conductivity $\delta\sigma$ in the region Ω_i

$$\delta Z = -\delta\sigma \iiint_{\Omega_i} \frac{\nabla\varphi}{I_\varphi} \cdot \frac{\nabla\psi}{I_\psi} d^3r \quad . \quad (3.41)$$

φ is the potential caused by injecting the current I_φ at the current electrodes. ψ and I_ψ denote the potential and the current taking the potential electrodes for current injection. Using the forward response $\rho_a = kZ$ and the parameter ρ , for the dimensionless sensitivity holds the relationship

$$S = \frac{\delta\rho_a}{\delta\rho} = \frac{k}{\rho^2} \iiint_{\Omega_i} \frac{\nabla\varphi}{I_\varphi} \cdot \frac{\nabla\psi}{I_\psi} d^3r \quad . \quad (3.42)$$

Homogeneous half-space sensitivities

For the case of a homogeneous half-space the potentials are known analytically. Hence for a four point measurement array holds

$$S_{hom} = \frac{\delta\rho_a}{\delta\rho} = \frac{k}{4\pi^2} \iiint_{\Omega_i} \left(\frac{\vec{r}_a}{r_a^3} - \frac{\vec{r}_b}{r_b^3} \right) \cdot \left(\frac{\vec{r}_m}{r_m^3} - \frac{\vec{r}_n}{r_n^3} \right) d^3r \quad , \quad (3.43)$$

where $\vec{r}_a = \vec{r} - \vec{r}_A$ etc. are the relative vectors of the position vector \vec{r} with regard to the electrode positions. Note that the sensitivity is independent of the half-space resistivity ρ . The integral has to be solved by numerical integration. The Gauss-Legendre integration proves to obtain accurate and robust results. A variable number

x=	y=	z=	1 point	2 ³ = 8 points	3 ³ = 27	5 ³ = 125	exact value
2-3	0-1	2-3	-0.8615	-0.8690	-0.8689	-0.8689	-0.8689
3-4	0-1	0.5-1	-0.8934	-0.7693	-0.7773	-0.7771	-0.7771
4-5	0-1	0-0.5	-9.7123	-12.7080	-12.7323	-12.6556	-12.5665

Table 3.5: Results (10^{-3}) of the Gauss-Legendre integration of the sensitivity for a hexahedral region with respect to a pole-pole measurement ($R_A = (-5, 0, 0)^T$, $R_M = (5, 0, 0)^T$)

of sampling points representing polynomial nulls is distributed for three directions. At every point the integrand is calculated, multiplied with the three weighting factors of the sampling point and summed up.

As example, Table 3.5 shows approximated sensitivity values for a variable number of sampling points. It can be seen, how for deeper cells an approximation of 2³ or 3³ points provides reasonable results. For most purposes a 5³ approximation is sufficiently accurate, for cells adjacent to electrodes up to 10³ points per cell are recommended.

Two-dimensional sensitivity integration

In two-dimensional inversion the model cells are of infinite length in y-direction. The unbounded integration with respect to y can be transformed to a bounded integration using elliptic integrals of first and second order K and E . Friedel (2000) showed, that the sensitivity function of a pole-pole array can be written as

$$S(x, z) = \frac{1}{4\pi^2} \{T_1[(x - x_A)(x - x_M) + z^2] + T_2\} \quad , \quad (3.44)$$

where x_A and x_M are the positions of the current and potential electrode. T_1 and T_2 read

$$T_1 = \frac{2}{ab^2(a^2 - b^2)^2} [(a^2 + b^2)E(\eta) - 2b^2K(\eta)] \quad \text{and}$$

$$T_2 = \frac{2}{a(a^2 - b^2)^2} [(a^2 + b^2)K(\eta) - 2a^2E(\eta)] \quad .$$

a^2 and b^2 are maximum and minimum of the values $(x - x_A)^2 + z^2$ and $(x - x_M)^2 + z^2$, respectively and $\eta = 1 - b^2/a^2$.

3.5.2 Sensitivities for Inhomogeneous Conductivities

The presence of conductivity contrasts in the model space changes the electric fields and thus also the sensitivities. In general, there exist three ways for the sensitivity calculation of the general case (Spitzer, 1998):

The perturbation method

This is a very obvious procedure often used for problems with few parameters. The partial derivative in equation (2.3) is approximated by the finite difference equation

$$S_{ij} \approx \frac{f_i(\mathbf{m} + \Delta m \delta_j) - f_i(\mathbf{m})}{\Delta m} .$$

δ_j is the discrete Dirac function with 1 at the position j and 0 elsewhere. Subsequently, all the model cells are perturbed with a small variation Δm and the difference in the forward response, related to the variation, yields one column of the sensitivity matrix. If the model response $\mathbf{f}(\mathbf{m})$ is already known, M additional forward calculations including all potential electrodes have to be carried out. Since the forward calculation is a time consuming procedure, this method seems quite impractical.

Sensitivity forward calculation

The forward problem is usually solved by finite differences or finite elements resulting in a system equations, which has to be solved for the potential vector φ .

$$\mathbf{K} \cdot \varphi = \mathbf{b} \tag{3.45}$$

The matrix \mathbf{K} denotes the coupling (FD) or stiffness (FE) matrix representing discretization and conductivity distribution, the vector \mathbf{b} is defined by the (primary or secondary) sources. Following Spitzer (1998), equation (3.45) can be differentiated with respect to the cell conductivity σ_i . While the source term \mathbf{b} is independent of σ_i , both \mathbf{K} and φ are not, yielding

$$\mathbf{K} \cdot \frac{\partial \varphi}{\partial \sigma_i} = -\frac{\partial \mathbf{K}}{\partial \sigma_i} \varphi . \tag{3.46}$$

Equation (3.46) reads as additional forward calculations with the same matrix and different source function. Since each element of the matrix \mathbf{K} contains the neighboring conductivities of the according node, sensitivity sources occur only at the nodes surrounding the element i . However, for each model cell all sources have to be considered, resulting in solving $M \times N$ single forward equations. The numerical effort can be significantly reduced using fast preconditioners. In multi-dimensional inversion the number of model parameters is usually higher than the number of data. Thus this method is seldom useful in practice.

Potential approximation

Assume that the potentials at the cell corner points are known from the DC forward calculation. Then equation (3.42) can be used replacing the partial derivatives by finite

differences, e.g. for

$$\frac{\partial \varphi_{i+1/2,j,k}}{\partial x} \approx \frac{1}{4\Delta x} (\varphi_{i+1,j,k} + \varphi_{i+1,j+1,k} + \varphi_{i+1,j,k+1} + \varphi_{i+1,j+1,k+1} - \varphi_{i,j,k} - \varphi_{i,j+1,k} - \varphi_{i,j,k+1} - \varphi_{i,j+1,k+1}) \quad . \quad (3.47)$$

This method yields satisfactory results for deep cells. Since the potential decreases under-linearly proportionally to $1/r$, both the gradients and the sensitivity are over-estimated for small distances to the electrode. Refining the grid can overcome this problem slightly resulting in a huge effort for the forward calculation. Another important issue is the fact, that the potential at the source points is singular. Generating a model grid excluding the electrodes handles this problem, but introduces modeling errors due to interpolation and therefore also affects the sensitivity calculation.

3.5.3 Secondary Potential Sensitivity Calculation

If the singularity removal technique is applied, one can use the decomposition of the potentials φ (and, by analogy, ψ)

$$\varphi = \varphi_p + \varphi_s \quad \text{and} \quad \psi = \psi_p + \psi_s \quad . \quad (3.48)$$

Inserting equations (3.48) into (3.42) yields

$$S = \frac{\delta \rho_a}{\delta \rho} = \frac{k}{I^2 \rho^2} \iiint_{\Omega_i} (\nabla \varphi_p + \nabla \varphi_s) \cdot (\nabla \psi_p + \nabla \psi_s) d^3 r \quad . \quad (3.49)$$

By multiplying the sums and changing integration and summation we obtain a sum of four integrals

$$S = \frac{k}{I^2 \rho^2} \left[\iiint_{\Omega_i} \nabla \varphi_p \cdot \nabla \psi_p d^3 r \quad + \quad \iiint_{\Omega_i} \nabla \varphi_p \cdot \nabla \psi_s d^3 r \quad + \right. \\ \left. \iiint_{\Omega_i} \nabla \varphi_s \cdot \nabla \psi_p d^3 r \quad + \quad \iiint_{\Omega_i} \nabla \varphi_s \cdot \nabla \psi_s d^3 r \right] \quad (3.50)$$

$$= S_1 + S_2 + S_3 + S_4 \quad . \quad (3.51)$$

Integral S_1 is constructed by the known primary potentials and represents the half-space sensitivity of equation (3.43). It contains the main sensitivity distribution and is generally of higher magnitude than the integrals S_2 to S_4 . Those are produced by conductivity deviations and can be calculated in addition to the integrated S_1 , which has already been obtained at the starting point of inversion.

Using Green's identity $\nabla \cdot (a\nabla b) = \nabla a \cdot \nabla b + a\nabla^2 b$ yields for S_2

$$S_2 = \iiint_{\Omega_i} \nabla \varphi_p \cdot \nabla \psi_s d^3r = \iiint_{\Omega_i} \nabla \cdot (\varphi_p \cdot \nabla \psi_s) d^3r - \iiint_{\Omega_i} \varphi_p \nabla^2 \psi_s d^3r \quad .$$

If $\nabla \psi_s$ is considered to be constant⁶ in Ω_i , the second term at the right hand side vanishes. The first term can be transformed from a volume integral into a surface integral using Gauss' theorem resulting in

$$S_2 = \iint_{\Gamma_i} \varphi_p \nabla \psi_s \cdot \vec{n} d\Gamma_i \quad , \quad (3.52)$$

where Γ_i is the surface of the model cell Ω_i and \vec{n} denotes its normal vector. For each face of the cell, independent of geometry, the gradient of ψ into the normal direction has to be multiplied with an integral value of φ_p . S_3 is calculated as S_2 interchanging φ and ψ . The integral S_4 caused by the secondary potentials φ_s and ψ_s is of the same type, but expected to have lower magnitude, because the secondary potentials are usually smaller than the primary ones.

In summary, the sensitivity is calculated by adding three terms of type (3.52) to the half-space sensitivity. The additional terms are much more accurate than with equation (3.42), because the potential is smoother than its gradient. Moreover, the surface integration can be performed with a smaller number of sampling points than the volume integral. Prerequisite are the secondary potentials φ_s and ψ_s , which are known from the forward calculation, if the potential of every electrode is calculated and kept in memory for the node positions.

3.5.4 Sensitivities for Logarithms

When using logarithms of the resistivities as model parameter, the sensitivity has to take this into account. Consider the model parameter $m_j = \log \rho_j$ and forward responses $f_i = \log \rho_i^a$. Thus by applying the chain rule for the sensitivity we yield

$$S_{ij}^{log} = \frac{\partial \log \rho_i^a}{\partial \log \rho_j} = \frac{\partial \rho_i^a}{\partial \rho_j} \cdot \frac{\partial \log \rho_i^a}{\partial \rho_i^a} / \frac{\partial \log \rho_j}{\partial \rho_j} = \frac{\rho_j}{\rho_i^a} \cdot S_{ij}^{lin} \quad . \quad (3.53)$$

Thus the sensitivities for linear parameters S^{lin} have to be multiplied with the cell resistivity and divided by the forward response. For the homogeneous half-space holds $S^{log} = S^{lin}$, because a constant resistivity ρ_0 causes a constant apparent resistivity of

⁶This can be physically motivated by the concept of positive and negative charge accumulations forming a capacitance with a constant electric field inside.

$\rho_a = \rho_0$. Therefore logarithms are generally applied to both data and forward responses to avoid a one-sided scaling.

Note that with the use of a lower logarithmic barrier resistivity ρ_l the sensitivity S^{low} of $f_i = \log(\rho_i^a - \rho_l)$ with respect to $m_j = \log(\rho_j - \rho_l)$ equation (3.53) changes to

$$S_{ij}^{low} = \frac{\rho_i^a - \rho_l}{\rho_j - \rho_l} \cdot S_{ij}^{lin} \quad .$$

3.5.5 Example

Figure 3.12 shows the section of a sample model. The model is discretized using $x = -10, -9, \dots, 10$ m, $y = -5, -4, \dots, 5$ m and increasing layer thicknesses of 0.3–1.6 m. Within a half-space of 100Ω m a rectangular body of 500Ω m is located between $y = -1$ and 1 m.

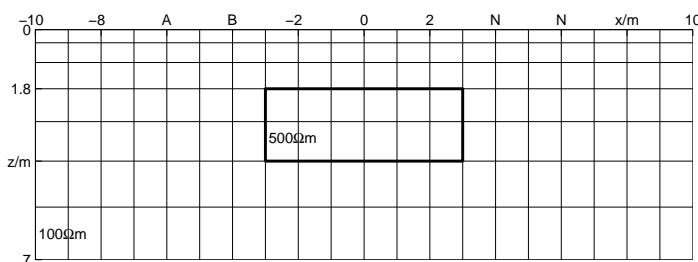


Figure 3.12: Model section at $y=0$ m

At first, the sensitivities are determined for homogeneous conductivity using Gauss-Legendre integration with an adaptive number of $2^3 - 10^3$ sampling points per direction (Figure 3.13). It shows the typical behavior with high sensitivities beneath the dipoles. Note that the sensitivity holds for both linear and logarithmic parameters and data.

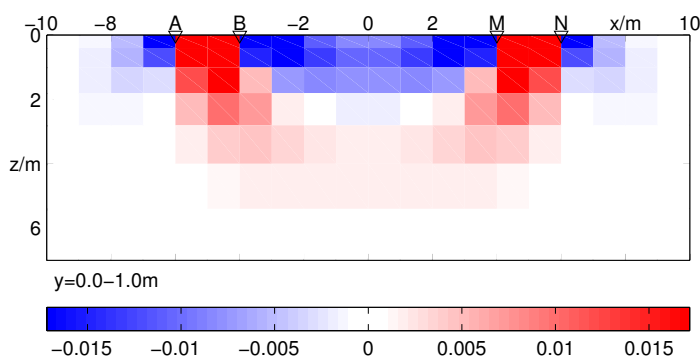


Figure 3.13: Section of the sensitivity distribution at $y=0-1$ m for the homogeneous half-space

Based on the forward calculation for the above depicted model the secondary potential technique is used to calculate sensitivities for the inhomogeneous case. Figure 3.14 shows the sensitivity distribution for the logarithmic parameters. It can clearly be seen, how the sensitivity in the body is increased. However, the main effect originates from the scaling operation and not from the secondary potentials.

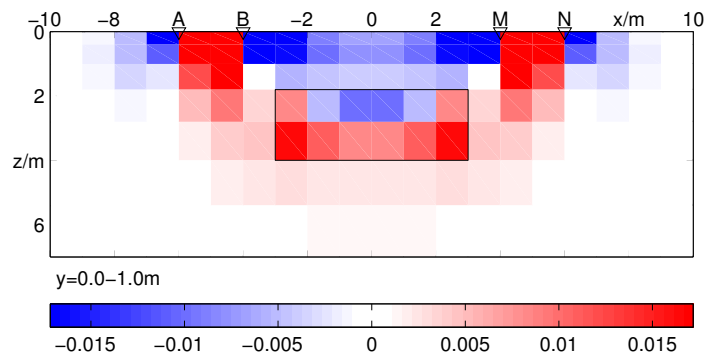


Figure 3.14: Sensitivity for the above described model, logarithmic data and parameter

3.6 Conclusions

Basis for a successful inversion of measured data are:

1. A comprehension of errors arising from DC measurements.
2. Procedures for an appropriate parameterization of the modeling domain.
3. A fast and accurate forward calculation for arbitrary resistivity distributions.
4. Methods for calculating the Jacobian matrix.

All of these aspects have been investigated in this chapter. Errors of DC data are caused by uncertainties of the measured voltages and the geometry of the measurement. Consequently, an error estimation was presented, which takes into account both configuration factors and electrode variations. If the electrode positions show variations, the data will correlate, which has to be considered in the inversion. It has been shown that both geometrical and voltage errors of profile data can be reduced by enlarging the dipole lengths for increased separations.

For flat-earth problems a grid-based parameterization is suggested. It enables the construction of model constraints as well as an easy forward calculation. Practical considerations and sensitivity studies are used to construct appropriate grids. For this purpose, two model types were presented.

The forward calculation for flat-earth problems can be carried out using finite difference methods. The resulting system of equations is efficiently solved by conjugate gradients using incomplete Cholesky preconditioners. The accuracy of the forward procedure affects the quality of the inversion result and can be assessed by considering the reciprocity of the forward response. The application of the singularity removal technique provides sufficient accuracy at moderate computing time. Nevertheless, the forward calculation is the most time-consuming procedure of inversion.

Finally, methods for the calculation of sensitivity were presented. For homogeneous models a Gauss-Legendre integration with variable number of sampling points is used. Furthermore it was shown, how the secondary potentials obtained by the singularity removal technique can be used for sensitivity approximation.

4 Inversion and Resolution Studies

In the following chapter systematic inversion and resolution studies are carried out. The inversion techniques of chapter 2 are combined with the DC resistivity methods of chapter 3. At this point the quest for optimal inversion strategy and setting parameters arises. The answers to the following questions are sought:

1. Is the Gauss-Newton method superior to the NLCG scheme concerning convergence speed, quality of inversion results and expressiveness?
2. Which regularization methods are suitable for the inverse resistivity problem?
3. How can the regularization parameter be estimated appropriately and in which way it has to be changed during inversion?
4. Is there a need for line search procedures and how can they be implemented?
5. How do local and global regularization differ in the course of iterations?
6. How do different noise conditions affect the inversion results and is it necessary to consider data covariances?

Model studies of well-defined parameter structures assuming realistic noise can yield valuable information. For reasons of clarity most of the systematic studies are based on the inversion of 2D profile data, whose results can be easily visualized. Since two-dimensional inverse problems are generally of small scale compared to the three-dimensional case, the resolution computation via the generalized singular value decomposition as denoted in section 2.4 can easily be applied. Considering resolution the following questions arise:

1. Does the defined information content represent an appropriate measure for the quality of the inversion result?
2. How do different regularization schemes affect resolution properties?
3. Which data sets are superior concerning information content and efficiency?
4. In which way does the model parameterization affect the resolution properties?

We know the resolution matrix to be a function of the model. In this point resides the general problems with resolution. To obtain a model independent resolution¹ we consider the resolution based on the half-space sensitivity, which is representative for small contrast problems. However, the crucial point is the choice of the regularization parameter, which strongly affects resolution. Therefore, the inverse procedure is linearized for reasons of comparability. As a result, the discrepancy principle can be applied to determine λ exactly. It allows an incorruptible estimation of information content and efficiency.

In section (4.2), three-dimensional problems are considered. Practical considerations for the three-dimensional inversion are figured out. The results of the 2D experimental design are applied to the 3D case. Two typical 3D electrode layouts, an electrode grid and a set of parallel profiles, are investigated regarding inversion results and resolution properties. Different parameterization techniques are compared concerning inversion results and information content.

Finally, in section 4.3 an inversion scheme is presented, which is able to incorporate topography. A triple-grid technique is used to combine a resolution-dependent model parameterization with accurate forward calculations using finite elements. The inversion scheme is proved by a synthetic study of a flood protection dike problem.

4.1 Two-dimensional Inversion Studies

4.1.1 Numerical Setup

Model Parameterization

Figure 4.1 shows the synthetic model investigated in this section. It is equally discretized in x from -1 to 42 m and in z from 0 to 6 m using block sizes of $1\text{m} \times 1\text{m}$ and contains six bodies within a homogeneous background of $100 \Omega \text{ m}$ with a contrast factor fak . The anomalies possess resistivities of $100/fak$ and $100 \cdot fak \Omega \text{ m}$, respectively.

Model parameters and data are the logarithms of the cell resistivities and the apparent resistivities, respectively.

Data and noise

The following numerical investigations employ 42 equally spaced surface electrodes at the positions $x = 0, 1, \dots, 41 \text{ m}$. Note that the results of the simulations can easily be

¹The term "resolution" is often used, in analogy with "sensitivity", in connection with the homogeneous half-space.

transferred to larger problems by a scale factor, which changes the configuration factor k and therefore the error level $\delta\rho^a$.

4.1.2 Inversion Methods and Resolution Analysis

Inversion methods

At first, the non-linear conjugate gradient method is compared to the regularized Gauss-Newton scheme. A Wenner- β data set (c-c-p-p) with $n = 1 - 13$ is simulated resulting in 273 single data. The used contrast factor $fa_k = 5$ provides data anomalies that can clearly be distinguished from noise. The sensitivity matrix is updated using the BFGS method. Regularization is achieved by smoothness constraints of first order. The Gauss-Newton minimization is applied in two ways using global and local regularization. The regularization parameters are chosen such that the data are fitted well ($\lambda = 30$ for the local and $\lambda = 5$ for the global scheme). For all three methods exact line search is carried out. It is based on χ^2 for NLCG and the local regularization scheme, whereas the line search of the global regularization uses the total functional Φ to be minimized.

Figure 4.2 shows the development of χ^2 in the course of iterations for all methods. Clearly the NLCG method shows slow convergence, whereas for both local and global scheme a reasonable data fit is reached by iteration 5. This is evidenced by the first iteration models as displayed in Figure 4.3. Whereas the Gauss-Newton inversion shows the essential model characteristics, the NLCG model can only provide a crude image of the subsurface.

The NLCG method needs a lot of inversion steps with one forward calculation per iteration. Therefore it is disadvantageous if the forward process is time-consuming as in three-dimensional DC inversion. Moreover, the model cannot be controlled actively and resolution properties cannot be easily derived. Therefore, in the following the Gauss-Newton method is used.

Figure 4.4 displays the final inversion results of local and global regularization. Both indicate the anomalous structures well. Note that the global scheme shows few artifacts near the boundary and a more compact image of the deep conducting body, whereas the anomalies are stronger for the local scheme. This seems reasonable, because the global scheme really tries to keep the model smooth and thus the solutions show less variation.

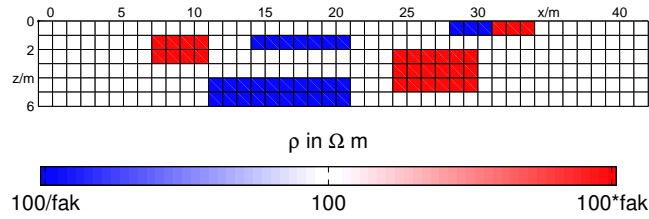


Figure 4.1: Model parameterization and synthetic model consisting of six bodies of $100/fak \Omega m$ or $100*fak \Omega m$ in a homogeneous half-space of $100 \Omega m$

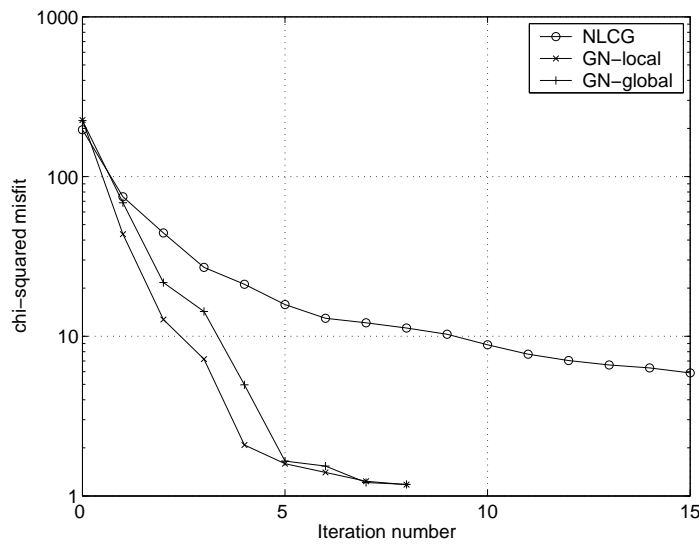


Figure 4.2: Convergence of NLCG and the Gauss-Newton method, local and global regularization

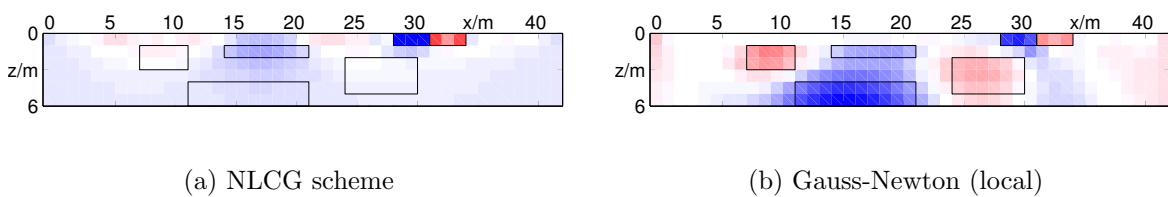


Figure 4.3: First iteration models for NLCG and Gauss-Newton

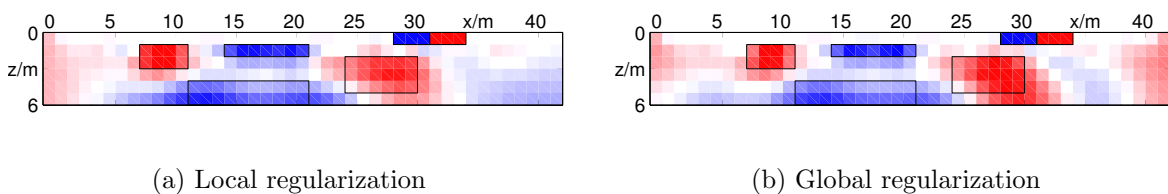


Figure 4.4: Inversion results for regularization Gauss-Newton scheme

Line search

The more non-linear the problem is, the more a line search procedure is needed to ensure convergence. Ongoing from the model step $\Delta \mathbf{m}^k$ the value $0 < \tau < 1$ minimizing $\chi^2(\mathbf{m}^k + \tau \Delta \mathbf{m}^k)$ is sought. Three approaches are presented:

Exact line search For a series of τ_n the model responses are calculated exactly. This results in many forward calculations, which is prohibitive for large-scale problems.

2-point parabola The model responses are calculated for $\tau = 0.5$ and $\tau = 1$ and are used to determine $\chi^2(\tau = 0.5)$ and $\chi^2(\tau = 1)$. Since $\chi^2(\tau = 0)$ is already known, a parabola is uniquely defined by the three points². Its minimum can be determined easily. At most 2 additional forward responses are needed.

linear interpolation The individual model responses for a series of τ_n are determined by linear interpolation between $\mathbf{f}(\tau = 0)$ and $\mathbf{f}(\tau = 1)$. By the interpolated $\mathbf{f}(\tau) \approx \mathbf{f}(0) + \tau(\mathbf{f}(1) - \mathbf{f}(0))$ the χ^2 values are defined, whose minimum has to be found. Also, a bisection scheme can be used. At most one additional forward routine is applied.

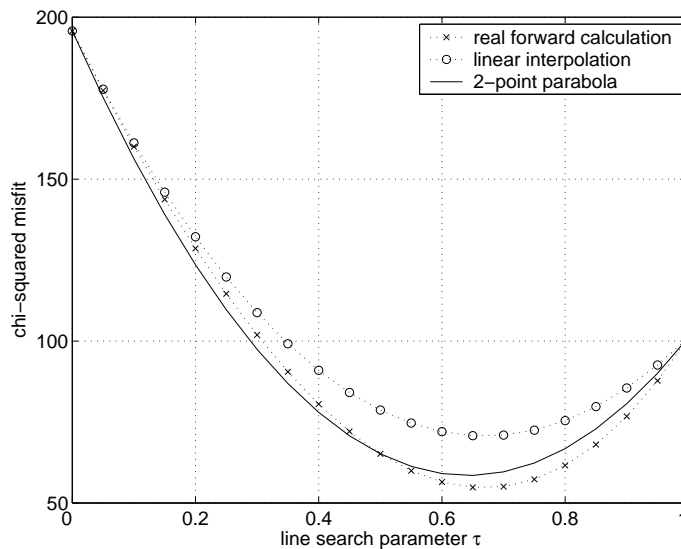


Figure 4.5: Comparison of line search procedures, the exact line search (crosses) is approximated by a two-point parabola (straight line) and by linear interpolation (circles)

Figure 4.5 shows the chi-squared misfit as a function of τ for the three methods. All curves show a broad minimum close to $\tau \approx 0.65$ indicating an essential non-linearity

²The parabola is defined if the three points do not lie on a straight line.

of the problem. Although the two-point parabola approximates the exact values much better, the linear interpolation provides sufficient results for the estimation of the line search parameter as well. Therefore, the latter method is used subsequently. Note that the model response has to be re-calculated if $\tau < 1$.

The L–Curve

Assuming solutions of the inverse subproblem for a set of different λ_i are known. To find a trade-off between data fit and model roughness, the model responses have to be calculated. Since the local regularization treats the sub-problem as independent inverse problem, we can investigate the two norms

$$\|\mathbf{D}(\mathbf{S}\Delta\mathbf{m} - \Delta\mathbf{d})\| \quad \text{and} \quad \|\mathbf{C}\Delta\mathbf{m}\| \quad .$$

Figure 4.6 shows the L–curve as arising in the first iteration. The curvature of the curve clearly indicates a maximum for $\lambda = 43.98$, which is not far from the above used value of 30. However, it seems practical only for well linearizable problems, because

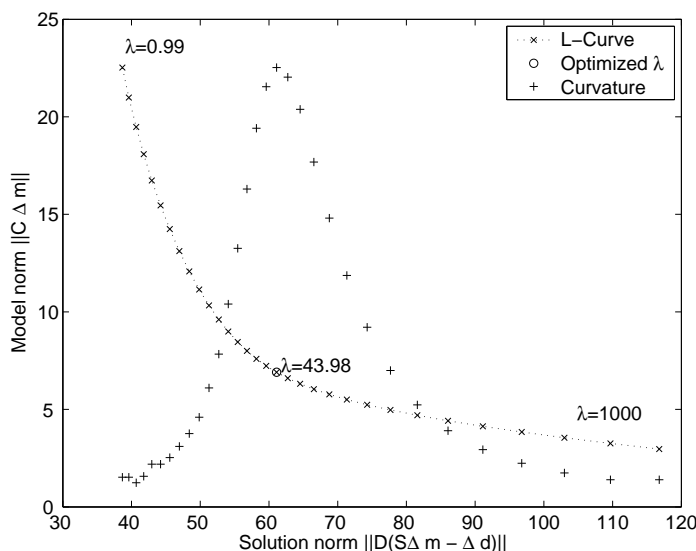


Figure 4.6: The L–curve of the linear sub-problem

the line search procedure is not included into the L–curve. The more non-linearity there is in the problem, the more the determined λ is over-estimated due to better data fits of the damped solutions.

For global regularization the L–curve has to consider the exact model responses, which are also affected by the line search parameter. In Figure 4.7 the data misfit (local regularization) and the functional Φ (global regularization) of the first inversion step are displayed for varying λ and τ . Note that the main structure of Φ is determined by

the data misfit χ^2 . However, the minimum of the the global scheme is shifted towards smaller λ . Dependent on a fixed λ the minimum of each column denotes the optimum

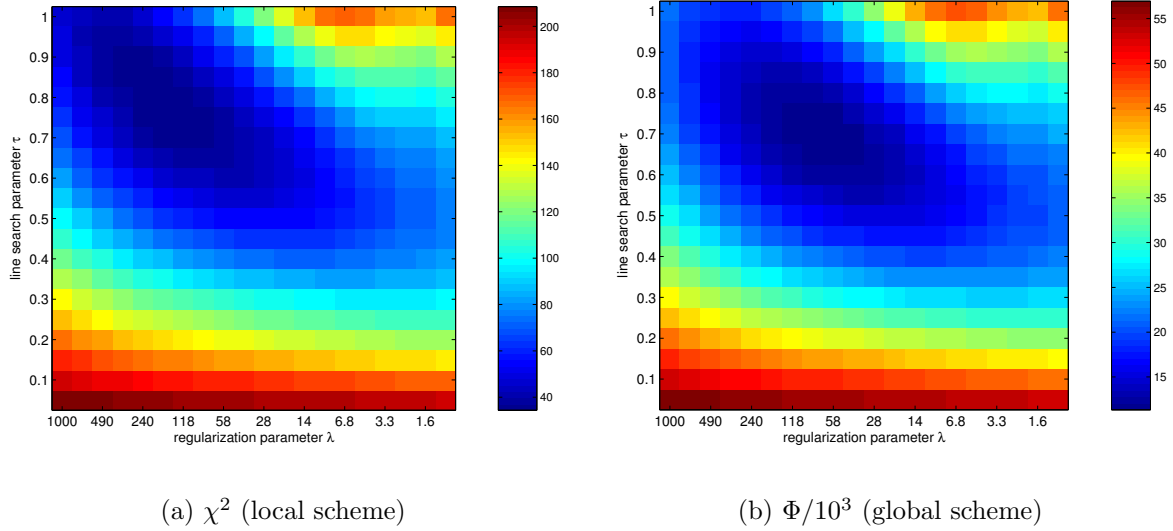


Figure 4.7: Data misfit χ^2 and total functional Φ as a function of λ and τ

line search. For strong regularization (large λ) the problem is nearly linear ($\tau \approx 1$), whereas for small regularization parameters the line search is essential to accelerate convergence. Note that the overall minima of both functions indicate too large values for λ , which do not lead to appropriate data fit. However, a bivariate search for optimum values of λ and τ is prohibitive. Consequently, the author recommends the use of the discrepancy principle as follows: A global minimization including line search using a fixed λ is carried out until convergence is reached. The final χ^2 reveals if the regularization parameter has been chosen appropriately. If χ^2 exceeds 1, λ has to be decreased, otherwise it has to be increased. Thus, a global L-curve is constructed and the process is stopped if $\chi^2 \approx 1$.

Resolution analysis

Resolution properties are investigated in the following. Basis are the sensitivity and the regularization parameter of the global minimization. The model resolution matrix \mathbf{R}^M is calculated using the generalized singular value decomposition as described in section 2.4. Figure 4.8 shows the model resolution, which constitutes of the main diagonal elements \mathbf{R}_{ii}^M . Those can be interpreted as individual reconstructabilities. Whereas near-surface cells are resolved almost perfectly, the resolution decreases drastically with depth and outside the electrodes.

The individual columns of \mathbf{R}^M reveal, how anomalies in the individual model cells

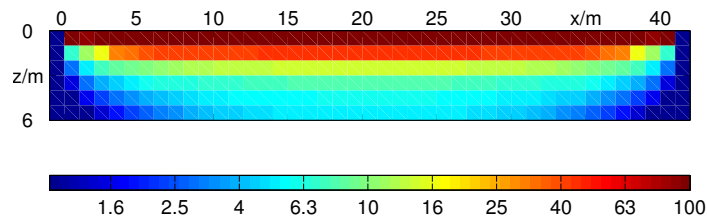


Figure 4.8: Model resolution of the individual model cells (in %)

are mapped into our model estimate. Figure 4.9 displays the cell resolutions for four selected model cells. The top cell is reproduced very well, whereas the shallow cell shows slight variations at left, right and bottom. Generally, resolution decreases with depth. For the deep cell the vertical resolution is better than the horizontal one. The artifacts arising for the resolution of the boundary cell point out, how careful the model boundaries have to be treated.

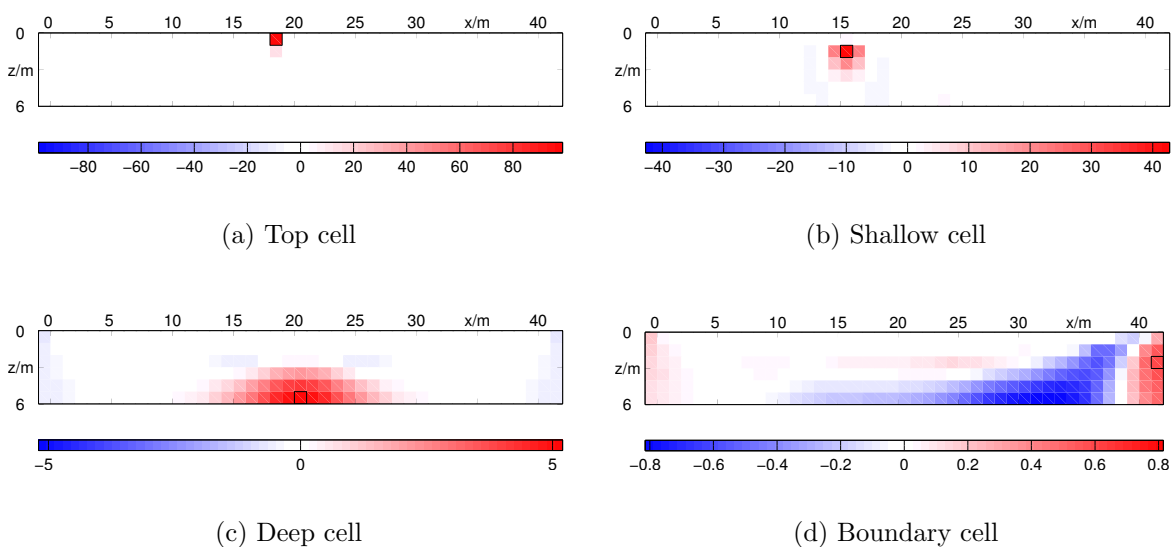


Figure 4.9: Model cell resolutions (in %) for 4 selected parameters, the cells are marked by black rectangles

From the resolution equation (2.42) we know the model being filled up with constraint information. Thus, badly resolved cells show a smooth resolution kernel. The sharpness of the model resolution is affected by the regularization strength. Figure 4.10 displays the resolution degree, which is the information content divided by the number of model parameters, as a function of the regularization parameter λ . For logarithmically increasing λ it decreases nearly linearly.

To make quantitative statements regarding information content and resolution properties, the regularization parameter has to be determined appropriately. The essential

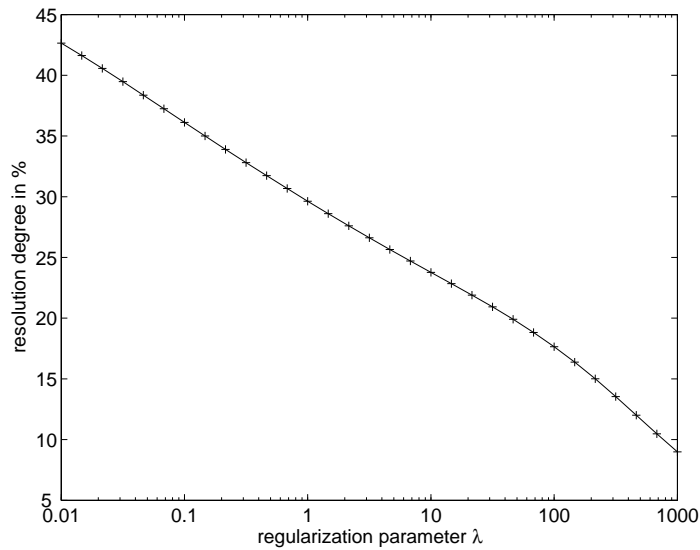


Figure 4.10: Resolution degree as a function of the regularization parameter

argument using synthetic data with known noise is to apply the discrepancy principle such that the data are fitted within noise ($\chi^2 = 1$). However, in non-linear inversion the data cannot be fitted in one inversion step. Therefore, the inversion scheme is linearized to avoid the dependence on regularization and line search parameters.

4.1.3 Comparative Linearized Resolution Study

Linearization procedure

Following the scheme of Friedel (2003), the Jacobian is assumed to be invariant of the model and determined by the homogeneous half-space. It is calculated by numerical integration and used to produce the synthetic data by forward approximation. Hence the non-linear problem is transferred into a linear one and can be solved in one iteration step without using line search. The global and the local regularization scheme are identical. Moreover, the resolution measures become independent from the model. Note that a linearization is valid only for small contrasts. Therefore, a relative small parameter contrast of $fak = 2$ is applied. Nevertheless, the linearized inversion reveals the main resolution properties.

Investigated data sets

Several widely used electrode configurations are investigated: pole-pole, pole-dipole, dipole-dipole, Schlumberger and Wenner (in form of α , β and γ arrays). Also, data sets with increased dipole lengths of $\overline{MN} = \text{mod}(n + 3, 4) \cdot a$ are considered. This holds

for pole-dipole, dipole-dipole and Schlumberger types and is connected with a data reduction, because the dipole length equals the moving distance between subsequent measurements. While information is lost by reducing the data set in such a way, additional data can be obtained by measuring so-called circulated data as presented by Friedel (2000).

Table 4.1 shows the electrode arrangements investigated. For reasons of comparability, the maximum separation is chosen such that the number of data are approximately equal (269-300). The synthetic data are displayed in Figure 4.11. Note that the pole-dipole and dipole-dipole types generally show the largest anomalies.

Dataset	tag	n	data	max(k)	min/max ρ_a
Pole-pole	c-p	1-8	300	50.3	79.3/143.4
Pole-dipole	c-p.p	1-8	292	452.4	50.0/211.2
(increased dipoles)	c-p:p	1-15	280	447.7	50.0/210.8
(+circulated data)	c-p:p+c	1-10	289	449.2	50.0/210.8
Dipole-dipole	c.c-p.p	1-8	284	2261.9	41.6/238.9
(increased dipoles)	c:c-p:p	1-15	265	1287.1	41.6/238.9
(+circulated data)	c:c-p:p+c	1-10	271	1319.5	41.6/238.9
Schlumberger	c-p.p-c	1-10	300	345.6	61.1/181.8
(increased dipoles)	c-p:p-c	1-15	285	1287.1	73.1/167.6
Wenner- α	c-p-p-c	1-13	273	81.7	63.1/165.1
Wenner- β	c-c-p-p	1-13	273	122.5	41.9/281.2
Wenner- γ	c-p-c-p	1-13	273	245.0	41.6/238.9

Table 4.1: Definition of data sets according to the above depicted classification

Simulation

The data are contaminated with Gaussian noise of standard deviation σ , which is made up of a relative error ε and a voltage error δU for a given current I .

The error-weighted sensitivity \mathbf{DS} is decomposed using the SVD or the GSVD. Once the eigenvector matrices are obtained, solutions of the inverse subproblem can be calculated easily. By a bisection scheme the parameter λ is determined in such a way that $\chi^2 = 1$. The procedure of adding noise and determining λ is repeated 10 times to get independent of the noise realization. The mean λ is used to compute model and resolution matrix.

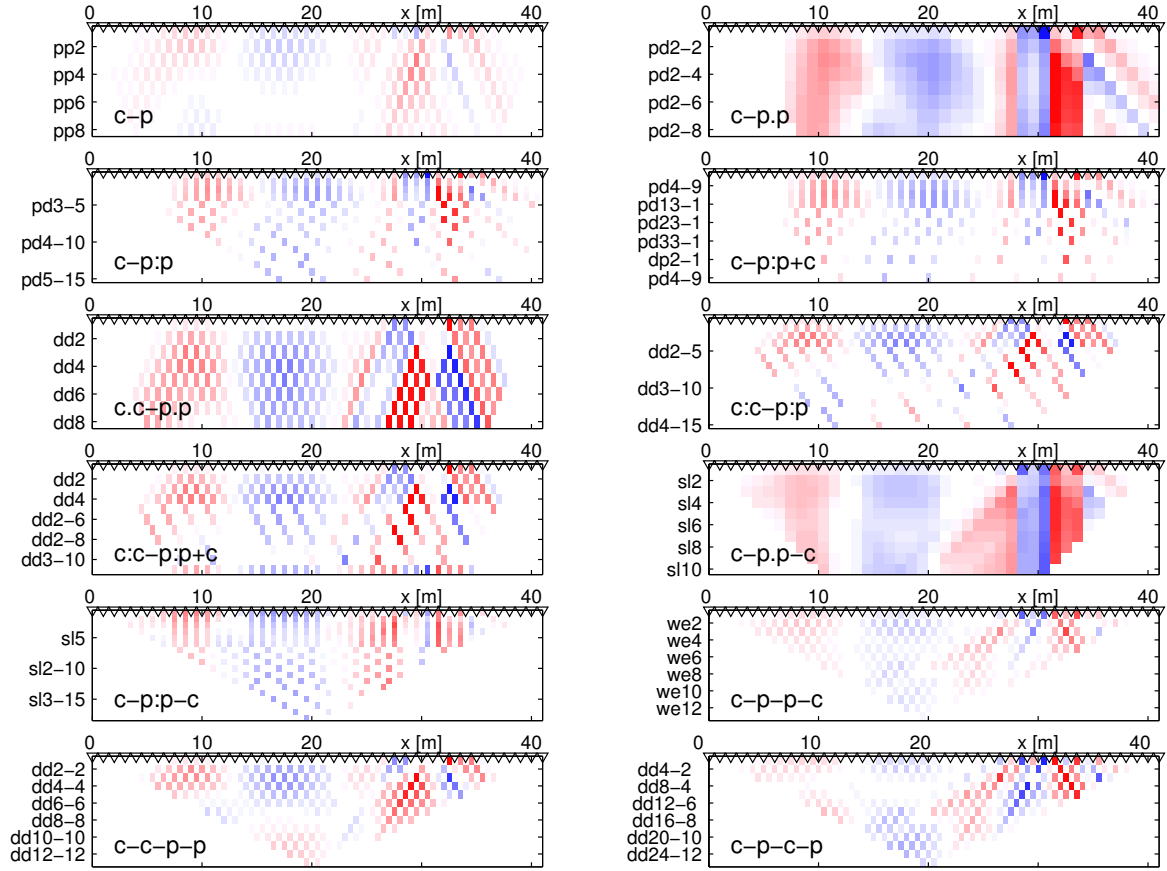


Figure 4.11: Synthetic data sets for the investigated data

Comparison of inversion schemes

For a fixed data set (c:c-p:p) different inversion or regularization schemes are compared with respect to information content. Besides the truncated SVD scheme and the classical Tikhonov regularization ($\mathbf{C} = \mathbf{I}$) a second order smoothness constraint matrix is investigated, which represents a discrete differential operator with Dirichlet boundary conditions. Additionally, a diagonal weighting matrix is applied using the square root of the coverage (or cumulative sensitivity), whose elements are the sums of the absolute sensitivity values over all data points with respect to the individual model cell:

$$\mathbf{C} = \text{diag}(\sqrt{\text{cov}}) \quad \text{with} \quad \text{cov}_j = \sum_{i=1}^N \|S_{ij}\| \quad .$$

Table 4.2 shows the results of the numerical simulation for low data noise. All schemes possess comparable information contents of around 70. To obtain a measure of the inversion quality, the RMS error between the result and the synthetic model is calculated, denoted with "deviation". Big IC values correspond to small deviations between synthetic and estimated model and vice versa.

Inversion scheme	regularization	IC	deviation
TSVD	$p = 69$	69.0	17.9%
Tikhonov	$\lambda = 39.3$	71.5	16.5%
weighting by coverage	$\lambda = 27.7$	70.0	16.0%
smoothness constraints	$\lambda = 58.1$	67.3	18.5%

Table 4.2: Comparison of different inversion methods, data set c:c-p:p, $\varepsilon=1\%$, $\delta U = 100 \mu\text{V}$, $I = 100 \text{ mA}$, $a = 1 \text{ m}$, deviation is RMS of estimated and synthetic model

The inversion results are displayed in Figure 4.12. TSVD and Tikhonov regularization produce noise artifacts resulting from the third term at the right hand side of the resolution equation (2.42). In contrast, smoothness constraints and weighting by coverage produce constraint artifacts near the boundaries.

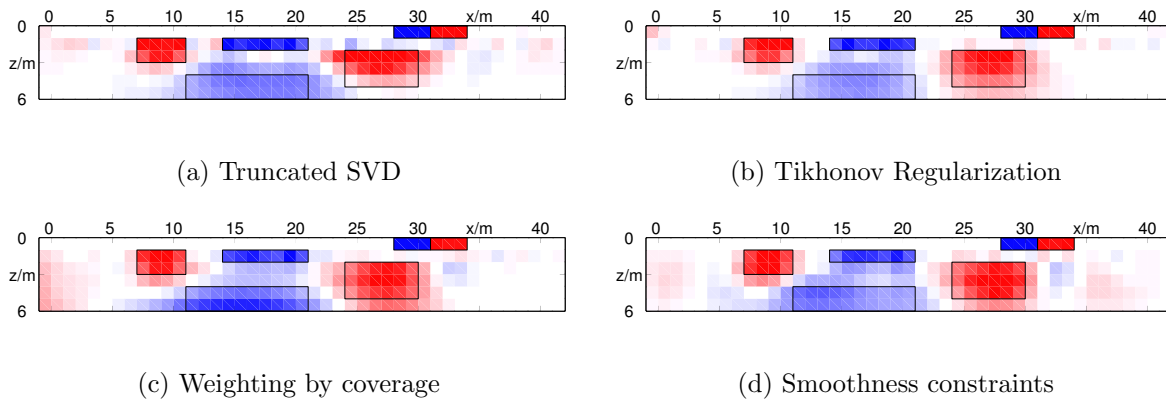


Figure 4.12: Inversion results for different inversion schemes using data set c:c-p:p

Comparison of data sets

In the following, the data denoted above are compared regarding information content and efficiency. For all simulations the second order smoothness constraints are applied. Figure 4.13 shows the individual inversion results. All data sets are able to locate the anomalous bodies, but with different quality.

In Table 4.3 it is to be seen, that the resolution properties correspond to the model quality. The largest IC values are achieved using pole-dipole and dipole-dipole type sets, whereas pole-pole and Wenner obtain the lowest values. In all three cases the dipole enlargement significantly increases the information content. Also, the added circulated data increase the IC for dipole-dipole type.

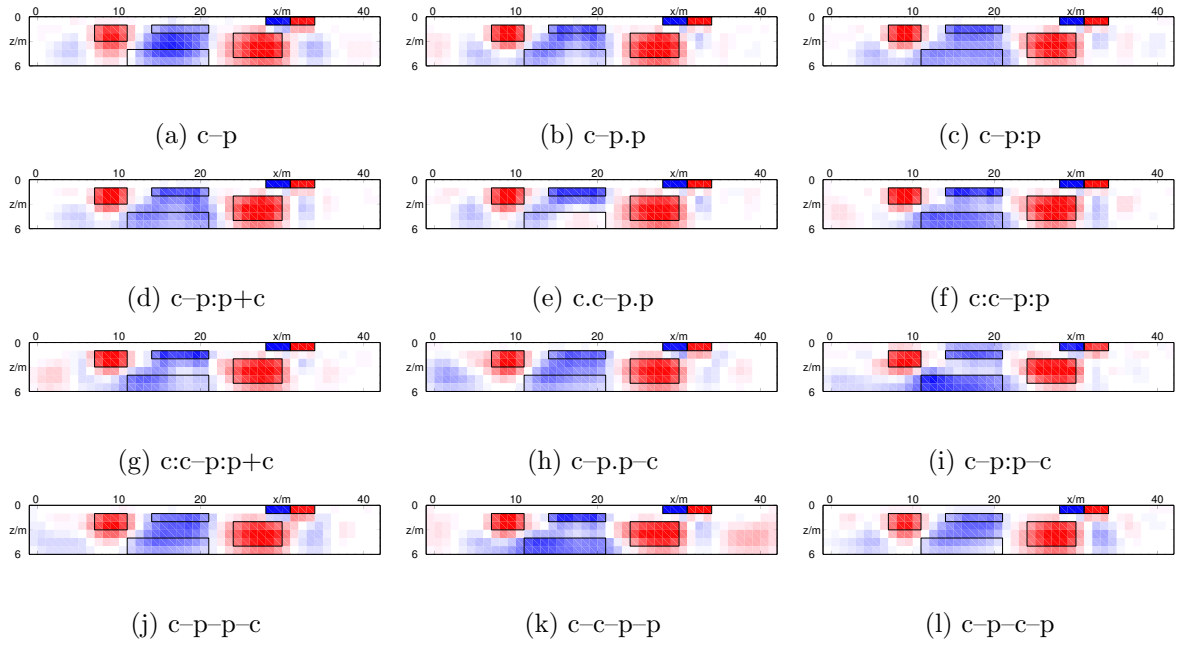


Figure 4.13: Inversion results for different data sets using $\varepsilon=1\%$, $\delta U = 100 \mu\text{V}$ and $I = 100 \text{ mA}$

data set	IC	IE	deviation	λ	maxerr
c-p	61.7	20.6	21.9	33.64	1.1
c-p.p	70.8	24.2	18.9	39.63	1.7
c-p:p	73.4	26.2	16.4	27.35	1.6
c-p:p+c	72.4	25.1	18.7	37.23	1.6
c.c-p.p	71.7	25.2	18.8	47.81	5.0
c:c-p:p	75.3	28.4	15.2	35.11	2.6
c:c-p:p+c	78.8	29.1	16.0	36.23	2.6
c-p.p-c	65.6	21.9	20.0	34.70	1.6
c-p:p-c	81.5	28.6	15.4	29.92	1.5
c-p-p-c	63.6	23.3	18.7	25.39	1.1
c-c-p-p	69.2	25.3	15.5	31.97	1.2
c-p-c-p	64.7	23.7	20.7	52.51	1.1

Table 4.3: Comparison of different data sets corresponding to Figure 4.13

Most information is obtained using the c-p:p-c set, whereas the c:c-p:p+c set yields the largest efficiency. The efficiency values, ranging between 20 and 30%, show a large variability. It can clearly be seen, that good model estimates denoted by small deviations correspond to large information contents and vice versa. Thus, the defined information content proves to be a reliable measure of inversion quality.

Resolution radii

The individual model resolutions for each cell i can be used to estimate a model resolution radius r_i . Assuming the model resolution to be locally constant, it is defined by the equivalent circle of model resolution 1, yielding for the cell dimensions Δx_i and Δz_i

$$r_i = \sqrt{\frac{\Delta x_i \Delta z_i}{\pi R_{ii}^M}} .$$

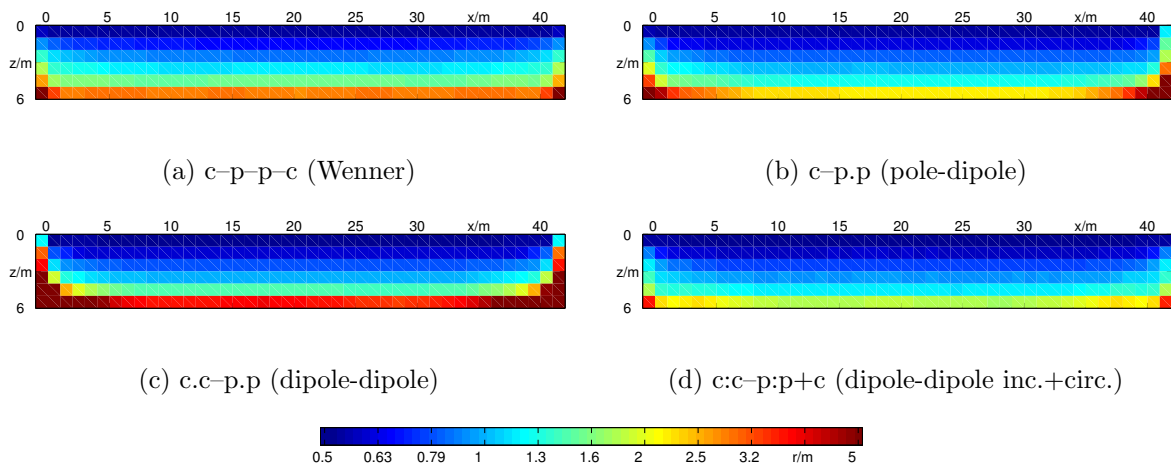


Figure 4.14: Resolution radii for data sets c-p-p-c, c-p.p, c.c-p.p and c:c-p:p+c

Figure 4.14 shows that for all selected configurations the resolution radius is about equal near the surface (0.5 m) and increases rapidly with depth up to 3–4 m. However, the decrease of resolution is different for the individual data sets. The Wenner array (a) has moderate resolution at depth while the dipole-dipole (c) array resolves best at medium depth. This is strongly improved by increasing dipole lengths and adding circulated data.

Effect of increased noise

Table 4.4 shows the resolution properties for an assumed noise of $\varepsilon=2\%$ and a voltage resolution of $\delta U = 1 \text{ mV}$ at $I = 100 \text{ mA}$ current, or, equivalently $\delta U = 100 \mu\text{V}$ at

$I = 10$ mA driving current.

data set	IC	IE	deviation	λ	maxerr
c-p	49.3	16.4	23.0	39.63	2.5
c-p.p	56.9	19.5	22.7	43.70	8.6
c-p:p	58.6	20.9	20.9	35.25	7.7
c-p:p+c	60.1	20.8	22.2	34.03	8.0
c.c-p.p	59.0	20.8	22.8	35.09	42.0
c:c-p:p	62.8	23.7	20.4	20.56	17.9
c:c-p:p+c	61.3	22.6	21.8	39.79	18.4
c-p.p-c	54.0	18.0	22.8	32.35	7.7
c-p:p-c	64.9	22.8	18.3	22.32	7.1
c-p-p-c	50.2	18.4	23.7	41.37	2.9
c-c-p-p	60.7	22.3	18.2	21.89	4.5
c-p-c-p	56.5	20.7	22.3	41.86	3.4

Table 4.4: Comparison of different data sets for $\varepsilon=2\%$, $\delta U = 1$ mV at $I = 100$ mA

As expected, higher noise leads to worse inversion results and lower information content for all data sets. Again, the increased dipole-dipole and Schlumberger configurations yield the best results. As the noise level rises more, pole-dipole arrays and the Wenner- β configuration perform better.

Effect of the parameterization

The question arises regarding how a variable model parameterization can affect the model resolution. Due to the increasing resolution radii, the horizontal grid lines are now defined by the values

$$z_i = [0 \quad 0.5 \quad 1.2 \quad 2.1 \quad 3.2 \quad 4.5 \quad 6] \quad .$$

There exist boundary cells showing bad resolution properties. If these cells bear no important structures, they can be excluded from the inversion process resulting in a better-posed inverse problem. Another possibility of lowering the degrees of freedom is a combination of badly resolved cells to form larger cells. Generally, the increasing of layer thicknesses results in cells with $\Delta x > \Delta z$ at the surface and $\Delta x < \Delta z$ at depth, which does not conform to the shape of the model cell resolution. In the following, deep cells are combined in such a way that the arising cells show approximately equally-spaced cells ($\Delta x \approx \Delta z$).

Both methods can be applied independently using a parameter mapping matrix \mathbf{P} as depicted in section 2.3. This has the advantage of using different mapping matrices without changing the sensitivity matrix. Table 4.5 shows the inversion parameters of the three possible parameter reduction methods compared to the unchanged problem.

Parameterization	parameter	λ	IC	deviation
Unchanged	258	44.0	92.0	13.4%
Combine	236	48.9	90.0	12.8%
Delete	224	43.5	90.3	13.4%
Combine&Delete	211	44.7	89.5	12.4%

Table 4.5: Comparison of different parameterization techniques for data set c:c-p:p

All methods obtain similar regularization parameters. The information content values are slightly, but significantly reduced by lowering the degrees of freedom. However, the quality of the inversion result can be improved, particularly for cell combinations. This can also be observed in Figure 4.15 showing the inversion results for the unchanged problem and the Combine&Delete method. The artifacts near the boundaries are reduced and the anomalous bodies seem to be more compact.

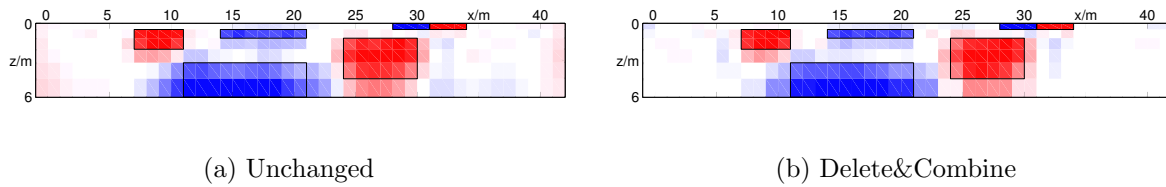


Figure 4.15: Iteration results of unchanged and reduced parameters

4.1.4 Consequences of Electrode Errors

In section 3.2 two methods of error-weighting have been presented: An uncorrelated weighting using the diagonal matrix $\mathbf{D} = \text{diag}(1/\epsilon_i)$ and a data covariance matrix \mathbf{C}_d resulting from varying electrode positions. Note that the inverse of \mathbf{C}_d replaces the term $\mathbf{D}^T \mathbf{D}$ in the preceding equations.

To find out whether the use of the more complicated covariance matrix is necessary, a synthetic study is carried out. The same synthetic model and electrode layout is used, the contrast factor is set to 5. As it is most sensitive to electrode variations, the classical dipole-dipole configuration was selected. The electrodes are considered to

show variations of 2 cm. Additionally, a relative deviation of 3% and a voltage error of $100 \mu\text{V}$ are superposed. This results in errors up to 8% for the first separation.

The chi-squared data misfit is generally computed by

$$\chi^2 = [\mathbf{d} - \mathbf{f}(\mathbf{m})]^T \mathbf{C}_d^{-1} [\mathbf{d} - \mathbf{f}(\mathbf{m})] / N \quad .$$

In the correlated case \mathbf{C}_d represents the full covariance matrix, while in the uncorrelated case off-diagonal elements are neglected. A global minimization scheme with smoothness constraints of 1st order and a regularization strength of $\lambda = 3$ is applied to both variants. The results are shown in Figure 4.16.

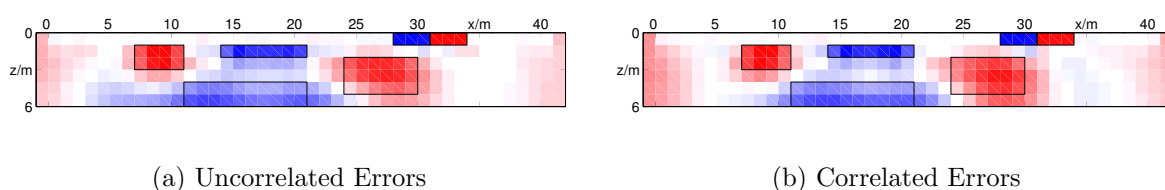


Figure 4.16: Inversion results for uncorrelated and correlated errors

The models show only slight differences. Hence, the off-diagonal elements are not necessary under practical circumstances. However, the knowledge of error values is essential to avoid over- or under-fitting. Particularly for small-scale tasks as many environmental problems represent, the error due to varying electrode positions must not be disregarded. Often small-scale variations are observed in the first separations of the pseudo-section, which are affected by electrode variations. An underestimation of these errors can therefore lead to over-interpreted structures in the first layers.

4.2 Three-dimensional Inversion Studies

Until now, three-dimensional measurements are carried out rarely. With developing multi-electrode devices and computer power a three-dimensional data acquisition becomes more and more interesting. Whereas some papers have been presented investigating optimizing profile data (Friedel, 2003; Stummer *et al.*, 2004), only few works deal with techniques for three-dimensional data acquisition (Loke and Barker, 1996a). In the following, two typical layouts of 3D measurements are discussed:

1. A grid of $E_x \times E_y$ electrodes is used allowing for arbitrary combinations. Generally, profile measurements on x- and y-profiles can be applied as well as on diagonal lines. A special case of grid measurements is the E-SCAN type measuring the pole-pole potentials for all electrode combinations (Li and Oldenburg, 1992).

2. In most cases data are obtained using several (not necessarily parallel) profiles. Each profile applies a two-dimensional investigation yielding a pseudo-section of the apparent resistivities.

Many multi-electrode systems available are engineered to carry out two-dimensional profile measurements quickly, e.g. by the use of multi-channel recorders. Furthermore, the experiences from the inversion of profile data have to be transferred to 3D inversion. Therefore, it is concentrated on a combination of profile data both on separated profiles and electrode grids.

4.2.1 The 11x11 Electrodes Grid

Most multi-electrode systems available have a limited number of electrodes, usually from 64 to 256. In the following, a typical grid of $11 \times 11 = 121$ electrodes is considered. The electrode distance in both directions is $a = 1$ m spanning an area of 10×10 m. The lower left electrode is placed at (0;0), the parameter domain is considered to reach from -1 to 11 m for x and y . Due to the limited electrode distances, the maximum penetration depth is estimated by sensitivity to $z_{max} = 4$ m. The layer thicknesses Δz increase from 0.4 to 1 m forming six layers.

Figure 4.17 presents the parameterization and the synthetic model used for the two denoted model types. Bodies of 20 and $500 \Omega \text{ m}$ are placed within a homogeneous half-space of $100 \Omega \text{ m}$. The Grid model (a) consists of (in x - and y -direction) equidistant model cells of 1×1 m grid size. In summary, the number of cells equals $12 \times 12 \times 6 = 864$. To the contrary, the Para model (b) constitutes of cells with varying size. Based on the calculated resolution radii for 2D sets, the grid sizes $\Delta x = \Delta y$ are increased from 0.5 m near the surface up to 3 m at the bottom. Thus, the total number of cells is $24 \times 24 + 12 \times 12 + 2 \cdot 8 \times 8 + 6 \times 6 + 4 \times 4 = 900$.

Pole-pole, pole-dipole and dipole-dipole data are simulated on the profile lines in x - and y -direction. The errors are considered to contain variations of 3% and a voltage error of $\delta U = 500 \mu\text{V}$ resulting in errors of 3.0 - 5.7% . A linearized inversion scheme as depicted in section 4.1 is applied to determine resolution measures and inversion results. Smoothness constraints were implemented for both model types. Since the first order smoothness matrix \mathbf{C} cannot be calculated explicitly for the para model type, second order smoothness constraints with Dirichlet boundaries were used. The vertical derivatives were neglected for reasons of comparability. An overview on inversion and resolution parameters is given in Table 4.6

The used regularization parameters range from 5 to 7 for the Grid model and from 20 to 30 for the Para model. Despite the smaller λ 's the information content and efficiency values of the Grid model are about 20 - 30% smaller than the ones of the Para

Data set	dir	data	Grid Model			Para Model		
			λ	IC	IE	λ	IC	IE
c-p	x	605	5.04	145.7	24.1%	21.4	172.1	28.4%
c-p	xy	1210	5.53	199.7	16.5%	29.3	267.2	22.1%
c-p.p	x	495	6.00	164.5	33.2%	23.5	193.0	39.0%
c-p.p	xy	990	6.83	231.4	23.4%	29.4	308.3	31.1%
c.c-p.p	x	396	6.26	153.0	38.6%	24.4	180.5	45.6%
c.c-p.p	xy	792	6.73	225.6	28.5%	25.5	299.1	37.8%
c:c-p:p+c	x	396	7.08	177.0	44.7%	25.4	213.4	53.9%
c:c-p:p+c	xy	792	7.22	246.0	31.1%	29.1	334.1	42.2%

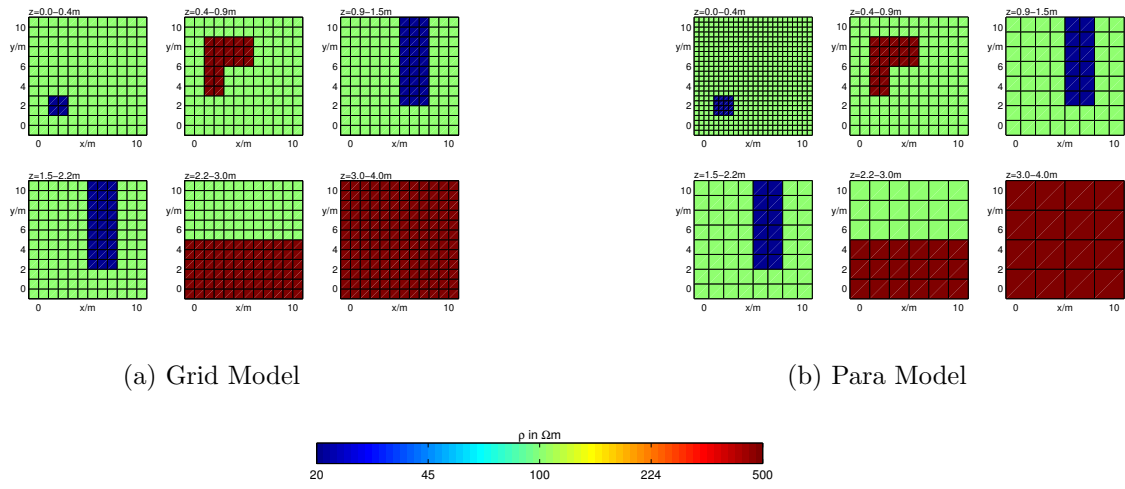
Table 4.6: Comparison of 3D data sets concerning information content (IC) and efficiency (IE)

model. Hence, the parameterization has effects on the resolution properties. The use of perpendicular profiles (xy) yields significantly improved IC values for all configurations and model types. However, due to the doubled data number the efficiency is reduced drastically.

A comparison of the individual configuration types yields the worst results for pole-pole. Differing from the 2D comparison the pole-dipole data are slightly superior to classical dipole-dipole data due to the relatively large voltage error. However, the improved dipole-dipole set with enlarged dipoles and circulated dipoles provides the best results of all.

The inversion results for the latter data set are displayed in Figure 4.18. Both model types are able to locate the synthetic bodies. However, in the deeper layers the resistivities are forced towards the starting model due to the boundary conditions in the constraints. This hold particularly for the Grid model. To the contrary, the Para model shows more artifacts in the shallow layers. The differing resolution properties are not to be seen comparing both model types. One reason for this is the relatively smooth synthetic model showing only small variations in the surface layer. The individual model resolutions are visualized in Figure 4.19.

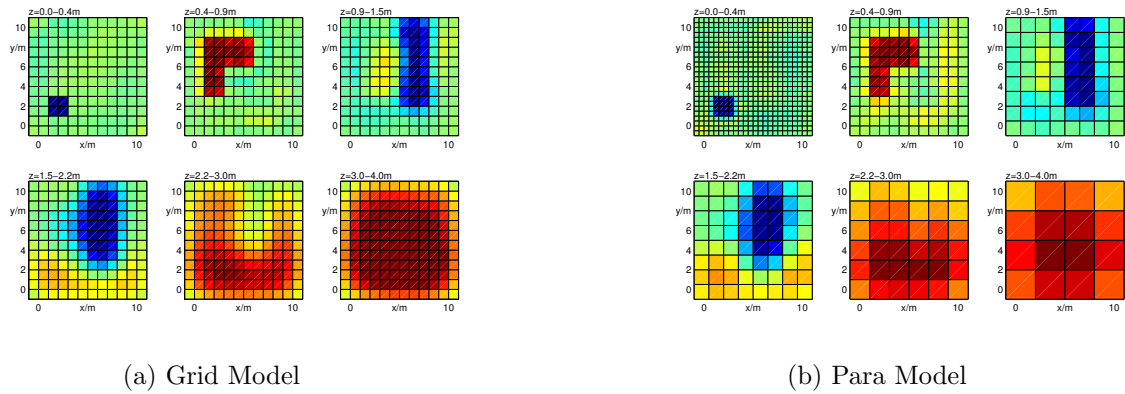
In the first layer the model resolution of the Para model is slightly smaller than for the Grid model. On the contrary, the deep Para model cells are resolved much better due to their increased size. This results in larger resistivity magnitudes. However, the blocky cells hardens the delineation of structures. In summary, the Para model type represents a more flexible and memory saving alternative respecting the resolution properties. This is particularly advantageous if the top layer shows fine variations.



(a) Grid Model

(b) Para Model

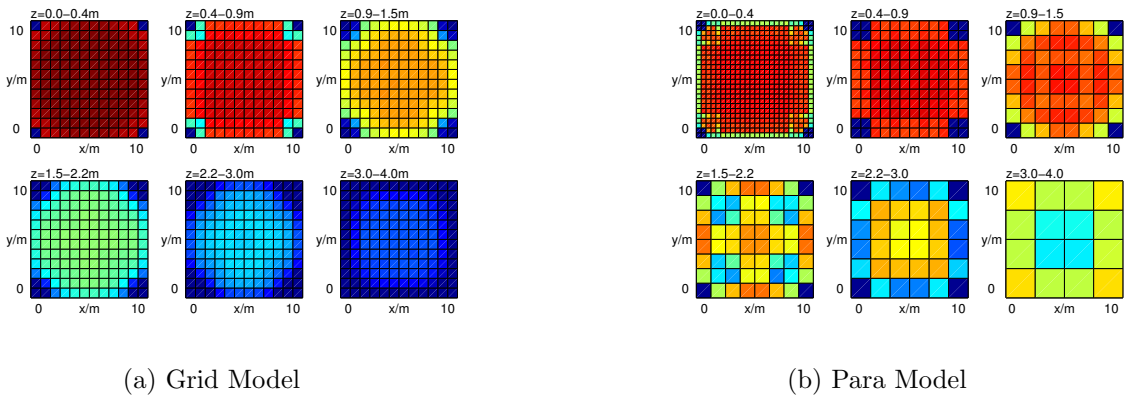
Figure 4.17: Model parameterization and synthetic model for Grid (a) and Para (b) type of model



(a) Grid Model

(b) Para Model

Figure 4.18: Inversion models of the c:c-p:p+c data set



(a) Grid Model

(b) Para Model

Figure 4.19: Model resolution for Grid (a) and Para (b) type of model

4.2.2 The Parallel Profiles

If large areas are investigated, an electrode grid cannot be applied due to limitations of the electrode number. In most cases profile arrays are measured allowing for independent 2D inversion. However, in the case of 3D structures the data have to be interpreted three-dimensionally. Often the lines are parallel, even though it is not necessary, because the singularity removal allows for grid independent electrode positioning.

The following study employs five parallel profiles. Each profile consists of 21 electrodes of electrode distance 1 m. The profile distance can be increased up to 4 m for the investigation of resolution properties. Figure 4.20 displays the synthetic model used.

From the 2D studies we know the data will be interpreted well if only one profile direction is chosen. Furthermore, it shows that the optimized 2D data sets obtain the best results in three dimensions, too. For that reason, an improved dipole-dipole configuration was chosen.

The data are considered to have variations of 1% plus a voltage error of $500 \mu\text{V}$. Smoothness constraints of first order were applied. At first, the linearized scheme is applied again to estimate resolution properties. Since the number of data and model parameter is too big for a GSVD decomposition, a single parameter resolution is determined using resolution approximation. It is displayed for profile distances of 1 m, 2 m and 4 m in Figure 4.21.

The relevant main diagonal elements are all below 10%, which is quite low. An anomaly in the model cell is distributed over the neighboring cells due to the smoothness constraints. Note that the shape of the cell resolution is nearly identical for 1 m and 2 m electrode distance, even though the anomaly is slightly reduced for the latter. An electrode distance of 4 m leads to a distorted resolution structure of diminished magnitude. The maximum values are obtained at cells underneath the nearest electrode lines due to their increased coverage.

It is therefore recommended not to exceed the factor 2 between profile distance and electrode distance for a detailed three-dimensional mapping of the ground. Otherwise the cells have to be defined non-equidistant. However, larger profile distances can be accepted, if the subsurface shows a nearly two-dimensional structure with a known strike direction as many geological problems represent.

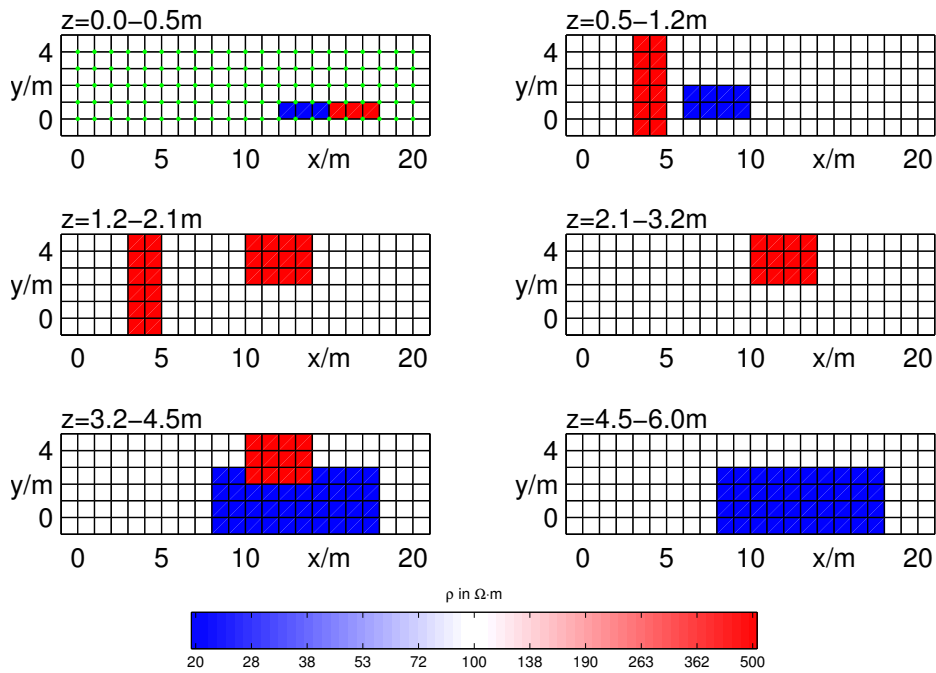


Figure 4.20: Synthetic model used, the green dots denote the electrodes of the five profiles

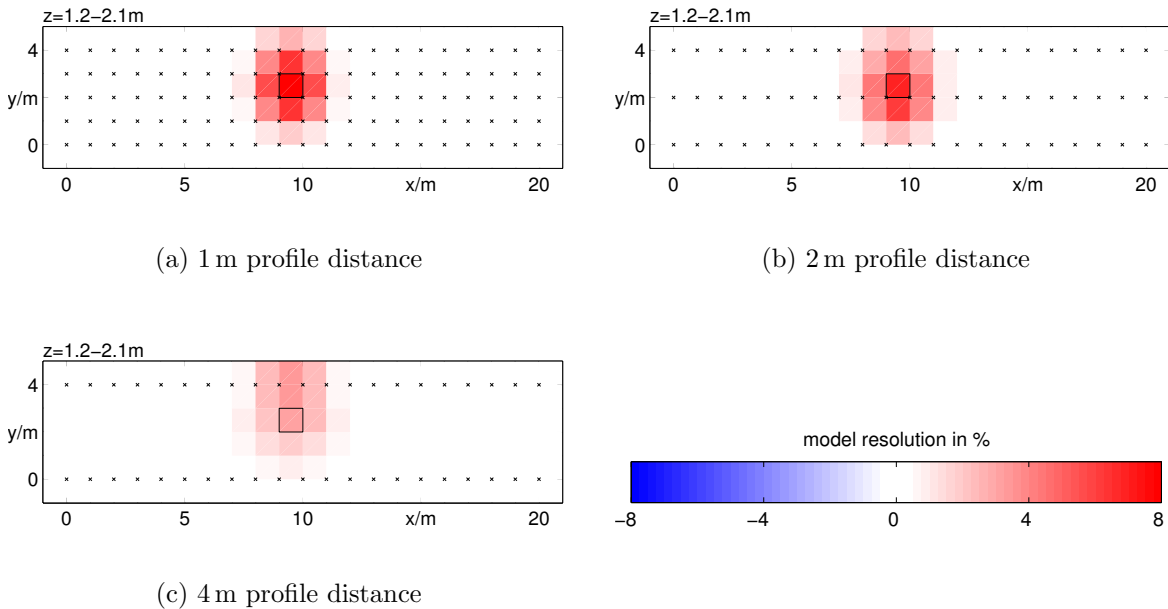


Figure 4.21: Second layer plan view of a single parameter resolution (in %) for varying profile distance, the according cell is indicated by black lines

4.3 3D–Inversion Including Topography

4.3.1 Motivation

In many cases an existing topography of the earth’s surface has a strong effect on the current distribution in the earth and therefore onto the measured apparent resistivities. The above depicted methods of parameterization, regularization and forward calculation are not directly applicable and have to be adapted.

Only few approaches have been presented that incorporate a three-dimensional topography into the inversion process Sasaki (1994); Sugimoto (1999, e.g.). Some of them are based on undulated FD-like grids, which are not very flexible and impractical for large-scale problems. In the following, we³ present a triple-grid inversion approach on the basis of unstructured tetrahedral grids motivated by two considerations.

First, a model parameterization has to be found that considers topography. This holds also for the construction of smoothness constraints based on neighboring relations between the individual model cells. The second problem is the application of the forward routine. Although there exist approaches to incorporate conductivity free cells into FD grids, they are generally not flexible and are time-consuming.

The finite element technique is independent of the geometrical form of the elements. The tetrahedron is the most flexible body for the description of arbitrary three-dimensional geometry. For the construction of tetrahedral grids respecting a given geometry a grid generator has to be used. There exist commercially available grid generators. In contrast, TetGen (Si, 2003) represents a free and flexible program for the generation of unstructured tetrahedral grids.

Even in the case of constant conductivity the potentials are not known analytically⁴. Therefore the technique of singularity removal cannot be applied as depicted above. Since the total potential shows a large curvature near the electrodes, the model has to be highly refined to obtain accurate potentials.

However, a fine discretization results in a huge increase of computer time and memory for every iteration step. This leads to the idea of a double-grid forward calculation: The primary potential is determined once using a homogeneous conductivity distribution on a very fine grid. On a coarse grid the secondary potential can be calculated for arbitrary conductivities as arising in the individual iteration steps.

To ensure accurate modeling results, appropriate boundaries have to be added. Furthermore, the grid lengths have to be smaller than a reasonable model parameterization suggests. Generally, the forward grid is created by global refinement of the elongated

³This section is based on a collaboration with Carsten Rucker from the University of Leipzig.

⁴Also, configuration factors transferring the impedances into apparent resistivities are not known.

parameter grid completing the triple-grid technique.

4.3.2 The Synthetic Model

A popular example for topography is the investigation of flood protection dikes, which came into the focus of public interest since the flood disaster in Eastern Germany in 2002. In critical areas it is important to obtain information about the internal structure of the dike. This can be accomplished using DC resistivity measurements.

Figure 4.22 displays the synthetic model used in the following. The topography originates from a real dike at the river Elbe near Torgau.

The model consists of four units: The dike body of resistivity $100 \Omega \text{ m}$ is placed over a basement of $300 \Omega \text{ m}$ representing unconsolidated sediments. Inside the dike an old dike of high resistivity ($500 \Omega \text{ m}$) exists. Finally, a conductive ($50 \Omega \text{ m}$) covering layer of 0.5 m thickness is assumed on top of the dike body.

Several 2D profiles are assumed, two at the edges of the dike's top ($x=50.6\text{-}52.6\text{m}$), three at the water side ($x=56.3\text{-}63.3\text{m}$) and two at the land side ($x=45\text{-}47.8\text{m}$). Each profile has 31 electrodes with an electrode distance of 1 m, and the dipole-dipole configuration is simulated. To deal with the large errors arising for big dipole separations, the dipole lengths were successively enlarged. Such an optimization increases the efficiency of the data and makes the three-dimensional investigation of the ground interesting from economical point of view.

In summary, 1526 data points were simulated using a grid, which is completely independent of the inversion grids. The data are displayed in Figure 4.23 in the form of pseudo-sections.

4.3.3 The Three Grids of Inversion

In the following, the three inversion grids (Figs. 4.24, 4.25 and 4.26) are described in detail. The figures were created using the mesh viewer MEDIT (INRIA-Rocquencourt, 2003).

The Parameter Grid defines the model parameters (Figure 4.24). It respects the topography of the modeling domain. While near the surface the edge lengths are smaller than the electrode distance, the size of the tetrahedrons increases successively with increasing depth. The largest cells are found at the lower model boundary. For this purpose, the resolution analysis helps to estimate reasonable resolution radii, which are used as input for the grid generation. In doing so, the number of degrees of freedom is kept small while resolution is ensured.

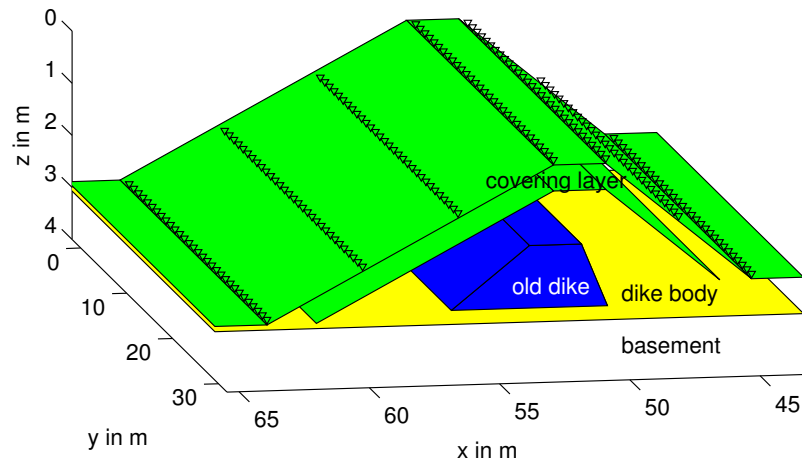


Figure 4.22: Synthetic model of a flood protection dike

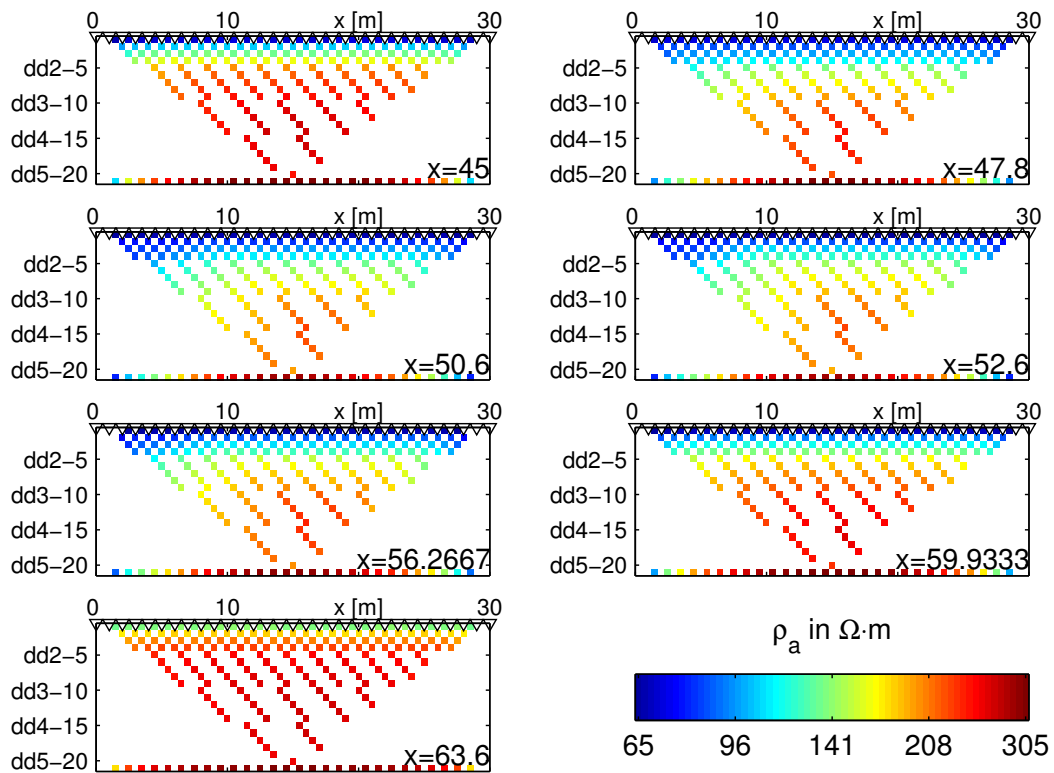


Figure 4.23: Pseudo-sections of the synthetic data used for inversion representing dipole-dipole data with increased dipole lengths. The profile location is indicated in the bottom right-hand corner of each section.

The Secondary Field Grid is used for the forward calculation with the singularity removal technique. It is created by a global refinement of the Parameter Grid (Figure 4.25) such that the edges of every element are halved resulting in 8 small tetrahedrons. This procedure may be repeated for large resistivity contrasts resulting in better accuracy. The prerequisite for the singularity removal are the primary potentials at the grid nodes.

The Primary Field Grid is a fine grid, which is used to calculate the primary fields. For this purpose, a homogeneous conductivity of $\sigma_0 = 1 \text{ S/m}$ is applied. The advantage is to avoid modeling errors due to conductivity contrasts. For the singularity removal calculation the potential can be scaled down by the local electrode conductivity. The primary potentials are also used to calculate configuration factors in such a way that $\rho_a = 1/\sigma_0 = 1 \text{ } \Omega \text{ m}$. To yield accurate results, the primary grid is highly refined near the electrodes (Figure 4.26).

Table 4.7 shows the number of elements and nodes for the three grids. 17727 model parameters are surrounded by 3794 nodes. The node and element numbers of the secondary field grid are much bigger due to the prolongation of the boundaries. Although the primary field grid is refined only locally, it contains much more nodes to ensure accurate primary potentials, which are essential for a successful inversion.

	Parameter grid	Secondary field grid	Primary field grid
Nodes	3794	64881	497243
Elements	17727	351616	2748100

Table 4.7: Node and element numbers of the three grids

4.3.4 Inversion Method

A global regularization scheme is applied as described in chapter 2. Since the problem is highly under-determined, smoothness constraints seem to be the method of choice. Due to the unstructured grid, classical discrete Laplacian operators cannot be applied. Therefore, a special roughness operator has to be implemented. For every tetrahedron i the individual neighboring cells j are determined and, independent of the relative direction, values of 1 and -1 are added to the constraint matrix elements $C_{i,i}$ and $C_{i,j}$, respectively. Note that for an equidistant FD grid this method yields the identical matrix \mathbf{C} as the first order smoothness measure described in section 2.2.

Since the potentials are not known analytically, the sensitivities for the homogeneous half-space cannot be determined via Gauss-Legendre integration. The use of finite

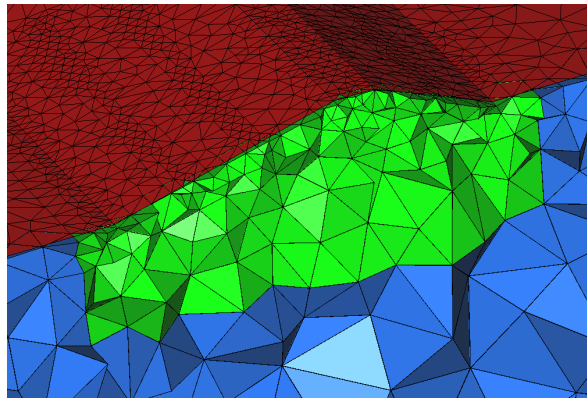


Figure 4.24: Section of the parameter grid, red is the earth's surface, the green tetrahedrons are the model parameter and the boundary elements are blue.

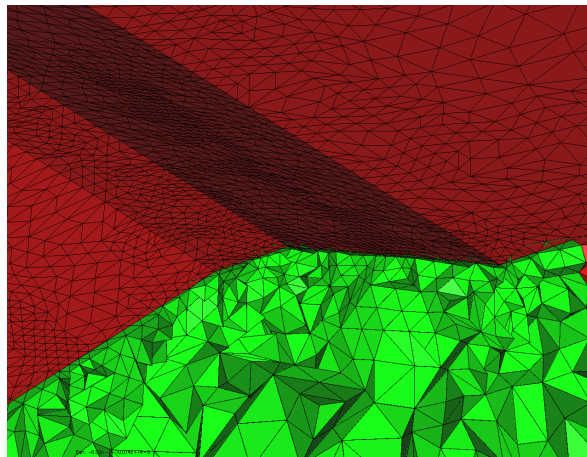


Figure 4.25: The secondary field grid, derived by global refinement.

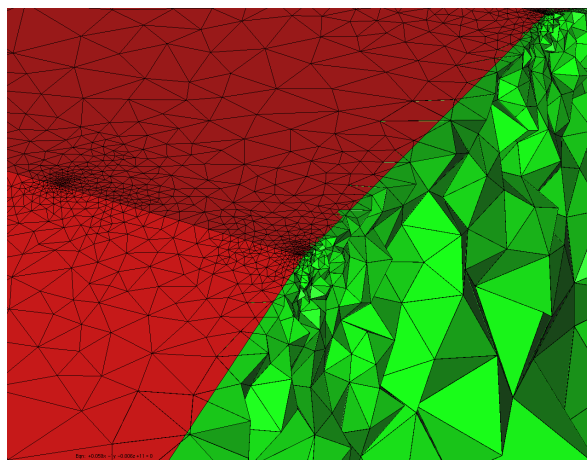


Figure 4.26: Section of the primary field grid. The local refinement at the electrodes makes their position clearly visible.

elements suggests a sensitivity calculation based on the reciprocity principle, which makes use of the FE stiffness matrix (Kemna *et al.*, 2000). This method can be used in every iteration step. However, the sensitivity values are expected to be only as accurate as the (total) potential approximation on the secondary field grid.

In the first iteration an L-curve scheme is applied to determine the λ that provides optimum regularization. Then, the regularization parameter stays constant. The entire inversion scheme reads as follows:

1. Reading of data files and topography
2. Grid generation with TetGen and by global refinement
3. Calculation of primary potentials and interpolation to secondary field grid
4. Determination of configuration factors and estimation of data errors
5. Assembling of data and model constraint matrices
6. Choice of an appropriate starting model
7. Sensitivity calculation
8. Parallel solution of the inverse subproblem for many λ
9. Choice of one solution using L-curve criterion
10. Finite Element forward calculation using singularity removal
11. Solution of single- λ inverse subproblems and forward calculation until convergence is reached

The starting model is a homogeneous resistivity distribution. Since the resolution at depth is supposed to be bad, the basement value of $300 \Omega \text{ m}$ is used. Smoothness constraints of second order were applied in form of a discrete differential operator. The boundary conditions used were of Dirichlet type at the subsurface boundaries and of Neumann type at the earth's surface.

4.3.5 Inversion Results

Table 4.8 shows the inversion parameters in the course of iterations. Both the chi-squared misfit and the RMS error decrease rapidly within 4 iteration steps down to the target value of $\chi^2 = 1$. Whereas the maximum resistivity increases during the iteration, the minimum value is forced back by the global smoothness constraints after the first

Iter.	line search τ	χ^2 misfit	RMS (%)	reciprocity (%)	min/max(ρ/Ω m)
0	-	129.9	62.3	-	300/300
1	0.93	28.9	10.8	3.7	19/418
2	0.69	2.7	3.6	1.2	41/468
3	1	1.18	2.4	1.1	42/462
4	1	0.89	2.1	1.0	42/470

Table 4.8: Inversion parameters in the course of inversion

inversion step. This affects also the reciprocal accuracy of the forward calculation, whose standard deviation is significantly reduced in the second iteration. Note that the line search parameter τ indicates the inverse problem being nearly linear at the first and last iteration, whereas the second iteration requires a significant damping of the model update.

In Figure 4.27 the final inversion result is displayed. Besides two slices all tetrahedrons with resistivities of more than 260 or less than 60 Ω m are patched into the image. It can be seen, that the boundary between the dike and the basement is determined well. Also, the position of the resistive old dike can be clearly indicated by increased resistivities.

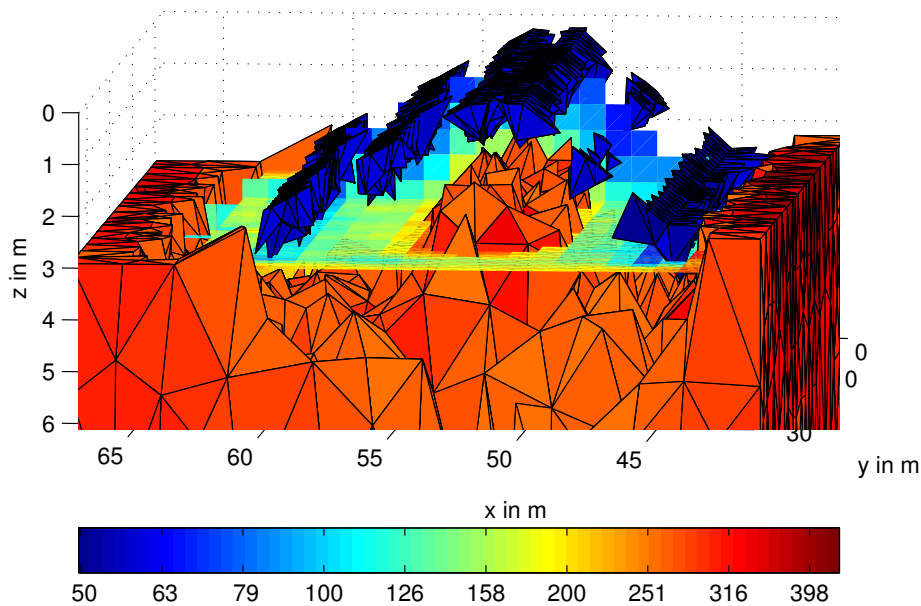


Figure 4.27: Inversion result of the synthetic data. All tetrahedrons with resistivities above 260 and below 60 Ω m are displayed.

The conductive covering layer below the surface is clearly indicated by low resistivities. However, the regions of increased conductivity are concentrated around the profile lines. Obviously the model is equivalent with respect to the measurements. Probably the electrode coverage perpendicular to the profiles is not fine enough, which can be improved by smaller profile distances or perpendicular profiles.

To summarize, a three-dimensional DC inversion for arbitrary topography is possible. The above described approaches were applied successfully using a triple-grid inversion technique. However, the generation of appropriate grids represents a task that cannot be run automatically at the current stage. Furthermore, compared to flat-earth 3D-inversion a lot of computing time and memory is needed to get accurate results.

4.4 Conclusions

It had been shown, that the Gauss-Newton method can be applied to the inversion of synthetic data. Its convergence is fast compared to the non-linear conjugate gradient technique. It is necessary to control the strength of the model regularization. This can be achieved by means of the L-curve. However, the essential method for the determination of the regularization parameter is the discrepancy principle.

Furthermore, a line search improves the convergence of the minimization procedure significantly. It can be shown, that one additional forward calculation is sufficient for determining appropriate line search steps. In summary, an effective combination of line search and constraint balancing yields accurate results within a small number of iterations. The use of a constant, global regularization respecting the discrepancy principle is recommended. For this scheme the regularization properties can easily be obtained, which are useful for the estimation of uncertainties.

The introduced resolution measures strongly depend on the regularization strength and the line search parameter. To circumvent these problems, a linearized scheme using the discrepancy principle was applied. On this basis several regularization techniques were compared. All provide a similar quality but emphasize variations at the surface or at the outer boundaries. Additionally, different data sets were investigated concerning information content and efficiency. Compared to classical configurations, data sets with increased dipole lengths yield superior results. Amongst the investigated data sets, dipole-dipole and Schlumberger configurations with increased dipole lengths obtained the largest resolution qualities for both 2D and 3D data acquisition. Noise and resistivity contrasts are important factors for the quality of inversion.

It can be seen, that the presented quantity known as information content is an appropriate measure of the reliability of a model, which can be applied to further experimental

design. However, it has to be proved if this applies to the full non-linear case as well. The crucial point in non-linear inversion is the choice of the regularization parameter, which strongly influences resolution properties and has to be chosen carefully. The presented inversion and experimental optimization techniques have to be validated in practice.

5 Application to field data

5.1 The Rothschönberg Gallery

In the following section data from the Rothschönberg test field close to Freiberg are investigated. Since the subsurface has a simple and well-known structure, it suits well for testing purposes. The investigation area represents fairly conductive background with lowest resistivities ($100 \Omega \text{ m}$) near the surface and resistivities of $300 \Omega \text{ m}$ below. At a depth of 5–8 m the resistive ($> 1000 \Omega \text{ m}$) Gneiss basement is found.

The Rothschönberg Gallery crosses the test field at a depth of 2–4 m. It has a lateral extent of about 2.5–3 m. In the past it has been used for the drainage of the Freiberg mining district. Figure 5.1 shows a section of the investigated gallery.

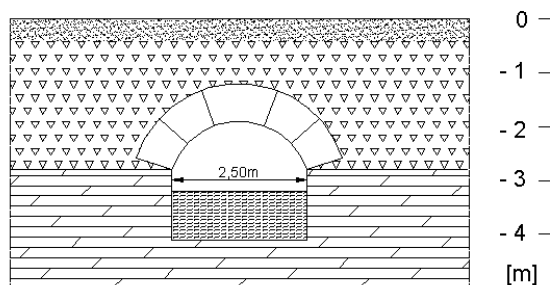


Figure 5.1: Section of the investigated Rothschönberg gallery

Next to the gallery another anomaly of undefined origin is known. It may be a side cave as well as a basement bulge.

5.1.1 2D–Inversion and the Global L–curve

First, a profile array of 21 electrodes with an electrode distance of 2 m was applied. Since the course of the gallery is known, the profile was arranged perpendicular to its strike direction. Thus the data can be interpreted two-dimensionally. Figure 5.2 displays the measured data set. It constitutes of classical dipole-dipole configurations using separation factors of $n=1-8$.

The 116 data show variations from 80 to $370 \Omega \text{ m}$. The data noise is estimated at a rate of 1% plus a voltage resolution of $100 \mu\text{V}$ at $I = 100 \text{ mA}$, which results in errors up to

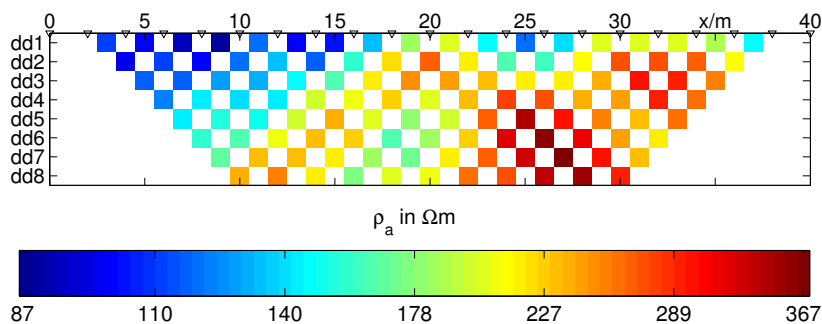


Figure 5.2: Field data of the original Rothschnberg profile representing a dipole-dipole pseudo-section of 21 electrodes

8% for the largest separations. The model is parameterized by grid lines in x-direction from -4 to 44 m with 2 m spacing. The thicknesses of the 8 layers increase from 0.5 to 1.7 m considering the decreasing resolution with depth. The maximum depth of 6.2 m was determined by sensitivity analysis. In summary, the model consists of 192 cells with varying size.

In the following, a global regularization scheme is applied using first order smoothness constraints. Starting model is the homogeneous half-space of $200 \Omega\text{m}$, which is the median of the data. Logarithmic resistivities without lower bound are used for data and model parameters. The regularization parameter is considered to be constant in the course of iterations. Independent solutions are obtained for a series of λ_i starting from 1000 with a factor of 0.8 down to 0.4.

Whereas for large λ the inversion converged, the data misfit increased again for $\lambda < 10$. Therefore, an inexact line search procedure was implemented to ensure convergence. Note that the line search is not based on χ^2 , but the total misfit functional Φ to be minimized. The inversion was stopped, when the relative change of Φ reached 1%. As a result, the models show reducing data misfits and increasing model roughnesses for decreasing regularization parameters. Figure 5.3 displays data misfit and model roughness for all solutions, yielding a global¹ L-curve.

Respecting the discrepancy principle the point has to be sought, where the L-curve crosses the line of $\chi^2 = 1$. This procedure yields a value of $\lambda_{19} = 18.0$. To the contrary, the L-curve criterion using the curvature of the parametric curve obtains $\lambda_9 = 54.9$ for optimum regularization.

The inversion results are displayed for six selected, nearly logarithmically equidistant, regularization strengths (1000, 167.8, 54.9, 18.0, 3.02 and 0.51) in Figure 5.4. All models show the main structures of two resistive bodies in a conductive background.

¹As the (local) L-curve was introduced for linearized inverse problems, the presented curve is depicted as "global L-curve"

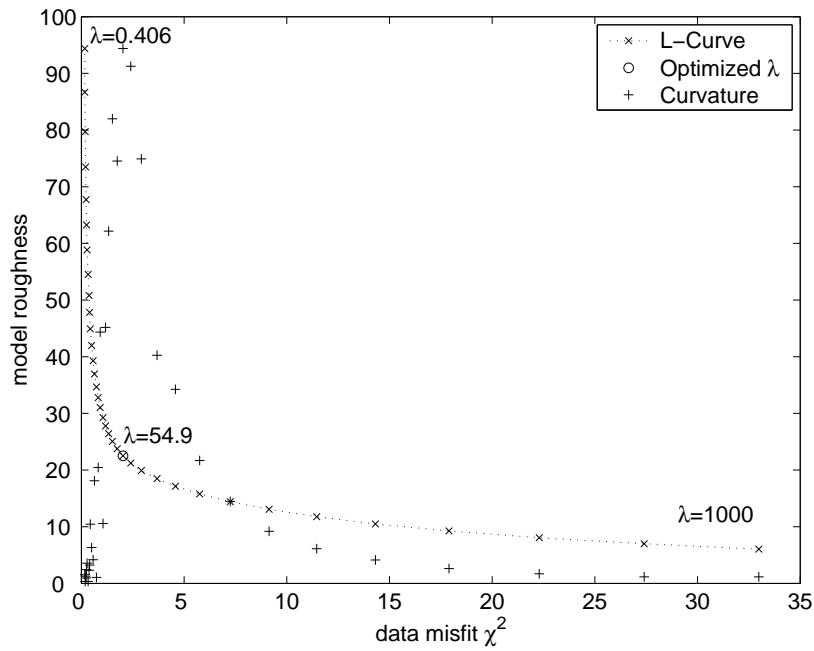


Figure 5.3: Global L-curve displaying data misfit χ^2 and model roughness for varying λ , the curvature of the curve has a clear maximum defining the optimum λ

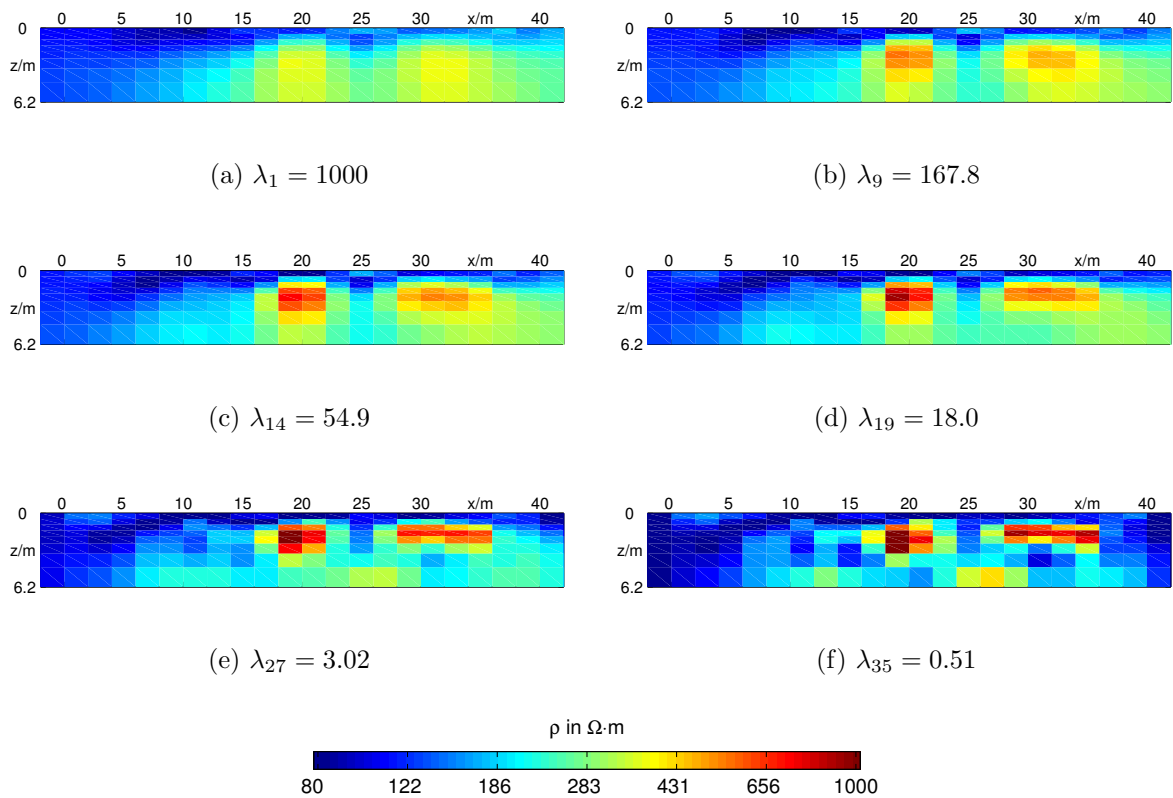


Figure 5.4: Inversion results for six selected values of λ

From (a) to (f) more and more complexity appears in the model due to the decreasing influence of the smoothness constraints. The models (a) and (b) are too smooth to delineate the boundaries of the anomalous structures well. Since the data are not explained accurately, they are under-fitted. In contrast, in the models (e) and (f) artifacts resulting from errors are visible, which can be interpreted as over-fitting.

The sub-figures (c) and (d) show reasonable structures without artifacts. Model (c) corresponds to the optimum regularization determined by the curvature of the global L-curve, whereas model (d) represents the solution associated with the discrepancy principle. As the global L-curve is generally not known, the local L-curve of the linearized subproblem in the first iteration can be used for estimating an optimized λ . In doing so, we yield $\lambda_{18} = 22.5$ for the optimum value, which is close to that of the discrepancy principle. Although the discrepancy principle is finally the less sophisticated choice, the local L-curve method produces reasonable estimates as well.

Figure 5.5 displays the misfit function and its histogram for λ_{19} . It represents the difference between the logarithmized data and forward response, weighted by the estimated errors. It can be seen, that neither single data outliers nor systematic structures are present. The reasons for a systematic misfit function may be a misappropriate model discretization as well as present 3D effects, which cannot be interpreted by 2D models. The histogram shows almost a Gaussian distribution, which is essential for successful least squares inversion.

Note that the form of the L-curve and hence the resulting optimum λ strongly depends on the error estimates or measures. Knowledge of the errors or a reasonable estimate is an essential prerequisite for the interpretation of geophysical data.

5.1.2 Optimization of Profile Data

The optimization techniques described in the last chapter are to be proved in practice. Therefore, different conventional and non-conventional data sets were measured on an electrode array in less than 1 m distance, parallel to the original profile. To increase the accuracy, the electrode spacing was reduced to 1 m. 36 electrodes were placed equidistantly and the measurements were applied subsequently without changing the electrode positions.

A total of 2488 single measurements includes all conventional configuration types with small and increased dipoles. From that, four data sets were selected and depicted in Table 5.1. In addition to the classical Wenner and dipole-dipole arrays, improved sets of Schlumberger and (circulated) dipole-dipole type with increasing dipole lengths were chosen to verify the good results of the latter in the linearized study (section 4.1). To keep the number of data relatively small, the maximum separation was limited.

Data set	tag	data	n	min/max ρ_a	min/max err
Wenner	c-p-p-c	196	1-11	112.4/450.6	1.2/3.0
Schlumberger(inc)	c-p:p-c	186	1-15	112.6/464.4	1.5/3.0
Dipole-dipole	c.c-p.p	285	1-10	81.1/530.3	3.3/6.7
Dipole-dipole(inc)+circ	c:c-p:p+c	259	1-15	81.1/585.9	2.0/4.3

Table 5.1: The four data sets used for comparison

The error constitutes a rate of 1%, a voltage error of 0.1 mV and a positioning error of 0.01 m without covariances. Note that both dipole-dipole configurations show relatively large error values compared to the Wenner/Schlumberger arrays. However, increasing the dipole lengths reduces both voltage and positioning error by 30%. The largest anomalies are obtained by the dipole-dipole type arrays.

Due to the large number of electrodes and separations the model was discretized by 9 layers of increasing thickness up to a maximum depth of 7 m. The horizontal grid spacing was set to $\Delta x = 1$ m. Again, the inversion scheme applied global smoothness constraints of first order. Unlike in the last example, the derivatives with respect to z were weighted by the factor of $\alpha_z = 0.3$ to enhance the results slightly. The regularization parameters were chosen to satisfy convergence at $\chi^2 \approx 1$. The concrete values were comparable and ranged from 10 to 30.

In Figure 5.6 the inversion results are plotted using filled contour lines. All four data sets are able to locate the two known resistive anomalies at $x = 16$ and 26 m, but at different quality. Whereas the Wenner/Schlumberger arrays are not able to delineate the boundaries of the gallery, the dipole-dipole measurements are able to do so. However, the inversion result of the classical dipole-dipole shows bad resolution at depth, which is also the case of the Schlumberger model. In contrast, the other two models are able to map out the course of the resistive basement. The result of the improved dipole-dipole set shows the most detailed image with the largest parameter contrasts.

To compare the resolution properties, in Figure 5.7 the model cell resolutions of the cell 174 are plotted, which is the center of the gallery's position. For all data sets the cell resolutions display an approximate Gaussian distribution, which is related to the concept of averaging resolution kernels (Backus and Gilbert, 1968). Clearly the resolution kernels for the individual data sets correspond to the particular inversion results. The very good result of the improved dipole-dipole set is indicated by a relatively sharp resolution. Also, the classical dipole-dipole and the Wenner data set show reasonable resolution kernels. To the contrary, the cell resolution of the Schlumberger data set indicates uncertainties in the vertical direction.

To sum it up, the improved dipole-dipole data set shows the most detailed inversion

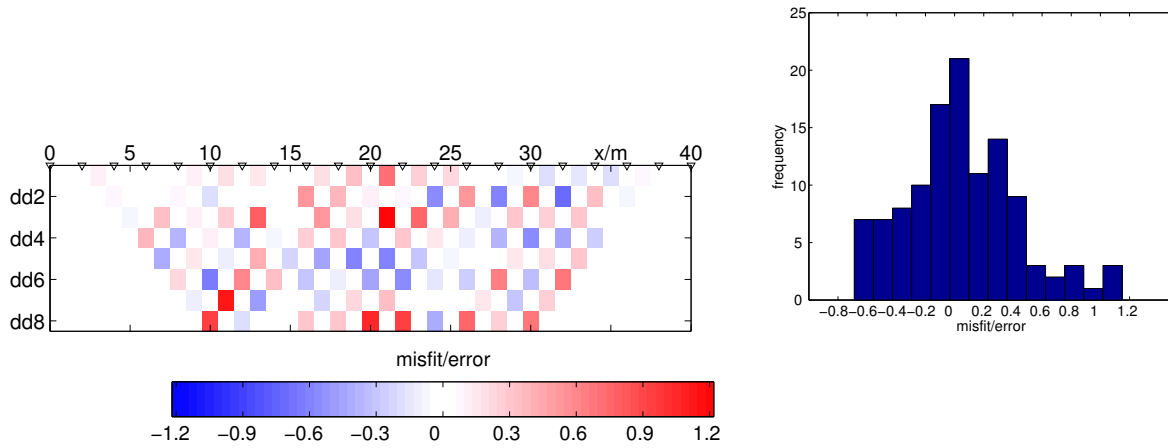


Figure 5.5: Error-weighted misfit as pseudo-section and histogram for model (d) using $\lambda_{19} = 18.0$

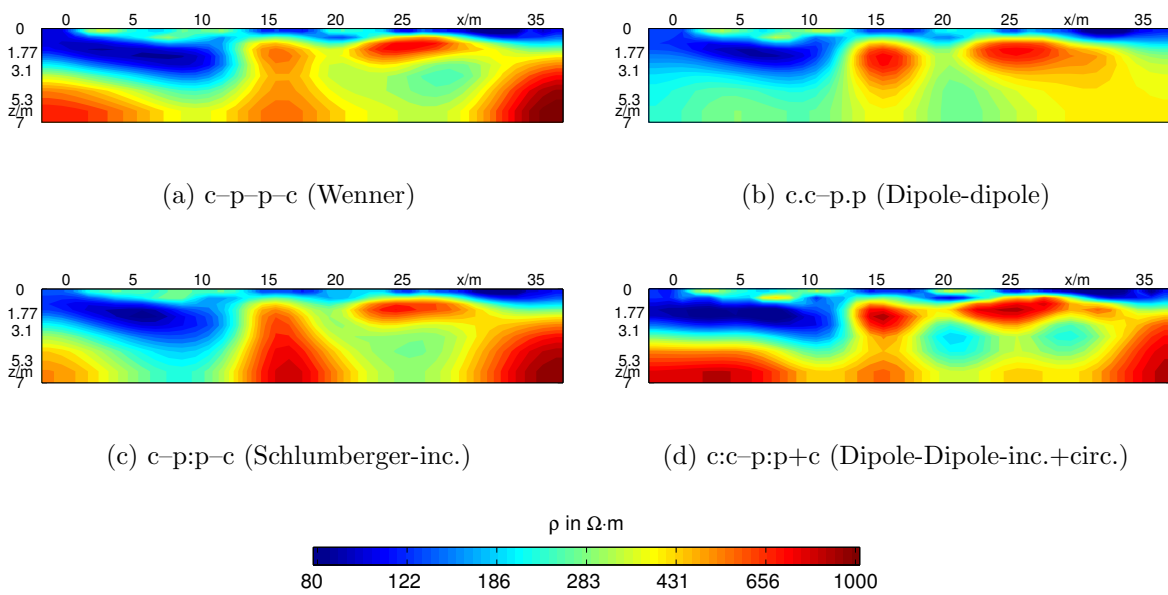


Figure 5.6: Inversion results of the depicted data sets

result in all model regions. The improvement represents a significant enhancement of information content compared to classical dipole-dipole sets. In contrast, the improved Schlumberger array does not yield a reasonable model and lacks vertical resolution.

5.1.3 3D–Inversion

The test field was also used for the application of the three-dimensional resistivity inversion. Figure 5.8 shows the measuring field. The data were acquired by Dr. Folker Donner (Institute of Geophysics, Freiberg).

The multi-electrode line was placed to form a grid of 14×9 electrodes using distances of 2.5 m in both directions. On all profiles in x- and y-direction dipole-dipole measurements were applied using the separation factors $n=1-6$. This resulted in a number of 753 single data that are visualized as pseudo-sections in Figure 5.9. The data show variations from 120 to $490 \Omega \text{ m}$. A relative error of 3% and a minimum voltage of $500 \mu\text{V}$ were used to estimate errors of 6% maximum.

The model was parameterized using cells of $\Delta x = \Delta y = 2.5 \text{ m}$. According to the six separations, six layers with increasing thickness up to maximum depth of 6 m were defined. The total number of $15 \times 10 \times 6 = 900$ cells is thus a small-scale 3D problem. Since the data show increasing resistivities with depth, a one-dimensional inversion was carried out to estimate a layered starting model. It represents the main resistivity distribution of a conductive overburden, sediments of moderate resistivity and the resistive basement at depth (see Table 5.2). In doing so, the data misfit could be reduced significantly.

z/m	0	1.1	2.3	3.4	5	7.3	9.6
$\rho/$	109	109	255	386	229	229	666
$\Omega \text{ m}$	Overburden			Sediments			base

Table 5.2: Layered starting model for the 3D Rothschönberg data

In the further course of inversion, smoothness constraints of first order using a regularization parameter of $\lambda = 10$ were applied globally. Convergence was reached within 5 iterations reducing the chi-squared misfit from 89.1 (RMS=33%) to 1.8 (RMS=4.3%). The final inversion model is displayed in Figure 5.10. The gallery is clearly to be seen in the second and third layers. In the same depth another resistive anomaly is visible, which can be followed down to model boundary. This supports the theory of a basement bulge.

In Figure 5.11 the model is visualized three-dimensionally. In addition to some slices in x- and y-direction an iso-body with a value of $450 \Omega \text{ m}$ is drawn. It provides a detailed

image of the Rothsönberg gallery and the neighboring anomaly.

Whereas the (two-dimensional) gallery is reproduced well in both 2D and 3D inversion, the neighboring anomaly leads to contradictory interpretations. Since the electrode distance of the 2D profile is much smaller, its resolution properties are better than for the 3D-inversion. However, the 2D inversion results has to be treated carefully, because the side anomaly is actually three-dimensional.

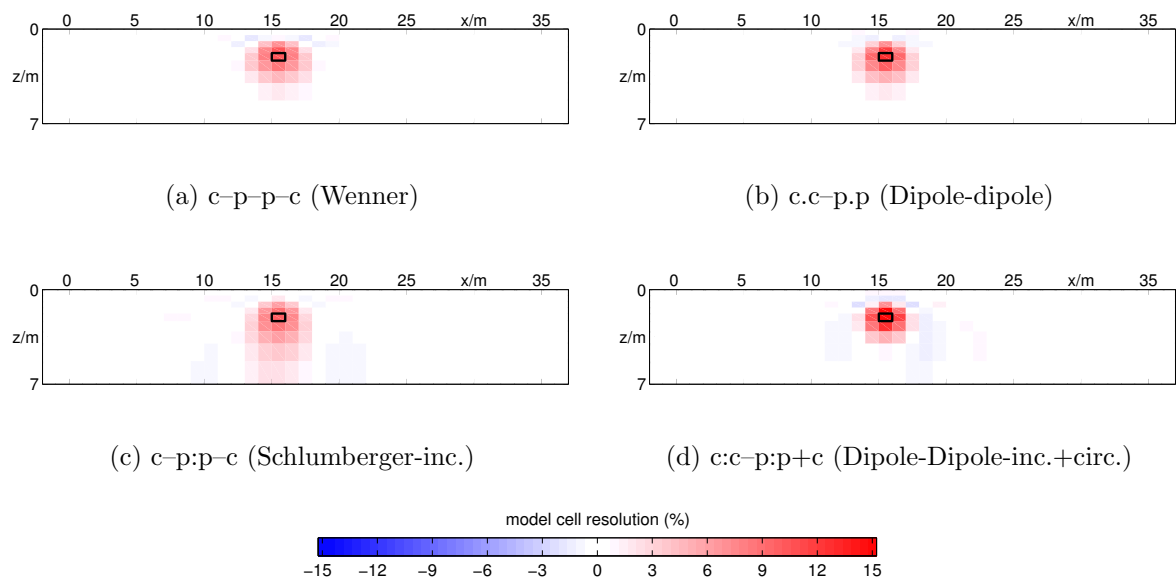


Figure 5.7: Cell resolutions of the cell 174 (anomaly center) in %. The model cell is marked by a black rectangle.

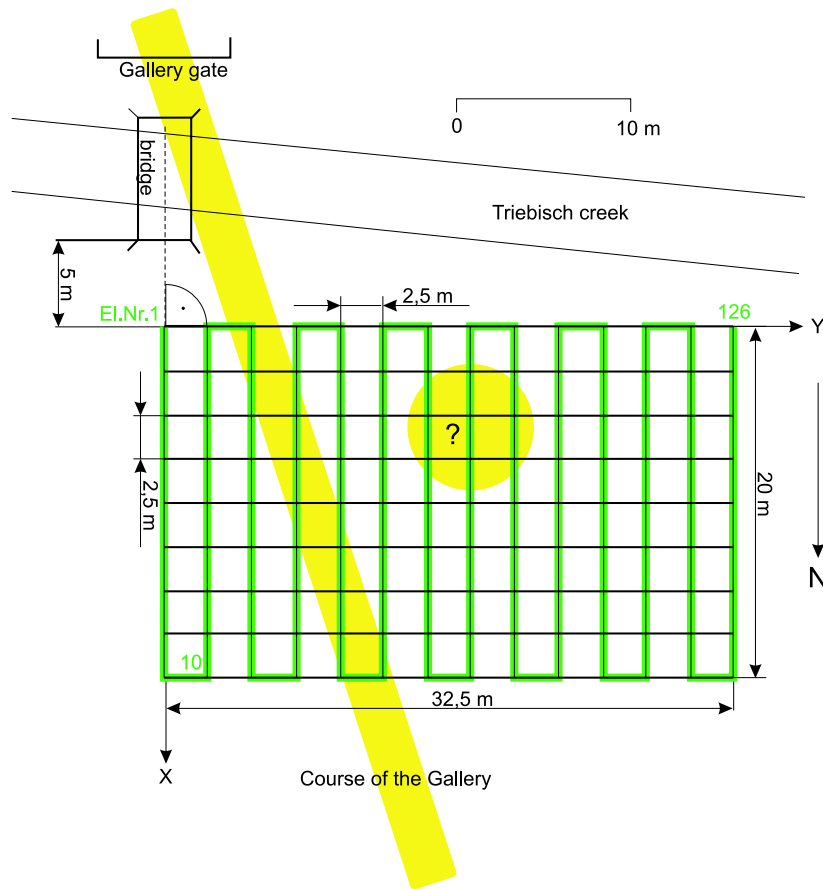


Figure 5.8: Electrode arrangement of the 3D measurement. The green line denotes the meander-like positioned multi-electrode, the yellow line denotes the course of the gallery

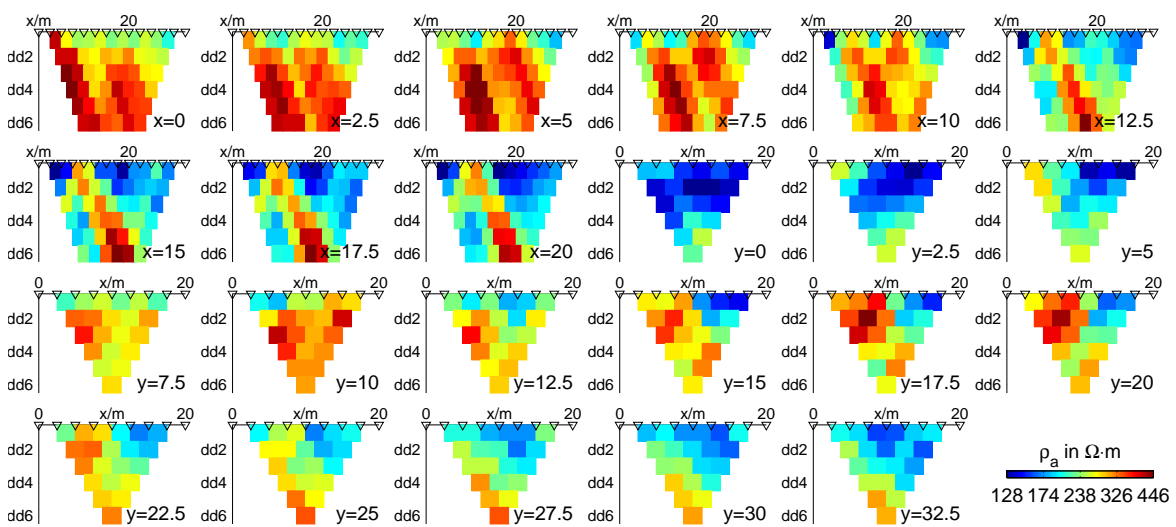


Figure 5.9: Measured data in form of pseudo-sections

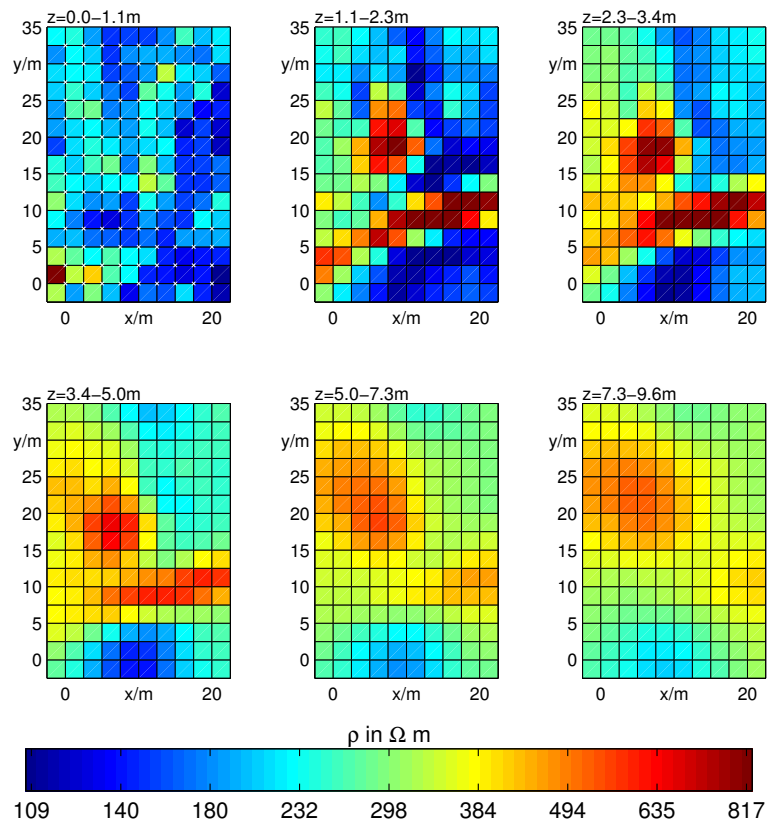


Figure 5.10: Plan view slices of the 3D inversion result

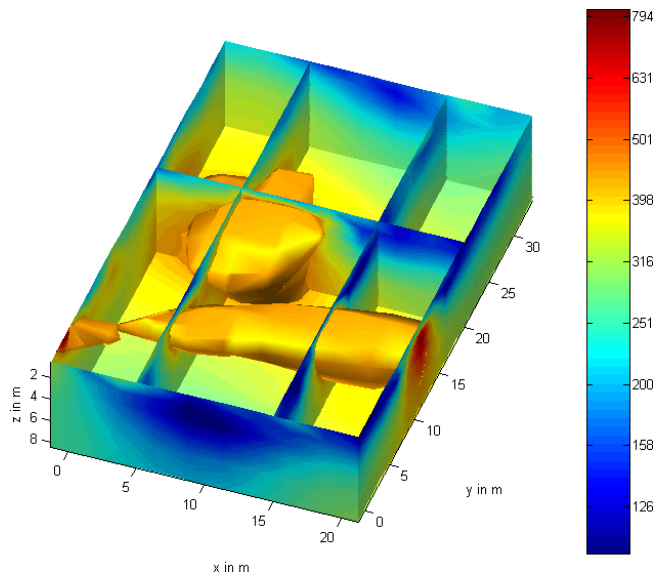


Figure 5.11: 3D-visualization of the inversion result, the iso-surface defined by the value of 450 $\Omega \text{ m}$ is drawn

5.2 Ground Fall Problems on Salt-waste Dumps

5.2.1 Motivation

In northern Thuringia salt layers are found at shallow depths. They have been commercially mined since the late 19th century. The wet residue of the salt production process was put on salt dumps from the middle of the 1960s on. It mainly contains chlorides (NaCl and MgCl₂) and sulphates (CaSO₄ and MgSO₄), which can dissolve in the presence of water leading to severe environmental problems.

To deal with the dissolution processes, the waste dumps have been covered since the early 1990s. Generally, three layers were established: Crude construction waste was disposed on top of the salt preventing fluids from capillary rise. Then, a layer of fine materials like soil and sewage sludge was added to compensate topographical undulations. The thickness of each layer may not have exceeded 5 m. Finally, a humus layer of at most 0.5 m was used to cover the waste dump allowing for planting vegetation.

The coverage diminishes the emergence of salt dissolutions. However, dissolution mechanisms are still in process. In recent years subsidences occurred that are obviously caused by dissolution processes in the salt waste. Figure 5.12 shows one of the ground falls, which can have radii of 10 m and depths of 5 m.



Figure 5.12: A recent ground fall at the salt-waste dump Menteroda

Ground falls represent a serious danger for both people and devices of the waste dump management. Using DC resistivity measurements the structure of known ground falls and the existence of possibly critical areas had to be investigated. The data were acquired at the salt-waste dump Menteroda by Marcus Walther within the scope of his diploma thesis in collaboration with the K-UTec GmbH Sondershausen.

5.2.2 Two-dimensional Investigation

At first, a profile array was applied crossing the filled ground fall. It employed 64 electrodes with an electrode distance of 1 m using a Schlumberger array type and systematically increasing dipole lengths. Figure 5.13 displays the pseudo-section of the measured data showing variations from 8 to 65 Ω m.

The model uses a regular horizontal grid spacing of 1 m. 12 layers with increasing thickness are defined reaching a maximum depth of 12 m as derived by the one-dimensional sensitivity distribution. The inversion routine applied global smoothness constraints of second order with Dirichlet boundaries. The regularization parameter of $\lambda = 3.0$ was chosen manually. A cell combination technique was used to prevent asymmetrical cell dimensions.

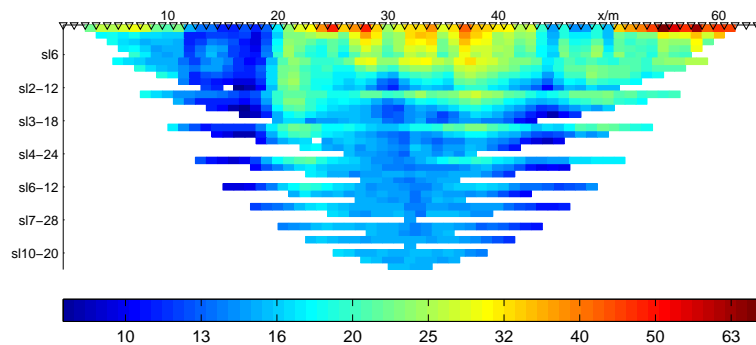


Figure 5.13: Pseudo-section of the measured data in Ω m

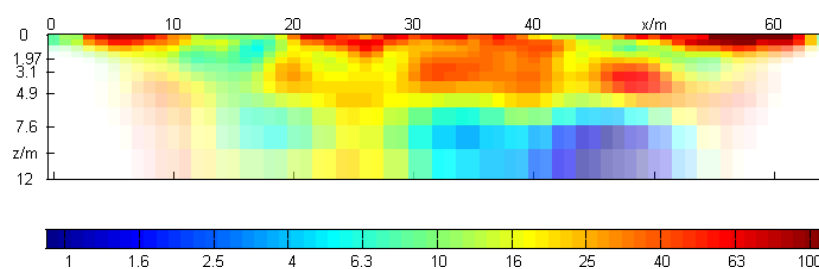


Figure 5.14: Inversion result (in Ω m) using global smoothness constraints (second order, Dirichlet boundaries) with $\lambda = 3$

The resulting model is displayed in Figure 5.14. It shows detailed structures of high and low resistivities in the range of 1–100 Ω m. The model resolution was used for alpha-mapping such that badly resolved model cells will not be interpreted. Generally, the upper layers are resistive, except for two conductive structures at $x = 17$ m and $x = 44$ m, which can be assigned to penetrating surface fluids. At a depth of 6 m the salt surface is visibly indicated by low resistivities. The location of the ground fall at $x = 27$ m is clearly visible due to a deflection of the layered structures near the surface.

At this position, increased resistivities are present down to the model boundary of $z=12$ m. The resistive material at depth is interpreted as loose covering material in a formerly rinsed cavity.

Hence, the objects of further investigations are either cavities at the salt surface or fluid paths towards depth, which can be followed by dissolution processes. However, the results of the 2D-inversion have to be handled with care, because the subsurface actually shows three-dimensional characteristics. From sensitivity analysis we know structures beside the electrode array being mapped into depth. Thus, for a reliable investigation, the subsurface has to be mapped by a surface electrode spread that allows for a 3D-inversion of resistivity.

5.2.3 Three-dimensional Investigation

The following data were acquired in the central part of the above discussed profile to investigate the three-dimensional structure of the ground fall and the surrounding area. The measurements were carried out on 11 parallel profiles of 1 m distance, shown in Figure 5.15. Each profile applied 64 electrodes of 0.5 m spacing. Note that two of the profiles were shifted by 0.25 m, which did not affect the inversion due to the possibility of independent electrode positioning.

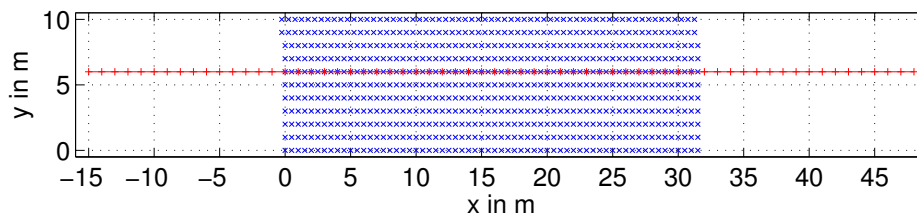


Figure 5.15: Layout of the 3D measurement. The "x" denote the electrodes of the 11 profiles, the "+" denote the electrodes of the 2D profile array

To ensure a high resolution, the dipole-dipole configurations were chosen. The above depicted improvement by a systematic enlargement of the dipole lengths was applied, each profile representing a combination of two shifted improved sets with 790 single data. Using a multi-channel receiver the total number of 8690 data was obtained in one day and could be used without pre-processing. Errors were estimated to be 3% plus a voltage accuracy of $100 \mu\text{V}$ at 50 mA, which obtained 11% at maximum. The measured pseudo-sections are displayed in Figure 5.16. They show variations from $3 \Omega\text{m}$ (large penetrations) to $70 \Omega\text{m}$ (shallow data).

The model consists of model cells with 0.5 m horizontal grid spacing and 8 layers of increasing thickness with a maximum depth of 8.7 m. This results in a total of $65 \times 22 \times 8 = 11440$ cells. The inversion routine applied smoothness constraints of

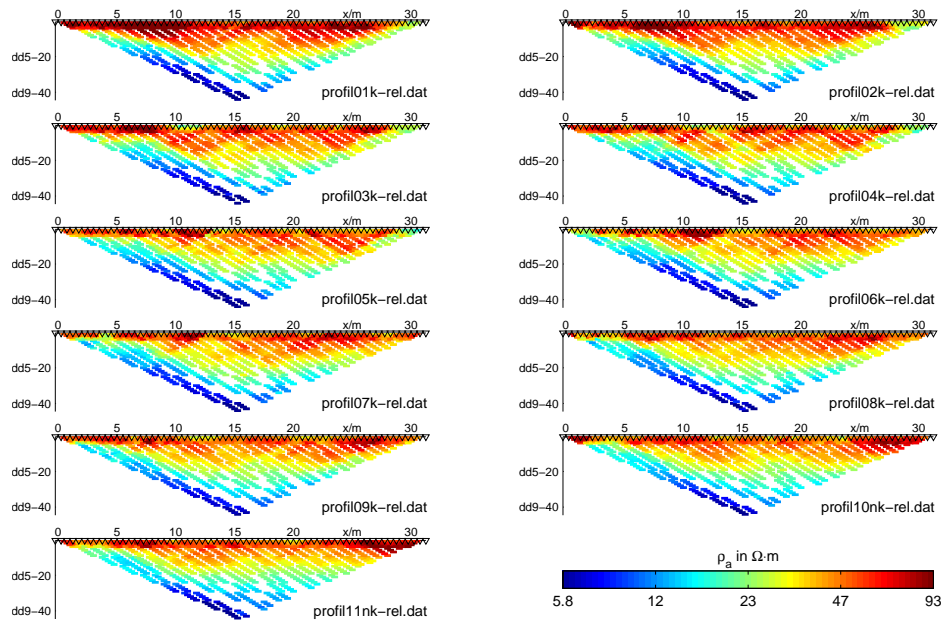


Figure 5.16: Pseudo-sections of the 11 measured profiles

second order with Dirichlet boundaries globally using the regularization parameter $\lambda=10.0$. As a result, the chi-squared data misfit has been decreased using 5 iterations from 406.6 (RMS=131%) down to 5.6 (RMS=7.8%).

Figure 5.17 displays horizontal sections of the inversion model. While the top layers show a complicated image of resistive material, the highly conductive salt surface is indicated in the deepest layers. Another conductive anomaly on the left side can be interpreted as mineral sewage sludge as known from borehole material. In layers 4-7 ($z = 1.3 - 6.6$ m) two resistive anomalies can be identified: The known ground fall is situated at the position defined by $x = 11$ m and $y = 10$ m. This anomaly can be followed down to 6 m. Close to the right model boundary ($x = 30$ m, $y = 8$ m) a huge resistive body can be seen with depth extending from 2 m down to 8.7 m. It can possibly be interpreted as existing cavity. This idea is supported by the low resistivities above, which may indicate fluid paths as the origins of salt dissolution. However, this has to be proved by further investigations, because the resolution at the boundary of the model is generally low. Nevertheless, Figure 5.18 provides a three-dimensional concept of the subsurface as being determined from DC resistivity measurements.

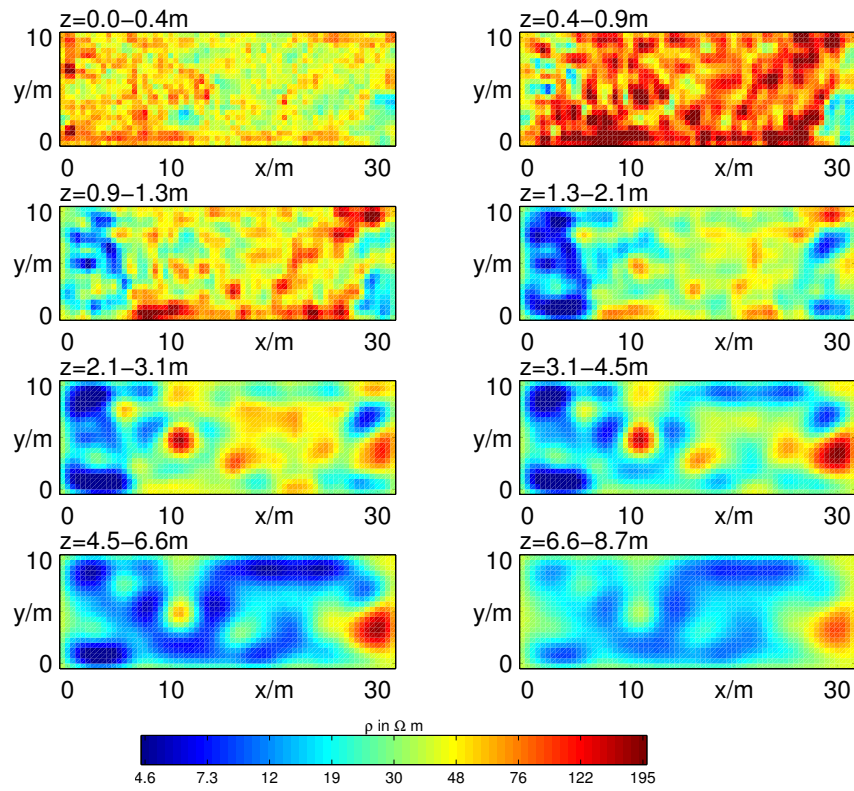


Figure 5.17: Plan view sections of the inversion result

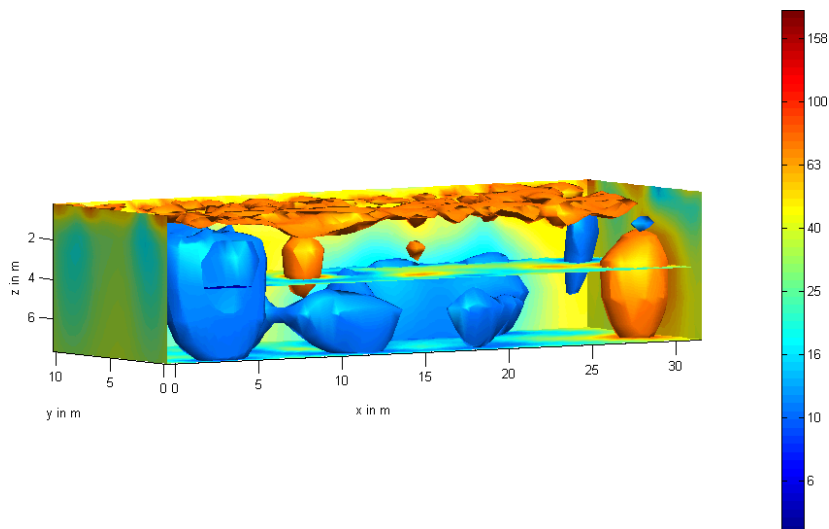


Figure 5.18: 3D-Visualization of the inversion result. The iso-surfaces created by the resistivities of 10 $\Omega \text{ m}$ and 100 $\Omega \text{ m}$ are displayed

5.3 Archaeological Investigation at the Walbeck Castle

Geophysical methods have been successfully used for the investigation of archaeological sites. They help to delineate traces of history without destruction and save time that may be used for a more careful excavation. Generally, the methods of choice are geo-magnetics, ground penetrating radar (GPR) and resistivity tomography used for both broad studies and detailed two/three-dimensional investigations. To delineate archaeological objects, they must show petrophysical parameters contrasting to the background. This is often the case for walls and foundations, which generally show large resistivities if still in good state. However, there are also cases of conductive anomalies due to water accumulation on top of the foundations.

The following example involves data collected by Dr. Donner at the Walbeck castle near Magdeburg. Next to the existing collegiate church ancient side buildings were supposed to be present. To detect wall foundations the area had been investigated using GPR and DC measurements. Figure 5.19 displays a plan view of the area and the electrode layout.

Dipole-dipole measurements were carried out on 29 profiles trending East-West and 17 profiles striking from North to South. Both electrode and profile distance have been set to 1 m. Thus, an area of 30×25 m was covered allowing for three-dimensional resistivity reconstruction. Figure 5.20 shows the collected data in the form of pseudo-sections. The separation factor between the two dipoles was varied from 1 to 6. Therefore a dipole enlargement was not necessary.

In summary, a total number of 5196 data points was acquired, all of them showing excellent quality. The errors were estimated to be 2% plus $100 \mu V$ at a driving current of 100 mA. It is noteworthy that archaeological data often show only small variances ($100 - 300 \Omega m$) compared to other problems. However, due to the shallow depths of the investigated objects a good data quality allows for accurate results anyway.

The model domain constitutes of six layers, the layer thicknesses vary from 0.5 to 1.0 m yielding a maximum depth of 4 m. Due to the excellent data coverage, a grid with horizontal cell dimensions of $\Delta x = \Delta y = 0.5$ m could be applied. A cell grid of $60 \times 50 \times 6 = 18000$ model parameters was used for inversion. Smoothness constraints of first order were used in a global regularization scheme of constant regularization parameter. The value of $\lambda = 10$ was adjusted manually considering the discrepancy principle.

The inversion parameters are displayed in Table 5.3. Note that the first two inversion steps show fast convergence, whereas the following iterations can improve data misfit only slightly. A possible reason for the slow convergence may be forward calculation errors, which rise up to the order of the RMS error.

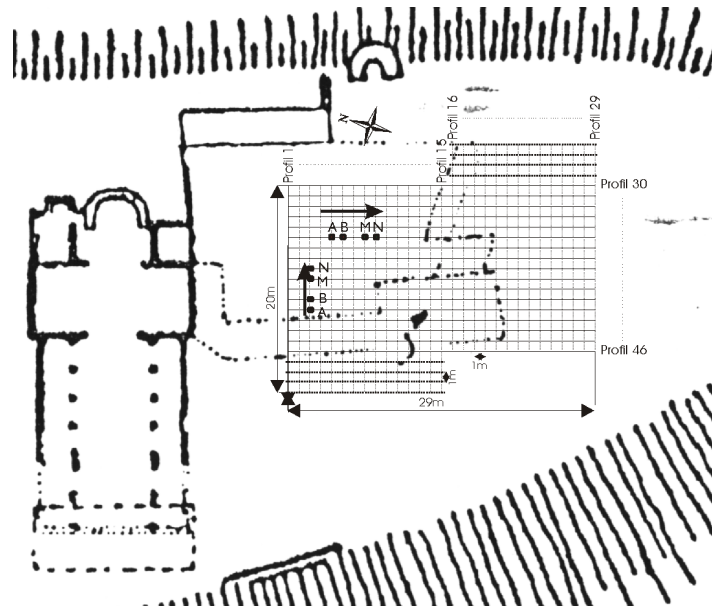


Figure 5.19: Area of measurement and electrode arrays. Black dots denote the position of known or guessed walls based on the historical site plan

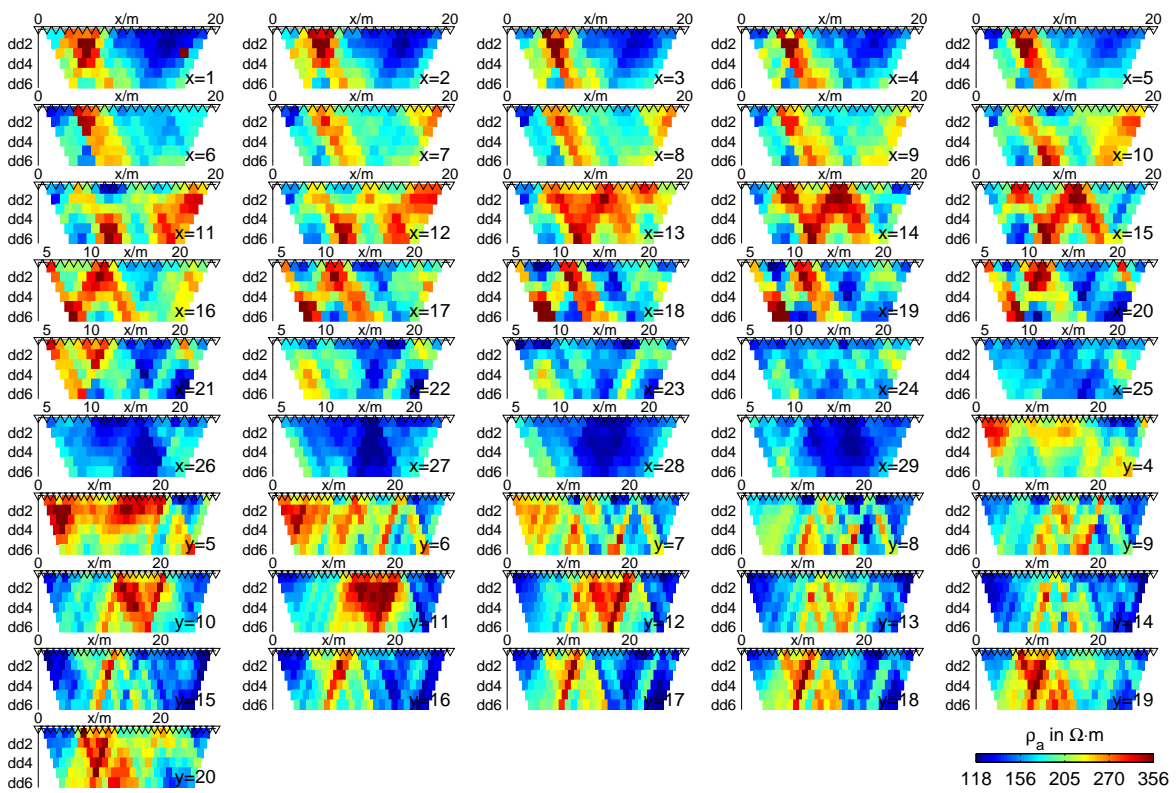


Figure 5.20: Pseudo-sections of the measured data

Iteration	0	1	2	3	4	5	6
χ^2	70.2	4.1	1.9	1.6	1.5	1.4	1.3
RMS error in %	25.6	5.3	4.1	3.7	3.6	3.5	3.4
Reciprocity in %		1.44	1.97	2.33	2.58	2.77	2.94

Table 5.3: Data fit and reciprocity in the course of iterations

The final inversion result is displayed in Figure 5.21. Whereas in the first layer only small anomalies are present, the layers 2-4 show more detailed variations. The resistive linear structures can be associated with wall foundations with the utmost probability. Note that the resolution is decreased in deeper layers as well as in the regions of missing electrodes.

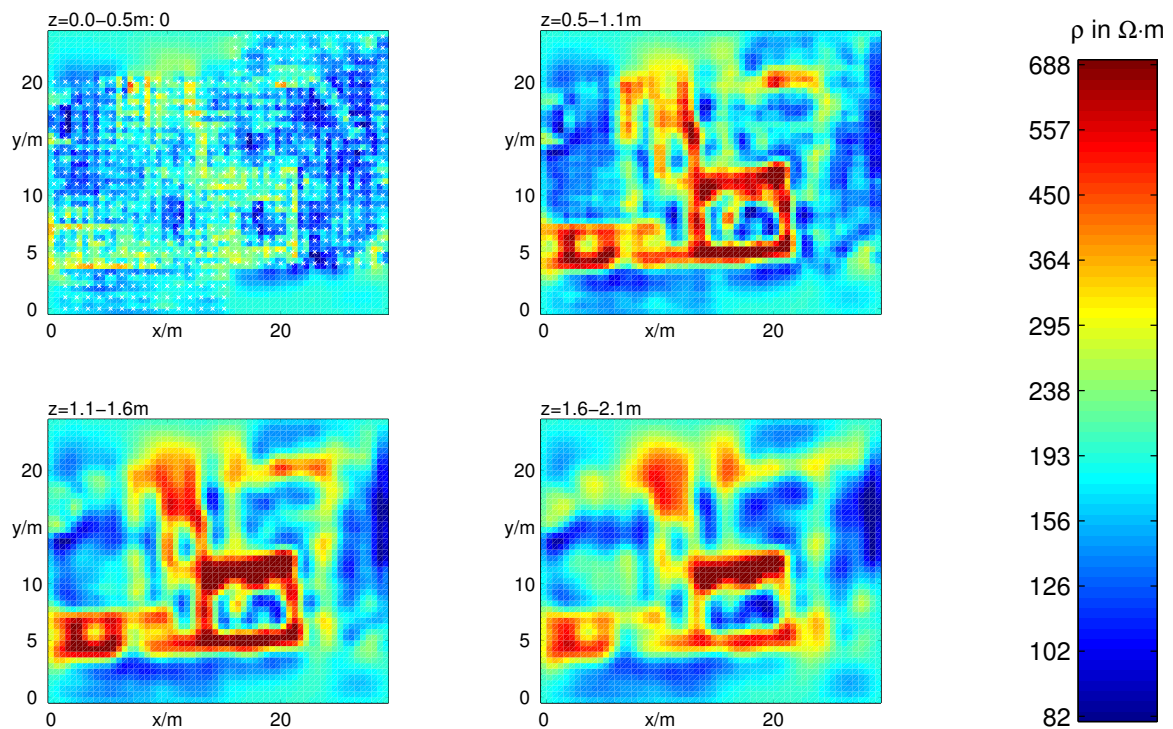


Figure 5.21: Inversion result sections of the upper four layers

To verify the results, Figure 5.22 shows an overlap of the second layer ($z = 0.5 - 1.1$ m) section and the original sketch. Most of the expected structures are mapped well by the resistivity, whereas others are not. Moreover, systematically situated resistivity structures indicate additional foundations. To sum it up, the use of the resistivity inversion provides a very detailed image of the subsurface, which helps to obtain a concept of the interior structures without destruction.

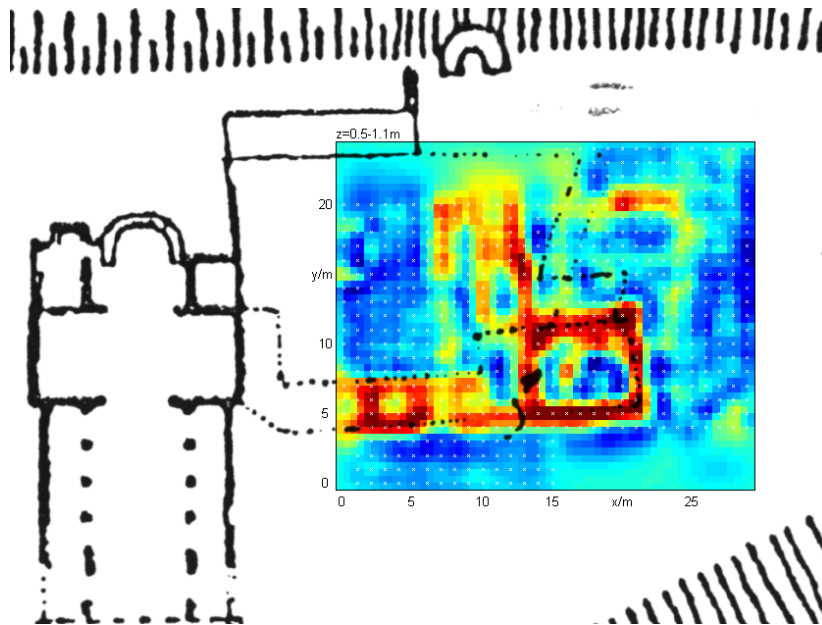


Figure 5.22: Second layer ($z=0.5-1.1$ m) plan view of the inversion result compared to the original sketch of the site plan

6 Discussion and Conclusions

The proposed minimization techniques can be applied to the inversion of DC resistivity data. In doing so, the knowledge of errors is important for weighting the data and controlling the inversion. Uncertainties in the measured voltages and the electrode positions have to be considered to obtain appropriate estimates. Note that a misjudgment of the errors can lead to erroneous structures. It was shown that covariances in the data due to varying electrode positions can be neglected.

The Gauss-Newton technique proves to be the method of choice for multidimensional problems. Different regularization techniques were successfully applied to circumvent the non-uniqueness. The author recommends the use of explicit global regularization schemes by the introduction of model restrictions such as smoothness constraints or a-priori information. This is supported by the resolution equation (2.42) for global regularization schemes. It predicts the model to be constructed of the true model and the starting model independent of the path the iterations took.

Consequently, the regularization parameter used for weighting the constraints has to be constant. The value can be estimated using the L-curve criterion in the first iteration. Finally, it is determined by the discrepancy principle which ensures to fit the data within their errors. As presented, a global L-curve is constructed along which the optimum regularization is to be sought. The use of an inexact line search procedure helps to accelerate convergence.

The non-linear resolution analysis played a central role in this thesis. The concept of Friedel (2003) based on an truncated SVD scheme could be broadened to arbitrary regularization schemes using the generalized singular value decomposition (GSVD). However, both SVD and GSVD cannot be applied to large-scale problems due to limitations of computer time and memory. Alternatively, each row of the model resolution matrix can be approximated solving a linear sub-problem as done by Alumbaugh and Newman (2000). To obtain the complete resolution matrix, a large computational effort is necessary. Preconditioners can help to overcome this problem slightly. Other methods have to be used to approximate the model resolution matrix.

The derived measure of information content represents an appropriate measure for the quality of inversion results. It is nearly independent of the used regularization scheme. To circumvent the dependence on the regularization parameter, a linearization provides the basis for a comparison of data sets. As a result, improved dipole-dipole,

pole-dipole and Schlumberger type data sets show superior quality. The optimization strategies have already been applied in practice. Whereas the improved dipole-dipole data set show excellent results, the Schlumberger data lack vertical resolution. Further resolution studies with large electrode spacings and increased noise levels have to be carried out. Moreover, the efficiency considerations shall be broadened to involve multi-channel recorders.

The resolution analysis is extended to three-dimensional inversion by transferring the 2D optimization techniques to the experimental design of 3D data. As a result of the synthetic studies and the field data examples, a successful 3D interpretation requires profiles of small distances or perpendicular directions.

The presented forward calculation with finite differences and the singularity removal technique has been enhanced using incomplete Cholesky preconditioners. The reciprocity proves to be an appropriate measure of modeling accuracy. Since inaccurate forward calculation limit the convergence of the inversion, the reciprocity has to be observed for adaptive refinement of the grid. However, the forward procedure represents the most time-consuming part of the inversion.

A consequent use of finite element techniques is proposed. Advantages are the incorporation of topography and the construction of a parameterization that considers the resolution abilities. Also, the presented regularization techniques have to be developed. For this purpose, adaptive constraint weights as presented by Yi *et al.* (2003) or Haber and Tenorio (2003) may be introduced.

Equation solvers have been presented to solve the linear sub-problem on the basis of conjugate gradients techniques. A further improvement of efficiency can probably be achieved by preconditioning. The multi-lambda solvers needed for choosing the regularization parameter are to be accelerated. Particularly, the fast equation solvers as presented by Frommer and Maass (1999) have to be generalized for arbitrary constraints.

The future requirements of DC resistivity inversion involve the following tasks:

1. Implementation of robust inversion schemes for model and data weighting.
2. Application to cross-hole and hole-to-surface data, including the estimation of errors and the optimization of data sets.
3. Inversion of complex resistivity (IP); frequency constraints are needed for the use of spectral IP.
4. 4D inversion for time dependent processes using appropriate constraints in time.

List of Figures

2.1	Typical L–curve as arising in ill-posed problems for the damped least squares solution of $\mathbf{Ax} = \mathbf{b}$	23
2.2	L–curve and its curvature for a reasonable interval of λ , the optimum regularization parameter is denoted by a circle	24
2.3	Singular values and filter factors for a typical ill-posed problem, note that the filter factors decrease faster	32
2.4	Transform function of Tikhonov ($\lambda = 10, \lambda = 5$) and TSVD ($r_{TSVD} = 90$) regularization for the sample problem of Figure 2.3	34
3.1	Sketch of a four-point DC measurement using surface electrodes	45
3.2	Relative geometric errors of several configurations with electrode spacing $x = 1$ m, $\delta x = 1$ cm	50
3.3	Data covariance structure for one configuration out of a dipole-dipole data set	51
3.4	Relative errors due to voltage accuracy of several configurations with electrode distance 1 m, $\delta U = 500 \mu\text{V}$, $I = 100$ mA, $\rho_a = 100 \Omega \text{ m}$	52
3.5	One-dimensional sensitivity distribution of a dipole-dipole measurement using $a = 2$ m and $n = 4$	55
3.6	Penetration depth (90% sensitivity) per electrode spacing a for several configurations as a function of separation factor n	56
3.7	Sample two-dimensional Grid and Para type models	56
3.8	Section of the FD grid for the grid node ijk	58
3.9	Plan view of the line between the grid nodes (i,j,k) and $(i,j,k+1)$, the dashed line represents the equivalent area with the weighted conductivity	59
3.10	Comparison of results for the conductive dike, (a) apparent resistivity, (b) accuracy and reciprocity	65
3.11	Synthetic model used for the accuracy investigations	66
3.12	Model section at $y = 0$ m	76
3.13	Homogeneous sensitivity distribution	76
3.14	Inhomogeneous sensitivity distribution	77
4.1	Model parameterization and synthetic model	82

4.2	Convergence of NLCG and the Gauss-Newton method, local and global regularization	82
4.3	First iteration models for NLCG and Gauss-Newton	82
4.4	Inversion results for regularization Gauss-Newton scheme	82
4.5	Comparison of line search procedures, the exact line search curve (crosses) is approximated by a two-point parabola (straight line) and by linear interpolation (circles)	83
4.6	The L-curve of the linear sub-problem	84
4.7	Data misfit χ^2 and total functional Φ as a function of λ and τ	85
4.8	Model resolution of the individual model cells (in %)	86
4.9	Model cell resolutions (in %) for 4 selected parameters, the cells are marked by black rectangles	86
4.10	Resolution degree as a function of the regularization parameter	87
4.11	Synthetic data sets for the investigated data	89
4.12	Inversion results for different inversion schemes using data set c:c-p:p	90
4.13	Inversion results for different data sets using $\varepsilon=1\%$, $\delta U = 100 \mu\text{V}$ and $I = 100 \text{ mA}$	91
4.14	Resolution radii for data sets c-p-p-c, c-p.p, c.c-p.p and c:c-p:p-c	92
4.15	Iteration results of unchanged and reduced parameters	94
4.16	Inversion results for uncorrelated and correlated errors	95
4.17	Model parameterization and synthetic model for Grid (a) and Para (b) type of model	98
4.18	Inversion models of the c:c-p:p+c data set	98
4.19	Model resolution for Grid (a) and Para (b) type of model	98
4.20	Synthetic model used, the green dots denote the electrodes of the five profiles	100
4.21	Second layer plan view of a single parameter resolution (in %) for varying profile distance, the according cell is indicated by black lines	100
4.22	Synthetic model of a flood protection dike	103
4.23	Pseudo-sections of the synthetic data used for inversion representing dipole-dipole data with increased dipole lengths. The profile location is indicated in the bottom right-hand corner of each section.	103
4.24	Section of the parameter grid, red is the earth's surface, the green tetrahedrons are the model parameter and the boundary elements are blue.	105
4.25	The secondary field grid, derived by global refinement.	105
4.26	Section of the primary field grid. The local refinement at the electrodes makes their position clearly visible.	105
4.27	Inversion result of the synthetic data. All tetrahedrons with resistivities above 260 and below 60 $\Omega \text{ m}$ are displayed.	107

5.1	Section of the investigated Rothschnberg gallery	111
5.2	Field data of the original Rothschnberg profile representing a dipole-dipole pseudo-section of 21 electrodes	112
5.3	Global L-curve displaying data misfit χ^2 and model roughness for varying λ , the curvature of the curve has a clear maximum defining the optimum λ	113
5.4	Inversion results for six selected values of λ	113
5.5	Error-weighted misfit as pseudo-section and histogram for model (d) using $\lambda_{19} = 18.0$	116
5.6	Inversion results of the depicted data sets	116
5.7	Cell resolutions of the cell 174 (anomaly center) in %. The model cell is marked by a black rectangle.	119
5.8	Rothschnberg, electrode arrangement of the 3D measurement	120
5.9	Measured data in form of pseudo-sections	120
5.10	Plan view slices of the 3D inversion result	121
5.11	3D-visualization of the inversion result, the iso-surface defined by the value of $450 \Omega \text{ m}$ is drawn	121
5.12	A recent ground fall at the salt-waste dump Menteroda	122
5.13	Pseudo-section of the measured data in $\Omega \text{ m}$	123
5.14	Inversion result (in $\Omega \text{ m}$) using global smoothness constraints (second order, Dirichlet boundaries) with $\lambda = 3$	123
5.15	Layout of the 3D measurement. The "x" denote the electrodes of the 11 profiles, the "+" denote the electrodes of the 2D profile array	124
5.16	Pseudo-sections of the 11 measured profiles	125
5.17	Plan view sections of the inversion result	126
5.18	3D-Visualization of the inversion result. The iso-surfaces created by the resistivities of $10 \Omega \text{ m}$ and $100 \Omega \text{ m}$ are displayed	126
5.19	Area of measurement and electrode arrays. Black dots denote the position of known or guessed walls based on the historical site plan	128
5.20	Pseudo-sections of the measured data	128
5.21	Inversion result sections of the upper four layers	129
5.22	Second layer ($z = 0.5 - 1.1 \text{ m}$) plan view of the inversion result compared to the original sketch of the site plan	130

List of Tables

2.1	Classification of regularization schemes regarding the effect on model or model update and explicit or implicit formulation	16
3.1	Classification of 2D multi-electrode data sets	47
3.2	Reciprocity (in %) as a function of refinement and solving accuracy, contrast factor $fa_k=5$	66
3.3	Reciprocity (in %) as a function of the resistivity contrast	67
3.4	Comparison of different preconditioners	67
3.5	Gauss-Legendre integration of the sensitivity for a hexahedron	72
4.1	Definition of data sets according to the above depicted classification	88
4.2	Comparison of different inversion methods, data set c:c-p:p, $\varepsilon=1\%$, $\delta U=100\ \mu\text{V}$, $I=100\ \text{mA}$, $a=1\ \text{m}$	90
4.3	Comparison of different data sets corresponding to Figure 4.13	91
4.4	Comparison of different data sets for $\varepsilon=2\%$, $\delta U=1\ \text{mV}$ at $I=100\ \text{mA}$	93
4.5	Comparison of different parameterization techniques for data set c:c-p:p	94
4.6	Comparison of 3D data sets concerning information content (IC) and efficiency (IE)	97
4.7	Node and element numbers of the three grids	104
4.8	Inversion parameters in the course of inversion	107
5.1	The four data sets used for comparison	115
5.2	Layered starting model for the 3D Rothsönberg data	117
5.3	Data fit and reciprocity in the course of iterations	129

Bibliography

- Alumbaugh, D. L. and Newman, G. A. (2000). Image appraisal for 2-d and 3-d electromagnetic inversion. *Geophysics*, **65**(5), 1455–1467.
- Backus, G. and Gilbert, F. (1968). The resolving power of gross earth data. *Geophys. J. R. Astr. Soc.*, **16**, 169–205.
- Barker, R. D. (1989). Depth of investigation of collinear symmetrical four-electrode arrays. *Geophysics*, **54**(8), 1031–1037.
- Beard, L. B., Hohmann, G. W., and Tripp, A. C. (1996). Fast resistivity/ip inversion using a low contrast approximation. *Geophysics*, **61**, 169–179.
- Berryman, J. G. (2000). Tomographic resolution without singular value decomposition. Lawrence Livermore National Laboratory.
- Bing, Z. and Greenhalgh, S. A. (2001). Finite element three-dimensional direct current resistivity modelling: Accuracy and efficiency considerations. *Geophys. J. Int.*, **145**, 679–688.
- Brewitt-Taylor, C. R. and Weaver, J. T. (1976). On the finite difference solution of two-dimensional induction problems. *Geophys. J. R. astr. Soc.*, **47**, 375–396.
- Broyden, C. G. (1972). Quasi-Newton methods. In W. Murray, editor, *Numerical methods for unconstrained optimization*. Academic Press.
- Claerbout, J. F. and Muir, F. (1973). Robust modeling with erratic data. *Geophysics*, **38**(1), 826–844.
- Coggon, J. H. (1971). Electromagnetic and electrical modeling by the finite element method. *Geophysics*, **36**, 132–155.
- Cole, K. S. and Cole, R. H. (1941). Dispersion and absorption in dielectrics, I: Alternating current fields. *J. Chem. Phys.*, **1**, 341–351.
- Constable, S. C., Parker, R. L., and Constable, C. G. (1987). Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, **52**, 289–300.

- Dahlin, T. and Loke, M. H. (1998). Resolution of 2d Wenner resistivity imaging as assessed by numerical modelling. *Journal of Applied Geophysics*, **38**, 237–249.
- Daily, W. and Owen, E. (1991). Cross-borehole resistivity tomography. *Geophysics*, **56**, 1228–1235.
- Dey, A. and Morrison, H. F. (1979a). Resistivity modeling for arbitrarily shaped three-dimensional structures. *Geophysics*, **44**(4), 753–780.
- Dey, A. and Morrison, H. F. (1979b). Resistivity modelling for arbitrarily shaped two-dimensional structures. *Geophys. Prosp.*, **27**(1), 106–136.
- Dietrich, P. (1999). *Konzeption und Auswertung gleichstromgeoelektrischer Tracerver-suche unter Verwendung von Sensitivitätskoeffizienten*. Ph.D. thesis, Institut für Geologie und Paläontologie, Tübingen.
- Dorn, O., Bertete-Aguirre, H., Berryman, J. G., and Papanicolaou, G. C. (2002). Sensitivity analysis of a nonlinear inversion method for 3d electromagnetic imaging in anisotropic media. *Inverse Problems*, **18**(2), 285–317.
- Ellis, R. G. and Oldenburg, D. W. (1994). The pole-pole 3-d dc-resistivity inverse problem: a conjugate gradient approach. *Geophys. J. Int.*, **119**, 187–194.
- Farquharson, C. G. and Oldenburg, D. W. (1998). Non-linear inversion using general measures of data misfit and model structure. *Geophys. J. Int.*, **134**, 213–227.
- Farquharson, C. G. and Oldenburg, D. W. (2004). A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. *Geophys. J. Int.*, **156**, 411–425.
- Fox, R. C., Hohmann, G. W., Killpack, T. J., and Rijo, L. (1980). Topographic effects in resistivity and induced-polarization surveys. *Geophysics*, **45**(1), 75–93.
- Friedel, S. (1997). Hochauffösende Geoelektrik - Geoelektrische Impedanztomographie. In M. Beblo, editor, *Umweltgeophysik*, pages 131–151. Ernst und Sohn, Berlin.
- Friedel, S. (2000). Über die Abbildungseigenschaften der geoelektrischen Impedanztomographie unter Berücksichtigung von endlicher Anzahl und endlicher Genauigkeit der Messdaten. Dissertation, Shaker, Aachen. Universität Leipzig.
- Friedel, S. (2003). Resolution, stability and efficiency of resistivity tomography estimated from a generalized inverse approach. *Geophys. J. Int.*, **153**, 305–316.
- Frommer, A. and Maass, P. (1999). Fast cg-based methods for tikhonov-phillips regularization. *SIAM J. Sci. Computing*, **20**(5), 1831–1850.

- Geselowitz, D. B. (1971). An application of electrocardiographic lead theory to impedance plethysmography. *IEEE Trans. Biomed. Eng.*, **BME-18**(1), 38–41.
- Golub, G. H. and van Loan, C. F. (1996). *Matrix Computations*. The Johns Hopkins University Press.
- Gomez-Trevino, E. (1987). A simple sensitivity analysis of time-domain and frequency-domain electromagnetic measurements. *Geophysics*, **52**(10), 1418–1423.
- Günther, T. (2000). Vergleichende Analyse zum Auflösungsvermögen geoelektrischer und elektromagnetischer Verfahren. Unpublished diploma thesis, <http://www.geophysik.tu-freiberg.de/spitzer/download/>.
- Haber, E. and Tenorio, L. (2003). Learning regularization functionals - a supervised training approach. *Inverse Problems*, **19**, 611–626.
- Hansen, P. C. (1992). *Regularization Tools - A Matlab Packager for Analysis and Solution of discrete ill-posed problems*. Department of Mathematical Modelling, Technical University of Denmark, Lyngby, <http://www.imm.dtu.dk/~pch>, 3.0 edition.
- Hansen, P. C. and O’Leary, D. P. (1993). The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, **14**, 1487–1503.
- Hestenes, C. M. R. and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.*, **49**, 409–436.
- Inman, J. R. (1975). Resistivity inversion with ridge regression. *Geophysics*, **40**, 798–917.
- INRIA-Rocquencourt (2003). MEDIT: A user friendly mesh viewer, gamma project at INRIA-Rocquencourt. <http://www-rocq1.inria.fr/gamma/medit/medit.html>.
- Jegen, M. D., Everett, M. E., and Schultz, A. (2001). Using homotopy to invert geophysical data. *Geophysics*, **66**(6), 1749–1760.
- Kemna, A. (2000). Tomographic inversion of complex resistivity. Ruhr-Universität Bochum.
- Kemna, A., Binley, A., Ramirez, A., and Daily, W. (2000). Complex resistivity tomography for environmental applications. *Chemical Engineering Journal*, **77**, 11–18.
- Kemna, A., Vanderborght, J., Kulesa, B., and Vereecken, H. (2002). Imaging and characterization of subsurface solute transport using electrical tomography (ERT) and equivalent transport models. *Journal of Hydrology*, **267**, 125–146.

- Kemna, A., Binley, A., and Slater, L. (2004). Crosshole ip imaging for engineering and environmental applications. *Geophysics*, **69**(1), 97–107.
- Kilmer, M. E. and O’Leary, D. P. (2001). Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM Journal on matrix analysis and applications*, **22**, 12004–1221.
- LaBreque, D. J. and Ward, S. H. (1990). Two-dimensional cross-borehole resistivity model fitting. In S. Ward, editor, *Geotechnical and environmental geophysics*, volume III, pages 51–74. Soc. Expl. Geophys.
- Li, Y. and Oldenburg, D. W. (1992). Approximate inverse mappings in dc resistivity problems. *Geophys. J. Int.*, **109**, 343–362.
- Li, Y. and Oldenburg, D. W. (1999). 3-d inversion of dc resistivity data using an l-curve criterion. In *69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts*, pages 251–254.
- Li, Y. and Oldenburg, D. W. (2003). Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method. *Geophys. J. Int.*, **152**, 251–265.
- Li, Y. and Spitzer, K. (2002). Three-dimensional dc resistivity forward modelling using finite elements in comparison with finite-difference solutions. *Geophys. J. Int.*, **151**, 924–934.
- Loke, M. H. and Barker, R. D. (1996a). Practical techniques for 3d resistivity surveys and data inversion. *Geophysical Prospecting*, **44**, 499–523.
- Loke, M. H. and Barker, R. D. (1996b). Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method. *Geophysical Prospecting*, **44**, 131–152.
- Loke, M. H. and Dahlin, T. (2002). A comparison of the Gauss-Newton and quasi-Newton methods in resistivity imaging inversion. *Journal of Applied Geophysics*, **49**, 149–162.
- Lowry, T., Allen, M. B., and Shive, P. N. (1989). Singularity removal: A refinement of resistivity modeling techniques. *Geophysics*, **54**(6), 766–774.
- Mackie, R. L. and Madden, T. R. (1993). Three-dimensional magnetotelluric inversion using conjugate gradients. *Geophys. J. Int.*, **115**, 215–229.

- Marquardt, D. W. (1963). An algorithm for least-squares estimation of non-linear parameters. *J. Soc. Ind. App. Math.*, **11**, 431–441.
- Maurer, H., Boerner, D. E., and Curtis, A. (2000). Design strategies for electromagnetic geophysical surveys. *Inverse Problems*, **16**, 1097–1117.
- Meju, M. A. (1994a). Biased estimation: A simple framework for inversion and uncertainty analysis. *Geophys. J. Int.*, **119**, 521–528.
- Meju, M. A. (1994b). *Geophysical Data Analysis: Understanding Inverse Problem Theory and Practice*. Society of Exploration Geophysicists.
- Menke, W. (1989). *Geophysical Data Analysis: Discrete Inverse Theory*, volume 45 of *International Geophysics Series*. Academic Press Inc.
- Mufti, I. R. (1976). Finite-difference resistivity modeling for arbitrarily shaped two-dimensional structures. *Geophysics*, **41**, 62–78.
- Narayan, S. (1992). Vertical resolution of two-dimensional dipole-dipole resistivity inversion. In *62nd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts*, volume 92, pages 431–434. Paper withdrawn.
- Oldenburg, D. W. and Li, Y. (1994). Inversion of induced polarization data. *Geophysics*, **59**, 1327–1341.
- Oldenburg, D. W. and Li, Y. (1999). Estimating depth of investigation in dc resistivity and ip surveys. *Geophysics*, **64**(2), 403–416.
- Oldenburg, D. W., McGillivray, P. R., and Ellis, R. G. (1993). Generalized subspace methods for large-scale inverse problems. *Geophys. J. Int.*, **114**, 12–20.
- Oraintara, S., Karl, W. C., Castanon, D. A., and Nguyen, T. Q. (1999). A reduced computation method for choosing the regularization parameter for tikhonov problems. In J. H. University, editor, *Proceedings of the 1999 conference on information and system science*.
- Pain, C. C., Herwanger, J. V., Saunders, J. H., Worthington, M. H., and de Oliveira, C. R. E. (2003). Anisotropic resistivity inversion. *Inverse Problems*, **19**, 1081–1111.
- Park, S. K. and Van, G. P. (1991). Inversion of pole-pole data for 3-d resistivity structure beneath arrays of electrodes. *Geophysics*, **56**, 951–960.
- Portniaguine, O. and Zhdanov, M. S. (1999). Focusing geophysical inversion images. *Geophysics*, **64**(3), 874–887.

- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1988-1992). *Numerical Recipes in C: The art of scientific computing*. Cambridge University Press.
- Pridmore, D. F., Hohmann, G. W., Ward, S. H., and Sill, W. R. (1980). An investigation of finite-element modelling for electrical and electromagnetic data in three dimensions. *Geophysics*, **46**, 1009–1024.
- Rodi, W. and Mackie, R. L. (2001). Nonlinear conjugate gradients algorithm for 2-d magnetotelluric inversion. *Geophysics*, **66**(1), 174–187.
- Roy, A. and Apparao, A. (1971). Depth of investigation in direct current methods. *Geophysics*, **36**(5), 943–959.
- Rücker, C. (1999). *Die Methode der finiten Elemente mit adaptiver Gittererzeugung und Integration eines Mehrgitterverfahrens zum Lösen des gleichstromgeoelektrischen Vorwärtsproblems*. Diploma thesis, Universität Leipzig.
- Rücker, C. (2003). 3D-FEM Widerstandsmodellierung unter Verwendung von unstrukturierten Tetraederdiskretisierungen. In A. Junge and A. Hördt, editors, *Tagungsband Elektromagnetische Tiefenforschung*, Königstein.
- Saad, Y. (1996). *Iterative Methods for Sparse Linear Systems*. PWS, Boston.
- Sasaki, Y. (1992). Resolution of resistivity tomography inferred from numerical simulation. *Geophys. Prosp.*, **40**, 453–463.
- Sasaki, Y. (1994). 3-d resistivity inversion using the finite-element method. *Geophysics*, **59**(12), 1839–1848.
- Seigel, H. O. (1959). Mathematical formulation and type curves for induced polarization. *Geophysics*, **24**, 547–565.
- Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. <ftp://warp.cs.cmu.edu>.
- Si, H. (2003). TETGEN: A 3d delaunay tetrahedral mesh generator. <http://tetgen.berlios.de>.
- Siripunvaraporn, W. and Egbert, G. (2000). An efficient data-subspace inversion method for 2-d magnetotelluric data. *Geophysics*, **65**(3), 791–803.
- Spies, B. R. and Habashy, T. M. (1995). Sensitivity analysis of crosswell electromagnetics. *Geophysics*, **60**(3), 834–845.

- Spitzer, K. (1995). A 3-d finite-difference algorithm for dc resistivity modelling using conjugate gradient methods. *Geophys. J. Int.*, **123**, 903–914.
- Spitzer, K. (1998). The three-dimensional dc sensitivity for surface and subsurface sources. *Geophys. J. Int.*, **134**, 736–746.
- Spitzer, K. (1999). *Development of three-dimensional finite difference resistivity, sensitivity, and IP forward modeling techniques and their application to surface and borehole surveys*. Habilitation thesis, University of Leipzig.
- Spitzer, K. and Wurmstich, B. (1995). Speed and accuracy in 3d resistivity modeling. In *3D EM Symposium, Ridgefield, Connecticut*.
- Stummer, P., Maurer, H., and Green, A. G. (2004). Experimental design: Electrical resistivity data sets that provide optimum subsurface information. *Geophysics*, **69**(1), 120–139.
- Sugimoto, Y. (1999). Shallow high-resolution 2-d and 3-d electrical crosshole imaging. *The Leading Edge*, pages 1425–1428.
- Tarantola, A. (1978). *Inverse Problem Theory*. Elsevier, Amsterdam.
- Tarantola, A. (2001). Logarithmic parameters. http://web.ccr.jussieu.fr/tarantola/Files/Professional/Papers_PDF/Music.pdf.
- Tong, L. and Yang, C. (1990). Incorporation of topography into two-dimensional resistivity inversion. *Geophysics*, **55**(3), 354–361.
- Tripp, A. C., Hohmann, G. W., and Swift, C. M. (1984). Two-dimensional resistivity inversion. *Geophysics*, **49**(10), 1708–1717.
- van Loan, C. F. (1976). Generalizing the singular value decomposition. *SIAM J. Num. Anal.*, **13**, 76–83.
- Vogel, C. R. (1996). Non-convergence of the l-curve regularization parameter selection method. *Inverse Problems*, **12**, 535–547.
- Vogel, C. R. (2002). *Computational Methods for Inverse Problems*. Frontiers in applied mathematics. Society for Industrial and Applied Mathematics.
- Wu, X., Xiao, Y., Qi, C., and Wang, T. (2003). Computations of secondary potential for 3d dc resistivity modelling using an incomplete choleski conjugate-gradient method. *Geophys. Prosp.*, **51**, 567.

-
- Yi, M.-J., Kim, J.-H., and Chung, S.-H. (2003). Enhancing the resolving power of least-squares inversion with active constraint balancing. *Geophysics*, **68**(3), 931–941.
- Zhang, J., Mackie, R. L., and Madden, T. R. (1995). 3-d resistivity forward modeling and inversion using conjugate gradients. *Geophysics*, **60**(5), 1313–1325.
- Zhang, J., Rodi, W., Mackie, R. L., and Shi, W. (1996). Regularization in 3-d dc resistivity tomography. *66th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts*, pages 687–693.
- Zhao, S. and Yedlin, M. J. (1996). Some refinements on the finite-difference method for 3-d dc resistivity modeling. *Geophysics*, **61**(5), 1301–1307.
- Zhou, B. and Greenhalgh, S. (2002). Rapid 2-d/3-d crosshole resistivity imaging using the analytic sensitivity function. *Geophysics*, **67**(3), 755–765.

Acknowledgments

The research was partially supported by the German Research Foundation (Ja 590/18-1). Furthermore, I'd like to thank the trustees of the SEG foundation for the Hohmann award.

The presented work would not have been possible without the help of many people I like to thank now.

First of all, I thank my supervisor Klaus Spitzer for financial support and his trust in me all the time. I also like to acknowledge the other supervisors of the thesis, Prof. Weidelt and Dr. Meju who gave important hints for improving the thesis. Prof. Peter Benner gave me financial support and the chance to see my subject from another, more mathematical point of view, which was, although short, very important to me.

I want to express my gratitude to Dr. Ralph-Uwe Börner for working and thinking together so many years and for his valuable advises.

I like to thank Dr. Sven Friedel for our fruitful discussions and for his excellent dissertation, which guided me from the beginning to the last words.

Thanks to the other members of the "Free your mind" working group (Jens, Christoph, Antje, Birgit, Kristina, Marcus, Yuguo, Xiaoping) for the great time we worked together and the inspiring lunch discussions.

For the fruitful co-work I like to thank Carsten Rücker, the finite element part of the 3D DC project. Our meetings were always very fruitful and inspiring and will certainly lead to further cooperation.

I highly appreciate the patience of all reviewers who helped me to improve the thesis, first of all Katrin, Kristina and Ralph-Uwe, but also Pierre, Carsten, Thomas, Peter and Sven.

Thanks to all the users of my inversion programs, who helped me to develop usable software which was one of my main aims.

I acknowledge Dr. Folker Donner, Marcus Walther and the K-UTec GmbH for providing some of the field data presented in the thesis.

Last, but not least, I express my gratitude to my friends and family members that made me the person I am now. Particularly, I thank my girl-friend Katrin who needed a lot of patience, but always had trust into me.

Index

- 2D DC classification, 47
- BFGS method, 13
- Boundary conditions, 20, 59
- Broyden update, 13
- CGLS, 26
 - CGLAN, 30
 - CGLSCD, 28
 - CGLSCDP, 28
 - CGLSI, 27
- Conjugate gradients, 11, 26, 63
 - preconditioned, 63
- Continuity equation, 44
- Data covariance matrix, 10
- Direct current, 43–47
- Discrepancy principle, 22, 112
- Errors, 48–53
 - geometrical, 48
 - voltage, 51
- Finite Differences, 57
- Finite Elements, 68
- Forward calculation, 57–70
 - 2D, 69
 - FD, 57
 - FE, 68
- Gauss-Newton method, 12, 81
- GCV, 23
- Generalized SVD, 34
- Grid Model, 55
- Hessian, 12, 13
- Information content, 39, 41
- Information efficiency, 40
- Jacobian, 9
- L-curve, 22, 84–85, 112
- Lagrangian parameter, 17
- Line search, 10, 13, 83–84, 112
- Linearized study, 87
- Logarithmic parameters, 75
- Logarithmized parameters, 53
- Model covariance matrix, 38, 39
- Model resolution matrix, 36, 37
- Newton’s method, 12–14
- NLCG, 11–81
- NLGC, 16
- Occam’s principle, 15
- Para Model, 56
- Parameterization, 53–56
- Preconditioner, 63
 - Incomplete Cholesky, 63, 67
 - Jacobi, 63, 67
 - SSOR, 63, 67
- Preconditioners, 67

- Quasi-Newton method, 13
- Reciprocity, 64
- Regularization, 14–25
 - explicit, 15, 17
 - global, 15
 - implicit, 14
 - local, 15
 - parameter, 17, 21
 - types, 16
- Resolution, 36–41
 - approximation, 41
 - degree, 39
 - equation, 38
 - linear, 36
 - measures, 39
 - non-linear, 37
 - radius, 40, 92
 - svd, 40
- Resolution analysis, 85
- Sensitivity, 9, 70–77
 - 1d, 54
 - 2d, 72
 - forward calculation, 73
 - perturbation method, 73
 - potential approximation, 74
- Singular value decomposition, 31–35
- Singularity removal, 61
- SIRT, 16
- Smoothness constraints, 19
- Steepest Descent, 10
- Steepest descent method, 10
- Topography, 101–108
- Truncated least squares, 14, 16, 27
- TSVD, 16