Inverse Problems **14** (1998) 387–404. Printed in the UK PII: S0266-5611(98)92486-8

The role of nonlinearity in inverse problems

Roel Snieder

Department of Geophysics, Utrecht University, PO Box 80.021, 3508 TA Utrecht, Netherlands
and
Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA

Received 13 March 1998

Abstract. In many practical inverse problems, one aims to retrieve a model that has infinitely many degrees of freedom from a finite amount of data. It follows from a simple variable count that this cannot be done in a unique way. Therefore, inversion entails more than estimating a model: any inversion is not complete without a description of the class of models that is consistent with the data; this is called the appraisal problem. Nonlinearity makes the appraisal problem particularly difficult. The first reason for this is that nonlinear error propagation is a difficult problem. The second reason is that for some nonlinear problems the model parameters affect the way in which the model is being interrogated by the data. Two examples are given of this, and it is shown how the nonlinearity may make the problem more ill-posed. Finally, three attempts are shown to carry out the model appraisal for nonlinear inverse problems that are based on an analytical approach, a numerical approach and a common sense approach.

1. Introduction

An important aspect of the physical sciences is to make inferences about physical parameters from data. In general, the laws of physics provide a way for computing the values of the data for a given model. This is called the 'forward problem', see figure 1. In the inverse problem, one aims to reconstruct the model from a set of measurements, see figure 1. In the ideal case there is an exact theory that prescribes how the data should be treated in order to reproduce the model. For some selected examples such a theory exists assuming that the required infinite and noise-free data sets would be available. A quantum mechanical potential in one spatial dimension can be reconstructed when the reflection coefficient is known for all energies (Marchenko 1955, Burridge 1980). This technique can be generalized for the reconstruction of a quantum mechanical potential in three dimensions (Newton 1989), but in this case a redundant data set is required for reasons that are not well understood. The mass-density in a one-dimensional string can be constructed from the measurements of all eigenfrequencies of that string (Borg 1946), but due to the symmetry of this problem only the even part of the mass-density can be determined. If the seismic velocity in the Earth depends only on depth, the velocity can be constructed exactly from the measurement of the arrival time of seismic waves using an Abel transform (Herglotz 1907). Mathematically this problem is identical to the construction of a spherically symmetric quantum mechanical potential in three dimensions (Keller *et al* 1956). However, the construction method of Herglotz (1907) only gives a unique result when the velocity increases monotonically with depth (Gerver and Markushevitch 1966). This situation is similar in quantum mechanics where a radially symmetric smooth potential can only be constructed uniquely when the potential does not have local minima (Sabatier 1973).

0266-5611/98/030387+18$19.50

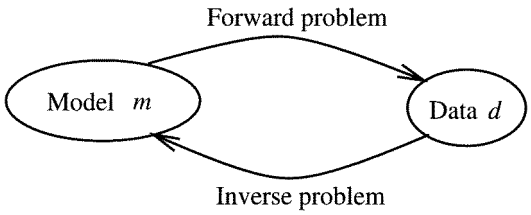

Figure 1. The conventional division of a problem into a forward and an inverse problem.

Despite the mathematical elegance of the exact nonlinear inversion schemes, they are of limited applicability. There are a number of reasons for this. First, the exact inversion techniques are usually only applicable for idealistic situations that may not hold in practice. For example, the Herglotz–Wiechert inversion presupposes that the velocity in the Earth depends only on depth and that the velocity increases monotonically with depth. Seismic tomography has shown that both requirements are not met in the Earth's mantle (Nolet *et al* 1994). Second, the exact inversion techniques are often very unstable. The presence of this instability in the solution of the Marchenko equation has been shown explicitly by Dorren *et al* (1994). However, the third reason is the most fundamental. In many inverse problems the model that one aims to determine is a continuous function of the space variables. This means that the model has infinitely many degrees of freedom. However, in a realistic experiment the amount of data that can be used for the determination of the model is usually finite. A simple count of variables shows that the data cannot have sufficient information to determine the model uniquely. In the context of linear inverse problems this point has been raised by Backus and Gilbert (1967, 1968) and by Parker (1994). However, this issue is also of relevance for nonlinear inverse problems. In practice one often replaces the model with infinitely many degrees of freedom by a model that is characterized by a finite number of parameters. It is shown by Trampert and Snieder (1996) that this may lead to a bias in the model estimate when the true model does not completely lie in the subspace spanned by the finite-dimensional models used in the inversion.

2. Inversion = estimation + appraisal

The fact that in realistic experiments one has a finite amount of data to reconstruct a model with infinitely many degrees of freedom necessarily means that the inverse problem is not unique in the sense that there are many models that explain the data equally well. The model obtained from the inversion of the data is therefore not necessarily equal to the true model that one seeks. This implies that the view of inverse problems shown in figure 1 is too simplistic. For realistic problems, inversion really consists of two steps. Let the true model be denoted by m and the data by d. From the data d one reconstructs an estimated model $\hat{m}$, this is called the *estimation problem*, see figure 2. Apart from estimating a model $\hat{m}$ that is consistent with the data, one also needs to investigate what the relation is between the estimated model $\hat{m}$ and the true model m. This is called the *appraisal problem*. In the appraisal problem one determines what properties of the true model are constrained by the estimated model. For example, one might find that for the seismic velocity in the Earth only the spatial fluctuations with a wavelength larger than a certain threshold are constrained by the data. The upshot of this discussion is that *inversion = estimation + appraisal*.

In general there are two reasons why the estimated model differs from the true model. The first reason is the non-uniqueness of the inverse problem that causes several (usually

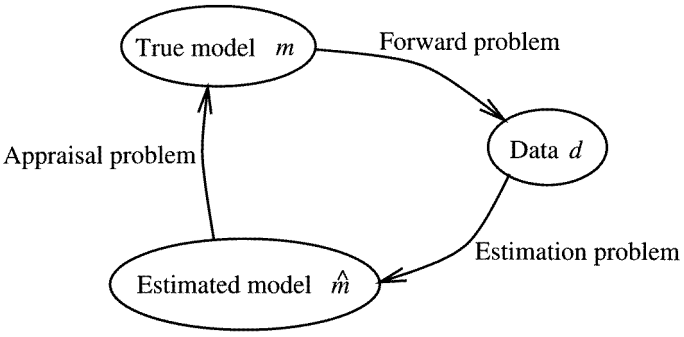

Figure 2. The division of a problem into a forward problem, an estimation problem and an appraisal problem that is appropriate for finite data sets.

infinitely many) models to fit the data. The second reason is that real data are always contaminated with errors and the estimated model is therefore affected by measurement errors as well. Therefore model appraisal has two aspects, non-uniqueness and error propagation. For *linear* inverse problems, methods exist to handle both aspects. The effect of non-uniqueness is accounted for in the theory of Backus and Gilbert (1968). For a linear problem, the ith datum d_i is related to the model $m(x)$ through a linear relation:

$$d_i = \int G_i(x)m(x)\,\mathrm{d}x. \tag{2.1}$$

Although the notation used here is one-dimensional, the arguments presented here are applicable in any number of dimensions. For a linear inverse problem, the estimated model $\hat{m}(x)$ at location x follows by making linear combinations of the data:

$$\hat{m}(x) = \sum_i a_i(x)d_i. \tag{2.2}$$

Backus and Gilbert (1968) provide a way to compute the coefficients $a_i(x)$. However, it should be noted that *any* linear estimator, including the conventional least-squares solution, can be written in the form (2.2) so that what follows holds for *any* linear estimator. Inserting equation (2.1) in (2.2) one can relate the estimated model to the true model:

$$\hat{m}(x) = \int R(x;x')m(x')\,\mathrm{d}x' \tag{2.3}$$

with

$$R(x;x') = \sum_i a_i(x)G_i(x'). \tag{2.4}$$

The function $R(x;x')$ is called the *resolution kernel*. In the ideal case, the estimated model $\hat{m}(x)$ and the true model $m(x)$ are equal. This is the case when the resolution kernel is a delta function: $R(x;x') = \delta(x-x')$. However, it follows from expression (2.4) that the resolution kernel can be viewed as the linear superposition of a finite amount of data kernels $G_i(x')$. Since a delta function cannot be written as a superposition of a finite amount of functions, this implies that the resolution kernel must have a finite width. In other words, the estimated model is a blurred version of the true model. Note that this holds for any linear inversion scheme, because any linear estimation procedure can be written in the form (2.2). A beautiful example of the use of resolution kernels is given by Tanimoto and Anderson (1985) who use resolution kernels to assess to what extent seismic surface waves constrain the seismic anisotropy in the Earth.

One should note that the theory given in equations (2.1)–(2.4) is actually quite subtle. It follows from expression (2.3) that $\hat{m}(x)$ cannot strictly be an estimate of the local value of the model. Backus and Gilbert (1967, 1968) carefully explain that $\hat{m}(x)$ should be interpreted as an estimate of a model functional, which at best can be considered to be a smoothed version of the true model. Only for such a functional can the error be finite, because an infinite model perturbation over an infinitesimal volume usually does not affect the data. The issues of resolution model error are therefore intricately linked together.

The second aspect of model appraisal, error propagation, lies in the realm of statistics. For linear inverse problems and for Gaussian data errors, the error propagation is relatively simple (e.g. Parker 1994). Bayesian techniques have gained a strong popularity in inverse problems. In this approach one prescribes the statistics of the data, and the statistics of the model before the measurements are taken (the *a priori* model statistics). Bayes' theorem then provides a way to estimate the statistics of the model when the information in the data has been used (e.g. Franklin 1970, Tarantola 1987). However, it is argued by Scales and Snieder (1997) that the prescription of the *a priori* model statistics is far from trivial and that Bayesian inversion is frequently abused in practical inversions. In addition, the theory for Bayesian inversion is only developed to the point where it can be applied for models that are characterized by a finite number of parameters. The formulation of Bayesian inversion for models with infinitely many degrees of freedom would require Wiener integration, which is not easily implemented for practical problems.

Despite these reservations, Bayesian inversion is an extremely valuable tool in inverse problems. One could, in fact, go one step further than indicated in figure 2 and view the inverse problem as a problem of statistical inference that aims at obtaining the complete *a posteriori* probability density function of the model as its final product. The problem with this approach is that it is not trivial to interpret this function; the interface with the human interpreter forms a major problem that needs to be solved. This is aggravated by the fact that many users of the models obtained from inversions of data prefer a single model they can work with (possibly supplemented with error estimates) rather than a more subtle answer in the form of a probability density function in a large-dimensional or infinite-dimensional model space.

The key point of this discussion is that the theory of model appraisal is only well developed up to the point that it can easily be used in practical inverse problems for linear inverse problems. The concept of resolution kernels to address the non-uniqueness of inverse problems is not easily generalized to general nonlinear inverse problems. The effect of error propagation and the use of Bayesian inversion is for most practical purposes restricted to linear (or linearized) inverse problems where Gaussian errors in the data lead to Gaussian errors in the model parameters. This does not imply that Bayesian inversion cannot be applied to nonlinear inverse problems, but these applications strongly rely on numerical techniques to sample model space. This leads to the paradox that part of the inverse problem community has invested a large amount of energy in developing exact theoretical inversion methods that cannot be applied to real data, and the another part of the community solves realistic problems, but that the theory for model appraisal for nonlinear inverse problems is poorly developed.

So how are inverse problems solved in practice? The estimation problem is usually solved by fitting the model to the data. Let the ith datum d_i be related to the model m through the relation

$$d_i = G_i(m) \tag{2.5}$$

where $G_i(m)$ is a nonlinear functional that maps the model on the datum d_i. The data fitting

can be achieved by minimizing the difference between the real data d_i and the estimated data $G_i(\hat{m})$ as a function of the estimated model $\hat{m}$. In the simplest approach one minimizes the unweighted least-squares misfit

$$S(\hat{m}) = \sum_i (d_i - G_i(\hat{m}))^2 \tag{2.6}$$

as a function of the estimated model $\hat{m}$. However, the resulting solution may be very unstable and it may not be very physical. In general one adds additional terms to expression (2.6) to make the solution well behaved. These terms can be used to impose ideas about the character of the solution that one seeks. For example, in Occam's inversion (Constable *et al* 1987) one seeks the smoothest solution that is consistent with the data by adding a term to (2.6) that penalizes the squared gradient of the model. Another important regularization technique is Bayesian inference which provides an elegant way to include statistical information of the data and the model in a misfit criterion (e.g. Tarantola 1987). The misfit function (2.6) possibly supplemented with regularization terms is usually referred to as the penalty function.

Finding the minimum of the penalty function is an art in itself. In general one either uses a descent method where one iteratively updates an estimate of the model that minimizes the penalty function by moving 'downhill' in one way or another, or one employs techniques that sample model space in many different locations as a way of minimization. Descent methods can relatively easily be implemented for large-scale inverse problems. The main drawback of this approach is that for nonlinear inverse problems the penalty function may have several minima. A descent method may lead to a model estimate that corresponds to a local estimate of the penalty function rather than the global minimum. It is for this reason that techniques that sample model space in a more global fashion have gained popularity (e.g. Sambridge 1998) despite their computational inefficiency.

For the model appraisal, one usually employs a local linearization of the relation between data and model to account for the error propagation in the estimated model (e.g. Tarantola 1987, Parker 1984). However, this is an oversimplification of the true error propagation, and the bias in the model estimate due to nonlinear error propagation (Jackson and Matsu'ura 1985, Dorren and Snieder 1997) is not accounted for in this approach.

An alternative to estimating models by data fitting is to use perturbation theory for the inverse problem where the nonlinearity is seen as a weak perturbation. This approach was originally used by Prosser (1968) for the determination of a quantum mechanical potential from scattering data. The technique has been generalized by Snieder (1990a, b) for general inverse problems where the data can be written as a perturbation series of the model. Generalizing expression (2.1) for linear problems this means that it is assumed that the data can be written as

$$d_i = \int G_i^{(1)}(x) m(x)\, \mathrm{d}x + \iint G_i^{(2)}(x_1, x_2) m(x_1) m(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 + \cdots . \tag{2.7}$$

In general, the model estimate is a nonlinear function of the data and can analogously to equation (2.2) be written as a Taylor series of the data that includes nonlinear terms:

$$\hat{m}(x) = \sum_i a_i^{(1)}(x) d_i + \sum_{i,j} a_{ij}^{(2)}(x) d_i d_j + \cdots . \tag{2.8}$$

When infinitely many data are available and one knows that the model can be recovered exactly from the data, the coefficients $a^{(i)}$ of expression (2.8) can be found by inserting (2.7) in (2.8) and by requiring that $\hat{m}(x) = m(x)$, see Prosser (1968) or Snieder (1990a, b) for details. However, in general the data set is finite, and it is not realistic to require that

the estimated model $\hat{m}(x)$ equals the true model. This case has been handled by Snieder (1991) in a nonlinear extension of Backus–Gilbert theory and will be discussed further in section 4.1.

It should noted that the series (2.7) is often difficult to obtain. A notable exception is the iterated form of a linear integral equation that can always be written in the form (2.7). An important example of this is the Neumann series in scattering theory that follows by iterating the integral equation of the scattering problem (e.g. Rodberg and Thaler 1967). The forward problem of multiple scattering of waves can thus always be written in the form of the series (2.7).

The reader may wonder why the issue of existence has not been invoked in the discussion. The existence problem is concerned with the question of whether there is at least one model that is consistent with the data. For linear inverse problems one can be sure that such a model exists (e.g. Parker 1994), but for nonlinear inverse problems one cannot be sure that a model exists that is consistent with a given data set. However, the existence issue is mostly a mathematical problem. It is crucial to make the distinction between an artificial data set or a data set that is physically generated by taking real mesurements. As an example of an artificial data set one can think of an imaginary reflection time series $r(t)$ that is an arbitrary function of time. For a nonlinear inverse problem, such a data set does not necessarily correspond to a model, and hence the existence issue is extremely relevant. However, when the data are obtained from real measurements the existence problem is not relevant because one knows that in that case the presence of the true model (that is responsible for the recorded data) guarantees that there is at least one model that is consistent with the data. This means that for this kind of data set the existence issue is not relevant. However, the situation is slightly more complex because in real measurements the data are always contaminated with errors. In general, there is no strict proof that there is a model that corresponds to the error-contaminated data. However, in the case of noise-contaminated data one will, in general, not aim at fitting the data perfectly and one can be sure that within the data errors there is a model that fits the data within the data error (namely the true model). When this is the case, the primary concern is not the existence issue but the problem of error propagation.

3. Nonlinearity and ill-posedness

Suppose for the moment that one aims at estimating a model and that the estimated model is constructed by minimizing the data misfit such as (2.6) or some more complex misfit function. If the forward problem is linear, and if an L_2-norm such as (2.6) is used for the calculation of the misfit, the misfit function has a parabolic dependence on the model parameters, and therefore the misfit function has a single minimum, this situation is depicted in figure 3(*a*). Any type of descent method will lead to this unique minimum. When the forward problem is nonlinear, the misfit function can have multiple minima, see figure 3(*b*). A practical optimization problem where multiple minima are a severe problem is the determination of static corrections in exploration seismics (e.g. Rothman 1985). The problem with the local minima is that search methods for the global minimum may misidentify a local minimum as the global minimum; in that case the estimated model is not the model that gives the best data fit. In sections 3.1 and 3.2 two examples show that this is an oversimplified view of the complications introduced by the nonlinearity of inverse problems. (However, this does not change the fact that local minima are a formidable problem in large-scale nonlinear inverse problems.)

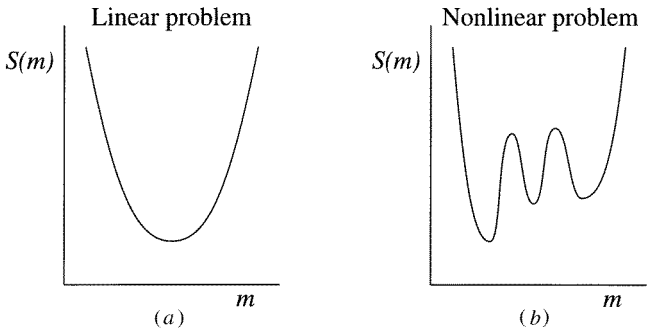

Figure 3. (*a*) The least-squares misfit function for a linear problem. (*b*) The conventional view of the misfit function for a nonlinear inverse problem.

3.1. Example 1, nonlinearity and the inverse problem for the Schrödinger equation

The inverse problem of the estimation of a quantum mechanical potential in one dimension from the measurement of the reflection coefficient of waves reflected by the potential is a very interesting tool for studying nonlinear inversion because the inverse problem has a stunningly simple solution (Marchenko 1955, Burridge 1980). This inverse problem is of direct relevance in the Earth sciences; both the inverse problem of geomagnetic induction (Weidelt 1972) as well as the seismic reflection problem (Burridge 1980, Newton 1981) can be reformulated as the problem treated in this section. For the Schrödinger equation the wavefield ψ satisfies the following differential equation:

$$\psi_{xx} + (k^2 - V(x))\psi = 0. \tag{3.1}$$

Let the reflection coefficient after a Fourier transform to the time domain be denoted by $R(t)$. The potential $V(x)$ at a fixed location x follows from the reflection coefficient by solving the Marchenko equation:

$$K(x,t) + R(x+t) + \int_{-t}^{x} K(x,\tau)R(\tau+t)\,\mathrm{d}\tau = 0 \tag{3.2}$$

for $K(x,t)$ and carrying out a differentiation

$$V(x) = -2\frac{\mathrm{d}K(x,x)}{\mathrm{d}x}. \tag{3.3}$$

In figure 4(*a*) the results of a numerical solution of the Marchenko equation are shown. The potential is shown by a full curve. For this potential the reflection coefficient is computed and the potential is reconstructed by numerically solving the Marchenko equation (3.2) and carrying out the differentiation (3.3), details of the calculation can be found in Dorren *et al* (1994). The reconstructed potential of figure 4(*a*) is, on the scale of the figure, indistinguishable from the true potential. In figure 4(*b*) the same synthetic experiment is shown, the only difference is that the potential has been multiplied by a factor of 2.5. If the problem was linear, the reflection coefficients would be 2.5 times as strong as for the potential in the top panel and the reconstructed potential would also be 2.5 times as strong. This would lead to a near-perfect reconstruction of the potential. However, the reconstructed potential is given by the broken curve. It can be seen that the left part of the potential (the side from which the waves are incident) is reconstructed quite well, but that the part of the potential on the right is very poorly reconstructed. According to the

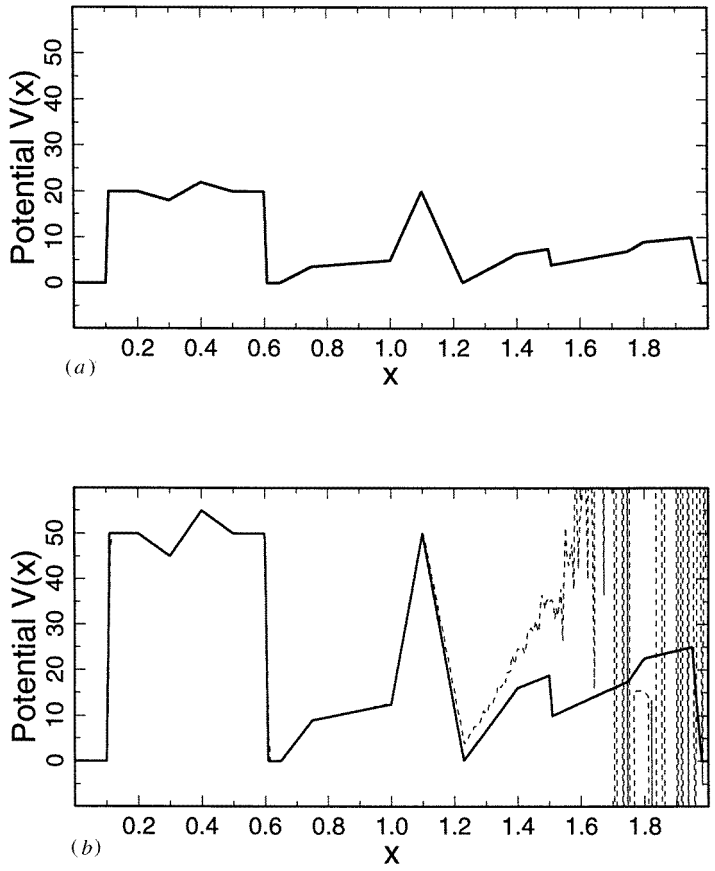

Figure 4. (*a*) The original potential (full curve) and the potential constructed by solving the Marchenko equation (broken curve). The two curves are indistinguishable on the scale of this figure. (*b*) The original potential (full curve) and the potential constructed by solving the Marchenko equation (broken curve) for a potential that is 2.5 times as strong as the potential in (*a*).

reasoning above, the potential would have been reconstructed quite well if the problem had been linear. This implies that the instability in the reconstructed potential is due to the nonlinearity of the problem.

The physical reason for this instability can be relatively easily understood. It follows from the Schrödinger equation (3.1) that the effective wavenumber is given by $\sqrt{k^2 - V(x)}$. When $k^2 < V(x)$ the wavenumber is complex which reflects the fact that the waves are evanescent when the potential energy is larger than the total energy. In that case the wavefield decays exponentially within the potential. For a fixed energy k^2 the wavefield is more evanescent for the potential in figure 4(*b*) than for the potential in figure 4(*a*), simply because the potential energy is 2.5 times as high. This implies that the wavefield penetrates deeper in the potential in figure 4(*a*) than in the potential in figure 4(*b*). Obviously, the potential in a certain region is not constrained by the recorded wavefield if the wavefield does not sample the potential in that region. In that case, the numerical details of the algorithm, including the numerical round-off error determine the reconstructed potential in that region. The essential point is that the values of the model parameters affect the way in

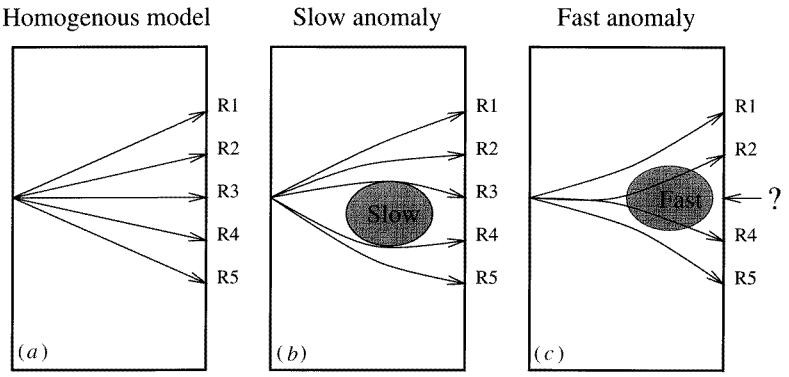

Figure 5. (*a*) Tomographic experiment where the velocity is homogeneous and the rays are straight. (*b*) Tomographic experiment where the rays curve around a low-velocity body. (*c*) Tomographic experiment where a high-velocity anomaly causes a shadow zone at the middle receiver.

which the wavefield interrogates the model.

Physically, the instability in the reconstructed potential in figure 4(*b*) can thus be understood. However, what does this imply for the inverse problem? What happens physically if the potential on the left side is high, is that the wavefield is prevented from sampling the potential on the right part. This means that for some values of the model parameters (that describe how high the potential is on the left), other model parameters (that describe the potential on the right) are unconstrained by the data. In terms of a misfit function this implies that the misfit does not depend on the model parameters that describe the right side of the potential (when the left side of the potential is high). In other words, the misfit function does not have a minimum, but has a broad plateau. Note that as an additional complexity this only occurs for certain values of other model parameters (that describe how high the potential is on the left).

3.2. Example 2, nonlinearity and seismic tomography

A second example of the ill-posedness introduced by the nonlinearity in inverse problems is seismic tomography. Consider a cross-borehole tomographic experiment where rays travel from a source in a well to a string of receivers in another well. The case where the velocity is homogeneous is shown in figure 5(*a*). In that case the rays are straight lines that travel from the source on the left to receivers $R1$ through $R5$ on the right. If the velocity is not homogeneous, the rays are curved. This implies that just as in the example of section 3.1 the values of the model parameters determine the way in which the rays interrogate the model.

Suppose that a low-velocity anomaly is present, see figure 5(*b*). Since the rays of first arrivals are curves of minimal travel time, the rays curve around the slow velocity anomaly. If the velocity anomaly is sufficiently slow, all the rays may completely curve around the slow velocity. In that situation the anomaly would not be sampled by any rays and the velocity within the anomaly cannot be determined because it is not sampled by any rays. At best one could derive an upper bound for the velocity within the anomaly. Just as in the previous section, the model parameters affect the way in which the probe (in this case the rays) samples the model. In terms of the misfit function this implies that for a certain range of model parameters, the misfit function does not depend on the model parameters at all. In other

words: the misfit function is completely flat over an extended range of parameter values.

Let us now consider the opposite situation where a high-velocity anomaly is present, see figure 5(*c*). Rays will be defocused by a high-velocity body and a shadow zone is formed behind the anomaly. This may mean that there is no ray that will hit the receiver $R3$, see the question mark in figure 5(*c*). This means that for some values of the model parameters it is impossible to compute the data because the travel time cannot be determined when there is no ray that hits the receiver. This means that for some values of the model parameters it is impossible to compute the corresponding data values given the theory that one uses. In this sense one can say that some values of the model parameters are 'forbidden'. It should be noted that this is not a complexity created in some exotic thought experiment; the problem of the 'missing rays' is a real problem in nonlinear travel time tomography (e.g. Sambridge 1990).

The critical reader might remark at this point that the fact that for certain values of the velocity model there are no rays hitting the receiver is due to the fact that ray theory is an approximation to the true theory (wave theory), and that wave theory predicts that some energy will diffract into the shadow zones. Although this is correct, one should also note that the wavefield in the shadow zones is very weak (this is why they are called shadow zones) and that in practice this diffracted wavefield is usually not detectable.

3.3. The landscape of the misfit function

In the examples in the previous sections the value of the model parameters have affected the way in which the probe (either the wavefield in example 1 or the rays in example 2) interrogates the model. This has led to the effect that for some range of model parameters the data are insensitive to some model parameters. This implies that the view that the main complexity of nonlinearity in inverse problems is the occurrence of multiple minima is oversimplistic. The examples of this section show that the misfit function may also have plateaus, i.e. there may be regions where the misfit is independent of the model parameters. This situation is sketched in figure 6. Note that when the data do not depend on a model parameter (in a certain range), the inverse problem is necessarily ill-posed. The nonlinearity may, therefore, actually render inverse problems ill-posed. In addition to this ill-posedness, the second example has also shown that some range of model parameters may actually be inconsistent with the data. This means that there are 'forbidden regions' of model space due to the nonlinearity of the problem. These two complexities introduced by the nonlinearity

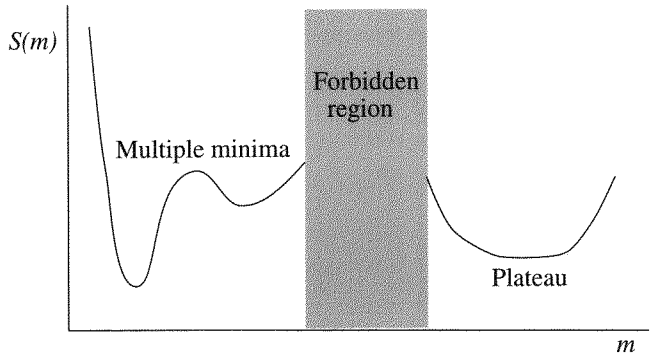

Figure 6. A misfit function that shows the complexity that may occur for a truly nonlinear inverse problem.

of the problem are indicated in figure 6. Together with the occurrence of multiple minima these complexities make model estimations for a truly nonlinear problem a challenging task. Further details on the complexity of the misfit function for nonlinear inverse problems are given by Sambridge (1998) in this issue. Deng (1997) gives examples of misfit functions that have pathologically many minima and provides a quantitative measure for the complexity of a misfit function.

4. Model appraisal for nonlinear inverse problems

In the previous section the effect of nonlinearity on model estimation has been discussed. In this section an attempt is made to describe the effect of nonlinearity on the appraisal problem where one wants to describe how the estimated model is related to the true model (see figure 2). However, it should be stressed that there is presently no general theory to deal with the appraisal problem for a truly nonlinear inverse problem with infinitely many degrees of freedom. In practice, one often linearizes the problem around the estimated model and then uses linear theory to make inferences about the resolution and reliability of the estimated model. The lack of a general theory for the appraisal problem should be seen as a challenge for theorists! In this section three attempts are described to carry out model assessment for a nonlinear inverse problem. These attempts follow the lines of formal theory (section 4.1), a numerical approach (section 4.2) and a pragmatic approach (section 4.3).

4.1. Nonlinear Backus–Gilbert theory

The beauty of Backus–Gilbert theory is that it leads for linear problems to a resolution kernel that describes how the estimated model and the true model are related, see expression (2.3). In the ideal case the resolution kernel is a delta function. Backus and Gilbert (1967, 1968) have shown that the criterion that the resolution kernel should resemble a delta function as much as possible can be used to determine the coefficients $a_i(x)$ in equation (2.2) that prescribe how datum d_i affects the estimated model at location x. This theory has been generalized by Snieder (1991) for the special case in which the forward problem can be written as a perturbation series as shown in equation (2.7). In a number of applications such a perturbation series arises naturally. Important examples are the Neumann series in scattering theory (e.g. Rodberg and Thaler 1967) where the scattering data are written as a sum of integrals that contain successively higher powers of the potential, or ray perturbation theory (e.g. Snieder and Sambridge 1993, Snieder and Aldridge 1995) where the travel time of rays is written as a sum of integrals with increasing powers of the slowness perturbation.

The key to nonlinear Backus–Gilbert theory is to insert the expansion of the forward problem (2.7) in the estimator (2.8). The result can then be written as

$$\hat{m}(x) = \int R^{(1)}(x; x_1) m(x_1)\, \mathrm{d}x_1 + \int\!\!\int R^{(2)}(x; x_1, x_2) m(x_1) m(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 + \cdots . \tag{4.1}$$

This expression generalizes the linear resolution kernel of equation (2.3) to nonlinear inverse problems. The kernel $R^{(1)}(x; x_1)$ describes to what extent the estimated model is a blurred version of the true model. The higher order kernels such as $R^{(2)}(x; x_1, x_2)$ can be interpreted as nonlinear resolution kernels that describe to what extent there is a spurious nonlinear mapping from the estimated model onto the true model in the inversion process.

In the ideal case, the estimated model is equal to the true model: $\hat{m}(x) = m(x)$. This is the case when the linear resolution kernel $R^{(1)}(x; x_1)$ is a delta function $\delta(x - x_1)$ and

when the nonlinear resolution kernels are equal to zero: $R^{(n)}(x; x_1, \ldots, x_n) = 0$ for $n \geqslant 2$. However, as in equation (2.4) the linear resolution kernel $R^{(1)}(x; x_1)$ can be written as a sum of a finite amount of data kernels $G^{(1)}(x_1)$. Since a delta function cannot be obtained by summing a finite amount of smooth functions, the linear resolution kernel can never truly be a delta function. This reflects the fact that with a finite amount of data the estimated model will be a blurred version of the true model. Snieder (1991) treats the inverse problem of the determination of the mass-density of a vibrating string from the eigenfrequencies of the string. He shows that if only a finite amount of eigenfrequencies are available, the nonlinear resolution kernels cannot be zero. This implies that the finiteness of the data set leads to a spurious nonlinear mapping from the true model to the estimated model. This can be related to the symmetries of the problem and to mode coupling 'off the energy shell'. The finite width of the linear resolution kernel and the fact that the nonlinear resolution kernels are nonzero imply not only that the estimated model is a blurred version of the true model, but also that the estimated model is biased. The reader is referred to Snieder (1991) for details. In that work it is also described how the coefficients $a^{(i)}$ in the estimator (2.8) can be determined.

Although nonlinear Backus–Gilbert theory is a new tool for dealing with the assessment problem for nonlinear inverse problems, one should realize that the theory can only be applied to weakly nonlinear problems where a (very) few orders are sufficient for an accurate description of both the forward and the inverse problem. In addition, the theory of Snieder (1991) is so complex that a reformulation is needed to make the theory applicable to the large-scale inverse problems that are being treated in practice.

It is important to realize that any regular (nonlinear) mapping from data d_i to an estimated model $\hat{m}(x)$ can be written in the form of expression (2.8). The details of the employed algorithm then determine the coefficients $a^{(n)}_{i_1 \ldots i_n}$. The resolution analysis shown in equation (4.1) and the subsequent discussion therefore is applicable to the estimated model. The conclusions concerning the linear and nonlinear resolution kernels can thus be used for *any* regular mapping from the data to the estimated model.

4.2. Generation of populations of models that fit the data

Another approach to asses the reliability of estimated models is to generate not a single model that fits the data within a certain tolerance but to obtain an ensemble of models that fit the data within a certain tolerance (e.g. Lomax and Snieder 1995a). An alternative approach is to compute the misfit for a very large class of models and to use the data fit, possibly in combination with Bayesian statistics to make inferences about the range of models that explain the data in a certain sense (e.g. Mosegaard and Tarantola 1995, Gouveia and Scales 1997, 1998, Mosegaard 1998). Obviously, this approach requires a numerical approach to create such ensembles, but with present day computers significant progress has been made.

An important concept in the generation of ensembles of models is the randomness in the search method that one employs. A descent method contains no element of randomness whatsoever, whereas a Monte Carlo search where one randomly samples model space is completely random. In between are algorithms which have both random components as well as some mechanism to prefer models that fit the data well. Examples of such algorithms are simulated annealing (Krikpatrick *et al* 1983, Rothman 1985) or genetic algorithms (Sambridge and Drijkoningen 1992, Sen and Stoffa 1992, Lomax and Snieder 1995b). A promising technique is the adaptive search (Mosegaard and Tarantola 1995) where in the process of carrying out a random search, information about the misfit function is built up and where this misfit function is used to drive the random search in an intelligent way.

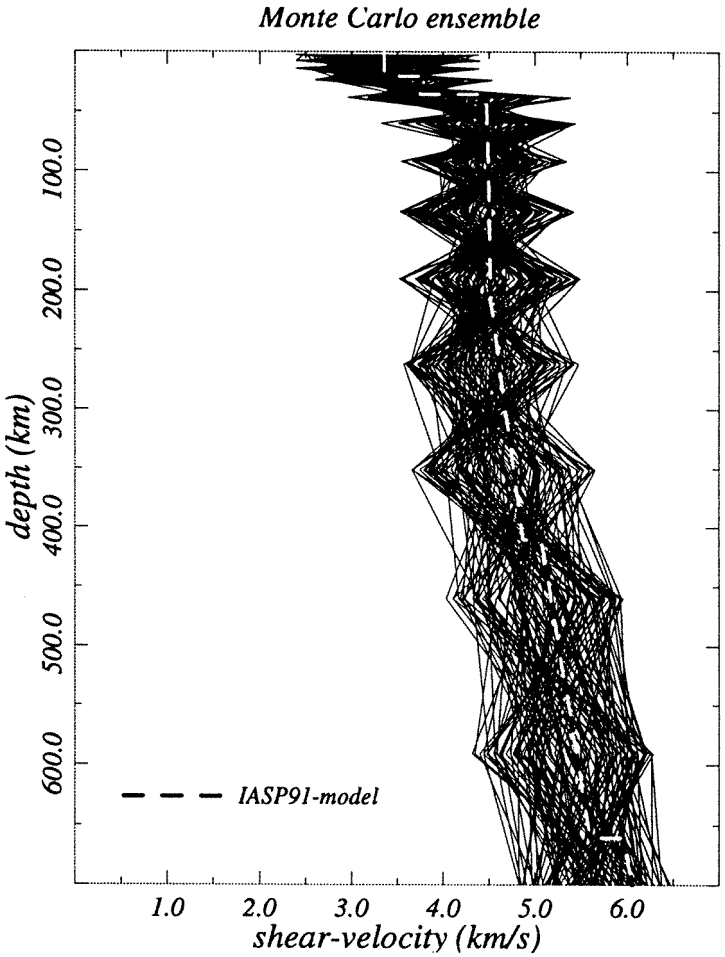

Figure 7. The true velocity model (broken curve) and an ensemble of models generated by a Monte Carlo search that fit the surface wave group-velocity data within a realistic tolerance (full lines).

The main merit of the algorithms that determine an ensemble of models together with information on how well each model explains the data is that this ensemble can be used to make inferences about the model. These inferences may or may not be statistical. This property of the algorithms that produce ensembles of models has not yet been fully exploited.

An example is shown from a study of Douma *et al* (1996). In their study synthetic group-velocity data of the fundamental mode Rayleigh wave that propagates along the Earth's surface were computed for periods between 10 and 300 s. The true velocity model is shown as the broken curve in figure 7 as a function of depth. A Monte Carlo search was used to find models of the S-velocity that were consistent with the data within a realistic measurement error. The resulting population of models is shown in figure 7. In this study, the S-velocity was deliberately overparametized. As a result, the resulting models are highly oscillatory and contain strong trade-offs. The aim of the study of Douma *et al* (1996) was to extract the robust features of the velocity model from the ensemble of models shown in figure 7. This was achieved by computing 'empirical orthogonal functions' (EOFs) from

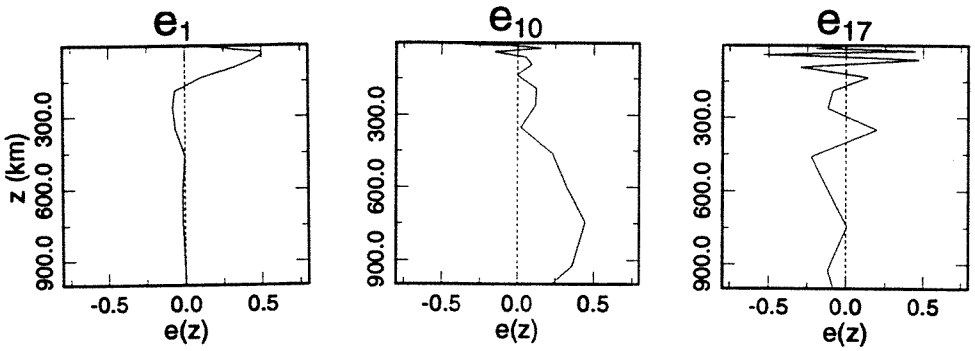

Figure 8. The empirical orthogonal function that corresponds to the pattern of smallest variability (left panel) and two examples of EOFs that correspond to patterns of large variability (middle and right panels) for the ensemble of models of figure 7.

this ensemble. These functions give patterns with different degrees of variability within an ensemble. Three representative EOFs extracted from the ensemble are shown in figure 8. The left panel shows the EOF with the smallest variability within the ensemble. The penetration depth of this function is about the same as the penetration depth of the surface waves that are used. In fact, the EOF in the left panel bears a striking resemblance to the Frechet derivative of the group velocity with respect to the shear velocity. Two EOFs that correspond to patterns with a larger degree of variability are shown in the middle and right panels of figure 8. It can be seen that these EOFs either are very oscillatory (right panel), or they carry much energy at great depth where the surface waves do not penetrate (middle panel). The EOFs that describe patterns of large variability within the population (middle and right panel) correspond to model perturbations that are poorly constrained by the data; these are the trade-offs of the inversion. The EOFs that correspond to patterns of small variability (left panel) account for the well-constrained aspects of the model.

The EOFs can be used to re-parametrize the model space in an intelligent way that reflects how well model perturbations are constrained by the data. Alternatively, the EOFs could be used to carry out a statistical analysis of the ensemble of models that explains the data. However, as noted by Douma *et al* (1996) the EOF techniques are only useful for inverse problems that are weakly nonlinear.

4.3. Using different inversion methods

In the previous sections, a theoretical and a numerical method for model appraisal were presented. Apart from these more formal approaches, 'common sense' is a powerful tool for carrying out model assessment. An important way to assess the reliability of a model is to determine the model in different ways. In the ideal case, different data sets are used by different research groups who use different theories to estimate the same properties. The agreement or disagreement between these models can be used as an indicator of the reliability of these models. It is admittedly difficult to quantify the sense of reliability that is thus obtained, but in the absence of an adequate theory to carry out model assessment for nonlinear inverse problems (recall that we do not have such a theory) this may be the best approach.

An example of this approach is shown in figure 9. Nonlinear waveform inversion using the partitioned waveform inversion of Nolet (1990) has been used by van der Hilst and

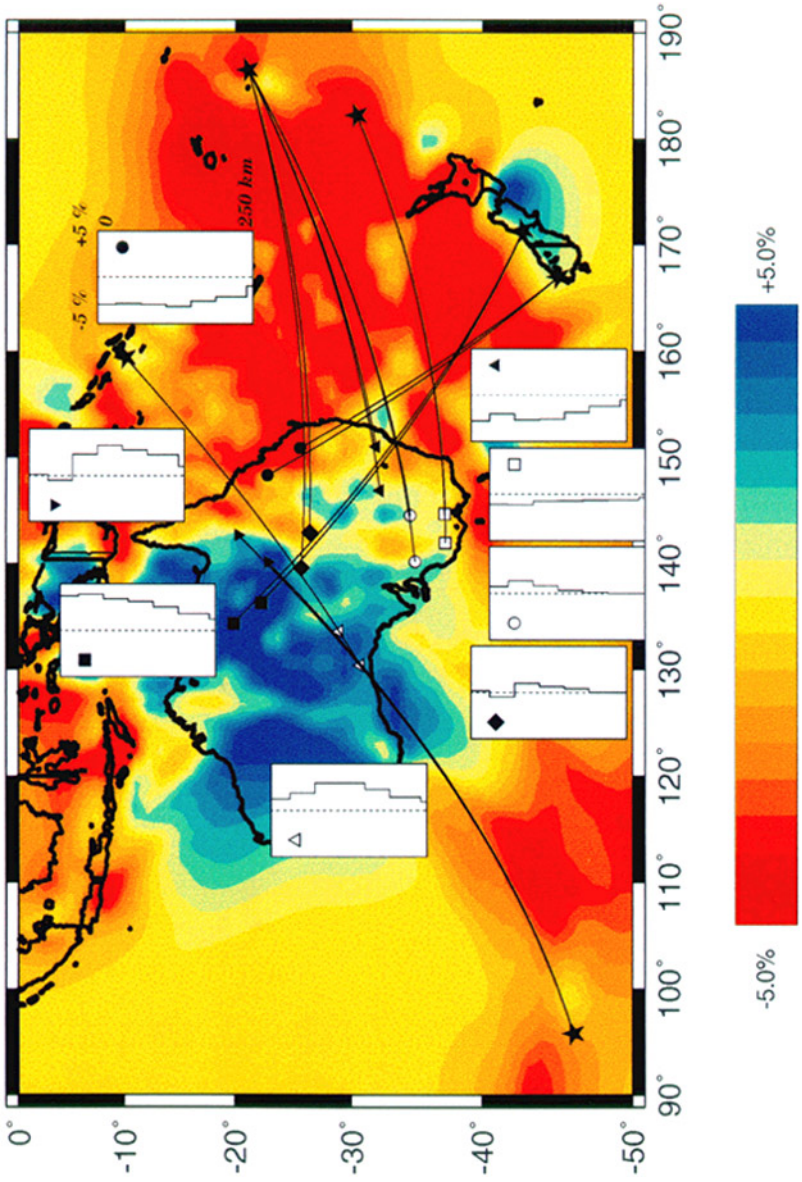

Figure 9. The S-velocity anomaly under Australia at a depth of 140 km obtained from surface wave tomography by van der Hilst and Kennett (1997) (colour contours with the contour levels shown by the colour bar at the bottom) and the interstation S-velocity between stations obtained by Passier *et al* (1997) shown in the panels. The S-velocity in the panels ranges from −5% to +5% and is shown from the surface down to a depth of 250 km. The employed stations for each panel are indicated with a symbol that corresponds with the symbol in the upper-right corner of each panel.

Kennett (1998) to determine a three-dimensional model of the S-velocity under Australia. A cross section of the model at a depth of 140 km is shown by the colours in figure 9. In such an inversion, averages of the Earth's structure along long paths are used to determine the local variations in the S-velocity. Effects such as errors in the source mechanisms of the employed earthquakes, seismic anisotropy that is not accounted for in the inversion, variations in the path coverage and the path direction over the domain of inversion can lead to artifacts in the inversion. As a way of verifying the reliability of the solution, Passier *et al* (1997) determined the local S-velocity using the waveform inversion method described by Passier and Snieder (1995). In this method, a consistency requirement of the waveforms recorded in two nearby stations is used to estimate the horizontally averaged velocity between the stations as a function of depth. Just as the technique used by van der Hilst and Kennett (1998) this technique may lead to models with artefacts, but in general these artefacts will be different, see Passier *et al* (1997) for a discussion.

The interstation S-velocity models are shown in the panels in figure 9. In these panels the depth range is from the surface down to 250 km, whereas the S-velocity ranges from -5% to $+5\%$. The symbol in each panel (circle, triangle, etc) corresponds to the employed stations that are marked on the map with the same symbol. The depth profile shown in each panel shows the S-velocity between the corresponding stations. By comparing the interstation velocity models shown in the panels with the velocity model shown in colour one can see that the locally determined S-velocity agrees well with the S-velocity obtained from the three-dimensional tomographic inversions. Consider, for example, the S-velocity between the stations in the northeast marked with full dots that is shown in the panel by the full dot. At a depth of about 140 km the interstation shear velocity is about -2.5%. This agrees well with the S-velocity obtained from three-dimensional tomography that is indicated with the colours. A comparison like this is very useful for assessing the reliability of models in a qualitative way. In this sense, a healthy competition between different research groups is an important ingredient in the appraisal of models.

5. Conclusion

The main point of this overview is that for realistic data sets, the inverse problem really consists of two problems: the estimation problem and the appraisal problem. For linear inverse problems powerful techniques exist both for the estimation problem as well as for the appraisal problem. For nonlinear inverse problems, techniques exist for many problems to estimate the model, but the tools for model appraisal are poorly developed. In fact, there is presently no proper theory that one can use for the assessment of infinite-dimensional models that are obtained from nonlinear inversion.

The inverse problems community is split into two parts. The theorists are mostly concerned with exact techniques to solve inverse problems. Valuable work is carried out in that field which has led to unexpected applications (such as the inverse scattering transform that is widely used for the solution of nonlinear evolution equations). However, these exact techniques often require data sets that are not realistic (infinitely many input data without noise), or they are in practice so unstable that they cannot easily be used for practical purposes. The other part of the inverse problems community makes progress in a more pragmatic way, constructing models from limited data sets while trying, at the same time, to understand the reliability of the models that are obtained. Unfortunately the different groups do not interact strongly. Such an interaction is very desirable. Nonlinear inversion using limited and noise-contaminated data is a problem of formidable proportions both from a theoretical as well as from a practical point of view. This area of research is of great

importance for those who actually carry out inversions, and it forms a major challenge for theorists. Fundamental progress in this field can only be expected when the two groups in the inverse problems community interact with each other.

Acknowledgments

Discussions with John Scales are very much appreciated. I thank Harm Dorren, Rob van der Hilst and Huub Douma for making figures available.

References

Backus G E and Gilbert J F 1967 Numerical applications of a formalism for geophysical inverse problems *Geophys. J. R. Astron. Soc.* **13** 247–76

——1968 The resolving power of gross Earth data *Geophys. J. R. Astron. Soc.* **16** 169–205

Borg G 1946 Eine Umkehrung der Sturm–Liouvillischen Eigenwertaufgabe. Bestimmung der Differentialgleichung durch die Eigenwerte *Acta Math.* **78** 1–96

Burridge R 1980 The Gel'fand–Levitan, the Marchenko, and the Gopinath–Sondi integral equations of inverse scattering theory, regarded in the context of the inverse impulse response problems *Wave Motion* **2** 305–23

Constable S C, Parker R L and Constable C G 1987 Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data *Geophys.* **52** 289–300

Deng H L 1997 A complexity analysis of generic optimization problems: Characterization of high-dimensional functions *PhD Thesis* (Colorado School of Mines) p 114

Dorren H J S, Muyzert E J and Snieder R K 1994 The stability of one-dimensional inverse scattering *Inverse Problems* **10** 865–80

Dorren H J S and Snieder R K 1997 Error propagation in nonlinear delay time tomography *Geophys. J. Int.* **128** 632–8

Douma H, Snieder R and Lomax A 1996 Ensemble inference in terms of emperical orthogonal functions *Geophys. J. Int.* **127** 363–78

Franklin J N 1970 Well-posed stochastic extensions of ill-posed linear problems *J. Math. Anal. Appl.* **31** 682–716

Gerver and Markushevitch 1966 Determination of a seismic velocity from a travel time curve *Geophys. J. R. Astron. Soc.* **11** 165–73

Gouveia W P and Scales J A 1997 Resolution of seismic waveform inversion: Bayes versus Occam *Inverse Problems* **13** 323–49

——1998 Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis *J. Geophys. Res.* **103** 2759–77

Herglotz G 1907 Über das Benndorfsche Problem für Fortplanzungsgeschwindigkeit der Eardbebensstralen *Z. Geophys.* **8** 145–7

Jackson J J and Matsu'ura M 1985 A Bayesian approach to nonlinear inversion *J. Geophys. Res.* **90** 581–91

Keller J B, Kay I and Shmoys J 1956 Determination of the potential from scattering data *Phys. Rev.* **102** 557–9

Kirkpatrick S, Gelatt C Jr and Vechhis M P 1983 Optimization by simulated annealing *Science* **220** 671–80

Lomax A and Snieder R 1995a Identifying sets of acceptable models to nonlinear geophysical inverse problems *Nonlinear Processes Geophys.* **2** 222–7

——1995b The contrast in upper mantle shear-velocity between the East European Platform and tectonic Europe obtained with a genetic algorithm of Rayleigh wave group dispersion *Geophys. J. Int.* **123** 169–82

Marchenko V A 1955 The construction of the potential energy from the phases of scattered waves *Dokl. Akad. Nauk* **104** 695–8

Mosegaard K 1998 Resolution analysis of general inverse problems through inverse Monte Carlo sampling *Inverse Problems* **14** 405–26

Mosegaard K and Tarantola A 1995 Monte Carlo sampling of solutions to inverse problems *J. Geophys. Res.* **100** 12 431–47

Newton R G 1981 Inversion of reflection data for layered media: A review of exact methods *Geophys. J. R. Astron. Soc.* **65** 191–215

——1989 *The Three-dimensional Inverse Problem* (New York: Springer)

Nolet G 1990 Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs *J. Geophys. Res.* **95** 8499–512

Nolet G, Grand S P and Kennett B L N 1994 Seismic heterogeneity in the upper mantle *J. Geophys. Res.* **99** 23 753–66

Parker 1994 *Geophysical Inverse Theory* (Princeton NJ: Princeton University)

Passier M L and Snieder R K 1995 Using differential waveform data to retrieve local S-velocity structure of path-averaged S-velocity gradients *J. Geophys. Res.* **100** 24 061–78

Passier M L, van der Hilst R D and Snieder R K 1997 Surface waveform inversions for local shear-velocity under eastern Australia *Geophys. Res. Lett.* **24** 1291–4

Prosser R T 1968 Formal solutions of inverse scattering problems *J. Math. Phys.* **10** 1819–22

Rothman D H 1985 Nonlinear inversion, statistical mechanics and residual statics estimation *Geophys.* **50** 2784–96

Rodberg LS and Thaler R M 1967 *Introduction to the Quantum Theory of Scattering* (New York: Academic)

Sabatier P C 1973 Discrete ambiguities and equivalent potentials *Phys. Rev.* A **8** 589–601

Sambridge M S 1990 Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D *Geophys. J. Int.* **102** 653–77

Sambridge M 1998 Exploring multidimensional landscapes without a map *Inverse Problems* **14** 427–40

Sambridge M and Drijkoningen G 1992 Genetic algorithms in seismic waveform inversion *Geophys. J. Int.* **109** 323–42

Scales J A and Snieder R 1997 To Bayes or not to Bayes? *Geophys.* **62** 1045–6

Sen M K and P L Stoffa 1992 Rapid sampling of model space using genetic algorithms: examples of seismic waveform inversion *Geophys. J. Int.* **108** 281–92

Snieder R 1990a A perturbative analysis of non-linear inversion *Geophys. J. Int.* **101** 545–56

——1990b The role of the Born approximation in nonlinear inversion *Inverse Problems* 247–66

——1991 An extension of Backus–Gilbert theory to nonlinear inverse problems *Inverse Problems* 409–33

Snieder R and Aldridge D F 1995 Perturbation theory for travel times *J. Acoust. Soc. Am.* **125** 796–812

Snieder R and Sambridge M 1993 The ambiguity in ray perturbation theory *J. Geophys. Res.* **98** 22 021–34

Tanimoto T and Anderson D L 1985 Lateral heterogeneity and azimuthal anisotropy of the upper mantle: Love and Rayleigh waves 100–250 *J. Geophys. Res.* **90** 1842–58

Tarantola A 1987 *Inverse Problem Theory* (Amsterdam: Elsevier)

Trampert J and Snieder R 1996 Model estimation biased by truncated expansions: possible artifacts in seismic tomography *Science* **271** 1257–60

van der Hilst R D and Kennett B L N 1998 Upper mantle structure beneath Australia from portable array deployments *American Geophysical Union 'Geodynamic series'* **26** 39–57

Weidelt P 1972 The inverse problem of geomagnetic induction *Z. Geophys.* **38** 257–89