I. The methods of geophysical inversion

The development of computer science broke a new path in engineering technology. The improved capacity and speed of computers facilitated to treat such complex geophysical problems that could not have been solved before. The theoretical and practical issues of geophysical information processing or geophysical inversion have been increasing in the scientific literature, too. These researches apply the methods of optimization theory and implement them in special geophysical problems. In geophysical inverse problems often various geophysical methods are applied simultaneously, which allows to study several kinds of physical responses measured above the same geological structure. This method is called joint inversion. In this course we review the theoretical backgrounds of geophysical inverse problems, introduce some important inversion methods and show those special tools that leads to a reliable and accurate estimate of petrophysical and structural properties of geological structures.

The solution of the inverse problem (i.e. inverse modeling) requires the assumption of an initial model. A decision must be made both on the type (e.g. horizontally layered, dipped or tectonic structure) and the quantitative (petrophysical and geometrical) properties of the elements of the model. It is supported, on the one hand, by the a priori geological and geophysical knowledge about the investigated geological structure and, on the other hand, the computational (hardware and software) facilities. The latter is of great importance in calculating the theoretical responses (i.e. calculated data) over complex structures by using proper physical principles and different measurement arrays. The problem of data prediction is called forward modeling, in which data are calculated by known values of model parameters. These calculated data are then compared to real (field) data measured over the geological structure in the phase of inversion. The initial model is progressively refined in an iteration procedure until a proper fit is achieved between the predictions and observations. In an iteration step the model parameters are modified to improve the fit between the observed and calculated data. The data misfit is characterized by an objective function, which is minimized by a proper optimization algorithm. The iteration procedure is continued until a stop criteria is met. The estimated model obtained in the last iteration step is accepted as the solution of the inverse problem, which represents the most probable geological structure. One can encounter with this type of inversion methodology in mathematics, natural sciences and engineering, for instance in curve fitting, factor analysis, data and image processing, medical and geophysical tomography, astrophysics, electron microscopy etc. Geophysical applications may be rooted

in such methods that were invented in different scientific fields. Similarly, the results of geophysical inversion, for instance noise suppression or improvement of signal-to-noise ratio, can be exploited in other scientific fields, too.

Let us assume that the geological model can be characterized by finite number of discrete parameters. The description of model parameters by continuous functions is not excluded, because continuous parameter distributions can be resolved by finite number of parameters by using a proper discretization scheme. In formulating the discrete inverse problem, the column vector of *M* number of model parameters is introduced as

$$\vec{m} = \{m_1, m_2, \dots, m_M\}^T , \tag{I-1.1}$$

where *T* denotes the matrix transpose. Similarly, the *N* number of data measured by geophysical surveys are collected into the data vector $\vec{d}_m$

$$\vec{d}_m = \{d_1^{(m)}, d_2^{(m)}, \dots, d_N^{(m)}\}^T , \tag{I-1.2}$$

In inverse modeling it is assumed that there exist a procedure to calculate such data from the elements of model vector (I-1.1) to given places, which would be measured only when the structure is characterized exactly by model $\vec{m}$ and noiseless data. Let these theoretical data be sorted into the following *N*-dimensional vector

$$\vec{d}_c = \{d_1^{(c)}, d_2^{(c)}, \dots, d_N^{(c)}\}^T , \tag{I-1.3}$$

which is an element of the Euclidean data space. The index *c* refers to calculated values. According to our fundamental assumption, there is a connection between vectors $\vec{d}_c$ and $\vec{m}$

$$\vec{d}_c = \vec{g}(\vec{m}) , \tag{I-1.4}$$

where $\vec{g}$ represents a non-linear tensor function in the general case. In the terminology function (I-1.4) is called response function. The overall error between the measured and calculated data sets is expressed by the deviation vector

$$\vec{e} = \vec{d}_m - \vec{d}_e , \tag{I-1.5}$$

which is not null vector because the individual errors cannot be made to zero simultaneously. To characterize the misfit between the two data sets

$$E = E\left(\vec{d}_m - g(\vec{m})\right) \tag{I-1.6}$$

is introduced, which is a scalar independent of the sign of e_k . Function *E* can be defined in different ways. Since the observed data are always contaminated some amount of differently

distributed noise, it is of great importance that how the data noise is transformed into the model space. This question can be answered by the mathematical tools of probability theory (i.e. Bayesian inversion). On the other hand, the inverse problem can be solved by deterministic data processing, where the law of error propagation is specified. In geophysical inversion the latter is preferred.

The solution of the inverse problem can be given by seeking the minimum of objective function (I-1.6). There are several techiques to find the extremum of function *E*. In the case of few model parameters to be determined, the model space can be spaced equidistantly (or non-uniformly) where function *E* is evaluated in each point to determine the extremum directly. This approach is usually not feasible, because the forward modeling phase of the inversion procedure is rather time-consuming. When the number of unknowns is high, the use of this approach is hopeless. With Monte Carlo simulation-based techniques the number of sampling points in the model space can be reduced significantly. For instance, with the Sumulated Annelaing method, which is based on the straightforward calculation of function *E*, it is possible to find the global optimum of the objective function. This global optimum-searching strategy is advantageous to use, because the traditional (linear) optimum searching methods tend to allocate the solution of the inverse problem in some local minimum of the same function.

The second group of inversion techniques is formed by the so-called gradient-based methods. The CPU time can be significantly reduced by the linearization of the forward problem. Although the performance of computers is increasing very rapidly, which improves the application of global optimization methods, the linearized procedures are preferably used because of their speed. By linearization the response functions can be greatly simplified. Linearization causes also truncation errors, which is considered as a risky procedure from this point of view. In the following chapter, the most frequently used linear inversion methods are presented. The common starting point of these methods is the linearization of function connection (I-1.4).

1. The methods of linearized geophysical inversion

Linear inversion methods are based on the solution of sets of linear equations, which are relatively fast procedures. These prevailing methods are used for several geophysical problems. The common starting point of these methods is the linearization of function connection (I-1.4).

Cosider $\vec{m}_o$ as a starting point in the model space

$$\vec{m} = \vec{m}_o + \delta\vec{m}, \tag{I-1.7}$$

where $\delta\vec{m}$ is the model correction vector. Let connection (I-1.4) be approximated by its Taylor series truncated at the first order additive term

$$d_{e_k} = g_k(\vec{m}_o) + \sum_{j=1}^{M}\left(\frac{\partial g_k}{\partial m_j}\right)_{\vec{m}_o}\delta m_j \qquad (k = 1,2,...N). \tag{I-1.8}$$

By introducing the Jacobi's matrix

$$G_{kj} = \left(\frac{\partial g_k}{\partial m_j}\right)_{\vec{m}_o} \tag{I-1.9}$$

and $d_k^{(o)} = g_k(\vec{m}_o)$, after which equation (I-1.8) can be written as

$$d_{e_k} = d_k^{(o)} + \sum_{j=1}^{M} G_{kj}\delta m_j$$

or in vectorial form

$$\vec{d}_e = \vec{d}^{(o)} + \underline{\underline{G}}\,\delta\vec{m}.$$

By applying the notation $\delta\vec{d} = \vec{d}_e - \vec{d}^{(o)}$ it can be seen that $\delta\vec{d} = \underline{\underline{G}}\,\delta\vec{m}$ is the linearized form of equation (I-1.4). The deviation vector $\vec{e}$ based on equation (I-1.5) is

$$\vec{e} = \vec{d}_m - \vec{d}^{(o)} - \underline{\underline{G}}\,\delta\vec{m}, \tag{I-1.10}$$

which is also linear with respect to $\delta\vec{m}$. By the solution of the linearized inverse problem, the elements of vector $\delta\vec{m}$ can be derived from equation (I-1.7). In the subsequent iteration steps $\vec{m}$ is chosen as $\vec{m}_o$ and the response function is linearized again. The iteration procedure is continued until a stop criteria is met. Thus, the nonlinear inverse problem is reduced to a solution of a set of linear inverse problems. For the sake of simplicity the vector $\vec{d}^m$ can be

used instead of $\vec{d}_m - \vec{d}^{(o)}$, and one can apply the substitution $\delta\vec{m} \to \vec{m}$. The deviation vector of the linearized inverse problem is

$$\vec{e} = \vec{d}^{\,m} - \underline{\underline{G}}\vec{m}\,. \tag{I-1.11}$$

When the number of data equals to that of the unknowns, equation (I-1.11) leads to the following inhomogeneous linear (algebraic) set of equations

$$\vec{d}^{\,m} = \underline{\underline{G}}\,\vec{m}$$

with a unique solution (when *M* is the rank of matrix $\underline{\underline{G}}$)

$$\vec{m} = \underline{\underline{G}}^{-1}\vec{d}^{\,m}\,, \tag{I-1.12}$$

where $\underline{\underline{G}}^{-1}$ represents the inverse matrix.

When the number of data is less that that of the unknowns, an underdetermined inverse problem is to be solved. Later, it will be demonstrated that in the underdetermined case the vector (I-1.11) is exacly zero and the inverse problem has infinite number of solutions. If the number of data exceeds that of the unknowns, the inverse problem is called overdetermined. In geophysical inversion there are even-determined problems in small numbers, when there is only one solution, and it has zero prediction error. In the followings, we study the solution of the under- and overdetermined inverse problems, then we discuss the so-called mixed-determined problem which shows some features of both types.

1.1 The overdetermined inverse problem

The overdetermined inverse problem leads to an inconsistent set of equations when the measurement data are charged with noise. There is no solution for this problem in the algebraic sense. When the observed data equal to the calculated ones, in equation $\underline{\underline{G}}\vec{m} = \vec{d}^{\,m}$ the *N*-by-*M* size Jacobi's matrix cannot be inverted.

The model vector estimated by the minimization of the scalar (I-1.6) is accepted as a solution of the overdetermined inverse problem. The choice of the type of the scalar is manifold. From the point of view of history the most important type is the L_2-norm of vector (I-1.5). The procedure, which minimizes it, is called the Gaussian Least Squares method.

1.1.1. The Gaussian Least Squares (LSQ) method

The objective function to be minimized is the squared L_2-norm of the deviation vector characterizing the misfit between the calculated and observed data

$$E = \vec{e}^T \vec{e} = \sum_{k=1}^{N} e_k^2 = \sum_{k=1}^{N} \left(d_k - \sum_{j=1}^{M} G_{kj} m_j \right) \left(d_k - \sum_{i=1}^{M} G_{ki} m_i \right),$$

which has optimum if the following set of equations fullfills for each $l = 1,2,\ldots,M$

$$\frac{\partial E}{\partial m_l} = 0 \,. \tag{I-1.13}$$

(Hereafter the index m in vector $\vec{d}^m$ is neglected, because the calculated data are given always by the linear term $\underline{\underline{G}}\vec{m}$.) The objective function $E = \vec{e}^T \vec{e}$ can be written in more details as the sum of three terms

$$E = \sum_{k}^{N} d_k^2 + \sum_{i=1}^{M} \sum_{j=1}^{M} m_i m_j \left(\sum_{k=1}^{N} G_{kj} G_{ki} \right) - 2 \sum_{j=1}^{M} m_j \left(\sum_{k=1}^{N} G_{kj} d_k \right). \tag{I-1.14}$$

The derivative of the first term with respect to m_l is zero, while that of the second term is

$$\frac{\partial}{\partial m_l} \left(\sum_{i=1}^{M} \sum_{j=1}^{M} m_i m_j \sum_{k=1}^{N} G_{ki} G_{kj} \right) = \sum_{i=1}^{M} \sum_{j=1}^{M} \left(m_i \delta_{jl} + m_j \delta_{ij} \right) \sum_{k=1}^{N} G_{ki} G_{kj} \,, \tag{I-1.15}$$

where

$$\delta_{jl} = \frac{\partial m_j}{\partial m_l} = \begin{cases} 1 & if \quad j = 1 \\ 0 & if \quad j \neq 1 \end{cases} \tag{I-1.16}$$

denotes the symbol of Kronecker Delta. The first term on the right of equation (I-1.15) can be written in the following form because of (I-1.16)

$$\sum_{i=1}^{M} \sum_{j=1}^{M} m_i \delta_{jl} \sum_{k=1}^{N} G_{ki} G_{kj} = \sum_{i=1}^{M} m_i \sum_{k=1}^{N} G_{ki} G_{kl} \,,$$

which gives also the second term when indices j and i are exchanged, that is

$$\frac{\partial}{\partial m_l} \left(\sum_{i=1}^{M} \sum_{j=1}^{M} m_i m_j \sum_{k=1}^{N} G_{ki} G_{kj} \right) = 2 \sum_{i=1}^{M} m_i \sum_{k=1}^{N} G_{ki} G_{kl} \,. \tag{I-1.17}$$

The third term on the right of equation (I-1.14) can be expressed as

$$\frac{\partial}{\partial m_l}\left(\sum_{j=1}^{M} m_j \sum_{k=1}^{N} G_{kj} d_k\right) = \sum_{j=1}^{M} \delta_{jl} \sum_{k=1}^{N} G_{kj} d_k = \sum_{k=1}^{N} G_{kl} d_k \,,$$

which leads to the

$$\sum_{i=1}^{M} m_i \sum_{k=1}^{N} G_{ki} G_{kl} = \sum_{k=1}^{N} G_{kl} d_k \qquad \text{(I-1.18)}$$

normal equation, which in the vectorial form is

$$\underline{\underline{G}}^T \underline{\underline{G}} \vec{m} = \underline{\underline{G}}^T \vec{d} \,. \qquad \text{(I-1.19)}$$

By minimizing the L_2-norm of deviation vector (I-1.11) an inhomogeneous set of linear equations is obtained, which can be solved numerically. The solution can be extracted by multiplying both sides of equation (I-1.19) with $(\underline{\underline{G}}^T \underline{\underline{G}})^{-1}$

$$\vec{m} = (\underline{\underline{G}}^T \underline{\underline{G}})^{-1} \underline{\underline{G}}^T \vec{d} \,.$$

It can be seen that by introducing the matrix

$$G^{-g} = (\underline{\underline{G}}^T \underline{\underline{G}})^{-1} \underline{\underline{G}}^T$$

the solution can be written in the following form

$$\vec{m} = \underline{\underline{G}}^{-g} \vec{d} \,.$$

Matrix $\underline{\underline{G}}^{-g}$ given in the above term plays a similar role that $\underline{\underline{G}}^{-1}$ does in equation (I-1.12). In case of the overdetermined problem matrix $\underline{\underline{G}}$ cannot be inverted in equation $\underline{\underline{G}}\vec{m} = \vec{d}^{\,m}$, but the model vector can be estimated by calculating matrix $\underline{\underline{G}}^{-g}$. The latter is called the generalized inverse of the Gaussian Least Squares method.

The condition number is an important property of matrix $(\underline{\underline{G}}^T \underline{\underline{G}})$

$$k = \frac{\lambda_{\max}}{\lambda_{\min}} \,,$$

where $\lambda_{\max}$ and $\lambda_{\min}$ denote the minimal and maximal eigenvalue, respectively. The set of equations with a high condition number (e.g. $k = 10^5$) indicates an ill-posed problem. In this case small differences in the Jacobi's matrix produce much bigger ones in the estimated parameters, which leads to an unstable inversion procedure. To avoid numerical instability the improvement of condition number (i.e. decrease of k) is of high practical importance.

1.1.2. The Weighted Least Squares (WLSQ) method

It is often encountered that the uncertainties of observed data are of different amount. In these cases care must be taken that a datum should contribute to the solution with a given weight proportional to its uncertainty. It is carred out by the application of a symmetric weighting matrix $\underline{\underline{W}}$, which includes the weighs of the data in its main diagonal. The solution is developed by the minimization of the following objective function

$$E = \vec{e}^T \underline{\underline{W}} \vec{e} = \sum_{k=1}^{N} \left(d_k - \sum_{i=1}^{M} G_{ki} m_i \right) \sum_{r=1}^{N} W_{kr} \left(d_r - \sum_{j=1}^{M} G_{rj} m_j \right),$$

which is the weighted norm of the deviation vector (I-1.11). With the fulfillment of the condition

$$\frac{\partial E}{\partial m_l} = 2 \sum_{i=1}^{M} m_i \sum_{k=1}^{N} \sum_{r=1}^{N} W_{kr} G_{ki} G_{rl} - 2 \sum_{k=1}^{N} d_k \sum_{r=1}^{N} W_{kr} G_{rl} = 0,$$

the following normal equation is obtained

$$\underline{\underline{G}}^T \underline{\underline{W}} \underline{\underline{G}} \vec{m} = \underline{\underline{G}}^T \underline{\underline{W}} \vec{d}. \qquad \text{(I-1.20)}$$

The solution of the weighted least squares method is

$$\vec{m} = (\underline{\underline{G}}^T \underline{\underline{W}} \underline{\underline{G}})^{-1} \underline{\underline{G}} \underline{\underline{W}} \vec{d}.$$

By introducing the generalized inverse

$$\underline{\underline{G}}^{-g} = (\underline{\underline{G}}^T \underline{\underline{W}} \underline{\underline{G}})^{-1} \underline{\underline{G}} \underline{\underline{W}},$$

the estimated model vector can be written as

$$\vec{m} = \underline{\underline{G}}^{-g} \vec{d}.$$

1.1.3. The Iteratively Reweighted Least Squares (IRLS) method

The inversion method minimizing the L_2-norm of the deviation vector can be generalized to arbitrary p

$$L_p = \left[\sum_{k=1}^{N} |e_k|^p \right]^{1/p}, \qquad \text{(I-1.21)}$$

which is called the L_p-norm of the same vector. In the overdetermined case it is sufficient to minimize the following objective function

$$E_p = \sum_{k=1}^{N} \left| d_k - \sum_{j=1}^{M} G_{kj} m_j \right|^p ,$$

that is to solve the set of equations

$$\frac{\partial E_p}{\partial m_l} = 0 , \qquad l = 1,2,\ldots,M. \tag{I-1.22}$$

Having completed the above operations the following can be obtained

$$\frac{\partial}{\partial m_l} \sum_{k=1}^{N} \left| d_k - \sum_{j=1}^{M} G_{kj} m_j \right|^p = -p \sum_{k=1}^{N} \left| e_k \right|^{p-1} sign(e_k) \sum_{j=1}^{M} G_{kj} \delta_{jl} ,$$

where

$$sign(e_k) = \frac{e_k}{|e_k|}$$

gives the sign of the k-th element of the deviation vector. After some modifications equation (I-1.22) leads to

$$\sum_{k=1}^{N} \left| e_k \right|^{p-2} \left(d_k - \sum_{i=1}^{M} G_{ki} m_i \right) G_{kl} = 0 .$$

By introducing the diagonal weighting matrix

$$W_{ks} = \left| e_k \right|^{p-2} \delta_{ks} \tag{I-1.23}$$

the above equation takes the form as

$$\sum_{i=1}^{M} m_i \sum_{k=1}^{N} G_{ki} \sum_{s=1}^{N} W_{ks} G_{sl} = \sum_{s=1}^{N} G_{sl} \sum_{k} W_{ks} d_k .$$

The above equation can be written in matrix form

$$\underline{\underline{G}}^T \underline{\underline{W}} \underline{\underline{G}} \vec{m} = \underline{\underline{G}}^T \underline{\underline{W}} \vec{d} , \tag{I-1.24}$$

which formally corresponds to (I-1.20). Since the diagonal of weigth matrix containing the unknowns m_i is

$$W_{kk} = \left| d_k - \sum_{i=1}^{M} G_{ki} m_i \right|^{p-2} ,$$

which forms a nonlinear set of equatons (I-1.24) when $p \neq 2$. For the case of $p = 2$ the weights are placed in an identity matrix, thus equation (I-1.24) reduces to equation (I-1.19)

$$\underline{\underline{G}}^T \underline{\underline{G}} \vec{m} = \underline{\underline{G}}^T \vec{d} , \tag{I-1.25}$$

which is the normal equation of the Gaussian LSQ method. The solutions of equation (I-1.24) for the cases of $p<2$ are of practical importance in geophysical applications. The weigths are inversely proportional to the deviations between the calculated and observed data. As a consequence, the noisier data contribute to the solution in a less degree.

The nonlinear set of equations is treated by the so-called iteratively reweighted least squares method, which assures that a linear set of equations is solved in each step of the iterative procedure. In the first step the weighting matrix is chosen as an identity matrix (p=2) and the model vector is estimated from (I-1.25)

$$\vec{m}^{(1)} = \left(\underline{\underline{G}}^T \underline{\underline{G}}\right)^{-1} \underline{\underline{G}}^T \vec{d} \,.$$

Let us generate the deviation vector from the above estimate as

$$\vec{e}^{(1)} = \vec{d}^{\,m} - \underline{\underline{G}}\,\vec{m}^{(1)} \qquad \text{(I-1.26)}$$

and weighting matrix $\underline{\underline{W}}^{(1)}$ according to (I-1.23). In the second step of the iteration procedure equation (I-1.24) is solved

$$\underline{\underline{G}}^T \underline{\underline{W}}^{(1)} \vec{m}^{(2)} = \underline{\underline{G}}^T \underline{\underline{W}}^{(1)} \vec{d} \,. \qquad \text{(I-1.27)}$$

Since weigth matrix $\underline{\underline{W}}^{(1)}$ based on equation (I-1.26) is connected to $\vec{m}^{(1)}$ and model vector $\vec{m}^{(2)}$ is unknown, thus, set of equations (I-1.27) is linear. The solution of the latter is

$$\vec{m}^{(2)} = \left(\underline{\underline{G}}^T \underline{\underline{W}}^{(1)} \underline{\underline{G}}\right)^{-1} \underline{\underline{G}}^T \underline{\underline{W}}^{(1)} \vec{d} \,.$$

The above procedure in the j-th iteration step leads to the nonlinear equation

$$\underline{\underline{G}}^T \underline{\underline{W}}^{(j-1)} \underline{\underline{G}} \vec{m}^{(j)} = \underline{\underline{G}}^T \underline{\underline{W}}^{(j-1)} \vec{d} \,. \qquad \text{(I-1.28)}$$

The actual weighting matrix is recalculated in each iteration step, therefore the procedure is called Iteratively Reweighted Least Squares (IRLS) method.

Among the inversion methods based on the minimization of L_p-norm the most frequently used one is represented by the case of p=1, which is called Least Absolute Deviations (LAD) method. The k-th diagonal element of weighting matrix in equation (I-1.28) for the case of p=1 is

$$W_{kk}^{(j-1)} = \frac{1}{\left|e_k^{(j-1)}\right|} \,.$$

It can be seen that the bigger the difference beween the predicted and observed data was obtained in the previous step, the less weight the data are given. Therefore, it is expected that

the L_1-norm based LAD method does suppress the effect of rarely appearing outliers in the data set. This is the most important advantage of the LAD technique against the traditional Gaussian LSQ method.

1.2. The underdetermined inverse problem

In a purely underdetermined inverse problem the number of data (equations) is less than-that of the unknowns. It might be thought that they can be avoided in geophysical applications, because we have means to measure enough data to set an overdetermined problem. Sometimes these conditions are not fulfilled, on the other hand there are such procedures (e.g. seismic tomography) that approximates the solution of an overdetermined inverse problem by solutions of a set of underdetermined problems. The partial underdetermination is also frequent in geophysical inversion.

By restricting ourselves to the linearized inverse problem

$$\vec{d} = \underline{\underline{G}}\,\vec{m}\,, \tag{I-1.29}$$

there is an infinite number of solutions of the underdetermined inverse problem. Unique solution can only be obtained when constraint relations (i.e. additional equations) are used, too. A simple solution can be given when the Euclidean norm of the model vector is minimal

$$L = \vec{m}^T \vec{m} = \sum_{i=1}^{M} m_i^2 = \text{ min}, \tag{I-1.30}$$

while the deviation vector

$$\vec{e} = \vec{d} - \underline{\underline{G}}\vec{m}$$

vanishes. To fulfill both requirements a solution can be worked out based on the method of the Lagrange multipliers. The objective function to be minimized is

$$\phi(\vec{m}) = L + \sum_{k=1}^{N} \lambda_k e_k \tag{I-1.31}$$

where λ_k is the k-th multiplier. The extreme value of equation (I-1.31) can be determined by the following equation

$$\frac{\partial \phi}{\partial m_l} = 2m_l - \sum_{k=1}^{N} \lambda_k G_{kl} = 0\,,$$

which in vectorial form is

$$2\vec{m} = \underline{\underline{G}}^T \vec{\lambda}. \tag{I-1.32}$$

Therewith, equation (I-1.29) must be satisfied, too. By combining the two equations the data vector is

$$\vec{d} = \frac{1}{2}\underline{\underline{G}}\,\underline{\underline{G}}^T\vec{\lambda}\ ,$$

whence the vector of multipliers can be derived as

$$\vec{\lambda} = 2\left(\underline{\underline{G}}\,\underline{\underline{G}}^T\right)^{-1}\vec{d}\ .$$

The above vector can be substituted into equation (I-1.32), which gives the simple solution of the underdetermined inverse problem

$$\vec{m} = \underline{\underline{G}}^T\left(\underline{\underline{G}}\,\underline{\underline{G}}^T\right)^{-1}\vec{d}\ . \qquad \text{(I-1.33)}$$

The estimated model vector can be defined also in this manner

$$\vec{m} = \underline{\underline{G}}^{-g}\vec{d}\ ,$$

where

$$\underline{\underline{G}}^{-g} = \underline{\underline{G}}^T\left(\underline{\underline{G}}\,\underline{\underline{G}}^T\right)^{-1}$$

is the generalized inverse matrix of the inverse problem.

1.2.1. The principles of generalized simple solution

The simple solution always satisfies the requirement of uniqueness. However, the additional conditions are advantageous to be considered variedly. The most frequently used method is the application of a symmetric weighting matrix in the model space. To define the condition of simplicity of the solution more generally, the following weighted norm is introduced

$$L^* = (\vec{m}, \underline{\underline{W}}\vec{m}) = \sum_{i=1}^{M} m_i \sum_{j=1}^{M} W_{ij} m_j = \text{ min.}$$

Since the inverse problem is purely underdetermined, it holds true that the deviation vector

$$\vec{e} = \vec{d} - \underline{\underline{G}}\vec{m}$$

vanishes. To fulfill both requirements a the following objective function is minimized

$$\phi(\vec{m}) = L^* + \sum_{k=1}^{N} \lambda_k e_k\ ,$$

where λ_k is the k-th multiplier. The extreme value can be calculated by specifying the following condition

$$\frac{\partial \phi}{\partial m_l} = 2\sum_{j=1}^{M} W_{lj} m_j - \sum_{k=1}^{N} \lambda_k G_{kl} = 0 ,$$

which in vectorial form is

$$2\underline{\underline{W}}\vec{m} = \underline{\underline{G}}^T \vec{\lambda} .$$

The equation of the forward problem (I-1.29) must also be fulfilled. By combining the two equations the data vector is

$$\vec{d} = \frac{1}{2}\underline{\underline{G}}\,\underline{\underline{W}}^{-1}\underline{\underline{G}}^T \vec{\lambda} ,$$

whence the vector of multipliers can be derived as

$$\vec{\lambda} = 2\left(\underline{\underline{G}}\,\underline{\underline{W}}^{-1}\underline{\underline{G}}^T\right)^{-1}\vec{d} .$$

The generalized simple solution of the underdetermined inverse problem is

$$\vec{m} = \underline{\underline{W}}^{-1}\underline{\underline{G}}^T\left(\underline{\underline{G}}\,\underline{\underline{W}}^{-1}\underline{\underline{G}}^T\right)^{-1}\vec{d} ,$$

which can be simplified to

$$\vec{m} = \underline{\underline{G}}^{-g}\vec{d}$$

with the generalized inverse of the problem

$$\underline{\underline{G}}^{-g} = \underline{\underline{W}}^{-1}\underline{\underline{G}}^T\left(\underline{\underline{G}}\,\underline{\underline{W}}^{-1}\underline{\underline{G}}^T\right)^{-1},$$

where $\underline{\underline{W}}$ is the weighting matrix defined in the model space.

1.3 The mixed-determined inverse problem

In the geophysical practice the number of measurement data can be increased as occasion requires. This circumstance does not result in automatically an overdetermined inverse problem. It is a frequent occurrence that we have a forward problem including some combinations (e.g. products) of model parameters. In that case only the combinations can be determined, while the model parameters cannot be resolved individually. If the goal is to determine each model parameter separately, which are coupled in the forward problem, the inverse problem will be underdetermined. It is unhelpful to increase the number of data (of the same kind). In

this case the inverse problem is mixed-determined (partly overdetermined and underdetermined in some measure).

1.3.1. Solution by minimizing the L_2-norm

There are infinite number of solutions of the mixed-determined inverse problem because of the partial underdetermination. The problem is partly overdetermined for some part of the unknowns at the same time, thus the deviation vector cannot be zero. In case of linear overdetermined inverse problems a solution can be given by minimizing the L_2-norm

$$\left(E = \sum_{k=1}^{N} e_k^2 \right).$$

Because of the underdetermined character of the same problem, one can be chosen from the infinite number of solutions by specifying

$$L = \sum_{j=1}^{M} m_j^2 = \text{min.}$$

The solution of the mixed-determined problem can be given by minimizing the following combined objective function

$$\Phi = E + \varepsilon^2 L ,$$

which in more detail is

$$\phi = \sum_{k=1}^{M} \left(d_k - \sum_{j=1}^{M} G_{kj} m_j \right) \left(d_k - \sum_{i=1}^{M} G_{ki} m_i \right) + \varepsilon^2 \sum_{j=1}^{M} m_j^2 = \text{min.}$$

In the above formula constant ε^2 is normally a small positive number, which can be used to formulate the ratio of the rate between over- and underdetermination. The extremum is defned by the following equations

$$\frac{\partial \phi}{\partial m_l} = 0 , \qquad (l = 1,2,\ldots,M),$$

that is

$$\frac{\partial \phi}{\partial m_l} = 2 \sum_{i=1}^{M} m_i \sum_{k=1}^{N} G_{ki} G_{kl} - 2 \sum_{k=1}^{N} G_{kl} d_k + \varepsilon^2 \sum_{j=1}^{M} 2 m_j \delta_{jl} = 0 .$$

The above equation can be written in this manner

$$\sum_{i=1}^{M} m_i \left(\sum_{k=1}^{N} G_{ki} G_{kl} + \varepsilon^2 \delta_{il} \right) = \sum_{k=1}^{N} G_{kl} d_k ,$$

or in vectorial form

$$\left(\underline{\underline{G}}^T \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}}\right)\vec{m} = \underline{\underline{G}}^T \vec{d}\ , \qquad \text{(I-1.34)}$$

where $\underline{\underline{I}}$ is the identity matrix. It is the normal set of equations of the mixed-determined inverse problem. The name of the method is mentioned in the literature variously, i.e. the Damped Least Squares (DLSQ) method, Levenberg-Marquardt method, Marquardt algorithm or Tikhonov's regularization procedure. The solution of equation (I-1.34) is

$$\vec{m} = \left(\underline{\underline{G}}^T \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}}\right)^{-1} \underline{\underline{G}}^T \vec{d}\ ,$$

which can be written also in the form of

$$\vec{m} = \underline{\underline{G}}^{-g} \vec{d}\ ,$$

where

$$\underline{\underline{G}}^{-g} = \left(\underline{\underline{G}}^T \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}}\right)^{-1} \underline{\underline{G}}^T$$

is the generalized inverse matrix of the problem.

A stable and convergent solution can be obtained by the proper selection of constant ε^2, which requires only small modifications in the program code of the LSQ method. It is seen that equation (I-1.34) differs from equation (I-1.19) only that the values of constant ε^2 are added to the elements of the main diagonal of matrix $\left(\underline{\underline{G}}^T \underline{\underline{G}}\right)^{-1}$. This operation results in the increase of the eigenvalues of the previous matrix. If the constant is chosen properly, the condition number of the matrix can be reduced by several orders of magnitude at good. By the same procedure one can transform an ill-conditioned set of equations to a better one. This is the numerical background of the efficiency of the DLSQ method.

1.3.2. Solution by the weighted least squares method

a.) Solution of the mixed-determined problem by the least squares method weighted in the data space

Let $\underline{\underline{W}}$ be a known weighting matrix defined in the data space. Similar to the derivation in Chapter 1.1.2, a weighted solution to the mixed-determined inverse problem can be derived. The weighted least squares method should be combined with the principles of the simple solution. The objective function to be minimized is

$$\Phi = \sum_{k=1}^{N}\left(d_k - \sum_{i=1}^{M} G_{ki} m_i \right)\sum_{r=1}^{N} W_{kr}\left(d_r - \sum_{j=1}^{M} G_{rj} m_j \right) + \varepsilon^2 \sum_{q=1}^{M} m_q^2 \ .$$

The solution is seeked by the following set of equations

$$\frac{\partial \Phi}{\partial m_l} = 0 \qquad (l = 1,2,\ldots,M).$$

The result of derivation without details is

$$\left(\underline{\underline{G}}^T \underline{\underline{W}}\, \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}}\right)\vec{m} = \underline{\underline{G}}^T \underline{\underline{W}}\vec{d} \ , \qquad \text{(I-1.35)}$$

of which solution is

$$\vec{m} = \left(\underline{\underline{G}}^T \underline{\underline{W}}\, \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}}\right)\underline{\underline{G}}^T \underline{\underline{W}}\vec{d} \ .$$

The generalized inverse matrix of the least squares method weighted in the data space is introduced as

$$\underline{\underline{G}}^{-g} = \left(\underline{\underline{G}}^T \underline{\underline{W}}\, \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}}\right)\underline{\underline{G}}^T \underline{\underline{W}} \ .$$

b.) Solution of the mixed-determined problem by the least squares method weighted in the data and model spaces

Let $\underline{\underline{W}}$ be a weighting matrix defined in the data space and $\underline{\underline{W}}^*$ be another one given in the model space. A weighted solution to the mixed-determined inverse problem can be given based on the generalized simple solution (Chapter 1.2.1.). The objective function to be minimized is

$$\Phi = \sum_{k=1}^{N}\left(d_k - \sum_{i=1}^{M} G_{ki} m_i \right)\sum_{r=1}^{N} W_{kr}\left(d_r - \sum_{j=1}^{M} G_{rj} m_j \right) + \varepsilon^2 \sum_{q=1}^{M} m_q \sum_{p=1}^{M} W_{qp}^{*}\, m_p \ .$$

The solution is seeked by the following set of equations

$$\frac{\partial \Phi}{\partial m_l} = 0 \qquad (l = 1,2,\ldots,M).$$

The result of derivation without details is

$$\left(\underline{\underline{G}}^T \underline{\underline{W}}\, \underline{\underline{G}} + \varepsilon^2 \underline{\underline{W}}^*\right)\vec{m} = \underline{\underline{G}}^T \underline{\underline{W}}\vec{d}$$

of which solution is

$$\vec{m} = \left(\underline{\underline{G}}^T \underline{\underline{W}}\underline{\underline{G}} + \varepsilon^2 \underline{\underline{W}}^*\right)\underline{\underline{G}}^T \underline{\underline{W}}\vec{d} \ .$$

The generalized inverse matrix of the damped least squares method weighted both in the data and model space is introduced as

$$\underline{\underline{G}}^{-g} = \left(\underline{\underline{G}}^T \underline{\underline{W}} \underline{\underline{G}} + \varepsilon^2 \underline{\underline{W}}^*\right) \underline{\underline{G}}^T \underline{\underline{W}} .$$

1.3.3. Solution by minimizing the L_p-norm

The solution of the mixed-determined problem can be given by the minimization of the L_p-norm of the deviation vector because of the partial overdetermination of the inverse problem. The objective function

$$\Phi = \sum_{k=1}^{N} \left| d_k - \sum_{j=1}^{M} G_{kj} m_j \right|^p + \varepsilon^2 \sum_{q=1}^{M} m_q^2$$

has an extreme value if the following set of conditions

$$\frac{\partial \Phi}{\partial m_l} = 0 \qquad (l = 1,2,\ldots,M)$$

is satisfied. Similarly to the procedure established in Chapter 1.1.3 a nonlinear set of equations can be derived

$$\left(\underline{\underline{G}}^T \underline{\underline{W}} \underline{\underline{G}} + \frac{2\varepsilon^2}{p} \underline{\underline{I}} \right) \vec{m} = \underline{\underline{G}}^T \underline{\underline{W}} \vec{d} , \qquad \text{(I-1.36)}$$

where the diagonal elements of the weighting matrix are

$$W_{kk} = \left| d_k - \sum_{i=1}^{M} G_{ki} m_i \right|^{p-2} .$$

Since the above weighting matrix contains the elements of the model vector, the set of equations (I-1.36) can be solved by the iteratively reweighted least squares method. In the first step of the procedure the following equation is solved (p=2)

$$\left(\underline{\underline{G}}^T \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}} \right) \vec{m}^{(1)} = \underline{\underline{G}}^T \vec{d}$$

and an initial weighting matrix is calculated. In the j-th iteration step a linear equation is solved

$$\left(\underline{\underline{G}}^T \underline{\underline{W}}^{(j-1)} \underline{\underline{G}} + \frac{\varepsilon^2}{p} \underline{\underline{I}} \right) \vec{m}^{(1)} = \underline{\underline{G}}^T \underline{\underline{W}}^{(j-1)} \vec{d} , \qquad \text{(I-1.37)}$$

which approximates the solution of the original nonlinear equation (I-1.36). In geophysical applications the cases of p<2 are advantageous to be used to reduce the strong effect of outlying data in the solution.

1.4. Quantities characterizing the quality of parameter estimation

The measurement data are always contaminated by some amount of noise. The measuring errors can be systematic or random. The former can be generally treated well, while the calculation of the latter is of great importance in the practice of geophysical inversion. The general equation

$$\vec{m} = \underline{\underline{G}}^{-g}\vec{d}$$

shows the transformation of data to the model space. According to that the data noise also will be mapped into the model space. The uncertainty of data causes the appearance of errors in parameter estimation. The most importan question is how the estimation errors can be quantified in the inversion procedure.

1.4.1. Data and model distances

For measuring the misfit between the measured and calculated data the relaive data distance is normally used

$$D_a = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{d_k^{(m)} - d_k^{(c)}}{d_k^{(m)}}\right)^2} * 100(\%),$$

where N is the number of inverted data, $d_k^{(m)}$, $d_k^{(c)}$ are the k-th measured and calculated data, respectively.

A new inversion method should be tried at first on synthetic data sets. The aim of the inversion of synthetic data is to test the performance of the inversion procedure. The experiment assumes that the geophysical model is exactly known. Synthetic data are calculated on the known model, which are then contaminated with different amounts of noise to simulate real measurements. A noisy data set provides as input for the inversion procedure. It can be studied how accurately and reliably the inversion parameters have been recovered. The relative model distance characterizes the goodness of the estimated model in cases of synthetic inversion experiments

$$D_m = \sqrt{\frac{1}{M}\sum_{i=1}^{N}\left(\frac{m_i^{(estimated)} - m_i^{(exact)}}{m_i^{(estimated)}}\right)^2} * 100(\%),$$

where M is the number of model parameters, $m_i^{(estimated)}$, $m_i^{(exact)}$ are the i-th estimated and exacly known model parameter, respectively. As it can be seen in the formalae, the model and data distances are normalized quantities. It is because that the data or model are of different orders of magnitude and measurement units. The normalization is necessary in case of joint inversion procedures.

1.4.2. The estimation errors

Random noises can be described by the tools of probability theory and mathematical statistics. The expected value of a continuous random variable x can be written with the knowledge of its probability density function $f(x)$

$$\hat{x} = \int_{-\infty}^{\infty} x f(x) dx,$$

and its variance is

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \hat{x})^2 f(x) dx.$$

When $k=1,...,N$ number of data is collected and function $f(x)$ depends on each data, the expected value and variance of the k-th data is

$$\hat{d}_k = \int_{1\cdots N}\int d_k f(d_1,\ldots,d_k,\ldots,d_N) d(d_1)\ldots d(d_N),$$

$$\sigma_k^2 = \int_{1\cdots N}\int (d_k - \hat{d}_k)^2 f(d_1,\ldots,d_k,\ldots,d_N) d(d_1)\ldots d(d_N),$$

where $dd_1\ldots dd_N$ denotes the infinitely small variation of data $d_1\ldots d_N$. The correlation between the variables is also important in geophical applications. In case of two data variables (d_k and d_s where $k \neq s$) the index number of co-variation is called covariance

$$cov_{ks}^{(d)} = \int_{1\cdots N}\int (d_k - \hat{d}_k)(d_s - \hat{d}_s) \cdot f(d_1,\ldots,d_N) d(d_1)\ldots d(d_N).$$

It can be seen that covariance equals to the variance of the k-th datum when $k=s$. The array including the covariances of N number of data is called the data covariance matrix. The N-by-

N size symmetric matrix is not diagonal when the data are correlated. The data covariance matrix can be written in a simpler form as

$$\mathrm{cov}_{ks}^{(d)} = \overline{(d_k - \hat{d}_k)(d_s - \hat{d}_s)}\,, \tag{I-1.38}$$

where the overline represents the expected value of the underneath quantity calculated by an integral. For the characterization of the strength of correlation between data the correlation matrix is used

$$corr_{ks}^{(d)} = \frac{\mathrm{cov}_{ks}^{(d)}}{\sqrt{\mathrm{cov}_{kk}^{(d)}\,\mathrm{cov}_{ss}^{(d)}}}\,.$$

In the model space a similar covariance matrix defined between the model parameters can be introduced

$$\mathrm{cov}_{ij}^{(m)} = \overline{(m_i - \hat{m}_i)(m_j - \hat{m}_j)}\,. \tag{I-1.39}$$

In the previuos chapters we defined the generalized inverse matrix and the solution of the inverse problem as $\vec{m} = \underline{\underline{G}}^{-g}\vec{d}$. The latter and its expected value can be represented in coordinates: $m_i = \sum_{k=1}^{N} G_{ik}^{-g} d_k$ and $\hat{m}_i = \sum_{k=1}^{N} G_{ik}^{-g}\hat{d}_k$. With these

$$m_i - \hat{m}_i = \sum_{k=1}^{N} G_{ik}^{-g}(d_k - \hat{d}_k)\,,$$

and

$$(m_i - \hat{m}_i)(m_j - \hat{m}_j) = \sum_{k=1}^{N}\sum_{s=1}^{N} G_{ik}^{-g}(d_k - \hat{d}_k)(d_s - \hat{d}_s)G_{js}^{-g}\,.$$

After the expected value have been calculated

$$\overline{(m_i - \hat{m}_i)(m_j - \hat{m}_j)} = \sum_{k=1}^{N}\sum_{s=1}^{N} G_{ik}^{-g}\,\overline{(d_k - \hat{d}_k)((d_s - \hat{d}_s)}\,G_{js}^{-g}\,,$$

which in other form is

$$\mathrm{cov}_{ij}^{(m)} = \sum_{k=1}^{N}\sum_{s=1}^{N} G_{ik}^{-g}\,\mathrm{cov}_{ks}^{(d)} G_{js}^{-g}\,. \tag{I-1.40}$$

Equation (I-1.40) gives a connection between the data and model covariance matrices in case of any linear inversion method, which is applicable to calculate the parameter estimation errors in the knowledge of data uncertanties. The matrix form of the equation is (Menke, 1984)

$$\underline{\underline{\mathrm{cov}}}^{(m)} = \underline{\underline{G}}^{-g}\,\underline{\underline{\mathrm{cov}}}^{(d)}(\underline{\underline{G}}^{-g})^T\,. \tag{I-1.41}$$

The estimation error of the *i*-th model parameter is derived from the *i*-th element of the main diagonal of the model covariance matrix

$$\sigma_i^{(m)} = \sqrt{\mathrm{cov}_{ii}^{(m)}} \qquad (i = 1,2,\ldots,M).$$

The inversion methods result in typically correlated parameter estmation. To quantify the strength of correlation between the model parameters the correlation matrix is used

$$corr_{ij}^{(m)} = \frac{\mathrm{cov}_{ij}^{(m)}}{\sqrt{\mathrm{cov}_{ii}^{(m)}\,\mathrm{cov}_{jj}^{(m)}}}. \qquad \text{(I-1.42)}$$

The correlation matrix characterizes the reliability of the inversion estimate to the effect that strongly correlated model parameters normally does not yield stable and acceptable accurate parameter estimation. The diagonal elements of $corr_{ij}^{(m)}$ are always 1. Outside the main diagonal the correlation between the *i*-th and *j*-th model parameters is indicated by a number in the range of -1 and 1. If the correlation coefficient is near to ±1, the solution of the inverse problem is unreliable. Reliable solution is indicated by a value between 0 and ±0.5. The correlation matrix can be characterized by one scalar

$$S = \sqrt{\frac{1}{M(M\text{-}1)}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(corr_{ij}^{(m)} - \delta_{ij}\right)^2},$$

which is called the mean spread (the symbol of Kronecker Delta). The range of this parameter is *[0;+1]*, which is also a measure of reliability.

References

Menke, W. 1984. Geophysical Data Analysis: Discrete Inverse Theory. Academic Press Inc.

II. The theory of global inversion methods

Linearized inversion methods are used to estimate the parameters of the model by minimizing the norm of the deviation vector characterizing the misfit between the measured and calculated data. These methods do not guarantee to find the absolute minimum of the objective function as they tend to assign the solution to some local minimum. This problem requires to apply such methods that can seek the global minimum of the objective function. Global optimization methods such as Simulated Annealing and Genetic Algorithms can be used effectively to find a global optimum in geophysical applications.

2.1. The Simulated Annealing method

The Simulated Annealing (SA) approach covers a group of global optimization methods. The earliest form of SA is called the Metropolis algorithm, which was further developed to improve principally the speed of the optimum-seeking procedure. The name of the newer methods refer to these faster algorithms (i.e. fast and very fast SA methods).

2.1.1. The Metropolis Simulated Annealing algorithm

The MSA method was developed by Metropolis et al. (1953). In metallurgy the removal of work-hardening is realized by a slow cooling manipulation from the temperature of liquid alloy state. This process reduces progressively the kinetic energy of a large number of atoms with high thermal mobility, which is followed by the starting of crystallization. Theoretically, the perfect crystal with minimal overall atomic energy can be produced by an infinitely slow cooling schedule. This is analogous with the stabilization of the inversion procedure at the global optimum of the objective function. A quick cooling process causes grating defects and the solid freezes in imperfect grid at a relatively higher energy state. It is similar to the trapping of the inversion procedure in a local minimum. However, the atoms may escape from the high energy state owing to a special process called annealing to achieve the optimal crystal grating by a slower cooling process. The MSA algorithm employs this technology to search the global optimum of the objective function which has been applied to several types of geophysical probles, for instance, in calculating seismic static corrections (Rothman, 1985; Rothman, 1986; Sen and Stoffa, 1997), global inversion of vertical electric sounding data collected

above a parallelly layered 1-D structure (Sen et al., 1993), joint inversion of magnetic and electric resistivity data (Dittmer and Szymansky, 1995), joint inversion of direct current and electromagnetic data (Sharma and Kaikkonen, 1999), joint inversion of direct current and seismic refraction data and resolving the problem of ambiguity (Kis, 1998; Kis, 2002).

The objective function

To solve the inverse problem an adequate objective function is defined, which is minimized during the optimization procedure. In most cases it is based on the L_2-norm of the difference between the measured and calculated data

$$E_2 = \frac{1}{N}\sum_{i=1}^{N}\left(d_i^{(m)} - d_i^{(c)}\right)^2, \qquad \text{(I-2.1)}$$

where N is the number of data, $\vec{d}^{\,m}$ and $\vec{d}^{\,c}$ are the measured and calculated data vectors, respectively. The latter does need to be linearized. The objective function is called energy function in the terminology. Geophysical inversion applications sometimes requires to use other types of norm than the Euclidean. A more general energy function for the solution of the mixed-determined problem is

$$\Phi = \left\|\vec{e}\right\|_p + \varepsilon^2\left\|\vec{m}\right\|_q = \sum_{k=1}^{N}\left| d_k^m - g_k\left(\vec{\mathrm{m}}\right)\right|^p + \varepsilon^2\sum_{j=1}^{M}\left|m_j\right|^q, \qquad \text{(I-2.2)}$$

where parameters p, q, ε may be chosen arbitrarily to the given inverse problem.

The method of parameter modification

The SA method is based on the modification of the elements of the model vector during an iteration procedure. The modification of the j-th parameter can be written as follows

$$m_j^{new} = m_j^{old} \pm b,$$

where b is the amount of changing. The simplest solution is that the given parameter is perturbed by a fixed value in each iteration step, which limits the variability of parameters, especially at the end of the procedure. It is more efficient when parameter b is changed randomly between $0 \le b \le b_{\max}$, while parameter b_{max} is renewed according to $b_{max} = b_{max} \cdot \varepsilon$, where parameter ε is a specified number from the interval of 0 and 1.

The acceptace criteria of estimated parameters

During the random seeking the energy function (I-2.1) is calculated and compared with the previous one in every iteration step (ΔE). The acceptance probability of the new model depends on the Metropolis-criteria

$$P(\Delta E,T)=\begin{cases}1 & , \ if \ \ \Delta E \le 0\\ e^{-\Delta E/T} & , \text{ otherwise}\end{cases}, \qquad \text{(I-2.3)}$$

where the model is always accepted when the value of energy function is lower in the new state than that of the previous one. Generalized temperture T represents an important control parameter, but it has not physical meaning. If the energy of the new model increased, there is also some probability of acceptance depending on the values of the energy function and control temperature T. If the following criteria

$$\alpha \le \exp(-\Delta E/T) \ . \qquad \text{(I-2.4)}$$

fulfills then the new model is accepted but otherwise it is rejected. Parameter α is a random number generated with uniform probability from the interval of 0 and 1. This criteria assures the escape from the local minimum.

The MSA method can be described by a simple and clear-cut scheme. The pseudo FORTRAN code is given by Sen and Stoffa (1997):

Start at a random location $\vec{m}_0$ with energy $E(\vec{\mathrm{m}}_0)$
loop over temperature (T)
loop over number of random moves/temperature

$\Delta E = E(\vec{\mathrm{m}}_1) - E(\vec{\mathrm{m}}_0)$

$P = \exp(-\frac{\Delta E}{T})$

if $\Delta E \le 0$ then

$\vec{m}_0 = \vec{m}_1$

$E(\vec{\mathrm{m}}_0) = E(\vec{\mathrm{m}}_1)$

end if

if $\Delta E \ge 0$ then

draw a random number $\alpha = U[0,1]$

if $P \ge \alpha$ then

$\vec{m}_0 = \vec{m}_1$

$$E(\vec{m}_0) = E(\vec{m}_1)$$

end if

end if

end loop

end loop

The rate of convergence is very sensitive to the cooling schedule, which cannot be either too fast or too slow. Fast cooling may cause the procedure to trap in a local minimum, while the slow one makes the procedure speed down and does not give better result. According to Rothman (1986) the found of the global minimum is successful when the following process is chosen. We set high initial temperature at the beginning of the procedure to try numerous energy states. Then a fast cooling schedule is applied until a low temperature, which was named later as critical temperature (T_c). At last the critical temperature is decreased by the method of geometrical attenuation (Dittmer and Szymansky, 1995)

$$T_i = T_c \, g^i ,$$

where g is an adequately chosen number between 0 and 1. The choice of temperature T_c is of crucial importance, to which a procedure was suggested by Basu and Frazer (1990).

2.1.2. The Fast Simulated Annealing algorithm

The SA procedure based on the Metropolis criteria is a widespread method for global optimum seeking. The advantages of MSA are initial model independence, simple and clear-cut program coding, exact mathematical treatment of the conditions of finding a global optimum (i.e. with the theory of Markov chains). The drawback is its computing time or slow rate of convergence, which sets a limit to the reduction of control temperature.

It is provable that the finding of global minimum is guaranteed only when the following cooling schedule is used

$$T(n) = \frac{T_0}{\ln(n)} , \qquad \text{(I-2.5)}$$

where T_0 denotes an initial temperature. The Fast Simulated Anneling (FSA) method only differs in one point from MSA. In equation (I-2.4) the probability distribution of random number α is non-uniform in case of FSA, i.e it is Cauchy-distributed. It is provable that the temperature in the n-th iteration step should be hereunder

$$T(n) = \frac{T_0}{n}, \tag{I-2.6}$$

where T_0 is the initial temperature.The above formula allows faster cooling compaed to equation (I-2.5) while the global optimum is still guaranteed.

2.1.3. The Very Fast Simulated Annealing algorithm

The unknown model parameters are modified in the same manner in the MSA and FSA inversion procedures. This is not always univocal, because the parameters can vary within different ranges. The parameters contribute to the change of objective function in different amount (i.e. different parameter sensitivity). This circumstance requires the improvement of the SA algorithm by modifying each parameter as optimal as possible. This problem can be solved by the VFSA algorithm.

Let the *i*-th model parameter change within the interval of $m_i^{\min} \le m_i \le m_i^{\max}$. In the *k*-th iteration step the modification of the same parameter is done by the formula of $m_i^{k+1} = m_i^k + y_i(m_i^{\max} - m_i^{\min})$, where the random number y_i from the interval of -1 and 1 is generated according to

$$y_i = sign(u - 0.5)T_i\left[(1 + \frac{1}{T_i})^{|2u_i - 1|} - 1\right],$$

where parameter *u* is a uniformly distributed random number from the interval of 0 and 1 and $sign(u - 0.5)$ is a randomly chosen sign. It can be seen that the above formula includes temperature T_i, which means that different temperatures are used to perturb the model parameters. Thus, each model parameter can be modified in a different amount.

The new value of the energy function is calculated by the new vector of model parameters in the VFSA procedure. The change in the energy function can be calculated accordingly. The acceptance rule of the new model is similar as that of the MSA procedure. It requires the use of a set of control parameters (i.e. temperature) that should be decreased progressively. To achieve a global minimum the *i*-th cooling schedule must be the following

$$T_i(n) = T_{0i}\exp(-c_i\sqrt[N]{n}), \tag{I-2.7}$$

where n is the number of iteration steps, T_{0i} is the i-th initial temperature, N is the number of unknowns of the inverse problem and c_i is an arbitrarily chosen constant. According to formula (I-2.7) the temperature can be decreased by an exponential function. Therefore, the VFSA procedure is faster that both the procedures of MSA and FSA.

The pseudo FORTRAN code of the VFSA method given by Sen and Stoffa (1997) is:

Start at a random location $\vec{m}_0$ with energy $E(\vec{\mathrm{m}}_0)$
loop over temperature (T)
    loop over number of random moves/temperature
        loop over model parameters i=1,N
            $u_i = U[0,1]$
            $y_i = sign(u-.5)\,T_i\left[(1+\frac{1}{T_i})^{|2u_i-1|}-1\right]$
            $m_i^{new} = m_i^{old} + y_i(m_i^{\max} - m_i^{\min})$
        end loop
        now we have a new model m^{new}
        $\Delta E = E(\vec{\mathrm{m}}^{\mathrm{new}}) - E(\vec{\mathrm{m}}_0)$
        $P = \exp(-\frac{\Delta E}{T})$
        if $\Delta E \le 0$ then
            $\vec{m}_0 = \vec{m}^{new}$
            $E(\vec{\mathrm{m}}_0) = E(\vec{\mathrm{m}}^{\mathrm{new}})$
        end if
        if $\Delta E \ge 0$ then
            draw a random number $\alpha = U[0,1]$
            if $P \ge \alpha$ then
                $\vec{m}_0 = \vec{m}^{new}$
                $E(\vec{\mathrm{m}}_0) = E(\vec{\mathrm{m}}^{\mathrm{new}})$
            end if
        end if
    end loop
end loop

References

BASU A. & FRAZER L.N., 1990: Rapid determination of critical temperature in simulated annealing inversion. *Science* **249**, 1409-1412.

DITTMER J.K. & SZYMANSKI J.E., 1995: The stochastic inversion of magnetics and resistivity data using the simulated annealing algorithm. *Geophysical Prospecting* **43**, 397-416.

KIS M., DOBRÓKA M. & AMRAN A., 1995: Joint inversion of geoelectric, refraction- and surface wave seismic data. *EAEG 57. International Conference*, Glasgow.

KIS M., 2002: Generalized Series Expansion (GSE) used in DC geoelectric – seismic joint inversion, *Journal of Applied Geophysics* **50**, 401– 416

METROPOLIS N., ROSENBLUTH M.N., TELLER A.H. & TELLER E., 1953: Equation of State calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087-1092.

ROTHMAN D.H., 1985: Nonlinear inversion, statistical mechanics, and residual static estimation. *Geophysics* **50**, 2784-2796

ROTHMAN D.H., 1986: Automatic estimation of large residual statics corrections. *Geophysics* **51**, 332-346.

SEN M.K., BHATTACHARYA B.B. & STOFFA P.L., 1993: Nonlinear inversion of resistivity sounding data. *Geophysics* **58**, 496-507.

SEN M.K. & STOFFA P.L., 1997: Advances in Exploration Geophysics **4**, Global optimization methods in geophysical inversion. *Elsevier Science Ltd.*

SHARMA S.P. & KAIKKONEN P., 1999: Appraisal of equivalence and suppression problems in 1D EM and DC measurements using global optimization and joint inversion. *Geophysical Prospecting* **47**, 219-249.

2.2 The Genetic Algorithm

The name of Genetic Algorithm (GA) refers to a subgroup of evolutionary algorithms that perform global optimum search by using stochastic rules. The GA method is based on the analogy to the process of natural selection. The first application for its use in solving optimization problems was proposed by John Holland (1975). The GA procedure improves a population of random models in an iteration process. For optimization problems the model can be considered as an individual of an artificial population. Each individual of a given generation has a fitness value, which represents its survival capability. The aim of the GA procedure eventually is to improve the subsequent populations by maximizing the average fitness of individuals. In the geophysical inverse problem the fitness function is connected to the distance between the observed data and theoretical data calculated on the actual petrophysical model. The most probable model can be estimated by the maximization of fitness function (i.e. minimization of data misfit). During the genetic process the fittest individuals reproduce more successfully in the subsequent generations than those who have smaller fitness. In the last generation, the individual with the maximum fitness corresponds to the solution of the inverse problem. In the classical GA procedure the model parameters are encoded using a binary coding scheme, which sets a limit to the resolution of the petrophysical model and the accuracy of the estimation results. The Float-Encoded GA computes with model parameters as real numbers, which is faster than binary coding, because it does not use coding-decoding phases and achieves higher resolution of the model space (Houck et al., 1996).

In geophysical inversion the following scalar characterizes the misfit between the observed and theoretical data calculated on the *i*-th model

$$E^{(i)} = E\left(\vec{d}^{(m)} - \vec{g}\left(\vec{m}^{(i)}\right)\right).$$

Since the aim of the inversion procedure is to find the minimum of norm *E*, thus the fitness function can be chosen as

$$F(\vec{m}^{(i)}) = \frac{1}{E^{(i)} + \varepsilon^2} = \max,$$

where ε^2 is a properly chosen positive scalar limiting the range of fitness. The fitness function defined in the *M*-dimensional model space is not always continuous (e.g. when some domains are excluded by the geophysical-geological laws). Moreover, the fitness is connected to the *N*-dimensional data space, which is built up of discrete points and always noisy. In these types of environments the linearized methods usually fail, because the calculation of the derivatives in Jacobi's matrix is essential and adequate a priori information is needed to avoid

local minima of the objective function. Compared to linear inversion the GA procedures are more efficient, robust and able to find the global maximum.

The workflow of the GA-based inversion procedure is different from SA algorithm. The latter improves only one model during the optimum search. After collecting measurement data and a priori information we set the searching intervals of the model parameters. Normally an initial population of 30-100 models is generated from the search space randomly. In the forward modeling phase of the inversion procedure theoretical data are calculated for each model. These data are then compared to real measurements. As long as the misfit between the observation and prediction is too high, the model population is improved by the use of the some random genetic operations, i.e. selection, crossover, mutation and reproduction. Instead of one point seach, several models are analyzed simultaneously. This assures to avoide the local optimum places in the model space. The GA technique is also advantageous because it does not require the calculation of derivatives and too much a priori information for the search. They are practically independent of the initial model. Only the bounds of the model parameters should be correctly given. The modern algorithms avoids low resolution coding schemes, they are calculating the fitness directly with real valued model parameters. The development of convergence of the GA procedure is highly dependent on the proper setting of control parameters of genetic operators. The searching process stops when a termination criterion is met. This can be defined by the maximum number of iterations (generations) or by a specified threshold in the distance between measured and calculated data. In the last iteration step we accept the fittest individual of the generation as the optimal solution.

2.2.1 The Classical Genetic Algorithm

The Classical Genetic Algorithm (CGA) operates with the model parameters in coded form. The genetic operations are applied to the codes directly. During the optimum search the fitness function is to be evaluated, thus the decoding of the models must be done in each iteration steps. The simplest method is binary coding, which converts the model to a finite sequence of numbers. The structure of the sequence is analogous to a chromosome, which uniquely identifies all individuals of the population. The geophysical model is represented by a bit string which is constructed from the codes of model parameters. The basic element of the string is the gene, which is represented by 0 or 1 bits. For instance, the seismic velocity from the interval of 1500-1810 m/s can be resolved with precision of 10 m/s by using 5 bits

$$
\begin{aligned}
&00000 = 1500\\
&00001 = 1510\\
&00010 = 1520\\
&00011 = 1530\\
&\quad\vdots\\
&11111 = 1810
\end{aligned}
\quad .
$$

It is easily noticable that the increase of velocity from 1530 to 1540 m/s requires the change-over of 3 bits. Binary coding makes the GA procedure slow down, because more bits have to be usually switch over to increase the model parameters with a unit value. From the point of view of optimal CPU time the so-called Gray- or other coding scheme is favourable (Forrest, 1993).

After the procedure of coding other genetic operations are applied to the model parameters. The first one is the selection, which sorts the individuals out of the population and reproduces them in proportion to their fitness values. As a result only the fittest models enter into a new generation, while individuals with low fitness die. The selection mechanism proportional to fitness was performed firstly by Roulette selection (Goldberg, 1989). There exist other selection methods as well.

Crossover operator is applied to a selected pair of individuals to make a partial information exchange between the parents. Using the above seismic problem a simple crossover produces the following pair of models (children) at k=4 random bit position

$$
\begin{aligned}
&m^{(parent1)} = 0110\,|\,1\\
&m^{(parent2)} = 1100\,|\,0\\
&\qquad\Downarrow\\
&m^{(child1)} = 01100\\
&m^{(child2)} = 11001
\end{aligned}
\quad .
$$

The above operation can be performed by multipoint crossover, where more bit positions are used at which the string is cut.

The last genetic operator is mutation, which switches over the value of a randomly chosen bit of the string. The rate of convergence of the GA procedure is highly dependent on the mutational rate (i.e ratio of the number of mutated individuals to population size). For instance, a uniform mutation at the fifth bit position gives

$$m^{(parent1)} = 01100$$
$$\Downarrow \quad .$$
$$m^{(child1)} = 01101$$

At the end of the CGA procedure the individual with the highest fitness in the last generation is accepted as the solution of the inverse problem.

2.2.2 The Float-Encoded Genetic Algorithm

The CGA is an effective but very rather time-consuming global optimization procedure. The solution of the forward problem requires the decoding of the actual model in each iteration steps. Many coding-decoding phases slows down the inversion procedure. The Float-Encoded Genetic Algorithm (FGA) suggested by Michalewicz (1992) processes the models coded directly as vectors of floating-point numbers, which provides higher resolution and much better CPU performance than CGA.

The principles of the FGA procedure are similar as those of the CGA method with a difference that FGA defines the genetic operations as real ones. In case of Roulette selection the probability of selecting the *i*-th model from the population is

$$P\left(\vec{m}^{(i)}\right) = \frac{F\left(\vec{m}^{(i)}\right)}{\sum_{j=1}^{S} F\left(\vec{m}^{(j)}\right)},$$

which is the ratio of the fitness of the actual model and the sum of fitnesses of all models (*S* is the population size). The *i*-th individual is copied into the new population only when the cumulative probability of the population

$$C_i = \sum_{j=1}^{i} P_j$$

fulfills the condition that $C_{i-1} < U \le C_i$, where *U* is a uniform random number in the range of 0 and 1. As a result of the selection process the worst models die, but the better ones are reproduced into the new generation more than once as good. There are other selection methods based on the above scheme. In case of normalized geometric ranking selection, the individuals are sorted according to their fitness values. The rank of the best individual is 1 and that of the worst is *S*, which is the size of the population. The probability of selecting the *i*-th individual can be calculated by

$$P_i = \frac{q}{1-(1-q)^S}(1-q)^{r_i-1},$$

where r_i is the rank of the i-th individual, q is the probability of selecting the best individual. The latter quantity is a control parameter, which has to be set at initialization.

In the next step a pair of individuals is selected from the population and a partial information exchange is made between the original individuals. The simple crossover operator recombines the individuals as

$$\mathrm{m}^{(1,\mathrm{new})} = \begin{cases} m_i^{(1,old)} & if\ i < x \\ m_i^{(2,old)} & otherwise, \end{cases}$$

$$\mathrm{m}^{(2,\mathrm{new})} = \begin{cases} m_i^{(2,old)} & if\ i < x \\ m_i^{(1,old)} & otherwise, \end{cases}$$

where x denotes the position of the crossing point, $m_i^{(old)}$ and $m_i^{(new)}$ are the i-th model parameter before and after crossover, respectively. The last genetic operator is mutation, which selects an individual from the population and changes one of its model parameters to a random number. In case of uniform mutation the new model is calculated by changing the value of the j-th model parameter as

$$\mathrm{m}^{(\mathrm{new})} = \begin{cases} u & if\ i = j \\ m_i^{(old)} & otherwise, \end{cases}$$

where u is a uniform random number generated from the range of the j-th model parameter. The above detailed three genetic operations are repeated until the end of the iteration procedure. In the last iteration step, we accept the fittest individual of the generation as the optimal geophysical model. Since the model parameters must not exceed their bounds, thus the applicability of the genetic operations must also be considered. The probability of feasibility is defined as

$$P_f = \left\{ \begin{matrix} 0 & , \forall i : m_i^{(new)} > \max(m_i) \parallel m_i^{(new)} < \min(m_i) \\ 1 & , otherwise, \end{matrix} \right\}, \qquad (\mathrm{i}=1,2,\ldots,\mathrm{M})$$

where M is the number of model parameters. When the proability P_f remains zero after numerous attempts the new model is chosen for the old one.

References

FORREST S., 1993: Genetic algorithms: Principles of adaptation applied to computation. *Science*, **261**, 872-878.

GOLDBERG D. E., 1989: Genetic Algorithms in search, optimization and machine learning. *Addison-Wesley Publishing Company*, Reading, Massachusetts, USA.

HOLLAND J. H., 1975: Adaptation in natural and artificial systems. *University of Michigan Press*, Ann Arbor, MI.

HOUCK C. R., JOINES J. A. & KAY M. G., 1995: A genetic algorithm for function optimization - A matlab implementation. *North Carolina State University*, Research, National Science Foundation DMI-9322834.

MICHALEWICZ Z., 1992: Genetic Algorithms plus data structures equals evolution programs. *Springer-Verlag Incorporation*, AI Series, New York.

III. The theory of joint inversion methods

In mixed-determined inverse problems the partial underdetermination is treated by the subsidiary minimization of the squared norm of the model vector. Such regularization scheme results in a deviation from the solution based on the minimization of the L_2- or L_1-norm. When it can be done, we can help the problem by the introduction of daming factor ε^2, which is continuously reduced to zero along the iteration procedure. However, several examples show that this approach is not always feasible. In these cases the regularization is successful only when additional data sets decrasing the amount of underdetermination are involved in the inversion procedure.

In the geophysical practice, it is fairly general that several surveying methods are used together in solving the geophysical problem. To solve the inverse problem these data sets can be treated in two ways. On the one hand, we can invert different data sets idependently, but the result of a given inversion procedure can be used up in the other one. For instance, the number of unknowns in the new inversion procedure can be reduced by fixing some parameters obtained from the other procedure. This is called co-operative inversion. This approach cannot be used to help the numerical problems of single procedures and resolve the problem of ambiguity when all inversion procedures are instable. The solution for this problem is the use of the joint inversion methodology, when different data sets are inverted in one inversion procedure simultaneously. The joint inversion of direct current (DC) and magnetotelluric (MT) data was performed first by Vozoff and Jupp (1975). This data set was completed by transient electromagnetic data (TEM) and jointl inverted by Yang and Tong (1988), which further improve the stability of the inversion procedure. In these tests, the data were dependent on the same physical parameters. Data measured by different physical principles are worth to be integrated in a joint inversion procedure when there are common model parameters that are dependent on more data types. The joint inversion of in-mine seismic and geoelectric data was elaborated by Dobróka et al. (1991). In spite of that the data sets depend on different physical parameters, i.e. on sesmic velocity and electric resistivity, the positions of layer-boundaries can be considered to be common parameters in the given problem. (There are many cases when this assumption cannot be tolerated.) Kis et al. published some results on the joint inversion of seismic refraction and guided-wave dispersion as well as geoelectric data. The joint inversion algorithm of surface geolelectric and guided-wave dispersion data was published by Hering et al. (1995). The results of the same algorithm was detailed by Misiek et al. (1997), Kis (2000) showed the application of geoelectric-seismic joint inversion

in the resolving the geoelectric ambiguity problem, Gallardo-Delegado et al. (2003) published about the joint inversion of gravity and magnetic data, Jegen et al. (2009) integrated gravity and magnetotelluric data sets in a joint inversion procedure. These examples show the widespread application of joint inversion methods in the geophysical inversion practice.

3.1. The algorithm of joint inversion

Consider *P* number of data sets based on different physical principles. They can be originated from different surveying methds or from the same method by different measurement arrays. (Two data sets are different when they can be calculated in distinct forward problems based on different algorithms or formulae.)

Let the measured and calculated data vectors of different data sets be denoted by $\vec{d}_m^{(i)}$ and $\vec{d}_e^{(i)}$ (i=1,2,…,P), where $\vec{d}_e^{(i)} = \vec{g}^{(i)}(\vec{m}^{(i)})$. The response function $\vec{g}^{(i)}$ represents the set of equations of the i-th forward problem, while model vector $\vec{m}^{(i)}$ includes the unknowns of the same problem. Let N_i be the number of elements of the i-th data set, then the i-th observed data vector is

$$\vec{d}_m^{(i)} = \left\{d_1^{(i)},...,d_{N_i}^{(i)}\right\}^T .$$

The conjunction of all data vector is represented by

$$\vec{d}_m = \left\{d_1^{(1)},...,d_{N_1}^{(1)},...,d_1^{(i)},...,d_{N_i}^{(i)},...,d_1^{(P)},...,d_{N_P}^{(P)}\right\}^T . \qquad \text{(I-3.1)}$$

A similar vector for the calculated data is

$$\vec{d}_e = \left\{g_1^{(1)},...,g_{N_1}^{(1)},...,g_1^{(i)},...,g_{N_i}^{(i)},...,g_1^{(P)},...,g_{N_P}^{(P)}\right\}^T , \qquad \text{(I-3.2)}$$

in which model vector $\vec{m}$ contains the M number of model parameters of all geophysical methods instead of $\vec{m}^{(i)}$. The response function of the i-th forward problem is

$$\vec{g}^{(i)} = g^{(i)}(\vec{m}) .$$

The data vector of the joint inversion problem is $\vec{d}_m$, while the calculated data vector is a nonlinear function of the model vector

$$\vec{d}_e = \vec{g}(\vec{m}) . \qquad \text{(I-3.3)}$$

The observed and calculated data vectors have $N = \sum_{i=1}^{P} N_i$ number of elements, respectively.

Similarly to the linearized (independent) inversion methods, let us linearize the problem near the point $\vec{m}_o$ in the model space. In the point $\vec{m} = \vec{m}_o + \delta\vec{m}$ we can write that

$$d_k^{(e)}(\vec{m}) = d_k^{(o)} + \sum_{j=1}^{M} \left(\frac{\partial g_k}{\partial m_j} \right)_{\vec{m}_o} \delta(m_j) \,, \tag{I-3.4}$$

where $d_k^{(o)} = g_k(\vec{m}_o)$. By introducing the *N*-by-*M* Jacobi's matrix

$$G_{kj} = \left(\frac{\partial g_k}{\partial m_j} \right)_{\vec{m}_o} \,,$$

with which equation (I-3.4) can be written as

$$\vec{d}_e = \vec{d}^{(o)} + \underline{\underline{G}}\delta\vec{m} \,.$$

Therewith the deviation vector is

$$\vec{e} = \vec{d}^{(m)} - \vec{d}^{(o)} - \underline{\underline{G}}\delta\vec{m} \,.$$

In the joint inversion problem the data sets can have different magnitudes and dimensions. Before minimizing the L_2- or L_1-norm of the over- or mixed-determined inverse problem the following normalized deviation vector is introduced

$$f_k = \frac{e_k}{d_k^{(o)}} = \frac{d_k^{(m)} - d_k^{(o)}}{d_k^{(o)}} - \sum_{j=1}^{M} \frac{1}{d_k^{(o)}} \left(\frac{\partial g_k}{\partial m_j} \right)_{\vec{m}_o} \delta m_j \,,$$

which may also stabilize the solution of the inverse problem numerically. By the above normalization the solution is prevented from being overweighted with the influence of a given data type. Similarly the model parameters can have different dimensions. Then it is advantageous to make normalization in the model space, too

$$f_k = \frac{d^{(m)} - d_k^{(o)}}{d_k^{(o)}} - \sum_{j=1}^{M} \frac{m_j^{(o)}}{d_k^{(o)}} \left(\frac{\partial g_k}{\partial m_j} \right)_{\vec{m}_o} \frac{\delta m_j}{m_j^{(o)}} \,. \tag{I-3.5}$$

By introducing the notations

$$y_k = \frac{d_k^{(m)} - d_k^{(o)}}{d_k^{(o)}} \,, \quad x_j = \frac{\delta m_j}{m_j^{(o)}} \tag{I-3.6}$$

$$G_{kj} = \frac{m_j^{(o)}}{d_k^{(o)}} \left(\frac{\partial g_k}{\partial m_j} \right)_{\vec{m}_o}$$

the relative deviation vector can be written as

$$\vec{f} = \vec{y} - \underline{\underline{G}}\vec{x} \,. \tag{I-3.7}$$

If the joint inversion problem is overdetermined, the minimization of the following objective function

$$E = \vec{f}^T \vec{f}$$

leads to the normal equation

$$\underline{\underline{G}}^T \underline{\underline{G}}\vec{x} = \underline{\underline{G}}^T \vec{y} \,. \tag{I-3.8}$$

By solving the above equation the elements of the model vector can be updated, for instance after the *r*-th iteration step they are calculated as

$$m_j^{(r+1)} = m_j^{(o)}\left(1 + x_j^{(r)}\right). \tag{I-3.9}$$

If weighting matrix $\underline{\underline{W}}$ defined in the data space is given, then the solution can be adjoined to the following norm

$$E = \left(\vec{f}, \underline{\underline{W}}\vec{f}\right),$$

which results in the solution

$$\underline{\underline{G}}^T \underline{\underline{W}}\, \underline{\underline{G}}\vec{x} = \underline{\underline{G}}^T \underline{\underline{W}}\, \vec{y} \,. \tag{I-3.10}$$

The correction of the mdel vector is performed by equation (I-3.9). If the inverse problem is mixed-determined the minimization of objective function

$$\phi = \vec{f}^T \vec{f} + \varepsilon^2 \vec{m}^T \vec{m}$$

leads to the equation

$$\left(\underline{\underline{G}}^T \underline{\underline{G}} + \varepsilon^2 \underline{\underline{I}}\right)\vec{x} = \underline{\underline{G}}^T \vec{y} \,.$$

If the solution is searched by the optimization of the L_p-norm of the deviation vector, we can use the IRLS method properly modified by equations (I-3.5) and (I-3.6). The details of derivation can be found in Chapter 1.1.3.

References

DOBRÓKA M., GYULAI Á., ORMOS T., CSÓKÁS J. & DRESEN L., 1991: Joint inversion of seismic and geoelectric data recorded in an underground coal mine. *Geophysical Prospecting* **39**, 643-665.

GALLARDO-DELAGDO L.A., PEREZ-FLORES M.A. & GOMEZ-TREVINO E., 2003: A Versatile Algorithm for Joint 3D Inversion of Gravity and Magnetic Data. *Geophysics* **68/3**, 949-959.

HERING A., MISIEK R., GYULAI Á., ORMOS T., DOBRÓKA M. & DRESEN L., 1995: A joint inversion algorithm to process geoelectric and surface wave seismic data. *Geophysical Prospecting* **43**, 135-156.

JEGEN M., HOBBS R.W., TARTIS P. & CHAVE A., 2009: Joint inversion of marine magnetotelluric and gravity data incorporating seismic constrains. Preliminary results of sub-basalt imaging of the farve shelf. *Earth and Planetary Science Letters*, **282/1-4**, 47-95.

KIS M., DOBRÓKA M. & AMRAN A., 1995: Joint inversion of geoelectric, refraction- and surface wave seismic data. *EAEG 57. International Conference*, Glasgow.

KIS M., 2002: Generalized Series Expansion (GSE) used in DC geoelectric – seismic joint inversion, *Journal of Applied Geophysics* **50**, 401– 416

MISIEK R., LIEBIG A., GYULAI Á., ORMOS T., DOBRÓKA M. & DRESEN L., 1997: A joint inversion algorithm to process geoelectric and surface wave seismic data. Part II: applications. *Geophysical Prospecting* **45**, 65-86.

VOZOFF K. & JUPP D.L.B., 1975: Joint inversion of geophysical data. *Geophysical Journal of the Royal Astronomical Society* **42**, 977-991.

YANG C.H. & TONG L.T., 1988: Joint inversion of DC, TEM and MT data. *58th Annual International Society of Exploration Geophysicists Meeting and Exposition*, Anaheim, 408-410.

IV. The series expansion-based inversion method

In case of complex geological structures the forward problem is solved by approximate numerical methods, e.g. finite difference (FDM) and finite element (FEM) methods. These methods designate the way of discretization, too. For instance, the 3-D space can be divided into properly sized blocks. The infinite half-space must be defined with a great number of cells. An adequately accurate calculation requires, e.g. in DC geoelectric modeling, 50-100 cells in horizontal direction and 100-200 cells in vertical direction. In the solution of the related inverse problem the physical parameters of the cells are assumed to be unknowns. The number of elements of the model vector can reach 1-2 millions, while that of the data is some hundreds or thousands, thus the inverse problem is underdetermined.

In Chapter I the purely underdetermined inverse problem and its solution were studied. It was concluded that a unique solution could be obtained only by assuming additional conditions, which cannot be coupled to measurements. These conditions have in common that they formulate an obvious requirement on the solution (e.g. simplicity, smoothness, magnitude of the derivatives etc.). In some cases the use of additional conditions can be advantageous, but in extremely underdetermined problems, where the number of arbitrary chosen conditions is order(s) of magnitude higher that that of the physical conditions, the solution is smearing.

4.1. Series expansion-based discretization of model parameters

The above detailed problems are rooted in the discretization of the model, to which a different method is suggested. At the Department of Geophysics, University of Miskolc the series expansion-based discretization scheme proved to be beneficial in several cases.

Consider the model parameter showing spatial dependency, such as layer-thickness or resistivity, in the form of series expansion

$$p(x,y,z)=\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_z}B_l\Psi_i(x)\Psi_j(y)\Psi_k(z)\,,$$

where $\Psi_1....\Psi_N$ are basis functions and N_x, N_y, N_z are their numbers requisite in the description of *x*, *y*, *z* dependencies, respectively ($l=i+(j-1)*N_x+(k-1)*N_x*N_y$). The type of basis fuctions can be different, e.g. power functions or Legendre polinomials. The choice

of the type may influence the stability of the inversion procedure. The basis functions constituting an orthonormal system are favourable.

The unknowns of the inverse problem (i.e. model parameters) are the series expansion coefficients B_l and their number is $M = N_x N_y N_z$. If the number of data is more that that of the model parameters ($N>M$), the inverse problem is overdetermined. In this problem, it is unnecessary to use additional constraints, because the result is dependent only on the data. Thus, the quality of the inversion result is hopefully better than that of the underdetermined problem.

The solution of the inverse problem is iterative, which requires the correction of expansion coefficients

$$\vec{B}^{new} = \vec{B}^{old} + \delta\vec{B}$$

and the solution of the forward problem in each step. To predict data the relevant physical parameter must be calculated

$$p^{new}(x,y,z) = \sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_z} B_l^{new}\,\Psi_i(x)\Psi_j(y)\Psi_k(z)\,. \qquad \text{(I-4.1)}$$

The advantage of the method is that one can calculate the values of the model parameters at any place in the half-space (e.g. in the cells defined by the FDM algorithm). In solving the forward problem one does not need to make concessions in the accuracy. The overdetermined inverse problem based on series expansion has the same accuracy as the underdetermined one. The only difference is that the unknowns of the inverse problem are the series expansion coefficients instead of the parameters of the cells.

4.2. Overdetermined series expansion-based inversion with a priori information

A priori information about the area of investigation is usually available for the interpretation. This knowledge is of great importance during the inversion procedure. Since the measurement data are contaminated by noise, the geological model is simplified and the solution of the forward problem is approximately valid, the inversion algorithms have internal uncertainty (instability, ambiguity). This can be reduced by the use of a priori information. In series expansion-based inversion procedures the situation is the same. In the inversion algorithm suggested in Chapter 4.1, we cannot assume anything for the structure itself, only the number of expansion coefficients can be specified. If we have information about the structure, for instance it is layered or else, we can make additional assumptions within the series expansion-

based inversion method, which can allow to reduce the number of unknowns of the inverse problem.

4.2.1. Layer-wise homogeneous model

In case of 3-D model geometry, the *q*-th layer-boundary can be described as a function $z = f_q(x, y)$, which can be discretized by series expansion

$$z = f_q(x, y) = \sum_{i=1}^{N_x^{(q)}} \sum_{j=1}^{N_y^{(q)}} C_l^{(q)} \Psi_i(x) \Psi_j(y), \qquad \text{(I-4.2)}$$

where $C_l^{(q)}$ represents the expansion coefficients, $l = L_q + i + (j-1) * N_x$, L_q is the initial index required in the *q*-th layer. The number of unknowns for a given layer-boundary is $N_x^{(q)} N_y^{(q)}$, while that of the *P*-layered model assuming one physical parameter per layer is

$$M = \sum_{q=1}^{P} N_x^{(q)} N_y^{(q)} + P + 1. \qquad \text{(I-4.3)}$$

For instance, the boundaries of a four-layered structure approximated by fifth degree polinomials along both horizontal dimensions can be described by *M*=4*5*5+4=104 number of expansion coefficients. Compared to the number of unknowns of underdetermined problems (typically ~10^6), it can be seen that the choosing of the discretization procedure is essential in inverse modeling. The number of 3-D measuring data exceeds that of the unknowns, thus the inverse problem is overdetermined.

4.2.2 Layer-wise inhomogeneous model

Assuming a layer-wise homogeneous model is often not adequate. In the following subsections such structures are presented whose physical properties are inhomogeneous between two layer-boundaries.

a.) Vertically inhomogeneous model

For the sake of simplicity, let us assume first that the model is only vertically inhomogeneous. The physical parameter of the *q*-th layer can be written as

$$p_q(z) = \sum_{i=1}^{N_q^{(p)}} D_l^{(q)} \Psi(z)\,, \tag{I-4.4}$$

where the number of unknowns including the layer-boundaries is

$$M = \sum_{q=1}^{P} (N_x^{(q)} N_y^{(q)} + N_q^{(p)}) + 1, \tag{I-4.5}$$

where $l = L_q + i$, L_q is the initial index in the *q*-th layer and the half-space is homogeneous.

For instance, the boundaries of a four-layered structure approximated by fifth degree polynomials along both horizontal dimensions and the vertical variations of physical parameters approximated by fifth degree polynomials can be described by M=4*(5*5+5)+1=121 number of expansion coefficients. Compared to the number of unknowns of underdetermined problems (typically ~10^6), it can be seen that the choosing of the discretization procedure is essential in inverse modeling. The number of 3-D measuring data exceeds that of the unknowns, thus the inverse problem is still overdetermined.

b.) Laterally inhomogeneous model

Assuming lateral inhomogeneity in each layer the series expansion-based discretization of the physical parameter is

$$p_q(x,y) = \sum_{i=1}^{N_{p,x}^{(q)}} \sum_{j=1}^{N_{p,y}^{(q)}} D_l^{(q)}\, \Psi_i(x) \Psi_j(y)\,, \tag{I-4.6}$$

where $D_l^{(q)}$ represents the expansion coefficients, $l = L_q + i + (j-1) * N_x$, L_q is the initial index required in the *q*-th layer. The number of unknowns for a given layer-boundary has been broadened with the term of $N_{p,x}^{(q)} N_{p,y}^{(q)}$ (compared to the layer-wise homogeneous model), thus for the *P*-layered model (and half-space) it is

$$M = \sum_{q=1}^{P} (N_x^{(q)} N_y^{(q)} + N_{p,x}^{(q)} N_{p,y}^{(q)}) + 1\,. \tag{I-4.7}$$

For instance, the boundaries of a four-layered structure and its physical parameters approximated by fifth degree polynomials can be described by M=4*(5*5+5*5)+1=201 number of expansion coefficients. Compared to the number of unknowns of underdetermined problems (typically ~10^6), it can be seen that the choosing of the discretization procedure is essential in inverse modeling. The number of 3-D measuring data exceeds that of the unknowns, thus the inverse problem is still overdetermined.

c.) Vertically and laterally inhomogeneous model

In a standard model the lateral and vertical inhomogeneity of the physical parameter must be considered together. In the present case equation (I-4.2) should be completed also with the discretization of the physical parameter

$$p_q(x,y,z) = \sum_{i=1}^{N_{p,x}^{(q)}} \sum_{j=1}^{N_{p,y}^{(q)}} \sum_{k=1}^{N_{p,z}^{(q)}} B_l^{(q)} \Psi_i(x) \Psi_j(y) \Psi_k(z),$$

where $l = L_q + i + (j-1) * N_{p,x}^{(q)} + (k-1) * N_{p,x}^{(q)} * N_{p,y}^{(q)}$, L_q is the initial index required in the *q*-th layer. The number of unknowns for a given layer-boundary has been broadened with the term of $N_{p,x}^{(q)} N_{p,y}^{(q)} N_{p,z}^{(q)}$ (compared to the layer-wise homogeneous model), thus for the *P*-layered model (and half-space) it is

$$M = \sum_{q=1}^{P} (N_x^{(q)} N_y^{(q)} + N_{p,x}^{(q)} N_{p,y}^{(q)} N_{p,y}^{(q)}) + 1. \qquad \text{(I-4.7)}$$

For instance, the boundaries of a four-layered structure approximated by fifth degree polinomials along both horizontal dimensions and the physical parameters approximated by fifth degree polinomials along three dimensions can be described by M=4*(5*5+5*5*5)+1=601 number of expansion coefficients. Compared to the number of unknowns of underdetermined problems (typically ~10^6), it can be seen that the choosing of the discretization procedure is essential in inverse modeling. The number of 3-D measuring data exceeds that of the unknowns, thus the inverse problem is still overdetermined.

The above chapter showed that all inversion algorithms used for the discretization of 3-D structures are greatly overdetermined and does not include arbitrary additional conditions. The suggested algorithm do allow to integrate a priori information into the inversion procedure as well as keep the computing procedures (e.g. FEM or FD). It does not require to make concessions, but supports the integration of modern computing techniques. It can be applied to the 3-D inversion of measurement data of any geophysical surveying method.