ORKUSTOFNUN
NATIONAL ENERGY AUTHORITY
GEOTHERMAL DIVISION

Knútur Árnason

CENTRAL LOOP TRANSIENT ELECTRO-MAGNETIC SOUNDINGS OVER A HORIZONTALLY LAYERED EARTH

OS-89032/JHD-06
Reykjavík, August 1989

ORKUSTOFNUN
NATIONAL ENERGY AUTHORITY
GEOTHERMAL DIVISION

GRENSÁSVEGUR 9,
108 REYKJAVÍK ICELAND

Knútur Árnason

CENTRAL LOOP TRANSIENT ELECTRO-MAGNETIC SOUNDINGS OVER A HORIZONTALLY LAYERED EARTH

OS-89032/JHD-06
Reykjavík, August 1989

ABSTRACT

In this report the principles of central loop transient electromagnetic soundings over a horizontally layered earth are discussed. The electromagnetic theory, on which the sounding method is based, is discussed in some details. This is followed by a discussion of the digital filter method which is used to perform the numerical integrations that are necessary in order to work out the sounding response of a layered earth. The non-linear least-squares inversion method is discussed. Finally a one-dimensional inversion program for inverting the central loop sounding results in terms of layered resistivity models is described. The program is an implementation of the concepts and formulas developed in the report. It is written in standard FORTRAN 77 and can be implemented on any machine supporting FORTRAN 77.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Resistivity methods have been used in geothermal exploration in Iceland since the fifties. The most common method has been Schlumberger soundings but in the last decade, resistivity profiling with half Schlumberger array has been used extensively. Direct current dipole methods have also been to some extent and the magnetotelluric (MT) method has been used for deep crustal studies. The Schlumberger soundings have been used in reconnaissance surveys for low temperature geothermal resources outside the volcanic zones and for mapping high temperature geothermal systems in the volcanic zone. Resistivity profiling has mainly been used for locating vertical or near vertical aquifers for low temperature geothermal waters and vertical resistivity boundaries in high temperature geothermal systems.

Even though resistivity surveys are much cheaper than drilling, their cost is often considerable and surface conditions are often such that use of direct current methods is very difficult and time consuming. This applies especially to many areas in the volcanic zone where large parts of the surface are covered with lava, making current injection into the ground almost impossible. Because of this, Orkustofnun (The National Energy Authority of Iceland) has been looking for resistivity methods which are cheaper than the Schlumberger soundings and can be used in areas with resistive surface conditions. Among the methods that have been considered is the transient electro-magnetic (TEM) method. The central loop TEM method was first tested for geothermal exploration in Iceland in the summer 1986 (Knútur Árnason et.al., 1987) and for mapping of saline ground water in the summer 1988 (Grímur Björnnson and Hjálmar Eysteinsson, 1988). This method turned out to be, in many respects, superior to the conventional Schlumberger soundings. The data collection is much cheaper and quicker and since no current has to be injected into the ground, this method makes many areas, which are practically unaccessible by Schlumberger sounding, easily accessible for geothermal exploration. The inversion of the sounding results in terms of subsurface resistivity structure is considerably more complicated for the TEM-soundings than for Schlumberger soundings. For TEM-soundings, only one-dimensional (horizontally layered earth) inversion is at the present commercially available but the central loop TEM method is less sensitive to lateral resistivity variations than Schlumberger soundings which makes one-dimensional inversion more justifiable. If very detailed information about relatively shallow resistivity structures, e.g. exact locations of near vertical resistivity boundaries is needed, then Schlumberger soundings along with head-on resistivity profiling and with joint two-dimensional modeling is preferable but a central loop TEM survey with one-dimensional inversion can give the general resistivity structure apart from the fine details.

The common principle of all resistivity sounding methods is to induce electrical current in the ground and monitor response signals, normally at the surface, generated by the current distribution. In conventional direct current (DC) soundings such as the Schlumberger soundings, this is done by injecting current into the ground through electrodes at the surface and the measured signal is the electric field (the potential difference over a short distance)

generated at the surface. In MT-soundings the current in the ground is induced by time variations in the Earth's magnetic field, and the measured signal is the electric field at the surface as in the DC-soundings.

In the central loop TEM-sounding method the current in the ground is, as in MT-soundings, generated by a time varying magnetic field. Yet, unlike the MT-soundings, the magnetic field is not the randomly varying natural field but a field of a controlled magnitude generated by a source loop. A loop of wire is placed on the ground and a constant magnetic field of known strength is built up by transmitting a constant current into the loop. The current is then abruptly turned off. The decaying magnetic field induces electrical current in the ground. The current distribution in the ground induces a secondary magnetic field decaying with time. The decay rate of the secondary magnetic field is monitored by measuring the voltage induced in a receiver coil (or a small loop) at the centre of the transmitter loop. The current distribution and the decay rate of the secondary magnetic field depends on the resistivity structure of the earth. The decay rate, recorded as a function of time after the current in the transmitter loop is turned off, can therefore be interpreted in terms of the subsurface resistivity structure.

The depth of penetration in the central loop TEM-soundings is dependent on how long the induction in the receiver coil can be traced in time before it is drowned in noise. For a transmitter loop with an area about 100.000 m^2 and current about 20 Ampers, the induction can be measured up to about 100 ms after the current is turned off. This gives penetration depth of about 500-1000 m, depending on the resistivity structure, which is similar to that of Schlumberger soundings with the maximum distance of 3 km between the current electrodes.

The central loop TEM-sounding method has, as mentioned earlier, several advantages over conventional DC-sounding methods. The most important one is that the transmitter couples inductively to the earth and no current has to be injected into the ground. This is of great importance in areas where the surface is highly resistive.

Secondly, the fact that the monitored signal is a decaying magnetic field rather than electric field at the surface, makes the results much less dependent on local resistivity conditions at the receiver site (Sternberg et.al., 1988). Distortions due to local resistivity inhomogeneities at the receiver site can be a severe problem in DC-soundings as well as in MT-soundings (Knútur Árnason, 1984).

Thirdly the central loop TEM method is less sensitive to lateral resistivity variations than the DC methods. The reason is that in the TEM method, the current induced in the ground can be visualized as a diffuse current ring which at early times, after the current in the transmitter loop is turned off, simulates the current in the transmitter loop (Hoversten and Morrison, 1982). As time goes on this current ring diffuses downwards and outwards resulting in increasing depth of penetration with time. The monitored signal at the surface is primarily dependent on the resistivity structure inside the diffuse current ring. The central loop TEM-soundings are thus much more downwards focused than the DC-soundings, where increased

depth of penetration is obtained by increasing the distance between the current injection electrodes and the receiver dipole, making the monitored electric field dependent on a much larger volume of rocks. One-dimensional inversion is therefore better justified in the interpretation of central loop TEM-soundings than in DC-soundings. Experience from geothermal resistivity surveys carried out in Iceland has shown that one dimensional interpretation of central loop TEM-soundings can give basically the same resolution as the much more time consuming and expensive two-dimensional modeling of DC data (Knútur Árnason et.al., 1987).

One more important difference between central loop TEM -soundings and DC-soundings is worth mentioning. In DC-soundings the monitored signal (the voltage difference) is low when the subsurface resistivity is low. It can therefore be very hard to obtain reliable DC data in geothermal areas where resistivity is very low because at a certain level the signal is drowned in telluric noise. In TEM-soundings the situation is the opposite, the lower the resistivity the stronger the signal.

The objective of this report is to discuss the principles of the central loop transient electro-magnetic method over horizontally stratified earth and to describe a one-dimensional inversion program for inverting the sounding results into horizontally layered resistivity models. An effort is made to keep the discussion self-contained and leave as few loose ends as possible. The discussion in this report is rather focused on the specific problem at hand. A more general discussion on electromagnetic methods in geophysical applications can e.g. be found in Nabigham (1988) and Kaufman and Keller (1983).

Chapters 2 and 3 deal with the physical basis, namely the electro-magnetic coupling between a circular current loop at the surface of a horizontally stratified earth and a coil at the centre of the loop. The problem is attacked backwards, so to speak, because we start, in chapter 2, by considering an oscillating magnetic dipole source at the surface and determine the electro-magnetic field due to this source.

In chapter 3 the electric field generated by the dipole is used to determine the mutual coupling, in the frequency domain, between the dipole and a loop with the dipole at the centre (central loop configuration) and also the coupling between a grounded electrical dipole and the magnetic dipole (grounded-dipole-coil configuration). The principle of reciprocity is then used to interchange the roles of the transmitter and the receiver. This indirect approach is used in order to keep the discussion as simple as possible and to get hold on the necessary formulas with minimum effort. The frequency domain coupling is then used to determine the coupling in the time domain for an arbitrary transmitted current function. The transient response in the receiver coil due to a sharp and a ramped step function current in the transmitter is worked out. The sharp step response for a homogeneous half-space is then used to introduce late time apparent resistivity. This is a useful representation of the transient response because for a homogeneous earth or layered earth with thick layers the late time apparent resistivity approaches the true resistivity for late times.

Chapter 4 discusses a computational technique, called the digital filter method, which is used to perform numerical integrations. The principles of the method are discussed in some details as well as methods for determining the filter coefficients. The digital filter method is a powerful tool that enables us to perform the extensive integrations that are necessary in order to work out the transient response of a layered earth. Computational speed in these numerical integrations is of outmost importance because they are the most time consuming calculations in the inversion program described in chapter 6.

Chapter 5 is devoted to inversion theory. The iterative non-linear least-squares inversion method is discussed and the Levenberg-Marquardt modification. Probability distributions and confidence limits for the model parameters resulting from the inversion are discussed. This chapter also discusses the problem of inverting singular or nearly singular matrices and the use of singular value decomposition to solve the problem.

A one-dimensional inversion program for central loop transient electro-magnetic soundings is described in chapter 6. The program is a non-linear least-squares inversion program of the Levenberg-Marquardt type and is an implementation of the concepts and formalism developed in chapters 2 through 5. The program is written in standard FORTRAN 77 and can be implemented on any machine supporting FORTRAN 77. An effort was made to make the program as fast as possible without loosing accuracy. It runs with an acceptable speed on personal computers with 8087 coprocessor but is un-practically slow without the coprocessor.

2. THE ELECTROMAGNETIC FIELD DUE TO AN OSCILLATING VERTICAL MAGNETIC DIPOLE OVER A HORIZONTALLY STRATIFIED EARTH.

In this section we will study the electromagnetic field generated by an oscillating vertical magnetic dipole over a horizontally stratified earth where each layer (stratum) is electrically homogeneous and isotropic. By this we mean that the conductivity (σ), dielectric permittivity (ε) and magnetic permeability (μ), which in the most general case are tensor functions of space coordinates, are scalars and constant in each layer.

The source dipole can be thought of as a small horizontal coil at the point $\boldsymbol{x}_0 = (x_0,y_0,z_0)$ carrying an alternating current $I = I_0\, e^{i\omega t}$, where ω is the angular frequency in radians per second. The magnetic dipole moment can be written as:

$$\boldsymbol{M} = M_0\, e^{i\omega t}\, \hat{\boldsymbol{z}}\, \delta(\boldsymbol{x} - \boldsymbol{x}_0) \tag{2.1}$$

where $\hat{\boldsymbol{z}}$ is a unit vector along the z-axis (parallel to the coil axis) and $M_0 = I_0\, n_s A_s$ where n_s is the number of windings, A_s is the cross-sectional area of the coil and $\delta(\boldsymbol{x} - \boldsymbol{x}_0)$ is the Dirac's delta "function".

We will state the Maxwell's equations that govern the electromagnetic field and introduce magnetic vector and scalar potentials as a computational aid to reformulate and solve the governing equations. The boundary conditions that the electromagnetic field and the potentials have to satisfy at layer boundaries and at the earth's surface will be stated. Finally we determine the electromagnetic field in the air and in each layer of the earth.

2.1 MAXWELL'S EQUATIONS AND MAGNETIC POTENTIALS.

The electromagnetic field caused by a magnetic dipole source, $\boldsymbol{M}$, is described by Maxwell's equations:

$$\nabla \times \boldsymbol{H} = \sigma \boldsymbol{E} + \frac{\partial \boldsymbol{D}}{\partial t} \tag{2.2}$$

$$\nabla \times \boldsymbol{E} = -\mu \frac{\partial \boldsymbol{H}}{\partial t} - \mu \frac{\partial \boldsymbol{M}}{\partial t} \tag{2.3}$$

$$\nabla \cdot \boldsymbol{D} = \eta \tag{2.4}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2.5}$$

where σ is the *electric conductivity* (S/m=1/ohmm in mks units), η is the *charge density* (coulombs/m^3 in mks units), $\boldsymbol{E}$ is the *electric field intensity* (volts/m in mks units) and $\boldsymbol{D}$ is the *electric displacement vector* (coulombs/ m^2 in mks units). $\boldsymbol{E}$ and $\boldsymbol{D}$ are related through the identity:

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \tag{2.6}$$

where ε is the *dielectric permittivity* (farad/m in mks units). $\boldsymbol{M}$ is the source *magnetic dipole moment* (ampere-turn/m in mks units), $\boldsymbol{H}$ is the *magnetic field intensity* (ampere-turn/m in mks units) and is related to the *magnetic induction, or flux density*, $\boldsymbol{B}$ (Weber/m^2 in mks units), through the identity:

$$\boldsymbol{B} = \mu(\boldsymbol{H} + \boldsymbol{M}) \tag{2.7}$$

where μ is the *magnetic permeability* (henry/m in mks units). In source free regions this relation simplifies to $\boldsymbol{B} = \mu \boldsymbol{H}$.

In the case we are discussing (a magnetic dipole source over a horizontally layered earth with homogeneous and isotropic layers) there is no charge density inside each layer, but there may be charge accumulation at the boundaries between layers. We will later see, that in the case of a vertical magnetic dipole source, there is no charge accumulation at layer boundaries. Inside each layer, and in the air, we can therefore by use of (2.6) write (2.4) in the form:

$$\nabla \cdot \boldsymbol{E} = 0 \ . \tag{2.4a}$$

The Maxwell's equations (2.2) and (2.3) are six coupled dynamical equations for the six components of the electric and magnetic vector fields. By a dynamical equation we mean an equation that relates time derivatives (evolution in time) to the momentary field distribution in space. The remaining Maxwell's equations, (2.4) or (2.4a) and (2.5), are not dynamical equations but should be considered as constraint equations relating the components of the vector fields. Since we have two constraint equations we must be able to describe the electromagnetic field by four field components obeying four dynamical equations instead of six. The standard way of achieving this is to introduce a three component vector potential and a scalar potential. In our case it is most natural to use vector and scalar potentials of the magnetic type.

Inspired by equation (2.4a) we introduce a magnetic vector potential $\boldsymbol{A}$ such that:

$$\boldsymbol{E} = \nabla \times \boldsymbol{A} \ . \tag{2.8}$$

Substituting this in (2.2) and rearranging a bit gives:

$$\nabla \times (\boldsymbol{H} - \sigma \boldsymbol{A} - \varepsilon \frac{\partial \boldsymbol{A}}{\partial t}) = 0 \tag{2.9}$$

which means that the quantity in the parenthesis can be written as a gradient of a scalar potential ϕ. We can therefore write:

$$\boldsymbol{H} = \sigma \boldsymbol{A} + \varepsilon \frac{\partial \boldsymbol{A}}{\partial t} - \nabla \phi \ . \tag{2.10}$$

The way we have introduced the magnetic vector and scalar potentials indicates that they are

to a certain extent arbitrary, that is to say, given $\boldsymbol{E}$ and $\boldsymbol{H}$, solutions to the Maxwell's equations, do not define $\boldsymbol{A}$ and ϕ uniquely. It is easy to verify from equations (2.8) and (2.10) that the electric and magnetic fields are unchanged under the transformation:

$$\boldsymbol{A} \rightarrow \boldsymbol{A}' = \boldsymbol{A} + \nabla\chi \tag{2.11a}$$

$$\phi \rightarrow \phi' = \phi + \sigma\chi + \varepsilon\frac{\partial\chi}{\partial t} \tag{2.11b}$$

where χ is an arbitrary differentiable function. The transformations described by equations (2.11a-b) are called gauge transformations and the invariance of the physically measurable fields, $\boldsymbol{E}$ and $\boldsymbol{H}$, under these transformations is called gauge invariance. The gauge invariance means that we still have some freedom to choose the magnetic vector and scalar potentials and this will be handy later.

Let us now establish the dynamical equations that $\boldsymbol{A}$ and ϕ have to satisfy. We used two of the Maxwell's equations, (2.2) and (2.4a), to relate the electric and magnetic fields to the potentials and will now use the remaining two to get dynamical equations for the potentials. By substituting (2.8) and (2.10) in (2.3) and using the identity $\nabla\times(\nabla\times\boldsymbol{A}) = \nabla(\nabla\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A}$ we get:

$$\nabla(\nabla\cdot\boldsymbol{A}) - \nabla^2\boldsymbol{A} + \mu\sigma\frac{\partial\boldsymbol{A}}{\partial t} + \mu\varepsilon\frac{\partial^2\boldsymbol{A}}{\partial t^2} - \nabla\frac{\partial\phi}{\partial t} = -\mu\frac{\partial\boldsymbol{M}}{\partial t}\ . \tag{2.12}$$

This looks like a dynamical equation for $\boldsymbol{A}$ (it has a second derivative with respect to time) but it contains the scalar potential ϕ. This we can repair because the gauge invariance gives us freedom to choose the potentials in such a way that ϕ drops out of (2.12). By demanding that the potentials satisfy the constraint equation or gauge condition:

$$\nabla\cdot\boldsymbol{A} = \mu\frac{\partial\phi}{\partial t} \tag{2.13}$$

equation (2.12) simplifies to:

$$\nabla^2\boldsymbol{A} - \mu\sigma\frac{\partial\boldsymbol{A}}{\partial t} - \mu\varepsilon\frac{\partial^2\boldsymbol{A}}{\partial t^2} = \mu\frac{\partial\boldsymbol{M}}{\partial t} \tag{2.14}$$

which is a dynamical equation for the vector potential $\boldsymbol{A}$.

To see explicitly that we can indeed impose the gauge condition (2.13) on the potentials, let us assume that the electromagnetic field is described by known potentials $\boldsymbol{A}$ and ϕ not satisfying (2.13). We want to determine a function χ such that, after a gauge transformation described by (2.11), the primed (transformed) potentials satisfy (2.13), that is:

$$\nabla\cdot\boldsymbol{A}' - \mu\frac{\partial\phi'}{\partial t} = \left(\nabla\cdot\boldsymbol{A} - \mu\frac{\partial\phi}{\partial t}\right) + \nabla^2\chi - \mu\sigma\frac{\partial\chi}{\partial t} - \mu\varepsilon\frac{\partial^2\chi}{\partial^2 t} = 0 \tag{2.13a}$$

This is a second order inhomogeneous differential equation which can be solved for χ because $\nabla \cdot \boldsymbol{A} - \mu \frac{\partial \phi}{\partial t}$ is a known function of space and time.

In order to get a dynamical equation for ϕ we use the last Maxwell's equation (2.5) which by use of (2.7) can be written in the form:

$$\nabla \cdot \boldsymbol{H} = - \nabla \cdot \boldsymbol{M} \tag{2.5a}$$

Substituting (2.10) for the magnetic field in (2.5a) and by using the gauge condition (2.13) we get:

$$\nabla^2 \phi - \mu\sigma \frac{\partial \phi}{\partial t} - \mu\varepsilon \frac{\partial^2 \phi}{\partial t^2} = \nabla \cdot \boldsymbol{M} \tag{2.15}$$

which is a dynamical equation for the scalar potential ϕ.

2.2 BOUNDARY CONDITIONS.

It is not enough to have the dynamical equations that the fields and potentials satisfy. In order to get the solutions to these equations we must determine the boundary conditions that they must satisfy. The boundary conditions for the electromagnetic field at the interfaces between layers and at the surface can be determined from Maxwell's equations (2.2)-(2.5).

Normal component of the *E* field: The conditions for the component of the electric field normal to the interfaces can be determined from (2.4). We consider a small "pill-box" volume with the larger surface areas, Δa, parallel to the interface between domains 1 and 2 with the electric parameters $\sigma_1, \varepsilon_1, \mu_1$ and $\sigma_2, \varepsilon_2, \mu_2$ as shown in figure 2.1. The "pill-box" is of thickness Δh and is cut in the middle by the interface. Integrating equation (2.4) over the volume of the "pill-box" we get:

$$\int_V \nabla \cdot \boldsymbol{D} dv = \int_S \boldsymbol{D} \cdot d\boldsymbol{s} = \int_V \eta dv = q \tag{2.16}$$

where Gauss divergence theorem has been used to write the volume integral of $\nabla \cdot \boldsymbol{D}$ as the integral of the flux of the $\boldsymbol{D}$ field out through the surface and q is the total charge inside the "pill-box". Since the "pill-box" is small, the field vector $\boldsymbol{D}$ is practically constant on each side and we can write:

$$\boldsymbol{n} \cdot (\boldsymbol{D}_1 - \boldsymbol{D}_2) \cdot \Delta a + \textit{term of order } \Delta h = q. \tag{2.17}$$

where $\boldsymbol{n}$ is an upwards pointing (into domain 1) unit normal to the interface. The first term on the left hand side of (2.17) is the total flux of the $\boldsymbol{D}$ field through the sides of the "pill-box" and the second term is the flux trough the edges. Dividing through (2.17) with Δa and letting the thickness of the "pill-box" tend to zero ($\Delta h \rightarrow 0$) this gives:

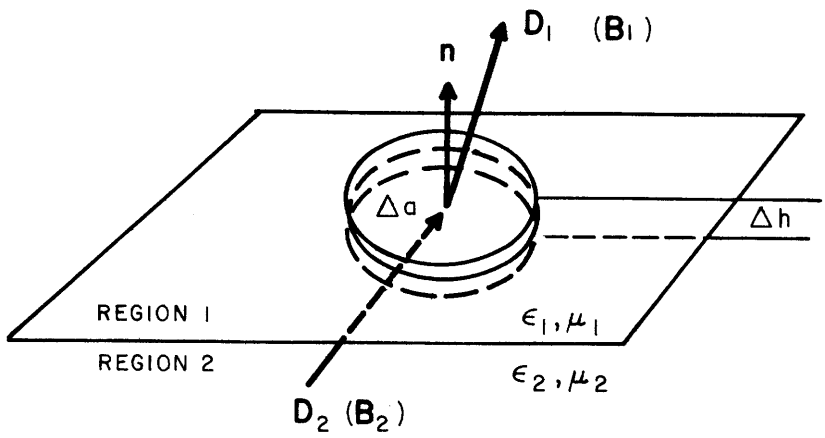

Figure 2.1 Boundary conditions for normal components of the E and H fields

$$n\cdot(D_1 - D_2) = n\cdot(\varepsilon_1 E_1 - \varepsilon_2 E_2) = \nu \tag{2.18}$$

where ν is the *surface charge density* at the interface (coulomb/m^2 in mks units). This equation states that the discontinuity in the D field across the interface is equal to the surface charge density at the interface.

Normal component of the H field: Starting from Maxwell's equation (2.5) and going through exactly the same procedure as for the normal component of the electric field we find that:

$$n\cdot(B_1 - B_2) = n\cdot(\mu_1 H_1 - \mu_2 H_2) = 0 \tag{2.19}$$

which states that the normal component of the B field is continuous a cross the interface.

Tangential component of the E field: The boundary conditions for the tangential components of the electric field can be deduced from equation (2.3). By using (2.7) in the right hand side of (2.3) and integrating both sides over the area of a small elongated rectangle cutting the interface and with the sides, of length Δl, parallel to the interface and the ends, of length Δh, perpendicular to the interface as shown in figure 2.2 we get:

$$\int_S \nabla\times E\cdot ds = \int_\Gamma E\cdot dl = -\int_S \frac{\partial B}{\partial t}\cdot ds = -\frac{\partial \Phi}{\partial t}. \tag{2.20}$$

In the second equality we have made use of Stokes theorem stating that a surface integral of the normal component of the curl of a vector field over a surface can be written as a line integral of the tangential component of the same vector field around the oriented contour , Γ, bounding the surface. Furthermore we have introduced the total magnetic flux, Φ, through the surface. Since the rectangle is small, the electric field is practically constant along the

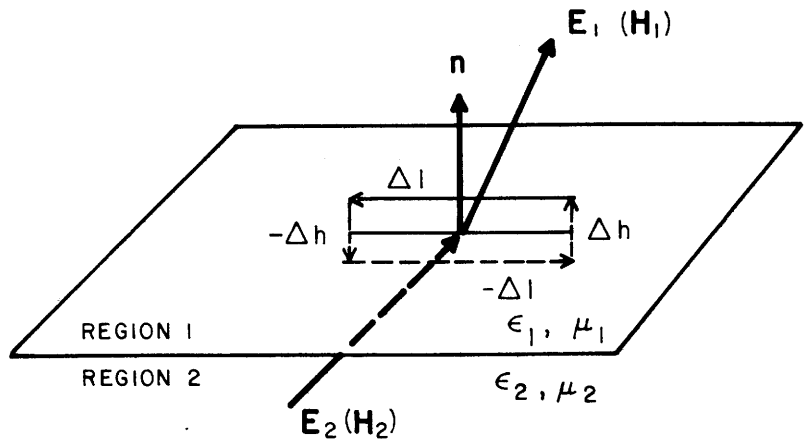

Figure 2.2 Boundary conditions for tangential components of the E and H fields

sides and since the magnetic flux density is finite at the boundary we can write this in the form:

$$\Delta l \cdot (E_1 - E_2) + \textit{term of order } \Delta h = \textit{term of order } \Delta h \Delta l \tag{2.21}$$

where Δl is the vector along the side of the rectangle in domain 1 and in the oriented (counter clockwise) direction. The first term on the left hand side of (2.21) is the difference between the components of the E field along the sides of the rectangle. The second term is a contribution from the ends and the right hand side is the total magnetic flux trough the rectangle. In the limit $\Delta h \rightarrow 0$ equation (2.21) reduces to:

$$\Delta l \cdot (E_1 - E_2) = 0 \tag{2.22}$$

Since the orientation of the rectangle was arbitrarily chosen, this means that the tangential component of the electric field in any direction is continuous over the interface. This can be conveniently expressed by the equation:

$$n \times (E_1 - E_2) = 0 \tag{2.22a}$$

where n is a unit normal to the interface and pointing into domain 1.

Tangential component of the H field: To get boundary conditions for the tangential component of the magnetic field we use Maxwell's equation (2.2). The right hand side of (2.2) expresses the *current density* (coulombs/m^2) which is a sum of the *conduction current density* and the *displacement current density*:

$$J = J_{cond} + J_{dis} \tag{2.23}$$

$$J_{cond} = \sigma E \ ; \ J_{dis} = \frac{\partial D}{\partial t}$$

By integrating (2.2) over the same rectangle as we did for the tangential component of the electric field we get:

$$\int_S \nabla \times H \cdot ds = \int_\Gamma H \cdot dl = \int_S J \cdot ds \tag{2.24}$$

Since the conductivity and the electric field are finite in domains 1 and 2 and at the interface the current density trough the rectangle is finite and (2.24) can be written as:

$$\Delta l \cdot (H_1 - H_2) + \textit{term of order } \Delta h = \textit{term of order } \Delta h \Delta l \tag{2.25}$$

where Δl is the same as in (2.21). In the limit $\Delta h \rightarrow 0$ equation (2.25) reduces to:

$$\Delta l \cdot (H_1 - H_2) = 0 \tag{2.26}$$

which means that the tangential component of the H field is continuous across the interface. This can also be expressed by the the equation:

$$n \times (H_1 - H_2) = 0 \tag{2.26a}$$

where n is a unit normal to the interface and pointing into domain 1.

By inserting (2.8) and (2.10) into the boundary conditions (2.18), (2.19), (2.22a) and (2.26a) for the electromagnetic field vectors we get the following boundary conditions for the potentials A and ϕ:

$$n \cdot (\varepsilon_1 \nabla \times A_1 - \varepsilon_2 \nabla \times A_2) = \nu \tag{2.27}$$

$$n \cdot (\mu_1 \sigma_1 A_1 + \mu_1 \varepsilon_1 \frac{\partial A_1}{\partial t} - \mu_1 \nabla \phi_1) = n \cdot (\mu_2 \sigma_2 A_2 + \mu_2 \varepsilon_2 \frac{\partial A_2}{\partial t} - \mu_2 \nabla \phi_2) \tag{2.28}$$

$$n \times (\nabla \times A_1) = n \times (\nabla \times A_2) \tag{2.29}$$

$$n \times (\sigma_1 A_1 + \varepsilon_1 \frac{\partial A_1}{\partial t} - \nabla \phi_1) = n \times (\sigma_2 A_2 + \varepsilon_2 \frac{\partial A_2}{\partial t} - \nabla \phi_2) \tag{2.30}$$

2.3 SOLUTION FOR AN OSCILLATING VERTICAL MAGNETIC DIPOLE OVER AN N-LAYERED EARTH.

The discussion so far has been rather general but let us now focus on the specific problem at hand, namely an oscillating vertical magnetic dipole above the surface of horizontally stratified earth. We consider the case of N layers with electrical parameters and thicknesses ε_i, μ_i, σ_i and d_i for $i = 1,...,N$. The thickness of the N'th layer is infinite and the depth to the

i'th layer is $h_i = \sum_{j=1}^{j=i-1} d_j$. For the air we take the dielectric permittivity and magnetic permeability to be equal to that of vacuum (ϵ_0, μ_0) and the conductivity to be zero. Resistivity structures like those we are considering here are often referred to as one dimensional resistivity structures because the resistivity only varies in one direction namely with depth.

We introduce a Cartesian coordinate system with the origin and the x-y plane at the earth's surface and the z-axis pointing downwards. Let the oscillating magnetic dipole be on the z-axis and at the height $|z_0|$ above the surface, that is to say the source coil is at the point $x_0 = (0,0,z_0)$, $(z_0 < 0)$, and with its axis parallel to the z-axis (see figure 2.3).

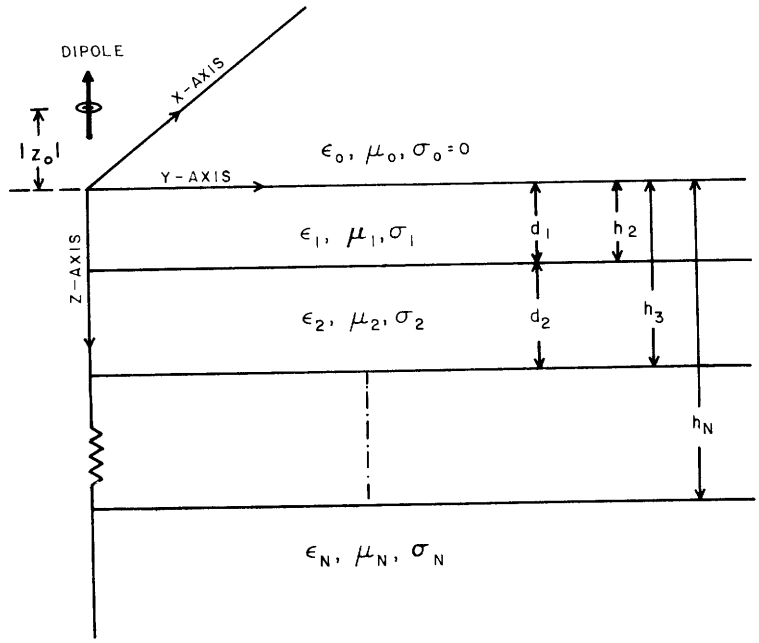

Figure 2.3 An N layered half-space

From the expression for the magnetic dipole source

$$\boldsymbol{M} = M_0 e^{i\omega t} \hat{\boldsymbol{z}} \delta(\boldsymbol{x} - \boldsymbol{x}_0) \tag{2.1}$$

we see that it has only a z-component. This means that, of the three differential equations for the components of the vector potential expressed by (2.14), only the equation for the z-component is nonhomogeneous. The equations for the x- and y-components are

homogeneous. This suggest that it might be sufficient to look for a vector potential with only a z-component. Because the electric and the magnetic fields have axial symmetry with respect to the z-axis it can be seen that the electric field will have no z-component and from equation (2.8) we see that this is trivially satisfied if the vector field has only a z-component. Therefore we assume that in each layer and in the air:

$$A_i = (0,0,A_i) \; ; \; i = 0,1,\ldots,N \tag{2.31}$$

(the zeroth layer is the air).

The boundary conditions for the potentials stated in equations (2.27)-(2.30) are greatly simplified by this. The left hand side of (2.27) is identically zero which means that there is no surface charge generated at the interfaces between layers. For the boundary between the layers i and $i + 1$ the remaining three conditions now become:

$$\mu_i(\sigma_i A_i + \varepsilon_i \frac{\partial A_i}{\partial t} - \frac{\partial \phi_i}{\partial z}) = \mu_{i+1}(\sigma_{i+1} A_{i+1} + \varepsilon_{i+1} \frac{\partial A_{i+1}}{\partial t} - \frac{\partial \phi_{i+1}}{\partial z}) \tag{2.28a}$$

$$\frac{\partial A_i}{\partial x} = \frac{\partial A_{i+1}}{\partial x} \; ; \; \frac{\partial A_i}{\partial y} = \frac{\partial A_{i+1}}{\partial y} \tag{2.29a}$$

$$\frac{\partial \phi_i}{\partial x} = \frac{\partial \phi_{i+1}}{\partial x} \; ; \; \frac{\partial \phi_i}{\partial y} = \frac{\partial \phi_{i+1}}{\partial y} \tag{2.30a}$$

(In the coordinate system that we have chosen, the upwards pointing normal in the boundary conditions (2.28)-(2.30) is given as $\boldsymbol{n} = (0,0,-1)$)

Equations (2.29a) and (2.30a) can be integrated over the boundary plane and since the potentials are only defined up to an additive constant we can take:

$$A_i = A_{i+1} \tag{2.31}$$

$$\phi_i = \phi_{i+1} \tag{2.32}$$

Things can be further simplified by making use of the simple time-dependence of the magnetic dipole source. Since the time-dependence of the source is given by $e^{i\omega t}$, all the fields and potentials will have this same dependence on time, that is:

$$\boldsymbol{E} = \boldsymbol{E}' e^{i\omega t} \; ; \; \boldsymbol{H} = \boldsymbol{H}' e^{i\omega t} \; ; \; A = A' e^{i\omega t} \; ; \; \phi = \phi' e^{i\omega t} \tag{2.33}$$

where the primed quantities are only functions of space coordinates. In the following we will use the primed and unprimed fields and potentials interchangeably, omitting the prime, and it will be clear from the context which quantity is meant.

Because of this simple time-dependence we can use the gauge condition to eliminate the potential ϕ from the problem, because from (2.13) we see that:

$$\phi = \frac{1}{i\mu\omega}\frac{\partial A}{\partial z} \tag{2.34}$$

and using this in the boundary condition (2.32) we find:

$$\frac{\partial A_i}{\partial z} = \frac{\partial A_{i+1}}{\partial z} \tag{2.35}$$

The source dipole is in the air above the surface (where $\sigma = 0$) so the differential equation for the z-component of the vector potential in the air is (see (2.14)):

$$\nabla^2 A_0 + k_0^2 A_0 = i\omega\mu_0 M \tag{2.36}$$

$$k_0^2 = \omega^2\mu_0\varepsilon_0 \; ; \; M = M_0\delta(x,y,z-z_0)$$

and since there is no source in the earth, the differential equation in the i'th layer is:

$$\nabla^2 A_i + k_i^2 A_i = 0 \tag{2.37}$$

$$k_i^2 = \omega^2\mu_i\varepsilon_i - i\omega\mu_i\sigma_i \; ; \; i = 1, \ldots, N$$

Now we have managed to reduce the problem of finding the electromagnetic field due to a harmonically oscillating vertical magnetic dipole over layered earth to the problem of solving the relatively simple time independent equations (2.36) and (2.37) for the z-component of the vector potential subject to the boundary conditions (2.31) and (2.35). When the vector potential has been determined we find the electric and magnetic fields from equations (2.8) and (2.10) where in the latter the scalar potential is eliminated by using (2.34).

We turn now to the task of solving the differential equations (2.36) and (2.37). Let us start with (2.36). The general solution to an inhomogeneous linear differential equation is the sum of the general solution of the homogeneous equation and a particular solution to the inhomogeneous equation. The homogeneous part of (2.36) is formally identical to the equations (2.37) in the layers of the earth and we will later solve them all together. Let us therefore start by finding the particular solution of the inhomogeneous equation (2.36) by using Fourier expansion. Let the solution $f(\boldsymbol{x})$ have the Fourier expansion:

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}}\int \tilde{f}(\boldsymbol{\xi})e^{i\boldsymbol{x}\cdot\boldsymbol{\xi}}d\boldsymbol{\xi} \tag{2.38}$$

where $\tilde{f}(\boldsymbol{\xi})$ is the Fourier transform of $f(\boldsymbol{x})$ defined by:

$$\tilde{f}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^{3/2}}\int f(\boldsymbol{x})e^{-i\boldsymbol{\xi}\cdot\boldsymbol{x}}d\boldsymbol{x} \tag{2.39}$$

The Fourier expansion (2.38) can be thought of as a linear expansion of $f(\boldsymbol{x})$ in terms of the infinite set of basis functions $\beta_{\boldsymbol{\xi}}(\boldsymbol{x}) = e^{i\boldsymbol{x}\cdot\boldsymbol{\xi}}/(2\pi)^{3/2}$ where the Fourier transform $\tilde{f}(\boldsymbol{\xi})$ is the expansion coefficients. We are here thinking of the space of integrable complex valued functions as an infinite dimensional inner product space where the inner product of the

ordered pair of functions $f(x)$ and $g(x)$ is given as:

$$<f \mid g> = \int f(x)g^*(x)dx \tag{2.40}$$

where * denotes complex conjugation. We see that the Fourier transform, $\tilde{f}(\xi)$, of the function $f(x)$ is given as the inner product with (projection along) the basis function $\beta_\xi(x)$:

$$\tilde{f}(\xi) = <f \mid \beta_\xi> . \tag{2.41}$$

The basis functions are "orthonormal" in the sense that

$$\frac{1}{(2\pi)^3}\int e^{ix\cdot\xi}e^{-ix\cdot\xi'}dx = \frac{1}{(2\pi)^3}\int e^{i(\xi-\xi')\cdot x}d\xi = \delta(\xi-\xi') \tag{2.42}$$

where $\delta(\xi-\xi')$ is Dirac's delta "function" with the property:

$$\delta(\xi-\xi') = 0 \text{ for } \xi \neq \xi' \; ; \int f(\xi)\delta(\xi-\xi')d\xi = f(\xi') \; ; \int \delta(\xi-\xi')d\xi = 1 . \tag{2.43}$$

By interchanging ξ and x in (2.42) we see that the Fourier transform of the delta "function" $\delta(x-x') \equiv \delta_{x'}(x)$, considered as a function of x, is given by:

$$\tilde{\delta}_{x'}(\xi) = \frac{1}{(2\pi)^{3/2}}e^{-ix'\cdot\xi} . \tag{2.44}$$

By differentiation on both sides of (2.38) we see that the Fourier transform of $\nabla f(x)$ is given by $i\xi\tilde{f}(\xi)$ and the transform of $\nabla^2 f(x)$ is given by $-\xi^2\tilde{f}(\xi)$. It is therefore seen that the Fourier transform converts differentiation into multiplication of the Fourier transform by a factor $i\xi$.

The usefulness of Fourier transforms in solving differential equations is based on this property of turning differentiations into algebraic manipulations. By putting the Fourier expansion of the solution (2.38), with unknown Fourier transform, into the differential equation and using the "orthonormality" of the basis functions to demand that the expansion functions on both sides are equal, we convert the differential equation into an algebraic equation for the Fourier transform. The resulting algebraic equation can then be solved for the Fourier transform and the solution is given by substituting the result into the Fourier expansion (2.38).

If we insert the Fourier expansion (2.38) into the inhomogeneous differential equation (2.36) we see that the Laplacian gives a multiplication factor $-\xi^2$. By using that the Fourier transform of the delta "function" on the right hand side of (2.36) is given by (2.44) (we write for the moment $(0,0,z_0) = x_0$), we get:

$$\frac{1}{(2\pi)^{3/2}}\int \tilde{f}(\xi)(-\xi^2 + k_0^2)e^{ix\cdot\xi}d\xi = \frac{i\omega\mu_0 M_0}{(2\pi)^3}\int e^{-ix_0\cdot\xi}e^{ix\cdot\xi}d\xi . \tag{2.45}$$

Since the basis functions $\frac{1}{(2\pi)^{3/2}}e^{ix\cdot\xi}$ are "orthonormal", the expansions functions on both

side must be equal for each value of ξ which means that:

$$\tilde{f}(\xi)(-\xi^2 + k_0^2) = \frac{i\omega\mu_0 M_0}{(2\pi)^{3/2}} e^{-ix_0\cdot\xi} . \tag{2.46}$$

This we solve for the Fourier transform and insert into the Fourier expansion (2.38) to get a formal expression for the particular solution to the inhomogeneous equation:

$$f(x) = \frac{-i\omega\mu_0 M_0}{(2\pi)^3} \int \frac{e^{i(x-x_0)\cdot\xi}}{\xi^2 - k_0^2} d\xi . \tag{2.47}$$

Because of the axial symmetry of the problem with respect to the z-axis, it is convenient to work in cylindrical rather than Cartesian coordinates. Let us therefore reexpress the solution (2.47) in cylindrical coordinates by writing:

$$x = (x,y,z) = (r\cos\theta, r\sin\theta, z) \tag{2.48a}$$

$$\xi = (\eta,\varsigma,\tau) = (\lambda\cos\psi, \lambda\sin\psi, \tau) \tag{2.48a}$$

where r and λ are the lengths of the projections of the vectors x and ξ into the xy-plane and θ and ψ are the angels between the projections and the x-axis (measured in the positive or counterclockwise direction). The volume element of integration is in cylindrical coordinates expressed by $d\xi = \lambda d\psi d\lambda d\tau$ and since $x_0 = (0,0,z_0)$, (2.47) becomes:

$$f(r,\phi,z) = \frac{-i\omega\mu_0 M_0}{(2\pi)^3} \int_0^{2\pi}\int_0^{\infty}\int_{-\infty}^{\infty} \frac{e^{i\lambda r\cos(\psi-\theta)} e^{i\tau(z-z_0)}}{\lambda^2 + \tau^2 - k_0^2} \lambda d\tau d\lambda d\psi . \tag{2.49}$$

From this equation it is easily seen that this particular solution is independent of the azimuth angle θ because we can change the integration variable ψ to $\psi' = \psi - \theta$, making the integral independent of θ without changing its value. This is of course only a manifestation of the cylindrical symmetry of the problem. By using the Sommerfeld's representation of the Bessel function of the zeroth order, J_0 :

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{ix\cos\psi} d\psi \tag{2.50}$$

and by writing:

$$\tau^2 + \lambda^2 - k_0^2 = (\tau + i\sqrt{\lambda^2 - k_0^2})(\tau - i\sqrt{\lambda^2 - k_0^2}) \tag{2.51}$$

(2.49) can be written as:

$$f(r,z) = \frac{-i\omega\mu_0 M_0}{(2\pi)^2} \int_0^{\infty}\int_{-\infty}^{\infty} \frac{e^{i\tau(z-z_0)}}{(\tau + i\sqrt{\lambda^2 - k_0^2})(\tau - i\sqrt{\lambda^2 - k_0^2})} J_0(\lambda r)\, \lambda d\tau d\lambda . \tag{2.52}$$

The integral in (2.52) can be further simplified by using Cauchy's integral formula to carry out the integration over τ. Cauchy's integral formula states that if $g(z)$ is an analytic function (with no poles) of the complex variable, $z = x + iy$, in a simply connected domain C, then for any point a and any closed path Γ in C :

$$\int_{\Gamma} \frac{g(z)}{(z-a)} dz = \begin{cases} 2\pi i\, g(a) & \text{if } a \text{ iside } \Gamma \\ 0 & \text{if } a \text{ outside } \Gamma \end{cases} \tag{2.53}$$

where Γ is oriented in the positive (counterclockwise) direction. In order to be able to use Cauchy's integral formula we must express (2.52) as an integral over a closed path in the complex plane. From (2.52) we see that for $z - z_0 > 0$ the numerator tends exponentially to zero if we take τ complex with positive imaginary part tending to infinity and for $z - z_0 < 0$ the numerator tends to zero if we take τ complex with negative imaginary part tending to infinity. The integration with respect to τ can therefore be carried out by considering the integral over the real axis as a part of a closed integration contour in the complex τ-plane.

For $z - z_0 > 0$ we close the integration path with a half circle in the upper half plane and let the radius tend to infinity. From the considerations above we see that the contribution to the integral along the half circle tends to zero as the radius tends to infinity. This integration contour has positive orientation and the integrand of (2.52) has the same form as the one in (2.53) with the singularity $a = i\sqrt{\lambda^2 - k_0^2}$ inside the integration contour.

For $z - z_0 < 0$ we close the integration path with a half circle in the lower half plane of the complex τ-plane. This integration contour encloses and picks up the singularity $a = -i\sqrt{\lambda^2 - k_0^2}$ but the contour is oriented in negative direction which gives an extra minus sign. The result for both of these cases can be written in the form:

$$f(r,z) = \frac{-i\omega\mu_0 M_0}{4\pi} \int_0^{\infty} \frac{e^{-\sqrt{\lambda^2 - k_0^2}\,|z - z_0|}}{\sqrt{\lambda^2 - k_0^2}} J_0(\lambda r)\, \lambda\, d\lambda \,. \tag{2.54}$$

(A careful reader may be worried that we get in trouble when $\lambda^2 \leq k_0^2$ because then the poles might sit on the real axis, but since the air actually has a finite but small conductivity, k_0^2 should be taken to have a small imaginary part (see (2.37)) which keeps the poles from the real axis).

We will in the forthcoming discussion use cylindrical coordinates so this is the form in which we will use the particular solution to the inhomogeneous equation (2.36) for the vector potential in the air. It is never the less instructive to work the solution out in spherical coordinates. The details will not be worked out here, but if we introduce spherical coordinates into the integral representation (2.47), the integration over the polar and azimuth angles can be carried out and the integration over the radial variable can be carried out by using Cauchy's integral formula in a similar way as we did above. The result is:

$$f(R) = \frac{-i\omega\mu_0 M_0}{4\pi}\frac{e^{-ik_0 R}}{R}\ ,\ R = \sqrt{x^2 + y^2 + (z - z_0)^2} \tag{2.55}$$

We have therefore established the identity:

$$\frac{e^{-ik_0\sqrt{r^2 + (z - z_0)^2}}}{\sqrt{r^2 + (z - z_0)^2}} = \int_0^\infty \frac{e^{-\sqrt{\lambda^2 - k_0^2}\,|z - z_0|}}{\sqrt{\lambda^2 - k_0^2}} J_0(\lambda r)\,\lambda\,d\lambda\ . \tag{2.56}$$

We will later make use of this formula.

Let us now turn to the homogeneous equations in the layers of the earth (2.37) and the homogeneous part of (2.36). Because of the cylindrical symmetry of the problem with respect to the z-axis, we work in cylindrical coordinates and demand that the solutions be independent of the azimuth angle. The homogeneous equations will there fore be on the form:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial A}{\partial r}\right) + \frac{\partial^2 A}{\partial z^2} + k^2 A = 0\,. \tag{2.57}$$

We will solve this equation by a method called separation of variables by finding a solution of the form $A(r,z) = F(r)G(z)$. Inserting this into (2.57), dividing by FG and rearranging a bit results in:

$$\frac{1}{F}\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial F}{\partial r}\right) = -k^2 - \frac{1}{G}\frac{\partial^2 G}{\partial z^2}\,. \tag{2.58}$$

The left hand side is only dependent on r while the right hand side is only dependent on z and this can only be satisfied if each side is equal to a constant which we call α. Because of this, equation (2.58) can, after minor algebraic manipulations, be written as two separate equations:

$$r\frac{\partial}{\partial r}\left(r\frac{\partial F}{\partial r}\right) - \alpha r^2 F = 0 \tag{2.59}$$

$$\frac{\partial^2 G}{\partial z^2} = (-\alpha - k^2)\,G\ . \tag{2.60}$$

By making the variable change, $r' = \sqrt{-\alpha}\,r = \lambda r$, equation (2.59) becomes Bessel's equation of order 0 and the general solution is:

$$F = A_\lambda J_0(\lambda r) + B_\lambda Y_0(\lambda r) \tag{2.61}$$

where J_0 and Y_0 are the Bessel functions of the first and second kind of order 0. Since we are looking for solutions which are finite at $r = 0$ and Y_0 is singular at this point we must demand that $B_\lambda = 0$. Furthermore we demand that the solution tends to zero when $r \to \infty$ because the electro-magnetic field tends to zero at infinite distance from the dipole source. This implies that λ is a real number (which we can take to be positive), because from the integral representation (2.50) it can be seen that $J_0(\lambda r)$ is finite for $r \to \infty$ only if λ is real

(and in that case it tends indeed to zero).

Since $\alpha = -\lambda^2$ equation (2.60) becomes:

$$\frac{\partial^2 G}{\partial z^2} = (\lambda^2 - k^2)G = m^2 G \ . \tag{2.62}$$

The general solution to this equation is:

$$G(z) = C_\lambda e^{mz} + D_\lambda e^{-mz} \tag{2.63}$$

$$m = \sqrt{\lambda^2 - k^2} \ .$$

For each $\lambda \geq 0$ a solution to the homogeneous equation, (2.57), is now obtained by multiplying (2.61) (with $B_\lambda = 0$) and (2.63) and a general solution is obtained by summing over λ, that is:

$$A(r,z) = \int_0^\infty [C(\lambda)e^{mz} + D(\lambda)e^{-mz}] J_0(\lambda r)\, d\lambda . \tag{2.64}$$

C and D, which are at this point arbitrary functions of λ, will be determined by the boundary conditions (2.31) and (2.35):

$$A_{i-1} = A_i \ ; \ \frac{\partial A_{i-1}}{\partial z} = \frac{\partial A_i}{\partial z} \ ; \ \text{at} \ z = h_i \ ; \ i = 1, \ldots, N . \tag{2.65}$$

The task of determining the unknown coefficient functions C and D is made easier by using the hyperbolic functions:

$$\cosh(z) = \frac{e^z + e^{-z}}{2} \ ; \ \sinh(z) = \frac{e^z - e^{-z}}{2} \tag{2.66}$$

in the solutions, instead of the exponential functions. By a simple redefinition of the coefficient functions C and D we can, instead of (2.64), write the solution in the i'th layer in the form:

$$A_i(r,z) = \frac{-i\omega\mu_0 M_0}{4\pi} \int_0^\infty [C_i(\lambda)\cosh(m_i(z - h_{i+1})) + D_i(\lambda)\sinh(m_i(z - h_{i+1}))] J_0(\lambda r)\, d\lambda$$

$$m_i = \sqrt{\lambda^2 - k_i^2} \ ; \ i = 1, \ldots, N-1 \ . \tag{2.67}$$

In the N'th layer (which is infinitely thick) we demand that the potential tends to zero when $z \to \infty$ which means (see (2.64)) that A_N must be exponentially decreasing with increasing z (z increases with depth). We take therefore:

$$A_N(r,z) = \frac{-i\omega\mu_0 M_0}{4\pi} \int_0^\infty D_N(\lambda)\, e^{-m_N(z-h_N)}\, J_0(\lambda r)\, d\lambda$$

$$m_N = \sqrt{\lambda^2 - k_N^2} \ . \tag{2.68}$$

In the air the potential is the sum of, (2.54), the particular solution to the inhomogeneous equation and, (2.64), the solution to the homogeneous equation. We demand that the potential A_0 tends to zero when $z \rightarrow -\infty$ (infinitely high above the ground). The particular solution (2.54) satisfies this condition and from (2.64) we see that the solution to the homogeneous equation can only contain a term with $e^{m_0 z}$. Therefore we write:

$$A_0(r,z) = \frac{-i\omega\mu_0 M_0}{4\pi} \int_0^\infty [\frac{\lambda}{m_0} e^{-m_0|z-z_0|} + C_0(\lambda)\, e^{m_0 z}]\, J_0(\lambda r)\, d\lambda$$

$$m_0 = \sqrt{\lambda^2 - k_0^2} \ . \tag{2.69}$$

In equations (2.67), (2.68) and (2.69) we have written the potential as a linear expansion, called Fourier-Bessel expansion, of the infinite set of basis functions $\gamma_\lambda(r) = J_0(\lambda r)$. It can be shown from the integral representation (2.50) of J_0 that the basis functions for different λ's are "orthonormal" in the sense that the inner product:

$$<\gamma_\lambda \mid \gamma_{\lambda'}> = \int_0^\infty J_0(\lambda r) J_0(\lambda' r)\, r dr = \frac{1}{\lambda}\delta(\lambda - \lambda') \ . \tag{2.70}$$

To determine the unknown functions $C_i(\lambda)$ and $D_i(\lambda)$ we insert the expansions (2.67), (2.68) and (2.69) into the boundary conditions (2.65). Because of the "orthonormality" of the basis functions J_0, we must demand that the expansion coefficients are equal for each value of λ. At the ground surface, $z = 0$, this gives:

$$\frac{\lambda}{m_0} e^{m_0 z_0} + C_0 = C_1 \cosh(m_1 d_1) - D_1 \sinh(m_1 d_1) \tag{2.71a}$$

$$-\frac{\lambda}{m_0} e^{m_0 z_0} + C_0 = -\frac{m_1}{m_0}[\, C_1 \sinh(m_1 d_1) - D_1 \cosh(m_1 d_1)\,] \ . \tag{2.71b}$$

At the layer interfaces at depths $z = h_i$, $i = 2, \ldots, N-1$ we have:

$$C_{i-1} = C_i \cosh(m_i d_i) - D_i \sinh(m_i d_i) \tag{2.72a}$$

$$D_{i-1} = -\frac{m_i}{m_{i-1}}[\, C_i \sinh(m_i d_i) - D_i \cosh(m_i d_i)\,] \ . \tag{2.72b}$$

At the interface between layer N - 1 and the basement, $z = h_N$, we have finally :

$$C_{N-1} = D_N \; ; \; D_{N-1} = -\frac{m_N}{m_{N-1}} D_N \; . \tag{2.73}$$

In these steps we have used that $\cosh(0) = 1$, $\sinh(0) = 0$, $\cosh'(mz) = m\sinh(mz)$, $\sinh'(mz) = m\cosh(mz)$ and that $h_i - h_{i+1} = -d_i$ where d_i is the thickness of the i'th layer.

From equations (2.72) and (2.73) we see that the coefficients C and D are defined recursively and it is furthermore clear that they will all be proportional to D_N. We write therefore:

$$C_i = S_i D_N \; ; \; D_i = T_i D_N \; ; \; i = 1, \ldots, N-1 \tag{2.74}$$

and get from (2.72) and (2.73) the following recursion relations for S and T:

$$S_{i-1} = S_i \cosh(m_i d_i) - T_i \sinh(m_i d_i) \tag{2.75a}$$

$$T_{i-1} = -\frac{m_i}{m_{i-1}} [\, S_i \sinh(m_i d_i) - T_i \cosh(m_i d_i) \,] \tag{2.75b}$$

$$S_{N-1} = 1 \; ; \; T_{N-1} = -\frac{m_N}{m_{N-1}} \; . \tag{2.75c}$$

Equations (2.71) now become :

$$\frac{\lambda}{m_0} e^{m_0 z_0} + C_0 = S_0 D_N \tag{2.76a}$$

$$-\frac{\lambda}{m_0} e^{m_0 z_0} + C_0 = T_0 D_N \; . \tag{2.76b}$$

Equations (2.76) can easily be solved for C_0 and D_N resulting in :

$$D_N = 2\frac{\lambda}{m_0} e^{m_0 z_0} \frac{1}{S_0 - T_0} \tag{2.77a}$$

$$C_0 = \frac{\lambda}{m_0} e^{m_0 z_0} \frac{S_0 + T_0}{S_0 - T_0} \; . \tag{2.77b}$$

Equations (2.74), (2.75) and (2.77) determine $C_i(\lambda)$ and $D_i(\lambda)$ as functions of λ and the electrical parameters for all the layers of the earth and for the air. The potentials are obtained by substituting these in (2.67), (2.68) and (2.69).

Let us work out the potential in the case of one layer (N = 1), that is a homogeneous half-space. From (2.75c) we see that $S_0 = 1$ and $T_0 = -m_1/m_0$. From (2.77) we then get:

$$D_1 = e^{m_0 z_0} \frac{2\lambda}{m_0 + m_1} \; ; \; C_0 = \frac{\lambda}{m_0} e^{m_0 z_0} \frac{m_0 - m_1}{m_0 + m_1} \; . \tag{2.78}$$

To determine the potential in the air above the half-space we insert this into (2.69) and get:

$$A_0(r,z) = \frac{-i\omega\mu_0 M_0}{4\pi}\int_0^\infty \frac{\lambda}{m_0}[e^{-m_0|z-z_0|} + \frac{m_0 - m_1}{m_0 + m_1}e^{m_0(z+z_0)}]J_0(\lambda r)\,d\lambda \;;\; z < 0 \qquad (2.79)$$

and in the half-space we get by inserting into (2.68) :

$$A_1(r,z) = \frac{-i\omega\mu_0 M_0}{4\pi}\int_0^\infty \frac{2\lambda}{m_0 + m_1}e^{m_0 z_0 - m_1 z}J_0(\lambda r)\,d\lambda \;;\; z > 0\;. \qquad (2.80)$$

Now when we have determined the potentials above and in the layers of the earth we can determine the electric and magnetic fields by inserting the potentials into equations (2.8) and (2.10) relating the fields and the potentials. Restating these equations we have:

$$E = \nabla\times A \qquad (2.81a)$$

$$H = \sigma A + i\varepsilon\omega A + \frac{i}{\mu\omega}\nabla(\nabla\cdot A) = \frac{i}{\mu\omega}k^2 A + \frac{i}{\mu\omega}\nabla(\nabla\cdot A) \qquad (2.81b)$$

where in the equation for the magnetic field we have made use of the gauge condition (2.13) and that, because of the simple harmonic time dependence, a time derivative only gives a multiplication factor $i\omega$ (see (2.33)). We have furthermore introduced $k^2 = \omega^2\mu\varepsilon - i\omega\mu\sigma$ as defined in (2.37). Because of the axial symmetry we want to express these equations in cylindrical coordinates. Doing this and making use of the fact that in each layer the vector potential is independent of the azimuth angle ψ and has only a z-component we get :

$$E_{ir} = 0 \;;\; E_{i\psi} = -\frac{\partial A_i}{\partial r} \;;\; E_{iz} = 0 \qquad (2.82a)$$

$$H_{ir} = \frac{i}{\mu_i\omega}\frac{\partial^2 A_i}{\partial r\partial z} \;;\; H_{i\psi} = 0 \;;\; H_{iz} = \frac{i}{\mu_i\omega}k_i^2 A_i + \frac{i}{\mu_i\omega}\frac{\partial^2 A_i}{\partial z^2} \qquad (2.82b)$$

$$i = 0,1,\ldots,N\;.$$

From this we see that the electric field has only azimuth- or ψ-component and that the magnetic field only has r- and z-components but no ψ-component. This is in agreement with the axial symmetry that the fields must satisfy.

Now we have completed the task of finding the electromagnetic field generated by a harmonically oscillating vertical magnetic dipole over an N-layered earth because now we only have to insert the potentials given by equations (2.67), (2.68) and (2.69) into (2.82). In geophysical applications of this formalism we usually measure the fields at the surface of the earth and due to a source also sitting at the surface. Let us therefore work out the fields at the surface in this case. Making use of the identity:

$$\frac{\partial J_0(\lambda r)}{\partial r} = -\lambda J_1(\lambda r) \tag{2.83}$$

where $J_1(x)$ is the Bessel function of the first kind and first order, and by using (2.77b) we get the electro-magnetic field at the surface by inserting (2.69) into (2.82a) and (2.82b) and setting $z = 0$ and finally letting $z_0 \rightarrow 0$. For the electric field this gives :

$$E_\psi(r) = \frac{-i\omega\mu_0 M_0}{2\pi} \int_0^\infty \frac{\lambda^2}{m_0} \frac{S_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda \ . \tag{2.84}$$

For the radial component of the magnetic field we get :

$$H_r(r) = \frac{-M_0}{2\pi} \int_0^\infty \lambda^2 \frac{T_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda \tag{2.85}$$

and finally for the z-component of the magnetic field :

$$H_z(r) = \frac{M_0}{2\pi} \int_0^\infty \frac{\lambda}{m_0} (k_0^2 + m_0^2) \frac{S_0}{S_0 - T_0} J_0(\lambda r)\, d\lambda \ . \tag{2.86}$$

S_0 and T_0 are given by the recursion relations (2.75) and r is the distance, along the surface, from the source dipole to the measuring point. We have dropped the subscript 0 because hereafter we will (if not otherwise stated explicitly) only be talking about the electromagnetic fields at the surface of the earth.

For a homogeneous half-space (N = 1) equation (2.75c) gives $S_0 = 1$ and $T_0 = -m_1/m_0$ and the electromagnetic field becomes :

$$E_\psi(r) = \frac{-i\omega\mu_0 M_0}{2\pi} \int_0^\infty \frac{\lambda^2}{m_0 + m_1} J_1(\lambda r)\, d\lambda \tag{2.87a}$$

$$H_r(r) = \frac{M_0}{2\pi} \int_0^\infty \frac{\lambda^2 m_1}{m_0 + m_1} J_1(\lambda r)\, d\lambda \tag{2.87b}$$

$$H_z(r) = \frac{M_0}{2\pi} \int_0^\infty \frac{\lambda}{m_0} (k_0^2 + m_0^2) \frac{m_0}{m_0 + m_1} J_0(\lambda r)\, d\lambda \ . \tag{2.87c}$$

In the next chapter, when discussing central loop and grounded dipole-coil transient soundings, we will concentrate on the azimuth component of the electric field. Let us therefore elaborate a little on the electric field for homogeneous earth in the so called *quasi-stationary approximation*. In this approximation we can evaluate the integral in (2.87a) and express the electric field in a closed form and the result will be useful later. In the quasi-

stationary approximation we neglect the displacement currents (the second term on the right hand side of the Maxwell equation (2.2)). To see that this is justified we note from (2.23) and (2.6) that for harmonically oscillating fields the total current is given as :

$$\boldsymbol{J} = \boldsymbol{J}_{cond} + \boldsymbol{J}_{dis} = \sigma(1 + i\omega\frac{\varepsilon}{\sigma})\boldsymbol{E} \ . \tag{2.88}$$

The dielectric permittivity is $\varepsilon = \varepsilon_r\varepsilon_0$ where $\varepsilon_0 = 8.854\cdot10^{-12}$ farad/m is the dielectric permittivity of vacuum and ε_r is the relative dielectric constant. For the water saturated rocks in the upper crust of the earth the conductivity σ is in the range $1 - 10^{-3}$ S/m and the dielectric constant ε_r is in the range $10^2 - 10^3$. This implies that the ratio ε/σ is in the range $10^{-9} - 10^{-5}$ showing that the angular velocity ω has to be of the order of 10^5 Hz or higher in order that the displacement current term in (2.88) is comparable to the conduction current. In transient soundings it is the low frequencies that penetrate to some depth and hence give information about the subsurface resistivity structure and we can therefore safely use the quasi-stationary approximation.

In the formalism that we have developed here the displacement current amount to give a term proportional to ε in the quantities k^2 defined in (2.36) and (2.37) and it is readily seen that the quasi-stationary approximation consists of dropping terms proportional to ε and taking :

$$k_0^2 = 0 \ ; \ k_i^2 = -i\omega\mu_i\sigma_i \tag{2.89}$$

Let us now turn to the electric field given by (2.87a). In the quasi-stationary approximation we have $m_0 = \lambda$ and the integral in (2.87a), let us call it $G(k_1,r)$, can be written as :

$$G(k_1,r) = \int_0^\infty \frac{\lambda^2}{\lambda + m_1} J_1(\lambda r)\, d\lambda = -\frac{\partial}{\partial r}F(k_1,r) \tag{2.90}$$

$$F(k_1,r) = \int_0^\infty \frac{\lambda}{\lambda + \sqrt{\lambda^2 - k_1^2}} J_0(\lambda r)\, d\lambda \ . \tag{2.90a}$$

Multiplying the numerator and the denominator of the integrand by $\lambda - \sqrt{\lambda^2 - k_1^2}$ we get :

$$F(k_1,r) = \frac{-1}{k_1^2}\int_0^\infty [\lambda\sqrt{\lambda^2 - k_1^2} - \lambda^2] J_0(\lambda r)\, d\lambda \ . \tag{2.91}$$

Recalling (2.56) we now define :

$$f(k,r,z) = \int_0^\infty \frac{\lambda\, e^{-\sqrt{\lambda^2 - k^2}z}}{\sqrt{\lambda^2 - k^2}} J_0(\lambda r)\, d\lambda = \frac{e^{-ik\sqrt{r^2 + z^2}}}{\sqrt{r^2 + z^2}} \ ; \ z \geq 0 \ . \tag{2.92}$$

By differentiating the integral representation of f with respect to z it is readily seen that:

$$F(k_1,r) = \frac{-1}{k_1^2}\lim_{z\to 0}\left[\frac{\partial^2 f(k_1,r,z)}{\partial z^2} - \lim_{k\to 0}\frac{\partial^2 f(k,r,z)}{\partial z^2}\right] . \tag{2.93}$$

Using the analytic expression for $f(k,r,z)$, given in (2.92), and taking the derivative and the limits we find that :

$$F(k_1,r) = \frac{-1 + (1 + ik_1r)e^{-ik_1r}}{k_1^2 r^3} . \tag{2.94}$$

Taking now the derivative with respect to r we find that :

$$G(k_1,r) = -\frac{\partial}{\partial r}F(k_1,r) = \frac{(3 + 3ik_1r - k_1^2r^2)e^{-ik_1r} - 3}{k_1^2 r^4} . \tag{2.95}$$

Substituting this in (2.87a) we get an analytic expression for the azimuth component of the electric field at the surface of a homogeneous half-space, in the quasi-stationary approximation :

$$E_\psi(r) = \frac{-i\omega\mu_0 M_0}{2\pi}\,\frac{(3 + 3ik_1r - k_1^2r^2)e^{-ik_1r} - 3}{k_1^2 r^4} . \tag{2.96}$$

The magnetic permeability of the rocks in the earth's crust only deviates very little (relative variation less than 10^{-5}) from the vacuum permeability. It is customary to take the magnetic permeability equal to that of vacuum ($\mu_i = \mu_0$). Using this and that $k^2 = -i\omega\mu\sigma$ in the quasi-stationary approximation (we drop the subscript 1), equation (2.96) further simplifies to :

$$E_\psi(r) = \frac{M_0\rho}{2\pi r^4}\,[(3 + 3ikr - k^2r^2)e^{-ikr} - 3] \tag{2.97}$$

where we have introduced the resistivity $\rho = 1/\sigma$.

Let us conclude this section by considering the electric field generated by the magnetic dipole source in the absence of the earth. To get the electric field for this case we consider the expression (2.87a) for the electric field at the surface of homogeneous half-space and take the electrical parameters of the half-space to be the same as for the air, that is we take $m_1 = m_0$, and get :

$$E_\psi^0(r) = \frac{-i\omega\mu_0 M_0}{4\pi}\int_0^\infty \frac{\lambda^2}{m_0} J_1(\lambda r)\, d\lambda . \tag{2.98}$$

In the quasi-stationary approximation ($m_0 = \lambda$) this becomes :

$$E_\psi^0(r) = \frac{-i\omega\mu_0 M_0}{4\pi r^2} \tag{2.99}$$

where we have used that :

$$r^2 \int_0^\infty \lambda J_1(\lambda r)\, d\lambda \;=\; \int_0^\infty x\, J_1(x)\, dx \;=\; 1 \tag{2.100}$$

which is easily deduced from (2.83).

Using (2.99) we can write the expression for the electric field at the surface of a homogeneous half-space in the quasi-stationary approximation (2.96) as :

$$E_\psi(r) \;=\; E_\psi^0(r)\, E_h^e(r) \tag{2.101}$$

where we have defined :

$$E_h^e(r) \;=\; 2\,\frac{(3 + 3ikr - k^2r^2)e^{-ikr} - 3}{k^2r^2}\;. \tag{2.102}$$

In a similar manner, the electric field at he surface of a horizontally layered earth, given by (2.84), can, in the quasi-stationary approximation, be written as :

$$E_\psi(r) \;=\; E_\psi^0(r)\, E^e(r) \tag{2.103}$$

where

$$E^e(r) \;=\; 2r^2 \int_0^\infty \lambda \frac{S_0}{S_0 - T_0}\, J_1(\lambda r)\, d\lambda\;. \tag{2.104}$$

Equations (2.101) and (2.103) express the electric field at the surface of a homogeneous and a layered earth as a product of the electric field that would be measured if the earth was absent and a earth response factor describing the the influence of the conducting earth.

3. CENTRAL LOOP AND GROUNDED-DIPOLE-COIL TRANSIENT SOUNDINGS.

In section 2 we worked out the electromagnetic fields at the surface of a horizontally stratified earth generated by a harmonically oscillating vertical magnetic dipole sited at the surface. In this section we will use the expression for the azimuth component of the electric field to determine the mutual coupling between a horizontal circular loop and a horizontal coil at the centre of the loop (central loop configuration) and the coupling between a grounded dipole and horizontal coil (grounded-dipole-coil configuration) as a function of frequency (in frequency domain) for harmonically oscillating excitation. This is often referred to as frequency domain coupling. The next step will then be to work out these mutual couplings as a function of time (in time domain) for excitations described by simple functions of time such as a step function and a ramp function. This is often referred to as time domain coupling.

3.1 CENTRAL LOOP AND GROUNDED-DIPOLE-COIL COUPLING IN THE FREQUENCY DOMAIN.

The electric and magnetic fields that we determined in section 2 describe the fields that would be measured at the surface of the earth, due to an oscillating magnetic dipole source. The z-component of the magnetic field could be measured by a coil with axis along the z-axis (a horizontal coil) and the radial component of the magnetic field by a coil with horizontal axis pointing towards the source dipole (a vertical coil). Actually we would measure the time derivative of the magnetic flux through the coils by measuring the the induced electromotive force in the coils as a function of time, but we could easily determine the fields, if we wanted to do so, by considering the geometry (cross-sectional area) and the number of windings in the coils and the simple harmonic variation of the fields.

As we stated above, we will not be so much concerned about the magnetic field but rather concentrate on the electric field. The electric field is measured by measuring the voltage generated in a small (infinitesimal) grounded dipole oriented in the azimuthal direction (perpendicular to the direction to the source dipole). If the dipole is of length dl then the voltage generated will be $dV = E_\psi\, dl$. Since the electric field has no radial component, it is easily seen that for an infinitesimal dipole oriented in such a way that the angle between the dipole and the direction to the source dipole is θ (see figure 3.1), the voltage generated in the dipole is:

$$dV = \sin(\theta)\, E_\psi(r)\, dl \tag{3.1}$$

where r is the distance from the Magnetic source dipole to the grounded dipole. If the grounded dipole is not infinitesimal then the voltage is obtained by integrating (3.1) along the dipole:

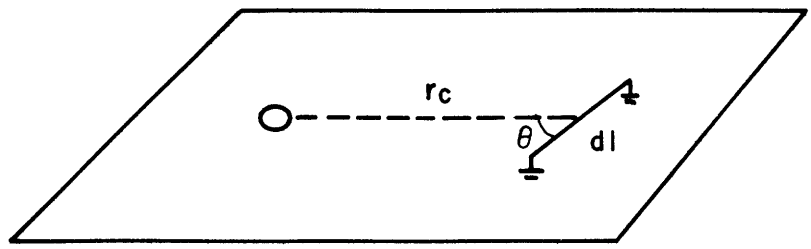

Figure 3.1 Grounded-dipole-coil configuration

$$V_D = \int_{dipole} \sin(\theta)\, E_\psi(r)\, dl \ . \tag{3.2}$$

If r_c, the distance from the magnetic source dipole to the centre of the grounded dipole, is much greater than the length of the grounded dipole, L, then the electric field and the angle θ are practically constant along the dipole and (3.2) can be approximated by:

$$V_D = \sin(\theta)\, E_\psi(r_c)\, L \ . \tag{3.3}$$

We will call this type of source-receiver configuration a *grounded-dipole-coil* (GDC) configuration.

In the other configuration that we will consider here, the electric field is not measured with a grounded dipole but in a circular loop of radius r_c and with the source dipole at the centre (see figure 3.2). The loop can be thought of as a large number of small grounded dipoles oriented perpendicular to the direction to the source and put end to end along the circle. The electric field is the same, $E_\psi(r_c)$, along all the dipoles and total voltage generated in the loop is obtained by multiplying the electric field with the length of the loop and the number of turns if the loop is wound with more than one turn. In this configuration, which we call *central loop* (CL) configuration, the voltage is therefore given as:

$$V_C = 2\pi r_c n_r E_\psi(r_c) = A_r n_r \frac{2\, E_\psi(r_c)}{r_c} \tag{3.4}$$

where we have introduced the area, $A_r = \pi r_c^2$, and the number of windings, n_r, in the loop. If we now use the expression (2.84) for the electric field in these formulas and remember that $M_0 = A_s n_s I_0$ we get the following expressions for the voltage due to the mutual coupling for the GDC configurations:

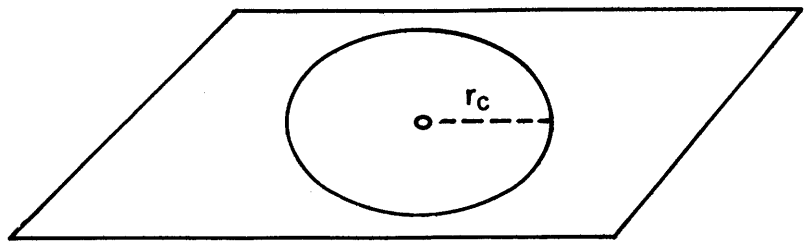

Figure 3.2 Central loop configuration

$$V_D(\omega,r) = \sin(\theta)\, LA_s n_s I_0\, e^{i\omega t}\, \frac{-i\omega\mu_0}{2\pi} \int_0^\infty \frac{\lambda^2}{m_0} \frac{S_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda \qquad (3.5)$$

and for the CL configurations:

$$V_C(\omega,r) = A_r n_r A_s n_s I_0\, e^{i\omega t}\, \frac{-i\omega\mu_0}{\pi r} \int_0^\infty \frac{\lambda^2}{m_0} \frac{S_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda\ . \qquad (3.6)$$

We have, for the sake of completeness, introduced again the factor $e^{i\omega t}$ describing the time variation of the current, that we had omitted for a while.

In the derivation of these couplings we assumed that the source was the magnetic dipole due to the harmonically oscillating current in the coil. Because of the **principle of reciprocity** we can interchange the roles of the source and the receiver. This means that (3.5) also describes the voltage generated in a coil due to an alternating current $I = I_0\, e^{i\omega t}$ transmitted into a grounded dipole of length L and at the distance r from the coil ($L << r$) and where θ is the angle between the dipole and the direction to the coil. In the same way (3.6) also describes the voltage generated in a coil at the centre of a loop of radius r in which an alternating current $I = I_0\, e^{i\omega t}$ is transmitted. This latter situation is much more common in geophysical application of induction and transient electromagnetic methods, that is one usually uses relatively large grounded dipole or loop as source and a coil or a small loop as a receiver.

Before proceeding further, it is convenient to introduce what we call the mutual impedance between the source and the receiver which is, by analogy with Ohm's law, defined as the ratio between the measured voltage and the transmitted current. From equations (3.5) and (3.6) we get the mutual impedances for the GDC and CL configurations as follows:

$$Z_D(\omega,r) = \frac{V_D(\omega,r)}{I_0\, e^{i\omega t}} = \sin(\theta)\, LA_s n_s \frac{-i\omega\mu_0}{2\pi} \int_0^\infty \frac{\lambda^2}{m_0} \frac{S_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda \qquad (3.7)$$

$$Z_C(\omega,r) = \frac{V_L(\omega,r)}{I_0\, e^{i\omega t}} = A_r n_r A_s n_s \frac{-i\omega\mu_0}{\pi r} \int_0^\infty \frac{\lambda^2}{m_0} \frac{S_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda\ . \qquad (3.8)$$

Now we have succeeded in deriving formulas for the induced voltage in a receiver coil, in the frequency domain, due to alternating current transmitted into a loop or a grounded dipole. The next step is now to use these formulas to work out the voltage in the time domain.

3.2 CENTRAL LOOP AND GROUNDED-DIPOLE-COIL COUPLING IN THE TIME DOMAIN.

In this section we will proceed to work out the time domain couplings for GDC and CL-soundings. We will begin by writing down a general expression for the coupling in the time domain by using Fourier expansion of the function describing the transmitted current. Then we will elaborate on two specific functions, a step function of time and a ramped step function.

Let the transmitted current be described by the function $I(t)$. We make a Fourier expansion of the current function:

$$I(t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \tilde{I}(\omega)\, e^{i\omega t}\, d\omega \qquad (3.9)$$

where

$$\tilde{I}(\omega) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} I(t)\, e^{-i\omega t}\, dt\ . \qquad (3.10)$$

Equation (3.9) is an expansion of the current function in terms of "orthonormal" (see eq. (2.42)) and hence linearly independent basis functions. The current function can therefore be thought of as an infinite sum of independent harmonic oscillations $e^{i\omega t}$ with amplitudes $I(\omega)$. Because the Maxwell's equations are linear in the sources, the response in the receiver is also a sum of the responses for the individual oscillations. From equations (3.7) and (3.8) we see therefore that the measured voltage as a function of time is given as:

$$V_{D,C}(t,r) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} Z_{D,L}(\omega,r)\, \tilde{I}(\omega)\, e^{i\omega t}\, d\omega\ . \qquad (3.11)$$

This equation is a general expression for the induced voltage in the receiver in terms of the mutual impedance as a function of frequency and the Fourier transform of the transmitted

current.

In transient soundings the transmitted current in the source is usually a square wave (or almost a square wave). Each time the transmitted current is abruptly changed an induced voltage is generated in the receiver. This induced transient voltage decays with time after the current switching. The duration of each square pulse of the transmitted current is usually made long enough so that the transient voltage in the receiver has dropped to zero well before the current changes again. We can therefore consider each abrupt current change in the square wave as an individual step function. If, for each transient, the time is measured relative to the time at which the current changed we only need to consider the step functions θ_+ and θ_- defined by:

$$\theta_+(t) = \begin{cases} 0 & t<0 \\ 1 & t>0 \end{cases} \quad ; \quad \theta_-(t) = \begin{cases} 1 & t<0 \\ 0 & t>0 \end{cases} . \tag{3.12}$$

These step functions have the Fourier expansion:

$$\theta_\pm(t) = \frac{\pm 1}{2\pi} \lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{e^{i\omega t}}{i\omega \pm \varepsilon} d\omega \quad ; \quad \varepsilon > 0 . \tag{3.13}$$

To see this we note that the integrand in (3.13) has a singularity at $\omega = i\varepsilon$ for $\theta_+(t)$ and at $\omega = -i\varepsilon$ for $\theta_-(t)$. For $t<0$ we can consider the integral as an integral along a closed contour in the complex ω-plane by adding the integral along a semi circle of infinite radius in the lower half plane giving no contribution. For $t>0$ we can likewise close the integration contour by an infinite semi circle in the upper half plane. For the step function $\theta_+(t)$ the singularity is outside the integration contour for $t<0$ and by Cauchy's integral formula (2.53) we get zero, but for $t>0$ the singularity is inside the contour and we get $\theta_+(t) = \lim_{\varepsilon \to 0} e^{-\varepsilon t} = 1$. For $\theta_-(t)$ the singularity is inside the integration contour for $t<0$ and we get $\theta_-(t) = 1$ (note that the contour is in this case oriented in the negative direction) but the singularity is out side the contour for $t>0$ giving zero. From (3.13) we see that the Fourier transforms of a current step functions $I_\pm = I_0\,\theta_\pm(t)$ are:

$$\tilde{I}_\pm(\omega) = \frac{1}{(2\pi)^{1/2}} \frac{\pm I_0}{i\omega} . \tag{3.14}$$

To get the measured transient voltage as a function of time after the current step we insert this into (3.11). Let us do that for $I_-(t)$, a steady current abruptly turned of at $t = 0$, (the case of the other step function is exactly the same except for a sign). This gives (we omit the subscripts D and L and also r because the following considerations apply equally to both configurations and all distances):

$$V_-(t) = \frac{-I_0}{2\pi} \int_{-\infty}^{\infty} \frac{Z(\omega)}{i\omega} e^{i\omega t} d\omega = \frac{I_0}{2\pi} \int_{-\infty}^{\infty} \Phi(\omega) e^{i\omega t} d\omega \qquad (3.15)$$

where we have for convenience defined:

$$\Phi(\omega) = \frac{Z(\omega)}{-i\omega} . \qquad (3.16)$$

(To the careful reader we point out that in (3.15) we don't have to worry about the singularity in the Fourier transform of the step function because the impedances defined in (3.7) and (3.8) are well behaved functions of ω multiplied by a factor $i\omega$ which cancels out the denominator giving the singularity).

From the recursion relations (2.75) defining the quantities S_0 and T_0 which determine the impedances (see (3.7) and (3.8)), we see that they only depend on ω through $m = \sqrt{\lambda^2 - k^2}$ where $k^2 = \omega^2\mu\varepsilon - i\omega\mu\sigma$ ($= -i\omega\mu\sigma$ in the quasi-stationary approximation). It is therefore clear from (3.7), (3.8) and (3.16) that $\Phi(\omega)$ only depends on ω through ω^2 and $i\omega$. This implies that:

$$\Phi^*(-\omega) = \Phi(\omega) \qquad (3.17)$$

where * denotes the complex conjugation and hence

$$\mathrm{Re}\,\Phi(-\omega) = \mathrm{Re}\,\Phi(\omega) \quad ; \quad \mathrm{Im}\,\Phi(-\omega) = -\mathrm{Im}\,\Phi(\omega) . \qquad (3.17a)$$

This can be used to simplify (3.15) because if we write it in terms of real and imaginary parts we get:

$$V_-(t) = \frac{I_0}{2\pi} \int_{-\infty}^{\infty} [\,\mathrm{Re}\,\Phi(\omega)\cos(\omega t) - \mathrm{Im}\,\Phi(\omega)\sin(\omega t)\,]\, d\omega$$
$$+ i\,\frac{I_0}{2\pi} \int_{-\infty}^{\infty} [\,\mathrm{Re}\,\Phi(\omega)\sin(\omega t) + \mathrm{Im}\,\Phi(\omega)\cos(\omega t)\,]\, d\omega . \qquad (3.18)$$

The symmetry relations (3.17) and the symmetry properties for sin and cos show that the integrand in the second integral is anti symmetric with respect to ω and the second integral is therefore zero. The integrand in the first integral is symmetric with respect to ω and we can therefore write:

$$V_-(t) = \frac{I_0}{\pi} \int_{0}^{\infty} [\,\mathrm{Re}\,\Phi(\omega)\cos(\omega t) - \mathrm{Im}\,\Phi(\omega)\sin(\omega t)\,]\, d\omega . \qquad (3.19)$$

This can be further simplified by noting that the voltage is zero before the current is abruptly turned off so we must have that:

$$0 = V_-(-t) = \frac{I_0}{\pi}\int_0^\infty [\,\mathrm{Re}\,\Phi(\omega)\cos(\omega t) + \mathrm{Im}\,\Phi(\omega)\sin(\omega t)\,]\,d\omega \quad ; \quad t>0\,. \qquad (3.20)$$

By adding or subtracting this from (3.19) we get two alternative and relatively simple expressions for the measured voltage as a function of time after the current in the transmitter is turned off namely:

$$V_-(t) = \frac{2I_0}{\pi}\int_0^\infty \mathrm{Re}\,\Phi(\omega)\cos(\omega t)\,d\omega \qquad (3.21a)$$

$$V_-(t) = \frac{-2I_0}{\pi}\int_0^\infty \mathrm{Im}\,\Phi(\omega)\sin(\omega t)\,d\omega\,. \qquad (3.21b)$$

Now when we have determined the step function response $V_-(t)$ we can derive an expression for the voltage response generated by an arbitrary current function $I(t)$ in terms of the step response and the time derivative of the current function. Such an expression will be useful to determine the response of current functions that have a simple derivative such as the ramped step function.

To begin with we note that the step functions have the following properties:

$$\theta_+(-t) = \theta_-(t) \quad ; \quad \frac{d}{dt}\theta_+(t) = \delta(t)\,. \qquad (3.22)$$

Using this and the properties of the delta function, (2.43), we can write for an arbitrary current function :

$$I(t) = \int_{-\infty}^\infty \delta(\tau - t)\,I(\tau)\,d\tau = \int_{-\infty}^\infty \frac{d}{d\tau}\theta_+(\tau - t)\,I(\tau)\,d\tau\,. \qquad (3.23)$$

Partial integration gives :

$$I(t) = \Big[\theta_+(\tau - t)\,I(\tau)\Big]_{\tau=-\infty}^{\tau=+\infty} - \int_{-\infty}^\infty \theta_+(\tau - t)\,\frac{d}{d\tau}I(\tau)\,d\tau = -\int_{-\infty}^\infty \theta_+(\tau - t)\,\frac{d}{d\tau}I(\tau)\,d\tau \qquad (3.24)$$

where we have used that $\theta_+(\tau)$ vanishes for $\tau<0$ and we assume that $I(\tau)$ vanishes for very large τ. The voltage response is given, in terms of the Fourier transform of the current function, by (3.11) so we take the Fourier transform of (3.24):

$$\tilde{I}(\omega) = -\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \theta_+(\tau - t)\, e^{-i\omega t}\, dt\, \frac{d}{d\tau} I(\tau)\, d\tau$$

$$= -\int_{-\infty}^{\infty} e^{-i\omega\tau} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \theta_+(\tau - t)\, e^{-i\omega(t-\tau)}\, dt\, \frac{d}{d\tau} I(\tau)\, d\tau \ . \qquad (3.25)$$

By using that $\theta_+(-t) = \theta_-(t)$ and by changing the integration variable for the inner integral to $t' = t - \tau$ we can write (3.25) as :

$$\tilde{I}(\omega) = -\int_{-\infty}^{\infty} e^{-i\omega\tau}\, \tilde{\theta}_-(\omega)\, \frac{d}{d\tau} I(\tau)\, d\tau \qquad (3.26)$$

where $\tilde{\theta}_-(\omega)$ is the Fourier transform of the step function $\theta_-(t)$. Inserting now (3.26) into (3.11) gives :

$$V(t) = -\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} Z(\omega) i_-(\omega) e^{i\omega(t-\tau)}\, d\omega\, \frac{d}{d\tau} I(\tau)\, d\tau$$

$$= -\int_{-\infty}^{\infty} v_-(t - \tau)\, \frac{d}{d\tau} I(\tau)\, d\tau \qquad (3.27)$$

where $v_-(t)$ is the voltage response for the current function $\theta_-(t)$, that is a constant current of value one abruptly turned of at $t = 0$.

Equation (3.27) states that the transient voltage response in the receiver for an arbitrary current function $I(t)$ is given as a the unit step response $v_-(t)$ convolved with the time derivative of the current. The unit step response, $v_-(t)$, vanishes for $t<0$ so that (3.27) can also be written as:

$$V(t) = -\int_{-\infty}^{t} v_-(t - \tau)\, \frac{d}{d\tau} I(\tau)\, d\tau \ . \qquad (3.28)$$

The unit step response is obtained by dividing equation (3.21a) with I_0 and is given as :

$$v_-(t) = \frac{2}{\pi} \int_0^{\infty} \mathrm{Re}\, \Phi(\omega) \cos(\omega t)\, d\omega \ . \qquad (3.29)$$

Let us apply (3.28) to determine the response due to a ramped step function with the ramp length, or turn-off time TOFF. Let the function be given as a constant current I_0 for $t<$TOFF, linearly decreasing to zero in the interval -TOFF$<t<0$ and zero for $t>0$ (see Figure 3.3).

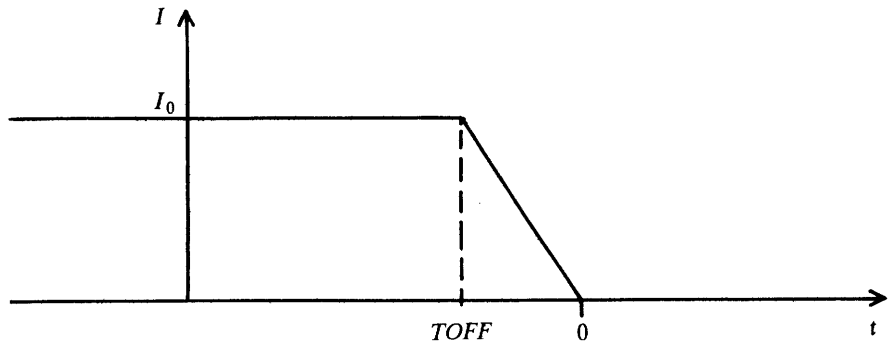

Figure 3.3 Ramped step function

The time derivative of the current function is given as :

$$\frac{dI}{dt} = \begin{cases} 0 & \text{for } t<\text{-TOFF} \\ \dfrac{-I_0}{\text{TOFF}} & \text{for -TOFF}<t<0 \\ 0 & \text{for } t>0 \end{cases} . \qquad (3.30)$$

Inserting this into (3.28) we find that :

$$V(t) = \frac{I_0}{TOFF} \int_{\text{-TOFF}}^{0} v_{-}(t-\tau)\, d\tau = \frac{I_0}{TOFF} \int_{t}^{t+\text{TOFF}} v_{-}(\tau)\, d\tau . \qquad (3.31)$$

The response due to the ramped step function is therefore given as the integral of the sharp step response over the ramp length.

We have now managed to derive expressions for the transient voltage generated in a receiver coil at the surface of a horizontally stratified earth due to current transmitted into a grounded dipole or a loop with the receiver at the centre. As we have seen, the measured voltage depends on the form of the transmitted current and on the resistivity structure of the subsurface. Since the transmitted current is known as a function of time, the measured voltage can be used to give information about the unknown resistivity structure for example by comparing the measured voltage as a function of time to calculated model curves for various layered resistivity structures and for the known current function. If the resistivity structure is not horizontally layered, things get much more complicated and the treatment of this more general case falls outside the scope of the present discussion. Fortunately it is, in most cases, a fair first approximation to assume that the resistivity structure is horizontally stratified.

3.3 THE UNIT STEP RESPONSE OF A LAYERED EARTH

It is now about time that we start to use some of the formulas that we have developed. Let us therefore study the unit step response (3.29) for a layered earth, but first wee will streamline

the notation a bit.

We will from now on work in the quasi-stationary approximation and by using the definition of the earth response factor

$$E^e(\omega,r) = 2r^2 \int_0^\infty \lambda \frac{S_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda \tag{2.104}$$

we can write (3.7) and (3.8) on the form:

$$Z_D(\omega,r) = -i\omega C_D E^e(\omega,r) \tag{3.32}$$

$$Z_C(\omega,r) = -i\omega C_C E^e(\omega,r) \tag{3.33}$$

where

$$C_D = \sin(\theta) L A_s n_s \frac{\mu_0}{4\pi r^2} \; ; \; C_C = A_r n_r A_s n_s \frac{\mu_0}{2\pi r^3} \; . \tag{3.34}$$

From (3.16) and (3.29) we see that the unit step response for the GDC and CL configurations is given as:

$$v_{D,C}(r,t) = \frac{2}{\pi} C_{D,C} \int_0^\infty \mathrm{Re}\,[E^e(r,\omega)] \cos(\omega t)\, d\omega \; . \tag{3.35}$$

In order to work out the transient response we have to calculate the integral (2.104) for the earth response factor and the cosine transform integral (3.35). The quantities S_0 and T_0 in the integrand of (2.104) are easily determined by the recursion relations (2.75) but the integration is generally not an easy task. In the simple case of homogeneous half-space, we can calculate the integral analytically as we did at the end of section two, but in the general case the integral has to be evaluated numerically. The same applies to the cosine transform integral. We can and will in the next section evaluate the integral analytically in the case of a homogeneous half-space, by making a series expansion of the integrand, but in the general case it has to be calculated numerically.

To perform the numerical integration in (2.104) and (3.35), we will use the so called *digital filter method* developed by Gosh. This method will be discussed in some details in chapter 4, but let us here look at some results and start with the simplest case, a homogeneous half-space. First we look at the earth response factor as a function of angular frequency ω. Figure 3.4 shows the earth response factor of a homogeneous earth for three different resistivities, $\rho = 1/\sigma$, namely 1, 10 and 100 Ωm and for r = 100 m (r is the distance form the grounded dipole to the receiver coil in the GDC configuration or the radius of the transmitter loop with the receiver coil at the centre in the CL configuration), plotted as a function of ω. On Figure 3.4 we see that the real part of the response factor tends to one for small ω and to zero for

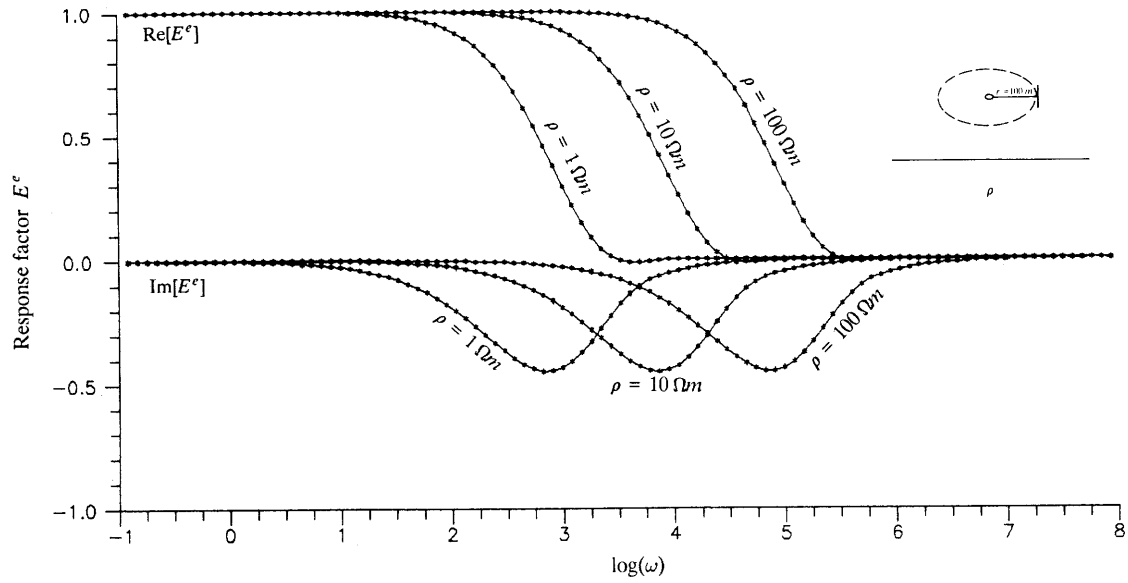

Figure 3.4 Earth response factor for homogeneous half-space

large ω. The transition between these limiting values extends over about three decades. In the transition interval the real part decreases monotonically with increasing ω in the beginning, gets a little bit negative and then levels on to zero. The imaginary part of the response factor tends to zero for both high and low values of ω and is negative in between. The interval in which the imaginary part deviates from zero is in the same frequency range as the transition interval for the real part, but it is a bit wider (about four decades). Figure 3.4 also shows that the curves describing the earth response factor are shifted towards higher frequency values as the resistivity of the half-space increases. We see that increasing the resistivity (lovering the conductivity) by a factor of 10 shifts the response curves one decade toward higher frequencies. This is easyly understood because, in the quasi-stationary approximation, the earth response factor given by (2.104) only depends on the conductivities and ω through k and hence only through the product $\sigma\omega$ (see (2.89)).

By performing the cosine transform (3.35) on the real parts of the earth response factors in Figure 3.4, we get the unit step response of the homogeneous half-space. The result is shown on Figure 3.5 (or actually $\pi\, v_{D,C}(r,t)/(2C_{D,L})$) . On figure 3.5 we see that the voltage response of a homogeneous half-spaces with different resistivities has the same general character and can be divided into three stages. In the *early stage* the induced voltage is constant in time. In the *intermediate stage* the voltage starts to decrease with time and with steadily increasing slope on log-log scale until the *late stage* is reached where the voltage response decreases with time in such a way that the logarithm of the induced voltage decreases linearly as a function of the logarithm of time. The slope of the response curve in

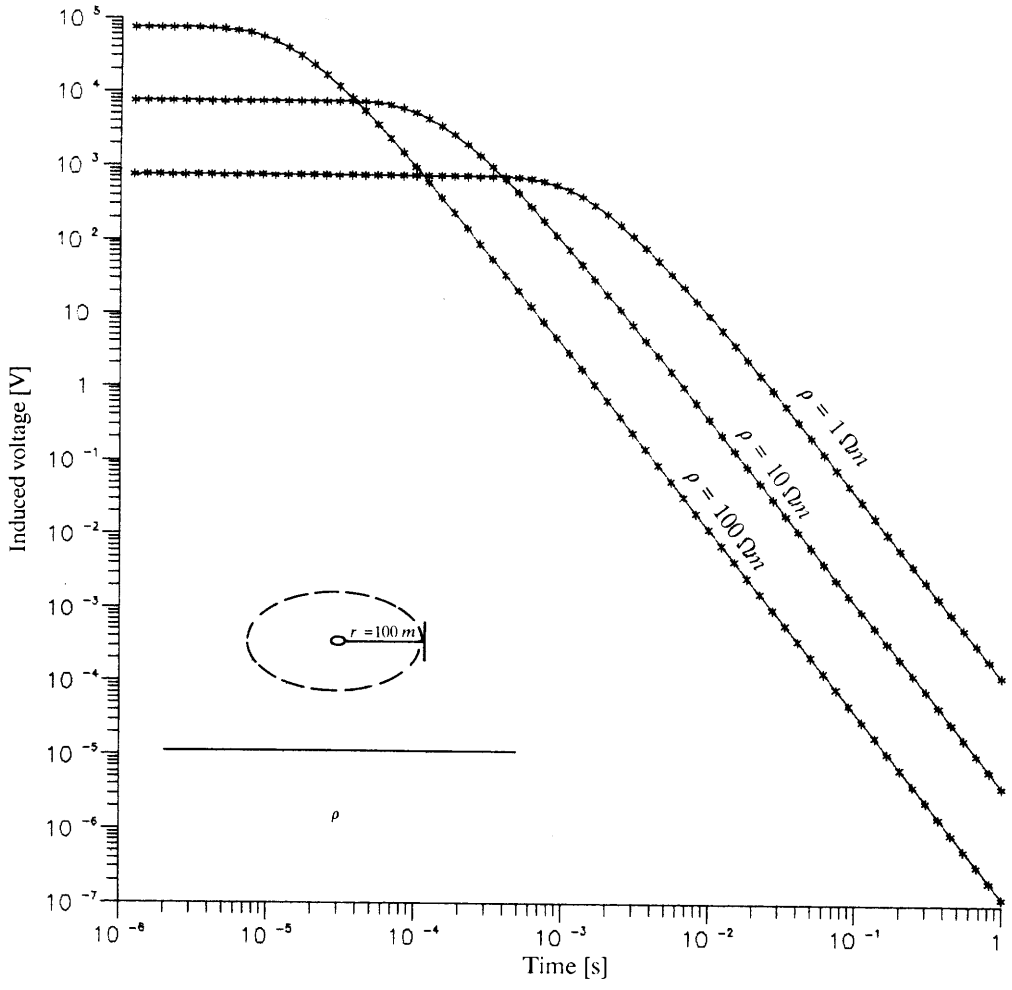

Figure 3.5 Voltage response for a homogeneous half-space

the late stage is easyly seen to be - 5/2, shoving that the voltage is proportional to $t^{-5/2}$. We see further more on figure 3.5 that the early stage response increases as the resistivity of the half-space increases and also that the transitions from early to intermediate and from intermediate to late stages gets shifted towards earlyer times but in such a way that the response curve has the same shape.

This form invariance of the voltage response curve can be easyly understood by considering the way in which the response depends on the conductivity of the layers of a horizontally stratified earth. As discussed earlyer, the earth response factor, in the quasi-stationary approximation, only depends on the conductivities of the layers through the product $\sigma\omega$. Equation (3.35) can therefore be written symbolically in the form :

$$v(\sigma,r,t) = \int_0^\infty f(r,\sigma\omega)\cos(\omega t)\,d\omega \tag{3.36}$$

where we, by writing $\sigma\omega$, are merely stating that in the earth response factor the angular frequency only appears as multiplied by conductivities. If we scale all conductivities by a common factor α then it is easy to see by a simple change of integration variable in (3.36) that

$$v(\alpha\sigma,r,t) = \frac{1}{\alpha} v(\sigma,r,t/\alpha)\ . \tag{3.37}$$

Equation (3.37) is a scaling law stating that if we increase all conductivities by a common factor α (decrease resistivities by $1/\alpha$) then the resulting voltage response is obtained from the unscaled voltage by scaling down both the time and the voltage by a factor $1/\alpha$. On log-log scale this is equivalent to shift the response curve both down and to the right by the amount $\log(\alpha)$.

Equation (3.37) is rather informative because for a homogeneous half-space we have in the early stage where the voltage is independent of time that

$$v(\alpha\sigma,r) = \frac{1}{\alpha} v(\sigma,r) \tag{3.37a}$$

which implies that $v(\sigma)$ is inversely proportional to the conductivity and hence proportional to the resistivity of the half-space. In the late stage, on the other hand, the voltage response is proportional to $t^{-5/2}$ and by (3.37) this means that

$$v(\alpha\sigma,r)\,t^{-5/2} = \frac{1}{\alpha} v(\sigma,r)\,(\frac{t}{\alpha})^{-5/2} = \alpha^{3/2}\,v(\sigma,r)\,t^{-5/2} \tag{3.37b}$$

showing that in the late stage the voltage is proportional to the conductivity of the half-space raised to the power 3/2. We will in the next section arrive at this same result when we solve the integrals defining the homogeneous earth response analytically by use of series expansion. Examples of the earth response factor for two layered earth and r = 100 m are shown on figures 3.6 and 3.7. Figure 3.6 shows the cases of a layer with resistivity of 100 Ωm and different thicknesses (1, 10 and 100 m) over a half-space with resistivity of 1 Ωm. On this figure we see that the earth response factor is hardly distinguishable from the response of a 1 Ωm homogeneous half-space when the resistive surface layer is very thin, d = 1 m. When the resistive surface layer gets very thick the response factor approaches that of a homogeneous half-space with 100 Ωm. For d = 100 m we see that the response is close to the 100 Ωm half-space response except for small deviations for low values of ω in the transition interval from the high to the low frequency limit. This is a manifestation of the fact that the low frequencies penetrate deeper than the high frequencies. For intermediate thicknesses of the resistive surface layer, the behaviour of the response factor is more complicated and it is seen that

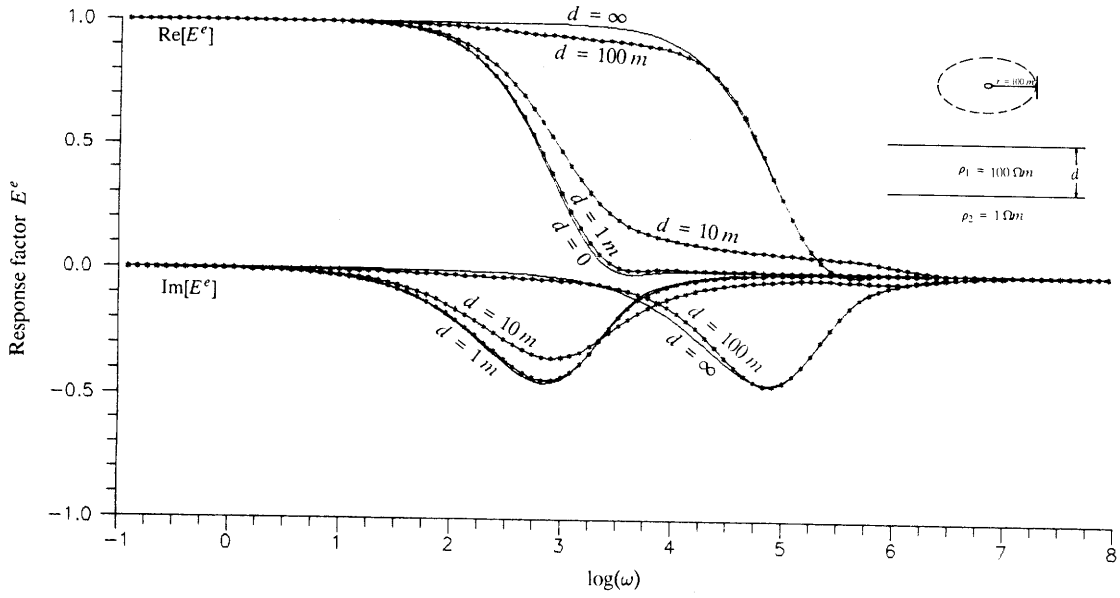

Figure 3.6 Response factor for a resistive layer on top of conductive half-space

deviations from the high frequency limit extend to higher frequency values than in the case of a 100 Ωm half-space.

Figure 3.7 shows the earth response factor in the case of a layer with resistivity of 1 Ωm and different thicknesses (1, 10 and 100 m) over a half-space with resistivity of 100 Ωm. In the case of a thick, $d \geq 100$ m, conductive layer on top of the 100 Ωm half-space, the response factor only deviates from the 1 Ωm half-space response at frequency values just before it levels on the low frequency limiting value. For thin conducting layer, d = 1 m, the response is still strongly affected by the layer and comparison with the case of a resistive layer above a conductive half-space shows that a conductive layer responds much more strongly than a resistive layer. This is of course to be expected because the oscillating magnetic field of the source dipole induces much stronger currents in a conductive layer than in a resistive one. Figure 3.8 shows the voltage response, as a function of time, in the case of a layer with resistivity of 100 Ωm and different thicknesses (1, 10, 100 and 1000 m) over a half-space with resistivity of 1 Ωm. On this figure we see that when the surface layer is thin, d < 10 m, the response deviates from that of a homogeneous half-space with 1 Ωm only at very early times. As the layer gets thicker, d > 100 m, the response follows that of a homogeneous half-space with 100 Ωm in the early and intermediate stages and approaches the late stage response of a homogeneous half-space with 1 Ωm at late times. In between it follows, for a while, the late stage response of a half-space with resistivity equal to that of the top layer, if the layer is thick enough and does so the longer the thicker the layer is. Before the response starts to rise to the

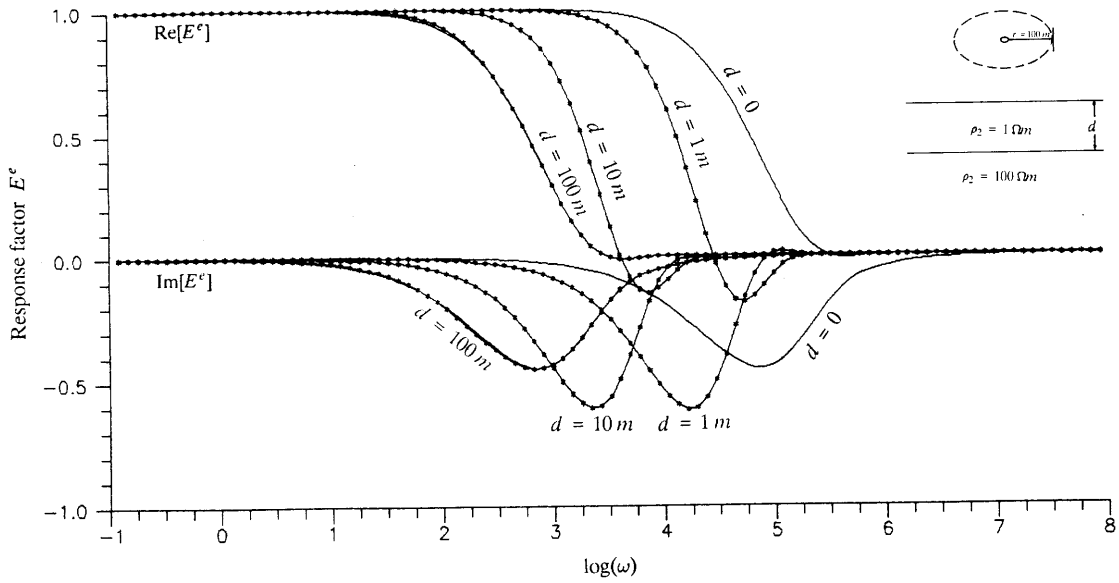

Figure 3.7 Response factor for a conductive layer on top of resistive half-space

1 Ωm half-space response it does a little undershoot under the late stage response of the top layer. Figure 3.9 shows the voltage response in the cases of a layer with resistivity of 1 Ωm and different thicknesses (1, 10 and 100 m) on top of a half-space with resistivity of 100 Ωm. When the top layer is very thin, d < 1 m, the response is, for very early times in between the early stage responses of half-spaces with resistivities 1 and 100 Ωm respectively. It then overshoots the underlaing 100 Ωm half-space response and finally approaches its late stage line for late times. As the top layer gets thicker, the voltage response traces the early stage response of the top layer for longer time and makes less overshooting at intermediate times. The convergence towards the late stage response of the underlaying 100 Ωm half-space also occurs at later times. When the top layer is thick, d > 100 m, the voltage response follows the 1 Ωm half-space response through the early and intermediate stages and a bit down the late stage and then starts to deflect towards the late stage response of the underlaing half-space.

3.4 LATE TIME APPARENT RESISTIVITY.

In the previous section we saw that, in the case of homogeneous earth, the measured transient voltage in the receiving coil can, for both of the CL and GDC configurations, be divided into three regimes or stages. In the early stage the induced voltage is constant in time but starts to decline with time in the intermediate stage. In the late stage the voltage decreases with time in such a way that the logarithm of the voltage decreases linearly as a function of the logarithm of time with slope 5/2.

For a homogeneous half-space with conductivity σ we can show explicitly that the induced transient voltage in the receiver is, for late times, proportional to $\sigma^{3/2}$ and $t^{-5/2}$ as we anticipated in the previous section. On the basis of this late time behavior we can define what is called a *late time apparent resistivity*. If the transient response is presented as late time apparent resistivity rather than induced voltage, the response will, in the case of homogeneous earth, approach the true resistivity of the earth for late times. In the case of layered earth, the late time apparent resistivity will approximately reach the true resistivity of the intermediate laers for some interval of time, if the laers are thick enough so that the voltage response approximately reaches, for some time interval, the linearly decreasing (on log-log plot) late stage behaviour corresponding to homogeneous earth with resistivity of the respective layers.

For homogeneous earth we found in (2.102) that the earth response factor is given as

$$E^{e_h} = 2\,\frac{(3 + 3ik - k^2r^2)e^{-ikr} - 3}{k^2r^2} \,. \tag{3.38}$$

Using the series expansion of e^{-ikr} we get after some trivial algebraic manipulations

$$E^{e_h} = 2\sum_{n=0}^{\infty} \frac{3(n+2) - (n+2)(n+1) - 3}{(n+2)!}(-ikr)^n \,. \tag{3.39}$$

By splitting the sum into two sums with even and odd powers we get :

$$\begin{aligned} E^{e_h} = {} & 2\sum_{n=0}^{\infty} \frac{6(n+1) - 2(2n+1)(n+1) - 3}{(2n+2)!}(ikr)^{2n} \\ & - 2\sum_{n=0}^{\infty} \frac{3(2n+3) - 2(2n+3)(n+1) - 3}{(2n+3)!}(ikr)^{2n+1} \,. \end{aligned} \tag{3.40}$$

Using the definition of k in the quasi-stationary approximation $k = (-i\omega\mu_0\sigma)^{1/2}$ we find that

$$ikr = i^{1/2}(\mu_0\sigma r^2\omega)^{1/2} \tag{3.41}$$

Turning now to the unit step response (3.35), let us calculate

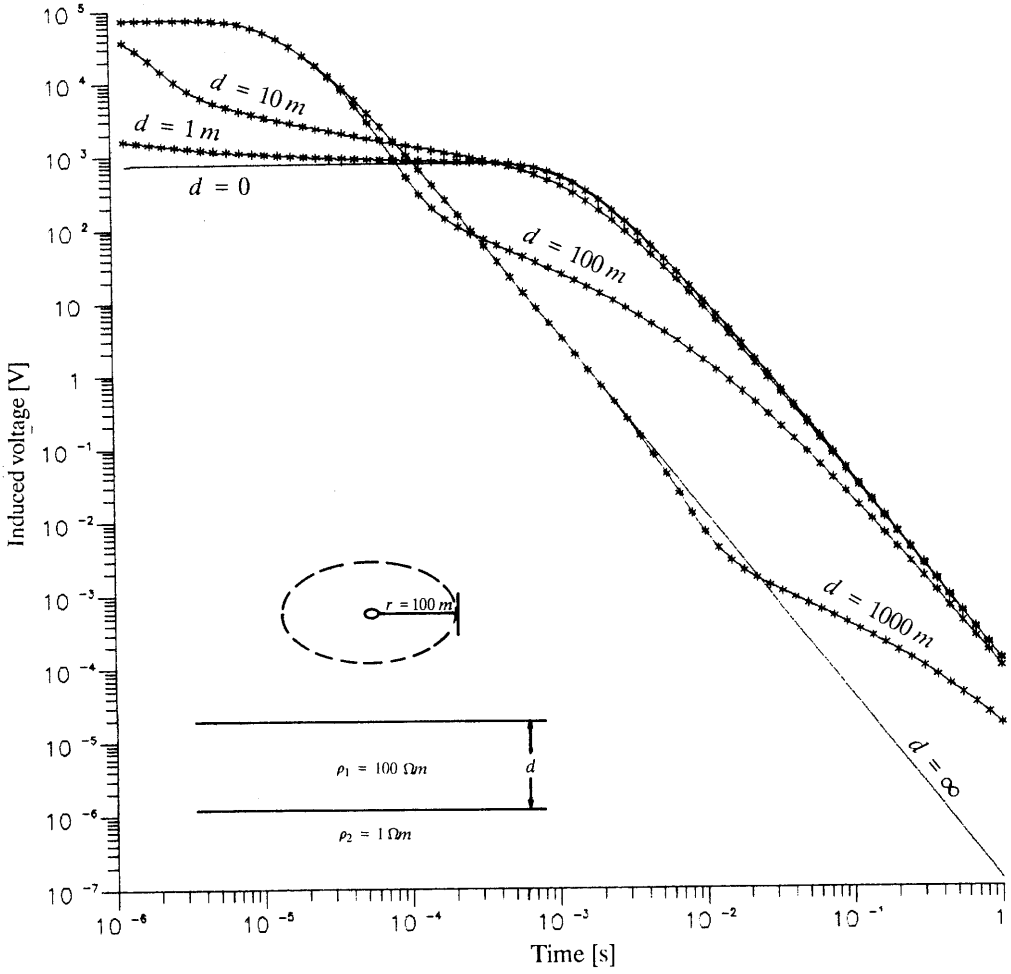

Figure 3.8 Voltage response for a resistive layer on top of conductive half-space

$$I = \frac{\pi\, v_{D,C}\text{-}(r,t)}{4C_{D,C}} = \frac{1}{2}\int_0^\infty \mathrm{Re}\,[\,E^{e_h}(\omega)\,]\cos(\omega t)\,d\omega \tag{3.42}$$

Using (3.40) and (3.41) we find that

$$I = \sum_{n=0}^{\infty} \frac{6(n+1) - 2(2n+1)(n+1) - 3}{(2n+2)!}(\mu_0\sigma r^2)^n\,\mathrm{Re}\,[\,i^n \int_0^\infty \omega^n\cos(\omega t)\,d\omega\,] \tag{3.43}$$

$$-2\sum_{n=0}^{\infty} \frac{3(2n+3) - 2(2n+3)(n+1) - 3}{(2n+3)!}(\mu_0\sigma r^2)^{n+1/2}\,\mathrm{Re}\,[\,i^{n+1/2} \int_0^\infty \omega^{n+1/2}\cos(\omega t)\,d\omega\,] \,.$$

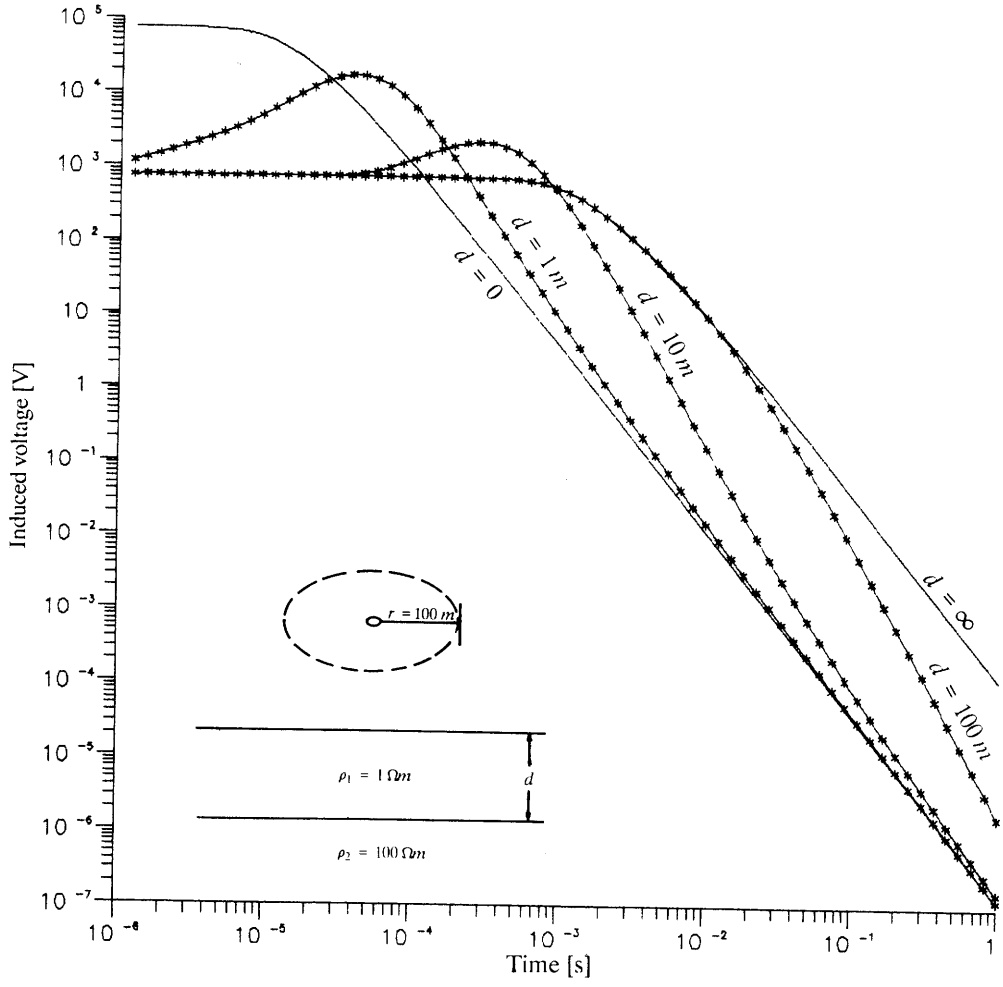

Figure 3.9 Voltage response for a conductive layer on top of resistive half-space

From this we see that we now have to calculate the integral :

$$g = \int_0^\infty \omega^s \cos(\omega t)\, d\omega \,. \tag{3.44}$$

Using the definition of the Γ function we find that

$$\int_0^\infty x^s e^{-ax}\, dx = \frac{\Gamma(s+1)}{a^{s+1}} \,. \tag{3.45}$$

If we take $a = it$ then we see that we can write

$$g = \frac{1}{2}\int_0^\infty \omega^s \left[e^{-it\omega} + e^{-(it)^*\omega}\right] d\omega = \frac{1}{2}\Gamma(s+1)\left[\frac{1}{(it)^{s+1}} + \frac{1}{(-it)^{s+1}}\right] . \qquad (3.46)$$

We have therefore that

$$\int_0^\infty \omega^s \cos(\omega t)\, d\omega = \frac{\Gamma(s+1)}{2t^{s+1}}\left[i^{s+1} + (-i)^{s+1}\right] . \qquad (3.47)$$

Considering now the first sum in (3.43) we see, by using (3.47) with $s = n$, that

$$\mathrm{Re}\left[i^n \int_0^\infty \omega^n \cos(\omega t)\, d\omega\right] = \frac{\Gamma(n+1)}{2t^{n+1}}\,\mathrm{Re}\left[i((-1)^n) - 1)\right] . \qquad (3.48)$$

The right hand side of (3.48) is zero for all $n \geq 0$ because the quantity in the bracket is purely imaginary and each term in the first sum in (3.43) is therefore zero.

Considering the second sum in (3.43) we find by using (3.47) with $s = n + 1/2$ we find that

$$\mathrm{Re}\left[i^{n+1/2} \int_0^\infty \omega^{n+1/2} \cos(\omega t)\, d\omega\right] \qquad (3.49)$$

$$= \frac{\Gamma(n+3/2)}{2t^{n+3/2}}\,\mathrm{Re}\left[i^{n+1/2}(i^{n+3/2} + (-i)^{n+3/2})\right] .$$

By using that $-i = i^3$ we can write

$$\mathrm{Re}\left[i^{n+1/2}(i^{n+3/2} + (-i)^{n+3/2})\right] = \mathrm{Re}\left[(i^{2n+2} + i^{4n+5})\right]$$

$$= \mathrm{Re}\left[i - (-1)^n\right] = (-1)^{n+1} . \qquad (3.50)$$

Using (3.50) and (3.49) and further more by noting that the numerator in the second sum in (3.43) is zero when $n = 0$, we finally find that I defined in (3.42) is given as:

$$I = \sum_{n=1}^{\infty} \frac{3(2n+3) - 2(2n+3)(n+1) - 3}{(2n+3)!}(-1)^n \frac{\Gamma(n+3/2)}{2} \frac{(\mu_0 \sigma r^2)^{n+1/2}}{t^{n+3/2}} \qquad (3.51)$$

and for the sake of completeness let us state that

$$\Gamma(n+3/2) = \frac{1 \cdot 3 \cdot 5 \cdot 7 \cdots (2n+1)}{2^{n+1}} \pi^{1/2} . \qquad (3.52)$$

From (3.42), (3.51) and (3.52) we see that the unit step response for the GDC and CL configurations, in the case of a homogeneous half-space with conductivity σ, has the following series expansion:

$$v_{D,C} = C_{D,C} \sum_{n=1}^{\infty} \frac{3(2n+3) - 2(2n+3)(n+1) - 3}{2^n(2n+3)(2\cdot4\cdot6\cdot8\cdots(2n+2))} \frac{(-1)^n}{\pi^{1/2}} \frac{(\mu_0\sigma r^2)^{n+1/2}}{t^{n+3/2}} . \quad (3.53)$$

For large t the first term in the series is dominant and we get

$$v_{D,C}(r,t) \approx \frac{C_{D,C}}{10\pi^{1/2}} \frac{(\mu_0\sigma r^2)^{3/2}}{t^{5/2}} \quad (3.54)$$

showing that the transient voltage is, for late times after the current in the transmitter is abruptly turned off, proportional to $\sigma^{3/2}$ and falls off with time as $-t^{-5/2}$. Because of this asymptotic behaviour of the transient voltage, it is customary to define a so called **late time apparent resistivity** obtained by solving (3.54) for the resistivity $\rho = 1/\sigma$ resulting in

$$\rho_{a\,D,C}(r,t) = \frac{\mu_0 r^2}{4} \left[\frac{4C_{D,C}}{5\pi^{1/2}t^{5/2}\, v_{D,C}(r,t)} \right]^{2/3} \quad (3.55)$$

or by using the definitions (3.34) of C_D and C_C:

$$\rho_{a\,D}(r,t) = \frac{\mu_0 r^2}{4\pi} \left[\frac{\sin(\theta) L A_s n_s \mu_0}{5r^2 t^{5/2}\, v_D(r,t)} \right]^{2/3} \quad (3.56)$$

for the GDC configuration and

$$\rho_{a\,C}(r,t) = \frac{\mu_0}{4\pi} \left[\frac{2\mu_0 A_r n_r A_s n_s}{5t^{5/2}\, v_C(r,t)} \right]^{2/3} \quad (3.57)$$

for the CL configuration. If the transient response for a homogeneous half-spaces is plotted as a late time apparent resistivity and as a function of time, it will approach asymptotically the true resistivity of the half-space for late times. This is seen on figure 3.10 where the transient response for uniform half-spaces with resistivities 1, 10 and 100 Ωm and for r = 100 m is shown as an apparent resistivity versus time according to equation (3.56) for GDC configuration or (3.57) for CL configuration. We see on this figure the apparent resistivity approaches the true resistivity of the half-space at later times as the resistivity gets lower. This is a reflection of the fact that the voltage response reaches the late stage at later times as the resistivity is lower as is seen on figure 3.5. Figures 3.11 and 3.12 show the transient response as late time apparent resistivity versus time for a two layered earth and r = 100 m. Figure 3.11 shows the case of a 100 Ωm layer with different thickness (1, 10, 100 and 1000 m) on top of a homogeneous half-space with resistivity 1 Ωm. On this figure we see that when the surface layer is thin, d < 10 m, the apparent resistivity shows only minor deviations from that of the homogeneous half-space below, except at very early times. When the surface layer gets thicker, d ≥ 100 m, the apparent resistivity coincides with the apparent resistivity for a homogeneous half-space with resistivity 100 Ωm for early times but approaches the resistivity of the underlaing half-space for late times. When the upper layer is thick, the apparent resistivity levels on the true resistivity of the upper layer because in this case the voltage

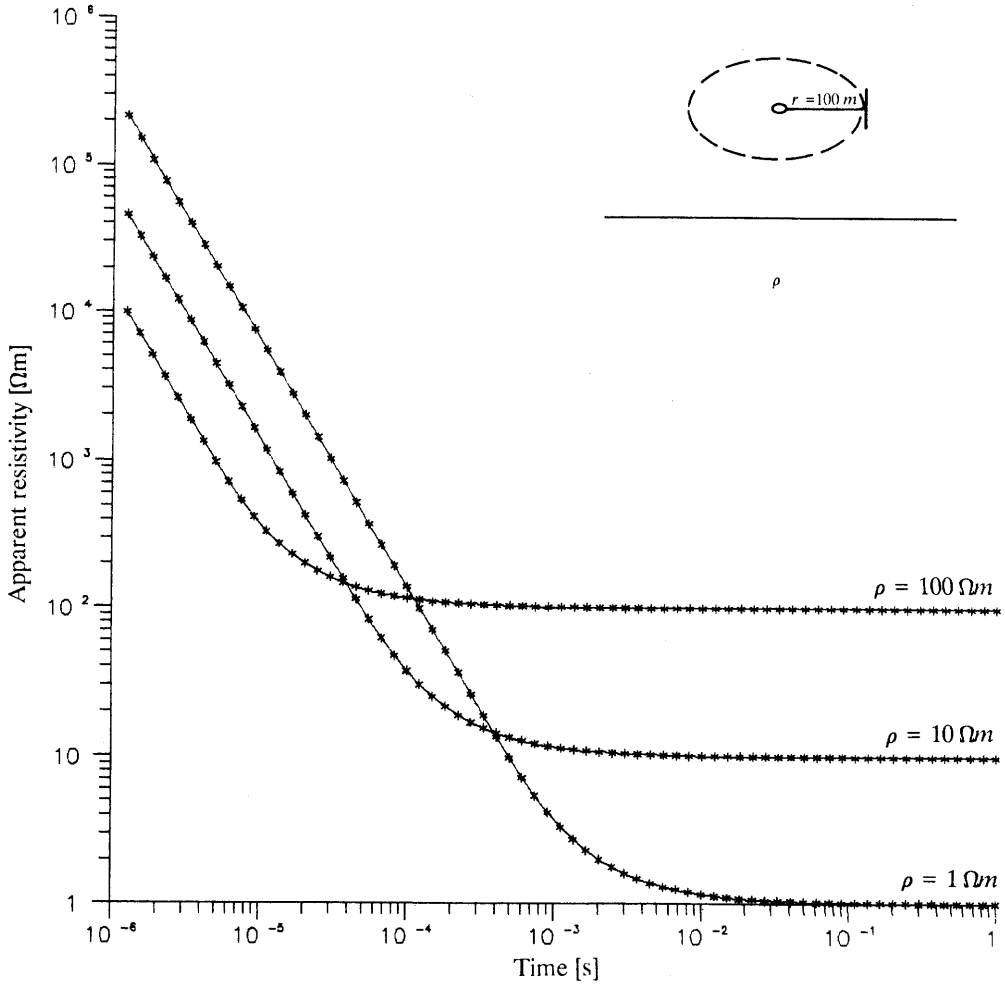

Figure 3.10 Late time apparent resistivity for homogeneous half-space

response coincides with the late time response for a homogeneous half-space with resistivity equal to that of the upper layer (see figure 3.8). We furthermore see that, before the apparent resistivity starts to decay down to the resistivity of the underlaing half-space it makes an overshoot of the upper layer resistivity reflecting the undershoot in the voltage response seen on figure 3.8.

Figure 3.12 shows the late time apparent resistivity for a two layered earth with layers of resistivity 1 Ωm and different thicknesses, 1, 10 and 100 m, over a homogeneous half-space with resistivity 100 Ωm and for r = 100 m. When the surface layer is thin, d $\leq$ 1 m, the apparent resistivity is in between the two half-space extremes for early times. Then it makes an undershoot and finally approaches the true resistivity of the underlaing half-space. When the laier gets thicker, the apparent resistivity coincides with the apparent resistivity of a homogeneous half-space with resistivity equal to that of the top layer for early times. Then it

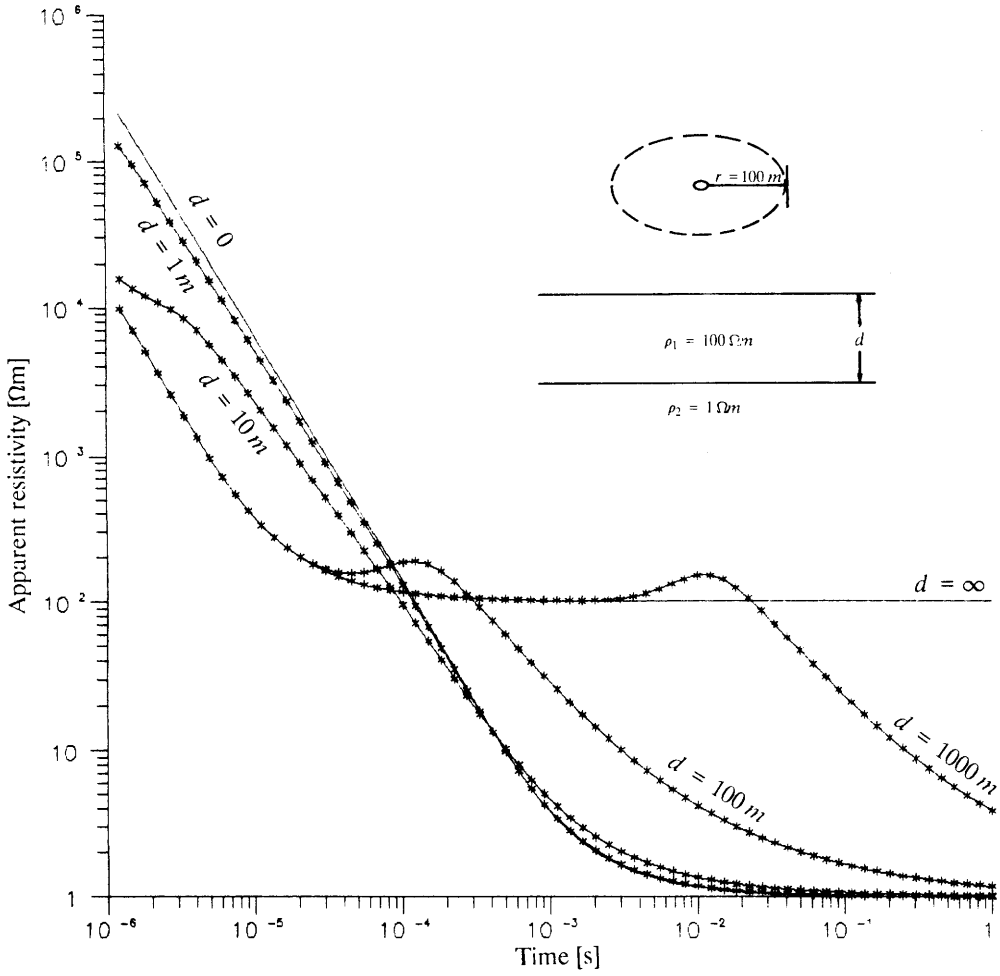

Figure 3.11 Late time apparent resistivity for two layered earth, resistive layer on top of conductive half-space

makes an under shot which decreases with increasing thickness of the top layer and for thick layer the under shot disappears and the apparent resistivity reaches the true resistivity of the top layer. Finally the apparent resistivity approaches the true resistivity of the under laing half-space at increasingly later time with increasing thickness of the top layer.

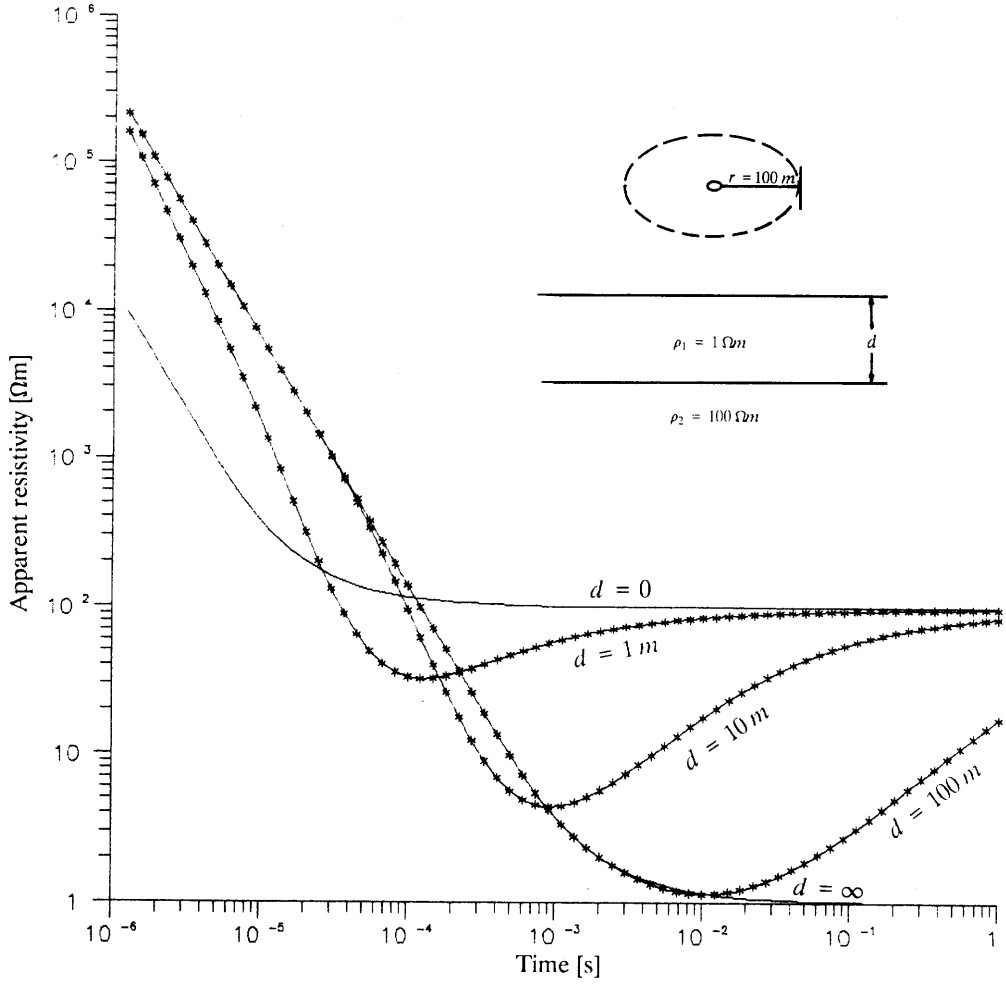

Figure 3.12 Late time apparent resistivity for two layered earth, conductive layer on top of resistive half-space

4. NUMERICAL INTEGRATION USING DIGITAL FILTERS

In section 3.3 we showed that the unit step response for the GDC and CL configurations are given in terms of the following integral :

$$v_{D,C}\text{-}(r,t) = \frac{2}{\pi} C_{D,C} \int_0^\infty \mathrm{Re}\,[E^e(\omega,r)] \cos(\omega t)\, d\omega \tag{3.35}$$

where $C_{D,C}$ is given by (3.34) and the earth response factor is given as :

$$E^e(\omega,r) = 2r^2 \int_0^\infty \lambda \frac{S_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda \,. \tag{2.104}$$

S_0 and T_0 in (2.104) are given by the recursion relations (2.75).

In the case of a homogeneous half space, we managed to calculate these integrals analytically but in the general case they have to be evaluated numerically. The task of this section is to describe a convenient and efficient method to perform numerical integration of integrals like (3.35) and (2.104). This method, usually called the **digital filter method**, was developed by Kunetz, Ghosh and Koefoed (Kunetz 1966, Ghosh 1971, 1971a and Koefoed 1972).

4.1 THE DIGITAL FILTER METHOD

The integrals in (3.35) and (2.104) can both be written on the form :

$$f(x) = x \int_0^\infty g(y)\, h(xy)\, dy \,. \tag{4.1}$$

In the case of (2.104) we have identified

$$f(r) = \frac{1}{r} E^e(\omega,r) \quad ; \quad g(\lambda) = 2\lambda \frac{S_0}{S_0 - T_0} \quad ; \quad h(\lambda r) = J_1(\lambda r) \tag{4.1a}$$

and in the case of (3.35) we have

$$f(t) = v_{D,C}\text{-}(r,t) \frac{t\pi}{2C_{D,C}} \quad ; \quad g(\omega) = Re\,[E^e(\omega,r)] \quad ; \quad h(\omega t) = \cos(\omega t) \,. \tag{4.1b}$$

The first step is to turn (4.1) into a convolution integral by changing the variables from the linear variables x and y to the logarithmic variables τ and η where :

$$x = e^\eta \; ; \; y = e^\tau \,. \tag{4.2}$$

Substituting this into (4.1) we get :

$$f(e^{\eta}) = \int_{-\infty}^{\infty} g(e^{\tau})\, h(e^{\eta+\tau})\, e^{\eta+\tau}\, d\tau \ . \tag{4.3}$$

This we write as :

$$F(\eta) = \int_{-\infty}^{\infty} G(\tau)\, H(\eta+\tau)\, d\tau \tag{4.4}$$

where we have defined :

$$F(\tau) = f(e^{\tau}) \ ; \ G(\tau) = g(e^{\tau}) \ ; \ H(\tau) = h(e^{\tau})e^{\tau} \ . \tag{4.4a}$$

The next step is to use the **sampling theorem** to express the function $G(\tau)$ in terms of discretely sampled values. The sampling theorem is a powerful theorem stating that if G is a band limited function, that is to say if $\tilde{G}$ is the Fourier transform of G and

$$\tilde{G}(\kappa) = 0 \quad \text{for} \quad |\kappa| > \frac{\pi}{\Delta} \tag{4.5}$$

for some $\Delta > 0$, then the function G can be written as :

$$G(\tau) = \sum_{n=-\infty}^{\infty} \frac{\sin[((\tau-\tau_0)/\Delta - n)\pi]}{((\tau-\tau_0)/\Delta - n)\pi}\, G(\tau_0 + n\Delta) \tag{4.6}$$

where τ_0 is arbitrary. This is a remarkable theorem stating that a band limited function can be completely regenerated as a sum of sinc functions multiplied by evenly spaced sample values of the function, provided that it is sampled densely enough. We will, in the next section, proof the sampling theorem and see what happens if the function G is not truly band limited or if it is not sampled densely enough. But for the moment we take (4.6) for granted and continue developing the digital filter method.

Assuming that the kernel function G in (4.4) is band limited and using (4.6), equation (4.4) becomes :

$$F(\eta) = \sum_{n=-\infty}^{\infty} G(\tau_0 + n\Delta) \int_{-\infty}^{\infty} \frac{\sin[((\tau-\tau_0)/\Delta - n)\pi]}{((\tau-\tau_0)/\Delta - n)\pi}\, H(\eta+\tau)\, d\tau \ . \tag{4.7}$$

The integrals in (4.7) are independent of the function G and and do only depend on the filter function H and on τ_0 , Δ and η, that is to say on the abscissa values at which G is sampled and the abscissa at which F is to be evaluated. It is therefore evident that if the function G, satisfies (4.5), then the integral in (4.4) can be written as :

$$F(\eta) = \sum_{n=-\infty}^{\infty} G(\tau_0 + n\Delta)\, b_n(\eta,\tau_0) \ . \tag{4.8}$$

where the **digital filter coefficients**

$$b_n(\eta,\tau_0) = \int_{-\infty}^{\infty} \frac{\sin[((\tau - \tau_0)/\Delta - n)\pi]}{((\tau - \tau_0)/\Delta - n)\pi} H(\eta + \tau)\, d\tau \ , \tag{4.9}$$

are independent of G and only depend on H, η, τ_0 and Δ.

The filter coefficients defined by (4.9) are in general not easily determined and we will later discuss methods for determining them. Since the coefficients do depend on η one might wonder if we have really gained much if we have to work out the whole set of filter coefficients for each abscissa value η at which we want to determine F. Fortunately we do not have to do so and that makes the digital filter method an extremely powerful method for calculating integrals of the type (4.4). Assume that we have determined the filter coefficients $b_n(\eta_0,\tau_0)$ so that

$$F(\eta_0) = \sum_{n=-\infty}^{\infty} G(\tau_0 + n\Delta)\, b_n(\eta_0,\tau_0) \tag{4.10}$$

and we want to evaluate $F(\eta_0 + \xi)$. Going back to (4.7) we find that

$$F(\eta_0 + \xi) = \sum_{n=-\infty}^{\infty} G(\tau_0 + n\Delta) \int_{-\infty}^{\infty} \frac{\sin[((\tau' - \tau_0)/\Delta - n)\pi]}{((\tau' - \tau_0)/\Delta - n)\pi} H(\eta_0 + \xi + \tau')\, d\tau' \ . \tag{4.11}$$

If we change the integration variable from τ to $\tau = \tau' + \xi$ we get

$$F(\eta_0 + \xi) = \sum_{n=-\infty}^{\infty} G(\tau_0 + n\Delta) \int_{-\infty}^{\infty} \frac{\sin[((\tau - \xi - \tau_0)/\Delta - n)\pi]}{((\tau - \xi - \tau_0)/\Delta - n)\pi} H(\eta_0 + \tau)\, d\tau \ . \tag{4.12}$$

But now we recall that the sampling theorem stated that τ_0 in (4.12) can be taken arbitrary. We replace therefore τ_0 by $\tau_0 - \xi$ and (4.12) becomes

$$\begin{aligned} F(\eta_0 + \xi) &= \sum_{n=-\infty}^{\infty} G(\tau_0 - \xi + n\Delta) \int_{-\infty}^{\infty} \frac{\sin[((\tau - \tau_0)/\Delta - n)\pi]}{((\tau - \tau_0)/\Delta - n)\pi} H(\eta_0 + \tau)\, d\tau \\ &= \sum_{n=-\infty}^{\infty} G(\tau_0 - \xi + n\Delta)\, b_n(\eta_0,\tau_0) \end{aligned} \tag{4.13}$$

We see therefore that we can use the same filter coefficients to calculate $F(\eta_0 + \xi)$ as we used to calculate $F(\eta_0)$ but we have to sample the kernel function G at the shifted abscissa

values $\tau_0 - \xi + n\Delta$ instead of $\tau_0 + n\Delta$. A convenient use can be made of this shifting property if we evaluate F at discrete, evenly spaced abscissa values with the same abscissa spacings, Δ, as we used to sample G that is to say at the abscissa values $\eta_0 + m\Delta$. In this case (4.13) becomes

$$F(\eta_0 + m\Delta) = \sum_{n=-\infty}^{\infty} G(\tau_0 + (n - m)\Delta)\, b_n(\eta_0,\tau_0) \ . \tag{4.14}$$

This equation demonstrates the power of the digital filter method to calculate convolution integrals of the form (4.4) for different but band limited kernel functions G. The filter coefficients b_n are the same for all kernel functions that are band limited by π/Δ and the same set of sample values of G can be used to evaluate F at evenly spaced abscissa values (with the same sample spacing, Δ, as for G) by shifting the sample values of G relative to the filter.

If the kernel function G is band limited, i.e. satisfying (4.5), then equations (4.13) and (4.14) are exact. The kernel functions that we encounter in practice are usually not truly band limited but their Fourier transform, $\tilde{G}(\kappa)$, usually tends asymptotically to zero for $|\kappa| \to \infty$. They can therefore be approximated, to the desired accuracy, by equation (4.6) if we take the sample spacing Δ sufficiently small. It is easy, for a given sample spacing, to check if the kernel function is represented to the desired accuracy by the sum in (4.6). We evaluate G at the sample points $\tau_0 + n\Delta$ and work out the sum for intermediate abscissa values, e.g. $\tau_0 + n\Delta + \Delta/2$, and compare the results with the exact values $G(\tau_0 + n\Delta + \Delta/2)$. If the sum and the exact values do not agree well enough, the sample spacing Δ is decreased until the desired accuracy is obtained.

The bandwidth of a function $g(y)$ is greatly influenced by changing from linear to logarithmic variables by the substitution (4.2). If the function is broad banded, that is to say if its Fourier transform $\tilde{g}(k)$ is non-zero or tends slowly to zero for large values of $|k|$ then $\tilde{G}(\kappa)$, the Fourier transform of $G(\tau)$, tends much faster to zero for large $|\kappa|$. On the other hand, if $g(y)$ is band limited, that is to say $\tilde{g}(k)$ is zero for $|k|$ larger than a finite value then $G(\tau)$ is generally not band limited.

To see that this is the case we assume that $g(y)$ has the Fourier expansion

$$g(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{g}(k)\, e^{iky}\, dk \ . \tag{4.15}$$

The functions involved in (4.1) are only defined for $y > 0$ but, in order for their Fourier transforms to have a meaning, they must also be defined for $y < 0$. The simplest thing to do is to define $g(y)$ as either symmetric or a-symmetric, that is $g(-y) = \pm g(y)$. It is a trivial exercise to see that $\tilde{g}(k)$ has the same symmetry as $g(y)$, that is $\tilde{g}(-k) = \pm\tilde{g}(k)$. Let us for

the moment take $g(y)$ symmetric. By changing to logarithmic variable by substituting $y = e^{\tau}$ in (4.15) and using (4.4a) we get

$$G(\tau) = g(e^{\tau}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{g}(k)\, e^{ike^{\tau}}\, dk \quad . \tag{4.16}$$

We now take the Fourier transform of (4.16) and get

$$\tilde{G}(\kappa) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(\tau)\, e^{-i\kappa\tau}\, d\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{g}(k)\, e^{ike^{\tau}} e^{-i\kappa\tau}\, d\tau\, dk \quad . \tag{4.17}$$

By going back to the linear variable $y = e^{\tau}$ and by changing the integration variable in the k-integration from k to $-k$ and using that $\tilde{g}(-k) = \tilde{g}(k)$ this can be written as

$$\tilde{G}(\kappa) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{g}(k) \int_{0}^{\infty} e^{-iky}\, y^{-i\kappa-1}\, dy\, dk \quad . \tag{4.18}$$

Using now the definition of the Γ function (3.45) we find that

$$\tilde{G}(\kappa) = \frac{\Gamma(-i\kappa)}{2\pi} \int_{-\infty}^{\infty} \tilde{g}(k)\, (ik)^{i\kappa}\, dk \tag{4.19}$$

which, by using the symmetry of $\tilde{g}$, can be written as

$$\tilde{G}(\kappa) = \frac{\Gamma(-i\kappa)}{2\pi} [i^{i\kappa} + (-i)^{i\kappa}] \int_{0}^{\infty} \tilde{g}(k)\, k^{i\kappa}\, dk = \frac{\Gamma(-i\kappa)}{\pi} \cosh(\kappa\pi/2) \int_{0}^{\infty} \tilde{g}(k)\, k^{i\kappa}\, dk \tag{4.20}$$

where we have in the last step used that $i = e^{i\pi/2}$ and $-i = e^{-i\pi/2}$. If we had taken $g(y)$ a-symmetric instead of symmetric we would have got

$$\tilde{G}(\kappa) = \frac{\Gamma(-i\kappa)}{\pi} \sinh(\kappa\pi/2) \int_{0}^{\infty} \tilde{g}(k)\, k^{i\kappa}\, dk \tag{4.20a}$$

To investigate the behaviour of $\tilde{G}(\kappa)$ we note that (see Abramowitz and Stegun 1970, p. 257)

$$|\Gamma(x + iy)| \approx \sqrt{2\pi}\, e^{-\frac{\pi}{2}|y|}\, |y|^{x-1/2} \quad \text{for} \quad |y| \to \infty \tag{4.21}$$

and we also note that

$$|\cosh(\kappa\pi/2)| \approx |\sinh(\kappa\pi/2)| \approx \frac{1}{2} e^{\frac{\pi}{2}|y|} \quad \text{for} \quad |\kappa| \to \infty \tag{4.22}$$

We have therefore that

$$|\tilde{G}(\kappa)| \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{|\kappa|}} \left| \int_0^\infty \tilde{g}(k)\, k^{i\kappa}\, dk \right| \quad \text{for} \quad |\kappa| \to \infty \,. \tag{4.23}$$

The factor $k^{i\kappa} = e^{i\kappa \ln k}$ in the integrand in (4.23) will oscillate wildly when $|\kappa|$ is large except for $\ln k \approx 0$. If $\tilde{g}(k)$ is a well behaved, smooth and bounded function of k then the oscillating factor will cause the integral to average out, except for $\ln k \approx 0$. The higher the value of $|\kappa|$ is, the more is the averaging and hence lower the value of the integral. This, along with the factor $1/\sqrt{|\kappa|}$, causes $|\tilde{G}(\kappa)|$ to tend rapidly to zero for large $|\kappa|$ even though $|\tilde{g}(k)|$ tends slowly to zero for large $|k|$. If for example $|\tilde{g}(k)|$ behaves as

$$\tilde{g}(k) \approx \frac{\tilde{g}_0\, k^s}{(a^2 + k^2)^r} \quad ; \; r > (r+1)/2 \tag{4.24}$$

for large $|k|$, then insertion into (4.23) gives

$$|\tilde{G}(\kappa)| \approx \frac{1}{\sqrt{2\pi}} \frac{|\tilde{g}_0|}{\sqrt{|\kappa|}} \left| \int_0^\infty \frac{k^{i\kappa + s}}{(a^2 + k^2)^r}\, dk \right| \quad \text{for} \quad |\kappa| \to \infty \,. \tag{4.25}$$

The integral in (4.25) is tabulated and the result is

$$\int_0^\infty \frac{k^{i\kappa + s}}{(a^2 + k^2)^r}\, dk = a^{i\kappa + s + 1 - 2r} \frac{\Gamma(i\kappa/2 + (s+1)/2)\, \Gamma(-i\kappa/2 - (s+1)/2 - r)}{2\,\Gamma(r)} \,. \tag{4.26}$$

By inserting this into (4.25) and using (4.21) we find that

$$|\tilde{G}(\kappa)| \approx |\tilde{g}_0| \sqrt{\frac{\pi}{2}} \frac{|a|^{s+1-2r}}{\Gamma(r)} \frac{e^{-\frac{\pi}{2}|\kappa|}}{\sqrt{|\kappa|}} \left|\frac{\kappa}{2}\right|^{r-1} \quad \text{for} \quad |\kappa| \to \infty \,. \tag{4.27}$$

We see therefore that if $\tilde{g}(k)$ tends to zero as $1/k^{2r-s}$ for large $|k|$ values then $|\tilde{G}(\kappa)|$ tends exponentially to zero for large values of $|\kappa|$, showing that $\tilde{G}$ tends much faster to zero for large abscissa values than $\tilde{g}$ does.

As an explicit example, let us consider the function $g(y) = e^{-a|y|}$. The Fourier transform of g is given as

$$\tilde{g}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-a|y| - iky}\, dy = \frac{1}{\sqrt{2\pi}} \frac{2a}{a^2 + k^2} \tag{4.28}$$

Equations (4.24) and (4.28) are identical if we take $s = 0$, $r = 1$ and $\tilde{g}_0 = 2a/\sqrt{2\pi}$. Inserting these values into (4.27) we find

$$|\tilde{G}(\kappa)| \approx \frac{e^{-\frac{\pi}{2}|\kappa|}}{\sqrt{|\kappa|}} \quad \text{for} \quad |\kappa| \to \infty \ . \tag{4.29}$$

It is worth noting that $G(\tau) = e^{-ae^{\tau}}$ has similar behaviour as the real part of the earth response factor (see figures 3.4, 3.6 and 3.7).

If the function $g(y)$ is band limited then it is obvious from (4.20) and (4.20a) that $\tilde{G}(\kappa)$ is in general not band limited. If for example we take $g(y) = x\cos(xy)$ (compare with equation (3.35)), it is easily seen that

$$\tilde{g}(k) = \sqrt{2\pi}\,\frac{x}{2}\,[\delta(k-x) + \delta(k+x)] \tag{4.30}$$

which is two delta spikes, at $k = -x$ and $k = x$, and hence g is band limited. From (4.20) we immediately see that

$$\tilde{G}(\kappa) = \frac{x^{i\kappa+1}}{\sqrt{2\pi}}\,\Gamma(-i\kappa)\cosh(\kappa\pi/2) \ . \tag{4.31}$$

From (4.21) and (4.22) we see now that

$$|\tilde{G}(\kappa)| \approx \frac{1}{2\sqrt{|\kappa|}} \quad \text{for} \quad |\kappa| \to \infty \ . \tag{4.32}$$

and $\tilde{G}(\kappa)$ is clearly not band limited.

The kernel functions $g(\lambda) = 2\lambda\frac{S_0}{S_0 - T_0}$ and $g(\omega) = Re\,[E^e(\omega,r)]$ that we will encounter in the convolution integral (4.1) are broad banded and their brodbandness or bandwidth will therefore be reduced by changing from linear to logarithmic variables. The reduction of the bandwidth makes it better justified to use the sampling theorem to express the kernel function in terms of discretely sampled values and does also allow greater sample spacing Δ. As we will see, it is sufficient, for the kernel functions that we will encounter here, to sample at five points per power of e, that is taking $\Delta = 0.2$.

For the kernel function $g(y) = e^{-|y|}$, we see that if we sample with $\Delta = 0.2$ then we have from (4.28) that for $|k| \geq \pi/\Delta = 5\pi$

$$\tilde{g}(k) \leq \frac{1}{\sqrt{2\pi}}\,\frac{2}{1+(5\pi)^2} = 3.22\cdot10^{-3} \tag{4.33}$$

but from (4.29) we see that for $|\kappa| \geq \pi/\Delta = 5\pi$

$$\tilde{G}(\kappa) \leq \frac{1}{\sqrt{5\pi}}\,e^{-\frac{5}{2}\pi^2} = 4.85\cdot10^{-12} \tag{4.33a}$$

and we see that (4.5) is much better approximated in the logarithmic than in the linear

representation.

Once the sampling spacing has been determined, the filter coefficients defined in (4.9), can be determined and we will later discuss how that is done. The sum in (4.13) or (4.14) has, in practical use, to be truncated at finite positive and negative values of n. The question about the behaviour of the filter coefficients for large positive and negative n is therefore important. In order to be able to truncate the sum, the terms for large positive and negative values of n, must become vanishingly small and we want to evaluate the sum to an acceptable accuracy with as few terms as possible. Fortunately it turns out that the filter coefficients for the filter functions in the integrals (3.35) and (2.104) do have a maximum in their absolute values for a given n and tail rapidly off towards zero for both increasing and decreasing n. This can be seen on figures 4.1 and 4.2.

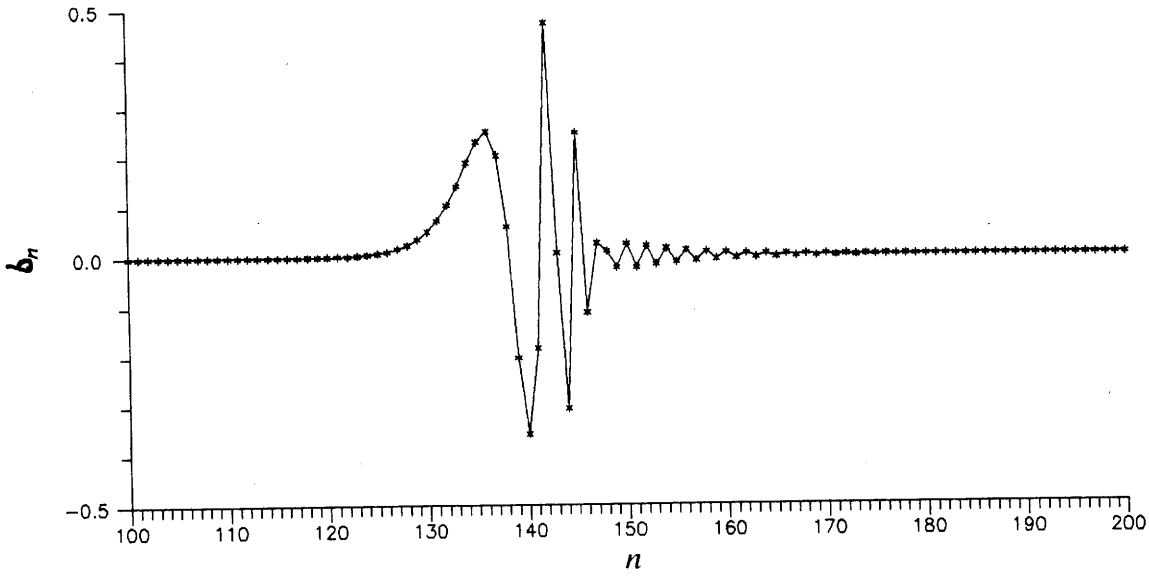

Figure 4.1 Digital filter for J_1-Hankel transform

Figure 4.1 shows a digital filter for performing the J_1-Hankel transform in (2.104) and figure 4.2 shows a digital filter for performing the cosine transform in (3.35). Both of these filters are constructed with the sample spacing of five points per power of e, that is with $\Delta = 0.2$. The Hankel transform filter was constructed with $\eta_0 = 0$ and $\tau_0 = -26.306681$ and the cosine transform filter with $\eta_0 = 0$ and $\tau_0 = -30.302512$.

In section 3.3 we saw that the real part of the earth response factor E^e is a well behaved and bounded function of the angular frequency ω and with well defined limiting values (1 and 0)

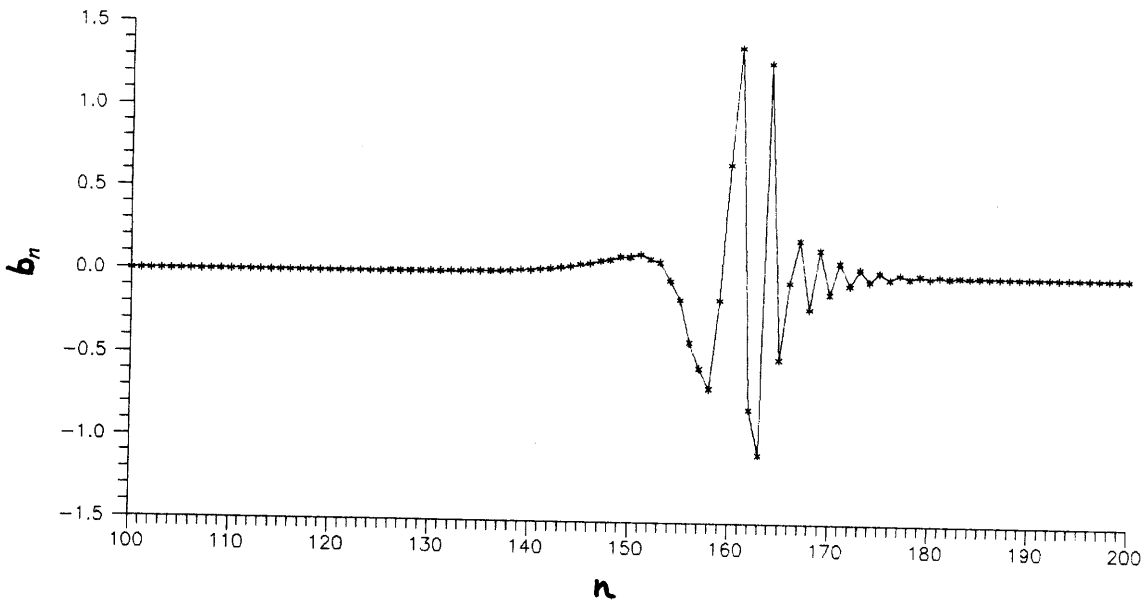

Figure 4.2 Digital filter for cosine transform

when $\ln(\omega) \rightarrow \pm \infty$ (see Figures 3.4, 3.6 and 3.7). It is therefore evident, when performing the cosine transform (3.35), that the sum in (4.13) or (4.14) can be truncated at finite values of $|n|$.

The situation is not quite as simple in the case of the Hankel transform because the kernel function $g(\lambda) = 2\lambda \dfrac{S_0}{S_0 - T_0}$ in (2.104) is not as well behaved as the earth response factor. For a homogeneous half-space we have for example (see equation (2.87a)) in the quasi-stationary approximation that $S_0 = 1$ and $T_0 = -\sqrt{\lambda^2 + i\omega\mu\sigma}\,/\lambda$ and hence

$$g(\lambda) = \frac{2\lambda^2}{\lambda + \sqrt{\lambda^2 + i\omega\mu\sigma}} \rightarrow \begin{cases} 0 & \text{for } \lambda \rightarrow 0 \\ \lambda & \text{for } \lambda \rightarrow \infty \end{cases} . \tag{4.34}$$

The asymptotic behaviour of $g(\lambda)$ is the same for layered earth. From this we see that the kernel function $G(\tau)$ for the Hankel transform is not bounded but behaves as e^{τ} for $\tau \rightarrow \infty$ and one might worry that the sum in (4.13) is not at all convergent. It turns out that it is convergent and can be truncated at finite values of $|n|$, but we need many terms to obtain reasonable accuracy.

We can avoid the unfortunate exponential behaviour of the kernel function and evaluate the earth response factor by a much faster converging sum by using a simple trick. Let us instead

of E^e evaluate $E^e - 1$. By using (2.100) and (2.104) we see that

$$E^e - 1 = r^2 \int_0^\infty \lambda [\frac{2S_0}{S_0 - T_0} - 1] J_1(\lambda r)\, d\lambda = r^2 \int_0^\infty \lambda \frac{S_0 + T_0}{S_0 - T_0} J_1(\lambda r)\, d\lambda \,. \qquad (4.35)$$

For a homogeneous half-space in the quasi-stationary approximation we see that instead of (4.34) and (4.1a) we now have

$$g'(\lambda) = \lambda \frac{\lambda - \sqrt{\lambda^2 + i\omega\mu\sigma}}{\lambda + \sqrt{\lambda^2 + i\omega\mu\sigma}} \quad \rightarrow \quad \begin{cases} 0 & \text{for } \lambda \rightarrow 0 \\ 0 & \text{for } \lambda \rightarrow \infty \end{cases} \qquad (4.36)$$

and again the asymptotic behaviour for layered earth is the same. We see therefore that the sample values $G(\tau_0 + n\Delta)$, of the kernel function tend to zero for large positive and negative values of n. The sum for evaluating $E^e - 1$ converges therefore much faster than the sum for E^e and can be truncated at much smaller values of n.

4.2 THE SAMPLING THEOREM

Since the sampling theorem is of such a fundamental importance in the theory of digital filters, it deserves a proof. The purpose of the proof is not just to maintain mathematical rigour in the development of the digital filter theory but also to learn how to use the theorem properly and see the possible pitfalls. By going through the proof we will see the importance of the band limitedness of the function to be sampled and learn how to adjust the sample spacing according to the bandwidth. We will also see the problem, called aliasing, which we run into if do not sample densely enough.

Before we can start to proof the sampling theorem, we need to proof a little lemma stating that the Fourier transform of the *sampling function*

$$S(t,t_0,\Delta) = \sum_{n=-\infty}^{\infty} \delta(t - t_0 - n\Delta) \qquad (4.37)$$

is given as

$$\tilde{S}(k,t_0,\Delta) = \frac{\sqrt{2\pi}}{\Delta} e^{-ikt_0} \sum_{n=-\infty}^{\infty} \delta(k - n2\pi/\Delta) = \frac{\sqrt{2\pi}}{\Delta} e^{-ikt_0} S(k,0,2\pi/\Delta) \,. \qquad (4.38)$$

The assertion stated in (4.38) is by definition of the delta "function" equivalent to the statement that

$$\tilde{S}(k,t_0,\Delta) = 0 \quad \text{if} \quad k \neq n\frac{2\pi}{\Delta} \qquad (4.39a)$$

$$\int_{-\infty}^{\infty} \tilde{S}(k,t_0,\Delta) f(k)\, dk = \frac{\sqrt{2\pi}}{\Delta} \sum_{n=-\infty}^{\infty} e^{-in2\pi/\Delta t_0} f(n2\pi/\Delta) \tag{4.39b}$$

where $f(k)$ is an arbitrary function and this we will show.

Taking the Fourier transform of (4.37) we find that

$$\tilde{S}(k,t_0,\Delta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikt} \sum_{n=-\infty}^{\infty} \delta(t - t_0 - n\Delta)\, dt = \frac{1}{\sqrt{2\pi}} e^{-ikt_0} \sum_{n=-\infty}^{\infty} e^{-ink\Delta} \ . \tag{4.40}$$

To evaluate the series in (4.40) we substitute it by two series, one for $n \geq 0$ and another one $n \leq 0$ and make them absolutely convergent by exponential damping for large n and write

$$\sqrt{2\pi}\, e^{ikt_0} \tilde{S}(k,t_0,\Delta) = \lim_{\varepsilon_1 \to 0} \sum_{n=0}^{\infty} e^{-n(ik\Delta + \varepsilon_1)} + \lim_{\varepsilon_2 \to 0} \sum_{n=0}^{\infty} e^{n(ik\Delta - \varepsilon_2)} - 1 \ . \tag{4.41}$$

We have to subtract one because we have counted the term for $n = 0$ twice. The two series in (4.41) are simple geometric series of the form

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \tag{4.42}$$

We can therefore write (4.41) on the form

$$\sqrt{2\pi}\, e^{ikt_0} \tilde{S}(k,t_0,\Delta) = \lim_{\varepsilon_1 \to 0} \frac{1}{1 - e^{-ik\Delta - \varepsilon_1}} + \lim_{\varepsilon_2 \to 0} \frac{1}{1 - e^{ik\Delta - \varepsilon_2}} - 1 \ . \tag{4.43}$$

For $k \neq n2\pi/\Delta$ the denominators are non-zero in the limit $\varepsilon \to 0$ and we have

$$\sqrt{2\pi}\, e^{ikt_0} \tilde{S}(k,t_0,\Delta) = \frac{1}{1 - e^{-ik\Delta}} + \frac{1}{1 - e^{ik\Delta}} - 1 = 0 \tag{4.44}$$

and we have established (4.39a). To show (4.39b) we use (4.39a) to write

$$\int_{-\infty}^{\infty} \tilde{S}(k) f(k)\, dk = \sum_{n=-\infty}^{\infty} \lim_{\delta \to 0} \int_{-\delta}^{\delta} \tilde{S}(n2\pi/\Delta + x) f(n2\pi/\Delta + x)\, dx \ . \tag{4.45}$$

Using (4.43) this can be written as

$$\int_{-\infty}^{\infty} \tilde{S}(k) f(k)\, dk = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \left[\lim_{\delta \to 0} \lim_{\varepsilon_1 \to 0} \int_{-\delta}^{\delta} \frac{f(n2\pi/\Delta + x)}{1 - e^{-ix\Delta - \varepsilon_1}} e^{-i(n2\pi/\Delta + x)t_0}\, dx \right. \tag{4.46}$$

$$+ \lim_{\delta \to 0} \lim_{\varepsilon_2 \to 0} \int_{-\delta}^{\delta} \frac{f(n2\pi/\Delta + x)}{1 - e^{ix\Delta - \varepsilon_2}} e^{-i(n2\pi/\Delta + x)t_0}\, dx$$

$$\left. + \lim_{\delta \to 0} \int_{-\delta}^{\delta} f(n2\pi/\Delta + x)\, e^{-i(n2\pi/\Delta + x)t_0}\, dx \right] .$$

The last integral under the sum in (4.46) vanishes in the limit $\delta \to 0$. In the limit $\delta \to 0$ and $\varepsilon_1 \to 0$, the integral in the first term in (4.46) can be written as

$$\int_{-\delta}^{\delta} \frac{f(n2\pi/\Delta + x)}{1 - e^{-ix\Delta - \varepsilon_1}} e^{-i(n2\pi/\Delta + x)t_0}\, dx \approx f(n2\pi/\Delta) e^{-int_0 2\pi/\Delta} \int_{-\delta}^{\delta} \frac{dx}{1 - (1 - ix\Delta - \varepsilon_1)}$$

$$= f(n2\pi/\Delta) e^{-int_0 2\pi/\Delta} \int_{-\delta}^{\delta} \frac{dx}{ix\Delta + \varepsilon_1} . \tag{4.47}$$

Since both δ and ε_1 are tending to zero we have in (4.47) only kept lowest order non-vanishing terms of δ and ε_1 both in the numerator and the denominator. We have therefore

$$\lim_{\delta \to 0} \lim_{\varepsilon_1 \to 0} \int_{-\delta}^{\delta} \frac{f(n2\pi/\Delta + x)}{1 - e^{-ix\Delta - \varepsilon_1}} e^{-i(n2\pi/\Delta + x)t_0}\, dx \tag{4.48}$$

$$= f(n2\pi/\Delta) e^{-int_0 2\pi/\Delta} \cdot \lim_{\delta \to 0} \lim_{\varepsilon_1 \to 0} \left[\frac{1}{i\Delta} \cdot \ln\left[\frac{\varepsilon_1 + i\delta\Delta}{\varepsilon_1 - i\delta\Delta} \right] \right]$$

$$= \frac{1}{i\Delta} f(n2\pi/\Delta) e^{-int_0 2\pi/\Delta} \ln(-1) = \frac{\pi}{\Delta} f(n2\pi/\Delta) e^{-int_0 2\pi/\Delta}$$

where we have used that the complex logarithm of -1 is $i\pi$. By going through exactly the same steps as above, we find that the second integral in (4.46) also gives $\Delta^{-1}\pi f(n2\pi/\Delta) e^{-int_0 2\pi/\Delta}$. Equation (4.46) can therefore be written as

$$\int_{-\infty}^{\infty} \tilde{S}(k) f(k)\, dk = \frac{\sqrt{2\pi}}{\Delta} \sum_{n=-\infty}^{\infty} f(n2\pi/\Delta) e^{-int_0 2\pi/\Delta} \tag{4.49}$$

which is exactly (4.39b) and we have established both (4.39a) and (4.39b) and hence proofed (4.38).

We are now ready to prove the sampling theorem stating that if the Fourier transform $\tilde{g}(k)$ of g

$$\tilde{g}(k) = 0 \quad \text{for} \quad |k| > \frac{\pi}{\Delta} \tag{4.5}$$

for some $\Delta > 0$, then the function g can be written as :

$$g(t) = \sum_{n=-\infty}^{\infty} \frac{\sin[((t - t_0)/\Delta - n)\pi]}{((t - t_0)/\Delta - n)\pi} g(t_0 + n\Delta) \tag{4.6}$$

To proof this we multiply g by the sampling function S and consider the function $I(t) = S(t,t_0,\Delta)g(t)$. Taking the Fourier transform and using the definition (4.37) of the sampling function we get

$$\begin{aligned} \tilde{I}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(t - t_0 - n\Delta) g(t) e^{-ikt}\, dt \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} e^{-i(t_0 + n\Delta)k} g(t_0 + n\Delta) \ . \end{aligned} \tag{4.50}$$

But by expressing S and g in terms of their Fourier integrals, $\tilde{I}(k)$ can also be expressed as

$$\begin{aligned} \tilde{I}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S(t) g(t) e^{-ikt}\, dt = \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{S}(\omega) \tilde{g}(s) e^{-i(k - \omega - s)t}\, dt d\omega ds \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{S}(\omega) \tilde{g}(s) \delta(\omega - (k - s))\, d\omega ds = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{S}(k - s) \tilde{g}(s)\, ds \ . \end{aligned} \tag{4.51}$$

Inserting now (4.38) for $\tilde{S}$ this can be written as

$$\begin{aligned} \tilde{I}(k) &= \frac{1}{\Delta} \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} e^{-i(k - s)t_0} \delta(k - n2\pi/\Delta - s) \tilde{g}(s)\, ds \\ &= \frac{1}{\Delta} \sum_{n=-\infty}^{\infty} e^{int_0 2\pi/\Delta} \tilde{g}(k - n2\pi/\Delta) \ . \end{aligned} \tag{4.52}$$

Equation (4.52) shows that $\tilde{I}(k)$ is, apart from a phase factor, obtained by duplicating $\tilde{g}(k)$ at positions shifted by $n2\pi/\Delta$ and we see that if $\tilde{g}(k) = 0$ for $|k| > \pi/\Delta$ then the duplicated spectra of $\tilde{g}(k)$ do not overlap and we have

$$\tilde{I}(k) = \frac{1}{\Delta} \tilde{g}(k) \quad \text{for} \quad |k| \leq \frac{\pi}{\Delta} \ . \tag{4.53}$$

But because $\tilde{g}(k) = 0$ for $|k| > \pi/\Delta$ we can of course write

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{g}(k) e^{ikt}\, dk = \frac{1}{\sqrt{2\pi}} \int_{-\pi/\Delta}^{\pi/\Delta} \tilde{g}(k) e^{ikt}\, dk \tag{4.54}$$

and since (4.53) is valid we have

$$g(t) = \frac{\Delta}{\sqrt{2\pi}} \int_{-\pi/\Delta}^{\pi/\Delta} \tilde{I}(k) e^{ikt}\, dk \ . \tag{4.55}$$

We do now finally insert (4.50) for $\tilde{I}(k)$ in (4.55) and perform the integration and get

$$\begin{aligned} g(t) &= \sum_{n=-\infty}^{\infty} \frac{\Delta}{2\pi} \int_{-\pi/\Delta}^{\pi/\Delta} e^{-i(t_0 - t + n\Delta)k}\, dk\, g(t_0 + n\Delta) \\ &= \sum_{n=-\infty}^{\infty} \frac{\sin[((t - t_0)/\Delta - n)\pi]}{((t - t_0)/\Delta - n)\pi} g(t_0 + n\Delta) \end{aligned} \tag{4.56}$$

which is exactly (4.6) and we have thus proofed the sampling theorem.

The crucial point in the proof is the bandlimitness of the kernel function. If $\tilde{g}(k) \neq 0$ for $|k| > \pi/\Delta$ then (4.53) does not hold true and the proof brakes down. Let us investigate what happens in this case. It is of help to display graphically the steps in the rather technical proof above in order to understand what is going on. But before we do so, let us develop the formalism a little bit further. First of all we recall that in (4.51) we saw that the Fourier transform of the product of two functions $f(t)g(t)$ is given as the convolution integral of their Fourier transforms

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) g(t) e^{-ikt}\, dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{f}(k - s) \tilde{g}(s)\, ds \ . \tag{4.57}$$

In exactly the same way as we deduced this identity we find that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{f}(k) \tilde{g}(k) e^{ikt}\, dk = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t - s) g(s)\, ds \ . \tag{4.58}$$

Let us now introduce the *sinc function* defined as

$$\mathrm{sinc}_a(x) = \mathrm{sinc}(ax) = \frac{1}{2a} \int_{-a}^{a} e^{ikx}\, dk = \frac{\sin(ax)}{ax} \ . \tag{4.59}$$

From this we see that the Fourier transform of the sinc function is given as

$$\tilde{\text{sinc}}_a(k) = \begin{cases} a^{-1}\sqrt{\pi/2} & \text{for } |k| < a \\ 0 & \text{for } |k| > a \end{cases} \tag{4.60}$$

By using the definition of the sinc function and the sampling function we can write (4.56) as

$$\begin{aligned} g_1(t) &= \sum_{n=-\infty}^{\infty} \text{sinc}[((t - t_0)/\Delta - n)\pi] g(t_0 + n\Delta) \\ &= \int_{-\infty}^{\infty} \text{sinc}[\frac{\pi}{\Delta}(t - s)] \sum_{n=-\infty}^{\infty} \delta(s - t_0 - n\Delta) g(s)\, ds \\ &= \int_{-\infty}^{\infty} \text{sinc}[\frac{\pi}{\Delta}(t - s)]\, S(s, t_0, \Delta) g(s)\, ds \quad . \end{aligned} \tag{4.61}$$

Similarly, by using (4.51) and (4.60), we can write (4.55) as

$$g_1(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[\tilde{\text{sinc}}_{\pi/\Delta}(k) \int_{-\infty}^{\infty} \tilde{S}(k - s) \tilde{g}(s)\, ds \right] e^{ikt}\, dk \quad . \tag{4.62}$$

We do here distinguish between g and g_1 because they are in general not identical. It is only in the case when $\tilde{g}(k) = 0$ for $|k| > \pi/\Delta$ and (4.53) holds true that we have $g_1 = g$.

Let us now introduce a short hand notation for the forward and inverse Fourier transforms and the convolution and write

$$F[f] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-ikt}\, dt = \tilde{f} \tag{4.63a}$$

$$F^{-1}[\tilde{f}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{f}(k) e^{ikt}\, dk = f \tag{4.63b}$$

$$f * g = \int_{-\infty}^{\infty} f(x - y) g(y)\, dy \quad . \tag{4.63c}$$

In this short hand notation (4.57) and (4.58) read

$$F[f \cdot g] = \frac{1}{\sqrt{2\pi}} \tilde{f} * \tilde{g} \tag{4.57a}$$

and

$$F^{-1}[\tilde{f}\cdot\tilde{g}] = \frac{1}{\sqrt{2\pi}} f*g \ . \tag{4.58a}$$

From this we immediately see that

$$F[f*g] = \sqrt{2\pi}\, \tilde{f}\cdot\tilde{g} \tag{4.64a}$$

and

$$F^{-1}[\tilde{f}*\tilde{g}] = \sqrt{2\pi}\, f\cdot g \ . \tag{4.64b}$$

In terms of this short hand notation (4.61) can be written as

$$g_1 = \mathrm{sinc}_{\pi/\Delta}*(S\cdot g) \tag{4.61a}$$

and (4.62) as

$$g_1 = F^{-1}\left[\,\tilde{\mathrm{sinc}}_{\pi/\Delta}\cdot(\tilde{S}*\tilde{g})\,\right] \ . \tag{4.62a}$$

It is a trivial exercise in using (4.57a), (4.58a), (4.64a) and (4.64b) to show that the two expressions for g_1, (4.61a) and (4.62a), are equivalent.

We are now prepared to display graphically the proof of the sampling theorem and see what goes wrong if the sampled function fails to be band limited or if the sample spacing Δ is to large. Parts a and b, c and d, e and f on Figure 4.3 display a function and its Fourier transform for the function g to be sampled, the sampling function S and the sinc function $\mathrm{sinc}_{\pi/\Delta}$. Part g) on Figure 4.3 shows the product of the sampling function S and the function g and part h the convolution of their Fourier transforms. Finally part i shows the sinc function convolved with $S\cdot g$ and part j shows $\tilde{S}*\tilde{g}$ multiplied by the Fourier transform of the sinc function. It is evident from Figure 4.3 that if $\tilde{g}(k) = 0$ for $|k| > \pi/\Delta$, then the repeated spectra on 4.3 h) do not overlap and $\tilde{g}$ is reproduced exactly in 4.3 j. An inverse Fourier transform will therefore regenerate g and g_1 in (4.61a) and (4.62a) will therefore be equal to g. If on the other hand $\tilde{g}(k) \neq 0$ for $|k| > \pi/\Delta$ as shown by the pointed curve on figure 4.3 b, then the repeated spectra in 4.3 h do overlap resulting in the spectrum shown by the broken lines. Multiplication by the Fourier transform of the sinc function will not reproduce $\tilde{g}$ in this case. The tails of $\tilde{g}$, laying outside the interval from $-\pi/\Delta$ to π/Δ are wrapped around into the interval. The tail laying above π/Δ appears above $-\pi/\Delta$ and the tail below $-\pi/\Delta$ appears below π/Δ. This is called **aliasing** and it obviously has the effect of replacing the high frequency components, with $|k| > \pi/\Delta$, of the Fourier expansion of g by lower frequencies. This is demonstrated on Figure 4.4. The Fourier transform of the function $g(t) = \cos[t3\pi/(2\Delta)]$ is two delta "function" spikes at $\pm 3\pi/(2\Delta)$. When this function is sampled with the sampling function $s(t, 0, \Delta)$, the delta "function" spikes are aliased to $\pm\pi/(4\Delta)$ and the resulting function is $g(t) = \cos[t\pi/(4\Delta)]$ of much lower frequency. The reason for the aliasing is of course that we do not sample g densely enough to be able to reconstruct its high frequency components and we say that the function is under sampled. To

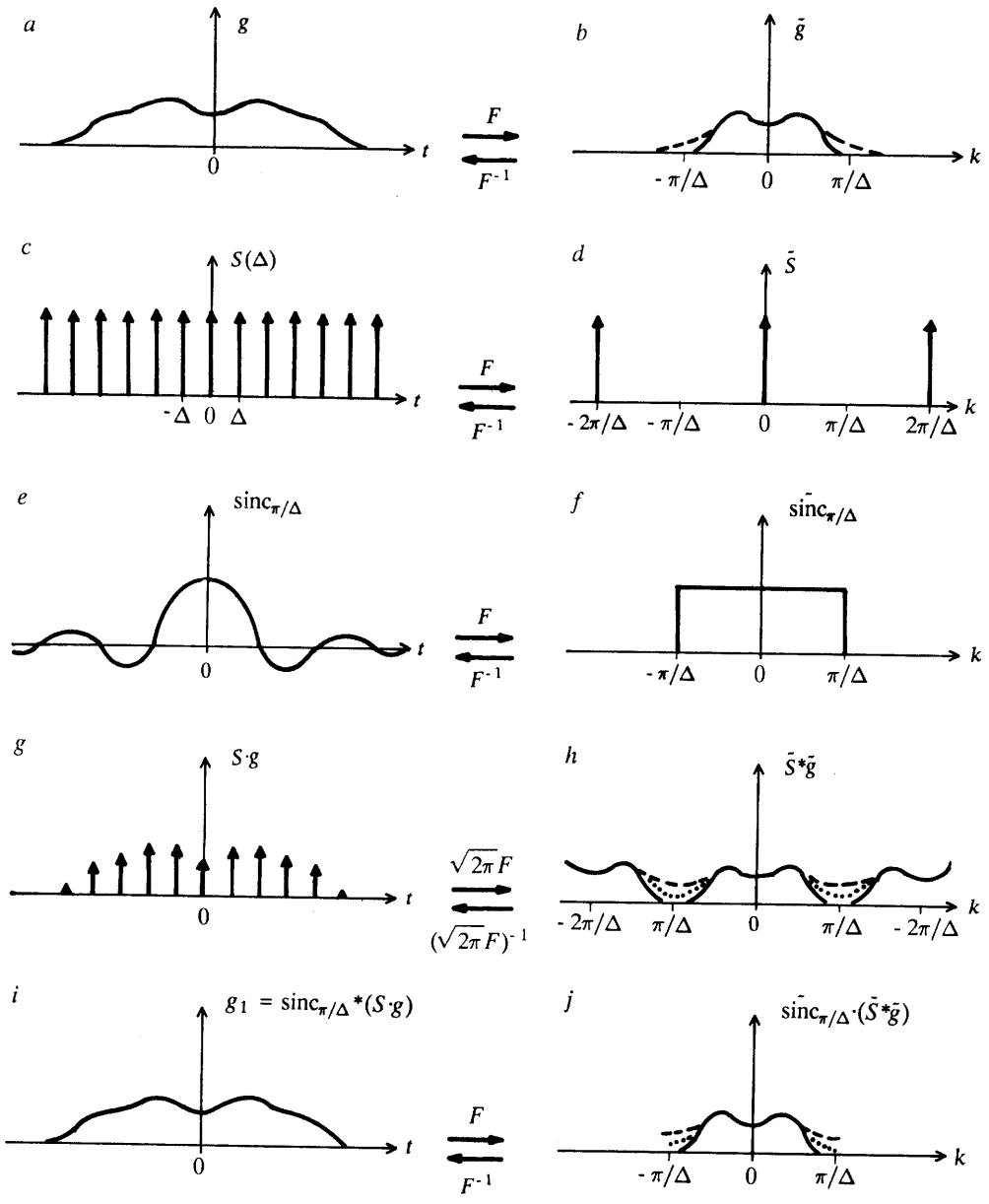

Figure 4.3 Graphical display of the steps in the proof of the sampling theorem

each sample spacing Δ there corresponds therefore a cut-off angular frequency, called the **Nyquist frequency** $k_N = \pi/\Delta$, such that all frequencies $|k| > k_N$ are aliased.

If the function g is band limited, then aliasing can be avoided by decreasing the sample spacing Δ until $\tilde{g}(k) = 0$ for $|k| > \pi/\Delta$. Decreasing the sample spacing increases the distance between the spikes of $\tilde{S}$ on Figure 4.3 d) and also the width of the box shaped

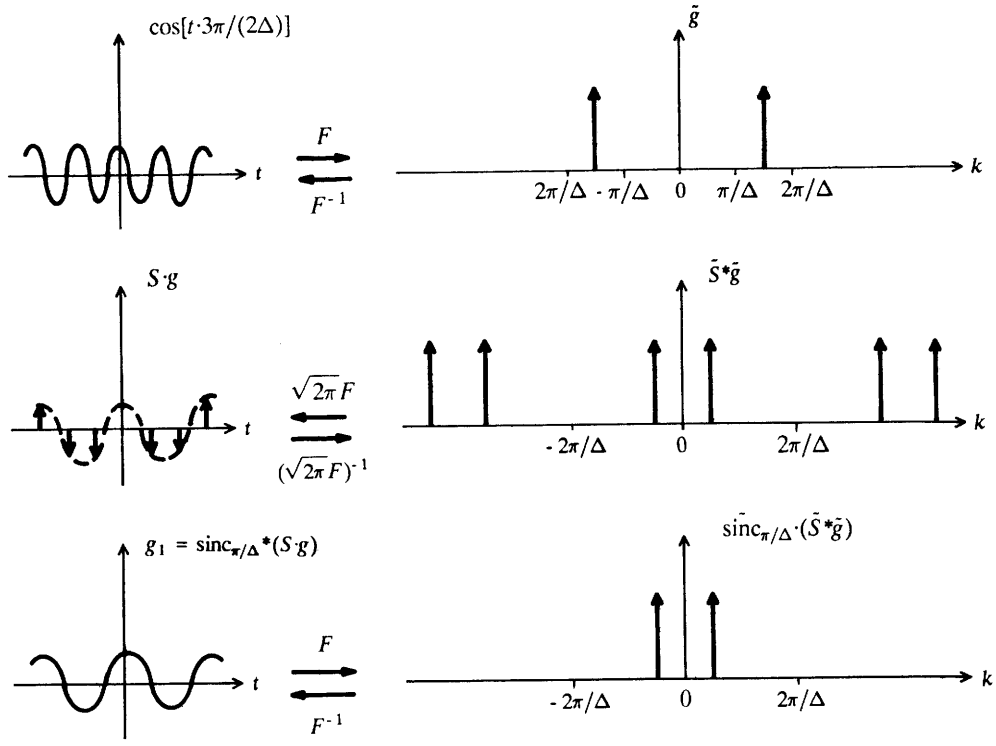

Figure 4.4 Aliasing due to unsufficient sampling

Fourier transform of the sinc function on Figure 4.3 f). This results in a pulling apart of the repeated spectra of g on Figure 4.3 h) and eventually they cease to overlap and g is no longer aliased by the sampling. If g is not band limited then sampling with any sample spacing will cause aliasing. But in the case when $|\tilde{g}(k)| \rightarrow 0$ for $|k| \rightarrow \infty$, the aliasing can be made small, provided that Δ is taken small enough and the effect of the aliasing will be to smooth high frequency oscillations and round off sharp edges in g. These considerations make clear the importance of the fact that using logarithmic instead of linear variables makes the Fourier transform of non-band limited function tend much faster to zero for large frequency values as discussed in section 4.1 .

When discussing Fourier transforms and Fourier expansions of functions in section 2.3 we interpreted the equation

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{g}(k) e^{ikt}\, dk \tag{4.65}$$

as an expansion of g in terms of an infinite set of "orthonormal" basis functions $\beta_k(t) = (2\pi)^{-1/2} e^{ikt}$ where the expansion coefficients $\tilde{g}(k)$ are the projections of g onto the

basis functions (see equations (2.38) - (2.42)). If g is band limited with $\tilde{g}(k) = 0$ for $|k| > \pi/\Delta$ then it can be expressed as a linear combination of a small, but still infinite, subset of the basis functions β_k. According to the sampling theorem g can also be written as

$$g(t) = \sum_{n=-\infty}^{\infty} \frac{\sin[((t - t_0)/\Delta - n)\pi]}{((t - t_0)/\Delta - n)\pi} g(t_0 + n\Delta) \quad . \tag{4.6}$$

This can also be interpreted as an expansion of g in terms of an infinite but discrete set of orthonormal basis functions. If we define

$$\alpha_n(t) = \frac{\sin[((t - t_0)/\Delta - n)\pi]}{((t - t_0)/\Delta - n)\pi} = \frac{\Delta}{2\pi} \int_{-\pi/\Delta}^{\pi/\Delta} e^{-i(t - t_0 - n\Delta)k} \, dk \tag{4.66}$$

then it is a simple exercise to show that the inner product

$$<\alpha_m \mid \alpha_n> = \int_{-\infty}^{\infty} \alpha_m(t)\alpha_n(t) \, dt = \Delta \, \delta_{m,n} \tag{4.67}$$

where $\delta_{m,n}$ is the Kronecker delta, defined as $\delta_{m,n} = 0$ if $m \neq n$ and $\delta_{n,n} = 1$. It is also straight forward to see that the expansion coefficients $g(t_0 + n\Delta)$ are the projections of g onto the basis functions because

$$\frac{1}{\Delta}<\alpha_n \mid g> = \frac{1}{\Delta} \int_{-\infty}^{\infty} \frac{\sin[((t - t_0)/\Delta - n)\pi]}{((t - t_0)/\Delta - n)\pi} g(t) \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\pi/\Delta}^{\pi/\Delta} g(t) e^{-i(t - t_0 - n\Delta)k} \, dk dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\pi/\Delta}^{\pi/\Delta} \tilde{g}(k) e^{i(t_0 + n\Delta)k} \, dk \quad . \tag{4.68}$$

But since $\tilde{g}(k) = 0$ for $|k| > \pi/\Delta$ we see that

$$\frac{1}{\Delta}<\alpha_n \mid g> = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{g}(k) e^{i(t_0 + n\Delta)k} \, dk = g(t_0 + n\Delta) \tag{4.69}$$

and the sampling theorem states therefore that

$$g(t) = \sum_{n=-\infty}^{\infty} \alpha_n(t) \frac{<\alpha_n \mid g>}{\Delta} = \sum_{n=-\infty}^{\infty} \alpha_n(t) g(t_0 + n\Delta) \quad . \tag{4.70}$$

Mathematically the sampling theorem states therefore that the two sets of basis functions $B_1 = \{ \beta_k(t) \mid |k| \leq \pi/\Delta \}$ and $B_2 = \{ \alpha_n(t) \mid n \in Z \}$ span the same subset of complex valued functions. This is a remarkable statement because B_1 is not a denumerable set but B_2 is and is therefore a much "smaller" set.

4.3 METHODS FOR DETERMINING THE FILTER COEFFICIENTS

We have in the last two sections discussed the theory of digital filters and their use in numerical integration. But we still need to discuss one important point, namely how we actually determine the filter coefficients. Let us now turn to that problem. We will discuss two different methods. The first one is based on the analytical definition of the filter coefficients in (4.9) and uses the theory of Fourier transforms to evaluate them. The second method is based on operational properties of the coefficients as expressed in equation (4.14) and they are determined by a least-squares method.

Let us begin by the Fourier transform method and restate the analytical definition of the filter coefficients

$$b_n(\eta_0,\tau_0) = \int_{-\infty}^{\infty} \frac{\sin[((\tau - \tau_0)/\Delta - n)\pi]}{((\tau - \tau_0)/\Delta - n)\pi} H(\eta_0 + \tau)\, d\tau \quad . \tag{4.9}$$

By the end of the last section we saw in (4.68) that this can be written as

$$b_n(\eta_0,\tau_0) = \frac{\Delta}{\sqrt{2\pi}} \int_{-\pi/\Delta}^{\pi/\Delta} \tilde{H}(\kappa) e^{i(\eta_0 + \tau_0 + n\Delta)\kappa}\, d\kappa \tag{4.71}$$

where $\tilde{H}(\kappa)$ is the Fourier transform of the filter function $H(\tau)$. If $\tilde{H}(\kappa)$ is known, then the filter coefficients can be determined by performing the integration in (4.71), either analytically or, as is usually needed, numerically by standard numerical integration methods. The Fourier transform of the filter function can be determined by calculating the integral

$$\tilde{H}(\kappa) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H(\tau) e^{-i\kappa\tau}\, d\tau \tag{4.72}$$

either analytically or numerically. Sometimes it is though easier to use a different approach. Suppose that we know two functions F and G satisfying

$$F(\eta) = \int_{-\infty}^{\infty} G(\tau)\, H(\eta + \tau)\, d\tau \tag{4.4}$$

and that the Fourier transforms $\tilde{F}$ and $\tilde{G}$ are easily determined. By changing the integration variable in (4.4) to $\tau' = -\tau$ and comparing the result with (4.63c) and (4.64a) we immediately see that

$$\tilde{H}(\kappa) = \frac{1}{\sqrt{2\pi}} \frac{\tilde{F}(\kappa)}{\tilde{G}(-\kappa)} \quad . \tag{4.73}$$

The abscissa values η_0 and τ_0 in (4.71) are arbitrary. As pointed out by Koefoed (Koefoed

1972), the filter coefficients can be made to tail more rapidly off by a special choice of $\eta_0 + \tau_0$. To see this we make a partial integration of (4.71) and write

$$b_n(\eta_0,\tau_0) = \frac{\Delta}{\sqrt{2\pi}} \frac{\tilde{H}(\pi/\Delta)e^{i(\eta_0 + \tau_0)/\Delta + n)\pi} - \tilde{H}(-\pi/\Delta)e^{-i((\eta_0 + \tau_0)/\Delta + n)\pi}}{i(\eta_0 + \tau_0 + n\Delta)}$$

$$+ \frac{-1}{\sqrt{2\pi}} \frac{1}{i((\eta_0 + \tau_0)/\Delta + n)} \int_{-\pi/\Delta}^{\pi/\Delta} \frac{d\tilde{H}(\kappa)}{d\kappa} e^{i(\eta_0 + \tau_0 + n\Delta)\kappa} d\kappa \ . \qquad (4.74)$$

We will now show that the first term in (4.74) can be made to vanish by making a proper choice of $\eta_0 + \tau_0$. Since $H(\tau)$ is a real valued function we see from (4.72) that

$$\tilde{H}^*(\kappa) = \tilde{H}(-\kappa) \qquad (4.75)$$

where $\tilde{H}^*(\kappa)$ is the complex conjugate of $\tilde{H}(\kappa)$. We do now express $\tilde{H}$ in terms of its modulus A and argument Φ and write

$$\tilde{H}(\kappa) = A(\kappa)e^{i\Phi(\kappa)} \qquad (4.76)$$

and from (4.75) we see that

$$A(-\kappa) = A(\kappa) \qquad \text{and} \qquad \Phi(-\kappa) = -\Phi(\kappa) \ . \qquad (4.77)$$

Inserting (4.76) and using (4.77) we see that the first term in (4.74) can be written as

$$\frac{\Delta}{\sqrt{2\pi}} \frac{A(\kappa)}{(\eta_0 + \tau_0 + n\Delta)} 2\sin[\Phi(\pi/\Delta) + ((\eta_0 + \tau_0)/\Delta + n)\pi] \ . \qquad (4.78)$$

The sine in (4.78) will be equal to zero if we take

$$\Phi(\pi/\Delta) + \pi(\eta_0 + \tau_0)/\Delta = m\pi \qquad \text{or} \qquad \eta_0 + \tau_0 = m\Delta - \frac{\Delta}{\pi}\Phi(\pi/\Delta) \ . \qquad (4.79)$$

The filter coefficients are now given as

$$b_n(\eta_0,\tau_0) = \frac{1}{\sqrt{2\pi}} \frac{i}{n + m - \pi^{-1}\Phi(\pi/\Delta)} \int_{-\pi/\Delta}^{\pi/\Delta} \frac{d\tilde{H}(\kappa)}{d\kappa} e^{i\Delta(n + m - \pi^{-1}\Phi(\pi/\Delta))\kappa} d\kappa \qquad (4.80)$$

which has the same form as (4.71) except that the integrand now contains the derivative of $\tilde{H}$ and the integral is multiplied by a factor that decreases as $1/n$ for large positive and negative values of n. We see therefore that by choosing $\eta_0 + \tau_0$ according to (4.79), we make the filter coefficients tail faster off to zero for large values of $|n|$. The coefficients can now be determined by performing the integration in (4.80) or (4.71).

The filter functions that we face in (4.4) in the case of transient electro-magnetic soundings over layered earth are

$$H_J(\tau) = J_1(e^\tau)e^\tau \quad \text{and} \quad H_C(\tau) = \cos(e^\tau)e^\tau \ . \tag{4.81}$$

The Fourier transforms of these functions are

$$\begin{aligned}\tilde{H}_J(\kappa) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} J_1(e^\tau)e^\tau e^{-i\kappa\tau} d\tau \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} J_1(y)y^{-i\kappa} dy = \frac{2^{-i\kappa}}{\sqrt{2\pi}} \frac{\Gamma(1 - i\kappa/2)}{\Gamma(1 + i\kappa/2)}\end{aligned} \tag{4.82}$$

(see Abramowitz and Stegun 1970, p. 486) and

$$\tilde{H}_C(\kappa) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(e^\tau)e^\tau e^{-i\kappa\tau} d\tau = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(y)y^{-i\kappa} dy \tag{4.83}$$

which by use of (3.47) is given as

$$\tilde{H}_C(\kappa) = \frac{1}{\sqrt{2\pi}} \frac{1}{2} \Gamma(1 - i\kappa) [(i)^{1-i\kappa} + (-i)^{1-i\kappa}] = \frac{i}{\sqrt{2\pi}} \Gamma(1 - i\kappa) \sinh(\pi\kappa/2) \ . \tag{4.83a}$$

To write $\tilde{H}_J(\kappa)$ and $\tilde{H}_C(\kappa)$ in terms of modulus and argument on the form (4.76) we note (Abramowitz and Stegun 1970, p. 256) that

$$\arg[\Gamma(1 + ix)] = -\gamma x + \sum_{n=1}^{\infty} [x/n - \arctan(x/n)] \tag{4.84a}$$

where $\gamma = 0.57722$ is the Euler's constant and that

$$|\Gamma(1 + ix)| = |\Gamma(1 - ix)| = \left[\frac{\pi x}{\sinh(\pi x)}\right]^{1/2} \ . \tag{4.84b}$$

We see therefore that we can write

$$\tilde{H}_J(\kappa) = A_J(\kappa)e^{i\Phi_J(\kappa)} \tag{4.85a}$$

where

$$A_J(\kappa) = \frac{1}{\sqrt{2\pi}} \ ; \ \Phi_J(\kappa) = -\kappa\ln(2) + \gamma\kappa - \sum_{n=1}^{\infty} [\kappa/n - 2\arctan(\kappa/(2n))] \tag{4.85b}$$

And similarly by noting that

$$i\sinh(\pi\kappa/2) = e^{i\frac{\pi}{2}\frac{\kappa}{|\kappa|}} |\sinh(\pi\kappa/2)| \tag{4.86}$$

and using (4.84a) and (4.84b) we see that

$$\tilde{H}_C(\kappa) = A_C(\kappa)e^{i\Phi_C(\kappa)} \tag{4.87a}$$

where

$$A_C(\kappa) = \left[\frac{\kappa\sinh^2(\frac{\pi}{2}\kappa)}{2\sinh(\pi\kappa)}\right]^{1/2}$$

$$\Phi_C(\kappa) = \frac{\pi}{2}\frac{\kappa}{|\kappa|} + \gamma\kappa - \sum_{n=1}^{\infty}[\kappa/n - \arctan(\kappa/n)] \ . \tag{4.87b}$$

Armed with the analytical expressions (4.85b) and (4.87b) for the modulus and argument for $\tilde{H}_J$ and $\tilde{H}_C$, it is a straight forward matter to determine the optimal shift $\eta_0 + r_0$ according to (4.79) and work out the filter coefficients for the cosine and Hankel transforms by performing the integration in (4.71) or (4.80).

It is often a tedious and time consuming work to determine the filter coefficients by the Fourier transform method and care must be taken in the numerical integration in (4.71) or (4.80) in order to get the coefficients accurate enough so that an acceptable accuracy is ensured in sums (4.13) and (4.14). As was mentioned at the beginning of this section, there exists another method for determining the filter coefficients which is usually much easier and results in more accurate coefficients. This method starts from the desired operational properties of the filter coefficients as expressed in equation (4.14). We want to determine filter coefficients to evaluate numerically the integral

$$f(x) = x\int_0^{\infty} g(y)h(xy)\,dy \tag{4.1}$$

for a given filter function h and for different kernel functions g, by the sum

$$F(\eta_0 + m\Delta) = \sum_{n=-\infty}^{\infty} G(r_0 + (n-m)\Delta)\, b_n(\eta_0, r_0) \tag{4.14}$$

where

$$F(r) = f(e^r) \ ; \ G(r) = g(e^r) \ . \tag{4.14a}$$

The filter coefficients are, as we emphasised earlier, independent of the kernel function and do only depend on the filter function h, the shift $\eta_0 + r_0$ and the sample spacing Δ which has been adjusted so that the sampling causes no or negligible aliasing of the kernel function. Let f_0 and g_0 be a pair of functions that satisfy (4.1) for the filter function at hand. Such a pair of functions are usually easily found and it is preferable to choose f_0 and g_0 such that $F_0(r)$ and $G_0(r)$ tend to zero or at least to some well defined limiting values as $r \to \pm\infty$. The idea is now to use (4.14) to determine b_n such that $F_0(r)$ is approximated as well as

possible by the sum in (4.14) but truncated at finite positive and negative values of n. But first we want to determine an optimal shift $\eta_0 + \tau_0$ as described by (4.79) in order to make the filter tail more rapidly off. To that end we determine the Fourier transform of $H(\tau) = h(e^{\tau})e^{\tau}$. This we do either directly using (4.72) or indirectly by using (4.73) since we know f_0 and g_0 satisfying (4.1).

The number, N, of filter coefficients that we need depends primarily on the filter function H and the desired accuracy, the higher the accuracy the higher N must be. We take N so large that the filter coefficients have tailed practically to zero (within the desired accuracy) at both ends. To be able to determine the N filter coefficients we need to evaluate $F_0(\eta_0 + m\Delta)$ for M values of m where $M \geq N$ and from (4.14) we see that we have to evaluate $G(\tau_0 + k\Delta)$ for $N + M - 1$ values of k. Normally we take M considerably larger than N. We can always select η_0 and τ_0 in such a way that the function values $F_0(\eta_0 + m\Delta)$ are numbered from 1 to M and the filter coefficients from 1 to N and still having $\eta_0 + \tau_0$ an optimal shift according to (4.79). We will further take N and M large enough so that $F_0(\eta_0 + \Delta)$, $F_0(\eta_0 + M\Delta)$, $G_0(\tau_0 + (1 - M)\Delta)$ and $G_0(\tau_0 + (N - 1)\Delta)$ are close to the limiting values, preferably zero, of F_0 and G_0 for large positive and negative arguments. Let us now write

$$f_m = F_0(\eta_0 + m\Delta) \quad ; \quad m = 1,....,M \tag{4.88}$$

$$G_{m,n} = G_0(\tau_0 + (n - m)\Delta) \quad ; \quad m = 1,....,M \quad ; \quad n = 1,....,N \ .$$

We do now want to determine the N filter coefficients such that f_m is approximated as well as possible, in the least-squares sense, by the truncated sum in (4.14), that is we want to determine b_n such that we minimize

$$\chi^2 = \sum_{m=1}^{M} [f_m - \sum_{n=1}^{N} G_{m,n} b_n]^2 \ . \tag{4.89}$$

The quantity χ^2 in (4.89), which is a function of the filter coefficients, has obviously no maximum value but it must have a minimum value ≥ 0. To find the b_n's that minimize χ^2, we demand that the partial derivatives with respect to the filter coefficients vanish, that is

$$\frac{\partial \chi^2}{\partial b_k} = 2 \sum_{m=1}^{M} [f_m - \sum_{n=1}^{N} G_{m,n} b_n] \, G_{m,k} = 0 \quad ; \quad k = 1,....,N \ . \tag{4.90}$$

These are N equations that can be solved for the N filter coefficients. It is convenient to develop (4.90) a little further and introduce vector notation. Let us represent the function values f_n as an M dimensional vector $\boldsymbol{f}$, the filter coefficients as an N dimensional vector $\boldsymbol{b}$ and $G_{m,n}$ as an MxN dimensional matrix $\boldsymbol{G}$. Equation (4.90) can now be written as

$$\boldsymbol{G}^T \cdot \boldsymbol{f} - \boldsymbol{G}^T \cdot \boldsymbol{G} \cdot \boldsymbol{b} = 0 \tag{4.91}$$

where $\boldsymbol{G}^T$ is the transpose of $\boldsymbol{G}$. This can also be written as

$$\boldsymbol{c} = \boldsymbol{A}\cdot\boldsymbol{b} \quad \text{where} \quad \boldsymbol{c} = \boldsymbol{G}^T\cdot\boldsymbol{f} \;\; ; \;\; \boldsymbol{A} = \boldsymbol{G}^T\cdot\boldsymbol{G} \;\; . \tag{4.92}$$

The vector $\boldsymbol{c}$ and the symmetric matrix $\boldsymbol{A}$ are known and equation (4.92) can be solved for $\boldsymbol{b}$ by standard methods. The only problem that we might run into is that $\boldsymbol{A}$ might turn out to be singular and hence not invertible. We will in the next section, when discussing the non-linear least-squares inversion technique, learn how to deal that problem and how to solve (4.92).

We conclude this section by giving some pairs of functions, f and g, that satisfy (4.1) for the filter functions $J_1(y)$ and $\cos(y)$ that we encounter in the theory of transient electromagnetic soundings over layered earth. Let us first consider $h(y) = J_1(y)$. If we take $k_0 = 0$, $z_0 = 0$ and $z \geq 0$ in (2.56) and differentiate with respect to r and finally multiply by r we get

$$\frac{r^2}{(r^2 + z^2)^{3/2}} = r\int_0^\infty \lambda e^{-\lambda z} J_1(\lambda r)\, d\lambda \;\; . \tag{4.93a}$$

We see therefore that the functions

$$f(x) = \frac{x^2}{(x^2 + z^2)^{3/2}} \;\; ; \;\; g(y) = ye^{-yz} \tag{4.94a}$$

satisfy (4.1) for $h = J_1$. We have also (see Abramowitz and Stegun 1970, p. 486) that

$$\frac{x^2}{4} e^{-\frac{x^2}{4}} = x\int_0^\infty y^2 e^{-y^2} J_1(xy)\, dy \tag{4.93b}$$

so that

$$f(x) = \frac{x^2}{4} e^{-\frac{x^2}{4}} \;\; ; \;\; f(y) = y^2 e^{-y^2} \tag{4.94b}$$

also satisfy (4.1) for $h = J_1$. In the case when $h(y) = \cos(y)$ we have for example (see S. M. Selby 1974, pp. 455-456)

$$\frac{x}{1 + x^2} = x\int_0^\infty e^{-y}\cos(xy)\, dy \tag{4.95a}$$

and

$$\frac{x(1+x^2)}{(1+x^2)^2} = x\int_0^\infty ye^{-y}\cos(xy)\,dy \tag{4.95b}$$

and furthermore

$$\frac{\sqrt{\pi}}{2}xe^{-\frac{x^2}{4}} = x\int_0^\infty e^{-y^2}\cos(xy)\,dy \quad . \tag{4.95c}$$

We see therefore that the pairs of functions

$$f(x) = \frac{x}{1+x^2} \quad ; \quad g(y) = e^{-y} \tag{4.96a}$$

$$f(x) = \frac{x(1+x^2)}{(1+x^2)^2} \quad ; \quad g(y) = ye^{-y} \tag{4.96b}$$

$$f(x) = \frac{\sqrt{\pi}}{2}xe^{-\frac{x^2}{4}} \quad ; \quad g(y) = e^{-y^2} \tag{4.96c}$$

satisfy (4.1) with h = cos. As we have seen, we can usually find many pairs of functions satisfying (4.1). When the filter coefficients are determined by the least-squares method, it is desirable to choose a function pair such that g has similar behaviour as the kernel functions to be integrated by the filter.

5. INVERSION THEORY

Until now we have been concerned with the problem of determining the transient electro-magnetic response of a layered earth. We have assumed that we know the resistivity structure of the layered earth, that is the number of layers and their conductivities and thicknesses. We have learned to calculate the transient response for the central loop and the grounded dipole loop configurations, both in terms of the induced transient voltage and in terms of late time apparent resistivity. We have thus solved the **forward problem** of finding the response of a given model.

In practical use of transient electro-magnetic soundings we are, like in experimental physics in general, faced with the more difficult **inverse problem**. We are measuring the response of a physical system due to known excitations and we want to determine the relevant physical parameters characterizing the system from the measured response. In transient electro-magnetic soundings we are exciting the earth using a transmitter and we want to determine the resistivity structure of the subsurface from the measured response in a receiver. In order to be able to solve the inverse problem we must obviously know, or at least have a candidate for, how the measured response variables depend on the parameters to be determined, that is to say we must know how to solve the forward problem.

At the end of section 3.2 we anticipated a possible approach to the inverse problem, namely to calculate the theoretical response of the system for different values of the model parameters and compare the results with the measured response and choose the "best" model by visual inspection. Even though this trial and error approach is tedious and results often only in a crude estimates of the relevant model parameters, it was, until the beginning of the seventies, the most generally used inversion method for geophysical data. Since then there has been, mainly because electronic computers have become generally available, a great progress in the development of inversion methods resulting in automatic and much more accurate procedures. The most widely used inversion method, at least for inversion of geoelectric soundings, is the **least-squares inversion method**. In the case when the response depends non-linearly on the model parameters this method is referred to as non-linear least-squares inversion.

The task of this chapter is to study the non-linear least-squares method and see how it can be used to solve the inverse problem of determining the resistivity structure of a layered earth from the measured transient electro-magnetic responses at the surface. But before doing so, let us state the problem precisely and make clear what our goal is and which assumptions we have to make. We have measured the late time apparent resistivity, ρ_a, defined by (3.56) or (3.57) at the time values t_i , $i = 1,....,M$ after the constant current in the transmitter is turned off. From the measured apparent resistivities we want to gain information about the resistivity structure of the earth beneath the sounding site. Since we only know how to solve the forward problem for layered earth, we make the fundamental assumption that the resistivity structure

beneath the sounding site is horizontally stratified. This assumption does in most cases not hold strictly true and may therefore seem to be a rather bold one, but it is usually a fair first approximation and the more general case falls out side the scope of the present discussion. We do therefore want to determine the layered earth model whose response reproduces the measured apparent resistivity values "as well as possible". In the next section we will state more precisely what we mean by "as well as possible".

Since the apparent resistivity values are obtained by measuring the induction of a time varying magnetic flux through a receiver coil, they will contain a random noise due to the variations in the Earth's magnetic field. Deviations of the measured apparent resistivity values, due to non-horizontal resistivity structures, from values obtainable from horizontal layers will in our case also be considered as a "noise". This "noise" will of course not be random as the external noise but let us anyway assume it is. We are now ready to start the discussion of the least-squares inversion method.

5.1 THE NON-LINEAR LEAST-SQUARES INVERSION METHOD

We are studying a physical system by exciting it in a carefully controlled way and measure its response y_i at the abscissa values x_i for $i=1,....,M$. In our case x is the time after the current in the transmitter is turned off and y is the late time apparent resistivity. We assume that we have solved the forward problem and have a model function $f(x,b_1,.....,b_N)$ which gives the response of the system at the abscissa value x and for given values of the state variables b_j, $j=1,.....,N$. The state variables are in our case the conductivities and thicknesses of the resistivity layers. We want to determine the state of the system from the measured responses. But since the responses are subject to a random noise, we can clearly not determine the state variables with absolute certainty. If the responses y_{0i} give the state variables b_{0j} then a repeated measurement giving y_{1i} would result in the state variables b_{1j}, not exactly the same as b_{0j} because y_{1i} are not the same as y_{0i}, due to the random noise. If we repeat this many times we will eventually get a distribution of the possible state vectors. We will use component and vector notation interchangeably from now on. An N-dimensional state vector will either be represented by its components b_j or as $\boldsymbol{b}$ and an M-dimensional response vector either as y_i or $\boldsymbol{y}$. The possible state vectors will be distributed around the true state vector of the system $\boldsymbol{b}_t$ and it is this statistical distribution of the possible state vectors that we want to determine. We would rather not have to repeat the measurement many times and the question is therefore if we can from a single measured response vector $\boldsymbol{y}_0$ determine the most probable state vector $\boldsymbol{b}_0$ and say something about the statistical distribution of the possible state vectors of the system.

If we had repeated the measurement many times, the resulting response vectors would have been distributed around the noise free response of the system, $f(x_i,\boldsymbol{b}_t)$, and the distribution would reflect the statistical distribution of the noise. Let us now assume that the noise in the individual y_i values is independent of the noise in the other response values. (This is surely

not the case when we are modeling transient electro-magnetic response of a non-horizontal resistivity structure by horizontally layered models !) We furthermore assume that the noise is distributed according to a normal distribution. Then the probability for obtaining the responses in the intervals $y_i \pm dy_i$ is given as

$$P'(y_1,.....,y_M)\, dy_1 \cdots dy_M \;=\; C' e^{-\sum_{i=1}^{M}(y_i - f(x_i,b_t))^2/\sigma_i^2} dy_1 \cdots dy_M \qquad (5.1)$$

where C' is defined so that

$$\int P'(\boldsymbol{y})\, d\boldsymbol{y} \;=\; 1 \qquad (5.1a)$$

and σ_i are the standard deviations for the response values (and should not be mixed with the layer conductivities).

Let us now go back to the real situation which is that we have a measured response vector $\boldsymbol{y}_0$ and we do not know the true state vector of the system $\boldsymbol{b}_t$. Inspired by the probability distribution in (5.1) we now take a different point of view. We assume that the response vector $\boldsymbol{y}_0$ that we measured is the most probable one and that if we take the response vector in the probability density in (5.1) to be $\boldsymbol{y}_0$ and the state vector to be arbitrary instead of $\boldsymbol{b}_t$, then we get a probability density for the possible state vectors. We assume therefore that if we measure the response vector $\boldsymbol{y}_0$, then the probability that the system actually has the components of the state vector in the intervals $b_i \pm db_i$ is given as

$$P(b_1,.....,b_N)\, db_1 \cdots db_N \;=\; C\, e^{-\sum_{i=1}^{M}(y_{0i} - f(x_i,b))^2/\sigma_i^2} db_1 \cdots db_N \qquad (5.2)$$

where C is defined so that

$$\int P(\boldsymbol{b})\, d\boldsymbol{b} \;=\; 1 \qquad (5.2a)$$

Now when we have established a candidate for the probability density for the state vectors, it is a rather straightforward matter to determine the most probable state vector $\boldsymbol{b}_0$ of the system corresponding to the measured response vector $\boldsymbol{y}_0$. It is simply the state vector that makes the probability density in (5.2) take maximum value, or equivalently, makes the **chi-square** sum

$$\chi^2(\boldsymbol{b}) \;=\; \sum_{i=1}^{M}(y_{0i} - f(x_i,\boldsymbol{b}))^2/\sigma_i^2 \qquad (5.3)$$

take minimum value. $\chi^2(\boldsymbol{b})$ in (5.3) will in general have no well defined maximum value but it will have a minimum value ≥ 0. The state vector $\boldsymbol{b}_0$ that minimizes χ^2 will therefore make all the partial derivatives of χ^2 with respect to the state variables vanish, that is

$$\frac{\partial \chi^2(\boldsymbol{b}_0)}{\partial b_j} = -2\sum_{i=1}^{M}\frac{1}{\sigma_i^2}[y_{0i} - f(x_i,\boldsymbol{b}_0)]\frac{\partial f(x_i,\boldsymbol{b}_0)}{\partial b_j} = 0 \quad ; \quad j = 1,....,N \ . \tag{5.4}$$

Until now we have made no explicit statement about the model function $f(x,\boldsymbol{b})$ other than it is a solution to the forward problem that enables us to calculate the theoretical response of the system for any state vector $\boldsymbol{b}$ and for arbitrary abscissa value x. If the model function depends linearly on the state vector, then we have

$$f(x_i,\boldsymbol{b}) = g(x_i) + \sum_{k=1}^{N} h_k(x_i)b_k \tag{5.5}$$

and (5.4) becomes

$$\frac{\partial \chi^2(\boldsymbol{b}_0)}{\partial b_j} = -2\sum_{i=1}^{M}\frac{1}{\sigma_i^2}[y_{0i} - g(x_i) - \sum_{k=1}^{N} A_{i,k}'b_{0k}]A_{i,j}' \quad ; \quad A_{i,j}' = h_j(x_i) \ . \tag{5.6}$$

These are N equations which can be solved for the N unknown components of the state vector $\boldsymbol{b}_0$ and we can thus minimize χ^2 in a single step. (Note that equation (5.6) is basically the same as (4.90).)

The situation is not quite as simple if $f(x_i,\boldsymbol{b})$ is a non-linear function of the state vector. In that case we can not solve (5.4) and hence minimize (5.3) in a single step but have to do it iteratively. To start the iteration process we pick a starting value $\boldsymbol{b}$ for the state vector, which we assume is not too far from the desired vector $\boldsymbol{b}_0$. We want to determine an increment $\delta\boldsymbol{b}$ such that $\boldsymbol{b}_0 = \boldsymbol{b} + \delta\boldsymbol{b}$. To achieve this we make a first order Taylor expansion of the model function around $\boldsymbol{b}$ and write

$$f(x_i,\boldsymbol{b} + \delta\boldsymbol{b}) = f(x_i,\boldsymbol{b}) + \sum_{k=1}^{N}\frac{\partial f(x_i,\boldsymbol{b})}{\partial b_k}\delta b_k \ . \tag{5.7}$$

Inserting this for $f(x_i,\boldsymbol{b}_0)$ in (5.4) and evaluating the partial derivative of the model function in (5.4) at $\boldsymbol{b}$ instead of at $\boldsymbol{b}_0$ we get the equation

$$\sum_{i=1}^{M}\frac{1}{\sigma_i^2}[y_{0i} - f(x_i,\boldsymbol{b}) - \sum_{k=1}^{N}\frac{\partial f(x_i,\boldsymbol{b})}{\partial b_k}\delta b_k]\frac{\partial f(x_i,\boldsymbol{b})}{\partial b_j} = 0 \tag{5.8}$$

Let us, for convenience, introduce the $M\times N$-dimensional partial derivative matrix $\boldsymbol{A}$ and the M-dimensional vector $\boldsymbol{f}$ defined as

$$\boldsymbol{A}(\boldsymbol{b}) = \left(A_{i,j}(\boldsymbol{b})\right) = \left(\frac{1}{\sigma_i}\frac{\partial f(x_i,\boldsymbol{b})}{\partial b_j}\right)$$

$$\boldsymbol{f}(\boldsymbol{b}) = \left(f_i(\boldsymbol{b})\right) = \left(f(x_i,\boldsymbol{b})\right) \quad . \tag{5.9}$$

Equation (5.8) can now be written as

$$\sum_{i=1}^{M}[\sigma_i^{-1}(y_{0i} - f(x_i,\boldsymbol{b}))A_{i,j}(\boldsymbol{b}) - A_{i,j}(\boldsymbol{b})A_{i,k}(\boldsymbol{b})\delta b_k] = 0 \tag{5.10}$$

and by defining

$$\boldsymbol{c} = \left(\frac{y_{0i} - f(x_i,\boldsymbol{b})}{\sigma_i}\right) \tag{5.9a}$$

equation (5.10) can be written in vector notation as

$$\boldsymbol{A}^T\cdot\boldsymbol{c} - \boldsymbol{A}^T\cdot\boldsymbol{A}\cdot\delta\boldsymbol{b} = 0 \quad . \tag{5.10a}$$

Equations (5.10) and (5.10a) consist of N equations where everything is known except $\delta\boldsymbol{b}$ and they can therefore be solved for $\delta\boldsymbol{b}$. In the next section we will discuss a powerful method for solving a set of equations like (5.10a) and (4.91). The vector $\boldsymbol{b}' = \boldsymbol{b} + \delta\boldsymbol{b}$ will now, hopefully, be closer to $\boldsymbol{b}_0$ than $\boldsymbol{b}$ is and hence $\chi^2(\boldsymbol{b}') \leq \chi^2(\boldsymbol{b})$. To reduce χ^2 further, we repeat the above procedure with $\boldsymbol{b}$ replaced by $\boldsymbol{b}'$ and continue this iterative process until χ^2 ceases to decrease and we conclude that we have minimized χ^2 and reached $\boldsymbol{b}_0$.

In practice things are not as straightforward as we have anticipated here because the iterative process that we have been describing often fails to converge. There are mainly two things that may go wrong. The first thing is that if $\boldsymbol{b}$ is a poor initial guess we may well find that $\chi^2(\boldsymbol{b} + \delta\boldsymbol{b}) \geq \chi^2(\boldsymbol{b})$. The other thing that frequently goes wrong is that the matrix $\boldsymbol{A}^T\cdot\boldsymbol{A}$ in (5.10a) is singular, or very close to being singular, resulting in unreasonable large values of some of the components of $\delta\boldsymbol{b}$. These problems make the above discussed iterative process too vulnerable to be used in practice. In section 5.3 we will discuss an elegant modification of this iterative process which solves both the above mentioned problems and makes the process very robust and almost certain to converge.

Let us conclude this section by discussing a little bit the probability distribution of the possible state vectors $\boldsymbol{b}$ corresponding to the measured state vector $\boldsymbol{y}_0$. In equation (5.2) we wrote an anticipated expression for the probability that the state vector of the system has its components in the intervals $b_i \pm db_i$ given as

$$P(\boldsymbol{b})\,d\boldsymbol{b} = Ce^{-\chi^2}\,d\boldsymbol{b} \tag{5.2}$$

where χ^2 is given by (5.3). We assume that we have determined the most probable state vector $\boldsymbol{b}_0$ that minimizes χ^2 and hence maximizes the probability density in (5.2). We want

to find an explicit expression for the probability that the state vector is not $\boldsymbol{b}_0$ but rather $\boldsymbol{b}_0 + \delta\boldsymbol{b}$ where $\delta\boldsymbol{b}$ is small. To do that we make a second order Taylor expansion of χ^2 around $\boldsymbol{b}_0$ and write

$$\chi^2(\boldsymbol{b}_0 + \delta\boldsymbol{b}) \approx \chi^2(\boldsymbol{b}_0) + \sum_{k=1}^{N} \frac{\partial\chi^2(\boldsymbol{b}_0)}{\partial b_k}\delta b_k + \frac{1}{2}\sum_{k,l=1}^{N} \frac{\partial^2\chi^2(\boldsymbol{b}_0)}{\partial b_k \partial b_l}\delta b_k \delta b_l \quad . \tag{5.11}$$

We are assuming that $\boldsymbol{b}_0$ minimizes χ^2 and therefore we have $\partial\chi^2(\boldsymbol{b}_0)/\partial b_k = 0$. To determine the second derivative of χ^2 with respect to the state vector components we differentiate equation (5.4) and get

$$\frac{\partial^2\chi^2(\boldsymbol{b}_0)}{\partial b_k \partial b_l} = 2\sum_{i=1}^{M} \frac{1}{\sigma_i^2}\left[\frac{\partial f_i(\boldsymbol{b}_0)}{\partial b_k}\frac{\partial f_i(\boldsymbol{b}_0)}{\partial b_l} - (y_{0i} - f_i(\boldsymbol{b}_0))\frac{\partial^2 f_i(\boldsymbol{b}_0)}{\partial b_k \partial b_l}\right] \quad . \tag{5.12}$$

The second term under the sum can usually be neglected because we have adjusted $\boldsymbol{b}_0$ such that $y_{0i} - f_i(\boldsymbol{b}_0)$ is merely the random noise and usually the second term will therefore practically average out when the summation is taken. The second derivative of χ^2 can therefore be written as

$$\frac{\partial^2\chi^2(\boldsymbol{b}_0)}{\partial b_k \partial b_l} \approx 2\sum_{i=1}^{M} \frac{1}{\sigma_i^2}\frac{\partial f_i(\boldsymbol{b}_0)}{\partial b_k}\frac{\partial f_i(\boldsymbol{b}_0)}{\partial b_l} = 2\sum_{i=1}^{M} A^T{}_{k,i}(\boldsymbol{b}_0)A_{i,l}(\boldsymbol{b}_0) \tag{5.13}$$

where we have introduced the partial derivative matrix defined in (5.9). Going now back to (5.11) we see that we can write

$$\chi^2(\boldsymbol{b}_0 + \delta\boldsymbol{b}) \approx \chi^2(\boldsymbol{b}_0) + \delta\boldsymbol{b}^T \cdot \boldsymbol{A}^T \cdot \boldsymbol{A} \cdot \delta\boldsymbol{b} \tag{5.14}$$

and the probability density for the state vectors in the vicinity of $\boldsymbol{b}_0$ is therefore given as

$$P(\boldsymbol{b}_0 + \delta\boldsymbol{b}) \approx Ce^{-\chi^2(\boldsymbol{b}_0)}e^{-\delta\boldsymbol{b}^T \cdot \boldsymbol{D} \cdot \delta\boldsymbol{b}} \tag{5.15}$$

where we have defined

$$\boldsymbol{D} = \boldsymbol{A}^T \cdot \boldsymbol{A} \quad . \tag{5.16}$$

(A word about notation. We do here take $\boldsymbol{b}$ to mean a column vector and $\boldsymbol{b}^T$ to mean a row vector, that is

$$\boldsymbol{b} = \begin{pmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_N \end{pmatrix} \quad ; \quad \boldsymbol{b}^T = (b_1, \ldots, b_N)$$

and matrices are denoted by bold capital letters. This notation will be used through out this chapter.) Equation (5.15) can be used to estimate confidence limits for the state vector and

we will turn to this point later.

We have now investigated the principles of least-squares inversion and seen that this method can be thought of as the one that gives the most probable state vector corresponding to the measured response. In the case when the forward problem is non-linear, we saw that the inversion has to be performed iteratively. But we also anticipated that we would run into problems if we adapted the bare iterative process that has been developed in this section and that it has to be modified to make sure that it converges. These modifications will be the subject of section 5.3 .

We have have now been confronted twice with the problem of solving a set of N equations for N unknowns, both in equations (5.10) or (5.10a) in this section and in equations (4.90) or (4.91) when discussing the determination of digital filter coefficients for numerical integration. In both cases we have been aware of the problem that the matrices involved might be singular or nearly singular and that this problem needs a special attention. It is therefore about time that we turn to this point which will be the subject of the next section.

5.2 SOLUTION OF A SYSTEM OF LINEAR EQUATIONS BY SINGULAR VALUE DECOMPOSITION

The systems of linear equations that we have been faced with and have to solve, equations (5.10a) and (4.91), are of the form

$$\boldsymbol{A}^T \cdot \boldsymbol{z} = \boldsymbol{A}^T \cdot \boldsymbol{A} \cdot \boldsymbol{x} \tag{5.17}$$

where $\boldsymbol{z}$ is a known M-dimensional vector and $\boldsymbol{A}$ is a known $M \times N$-dimensional matrix where $M \geq N$ and we want to determine the N-dimensional vector $\boldsymbol{x}$. By defining

$$\boldsymbol{y} = \boldsymbol{A}^T \cdot \boldsymbol{z} \quad ; \quad \boldsymbol{D} = \boldsymbol{A}^T \cdot \boldsymbol{A} \tag{5.16a}$$

equation (5.17) can be written as

$$\boldsymbol{y} = \boldsymbol{D} \cdot \boldsymbol{x} \ . \tag{5.17a}$$

If the matrix $\boldsymbol{D}$ is not singular and hence invertible, the solution to (5.17a) is simply given as

$$\boldsymbol{x} = \boldsymbol{D}^{-1} \cdot \boldsymbol{y} \tag{5.18}$$

and there exist various methods to determine the inverse matrix $\boldsymbol{D}^{-1}$ and hence to determine $\boldsymbol{x}$. If $\boldsymbol{D}$ is singular then its inverse does not exist and (5.17a) has no well defined solution, in fact it has infinitely many solutions as we will see. But we do need a solution and must therefore either guarantee that $\boldsymbol{D}$ is not singular or find a sensible way of choosing the "best" solution from the infinite set of solutions in the case when $\boldsymbol{D}$ is singular. The first alternative is the right thing to do in the case of the non-linear least-squares inversion problem but for the determination of the digital filter coefficients the second alternative is the right one. Let us now analyse the problem of solving (5.17) and (5.17a) a little more closely.

It is evident from (5.16a) that $\boldsymbol{D}$ is an $N \times N$ symmetric and non-negative matrix. By non-negative we mean that $\boldsymbol{x}^T \cdot \boldsymbol{D} \cdot \boldsymbol{x} \geq 0$ for any N-dimensional vector $\boldsymbol{x}$. There is a theorem of linear algebra stating that if $\boldsymbol{D}$ is an $N \times N$ symmetric and non-negative matrix then it can be written as

$$\boldsymbol{D} = \boldsymbol{V} \cdot \boldsymbol{W} \cdot \boldsymbol{V}^T \tag{5.19}$$

where $\boldsymbol{V}$ is an $N \times N$ orthogonal matrix i.e.

$$\boldsymbol{V}^T \cdot \boldsymbol{V} = \boldsymbol{V} \cdot \boldsymbol{V}^T = \boldsymbol{I} \tag{5.20}$$

($\boldsymbol{I}$ is the $N \times N$ identity matrix) and $\boldsymbol{W}$ is an $N \times N$ diagonal matrix with non-negative diagonal elements, that is

$$\boldsymbol{W} = \text{diag}\,[w_1, \ldots\ldots, w_N] \quad , \quad w_i \geq 0 \quad i = 1, \ldots\ldots, N \ . \tag{5.21}$$

(For a proof of this theorem see e.g. Hoffman and Kunze 1971, pp. 311-341). The product on the right hand side of (5.19) is called a **singular value decomposition** of $\boldsymbol{D}$.

Armed with this theorem we immediately see that if all the diagonal elements of $\boldsymbol{W}$ are different from zero, then the inverse of $\boldsymbol{D}$ is given as

$$\boldsymbol{D}^{-1} = \boldsymbol{V} \cdot \boldsymbol{W}^{-1} \cdot \boldsymbol{V}^T \tag{5.22}$$

where

$$\boldsymbol{W}^{-1} = \text{diag}\,[w_1^{-1}, \ldots\ldots, w_N^{-1}] \ . \tag{5.23}$$

If on the other hand one or more of the diagonal elements of $\boldsymbol{W}$ are zero, then $\boldsymbol{W}^{-1}$ has no meaning and $\boldsymbol{D}$ is not invertible and hence singular. But we are now able to analyse the problem in all details and see exactly how $\boldsymbol{D}$ fails to be invertible. To see what is going on we look at $\boldsymbol{D}$ as a representation of a linear mapping F of an N-dimensional Euclidean vector space with the standard inner product

$$\langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle = \boldsymbol{x}^T \cdot \boldsymbol{y} = \sum_{i=1}^{N} x_i y_i \tag{5.24}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are arbitrary, onto itself. That $\boldsymbol{D}$ is a representation of F means that there exists a set of orthonormal basis vectors $B_1 = \{\boldsymbol{\beta}_1, \ldots\ldots, \boldsymbol{\beta}_N\}$ in the vector space such that the image of the vector $\boldsymbol{x}$ is given as

$$\boldsymbol{y} = F(\boldsymbol{x}) = \boldsymbol{D} \cdot \boldsymbol{x} \ . \tag{5.25}$$

Equations (5.19) and (5.20) now state that F has N orthonormal eigenvectors whose components are, in the basis B_1, given as the columns of the matrix $\boldsymbol{V}$ and the eigenvalues are the diagonal elements of $\boldsymbol{W}$. To see this we write

$$\boldsymbol{V} = \begin{bmatrix} V_{1,1} & \cdots & V_{1,N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ V_{N,1} & \cdots & V_{N,N} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{v}_1)_1 & \cdots & (\boldsymbol{v}_N)_1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ (\boldsymbol{v}_1)_N & \cdots & (\boldsymbol{v}_N)_N \end{bmatrix} = [\boldsymbol{v}_1,.....,\boldsymbol{v}_N] \tag{5.26}$$

where we have identified the i'th component of the j'th column vector $\boldsymbol{v}_j$ of the $\boldsymbol{V}$ with

$$(\boldsymbol{v}_j)_i = V_{i,j} \ . \tag{5.27}$$

$\boldsymbol{V}^T$ can now be written as

$$\boldsymbol{V}^T = \begin{bmatrix} \boldsymbol{v}_1{}^T \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{v}_N{}^T \end{bmatrix} \tag{5.28}$$

and equation (5.20) states that

$$\begin{bmatrix} \boldsymbol{v}_1{}^T \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{v}_N{}^T \end{bmatrix} \cdot [\boldsymbol{v}_1,.....,\boldsymbol{v}_N] = \boldsymbol{I} \quad \text{or} \quad \boldsymbol{v}_i{}^T \cdot \boldsymbol{v}_j = \delta_{i,j} \tag{5.29}$$

and hence that the vectors $\boldsymbol{v}_i$ are orthonormal. It is now evident from (5.19) that the vectors $\boldsymbol{v}_i$ are eigenvectors of $\boldsymbol{D}$ with eigenvalues w_i because

$$\boldsymbol{D} \cdot \boldsymbol{v}_i = [\boldsymbol{v}_1,.....,\boldsymbol{v}_N] \cdot \text{diag}\,[w_1,.....,w_N] \cdot \begin{bmatrix} \boldsymbol{v}_1{}^T \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{v}_N{}^T \end{bmatrix} \cdot \boldsymbol{v}_i = [\boldsymbol{v}_1,.....,\boldsymbol{v}_N] \cdot \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ w_i \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} = w_i \boldsymbol{v}_i \ . \tag{5.30}$$

If $\boldsymbol{x}$ is an arbitrary vector, then the components of the vector

$$\boldsymbol{x}' = \boldsymbol{V}^T \boldsymbol{x} = \begin{bmatrix} \boldsymbol{v}_1{}^T \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{v}_N{}^T \end{bmatrix} \cdot \boldsymbol{x} \tag{5.31}$$

are the projections of $\boldsymbol{x}$ onto the N orthonormal eigenvectors and we see that $\boldsymbol{x}$ and $\boldsymbol{x}'$ can be thought of as two representations of the same vector in terms of different sets of basis vectors, namely $B_1 = \{\boldsymbol{\beta}_1,.....,\boldsymbol{\beta}_N\}$ and $B_2 = \{\boldsymbol{v}_1,.....,\boldsymbol{v}_N\}$. Furthermore if we multiply (5.25) from the left by $\boldsymbol{V}^T$ we find by using (5.19) and (5.20) that

$$y' = V^T y = V^T \cdot V \cdot W \cdot V^T x = W x' \tag{5.32}$$

and we see that if we use the basis B_2 of orthonormal eigenvectors instead of B_1, then the linear mapping F is represented by the diagonal matrix $W = \text{diag}\,[w_1,.....,w_N]$.

Now we can see what goes wrong if one or more of the eigenvalues w_i are zero. If the eigenvalue w_l is zero, then we see from (5.30) and (5.32) that all vectors proportional to v_l, and hence the projection of any vector onto v_l, is transformed to the zero vector. We have therefore that if x is mapped to y then

$$y = D \cdot x = D \cdot (x + a v_l) \tag{5.33}$$

where a is arbitrary and there are infinitely many vectors mapped to y. The equation

$$y = D \cdot x \tag{5.25}$$

has therefore not a well defined unique solution but infinitely many solutions. The question is now, if we need a solution to (5.25) for a given y and D is singular with w_l equal to zero, is there any natural way of choosing a particular member from the infinitely many solutions? The answer is yes and the the most natural thing to do is to choose the solution vector x to have no component along the eigenvector v_l corresponding to the zero eigenvalue.

To see that this is indeed the most natural choice in our problem we go back to equation (5.17) and recall that $D = A^T \cdot A$ and $y = A^T \cdot z$. We can use the singular value decomposition of D, expressed in (5.19), to make a similar decomposition of the $M \times N$ matrix A. Since all the eigenvalues in the diagonal of W are non-negative, we can define the positive square root of W. We define

$$\Lambda = \text{diag}\,[\lambda_1,.....,\lambda_N] \quad , \quad \lambda_i = \sqrt{w_i} \quad i = 1,.....,N \tag{5.34}$$

and have therefore that

$$\Lambda^2 = W \ . \tag{5.35}$$

If all the eigenvalues are different from zero then Λ is invertible and

$$\Lambda^{-1} = \text{diag}\,[\lambda_1^{-1},.....,\lambda_N^{-1}] \tag{5.36}$$

Let us define the $M \times N$ matrix

$$U = A \cdot V \cdot \Lambda^{-1} \ . \tag{5.37}$$

We see, by using the singular value decomposition of $D = A^T \cdot A$ expressed in (5.19) that

$$U^T \cdot U = \Lambda^{-1} \cdot V^T \cdot A^T \cdot A \cdot V \cdot \Lambda^{-1} = \Lambda^{-1} \cdot V^T \cdot V \cdot W \cdot V^T \cdot V \cdot \Lambda^{-1} = \Lambda^{-1} \cdot W \cdot \Lambda^{-1} = I \tag{5.38}$$

where I is the $N \times N$ identity matrix. Equation (5.38) states that the N column vectors of U, which are M-dimensional, are orthonormal. By multiplying (5.37) from the right by Λ and

then by V^T we find that

$$A = U \cdot \Lambda \cdot V^T \tag{5.39}$$

which is a singular value decomposition of A and this equation does also hold true if some of the eigenvalues w_i and hence λ_i are zero.

Now we return to the question why it is most natural, in the case that D is singular with $w_l = 0$, to choose the particular solution x of (5.25) with zero component along the eigenvector v_l. Using the singular value decomposition (5.39) of A we can write

$$y = A^T \cdot c = V \cdot \Lambda \cdot U^T \cdot c \ . \tag{5.40}$$

Since $w_l = 0$ we also have that $\lambda_l = 0$ and if we take the projection of y onto v_l we find

$$v_l^T \cdot y = (v_l^T \cdot V) \cdot (\Lambda \cdot U^T \cdot c) = 0 \tag{5.41}$$

because the row vector $v_l^T \cdot V$ has all components equal to zero, except the l'th component which is one, and the column vector $\Lambda \cdot U^T \cdot c$ has the l'th component equal to zero because $\lambda_l = 0$. We see therefore that y has no component along v_l and it is therefore no reason to let the solution vector x have any component along v_l either. Even in the more general case when y is not on the form (5.40) and can therefore have component along v_l, it is still natural to take x to have no component along v_l, because then x will be the shortest vector that solves (5.25) in the sense that all other solutions q will have larger norm than x, that is to say, $|q| > |x|$ where

$$|q| = (q^T \cdot q)^{1/2} = \left[\sum_{i=1}^{N} q_i^2 \right]^{1/2} . \tag{5.42}$$

To see this we note that all solutions can be written on the form

$$q = x + a v_l \ . \tag{5.43}$$

(We are here assuming that we have only one eigenvalue equal to zero, but the argument is easily generalized to the case when more than one of the eigenvalues are zero.) If we express q in the orthonormal eigenvector basis $B_2 = \{v_1, \ldots, v_N\}$ then x' has the l'th component equal to zero and we have

$$|q|^2 = |q'|^2 = \sum_{\substack{i=1 \\ i \neq l}}^{N} x_i'^2 + a^2 = |x|^2 + a^2 \tag{5.44}$$

and it is obvious that x is the shortest solution vector.

The question is now how we actually get hold on the solution x which has no component along the eigenvector v_l corresponding to a zero eigenvalue. It is easy because we

immediately see that if we *replace the undefined diagonal element* w_l^{-1} *in* $\boldsymbol{W}^{-1}$ *as given in* (5.23) *by zero*, then

$$\boldsymbol{x} = \boldsymbol{V} \cdot \boldsymbol{W}^{-1} \cdot \boldsymbol{V}^T \cdot \boldsymbol{y} \tag{5.45}$$

will have no component along $\boldsymbol{v}_l$. We can call (5.45) a generalized inversion of of (5.25). The same game can of course be played if more than one of the eigenvalues are zero. It frequently happens that $\boldsymbol{D}$ is not strictly singular but one or more of the eigenvalues are not exactly zero but very small and $\boldsymbol{D}$ is very close to being singular. The corresponding diagonal elements of $\boldsymbol{W}^{-1}$ will be very large, resulting in unreasonably large components of the solution vector $\boldsymbol{x}$ in (5.45). This means that the vector $\boldsymbol{y}$ in (5.25) is almost independent of the components of $\boldsymbol{x}$ along the eigenvectors corresponding to the very small eigenvalues. In this case it is best to simply replace the diagonal elements w_i^{-1} in $\boldsymbol{W}^{-1}$ corresponding to the very small eigenvalues by zero and disregard the contribution of their eigenvectors to the solution $\boldsymbol{x}$.

We have now learned how to find a generalized "inverse" of the matrix $\boldsymbol{D}$, in (5.16a) in the case when the matrix is either strictly or very close to being singular. This solves the problem that we foresaw that we might encounter when solving equations (4.91) or (4.92) in order to determine the digital filter coefficients. But could we also use this approach in the iterative process of non-linear least-squares inversions, to solve equation (5.10a) when the matrix $\boldsymbol{D} = \boldsymbol{A}^T \cdot \boldsymbol{A}$ is singular or nearly singular? Yes we could, but the above described method has a property which becomes unfortunate when it is used in the non-linear inversion process. When we demand that the solution vector has no components along eigenvectors with zero or very small eigenvalues, we are in fact reducing the number of degrees of freedom in the solution space. In the non-linear least-squares inversion process this means that we cut down the number of independent model parameters. Furthermore if $\boldsymbol{D}$ is singular then we see from (5.39) that $\boldsymbol{A}$ is also singular. In the least-squares inversion problem $\boldsymbol{A}$ is the partial derivatives of the model function f at the abscissa values x_i with respect to the model parameters (see equation (5.9)). If $\boldsymbol{A}$ turns out to be strictly singular it means that the model function does not predict the system response uniquely in terms of the model parameters b_j and hence that the model function f is not a good one. This should not occur and we conclude that $\boldsymbol{D} = \boldsymbol{A}^T \cdot \boldsymbol{A}$ in (5.10a) will not become strictly singular, but it may become very close to being singular. Since we would rather not reduce the number of the independent model parameters by applying the generalized "inverse" of a nearly singular matrix to solve (5.10a) we must find a way to modify equation (5.10a) in such a way that $\boldsymbol{D}$ is never very close to being singular and this will be the subject of the next section.

5.3 THE LEVENBERG-MARQUARDT INVERSION METHOD

The iterative non-linear least-squares inversion process described in section 5.1 may, as we have discussed previously, fail to converge. We will now discuss modifications of this process, originally suggested by Levenberg and later improved by Marquardt, which will make the

process much more robust and almost certain to converge. There are, as was mentioned in section 5.1, mainly two reasons why the iterations may fail to converge. The first reason is that if the initial guess for the state vector $\boldsymbol{b}$ is far from the desired vector $\boldsymbol{b}_0$ that minimizes χ^2 in (5.3) then we may find that $\chi^2(\boldsymbol{b} + \delta\boldsymbol{b}) > \chi^2(\boldsymbol{b})$. The second reason is that if $\boldsymbol{D}$ in (5.10a) is very close to being singular then the resulting $\delta\boldsymbol{b}$ will have very large components leading the iterative process totally astray. These two reasons have one thing in common, namely that going from $\boldsymbol{b}$ to $\boldsymbol{b} + \delta\boldsymbol{b}$ does not bring us closer to the desired state vector $\boldsymbol{b}_0$, usually because $\delta\boldsymbol{b}$ which is the solution to the equation

$$\boldsymbol{A}^T \cdot \boldsymbol{c} = \boldsymbol{A}^T \cdot \boldsymbol{A} \cdot \delta\boldsymbol{b} \tag{5.10a}$$

is too large. In the first case this is because $\boldsymbol{c} = ([y_{0i} - f_i(\boldsymbol{b})]/\sigma_i)$ is large and in the second case because one or more of the eigenvalues of $\boldsymbol{D} = \boldsymbol{A}^T \cdot \boldsymbol{A}$ are very small. If $\boldsymbol{c}$ is large we may say that the procedure described in section 5.1 is too greedy because it tries to bring us in large steps towards $\boldsymbol{b}_0$ and it might have been preferable to be more modest in approaching $\boldsymbol{b}_0$ by taking small steps in the opposite direction to the gradient of χ^2 evaluated at $\boldsymbol{b}$. That is to say it might have been better to take

$$\delta b_j = -a\frac{\partial \chi^2(\boldsymbol{b})}{\partial b_j} = a2\sum_{i=1}^{M} \frac{1}{\sigma_i}[y_{0i} - f_i(\boldsymbol{b})]\frac{1}{\sigma_i}\frac{\partial f_i(\boldsymbol{b})}{\partial b_j} = a2\boldsymbol{A}^T \cdot \boldsymbol{c} \tag{5.46}$$

where a is a small positive number, instead of the solutions to (5.10a). On the other hand if $\boldsymbol{D} = \boldsymbol{V} \cdot \boldsymbol{W} \cdot \boldsymbol{V}^T$ is close to being singular we might do well in replacing $\boldsymbol{W}$ by a diagonal matrix with slightly larger diagonal elements. The brilliant observation of Levenberg and Marquardt was to realize that both of these goals can be achieved by adding to $\boldsymbol{D}$ the $N \times N$ identity matrix multiplied by a positive constant and instead of (5.10a) consider the equation

$$\boldsymbol{A}^T \cdot \boldsymbol{c} = (\boldsymbol{D} + \alpha^2 \boldsymbol{I}) \cdot \delta\boldsymbol{b} \tag{5.47}$$

where α is an adjustable variable. If the left hand side of (5.47) is not too large and the eigenvalues of $\boldsymbol{D}$ not too small, we can take the *Marquardt parameter* α close to zero and equation (5.47) becomes practically the same as (5.10a). If the left hand side of (5.47) is large we can take α large and equation (5.47) becomes the same as (5.46). We can of course by using (5.19) write

$$\boldsymbol{D}_\alpha = \boldsymbol{D} + \alpha^2 \boldsymbol{I} = \boldsymbol{V} \cdot (\boldsymbol{W} + \alpha^2 \boldsymbol{I}) \cdot \boldsymbol{V}^T \tag{5.48}$$

and we see that if some of the eigenvalues of $\boldsymbol{D}$ are very small then we can ensure that $\boldsymbol{D} + \alpha^2 \boldsymbol{I}$ is non-singular by appropriate choice of $\alpha \neq 0$.

From (5.48) we immediately see that the inverse of $\boldsymbol{D} + \alpha^2 \boldsymbol{I}$ is given as

$$\boldsymbol{D}_\alpha^{-1} = (\boldsymbol{D} + \alpha^2 \boldsymbol{I})^{-1} = \boldsymbol{V} \cdot (\boldsymbol{W} + \alpha^2 \boldsymbol{I})^{-1} \cdot \boldsymbol{V}^T \tag{5.49}$$

where

$$(\boldsymbol{W} + \alpha^2\boldsymbol{I})^{-1} = \text{diag}\,[(w_1 + \alpha^2)^{-1},.....,(w_N + \alpha^2)^{-1}] \ . \tag{5.50}$$

We can call (5.49) a damped inverse of $\boldsymbol{D}$ and the solution to (5.47) is now given as

$$\delta\boldsymbol{b} = \boldsymbol{V}\cdot(\boldsymbol{W} + \alpha^2\boldsymbol{I})^{-1}\cdot\boldsymbol{V}^T\cdot\boldsymbol{A}^T\cdot\boldsymbol{c} \tag{5.51}$$

which, by use of (5.39) and the fact that the diagonal elements of $\boldsymbol{W}$ and $\boldsymbol{\Lambda}$ are related through $\lambda_i^2 = w_i$, can be written as

$$\delta\boldsymbol{b} = \boldsymbol{V}\cdot\boldsymbol{\Lambda}_\alpha^{-1}\cdot\boldsymbol{U}^T\cdot\boldsymbol{c} \tag{5.52}$$

where

$$\boldsymbol{\Lambda}_\alpha^{-1} = \text{diag}\left[\frac{\lambda_1}{\lambda_1^2 + \alpha^2},.....,\frac{\lambda_N}{\lambda_N^2 + \alpha^2}\right] \ . \tag{5.53}$$

The Levenberg-Marquardt modification of the iterative non-linear least-squares inversion method consist in using equation (5.52) to determine $\delta\boldsymbol{b}$ instead of solving (5.10a). In each iteration step we start by trying the increment $\delta\boldsymbol{b}$ obtained from (5.52) with the Marquardt parameter α relatively small. If $\chi^2(\boldsymbol{b} + \delta\boldsymbol{b}) \leq \chi^2(\boldsymbol{b})$ we go on to the next iteration step but if $\chi^2(\boldsymbol{b} + \delta\boldsymbol{b}) > \chi^2(\boldsymbol{b})$ we increase α and determine a new $\delta\boldsymbol{b}$ and continue to do so until $\chi^2(\boldsymbol{b} + \delta\boldsymbol{b})$ becomes less than $\chi^2(\boldsymbol{b})$ and then proceed to the next iteration step. Eventually we end up by not being able to decrease χ^2 any further no matter how we increase the Marquardt parameter and we conclude that χ^2 has been minimized and that the most probable state vector $\boldsymbol{b}_0$ corresponding to the measured response vector $\boldsymbol{y}_0$, has been determined.

Some final questions need attention, such as in what range the Marquardt parameter should be taken, what is an appropriate relatively small starting value of α, how much should it be increased if $\chi^2(\boldsymbol{b} + \delta\boldsymbol{b}) > \chi^2(\boldsymbol{b})$ and when should we stop trying to increase α ? The answers to these questions are to some extent dependent on the specific problem under consideration but from equations (5.47) and (5.48) we see that the original principles in section 5.1, on which the derivation of (5.10a) was based, are completely overridden if α^2 is much larger than the largest diagonal element of $\boldsymbol{W}$. On the other hand we remember that the point in replacing (5.10a) by (5.47) was to take α^2 considerable larger than the smallest eigenvalue in $\boldsymbol{W}$ in the case when that eigenvalue is very small. The smallest and the starting value α_0 can therefore for example be taken to be equal to the smallest value of $\lambda_i = \sqrt{w_i}$ but still never less than a certain minimum value α_{min}, that is $\alpha_0 = \max\{\alpha_{\text{min}},\min\{\lambda_i\}\}$. A reasonable value of α_{min} could for example be 10^{-4} . If this value of the Marquardt parameter does not result in decreased value of χ^2 then we could take α_1 equal to the smallest $\lambda_i > \alpha_0$ and if this does not work we could try α_2 as the smallest $\lambda_i > \alpha_1$ and so on. A reasonable criterion for stopping the process of increasing the Marquardt parameter could for example be to stop if $\alpha = \max\{\lambda_i\}$ does not result in decreased value of χ^2 .

5.4 CONFIDENCE LIMITS FOR THE MODEL PARAMETERS

We have discussed the non-linear least-squares inversion method and seen how it can be interpreted as a method to determine the most probable state vector of the system corresponding to a given measured response vector. Because of the inherent uncertainties in the problem, we can not state that the most probable state vector $\boldsymbol{b}_0$ is the true state vector of the system, but we expect it to be close to it. It is of interest to try to estimate how different from $\boldsymbol{b}_0$ the true state vector could be. The question is if we can determine a region around $\boldsymbol{b}_0$, in the N-dimensional model space, so that we can state with reasonable confidence that the true state vector is in this region. In section 5.1 we wrote down an expression for the probability that the true state vector has its components in the intervals $b_i \pm db_i$ and in equation (5.15) we gave an approximate expression for the probability density that the true state vector is given as $\boldsymbol{b}_0 + \delta\boldsymbol{b}$

$$P(\boldsymbol{b}_0 + \delta\boldsymbol{b}) = Ce^{-\chi^2(\boldsymbol{b}_0)}e^{-\Delta\chi^2} \tag{5.54}$$

where

$$\Delta\chi^2 = \delta\boldsymbol{b}^T\cdot\boldsymbol{D}\cdot\delta\boldsymbol{b} = \delta\boldsymbol{b}^T\cdot\boldsymbol{A}^T\cdot\boldsymbol{A}\cdot\delta\boldsymbol{b} \ . \tag{5.55}$$

We can use these equations to define confidence limits for the model parameters e.g. by determining the region around $\boldsymbol{b}_0$ such that the probability density for state vectors in that region is higher than a certain value, or equivalently, such that $\Delta\chi^2$ is less than a predefined value. The most frequently stated confidence limits for inversion problems like the one we are discussing here, are the socalled 68% confidence limits. They are defined as the limits to the region in which the probability density in (5.54) is larger than $P(\boldsymbol{b}_0)e^{-1}$, that is to say, the 68% confidence region in the model parameter space is the region $\boldsymbol{b}_0 + \delta\boldsymbol{b}$ such that

$$\Delta\chi^2 = \delta\boldsymbol{b}^T\cdot\boldsymbol{A}^T\cdot\boldsymbol{A}\cdot\delta\boldsymbol{b} < 1 \ . \tag{5.56}$$

The name originates from the fact that for a single parameter which is distributed according to normal distribution, there is 68.3% probability to observe a value less than one standard deviation from the central value of the distribution. This probability decreases considerably as the number of parameters (dimension of the solution space) increases and the term 68% confidence limit is in this case merely a name for the condition (5.56).

By using the singular value decomposition (5.39) of the matrix $\boldsymbol{A}$ we see that the boundary of the 68% confidence region is defined by the equation

$$\delta\boldsymbol{b}^T\cdot\boldsymbol{V}\cdot\boldsymbol{\Lambda}^2\cdot\boldsymbol{V}^T\cdot\delta\boldsymbol{b} = 1 \ . \tag{5.57}$$

If we express the vector $\delta\boldsymbol{b}$ in terms of the orthonormal basis of eigenvectors B_2 as in equation (5.31) and write $\delta\boldsymbol{b}' = \boldsymbol{V}^T\cdot\delta\boldsymbol{b}$, equation (5.57) can be written as

$$\sum_{i=1}^{N} \lambda_i^2 \delta b_i'^2 = 1 \qquad (5.57a)$$

which defines a surface of an N-dimensional ellipsoid. Therefore the 68% confidence region is the interior of an N-dimensional ellipsoid around $\boldsymbol{b}_0$. The axis of the ellipsoid are parallel to the eigenvectors $\boldsymbol{v}_i$ and from (5.57a) we see that the length of the half axis, from the centre to the surface, are given as λ_i^{-1}. The confidence intervals for the individual components of the state vector are now given as the orthogonal projections of the ellipsoid onto the coordinate axis as shown on figure 5.1 .

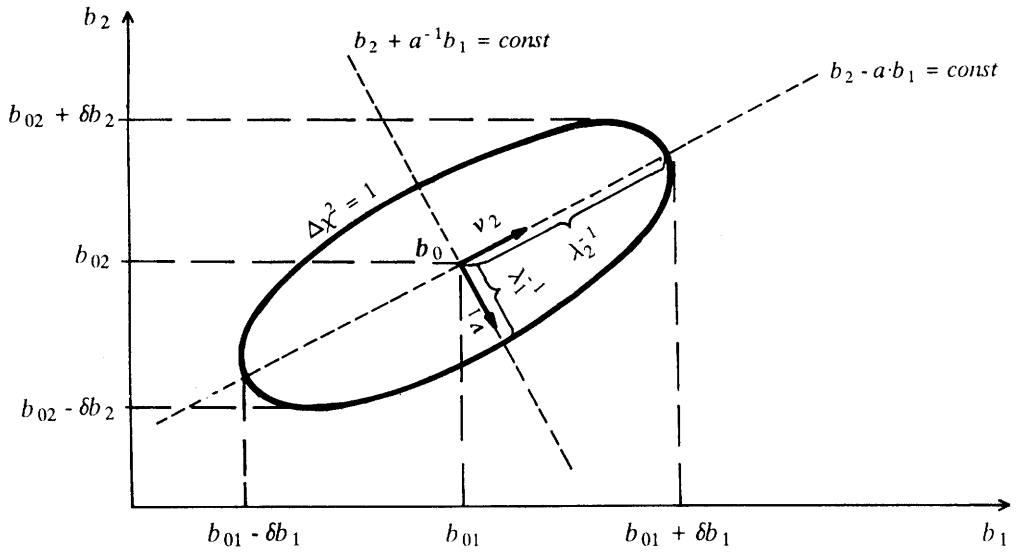

Figure 5.1 Confidence ellipsoid for the model vector

If one or more of the eigenvalues λ_i are very small we see that the confidence ellipsoid is elongated along the corresponding eigenvectors. If one of the eigenvalues, say λ_l, is very small then the ellipsoid is elongated along $\boldsymbol{v}_l$ and we see from figure 5.1 that if $\boldsymbol{v}_l$ has relatively large projection onto the k'th axis in the parameter space then the k'th component, b_k, of the state vector, will have wide confidence limits and is therefore not very well determined. Let us state this in another way. If some of the eigenvalues λ_i are small then the components of eigenvectors corresponding to the small eigenvalues tell us which parameters are not well determined. The model parameters corresponding to the largest components of the eigenvectors will not be well determined. If one component, say the k'th component, of an eigenvector corresponding to a small eigenvalue has considerable larger

absolute value than the others, then the elongated ellipsoid will be nearly parallel to the k-axis and it will only be the model parameter b_k that is not well determined. If, on the other hand, there are two or more of the components of the eigenvector that have high absolute values and of comparable size, then the confidence ellipsoid is tilted relative to the axis and it is a combination of the model parameters corresponding, to the large components that is not well determined. In this case we can say that the parameters are correlated. In the two-dimensional case shown on figure 5.1 the ellipsoid is elongated along ν_2 and tilted upwards. If the eigenvalue λ_2 is very small, then the ellipsoid is effectively a band around the line $b_2 - ab_1 = const$ (where $a = (\nu_2)_2/(\nu_2)_1$) and we see that it is only the composed parameter $b_2 + a^{-1}b_1$ that is determined with any confidence while the parameter $b_2 - ab_1$ could take almost any value. If on the other hand we had $\lambda_1 << \lambda_2$, the ellipsoid would be elongated along ν_1 and $b_2 + a^{-1}b_1$ would be well determined but $b_2 - ab_1$ ill determined. It is worth mentioning here that if we had, instead of the Levenberg-Marquardt method, used the generalized inversion method described in section 5.2 to solve equation (5.10a) in the non-linear least-squares inversion, we would have disregarded the ill determined parameters discussed above by setting them equal to zero and hence cut down the number of model parameters.

There is another way to study uncertainties and correlations among the components of the state vector $\boldsymbol{b}_0$ obtained by the inversion. This method is based on the concepts of *variance* and *covariance* of the model-parameters. Let us consider the case when we have N random variables x_i which have probability distribution described by the probability density

$$f(x_1,.....,x_N) \quad ; \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1,.....,x_N)\, dx_1 \cdots dx_N \;=\; 1 \tag{5.58}$$

and with the expectation values

$$\hat{x}_i \;=\; \int x_i f(\boldsymbol{x})\, d\boldsymbol{x} \quad . \tag{5.59}$$

The variance of the i'th variable is defined as the expectation value of the square of the difference between an observed value x_i an $\hat{x}_i$, that is

$$\mathrm{var}(x_i) \;=\; \int (x_i - \hat{x}_i)^2 f(\boldsymbol{x})\, d\boldsymbol{x} \tag{5.60}$$

and the covariance between the i'th and the j'th variables is defined as

$$\mathrm{cov}(x_i,x_j) \;=\; \int (x_i - \hat{x}_i)(x_j - \hat{x}_j) f(\boldsymbol{x})\, d\boldsymbol{x} \quad . \tag{5.61}$$

For those not familiar with probability theory we point out that the expectation value of a quantity $g(x_1,.....,x_N)$ which is a function of the random variables x_i is the value g that we

would expect to get and it is given as the weighted average of the values $g(x_1,.....,x_N)$ for the different values of x_i, where the weight function is the probability density, that is

$$\hat{g} = \int g(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} \quad . \tag{5.62}$$

From (5.60) we see that the variance of x_i is a measure of how much we can expect an observed value to deviate from the expected value $\hat{x}_i$. Equations (5.58) to (5.61) can now be interpreted in the following way. We do not know the values of the components of the vector $\boldsymbol{x}$ but we know that they have the probability density (5.58). We expect the components of $\boldsymbol{x}$ to have values given by (5.59) but we are not certain and the variances in (5.60) are measures of the uncertainties. The covariances in (5.61) measure how much we can expect two compounds to deviate simultaneously from their expected values.

Let us now consider the probability density for the deviations $\delta\boldsymbol{b}$ from the state vector $\boldsymbol{b}_0$ given by (5.54) and (5.55). If we set $\boldsymbol{x} = \delta\boldsymbol{b}$ we get

$$P(\boldsymbol{x}) = Be^{-\boldsymbol{x}^T \cdot D \cdot \boldsymbol{x}} \tag{5.63}$$

To determine the normalization constant B we demand that

$$B^{-1} = \int e^{-\boldsymbol{x}^T \cdot D \cdot \boldsymbol{x}}\, d\boldsymbol{x} \quad . \tag{5.64}$$

To evaluate the integral in (5.64) we use a well known formula of calculus stating that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(y_1(\boldsymbol{x}),.....,y_N(\boldsymbol{x}))\, dx_1 \cdots dx_N = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(y_1,.....,y_N) \det\left[\frac{\partial x_i}{\partial y_j}\right] dy_1 \cdots dy_N \quad .(5.65).$$

Now we have $D = V \cdot W \cdot V^T = V \cdot \Lambda^2 \cdot V^T$ and by using the orthonormal eigenvectors as basis vectors and changing the variables in the integral (5.64) to $\boldsymbol{x}' = V^T \cdot \boldsymbol{x}$ we find that

$$\int e^{-\boldsymbol{x}^T \cdot D \cdot \boldsymbol{x}}\, d\boldsymbol{x} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\sum_{i=1}^{N} \lambda_i^2 x_i'^2} \det V\, dx_1' \cdots dx_N'$$

$$= \int_{-\infty}^{\infty} e^{-\lambda_1^2 x_1'^2}\, dx_1' \cdots \int_{-\infty}^{\infty} e^{-\lambda_N^2 x_N'^2}\, dx_N' = \pi^{N/2} \left[\prod_{i=1}^{N} \lambda_i\right]^{-1} \tag{5.66}$$

where we have used that the determinant of an orthogonal matrix obeying (5.20) is one. The normalization constant is therefore given as

$$B = \pi^{-N/2} \prod_{i=1}^{N} \lambda_i \quad . \tag{5.67}$$

It is now easy to show that the expectation value of $\boldsymbol{x}$ is the zero vector. Let us evaluate the

components of the vector $\hat{\boldsymbol{x}}' = \boldsymbol{V}^T \cdot \hat{\boldsymbol{x}}$ which is given as

$$\hat{x_i}' = B \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i' e^{-\sum_{k=1}^{N} \lambda_k^2 x_k'^2} \det \boldsymbol{V}\, dx_1' \cdots dx_N'$$

$$= B \int_{-\infty}^{\infty} e^{-\lambda_1^2 x_1'^2} dx_1' \cdots \int_{-\infty}^{\infty} x_i' e^{-\lambda_i^2 x_i'^2} dx_i' \cdots \int_{-\infty}^{\infty} e^{-\lambda_N^2 x_N'^2} dx_N' = 0 \qquad (5.68)$$

because the integrand in the i'th integral is antisymmetric. But since all the projections of the vector $\hat{\boldsymbol{x}}$ onto the orthonormal basis of eigenvectors are zero we conclude that $\hat{\boldsymbol{x}} = 0$. We do now introduce the *covariance matrix* $\boldsymbol{C}$ whose components are defined as

$$C_{i,j} = \int x_i x_j P(\boldsymbol{x})\, d\boldsymbol{x} \quad . \qquad (5.69)$$

Since $\hat{x}_i = 0$ we see that the diagonal elements of $\boldsymbol{C}$ are the variances $\mathrm{var}(x_i)$ and the off-diagonal elements are the covariances $\mathrm{cov}(x_i,x_j)$. To evaluate $C_{i,j}$ we use a similar trick as we did when evaluating $\hat{x}_i$ and evaluate the i,j element of the matrix $\boldsymbol{V}^T \cdot \boldsymbol{C} \cdot \boldsymbol{V}$ and find

$$(\boldsymbol{V}^T \cdot \boldsymbol{C} \cdot \boldsymbol{V})_{i,j} = \pi^{-N/2} \prod_{k=1}^{N} \lambda_k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i' x_j' e^{-\sum_{k=1}^{N} \lambda_k^2 x_k'^2} dx_1' \cdots dx_N' = \frac{1}{2} \delta_{i,j} \frac{1}{\lambda_i^2} \qquad (5.70)$$

and we see that

$$\boldsymbol{V}^T \cdot \boldsymbol{C} \cdot \boldsymbol{V} = \frac{1}{2} \mathrm{diag}[\lambda_1^{-2}, \ldots\ldots, \lambda_N^{-2}] = \frac{1}{2} \boldsymbol{\Lambda}^{-2} \quad . \qquad (5.71)$$

and therefore that

$$\boldsymbol{C} = \frac{1}{2} \boldsymbol{V} \cdot \boldsymbol{\Lambda}^{-2} \cdot \boldsymbol{V}^T = \frac{1}{2} \boldsymbol{V} \cdot \boldsymbol{W}^{-1} \cdot \boldsymbol{V}^T \quad . \qquad (5.72)$$

(In (5.66), (5.68) and (5.70) we have used

$$\int_{-\infty}^{\infty} e^{-\lambda^2 x^2} dx = \frac{\sqrt{\pi}}{\lambda} \quad ; \quad \int_{-\infty}^{\infty} x e^{-\lambda^2 x^2} dx = 0 \quad ; \quad \int_{-\infty}^{\infty} x^2 e^{-\lambda^2 x^2} dx = \frac{\sqrt{\pi}}{2\lambda^3} \quad .)$$

We see therefore that the covariance matrix is immediately obtained from the singular value decomposition of the matrix $\boldsymbol{D}$. The diagonal elements of the covariance matrix indicate the uncertainties in the components of the state vector $\boldsymbol{b}_0$ and the off-diagonal elements are, according to (5.61), a measure of how much we can expect two different components to deviate simultaneously from their estimated values. If one of the eigenvalues, say λ_l , is very small and the confidence ellipsoid is hence elongated along the eigenvector $\boldsymbol{v}_l$ and if $\boldsymbol{v}_l$ has large components along the$i$'th and j'th coordinate axis in the parameter space, then we saw

previously that the model parameters b_i and b_j are correlated and that only one of the combinations $b_i - ab_j$ and $b_i + a^{-1}b_j$ is determined with any certainty while the other one is ill determined. But it is easy, under these conditions, to see from (5.72) that the matrix element $C_{i,j}$ of the symmetric covariance matrix is large. We see therefore that the off-diagonal elements of the covariance matrix indicate the correlation between different model parameters. The correlation can be displayed more clearly by considering still another matrix, called the *correlation matrix* $\boldsymbol{K}$ which is obtained by normalizing the covariance matrix such that all the diagonal elements are equal to one, that is to say

$$K_{i,j} = \frac{C_{i,j}}{\sqrt{C_{i,i}C_{j,j}}} \quad . \tag{5.73}$$

It can be shown that if the eigenvector $\boldsymbol{v}_l$ has large components along the i'th and the j'th coordinate axis in the parameter space and the eigenvalue λ_l tends to zero, then $K_{i,j}$ tends to ± 1 . If $K_{i,j}$ tends to 1 then $b_i + a^{-1}b_j$ is well determined and $b_i - ab_j$ is ill determined but if $K_{i,j}$ tends to -1 , then the combination $b_i - ab_j$ is well determined but $b_i + a^{-1}b_j$ ill determined.

We have now studied in some details how we can estimate confidence limits and identify correlation among the components of the state vectors resulting from the least-squares inversion. But one important point has to be stressed, namely that all the discussion has, so far, been based on the probability density expressed in (5.54) and (5.55). This probability density was derived from the more general expression (5.2) by making a second order Taylor expansion of $\chi^2(\boldsymbol{b})$ around $\boldsymbol{b}_0$ and actually more than that, because we omitted the second term in (5.12) which contained the contribution of the second derivative of the model function f to the second derivative of χ^2 with respect to the model parameters. The probability density in (5.54) and (5.55) is therefore, strictly speaking, based on first order linear Taylor expansion of the model function $\chi^2(\boldsymbol{b})$ around $\boldsymbol{b}_0$. The analysis presented in this section does therefore only hold for linear model functions. In the case of non-linear model functions as we are dealing with in transient electro-magnetic soundings, the formulas and concepts presented in this section have to be considered as approximation to the real situation. How good the approximation is depends on how severely the model function is non-linear in the neighbourhood of $\boldsymbol{b}_0$, that is to say how well it can be approximated by a first order Taylor expansion. The non-linearity of the model function can often be reduced, as pointed out by Johansen (1977), by using non-linear, e.g. logarithmic, instead of linear variables for either or both of the response and model parameters and this is in fact done in the non-linear least-squares inversion program described in the next chapter.

A more rigorous determination of the confidence limits of the model parameters can be based on the probability density in (5.2). We can for example try to determine the 68% confidence domain around $\boldsymbol{b}_0$ by applying some clever algorithm to determine the domain in the model parameter space such that $\chi^2(\boldsymbol{b})$ given in (5.3) is less than $\chi^2(\boldsymbol{b}_0) + 1$ and project the

resulting domain onto the parameter axis. The 68% confidence domain can have complicated shape if the model function f is severely non-linear and it is generally not an easy task to find a clever way to map the domain accurately. The benefit of such more advanced methods to determine the confidence limits is often questionable because there might be other and sometimes more serious uncertainties in the problem, such as the question if the model function is an appropriate one. In the case of transient electro-magnetic soundings this refers for example to the question whether it is justifiable to assume horizontal layering. The approximate methods described here are usually sufficient to show which parameters are well defined and which are not and to give rough estimates of the uncertainties involved.

5.5 REVIEW OF THE PROBABILITY DENSITY OF STATE VECTORS AND GENERALIZATION OF THE LEAST-SQUARES INVERSION

Let us conclude this chapter by reviewing the probability density for the possible state vectors. The argumentation leading up to the probability density (5.2) included a lot of assumptions and hand-wavings. The result may therefore seem to be vulnerable. But this is not really the case because there exist a rather general arguments for that the probability density should indeed have the form

$$P(\boldsymbol{b}) = Ce^{-r(\boldsymbol{b})} \tag{5.74}$$

where $r(\boldsymbol{b})$ is a "merit" function that measures the difference between the measured response values and the values calculated from the model function. The chi-square sum in (5.3) is among the simplest and most obvious candidates for such a merit function. The argument goes as follows. We want a probability density for the state vectors such that the expectation value of the merit function

$$\hat{r} = \int r(\boldsymbol{b})P(\boldsymbol{b})\,d\boldsymbol{b} \tag{5.75}$$

is minimized and we further demand that the probability density is normalized

$$\int P(\boldsymbol{b})\,d\boldsymbol{b} = 1 \ . \tag{5.2a}$$

In addition to these conditions we want the probability density to be as random or disordered as possible. Stated differently, since we cannot observe the state directly we want the probability density to be such that our information about the distribution of the possible state vectors is minimal but we want the expectation value of the merit function to be minimum and the density to be normalized. These are very modest demands, but we will show that they imply that the probability density is on the form (5.74).

Before we are able to show this, we need to introduce some new concepts. We have just stated that we want our information about the possible state vectors to be minimum. We obviously have to state more precisely what we mean by this phrase and therefore to define

what we mean by the information content of a probability density. To do that we imagine that we are playing a game with our friend. Let us suppose that we have a big box and close our eyes and throw a ball into the box. Our friend is standing by the box with his eyes open and he tells us that the ball is in the box. What is the information content of this statement? It is zero because we knew that the ball was in there, the probability that the ball is in the box is one. We now divide the box into two subboxes A and B. They may be different in size so that the probabilities that a ball thrown into the box ends up in subbox A or B may be different. Let p be the probability that the ball ends up in A and q the probability that that it ends up in B. We have that $p < 1$ and $q < 1$ and $p + q = 1$. Again we close our eyes and throw the ball into the box and ask our friend to tell us where it landed. He tells us that it landed in box A. This statement must contain some information because we did not know in which subbox the ball was. Let us call this information $I(p)$. If our friend had observed that the ball was in subbox B, we would have got the information $I(q)$. We have anticipated that the information content of the observed result is a function of its probability. This is reasonable because if the subbox B is much smaller than A and hence $q << p$, then we would have expected that the ball landed in A but would be surprised if our friend told us that it was indeed in B. We have much more information about the spatial position of the ball if we know that it is in subbox B but not in A because B is much smaller. We see therefore that $q < p$ implies $I(q) > I(p)$ and we have previously argued that $I(1) = 0$.

Let us now further subdivide A into two subboxes A_1 and A_2 and B into B_1 and B_2. Let the probabilities for that the ball ends up in A_1 or A_2, given that it is in A, be p_1 and p_2 respectively and probabilities that it ends up in B_1 or B_2, given that it is in B, be q_1 and q_2 respectively. The probability that the ball ends up in A_1 is therefore $p \cdot p_1$ and that it ends up in B_2 is $q \cdot q_2$ and so on. Once again we close our eyes and throw the ball and ask our friend where it is. He is not aware that we had subdivided A and B and tells us that the ball is in B, delivering the information $I(q)$. We point out to him that A and B are now subdivided and ask him to look closer and give us further information. He delivers the additional information $I(q_1)$ by saying that it is in B_1. The total amount of information that we have obtained is $I(q) + I(q_1)$ but this must be the same information $I(q \cdot q_1)$ that we would have got if our friend had told us at once that the ball was in B_1.

Let us summarize and state that we expect the information I obtained from an observation resulting in the observed value x_i to be a function of the probability p_i of observing the particular result x_i and we demand that

$$I(p_i) > I(p_j) \quad \text{if} \quad p_i < p_j \quad ; \quad I(1) = 0$$

$$I(p_i \cdot p_j) = I(p_i) + I(p_j) \tag{5.76}$$

where the probabilities p_i satisfy

$$0 \leq p_i \leq 1 \quad \text{and} \quad \sum_i p_i = 1 \ . \tag{5.77}$$

From these properties we immediately see that the information function must be given as

$$I(p_i) = -k \ln(p_i) \tag{5.78}$$

where k is a constant. Now we can ask the question how much information we expect to get by making an observation. The information we get by observing the value x_i is $-k \ln(p_i)$ and the expected information value S will be the weighted average of all the possible information values, weighted by the corresponding probabilities, that is

$$S = -k \sum_i \ln(p_i)\, p_i \ . \tag{5.79}$$

Those familiar with statistical mechanics will immediately recognize (5.79) as the expression for the entropy of an ensemble and recall that it is a measure of the disorder in the ensemble. If a system is highly disordered then we have little knowledge about its state and would therefore gain much information by determining its actual state x_i . But if we can not do so, then we can turn things around and interpret (5.79) as a measure of our lack of information about the system because we can not determine its state directly.

We have now established a measure of our lack of knowledge about the state of the system and can return to the problem of determining the probability density for the possible state vectors. We want to determine the probability density $P(\boldsymbol{b})$ which satisfies the demands of maximizing the lack of knowledge

$$S = -\int \ln(P(\boldsymbol{b}))P(\boldsymbol{b})\, d\boldsymbol{b} \tag{5.80}$$

minimizing the expectation value of the "merit" function

$$\hat{r} = \int r(\boldsymbol{b})P(\boldsymbol{b})\, d\boldsymbol{b} \tag{5.75}$$

and satisfies the normalization condition

$$\int P(\boldsymbol{b})\, d\boldsymbol{b} = 1 \ . \tag{5.2a}$$

To achieve this we make use of the method of Lagrange multipliers and maximize the following integral

$$F = \int [-\ln(P(\boldsymbol{b}'))P(\boldsymbol{b}') - r(\boldsymbol{b}')P(\boldsymbol{b}') + aP(\boldsymbol{b}')]\, d\boldsymbol{b}' \tag{5.81}$$

with respect to P and thereafter determine the Lagrange multiplier a so that (5.2a) is satisfied. To maximize F with respect to P we take the functional derivative with respect to

$P(\boldsymbol{b})$ and demand that the result is zero, that is

$$\frac{DF}{DP(\boldsymbol{b})} = -\ln(P(\boldsymbol{b})) - 1 - r(\boldsymbol{b}) + a = 0 \ . \tag{5.82}$$

For those who are not familiar with functional derivatives, we point out that we can think of the integral in (5.81) as the limit of a sum of the integrand evaluated at discrete points in the state vector space and the functional derivative is obtained by differentiating the sum with respect to $P(\boldsymbol{b})$. It is now trivial to solve (5.82) and the solution is

$$P(\boldsymbol{b}) = e^{a-1}e^{-r(\boldsymbol{b})} = Ce^{-r(\boldsymbol{b})} \tag{5.83}$$

which is the promised result stated in (5.74). The normalization constant C and hence the Lagrange multiplier a is determined by using the normalization condition (5.2a). We have thus seen that a probability density for the state vectors of the form (5.83) is of much more general validity than we would have expected from the hand-waving argumentation in section 5.1 . The central question in the inversion problem is therefore not about the form of the probability density for the solutions but about the merit function. In section 5.1 we saw that the assumption of a normal distribution of the noise in the measured responses led us naturally to merit function given by the chi-square sum (5.3). If the noise has some other statistical distribution we should use a different merit function.

It is often observed that using the chi-square sum as a merit function, causes the inversion to take too much notice of response values which are scattered far off from the rest. That is to say, if most of the measured responses are close to being on a smooth curve, but a few points are far off, then an inversion with the chi-square sum as a merit function takes too much notice of the far off points. The resulting state vector and the corresponding fitted response curve compromises between the smooth curve, on which most of the measured responses lay, and the far off points. The reason for this is of course that the noise in the measured data is not normally distributed and the normal distribution underestimates the possibility of far off points. In such cases the merit function should not be taken to be the chi-square sum but some other function which does not increase as steeply for large differences between measured and fitted response values as the chi-square sum does. Such an approach is sometimes referred to as a robust estimation.

From the definition of the chi-square sum

$$\chi^2(\boldsymbol{b}) = \sum_{i=1}^{M} (y_{0i} - f_i(\boldsymbol{b}))^2/\sigma_i^2 \tag{5.3}$$

we see that the standard deviations σ_i can be considered as weight factors for the individual response values, because the lower the σ_i is, the larger is the contribution of corresponding response value to the sum. If the measured response vector $\boldsymbol{y}_0$ is the average of many repeated measurements, then it is natural to take σ_i to be the standard deviation of the

repeated measurements of the i'th component of the response vector. But if we have some preassumed ideas of how the different response values should be weighted in the inversion or if we want to put more emphasis on some of the data points than others, we can do that by manipulating the σ_i's in the chi-square sum. A use is made of this possibility in the least-squares inversion program described in the next chapter.

6. ONE DIMENSIONAL INVERSION PROGRAM FOR CENTRAL LOOP TRANSIENT ELECTRO-MAGNETIC SOUNDINGS

We have now studied the principles of transient electro-magnetic soundings over horizontally stratified earth. In chapters 2 and 3 we developed the fundamental equations describing the transient response for the central loop and grounded-dipole-coil configurations. Chapter 4 describes an efficient method for evaluating numerically the integrals which have to be calculated in the forward problem of determining the response of a given layered model and chapter 5, which is devoted to the inverse problem, describes the iterative non-linear least-squares inversion method. We are now, at last, ready to apply all this machinery that we have been piling up.

This chapter describes the program TINV (Transient electro-magnetic INVersion) for one-dimensional inversion of central loop transient electro-magnetic resistivity soundings. TINV is an implementation of the principles developed in chapters 2 through 5. It is a non-linear least-squares inversion program using a Levenberg-Marquardt type of inversion algorithm together with a fast forward routine based on the digital filter method.

The inversion program assumes that the apparent resistivity data is collected with equipment where the current in the transmitter loop is turned off linearly from its maximum value to zero and that the time values, at which the apparent resistivity values are given, are measured in seconds after the current has become zero (t = 0 is at the end of the turn-off ramp). The program also assumes that the data is collected with a circular transmitter loop. If this is not the case the actual transmitter loop is simulated by a circular loop having the same area.

The program reads measured apparent resistivity data from an input file and prompts for an initial guess for a layered resistivity model. It iteratively adjusts the resistivity model to minimize the difference between the measured and the calculated apparent resistivity values. The results are written into two files, an output list file that can be typed on the screen and printed as a hard copy and an output plot file that can be plotted both on the screen and as a hard copy on a printer or a plotter.

The source code consists of 16 modules, the main program TINV.FOR and 15 subroutines. The modules are listed in Table 1. (In this chapter we will use capital letters to write program names and variable names which refer explicitly to the program). The program is written in standard FORTRAN 77 and can be made to run on any computer supporting FORTRAN 77. The run file TINV.EXE, referred to in section 6.2, was made by compiling the source files by the Microsoft FORTRAN Optimizing Compiler, version 4.01, and linking by the Microsoft Object Linker and can be run on IBM PC or PC-compatible computers with an 8087 coprocessor. The plotting utilities, described in section 6.3.2, use plot programs from Golden Software, Inc. and can only run on IBM PC and PC-compatibles.

MODULE	FUNCTION
TINV	Main program; inversion algorithm
TRDT	Reads input data from file and terminal
J1FLT	Stores J1 Hankel transform filter
COSFLT	Stores cosine transform filter
TFWD	Forward algorithm; calculates apparent resistivity curve from a given model
YCAL	Calculates frequency domain response for a given frequency
RECUR	Calculates kernel function for frequency domain response
SPLINE	Calculates second derivatives of frequency domain responses
SPLINT	Performs cubic spline interpolation between calculated frequency domain responses
TRF	Integrates sharp step response from t to t+TOFF to get ramped step response
CHSQ	Calculates chi-square sum
TDR	Calculates partial derivative matrix
SVDC	Performs a singular value decomposition on the partial derivative matrix
ORDW	Orders eigenvalues in increasing order
NEWP	Calculates increments to be added to model parameters to get a new model
WROT	Writes out results into output files

Table 6.1 The modules of the program TINV

A slightly modified version of the forward routine of the inversion program has been made into a separate program called TEM. TEM calculates late time apparent resistivity curve corresponding to a one-dimensional resistivity model, read from the terminal and for time values equally distributed on log-scale with five points per power of *e*, over an interval specified by the user. The results are written into an output file (that has the right format for an input file for TINV) and a plot file that can be plotted in the same way as the output plot files from the inversion program TINV. The program TEM is a by-product of the inversion program, meant for simple model calculations, and will not be discussed further.

6.1 THE STRUCTURE OF THE PROGRAM

The backbone of the program TINV is a general non-linear least-squares inversion algorithm of the Levenberg-Marquardt type described in chapter 5. The inversion algorithm is supplemented by routines for data input, TRDT, and output, WROT, a forward routine, TFWD, that calculates response values from a given layered resistivity model and a routine, TDR, that calculates the partial derivatives of the response values with respect to the model parameters. The general structure of the program is shown on Figure 6.1.

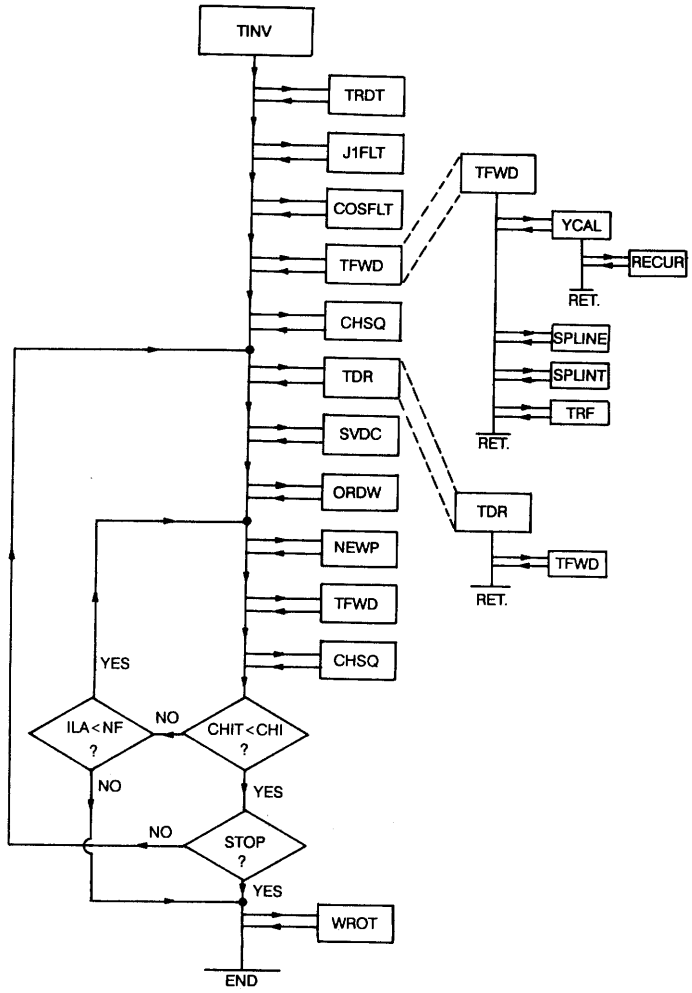

Figure 6.1 Structure of inversion program

The least-squares algorithm, the forward routine and the partial derivative routine will now be discussed briefly.

6.1.1 THE INVERSION ALGORITHM

The program TINV, like most inversion programs, works in such a way that it reads the measured data points (apparent resistivity curve) and prompts for a starting model. The interpreter guesses, by visual inspection of the data curve, the number of layers and initial model parameters i.e. the resistivity values and thicknesses of the layers. Each model parameter can either be taken to be a free or a fixed parameter. The program iteratively adjusts the values of the free model parameters to get the best fit, in the chi-square sense, between the measured curve and the curve calculated from the model. It is, in most cases, necessary to try models with different numbers of layers to find the model that best fits the data. It should also be kept in mind that the model resulting from the iterative inversion can depend on the initial guess. A poor initial guess can possibly lead the inversion process astray even though this has never been observed.

In the inversion algorithm, all computations are done with the data and model parameters on logarithmic form. The measured response values at the abscissa values $x_i = \ln(t_i)$ are $y_{0i} = \ln(\rho_{am}(t_i))$ and the model parameters are kept as $b_j = \ln(p_j)$ where p_j stand for the resistivity values and layer thicknesses. This is done because the non-linearity in the dependence of the response values on the model parameters is not as severe in the logarithmic as in the linear representation. The logarithmic representation furthermore prevents the occurrence of non-physical negative model parameters.

The quality of the fit between the measured and the calculated apparent resistivity values is measured by the chi-square sum which is calculated by the subroutine CHSQ. CHI, the square root of the chi-square sum, is given in terms of the natural logarithms of the measured and calculated apparent resistivity values, by the following formula:

$$\mathrm{CHI} = \left[\frac{1}{\mathrm{ND}} \sum_{1}^{\mathrm{ND}} [(\ln(\rho_{am}) - \ln(\rho_{ac}))\cdot \mathrm{WPM}]^2 \right]^{1/2} \qquad (6.1)$$

where ND is the number of data points and WPM are a weight factors (to be discussed later). The lower the CHI is, the better is the fit. The weight factors are normalized such that $\sum_{1}^{\mathrm{ND}} \mathrm{WPM} = \mathrm{ND}$ and if CHI is low, say lower than 0.1, then it can be interpreted as the weighted average fractional difference between the measured and the calculated apparent resistivity values.

The program always keeps the best model obtained and each iteration cycle starts with the determination of a temporary model to be tried next. To determine the temporary model, the partial derivatives of the calculated response values $f_i(\boldsymbol{b}) = \ln(\rho_{ac}(t_i,\boldsymbol{b}))$, with respect to the free model parameters, is calculated by the subroutine TDR. The partial derivative matrix $\boldsymbol{A}$ is decomposed, by the singular value decomposition routine SVDC, into a product of a data

eigenvector matrix U, a diagonal eigenvalue matrix Λ and the transpose of an orthogonal parameter eigenvector matrix V:

$$A = U \cdot L \cdot V^T \quad . \tag{5.39}$$

The increments δb to be added to the model parameter vector b in order to get the temporary model to be tried next, are obtained by solving the equation

$$A^T \cdot c = A^T \cdot A \cdot \delta b \tag{5.10a}$$

where

$$c = \left[\frac{y_{0i} - f_i(b)}{\sigma_i} \right] \tag{5.9a}$$

and σ_i^{-1} is equal to the square root of the weight parameter WPM for the i'th response value. Equation (5.10a) is solved for δb in the subroutine NEWP by applying the damped inverse of A and is given as

$$\delta b = V \cdot \Lambda_\alpha^{-1} \cdot U^T \cdot c \tag{5.52}$$

The Marquardt parameter α, which is denoted by XLA in the program, is taken to be equal to one of the eigenvalues which have been ordered in an increasing order in the subroutine ORDW. The ordered eigenvalues are numbered by the index ILA which runs from 1 to NF, the number of free parameters in the model.

The first temporary model tried in each iteration step is obtained by adding the increments resulting from the damped inversion using the smallest eigenvalue (but not smaller than 0.01, the index ILA shows which eigenvalue is used) as a Marquardt parameter. If the square root of the chi-square sum for the temporary model, CHIT, is lower than CHI then the temporary model is kept as the best model and a new iteration cycle is started. If CHIT is higher than CHI, another temporary model, obtained by inversion of equation (5.10a), with increased damping by a larger eigenvalue (ILA increased by one), is tried. This is continued until CHIT becomes lower than CHI or ILA gets higher than NF, in which case the iteration process is terminated.

There are four stop checks in the program. The iteration process terminates if:

a) The parameter ILA gets higher than NF. This is called no-convergence. If the program terminates on this stop check it does not necessarily mean that an acceptable fit to the measured data was not obtained. It simply states that the program could not further improve the fit.

b) The average fractional difference, CHI, between the measured and the calculated apparent resistivity values becomes lower than 10^{-3} i.e. an average difference less than 0.1%. This is called CHI-convergence.

c) The fractional decrease in CHI in the last iteration is less than 10^{-5}. This is called DCHI-convergence.
d) The program has performed the number of iterations that the operator asked for and he does not wish to perform more iterations. This stop check is called maximum of iterations.

A CHI-convergence is seldom obtained in inversion of real measured data and the iteration is usually terminated on no-convergence or DCHI-convergence (no further improvement) or on maximum of iterations.

The forward routine TFWD calculates the late time apparent resistivity values at time values equally distributed on logarithmic scale with five points per power of *e*. The inversion algorithm demands therefore that the data have this distribution. If the measured data do not fulfill this condition, an interpolation, on log-log scale, is performed (in the data input routine TRDT) in order to get data points, equally spaced in logarithm of time, with five points per power of *e*.

A different weight can be given to the data points in the inversion. The weighing is controlled by the weight parameter RW ($-1 \leq RW \leq 1$). Using RW=1, the data points are weighted with WPM proportional to $\ln(\rho_{am})$. This causes data points with higher apparent resistivity values to have more weight in the inversion. For WR=-1 the data points are weighted with WPM inversely proportional to $\ln(\rho_{am})$ giving the lower apparent resistivity values the higher weight. For WR in between -1 and 1 the data points are weighed with WPM proportional to $(\ln(\rho_{am}))^{WR}$ giving weights in between the above extreme cases. For WR=0 all data points have the same weight. The weight factors are normalized such that the sum of WPM is equal to ND, the number of data points. The different modes of weighings can be used to put emphasis on different parts of the measured data curve but normally it is recommended to give all data points similar weight and hence to take WR close to 0, giving WPM ≈ 1.

6.1.2 THE FORWARD ALGORITHM

The forward algorithm TFWD calculates late time apparent resistivity values for a given horizontally layered resistivity model, a given transmitter loop radius RP and turn-off time TOFF. The turn-off time is the time it takes the transmitter to turn the current in the transmitter loop linearly off from its maximum value to zero. The calculations are done in four steps. The first step is to calculate the frequency domain response i.e. the induced voltage in a receiver coil at the centre of a circular transmitter loop into which an alternating current, $I_0 e^{i\omega t}$, is transmitted. The next step is to calculate, from the frequency domain response, the transient voltage response due to a sharp current step in the transmitter loop. The third step is to take into account the turn-off time and calculate the response of a ramped step function with the ramp length TOFF. The fourth and final step is to calculate the late time apparent resistivity from the ramped step response.

The frequency domain response which is a function of the frequency ω, the transmitted current and the radius r (=RP) of the transmitter loop is calculated in the subroutine YCAL, according to the following formulas :

$$V(\omega,r) = (-i\omega)\,\Phi(\omega,r)\,I_0 e^{i\omega t} \quad ; \quad \Phi(\omega,r) = \int_0^\infty K(\omega,\lambda)\,J_1(\lambda r)\,d\lambda \tag{6.2}$$

(see equations (3.8) and (3.16)) . The kernel function K, which is calculated in the subroutine RECUR, is given as

$$K(\omega,\lambda) = \frac{A_s A_r \mu_0}{\pi r}\,\frac{\lambda^2}{m_0}\,\frac{S_0}{S_0 - T_0} \tag{6.3}$$

where A_s and A_r are the effective areas of the transmitter and receiver loops respectively (are times number of turns), m_0 is given by (2.36) and (2.69) and S_0 and T_0 are given in terms of the conductivities and thicknesses of the resistivity layers by the recursion relations (2.75). The Hankel transform integral is calculated by using the digital filter shown on Figure 4.1 in chapter 4. The filter is transferred through a common block from the subroutine J1FLT which is called by the main program.

The frequency domain response has typically to be determined at about 100 frequency values equally spaced on log-scale with five points per power of e. It is very time consuming to calculate each frequency domain response value by numerical integration. Fortunately it is a well behaved function of the frequency (see Figures 3.4, 3.6 and 3.7 in chapter 3). It is therefore, in most cases, sufficient to calculate the frequency domain response in the above described manner at every 2nd to 4th point. The response values at the intermediate points can then be determined by a spline between the calculated values. Use is made of this in the forward routine TFWD in order to reduce the running time of the program. The frequency domain response is calculated by a numerical integration, in the subroutine YCAL, at every MMth frequency value where MM is an integer supplied by the user. The response at the intermediate frequency values is obtained by a cubic spline performed by the subroutines SPLINE and SPLINT. How densely the response has to be calculated by numerical integration depends on the resistivity model and the radius of the transmitter loop. The higher the resistivity is, the smaller the MM must be. It is not recommended to take MM to be higher than 4, unless the model contains a thick layer of very low resistivity (less than 1 Ωm) at a depth not greater than about three times the transmitter loop radius. If MM is taken too high, the calculated late time apparent resistivity will show oscillations at late time values. If this is observed, MM has to be decreased.

The next step is to calculate the time domain response due to a sharp step function transmitted into the transmitter loop. The response is given as

$$V_{step}(r,t) = \frac{2I}{\pi} \int_0^\infty \mathrm{Re}\,\Phi(r,\omega)\cos(\omega t)\,d\omega \qquad (3.21a)$$

where I is the transmitted current strength and Φ is the Hankel transform integral for the frequency domain response in (6.2). The cosine transform integral is turned into a convolution integral by changing to logarithmic variables and the integration is performed numerically by using the digital filter shown on Figure 4.2 . The cosine transform filter is transferred through a common block from the subroutine COSFLT which is called by the main program. The forward routine TFWD calculates the time domain response for time values equally spaced on log-scale, for five values per power of e, by shifting the kernel function relative to the filter and performing the convolution summation as discussed in chapter 4.

The third step is to take into account the fact that the current turn-off is not described by a sharp step but rather by a ramped step function with the ramp length given by the turn-off time TOFF. The ramped step response is obtained by convolving the derivative of the current function with the sharp step response. This results in the integral:

$$V(r,t) = \frac{1}{TOFF} \int_t^{t+TOFF} V_{step}(r,\tau)\,d\tau \qquad (3.31)$$

which is calculated numerically in the subroutine TRF.

The fourth and final step in the forward algorithm is to calculate the late time apparent resistivity values from the ramped step response. This is done at the end of the routine TFWD, and according to the formula:

$$\rho_a(r,t) = \frac{\mu_0}{4\pi}\left[\frac{2\mu_0 A_s A_r I}{5t^{5/2}V(r,t)}\right]^{2/3} \qquad (3.57)$$

where A_s and A_r are the effective areas of the transmitter and receiver loops, respectively and I is the current in the transmitter loop.

6.1.3 THE PARTIAL DERIVATIVE ALGORITHM

The partial derivatives of the logarithms of the apparent resistivity values with respect to the free logarithmic model parameters are calculated by the subroutine TDR. A small increment is added to the free logarithmic parameters, one at the time, and the corresponding logarithmic apparent resistivity values calculated by the forward routine TFWD. The derivatives are then calculated as the difference between the logarithmic apparent resistivity values for the shifted and unshifted models and divided by the increment. The result is returned as the partial derivative matrix $\boldsymbol{A}$.

6.2 RUNNING THE PROGRAM

In order to run the inversion program, the file TINV.EXE must either be in the working directory or a directory specified in the PATH-list. The measured apparent resistivity curve is read from an input file. The input file must have one data point in each line. A data point consists of a pair of numbers separated by a comma (,), a space(s) or a tab. The first number is a time value t in seconds (t=0 is at the end of the turn-off ramp) and the second number is the corresponding measured late time apparent resistivity value ρ_a in Ωm.

The following is an example of an input data file:

```
0.89000E-04  526.9
0.10900E-03  422.8
0.14000E-03  320.7
0.17700E-03  249.8
0.21900E-03  199.6
0.28000E-03  157.6
0.35500E-03  124.3
0.44300E-03  101.8
0.56300E-03   82.0
0.71200E-03   67.7
0.87600E-03   55.7
0.10900E-02   47.2
0.14000E-02   39.1
0.17700E-02   33.3
0.21900E-02   29.1
0.28000E-02   24.6
0.35500E-02   21.1
0.44300E-02   18.4
0.56300E-02   16.0
0.71200E-02   14.1
0.87600E-02   12.4
0.10900E-01   11.2
0.14000E-01   10.1
0.17650E-01    9.5
0.21950E-01    9.0
0.28000E-01    8.5
0.35450E-01    8.2
0.44250E-01    7.9
0.56300E-01    7.9
0.71200E-01    8.1
```

When a data file containing the measured apparent resistivity curve has been created and having the above described format, the inversion program is run by going through the following steps. (All user's responses discussed below are to be followed by striking the return key):

1. The inversion program is started by typing TINV at the system prompt.

2. The program prompts for input file name. This is answered by typing the name of the file containing the measured data curve to be inverted.

3. The program asks for the weight parameter RW discussed above. This is answered by typing a number in the interval -1 to 1. It is generally recommended to use RW close to 0.

4. The program asks for the turn-off time TOFF. This is answered by typing the turn-off time in seconds.

5. The program asks for the effective radius RP of the transmitter loop. If the transmitter loop used for collecting the data was not circular, the radius of a circular loop with the same area is typed.

6. The program asks how densely the frequency domain response is to be calculated by numerical integration. This is answered by typing an integer for the parameter MM, causing the frequency domain response to be calculated at every MMth frequency value and be determined by cubic spline at intermediate frequency values. It is normally not recommended to take MM to be higher than 4 (see discussion above).

7. The program asks for the guessed initial model. It asks for the number of layers which is answered by typing the guessed number of layers. Next it asks for the model parameters for each layer (resistivity in Ohmm and thickness in m). Each parameter can either be a free parameter to be adjusted by the program or a fixed parameter not to be adjusted. For a free parameter the guessed value is typed. For parameters to be held fixed the corresponding parameter values are followed by a comma and an asterisk (e.g. 235,*).

8. The program prompts for the number of iterations to be performed. This is answered by typing the desired number of iterations.

9. Finally the program prompts for names of output list file and an output plot file into which the results from the iterative inversion are written. When these file names have been given, the program starts the iteration process.

During the iterations the program writes on the terminal the iteration number ITR, the parameter ILA telling which of the ordered eigenvalues is used as Marquardt parameter and its value XLA. It writes the best model parameters obtained so far (resistivity values, rho and layer thicknesses, d) and the temporary model parameters (rhot and dt) to be tried next. It also writes the lowest CHI-value obtained and the temporary CHIT obtained from the temporary model. If CHIT is lower than CHI then the temporary model is kept as the best model and the program proceeds to the next iteration with a new temporary model. If CHIT is higher than CHI, another temporary model is tried. This is repeated until CHIT becomes lower than CHI or ILA becomes higher than NF, the number of free parameters in the model, and the program terminates on the criterion of no-convergence. In addition to the above listed parameters, the program also writes on the terminal, each time the forward

routine TFWD is called, two numbers indicating the lowest and the highest frequency values for which the frequency domain response is calculated. These values are written merely for psychological reasons so that the operator can see that the program working. The minimum value should be considerably higher than 1 and the maximum value lower than 283.

If the iteration process has not stopped on the no-convergence, the CHI-convergence or the DCHI-convergence stop checks (see above) and the number of iterations specified (in step 8 above) has been performed, the program pauses and asks if the iteration process is to be continued. If this question is answered with N (no), the iteration process is terminated on maximum number of iterations. If it is answered with Y (yes), the program asks how many more iterations are to be performed.

6.3 OUTPUT FILES

The results from the inversion are written into two output files, the output list file and the output plot file. The content of these files will now be discussed shortly.

6.3.1 THE OUTPUT LIST FILE

The following is an example of an output list file:

```
 EFFECTIVE LOOP RADIUS, RP=169.3 m
 TURN-OFF TIME, TOFF=.240E-03 s
 WEIGHT PARAMETER, RW=  .00


INITIAL MODEL PARAMETERS:
rho: 1000.00    50.00     2.00     8.00
  d:  100.00    50.00   100.00
CHI= .49883E+00


ITR= 1  ILA= 2  CHI= .3910E+00  DCHI= .2161E+00
rho:  413.58     1.84     8.22    11.30
  d:  144.76    15.50   466.19


ITR= 2  ILA= 3  CHI= .1215E+00  DCHI= .6894E+00
rho:   86.66     6.66     4.34    14.42
  d:  112.20    19.67   145.75


ITR= 3  ILA= 2  CHI= .1060E+00  DCHI= .1270E+00
rho:  109.27    12.58     6.07     9.80
  d:  100.70    57.55   285.70


ITR= 4  ILA= 1  CHI= .1594E-01  DCHI= .8497E+00
rho:  128.61     9.30     4.88    12.17
  d:   98.65    67.27   290.69


ITR= 5  ILA= 1  CHI= .1118E-01  DCHI= .2988E+00
rho:  132.33     9.48     4.79    12.56
  d:   98.61    68.17   258.00


ITR= 6  ILA= 1  CHI= .1109E-01  DCHI= .7392E-02
rho:  132.48     9.45     4.76    12.45
  d:   98.64    68.83   255.69


*** THE PROGRAM TERMINATED  AFTER  7 ITRERATIONS ***


ISTOP=2 * MAX ITERATIONS *
```

FINAL CHI IS CHI= .1109E-01

THE FINAL MODEL PARAMETERS ARE:

rho: 132.26 9.43 4.76 12.39
d: 98.72 68.98 254.65

DATA EIGENVECTORS:

1	-.095	.061	-.421	.282	-.534	.450	-.064
2	-.135	.072	-.367	.212	-.228	.057	.014
3	-.167	.081	-.315	.143	.011	-.215	.063
4	-.188	.087	-.275	.084	.141	-.311	.071
5	-.198	.092	-.242	.033	.181	-.265	.046
6	-.202	.095	-.207	-.017	.181	-.158	.010
7	-.206	.099	-.166	-.069	.179	-.053	-.023
8	-.212	.105	-.117	-.123	.184	.026	-.046
9	-.218	.112	-.065	-.174	.178	.099	-.062
10	-.220	.117	-.016	-.210	.136	.187	-.075
11	-.220	.119	.031	-.226	.075	.252	-.076
12	-.218	.118	.075	-.218	.010	.258	-.055
13	-.215	.112	.110	-.185	-.058	.208	-.015
14	-.213	.103	.139	-.134	-.117	.104	.040
15	-.210	.090	.162	-.078	-.166	-.012	.096
16	-.205	.074	.176	-.028	-.205	-.092	.131
17	-.198	.056	.183	.009	-.222	-.131	.139
18	-.192	.035	.181	.040	-.209	-.156	.125
19	-.187	.012	.174	.072	-.169	-.180	.092
20	-.180	-.010	.164	.109	-.127	-.174	.032
21	-.171	-.031	.155	.154	-.085	-.151	-.046
22	-.163	-.051	.147	.194	-.031	-.103	-.140
23	-.156	-.071	.136	.221	.036	-.030	-.229
24	-.148	-.095	.120	.241	.094	.036	-.272
25	-.139	-.124	.098	.246	.138	.088	-.243
26	-.132	-.154	.075	.231	.164	.124	-.144
27	-.125	-.185	.053	.199	.170	.150	-.005
28	-.118	-.217	.027	.153	.164	.163	.133
29	-.111	-.247	-.001	.099	.147	.166	.218
30	-.104	-.276	-.028	.033	.106	.138	.264
31	-.098	-.302	-.052	-.048	.031	.068	.289
32	-.092	-.326	-.077	-.134	-.048	-.007	.244
33	-.086	-.344	-.102	-.217	-.117	-.070	.132
34	-.081	-.355	-.119	-.270	-.162	-.123	-.127
35	-.074	-.355	-.124	-.283	-.187	-.178	-.593
	1	2	3	4	5	6	7

PARAMETER EIGENVECTORS:

1	-.140	.104	-.590	.278	-.577	.458	-.033
2	-.246	-.083	.651	-.305	-.326	.553	-.055
3	-.140	-.935	-.157	.108	.205	.156	-.062
4	-.003	-.056	-.056	-.217	-.237	-.292	-.897
5	-.947	.160	-.079	-.003	.179	-.198	.016
6	-.058	-.161	.400	.644	-.455	-.425	.087
7	.022	.232	.184	.596	.472	.394	-.423
	1	2	3	4	5	6	7

PARAMETER EIGENVALUES:

.446E+01 .224E+01 .101E+01 .591E+00 .322E+00 .201E+00 .700E-01

COVARIANCE MATRIX:

1	.46E+01						
2	.39E+01	.50E+01					
3	.60E+00	.10E+01	.10E+01				
4	.19E+01	.35E+01	.49E+01	.84E+02			
5	-.17E+01	-.18E+01	-.32E+00	-.99E+00	.70E+00		
6	-.13E+01	-.28E+01	-.17E+01	-.62E+01	.78E+00	.47E+01	
7	.25E+01	.41E+01	.40E+01	.37E+02	-.13E+01	-.63E+01	.22E+02
	1	2	3	4	5	6	7

CORRELATION MATRIX:

1	1.000						
2	.821	1.000					
3	.277	.449	1.000				
4	.098	.171	.529	1.000			
5	-.923	-.942	-.376	-.130	1.000		
6	-.278	-.588	-.799	-.311	.429	1.000	
7	.251	.396	.844	.857	-.326	-.623	1.000
	1	2	3	4	5	6	7

I	T	Rhoam	Rhoac	WPM
1	.82724E-04	570.44	565.67	1.00
2	.10104E-03	459.09	460.93	1.00
3	.12341E-03	368.63	373.34	1.00
4	.15073E-03	296.43	300.71	1.00
5	.18411E-03	239.65	241.32	1.00
6	.22487E-03	194.59	193.86	1.00
7	.27465E-03	160.55	157.31	1.00
8	.33546E-03	131.54	129.61	1.00
9	.40973E-03	109.22	108.11	1.00
10	.50045E-03	91.19	90.85	1.00
11	.61125E-03	76.68	76.68	1.00
12	.74659E-03	64.74	64.85	1.00
13	.91188E-03	54.03	55.03	1.00
14	.11138E-02	46.44	47.11	1.00
15	.13604E-02	39.95	40.57	1.00
16	.16616E-02	34.77	35.00	1.00
17	.20294E-02	30.54	30.33	1.00
18	.24788E-02	26.74	26.41	1.00
19	.30276E-02	23.39	23.13	1.00
20	.36979E-02	20.57	20.37	1.00
21	.45166E-02	18.19	18.04	1.00
22	.55166E-02	16.19	16.09	1.00
23	.67379E-02	14.52	14.43	1.00
24	.82297E-02	12.89	13.03	1.00
25	.10052E-01	11.63	11.83	1.00
26	.12277E-01	10.66	10.82	1.00
27	.14996E-01	9.92	9.99	1.00
28	.18316E-01	9.41	9.33	1.00
29	.22371E-01	8.96	8.82	1.00
30	.27324E-01	8.55	8.47	1.00
31	.33373E-01	8.28	8.23	1.00
32	.40762E-01	8.01	8.10	1.00
33	.49787E-01	7.90	8.03	1.00
34	.60810E-01	7.97	8.00	1.00
35	.74274E-01	8.14	8.04	1.00

The program begins by writing the effective transmitter loop radius RP, the turn-off time TOFF and the weight parameter RW, as well as the initial model and the corresponding

square root of the chi-square sum CHI. During the iteration process the program writes for each iteration the iteration number ITR, the Marquardt parameter counter ILA, the square root of the chi-square sum CHI for the model obtained in the present iteration, the fractional decrease DCHI of CHI between the present and the last iteration and the model obtained in the present iteration step. Model parameters that are held fixed are followed by an asterisk. When the iteration process is finished the program writes the number of iterations performed and the stop check on which the program stopped. Then it writes the final CHI and the best model obtained.

Next the program writes out information on how changes in the model parameters affect the calculated apparent resistivity values. This is described by the partial derivative matrix which has been decomposed by singular value decomposition into $\boldsymbol{A} = \boldsymbol{U}\cdot\boldsymbol{\Lambda}\cdot\boldsymbol{V}^T$. $\boldsymbol{U}$ is an NDxNF matrix whose column vectors are the data eigenvectors listed (as columns) in the output list file (ND and NF are the number of data points and free model parameters). $\boldsymbol{\Lambda}$ is an NFxNF diagonal matrix whose diagonal elements are the parameter eigenvalues written in the list file. $\boldsymbol{V}$ is an NFxNF orthogonal matrix whose column vectors are the listed parameter eigenvectors (as columns).

If no parameter is fixed, the first NL components of NF-dimensional vector $\boldsymbol{b}$ are the natural logarithms of the resistivity values of the layers (NL is the number of layers in the model, NF=2·NL-1 if no parameter is fixed). The remaining NL-1 components are the natural logarithms of the layer thicknesses. Fixed parameters are not included in the vector $\boldsymbol{b}$. For example, if we take a three layered model and fix the resistivity of the second layer then NF=4 and b_1 and b_2 are the natural logarithms of the resistivity values of the first and the third layer and b_3 and b_4 are the natural logarithms of the thicknesses of the first and the second layer. If we change the model vector $\boldsymbol{b}$ by a small amount $\delta\boldsymbol{b}$ then the calculated ND-dimensional response vector $\boldsymbol{f}$ (the natural logarithms of the apparent resistivity values) will be changed by $\delta\boldsymbol{f}$ according to the equation

$$\delta\boldsymbol{f} = \boldsymbol{U}\cdot\boldsymbol{\Lambda}\cdot\boldsymbol{V}^T\cdot\delta\boldsymbol{b} \tag{6.4}$$

From this equation the significance of the data and parameter eigenvectors and the eigenvalues can be deduced. In the first place we see that the increments $\delta\boldsymbol{f}$ are proportional to the eigenvalues. The first parameter eigenvector shows which model parameters have the strongest association with the first eigenvalue which is normally the highest one. The degree of association is shown by the absolute value of the components of the eigenvector, the higher the absolute value the greater the contribution of the corresponding model parameter. Likewise the eigenvector corresponding to the smallest eigenvalue shows which model parameters have the least influence on the calculated apparent resistivity. The model parameters that are most strongly associated to the highest eigenvalue are the most reliable ones whereas the parameters associated to the smallest eigenvalue are the most uncertain parameters in the final model as discussed in section 5.4 . The relative contribution of the

parameter eigenvectors to the different components of the response vector is described by the data eigenvectors.

In the four layer example above, it is seen that the thickness of the first layer is the best determined model parameter because the eigenvector corresponding to the highest eigenvalue has the absolute value of the fifth component close to one (0.947) while the other components are relatively small. The resistivity value of the third layer is also well determined because the eigenvector corresponding to the second highest eigenvalue has the absolute value of the third component close to one (0.935). It is also seen that the resistivity value of the fourth layer is not well determined because the absolute value of the fourth component of all the eigenvectors is small except for the eigenvector corresponding to the smallest eigenvalue. This is also seen by considering the covariance matrix which is written in the output list file. We see that the fifth and the third diagonal elements are are considerably smaller than the others showing that the corresponding model parameters are the best determined ones. The seventh diagonal element is considerably larger than the others, showing that the seventh model parameter is not well determined.

The program also writes the correlation matrix into the output list file. This matrix indicates, as was discussed in section 5.3, how the different model parameters are correlated. If an off-diagonal element of the correlation matrix has absolute value close to one, it means that the corresponding model parameters are correlated. If an off-diagonal element is close to plus one, then the sum of the corresponding logarithmic parameters and hence the product of the linear parameters is a well defined parameter. If on the other hand an off-diagonal element is close to minus one then the difference between the corresponding logarithmic model parameters and hence the ratio of the linear parameters is a well defined parameter. There is a weak correlation between the first and the fifth model parameters (resistivity and thickness of the first layer; $C_{5,1}$=-0.923) in the above example. There is also a weak correlation between the second and fifth model parameters (resistivity of the second layer and the thickness of the first layer; $C_{5,2}$=-0.942). Two model parameters are not strongly correlated unless the corresponding correlation matrix element has absolute value very close to one (≥ 0.99).

Finally the program writes into the output list file the time values T for the data points, the measured apparent resistivity values Rhoam, the apparent resistivity values Rhoac calculated from the final model and the weight factors for the data points.

6.3.2 THE OUTPUT PLOT FILE

The inversion program writes the measured data points, the calculated apparent resistivity values, the final model and the value of the square root of the chi-square sum into the output plot file. The content of this file can be plotted both on the screen and as a hard copy on a printer or a plotter. To plot the results the files SPLOT.BAT (screen plot), PPLOT.BAT

(paper plot), RES.GRD (grid file) and CENTERED.SYM (plot symbols) must be in the working directory and the files VIEW.EXE (screen plot) and PLOT.EXE (paper plot) must be either in the working directory or a directory specified in the PATH-list.

The measured apparent resistivity values are plotted against the square root of time, measured in microseconds (i.e. $1000 \cdot \sqrt{t}$ where t is in seconds), as small circles on a double logarithmic plot and the calculated apparent resistivity curve is drawn as an unbroken line. The resistivity model is displayed both numerically as resistivity values (Ohmm) and layer thicknesses (m) and also as a histogram where the x-axis shows the depth in meters and the y-axis the resistivity values. The value of the square root of the chi-square sum is also displayed on the plot. The plot is marked by a station identification which is identical to that part of the output plot file name which is in front of the point (.). If e.g. the output plot file is given the name HT05.PLT, then the plot will be marked as STATION: HT05.

To plot the results on the terminal, simply type SPLOT followed by the name of the output plot file and press return. This initiates a command procedure that appends the plot file to the grid file RES.GRD and plots the results on the screen by the program VIEW.EXE. The plot can be zoomed in by striking the + key and shifted both in horizontal and vertical directions by striking the arrow keys. To return to the DOS prompt, strike the Esc-key, then q and return.

To plot a hard copy on a printer or a plotter, simply type PPLOT followed by the name of the output plot file and return. This initiates a command procedure that appends the plot file to the grid file and plots the results by the program PLOT.EXE. The plot program asks if the plot is to be shifted, and if confirmed, how much in each direction.

The first time the plot utilities are used to plot the results on the screen or as a hard copy or if the output device is changed, it may be necessary to reset the output device specification in VIEW.EXE and/or PLOT.EXE. This is done by changing the working directory to the directory containing the plot program and typing VIEW/I or PLOT/I depending on which plot mode is to be initialized. The plot program displays the current output device specification and asks if it is to be changed. If this is answered positively, it displays a list of possible output devices and the appropriate choice can be made and saved by following a step by step procedure conducted by the plot program. An example of an output plot, plotted on a printer, is shown on Figure 6.2.

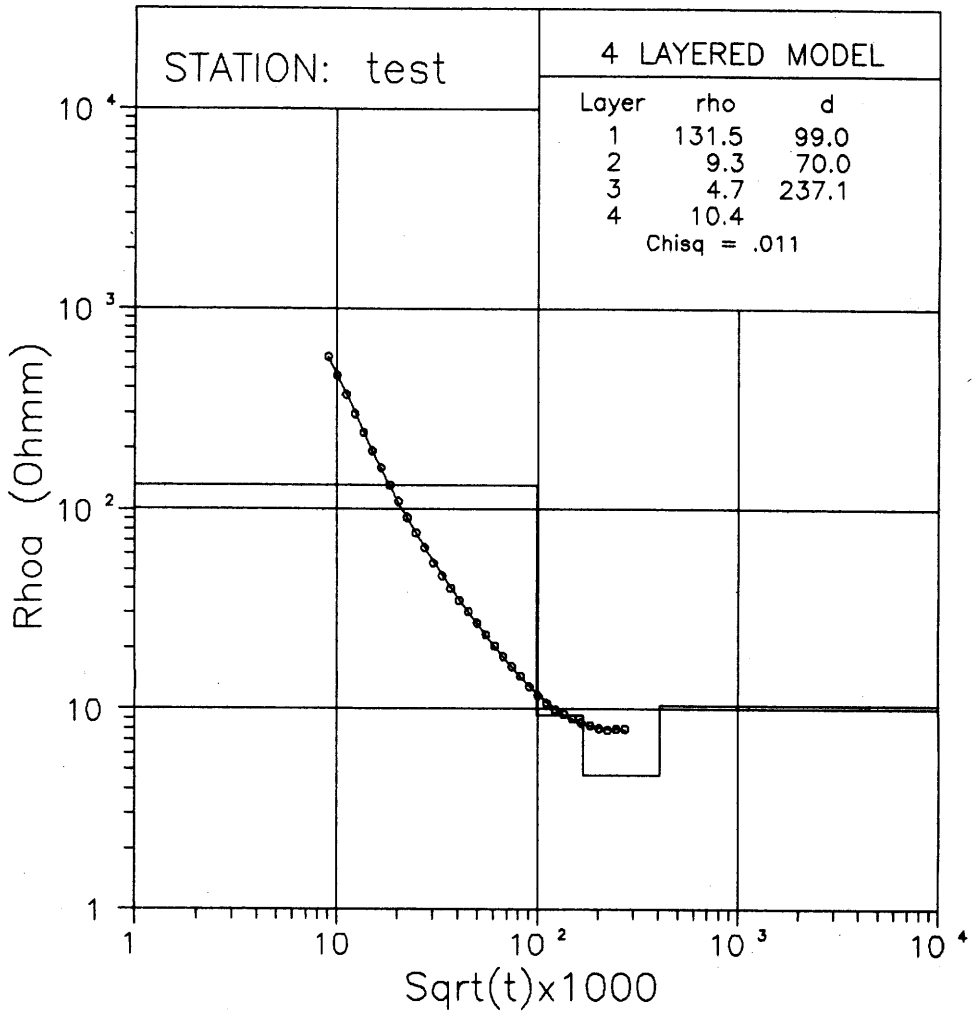

Figure 6.2. An output plot from TINV

REFERENCES:

Abramowitz, M. and Stegun, I. A., 1972. Handbook of Mathematical Functions. Dover Publications, New York, N.Y., 1045 pp.

Árnason, Knútur, 1984. The Effect of Finite Potential Electrode Separation on Schlumberger Soundings. 54th Annual International SEG Meeting, Atlanta, Extended abstracts, 129-132.

Árnason, Knútur, et al. 1987. Nesjavellir-Ölkelduháls, yfirborðsrannsóknir 1986. A report from Orkustofnun on geophysical an geological investigation of the Nesjavellir high temperature geothermal field in SW-Iceland (In Icelandic with English summary). OS-87018/JHD-02, 112 pp.

Björnsson, Grímur and Hjálmar Eysteinsson, 1988. Viðnámsmælingar á Meðallandssandi í júlí 1988. A report from Orkustofnun on prospecting for saline ground water at the south coast of Iceland (In Icelandic). OS-88061/JHD-31 B, 7 pp.

Ghosh, D. P., 1971. The Application of Linear Filter Theory to the Direct Interpretation og Geoelectrical Resistivity Sounding Measurements, Geophysical Prospecting 19, 192-217

Ghosh, D. P., 1971a. Inverse Filter Coefficients for the Computation of Apparent Resistivity Standard Curves for a Horizontally Stratified Earth, Geophysical Prospecting 19, 769-775

Hoffman, K. and Kunze, R., 1971. Linear Algebra, Prentice-Hall, New Jersey, 407 pp.

Hoversten, G. M. and Morrison, H. F., 1982. Transient Fields of a Current Loop Source above a Layered Earth, Geophysics, 47, 1068-1077

Johansen H. K., 1972. A Man/Computer Interpretation System for Resistivity Soundings over a Horizontally Stratified Earth, Geophysical Prospecting, 25, 667-691

Kaufman, A.A. and Keller, G.V., 1983. Frequency and Transient Soundings, Elsevier, Amsterdam, 685 pp.

Koefoed, O., 1972. A Note on the Linear Filter Method of Interpreting Resistivity Sounding Data, Geophysical Prospecting, 20, 403-405

Kunetz, G., 1966. Principles of Direct Current Resistivity Prospecting, Borntraeger, Berlin, 101 pp.

Nabigham, M.N. (ed.), 1988. Electromagnetic Methods in Applied Geophysics, Vol. 1, Theory, Society of Exploration Geophysicists, Tulsa Oklahoma, 513 pp.

Selby, S. M., 1974. Standard Mathematical Tables, CRC Press, 706 pp.

Sternberg, B. K., Washburne, J. C. and Pellerin, L., 1988. Correction for the Static Shift in Magnetotellurics Using Transient Electromagnetic Soundings, Geophysics, 53, 1459-1468