

An extension of Backus–Gilbert theory to nonlinear inverse problems

Roel Snieder

Department of Theoretical Geophysics, University of Utrecht, PO Box 80.021,
3508 TA Utrecht, The Netherlands

Received 13 August 1990, in final form 16 December 1990

Abstract. The inverse problem where one wants to estimate a continuous model with infinitely many degrees of freedom from a finite data set is necessarily ill-posed. Although some examples exist of exact nonlinear inversion schemes for infinite data sets, there exists apart from data-fitting procedures no theory for nonlinear inversion that takes into account that a real data set is finite. A nonlinear perturbation theory is presented for the optimal determination of a model from a finite data set which generalizes Backus–Gilbert theory for linear inverse problems to include nonlinear effects. The extent to which the reconstructed model resembles the true model is described by linear and nonlinear resolution kernels. In this way, it is possible to evaluate to what degree the reconstructed model resembles the true model. A statistical analysis reveals the effects of errors in nonlinear inversion. The most dramatic effect is that if the data have no bias, the reconstructed model may suffer from a bias due to the nonlinearity of the problem. The theory is extended for the special case of infinite data sets which are of mathematical interest. As an example, it is shown that the Newton–Marchenko method for the inverse problem of the 3D Schrödinger equation requires a redundant data set, even if the nonlinearities are taken into account.

1. Introduction

The aim of inverse theory is twofold. In many areas of science and technology, such as quantum mechanics, seismology, non-destructive testing, one aims to get estimates of a model from a given data set. However, exact inverse problems are also intensively studied in mathematics and theoretical physics. There is a large discrepancy between the development of inverse theory for linear problems and for nonlinear problems. Linear inverse theory can use the powerful tools of linear algebra, and the linear inverse problem is well understood (e.g. Backus and Gilbert 1967, 1968, Franklin 1970, Jackson 1972, Wiggins 1972, Parker 1977a). In contrast to this, there exists no general treatment for nonlinear inverse problems. For some special nonlinear inverse problems where one has an infinite data set, exact inversion methods exist. Examples are the determination of the potential from the reflection coefficient of the one-dimensional Schrödinger equation using the Marchenko equation (Agranovich and Marchenko 1963, Chadan and Sabatier 1989), travel-time inversion using the Herglotz–Wiegert method in seismology (Aki and Richards 1980), or the inversion of Love-wave dispersion data (Takahashi 1955). In practical problems one frequently has to resort to optimization methods where one

minimizes the misfit between the measurements and synthetic data as a function of the desired model (Krappe and Lipperheide 1985, Tarantola 1987, Snieder and Tarantola 1989).

For general nonlinear inverse problems there is no systematic way to reconstruct the model. However, when the forward problem can be expanded in a regular perturbation series, it is possible to use perturbation theory for the inverse problem. This is shown in Snieder (1990a) for the Marchenko equation and some other examples in inverse theory, whereas in Snieder (1990b) a general perturbative treatment of nonlinear inverse problems is presented for inverse problems where the model can be reconstructed from the data in an exact and unique way. A major conclusion of these papers is that when the data are considered as a component that is related in a linear way to the model (such as the first Born approximation in quantum scattering theory) plus components that depend in a nonlinear way on the model, only the linear components give a non-zero contribution to the reconstruction of the model. In the process of inversion, the nonlinear components of the data are subtracted either explicitly or implicitly, so that these components give no net contribution to the reconstructed model. The reader is referred to Snieder (1990b) for details.

However, both exact nonlinear inversion methods and the perturbative treatment of Snieder (1990a, b) are unrealistic in the sense that they ignore the fact that a physical experiment always leads to a finite amount of data, whereas the desired model is in general an element of an infinite-dimensional function space. For linear inverse problems, this was recognized by Backus and Gilbert (1967), who showed that linear inverse problems have either no solution or infinitely many solutions. In the latter case one can only determine some linear functionals of the true model. Backus and Gilbert (1967, 1968) show how these functionals can be designed so that they represent local averages of the true model in an optimal way. Their theory also allows for the computation of resolution kernels that describe how the model functionals depend on the true model.

In this paper, the theories of Backus and Gilbert (1968, 1970), and Snieder (1990b) are synthesized to a formalism which leads to nonlinear model functionals that represent optimal local averages of the true model. This is achieved using perturbation analysis. The theory generalizes the concept of linear resolution kernels by using higher-order nonlinear resolution kernels. The first-order resolution kernels indicate to what extent the model functionals are local averages of the true model, whereas the higher-order resolution kernels indicate to what degree there is a nonlinear mapping from the true model to the model functionals. In the ideal case, the model functionals resemble the true model as well as possible, which implies that the nonlinear components of the model functionals should be as small as possible. An algorithm is presented in section 2 that achieves this in an optimal way.

In section 3, a statistical analysis is applied to the model functionals derived in section 2. It shows that the statistics are dramatically affected by the nonlinearity of the problem. In the linear theory of Backus and Gilbert (1970) there is a fundamental trade-off between variance and resolution. In nonlinear inverse problems this trade-off is also present; it is shown in section 3 that there is also a trade-off between variance and the removal of spurious nonlinear components in the model functionals. Although it is for most nonlinear inverse problems not possible to eliminate the spurious nonlinear mapping from the true model to the estimated model, it is shown in section 4 that for a special class of nonlinear inverse problems the nonlinearities can be removed completely from the data. A numerical example of the theory of the sections 2 and 3 is shown in section 5.

In section 6 it is shown how the formalism of section 2 can be generalized to inverse problems with an infinite data set, which arise in mathematical problems. Inverse scattering for the 3D Schrödinger equation is treated as an example. This example shows that the theory of this paper cannot only be used for model estimation, but that it can also be used to investigate the properties of exact nonlinear inverse methods. Throughout the paper the summation convention is used.

2. Derivation of the algorithm

Suppose we have a model $\varepsilon m(x)$, and that we have data d_i which are related to the model through some nonlinear relation:

$$d_i = G_i(\varepsilon m). \quad (1)$$

The parameter ε is attached to the model to facilitate a systematic perturbation analysis of the inverse problem. In the notation of this paper the model is a function of one space variable, extensions to higher dimensions involve merely a change of notation. The data kernels G_i are nonlinear functionals that produce the data for a given model. The model, the data and the data kernels may be complex. Note that the model is a function of the space variables, but that a discrete index i is used to enumerate the data. The reason for this is that for realistic inverse problems with real data, the amount of data is always finite. The inverse problem consists of the determination of the model from the data d_i .

A general solution of the inverse problem for arbitrary nonlinear data kernels G_i is not possible. However, if the forward problem has a regular perturbation expansion with non-zero first-order terms it is possible to derive a general algorithm for the inverse problem. Consider therefore the following perturbation expansion for the forward problem.

$$d_i = \varepsilon \int G_i^{(1)}(x)m(x) dx + \varepsilon^2 \int G_i^{(2)}(x_1, x_2)m(x_1)m(x_2) dx_1 dx_2 + \dots \quad (2a)$$

or

$$d_i = \sum_{n=1}^{\infty} \varepsilon^n \int G_i^{(n)}(x_1, \dots, x_n)m(x_1) \dots m(x_n) dx_1 \dots dx_n. \quad (2b)$$

An example of the expansion (2) is the Born series in quantum scattering theory (Rodberg and Thaler 1967). The reader is referred to section 6 for an example of the Born series for the Schrödinger equation in three dimensions.

It should be noted that the forward problem cannot always be expanded in a regular perturbation series. For examples, travel times of elastic waves in the Earth may be extremely sensitive to perturbations in the velocity structure. Near the triplication a minor change in the velocity field leads to catastrophic changes in the ray paths, a singular phenomenon (Zeeman 1977). In tunnelling experiments in quantum mechanics the transmission coefficient T is in the WKB approximation related to the potential V by

$$T = \exp\left(-2 \int_{x_L}^{x_R} (V(x) - E)^{1/2} dx\right) \quad (3)$$

where E is the energy and x_L and x_R are the turning points (Bender and Orszag, 1978). When $V \approx E$ this leads to a singular perturbation expansion of the forward problem. The theory of this paper is restricted to situations where the forward problem has a regular perturbation expansion of the form (2).

For a finite data set, it is impossible to infer the model in a unique way; the null-space is non-zero and one is forced to accept that it is not possible to reconstruct the model from the data in all detail. However, as in the theory of Backus and Gilbert (1968, 1970) for linear inverse problems, one can look for functionals that characterize certain properties of the true model. For example, given the orbital parameters of the Earth, it is impossible to infer the density structure in all detail, but there is a functional of the density model, the Earth's mass, that can be determined from the orbital parameters and which contains relevant information. As shown by Backus and Gilbert (1968, 1970), one can design model functionals that effectively are local smoothed averages of the true model. For the linear problems discussed by Backus and Gilbert, these model functionals are linear functionals. For the nonlinear inverse problems treated here, the model functionals are necessarily nonlinear functionals.

Since the data are nonlinear functionals of the true model, the inversion leads to functionals of the true model. Now let us consider a model functional $\hat{m}(x_0)$, which depends in a nonlinear way on the data. In general, this model functional can be expressed as a power series of the data:

$$\hat{m}(x_0) = a_i^{(1)}(x_0)d_i + a_{ij}^{(2)}(x_0)d_i d_j + \dots + a_{i_1 \dots i_n}^{(n)}(x_0)d_{i_1} d_{i_2} \dots d_{i_n} + \dots \quad (4)$$

The model functional is labelled with a variable x_0 to indicate that it represents a local estimate of the model around position x_0 . For any desired position x_0 one can determine the functional $\hat{m}(x_0)$. For arbitrary coefficients $a^{(n)}$ this model functional may not contain any useful information. The inverse problem consists of finding coefficients $a^{(n)}(x_0)$ in such a way that the model functional $\hat{m}(x_0)$ constitutes a local average of the true model $m(x)$ around $x = x_0$. Note that the subscripts of the $a^{(n)}$ can be permuted without changing the sum (4).

By inserting (2) in (4) one finds that the model functional and the true model are related by

$$\begin{aligned} \hat{m}(x_0, \varepsilon) = & \varepsilon a_i^{(1)}(x_0) \int G_i^{(1)}(x_1) m(x_1) dx_1 \\ & + \varepsilon^2 \int \{ a_{ij}^{(2)}(x_0) G_i^{(1)}(x_1) G_j^{(1)}(x_2) + a_i^{(1)}(x_0) G_i^{(2)}(x_1, x_2) \} m(x_1) m(x_2) dx_1 dx_2 \\ & + \dots \end{aligned} \quad (5a)$$

or in a more general notation

$$\begin{aligned} \hat{m}(x_0, \varepsilon) = & \varepsilon a_i^{(1)}(x_0) \int G_i^{(1)}(x_1) m(x_1) dx_1 + \sum_{n=2}^{\infty} \varepsilon^n \sum_{j=1}^n a_{q_1 \dots q_j}^{(j)}(x_0) \int \sum_{i_1 + \dots + i_j = n} G_{q_1}^{(1)}(x_1, \dots, x_{i_1}) \\ & \times G_{q_2}^{(i_2)}(x_{i_1+1}, \dots, x_{i_2}) \dots G_{q_j}^{(i_j)}(x_{i_{j-1}+1}, \dots, x_{i_n}) m(x_1) \dots m(x_n) dx_1 \dots dx_n. \end{aligned} \quad (5b)$$

Observe that the model functional $\hat{m}(x_0, \varepsilon)$ depends on ε in an arbitrary (nonlinear) way. The reason for this is that for a general nonlinear inverse problem with a finite data set there is no guarantee that even with optimal coefficients $a^{(n)}(x_0)$ the model functional $\hat{m}(x_0, \varepsilon)$ is equal to the true model $\varepsilon m(x_0)$ (which would imply a linear relation). The relation (5) between the model functional and the true model can be written as

$$\begin{aligned} \hat{m}(x_0, \varepsilon) = & \varepsilon \int R^{(1)}(x_0; x_1) m(x_1) dx_1 \\ & + \sum_{n=2}^{\infty} \varepsilon^n \int R^{(n)}(x_0; x_1, \dots, x_n) m(x_1) \dots m(x_n) dx_1 \dots dx_n \end{aligned} \quad (6)$$

with

$$R^{(1)}(x_0; x_1) = a_i^{(1)}(x_0)G_i^{(1)}(x_1) \tag{7a}$$

$$R^{(n)}(x_0; x_1, \dots, x_n) = \sum_{j=1}^n a_{q_1, \dots, q_j}^{(j)}(x_0) \sum_{i_1 + \dots + i_j = n} G_{q_1}^{(i_1)}(x_1, \dots, x_{i_1}) \dots G_{q_j}^{(i_j)}(x_{i_{j-1}+1}, \dots, x_n). \tag{7b}$$

The $R^{(n)}$ are the nonlinear generalization of the resolution kernels introduced by Backus and Gilbert (1968, 1970). The first-order resolution kernel $R^{(1)}$ describes to what degree there is a smearing in the linear mapping from the true model to the model functional. The higher-order resolution kernels describe to what extent there is a spurious nonlinear mapping from the true model to the model functional. In the ideal case, the model functional $\hat{m}(x_0, \varepsilon)$ represents the true model at $x = x_0$: $\hat{m}(x_0, \varepsilon) = \varepsilon m(x_0)$. This is only the case when

$$R^{(1)}(x_0; x_1) = \delta(x_0 - x_1) \tag{8a}$$

$$R^{(n)}(x_0; x_1, \dots, x_n) = 0. \tag{8b}$$

In this ideal situation, there is no blurring of the true model in the inversion (equation (8a)), and there is no spurious nonlinear mapping from the true model on the model functional (equation (8b)). These relations can be used to derive optimal values for the coefficients $a^{(n)}$.

The first-order coefficients $a^{(1)}(x_0)$ follow from the requirement that the first-order resolution kernel $R^{(1)}(x_0, x_1)$ resembles the delta function $\delta(x_0 - x_1)$ as closely as possible. The requirement (8a) and the relation (7a) are the same as in the linear theory of Backus and Gilbert (1968, 1970). These authors provide several methods to satisfy (8a) in an optimal way and discuss the trade-off between resolution and variance in great detail. The coefficients $a^{(1)}$ that lead to the minimum-norm solution are found by using the projection operator on the dual space of the $G_i^{(1)}$ as described in appendix B of Backus and Gilbert (1968). The solution is given by

$$a_i^{(1)}(x_0) = \Gamma_{ij}^{-1} G_j^{(1)*}(x_0) \tag{9}$$

where the Gram matrix Γ is defined by

$$\Gamma_{ij} = \int G_i^{(1)*}(x) G_j^{(1)}(x) dx. \tag{10}$$

In general one may assume that the first-order data kernels are independent. If they are not, one may form combinations of the data that are independent (Parker 1977a). This ensures that the Gram matrix Γ can be inverted. However, in many cases the Gram matrix is ill-conditioned because the data are nearly dependent and thus has a large condition number, which may give rise to large uncertainties in the inversion. In practice, one adds a damping term η_g and computes

$$\Gamma^{-g} \equiv (\Gamma + \eta_g I)^{-1}. \tag{11}$$

This regularizes the solution for the $a^{(1)}$ in (9). This damping term effectively regulates the trade-off between the resolution and the variance of the linear inversion; the reader is referred to Backus and Gilbert (1970) for details.

The solution (9) leads to resolution kernels with deep negative sidelobes (Backus and Gilbert 1968). One can avoid these sidelobes by determining the $a^{(1)}$ from the condition

that the following function is minimized (Backus and Gilbert 1970, Aki and Richards 1980)

$$K = 12 \int (x - x_0)^2 |R^{(1)}(x_0, x)|^2 dx \quad (12)$$

subject to the unimodularity constraint

$$\int R^{(1)}(x_0, x) dx = 1. \quad (13)$$

In practical inversions one adds a regularization term to (12) and minimizes

$$K_\eta = 12 \int (x - x_0)^2 |R^{(1)}(x_0, x)|^2 dx + \eta a_i^{(1)*} C_{ij}^d a_j^{(1)}. \quad (14)$$

The last term regulates the trade-off between resolution and variance, C_{ij}^d denotes the data covariance and η is the trade-off parameter. The coefficients $a_i^{(1)}$ that minimize (14) subject to the constraint (13) are (Aki and Richards 1980)

$$a_i^{(1)}(x_0) = \frac{W_{ij}^{-1}(x_0) u_j^*}{u_k W_{kj}^{-1}(x_0) u_j^*} \quad (15)$$

with

$$W_{ij}(x_0) = 12 \int (x - x_0)^2 G_i^{(1)*}(x) G_j^{(1)}(x) dx + \eta C_{ij}^d \quad (16)$$

and

$$u_i = \int G_i^{(1)}(x) dx. \quad (17)$$

In nonlinear inversion, the higher-order resolution kernels for the nonlinear terms should be as close as possible to zero, because one wants to minimize the spurious nonlinear mapping from the true model to the model functional. Just as with the linear terms $a^{(1)}$, one can determine optimal solutions to (8b) using different criteria. The number of coefficients $a^{(n)}$ that needs to be determined is staggering, and a simple recursive method is presented here for the computation of these coefficients, where (8b) is satisfied in an optimal way by minimizing the L_2 -norm of $R^{(n)}$:

$$\|R^{(n)}\|_2^2 = \int |R^{(n)}(x_0; x_1, \dots, x_n)|^2 dx_1 \dots dx_n. \quad (18)$$

Inserting (7b) in (18) one finds that in the minimization criterion for $\|R^{(n)}\|_2^2$, the coefficients $a^{(j)}$ are present for $j = 1, \dots, n$. However, suppose that the $a^{(j)}$ are known for $j = 1, \dots, n-1$, then one can use (18) for the determination of the $a^{(n)}$. This can be done by inserting (7b) in (18), and by splitting the sum from $j = 1$ to n in a sum from $j = 1$ to $n-1$ and a term $j = n$. Ignoring terms that do not contain the $a^{(n)}$, and which are therefore irrelevant for the minimization of $\|R^{(n)}\|_2^2$ with respect to the $a^{(n)}$, gives

$$\begin{aligned} \|R^{(n)}\|_2^2 = \int dx_1 \dots dx_n & \left\{ \sum_{j=1}^{n-1} a_{q_1 \dots q_j}^{(j)}(x_0) \sum_{i_1 + \dots + i_j = n} G_{q_1}^{(i_1)}(x_1, \dots, x_{i_1}) \dots G_{q_j}^{(i_j)}(x_{i_{j-1}+1}, \dots, x_n) \right. \\ & \times a_{p_1 \dots p_n}^{(n)*}(x_0) G_{p_1}^{(1)*}(x_1) \dots G_{p_n}^{(1)*}(x_n) \\ & + \sum_{j=1}^{n-1} a_{q_1 \dots q_j}^{(j)*}(x_0) \sum_{i_1 + \dots + i_j = n} G_{q_1}^{(i_1)*}(x_1, \dots, x_{i_1}) \dots G_{q_j}^{(i_j)*}(x_{i_{j-1}+1}, \dots, x_n) \\ & \times a_{p_1 \dots p_n}^{(n)}(x_0) G_{p_1}^{(1)}(x_1) \dots G_{p_n}^{(1)}(x_n) \\ & \left. + a_{q_1 \dots q_n}^{(n)}(x_0) a_{p_1 \dots p_n}^{(n)*}(x_0) G_{q_1}^{(1)}(x_1) \dots G_{q_n}^{(1)}(x_n) G_{p_1}^{(1)*}(x_1) \dots G_{p_n}^{(1)*}(x_n) \right\}. \quad (19) \end{aligned}$$

Differentiating this expression with respect to $a_{r_1 \dots r_n}^{(n)*}$ and carrying out the integration $\int dx_1 \dots dx_n$ gives

$$\sum_{j=1}^{n-1} \sum_{i_1 + \dots + i_j = n} \Gamma_{r_1 \dots r_{i_1} q_1}^{(i_1)} \dots \Gamma_{r_{j-1} + 1 \dots r_n q_j}^{(i_j)} a_{q_1 \dots q_j}^{(j)}(x_0) + \Gamma_{r_1 q_1} \dots \Gamma_{r_n q_n} a_{q_1 \dots q_n}^{(n)}(x_0) = 0. \quad (20)$$

In this expression the Gram matrix Γ is defined in (10). The generalized Gram matrices $\Gamma^{(n)}$ are defined by

$$\Gamma_{p_1 \dots p_n q}^{(n)} \equiv \int G_{p_1}^{(1)*}(x_1) \dots G_{p_n}^{(1)*}(x_n) G_q^{(n)}(x_1, \dots, x_n) dx_1 \dots dx_n. \quad (21)$$

Equation (20) is a linear equation for the $a^{(n)}$ that can be solved efficiently by premultiplying n times with Γ^{-g} on the left, this gives

$$a_{p_1 \dots p_n}^{(n)}(x_0) = -\Gamma_{p_1 r_1}^{-g} \dots \Gamma_{p_n r_n}^{-g} \sum_{j=1}^{n-1} \sum_{i_1 + \dots + i_j = n} \Gamma_{r_1 \dots r_{i_1} q_1}^{(i_1)} \dots \Gamma_{r_{j-1} + 1 \dots r_n q_j}^{(i_j)} a_{q_1 \dots q_j}^{(j)}(x_0). \quad (22)$$

Note that (22) follows from (20) when $\Gamma^{-g} = \Gamma^{-1}$ (hence when $\eta_g = 0$). For non-zero values for η_g (22) does not lead exactly to the minimum of (19).

It can be seen explicitly that the $a^{(n)}$ can be computed once the $a^{(j)}$ ($j = 1, \dots, n - 1$) are known, so that these coefficients can be computed recursively. Note that it is necessary to compute the inverse Γ^{-g} of the first-order Gram matrix (11) once, but that it is not necessary to invert a different matrix for every order n . This is important, because the number of coefficients increases rapidly with the order n of the nonlinearity. Once the $a^{(n)}$ are determined, the desired model functional can be computed from (4). The non-linear resolution kernels can be obtained from (7b).

One can show that when the true model has a finite L_2 -norm and is Lipschitz continuous the model functional $\hat{m}(x_0)$ approaches the true model $m(x_0)$ when the width (12) of the first-order resolution kernel and the L_2 -norm of the higher-order resolution kernels go to zero. To see this, define the following norm in model space:

$$\|m\|_{x_0}^2 \equiv \int \left| \frac{m(x) - m(x_0)}{x - x_0} \right|^2 dx. \quad (23)$$

Using the unimodularity constraint (13) one can write (6) as

$$\begin{aligned} \hat{m}(x_0) - m(x_0) &= \int (m(x) - m(x_0)) R^{(1)}(x_0; x) dx \\ &+ \sum_{n=2}^{\infty} \int R^{(n)}(x_0; x_1, \dots, x_n) m(x_1) \dots m(x_n) dx_1 \dots dx_n. \end{aligned} \quad (24)$$

With the help of Schwartz inequality one can derive that

$$\begin{aligned} |\hat{m}(x_0) - m(x_0)|^2 &\leq \frac{1}{12} \|m\|_{x_0}^2 K \\ &+ \frac{1}{8} \|m\|_{x_0} \sum_{n=2}^{\infty} \|R^{(n)}\|_2 \|m\|_2^n + \left(\sum_{n=2}^{\infty} \|R^{(n)}\|_2 \|m\|_2^n \right)^2. \end{aligned} \quad (25)$$

This implies that when (12) and (18) become smaller the model functional $\hat{m}(x_0)$ resembles the true model in a better way. Note that we tacitly assumed here that for a given data set one can find higher-order resolution kernels which have a sufficiently small L_2 -norm for the series $\sum_{n=2}^{\infty} \|R^{(n)}\|_2 \|m\|_2^n$ to converge. Whether this condition is satisfied depends both on the data kernels that are at our disposal (because these quantities determine the $R^{(n)}$), and on the model norm $\|m\|_2$ (because this determines the strength of the nonlinear effects).

3. The propagation of errors in the inversion

In a realistic experiment, the data are contaminated with errors, and it is important to know how the model functional is affected by these errors. Very little is known about the propagation of errors in nonlinear inversion. The perturbative treatment of section 2 allows for the estimation of errors in the model functional. Let the data d_i be perturbed with errors Δd_i . The perturbation in the estimated model follows from (4)

$$\Delta \hat{m}(x_0) = \sum_{n=1}^{\infty} a_{i_1 \dots i_n}^{(n)}(x_0) \{(d_{i_1} + \Delta d_{i_1}) \dots (d_{i_n} + \Delta d_{i_n}) - d_{i_1} \dots d_{i_n}\}. \quad (26)$$

It follows that the n th-order term in this sum depends on products of the errors Δd in powers up to the n th-order. For a statistical analysis of the model functional it is therefore necessary to prescribe the higher moments of the distribution of the errors. For realistic data, it is usually difficult to determine the mean and the variance of the data; estimating higher-order moments such as the skewness and kurtosis is virtually impossible.

In order to obviate the need for the higher-order moments, it is assumed here that the errors in the data are much smaller than the data themselves:

$$|\Delta d_i| \ll |d_i|. \quad (27)$$

It is then realistic to take only the first- and second-order effects of the data errors into account. Using the fact that the subscripts of the $a^{(n)}$ can be permuted and that a sum over the subscripts is implied, one finds that

$$\Delta \hat{m}(x_0) = \sum_{n=1}^{\infty} a_{i_1 \dots i_n}^{(n)}(x_0) \{n(\Delta d_{i_1})d_{i_2} \dots d_{i_n} + n(n-1)(\Delta d_{i_1})(\Delta d_{i_2})d_{i_3} \dots d_{i_n}\} + O(\Delta d)^3. \quad (28)$$

Let an overbar denote the statistical average. If the data errors have zero mean ($\overline{\Delta d_i} = 0$), the expectation value of the error in the model functional is given by

$$\overline{\Delta \hat{m}(x_0)} = \sum_{n=1}^{\infty} n(n-1)a_{i_1 \dots i_n}^{(n)}(x_0)C_{i_1 i_2}^d d_{i_3} \dots d_{i_n} + O(\Delta d)^3 \quad (29)$$

where C^d is the data covariance:

$$C_{ij}^d = \overline{\Delta d_i \Delta d_j}. \quad (30)$$

This means that if the data have no bias ($\overline{\Delta d_i} = 0$), the model functional may suffer from a non-zero bias because of the nonlinear terms in expression (29). For linear inverse problems there is no bias in the reconstructed model if the data are unbiased; this is reflected in (29) by the fact that the linear term $n = 1$ is equal to zero. Note that the magnitude of the bias in the model functional depends both on the data covariance and on the data themselves. This is due to the fact that the model bias is a purely nonlinear effect. If the nonlinearity is stronger, and hence if the magnitude of the d_i is larger, the bias is also stronger.

The covariance of the model functional follows from (28). Up to second order in the data errors the model covariance is given by:

$$\overline{\Delta \hat{m}(x) \Delta \hat{m}(x')} = \sum_{n=1}^{\infty} \sum_{r=1}^{\infty} nr a_{i_1 \dots i_n}^{(n)}(x) a_{j_1 \dots j_r}^{(r)}(x') d_{i_2} \dots d_{i_n} d_{j_2} \dots d_{j_r} C_{i_1 j_1}^d + O(\Delta d)^3. \quad (31)$$

In contrast to linear inversion, the model covariance depends not only on the errors in the data, but also on the data themselves. (This is only the case for the nonlinear

terms (n and $r \geq 2$). This can be understood as follows. It is shown in Snieder (1990a,b) that in nonlinear inversion only the linear part of the data gives a non-zero contribution to the reconstruction of the model. In the process of nonlinear inversion the nonlinear components in the data are being subtracted. If the data are larger, and the nonlinearity is stronger, this subtraction is more important. However, as shown in Snieder (1990b), this process of subtraction can be highly unstable. This means that for a given data covariance, the model covariance is larger when the nonlinearity is stronger.

It is shown by Backus and Gilbert (1970) that the parameter η in (14) regulates the trade-off between the width of the first-order resolution kernel and the variance of the model functional. It follows from the results of Backus and Gilbert (1970) that increasing the trade-off parameter η leads to decreasing values of the magnitude of the coefficients $a^{(1)}$ because the last term in (14) penalizes for solutions with large values of the sum $a_i^{(1)*} C_{ij}^d a_j^{(1)}$. It follows from (22) that all higher-order coefficients $a^{(n)}$ can be expressed as a linear combination of the first-order coefficients $a^{(1)}$, i.e. by applying (22) recursively one can write the $a^{(n)}$ in the form

$$a_{p_1 \dots p_n}^{(n)}(x_0) = B_{p_1 \dots p_n q_1 \dots q_n} a_{q_1}^{(1)}(x_0) \dots a_{q_n}^{(1)}(x_0). \tag{32}$$

This implies that

$$|a_{p_1 \dots p_n}^{(n)}(x_0)|^2 \leq \left(\sum_{q_1, \dots, q_n} |B_{p_1 \dots p_n q_1 \dots q_n}|^2 \right) \left(\sum_q |a_q^{(1)}(x_0)|^2 \right)^n. \tag{33}$$

An upper bound on the $|a^{(n)}|^2$ therefore decreases when the L_2 -norm of the coefficients $a^{(1)}$ decreases. One can conclude from this that increasing the trade-off parameter η decreases the upper bound on $|a_{p_1 \dots p_n}^{(n)}(x_0)|$. Because of (31) this implies that an upper bound on the nonlinear variance decreases when the trade-off parameter η is increased. Note that this does not imply that the nonlinear variance goes to zero in a uniform way when η is increased. However, once the coefficients $a^{(n)}(x_0)$ are computed for the determination of the model functional it requires little numerical effort to compute the nonlinear variance from (31) and to investigate the effect of η on the nonlinear variance numerically.

When using either of the solutions (9) or (15) for the $a_i^{(1)}$, one needs the (generalized) inverse Γ^{-g} of the Gram matrix for the computation of the higher-order $a^{(n)}$. Increasing the value of the damping parameter η_g has two effects. First, since Γ is a positive definite matrix, one increases all eigenvalues of $(\Gamma + \eta_g I)$ by increasing η_g . This means that the norm $\|\Gamma^{-g}\|_2 = \|(\Gamma + \eta_g I)^{-1}\|_2$ decreases when η_g is increasing. From (22) one finds that this implies that an upper bound on $\sum_{p_1, \dots, p_n} |a_{p_1 \dots p_n}^{(n)}|^2$ decreases when η_g increases. With (31) this implies that the upper bound on the nonlinear variance decreases for increasing η_g . Again, the nonlinear variance need not decrease uniformly with increasing η_g , and numerical computations are needed to establish the detailed effects of η_g on the nonlinear variance.

Second, the L_2 -norm $\|R^{(n)}\|_2$ of the nonlinear resolution kernels is minimized for coefficients $a^{(n)}$ that satisfy (20). Premultiplying (20) n times with Γ^{-1} gives the $a^{(n)}$ that minimize $\|R^{(n)}\|_2$. If one premultiplies (20) with Γ^{-g} (with $\eta_g > 0$) one obtains different values for the $a^{(n)}$, these necessarily do not lead to the minimum of $\|R^{(n)}\|_2$, because the minimum is obtained by premultiplying (20) with Γ^{-1} . This means that taking $\eta_g > 0$ leads to an increased value of $\|R^{(n)}\|_2$ and hence to a suboptimal subtraction of the non-linear components in the data.

This implies that the parameter η_g regulates a trade-off between the upper bound of the nonlinear variance, and the nonlinear resolution. Since the nonlinear variance may

vary with η_g in a nonuniform way one needs to establish this trade-off numerically. Once the $a^{(n)}$ are computed this can be done with little computational effort.

4. The removability of the nonlinear terms in the inversion

In an ideal nonlinear inversion, there is no smearing in the linear part of the inversion (8a), and there is no spurious nonlinear mapping from the true model to the model functional (8b). The condition (8a) can be investigated with standard Backus–Gilbert theory (1968, 1970). The issue of the removal of the nonlinearities in the inversion (8b) is difficult to answer in general. However, it is shown in this section that for the special case that:

- (a) a unique inverse of the Gram matrix exists,
- and
- (b) the nonlinear data kernels can be expressed as a linear combination of the linear data kernels,

there is no spurious nonlinear mapping from the true model to the model functional.

To see this, split the sum (7b) in a sum $j = 1, \dots, n-1$ and a term $j = n$:

$$R^{(n)}(x_0; x_1, \dots, x_n) = \sum_{j=1}^{n-1} a_{q_1 \dots q_j}^{(j)}(x_0) \sum_{i_1 + \dots + i_j = n} G_{q_1}^{(i_1)}(x_1, \dots, x_{i_1}) \dots G_{q_j}^{(i_j)}(x_{i_{j-1}+1}, \dots, x_n) + a_{q_1 \dots q_n}^{(n)}(x_0) G_{q_1}^{(1)}(x_1) \dots G_{q_n}^{(1)}(x_n). \quad (34)$$

Inserting (22) in the last term leads to

$$R^{(n)}(x_0; x_1, \dots, x_n) = \sum_{j=1}^{n-1} \sum_{i_1 + \dots + i_j = n} \{G_{q_1}^{(i_1)}(x_1, \dots, x_{i_1}) \dots G_{q_j}^{(i_j)}(x_{i_{j-1}+1}, \dots, x_n) - \Gamma_{p_1 r_1}^{-g} \dots \Gamma_{p_n r_n}^{-g} \Gamma_{r_1 \dots r_{i_1} q_1}^{(i_1)} \dots \Gamma_{r_{i_{j-1}+1} \dots r_n q_j}^{(i_j)} G_{p_1}^{(1)}(x_1) \dots G_{p_n}^{(1)}(x_n)\} a_{q_1 \dots q_j}^{(j)}(x_0). \quad (35)$$

The extent to which the nonlinearities can be removed depends on the term within brackets in (35). Multiplying (35) with $G_{s_1}^{(1)*} \dots G_{s_n}^{(1)*}$, integrating over x_1, \dots, x_n , and using (21) gives:

$$\int G_{s_1}^{(1)*}(x_1) \dots G_{s_n}^{(1)*}(x_n) R^{(n)}(x_0; x_1, \dots, x_n) dx_1 \dots dx_n = \sum_{j=1}^{n-1} \sum_{i_1 + \dots + i_j = n} \{ \Gamma_{s_1 \dots s_{i_1} q_1}^{(i_1)} \dots \Gamma_{s_{i_{j-1}+1} \dots s_n q_j}^{(i_j)} - (\Gamma^{-g} \Gamma)_{r_1 s_1} \dots (\Gamma^{-g} \Gamma)_{r_n s_n} \Gamma_{r_1 \dots r_{i_1} q_1}^{(i_1)} \dots \Gamma_{r_{i_{j-1}+1} \dots r_n q_j}^{(i_j)} \} a_{q_1 \dots q_j}^{(j)}(x_0). \quad (36)$$

If the inverse of the Gram matrix exists ($\Gamma^{-g} \Gamma = I$) the right-hand side vanishes:

$$\int G_{r_1}^{(1)*}(x_1) \dots G_{r_n}^{(1)*}(x_n) R^{(n)}(x_0; x_1, \dots, x_n) dx_1 \dots dx_n = 0. \quad (37)$$

This means that the projection of the nonlinear resolution kernels $R^{(n)}$ on the linear data kernels is equal to zero; only parts of the $R^{(n)}$ outside the subspace of the linear data kernels $G_i^{(1)}(x)$ can be non-zero.

It follows from (7b), that the nonlinear resolution kernel $R^{(n)}$ is a combination from the nonlinear data kernels $G^{(j)}$, with $j = 1, \dots, n$. Now suppose that the nonlinear data

kernels are contained in the subspace of the linear data kernels, i.e. that they can be expanded with expansion coefficients α as

$$G_{q_1 \dots q_j}^{(j)}(x_1 \dots x_j) = \alpha_{q_1 r_1}^{(j)} \dots \alpha_{q_j r_j}^{(j)} G_{r_1}^{(1)}(x_1) \dots G_{r_j}^{(1)}(x_j) \quad j = 1, \dots, n. \quad (38)$$

In that case, because of (7b), the nonlinear resolution kernel $R^{(n)}$ has no components outside the subspace spanned by the linear data kernels. With (37) this implies that the nonlinear resolution kernel vanishes: $R^{(n)} = 0$, so that the model functional depends in a linear way on the true model. (The nonlinearities are completely subtracted in the inversion.)

Note that it is tacitly assumed here that the data kernels can be written as a direct sum of linear data kernels and nonlinear data kernels; i.e. that this decomposition is unique. The data kernels follow from a perturbation analysis of the forward problem. When for a particular problem a regular perturbation series of the form (2) can be derived in an unambiguous way, the decomposition of the data kernels in linear components and nonlinear components is unique.

In section 5 an example is shown where the second-order data kernel $G_n^{(2)}(x_1, x_2)$ is not contained in the subspace spanned by the combination $G_i^{(1)}(x_1)G_j^{(1)}(x_2)$, so that (38) cannot be satisfied. This leads to a non-zero spurious second-order mapping from the true model to the model functional. As a simple example of a finite data set where (38) is satisfied consider data defined by

$$d_i = f\left(\int H_i(x)m(x) dx\right) \quad (39)$$

where the $H_i(x)$ are linear kernels and where $f(\xi)$ is a nonlinear function with a Taylor expansion

$$f(\xi) = f_1 \xi + f_2 \xi^2 + \dots \quad (40)$$

It is assumed that the first-order term f_1 is non-zero. Inserting (40) in (39) leads in the notation of (2) to the following data kernels:

$$G_i^{(n)}(x_1, \dots, x_n) = f_n H_i(x_1)H_i(x_2) \dots H_i(x_n). \quad (41)$$

The first-order data kernel is given by $G_i^{(1)}(x) = f_1 H_i(x)$, and one readily verifies that the condition (38) for complete subtraction of nonlinearities in the inversion is satisfied. The fact that in this example the nonlinearities can be completely removed in the inversion is not surprising. Applying the inverse function f^{-1} to (39) and defining transformed data by $\tilde{d}_i = f^{-1}(d_i)$ one finds that

$$\tilde{d}_i = \int H_i(x)m(x) dx. \quad (42)$$

This is a linear problem, the application of f^{-1} produced a global linearization of the inverse problem. When using the perturbative inversion (4) one performs this linearization by implicitly subtracting out the nonlinearities that are present in the original data d_i . For the inverse problem of (39) one can determine the coefficients $a^{(n)}$ from the transformed linear problem (42). The Backus–Gilbert solution of the linearized problem (42) is given by

$$\hat{m}(x_0) = a_i^{(1)}(x_0)\tilde{d}_i = a_i^{(1)}(x_0)(h_1 d_i + h_2 d_i^2 + \dots) \quad (43)$$

where the coefficients h_n are Taylor expansion coefficients of the inverse of $f: f^{-1}(\mu) = h_1 \mu + h_2 \mu^2 + \dots$. Since we know that the application of f^{-1} removed the nonlinearity completely, we are assumed that (43) is the optimal inverse series of the form (4); hence $a_{j_1 \dots j_n}^{(n)}(x_0) = h_n a_i^{(1)}(x_0) \delta_{i,j_1} \delta_{i,j_2} \dots \delta_{i,j_n}$.

5. A numerical example

In this section an example is shown of the theory for nonlinear inversion. Consider a one-dimensional string of length L with fixed endpoints. The mass-density $\rho(x)$ varies along the string. The differential equation for the modes of vibration of the string is given by

$$u_{xx}(x) + \frac{\rho(x)\omega^2}{T}u(x) = 0. \quad (44)$$

In this expression T is the tension in the string. The condition that the endpoints are fixed implies that

$$u(0) = u(L) = 0. \quad (45)$$

These boundary condition are only satisfied for discrete frequencies ω_n . In the inverse problem that is considered here, the aim of the inversion is to determine the unknown density distribution $\rho(x)$, using a finite set of eigenfrequencies ω_n .

The first step is formulate a perturbation expansion of the form (2) for the forward problem. Consider a perturbation from a homogeneous reference string with constant mass-density ρ_0 . The model $\varepsilon m(x)$ is defined by

$$\rho(x) = \rho_0(1 + \varepsilon m(x)). \quad (46)$$

The eigenfunctions of the homogeneous reference string are given by

$$u_n^{(0)}(x) = \left(\frac{2}{L}\right)^{1/2} \sin\left(\frac{n\pi x}{L}\right) \quad (47)$$

and the unperturbed eigenfrequencies satisfy

$$\omega_n^{(0)} = \left(\frac{T}{\rho_0}\right)^{1/2} \left(\frac{n\pi}{L}\right). \quad (48)$$

The data d_n for this problem are the relative frequency shifts defined by

$$d_n \equiv \frac{\omega_n^2 - \omega_n^{(0)2}}{\omega_n^{(0)2}}. \quad (49)$$

A perturbation expansion of (49) can be formulated using non-degenerate Rayleigh-Schrödinger perturbation theory (e.g. Morse and Feshbach, 1953) and leads to the following perturbation expansion:

$$\omega_n^2 = \omega_n^{(0)2} - \varepsilon \omega_n^{(0)2} M_{nn} + \varepsilon^2 \left(\sum_{m \neq n} \frac{\omega_n^{(0)4}}{\omega_n^{(0)2} - \omega_m^{(0)2}} M_{nm}^2 + \omega_n^{(0)2} M_{nn}^2 \right) + \mathcal{O}(\varepsilon^3) \quad (50)$$

In this expression the matrix elements M_{nm} are defined by

$$M_{nm} = \int_0^L u_n^{(0)}(x) m(x) u_m^{(0)}(x) dx \quad (51)$$

with the eigenfunctions given by (47). Inserting (50) in the definition (49) of the data one finds that

$$d_n = -\varepsilon M_{nn} + \varepsilon^2 \left(\sum_{m \neq n} \frac{n^2}{n^2 - m^2} M_{nm}^2 + M_{nn}^2 \right) + \mathcal{O}(\varepsilon^3). \quad (52)$$

Using the definition (51) of the matrix elements M_{nm} it follows that the first- and second-order data kernels defined in (2) are given by

$$G_n^{(1)}(x) = -\frac{2}{L} \sin^2\left(\frac{n\pi x}{L}\right) \quad (53)$$

and

$$G_n^{(2)}(x_1, x_2) = \sum_{m \neq n} \frac{4}{L^2} \frac{n^2}{n^2 - m^2} \sin\left(\frac{n\pi x_1}{L}\right) \sin\left(\frac{m\pi x_1}{L}\right) \sin\left(\frac{n\pi x_2}{L}\right) \sin\left(\frac{m\pi x_2}{L}\right) + \frac{4}{L^2} \sin^2\left(\frac{n\pi x_1}{L}\right) \sin^2\left(\frac{n\pi x_2}{L}\right). \quad (54)$$

It is instructive to consider first the accuracy of the second-order expansion (52). As a test example, consider the situation where a point mass is attached to the string in $x = x_0$:

$$m(x) = m_0 \delta(x - x_0) \quad (55)$$

where m_0 is the ratio of the perturbing point mass to the mass of the unperturbed string. For this simple example, the eigenfrequencies obey the following transcendental equation:

$$m_0 k \sin[k(L - x_0)] \sin(kx_0) = \sin(kL) \quad \text{with } k = \sqrt{\frac{\rho_0}{T}} \omega. \quad (56)$$

The exact frequency perturbations, and the first- and second-order approximations are for the three gravest modes shown in figures 1(a, b, c) when the point mass is attached to the point $x_0 = 0.25$. (in the examples of this section the string is of unit length; $L = 1$.) For the fundamental mode ($n = 1$) the fully nonlinear frequency shift and the first- and second-order approximations are undistinguishable for mass perturbations less than 15%. For the modes 2 and 3 (figures 1(b, c)) the first-order approximation is only accurate for perturbations of a few percent, whereas the second-order approximation cannot be distinguished from the exact frequency shift for mass perturbations up to 8%. For larger mass perturbations the third- and higher-order effects are important in the forward problem. In this section, only first- and second-order effects in the expansion (4) for the inverse problem are taken into account. The perturbation expansions should therefore be regarded as asymptotic relations. Whether the obtained accuracy is acceptable or not depends on the data errors, the required accuracy and on the magnitude of the model perturbation (hence on the degree of nonlinearity). Extending the example of this section to higher than second order poses no conceptual problems.

In order to perform the nonlinear inversion to second order one needs to know the first- and second-order Gram matrices. From the definition (10) and the data kernels (53) one finds that

$$\Gamma_{nm} = \frac{1}{L} (1 + \frac{1}{2} \delta_{nm}). \quad (57)$$

The generalized inverse of this matrix is given by

$$\left(\Gamma + \frac{\eta_g}{L} I\right)_{nm}^{-1} = \frac{-2L}{(1 + 2\eta_g)(N + \frac{1}{2} + \eta_g)} + \frac{2L}{(1 + 2\eta_g)} \delta_{nm} \quad (58)$$

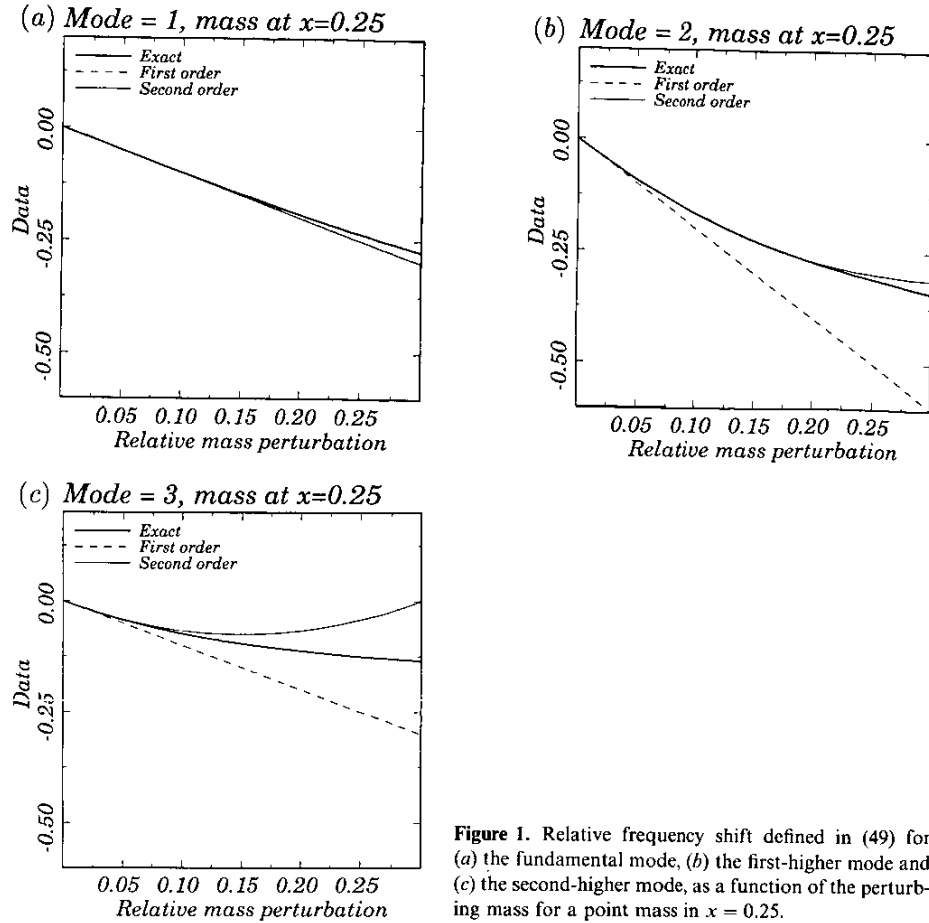


Figure 1. Relative frequency shift defined in (49) for (a) the fundamental mode, (b) the first-higher mode and (c) the second-higher mode, as a function of the perturbing mass for a point mass in $x = 0.25$.

where N is the number of data. The second-order Gram matrix follows by inserting (53) and (54) in (21) and performing the trigonometric integrations:

$$\begin{aligned} \Gamma_{kln}^{(2)} = & \frac{1}{4L^2} \sum_{m \neq n} \delta_{kl} (\delta_{m,|n-2k|} + \delta_{m,n+2k}) \frac{n^2}{n^2 - m^2} \\ & - \frac{i}{4L^2} \sum_{m \neq n} (\delta_{n,k+1} \delta_{m,|k-l|} + \delta_{n,|k-l|} \delta_{m,k+1}) \frac{n^2}{n^2 - m^2} \\ & + \frac{1}{L^2} (1 + \frac{1}{2} \delta_{nk} + \frac{1}{2} \delta_{nl} + \frac{1}{4} \delta_{nk} \delta_{nl}). \end{aligned} \tag{59}$$

Because of the delta functions in the right-hand side the summation over the modes m can be performed trivially.

In the inversions shown in this section the frequency shifts of the four gravest modes ($n = 1, 2, 3, 4$) are used to determine the estimated model at position $x_0 = 0.25$. The first-order resolution kernel obtained from the minimization of (14) is shown in figure 2. The resolution kernel has a finite width, for the undamped case ($\eta = 0$) this width is determined by the highest eigenfunction that is taken into account ($n = 4$ in this

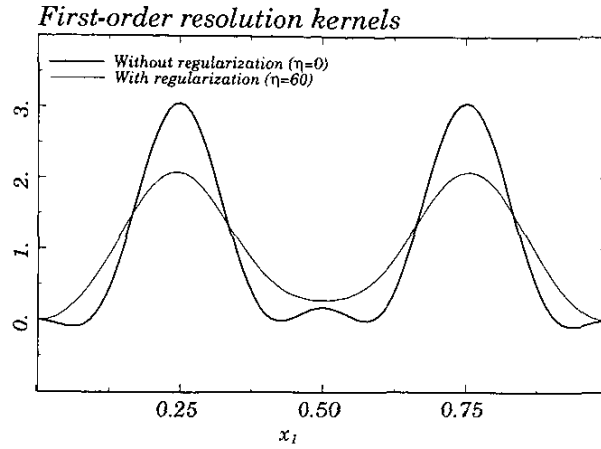


Figure 2. First-order resolution kernels $R^{(1)}(x_0; x_1)$ for $x_0 = 0.25$ for $\eta = 0$ (thick line) and $\eta = 60$ (thin line).

case). Using the solution (9) instead of (15) leads to a resolution kernel with deep negative sidelobes; this solution is not shown here.

The most pronounced aspect of the resolution kernel is the symmetry around the centre of the string. When one has recorded the eigenfrequencies of a string subject to the boundary conditions (45) one cannot determine the structure of the string in a unique way because reflecting the string around its centre leads to a string with the same eigenfrequencies as the original string. This ambiguity holds of course also for the nonlinear inverse problem. It was recognized by Borg (1946) that in order to determine the structure of the string uniquely one also needs to know the eigenfrequencies for other boundary conditions than (45). For example, if one also knows the boundary conditions when one side of the string is open ($u_x(L) = 0$) then the symmetry between the two halves of the string is broken and one can recover the non-symmetric part of the model. The resolution kernels in figure 2 reflect this symmetry property by the peak in the resolution kernel at the location $x_1 = 0.75$. Since it is known that one cannot remove this peak, the integration over x in (14) was computed only over the left half of the string ($0 < x < 0.5$).

According to (5a) or (7b) the second-order resolution kernel contains two terms; a term $a_i^{(1)}G_i^{(2)}(x_1, x_2)$ which describes the mapping of the quadratic nonlinearity by the linear estimators $a_i^{(1)}$, and a term $a_{ij}^{(2)}G_i^{(1)}(x_1)G_j^{(1)}(x_2)$ which serves to remove this quadratic nonlinearity as well as possible. The first term is shown in figure 3(a) (In the numerical example of this section, the value $x_0 = 0.25$ is used.) Suppose one would ignore the nonlinearity of the inverse problem and perform a linear inversion. This amounts to using only $a_i^{(1)}$ and setting all the $a^{(n)}$ equal to zero for $n \geq 2$. In that case the model functional $\hat{m}(x_0)$ is given by

$$\begin{aligned} \hat{m}(x_0, \varepsilon) = & \varepsilon a_i^{(1)}(x_0) \int G_i^{(1)}(x_1) m(x_1) dx_1 \\ & + \varepsilon^2 \int a_i^{(1)}(x_0) G_i^{(2)}(x_1, x_2) m(x_1) m(x_2) dx_1 dx_2 + \dots \end{aligned} \quad (60)$$

The function in figure 3(a) therefore shows the second-order mapping for the problem of the inhomogeneous string in case one performs a linear inversion without bothering

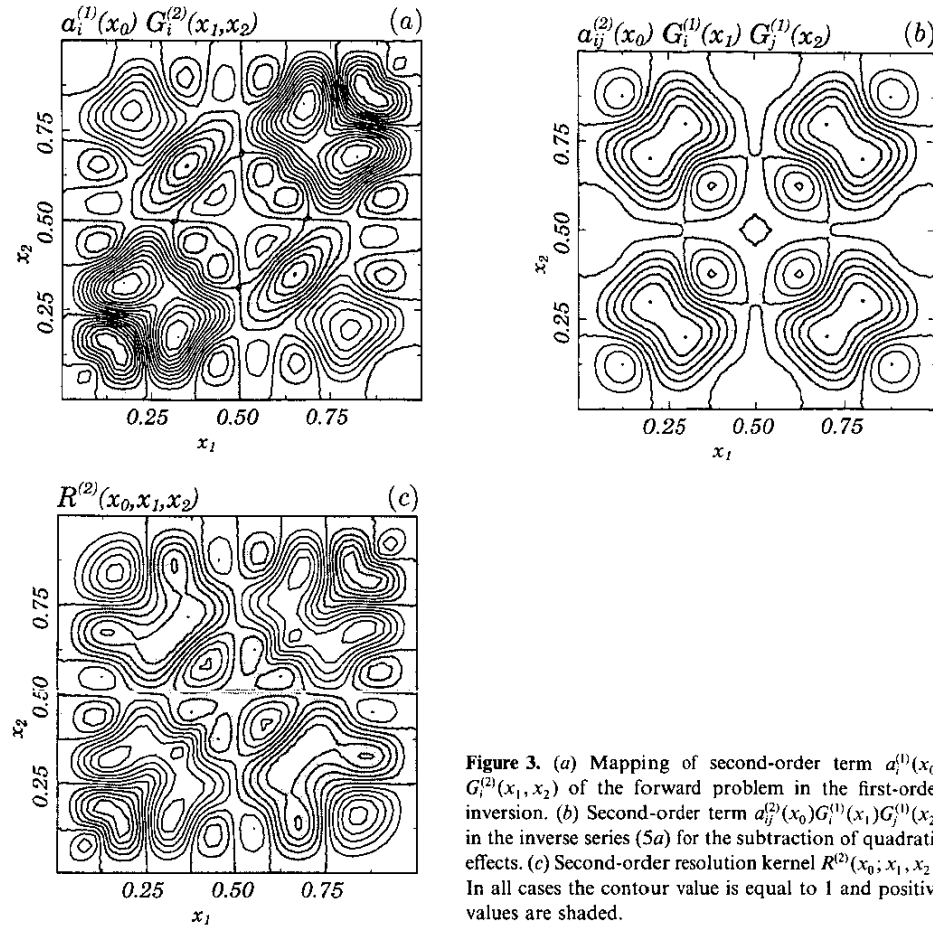


Figure 3. (a) Mapping of second-order term $a_i^{(1)}(x_0) G_i^{(2)}(x_1, x_2)$ of the forward problem in the first-order inversion. (b) Second-order term $a_{ij}^{(2)}(x_0) G_i^{(1)}(x_1) G_j^{(1)}(x_2)$ in the inverse series (5a) for the subtraction of quadratic effects. (c) Second-order resolution kernel $R^{(2)}(x_0; x_1, x_2)$. In all cases the contour value is equal to 1 and positive values are shaded.

about the nonlinearity of the forward problem. The function $a_{ij}^{(2)} G_i^{(1)}(x_1) G_j^{(1)}(x_2)$ which serves to subtract this spurious nonlinear mapping is shown in figure 3(b). The second-order resolution kernel $R^{(2)}(x_0; x_1, x_2)$, which is the sum of the functions of the figures 3(a) and 3(b) is shown in figure 3(c).

It is instructive to estimate the ratio of the spurious quadratic mapping to the linear mapping in the inversion. If the model has a characteristic magnitude M this ratio is approximately given by

$$\frac{|\varepsilon^2 \int R^{(2)}(x_0; x_1, x_2) m(x_1) m(x_2) dx_1 dx_2|}{|\varepsilon \int R^{(1)}(x_0; x_1) m(x_1) dx_1|} \approx \varepsilon M \|R^{(2)}\|_2 \quad (61)$$

where the unimodularity constraint (13) has been used for the estimation of the denominator. For the resolution kernel of figure 3(c), $\|R^{(2)}\|_2 = 2.55$. This means that for a model with a strength $\varepsilon M = 1\%$ the relative error by the spurious quadratic mapping is only 2.5%. However, for a model strength of $\varepsilon M = 20\%$, the error introduced by the spurious quadratic mapping is 50% of the linear mapping. The resolution kernel thus contains valuable information as to what extent the nonlinearities can or cannot be removed in the inversion.

If one simply performs a linear inversion of the data, and ignores that the forward problem is in reality nonlinear, one obtains a quadratic mapping from the true model to the model functional shown in figure 3(a). The L_2 -norm of this function is approximately 3.47. Since the L_2 -norm of the second-order resolution kernel in figure 3(c) is equal to 2.55 this means that for the employed data about one third of the quadratic nonlinearity in the forward problem is removed in the second-order inversion. There are two reasons why the spurious nonlinear mapping is only partly removed

(1) It can be seen from (54) that the second-order data kernel $G_n^{(2)}(x_1, x_2)$ contains a sum over all eigenfunctions $u_m^{(0)}$, regardless of whether or not one actually knows the perturbed eigenfrequencies of these eigenfunctions. This gives rise to smaller wavelength structures in $G_n^{(2)}(x_1, x_2)$ than are present in the first-order data kernels. Since a linear combination of the first-order data kernels is used to subtract the nonlinearity introduced by the higher-order data kernels this implies that this subtraction can only be partially achieved. One finds indeed that the function used for subtracting the second-order nonlinearity (figure 3(b)) is smoother than the second-order forward mapping (figure 3(a)) so that this subtraction cannot be fully achieved. The results of section 4 imply that this discrepancy in the wavelengths of these patterns is only removed when the first-order data kernels form a complete set, so that all wavelengths are present in the set of functions that is used to subtract the nonlinearity.

(2) The nonlinear forward mapping of figure 3(a) has a twofold symmetry, it is symmetric for reflections in the two diagonals of the (x_1, x_2) plane. In contrast to this, the term for subtracting the second-order nonlinearity in figure 3(b) has a fourfold symmetry; this function is symmetric for reflections in the two diagonals in the (x_1, x_2) plane and in the lines $x_1 = L/2$ and $x_2 = L/2$. These symmetries can be related to the symmetry properties of the data kernels defined in (53) and (54). When subtracting a function with a fourfold symmetry from a function with only a twofold symmetry, the result is in general non-zero. This inability to completely subtract the quadratic nonlinearity in the inversion is related to the result of Borg (1946) that one needs the eigenfrequencies of the string for two different boundary conditions for a complete reconstruction of the structure of the string. For example, if one also knows eigenfrequencies for the string with one open end, one breaks the symmetry of the first-order data kernels for reflection in the middle of the string. The fourfold symmetry of figure 3(b) is then broken in the same way, and the term used for subtracting the nonlinearity has the same twofold symmetry as the second-order mapping in the forward problem. This allows for a more complete subtraction of the nonlinearity.

In figure 4 the results of a second-order inversion are shown for exact data for the four gravest modes computed for a point mass in position $x_0 = 0.25$. In the ideal case there is no spurious nonlinear mapping and the model functional at the location of the point mass is given by

$$\hat{m}(x_0) = \int_0^L R^{(1)}(x_0, x)m(x) dx = m_0 R^{(1)}(x_0, x_0) \quad \text{for } m(x) = m_0 \delta(x - x_0). \quad (62)$$

This quantity is shown in figure 4 together with the model functionals determined with a first-order inversion and with a second-order inversion where the series (5b) was truncated at respectively $n = 1$ and $n = 2$. The first-order inversion is a normal Backus–Gilbert inversion using the criterion (12). It follows from figure 4 that for mass perturbations between 2.5% and 10% the second-order inversion is much more accurate than the first-order inversion. In this parameter range a substantial increase in the accuracy

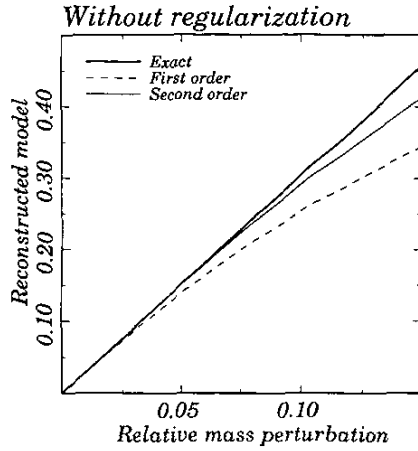


Figure 4. Ideal model estimate, as defined in (62), and the result of linear and second-order inversion for a point mass in $x = 0.25$ as a function of the relative mass-perturbation, m_0 .

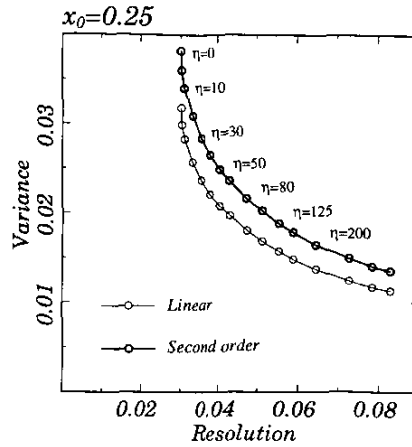


Figure 5. Trade-off curve for variance versus the width of the first-order resolution kernel as a function of the regularization parameter η . Linear variance is denoted with a thin line, the second-order variance with a thick line.

of the inversion is obtained by taking second-order effects into account. For larger values of the mass perturbation the ideal result from (62) and the result from the second-order inversion diverge. There are three reasons for this. First, as shown in figure 3(c), one cannot completely remove the second-order nonlinearity with this data set, so that there still is a spurious second mapping from the true model to the model functional. Second, the data used in the forward problem are computed from a numerical solution of the transcendental equation (56). This means that the used data also contain cubic and higher-order terms which are increasingly important for larger values of the mass perturbation. Third, even if the forward problem contained only first- and second-order terms, one needs to take third- and higher-order terms in the inverse series (4) into account. A simple example can clarify this. Suppose that d and m are simple scalars and that the forward problem is given by $d = \epsilon m + \frac{1}{2}(\epsilon m)^2$. This relation can be inverted to give $m = -\frac{1}{2} + \frac{1}{2}(1 + 2d)^{1/2} = d - \frac{1}{2}d^2 + \frac{1}{2}d^3 + \dots$. This last effect, and the effect of higher-order nonlinearities in the forward problem can be handled by using third- and higher-order terms in the inverse series (4). However, the fact that the second-order resolution kernel $R^{(2)}$ is not equal to zero is an effect of the finite amount of data that is used. Extending the inverse series to higher order does of course not remove this source of spurious second-order mapping in the inverse problem. This is analogous to the inherent inability to obtain a first-order resolution kernel with a vanishing width from a finite amount of data.

Up to this point, the effect of errors in the data has been ignored. Now suppose that the data are uncorrelated, and have the same variance σ :

$$C_{ij}^d = \sigma^2 \delta_{ij}. \tag{63}$$

For this case the variance in the model functional constructed from the linear and second-order estimators $a_i^{(1)}$ and $a_{ij}^{(2)}$ is according to (31) given by

$$\text{var}_{\hat{m}(x)} \equiv (\Delta \hat{m}(x) \Delta \hat{m}(x))^{1/2} = \sigma (a_i^{(1)} a_i^{(1)} + 4a_{ij}^{(2)} a_i^{(1)} d_j + 4a_{ij}^{(2)} a_{ik}^{(2)} d_j d_k)^{1/2}. \tag{64}$$

In the examples shown here the values $d_i = -0.1$ and $\sigma = 0.02$ are used.

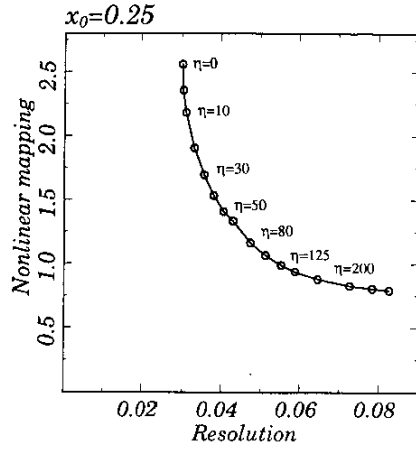


Figure 6. Trade-off curve for $\|R^{(2)}\|_2$ versus the width of the first-order resolution kernel as a function of the regularization parameter η .

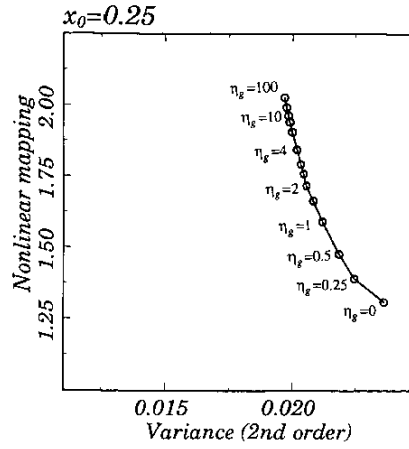


Figure 7. Trade-off curve for $\|R^{(2)}\|_2$ versus the second-order variance as a function of the regularization parameter η_g . Note the expanded scales in this figure.

By varying the parameter η in the criterion (14) one can regulate the trade-off between linear resolution (defined by (12)) and linear variance. The resulting trade-off curve is shown in figure 5. In this figure the variance is computed from (64) both for a linear inversion (where $a^{(2)} = 0$) and for a second-order inversion. Including second-order effects in the inversion increases the variance of the resulting model functional. This is due to the fact that the subtraction of the nonlinear components of the data is sensitive to data errors.

Increasing the regularization parameter η decreases the norm of the coefficients $a_i^{(1)}(x_0)$. Since the nonlinear components in the data are mapped to the model functional $\hat{m}(x_0)$ through the coefficients $a_i^{(1)}(x_0)$ one can reduce the spurious nonlinear mapping from the true model to the model functional through the linear estimator $a_i^{(1)}(x_0)$ by increasing η . This is shown in figure 6 where $\|R^{(2)}\|_2$ is shown as a function of the resolution of the first-order mapping. Note the strong trade-off between linear resolution and spurious nonlinear mapping. Interestingly, the optimal points on the trade-off curves in figures 5 and 6 are attained for similar values of η . This may not be so for other nonlinear inverse problems, and one has to decide how much importance one wants to attach to the factors variance, resolution and nonlinear contamination. The value $\eta = 60$ is selected here as an optimal choice; the resulting linear resolution kernel is shown in figure 2.

As shown in section 3, the parameter η_g can be used as an additional tool to regulate the trade-off between (nonlinear) variance and spurious nonlinear mapping. This trade-off is shown in figure 7. (In this example the value $\eta = 60$ is used for the computation of the linear coefficients $a^{(1)}(x_0)$.) Note the expanded scales that are used in this figure. It can be seen that increasing η_g has a minor effect on the variance, whereas the nonlinear mapping increases dramatically with η_g . For this example the value $\eta_g = 0.25$ would be a good choice. One should keep in mind that for this problem the Gram matrix (57) is well-conditioned. For problems where the data are nearly dependent and where the Gram matrix is poorly conditioned a stronger regularization using the parameter η_g may be advisable.

Trade-off curves as shown in figures 5, 6 and 7 make it possible to choose values for the trade-off parameters that optimize the conflicting requirements of minimizing the width of the linear resolution kernel, the (nonlinear) variance, and the spurious nonlinear mapping. The optimal choice depends on the strength of the nonlinearity. Note that for the computation of the trade-off curves one does not need to know the magnitude of the model; it suffices to use the data.

6. An example from quantum mechanics

The theory of the preceding section can be used for model estimation from a finite data set, and for the assessment of the artifacts that occur in model estimation. The example of this section shows that the theory can be extended for infinite data sets and that it can be used for the determination of the amount of data that are needed for an exact nonlinear inversion, a problem of mathematical interest.

Suppose that for the Schrödinger equation in three dimensions one wants to reconstruct an unknown potential from scattering data. In linear theory (the first Born approximation), it suffices to have the scattering coefficient for backscattering for all energies and all incoming directions. In contrast to this, the exact (nonlinear) Newton–Marchenko algorithm for inverse scattering requires the scattering coefficient for all energies, all incoming directions and all outgoing directions (Newton 1980, 1981). This means that the Newton–Marchenko algorithm requires a larger data set than a Born inversion. It is conjectured by Snieder (1990a) that this is due to the fact that the removal of the nonlinearities requires a larger data set than is needed for an exact Born inversion. The theory of sections 2 and 4 shows that this is not the case.

The analysis of this section leads to a reconstruction algorithm that is equivalent to the method derived by Prosser (1969). However, Prosser assumed *a priori* that the potential could be reconstructed exactly from backscattering data for all energies and incoming directions. His derivation relies completely on this assumption and cannot therefore be extended to data sets that are insufficient for a unique reconstruction of the potential. In general one cannot determine *a priori* if a given infinite data set is sufficient for an exact reconstruction of the model or not. This means that Prosser's results, although correct, are based on an unproven assumption. Using the method of section 2 one does not need the *a priori* assumption that the potential can be reconstructed exactly and one can verify *a posteriori* to what extent the nonlinearities can be removed by inspecting the nonlinear resolution kernels. For this, the theorem derived in section 4 is useful. The theory of this paper can thus not only be used to determine the potential (the construction problem), but it can also be used to determine to what extent the potential can be reconstructed from a given data set (the existence problem).

For the 3D Schrödinger equation, a perturbation treatment follows from the Born series. The scattering amplitude for wavenumber k , direction of incoming wave \hat{n} and direction of outgoing wave \hat{n}' is (Rodberg and Thaler 1967):

$$\begin{aligned}
 A_k(\hat{n}', \hat{n}) = & -\frac{\varepsilon}{4\pi} \int d^3 r \exp(-ik\hat{n}' \cdot \mathbf{r}) m(\mathbf{r}) \exp(ik\hat{n} \cdot \mathbf{r}) \\
 & + \frac{\varepsilon^2}{16\pi^2} \int d^3 r \int d^3 r' \exp(-ik\hat{n}' \cdot \mathbf{r}) m(\mathbf{r}) \\
 & \times \frac{\exp(ik|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} m(\mathbf{r}') \exp(ik\hat{n} \cdot \mathbf{r}') + O(\varepsilon^3).
 \end{aligned} \tag{65}$$

The potential is denoted here by the model $\epsilon m(\mathbf{r})$. Let the data consist of the scattering amplitude for backscattering:

$$d(\mathbf{k}, \hat{\mathbf{n}}) = A_k(-\hat{\mathbf{n}}, \hat{\mathbf{n}}). \quad (66)$$

Note that instead of the discrete data label i , the data are now a function of the continuous variables k and $\hat{\mathbf{n}}$. The example in this section shows that the theory of sections 2–4 can easily be generalized for an infinite data set. For convenience, the notation $\mathbf{k} \equiv k\hat{\mathbf{n}}$ is used in the following derivation.

Using (65) with $\hat{\mathbf{n}}' = -\hat{\mathbf{n}}$, one finds for the first- and second-order data kernels:

$$G^{(1)}(\mathbf{k}; \mathbf{x}) = \frac{-1}{4\pi} \exp(2i\mathbf{k} \cdot \mathbf{x}) \quad (67a)$$

$$G^{(2)}(\mathbf{k}; \mathbf{x}_1, \mathbf{x}_2) = \frac{1}{16\pi^2} \exp(i\mathbf{k} \cdot \mathbf{x}_1) \frac{\exp(ik|\mathbf{x}_1 - \mathbf{x}_2|)}{|\mathbf{x}_1 - \mathbf{x}_2|} \exp(i\mathbf{k} \cdot \mathbf{x}_2). \quad (67b)$$

Again, the discrete data indices of the data kernels are replaced by the continuous variable \mathbf{k} .

The Gram matrix Γ follows from (67a):

$$\Gamma(\mathbf{k}, \mathbf{k}') = \int G^{(1)*}(\mathbf{k}; \mathbf{x}) G^{(1)}(\mathbf{k}'; \mathbf{x}) d^3x = \frac{\pi}{16} \delta(\mathbf{k} - \mathbf{k}'). \quad (68)$$

Since this operator is diagonal, the inverse follows has the simple form

$$\Gamma^{-1}(\mathbf{k}, \mathbf{k}') = \frac{16}{\pi} \delta(\mathbf{k} - \mathbf{k}'). \quad (69)$$

Using the first-order coefficients from (9) one obtains

$$a^{(1)}(\mathbf{k}; \mathbf{x}_0) = \int d^3k' \Gamma^{-1}(\mathbf{k}, \mathbf{k}') G^{(1)*}(\mathbf{k}'; \mathbf{x}_0) = \frac{-4}{\pi^2} \exp(-2i\mathbf{k} \cdot \mathbf{x}_0). \quad (70)$$

Using this in (4) and ignoring higher-order terms is equivalent to performing a Born inversion of the data. By inserting (67a) and (70) in (7a) and replacing the summation over the data index by an integration over \mathbf{k} one finds that

$$R^{(1)}(\mathbf{x}_0; \mathbf{x}) = \delta(\mathbf{x}_0 - \mathbf{x}) \quad (71)$$

which reflects the well known fact that a Born inversion applied to the linear components of these data leads to a perfect reconstruction of the model.

It is interesting to see whether this data set is also sufficient for the removal of the nonlinearities in the inversion. The second-order terms $a^{(2)}$ follow from (22):

$$a^{(2)}(\mathbf{k}_1, \mathbf{k}_2; \mathbf{x}_0) = - \int d^3p_1 \int d^3p_2 \int d^3k \Gamma^{-1}(\mathbf{k}_1, \mathbf{p}_1) \Gamma^{-1}(\mathbf{k}_2, \mathbf{p}_2) a^{(1)}(\mathbf{k}; \mathbf{x}_0) \Gamma^{(2)}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{k}). \quad (72)$$

Using (69) and (70), and carrying out the integrations gives

$$a^{(2)}(\mathbf{k}_1, \mathbf{k}_2; \mathbf{x}_0) = \frac{1024}{\pi^4} \int d^3k \exp(-2i\mathbf{k} \cdot \mathbf{x}_0) \Gamma^{(2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}). \quad (73)$$

With the definition (21) and (67a, b), one obtains

$$\begin{aligned}\Gamma^{(2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) &= \int d^3 x_1 \int d^3 x_2 G^{(1)*}(\mathbf{k}_1; \mathbf{x}_1) G^{(1)*}(\mathbf{k}_2; \mathbf{x}_2) G^{(2)}(\mathbf{k}; \mathbf{x}_1, \mathbf{x}_2) \\ &= \frac{1}{16\pi^2} \int d^3 x_1 \int d^3 x_2 G^{(2)}(\mathbf{k}; \mathbf{x}_1, \mathbf{x}_2) \exp(-2i\mathbf{k}_1 \cdot \mathbf{x}_1) \exp(-2i\mathbf{k}_2 \cdot \mathbf{x}_2).\end{aligned}\quad (74)$$

Using this and inserting (69) and (70) in (72) leads to

$$\begin{aligned}a^{(2)}(\mathbf{k}_1, \mathbf{k}_2; \mathbf{x}_0) &= \frac{64}{\pi^6} \int d^3 k \int d^3 x_1 \int d^3 x_2 G^{(2)}(\mathbf{k}; \mathbf{x}_1, \mathbf{x}_2) \\ &\quad \times \exp(-2i\mathbf{k} \cdot \mathbf{x}_0) \exp(-2i\mathbf{k}_1 \cdot \mathbf{x}_1) \exp(-2i\mathbf{k}_2 \cdot \mathbf{x}_2).\end{aligned}\quad (75)$$

The second-order resolution kernel can be derived from (7b):

$$\begin{aligned}R^{(2)}(\mathbf{x}_0; \mathbf{x}_1, \mathbf{x}_2) &= \int d^3 k a^{(1)}(\mathbf{k}; \mathbf{x}_0) G^{(2)}(\mathbf{k}; \mathbf{x}_1, \mathbf{x}_2) \\ &\quad + \int d^3 k_1 \int d^3 k_2 a^{(2)}(\mathbf{k}_1, \mathbf{k}_2; \mathbf{x}_0) G^{(1)}(\mathbf{k}_1; \mathbf{x}_1) G^{(1)}(\mathbf{k}_2; \mathbf{x}_2).\end{aligned}\quad (76)$$

Inserting (70) and (75), and carrying out the integrals over \mathbf{k}_1 and \mathbf{k}_2 one obtains

$$\begin{aligned}\int d^3 k_1 \int d^3 k_2 a^{(2)}(\mathbf{k}_1, \mathbf{k}_2; \mathbf{x}_0) G^{(1)}(\mathbf{k}_1; \mathbf{x}_1) G^{(1)}(\mathbf{k}_2; \mathbf{x}_2) \\ = \frac{4}{\pi^2} \int d^3 k \exp(-2i\mathbf{k} \cdot \mathbf{x}_0) G^{(2)}(\mathbf{k}; \mathbf{x}_1, \mathbf{x}_2).\end{aligned}\quad (77)$$

Using this and (70) in the first term of (76) it follows that the second-order resolution kernel vanishes

$$R^{(2)}(\mathbf{x}_0; \mathbf{x}_1, \mathbf{x}_2) = 0 \quad (78)$$

This result could also have been obtained with the results of section 4. For the example of this section, the inverse of the Gram matrix exists; see (69). Furthermore, the first-order data kernels (67a) form a complete set in the sense that one can expand any function $f(\mathbf{x}_1, \dots, \mathbf{x}_n)$ in the first-order kernels

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = \int d^3 k_1 \dots d^3 k_n F(\mathbf{k}_1, \dots, \mathbf{k}_n) G_{k_1}^{(1)}(\mathbf{x}_1) \dots G_{k_n}^{(1)}(\mathbf{x}_n). \quad (79)$$

For the data kernels (67a) $F(\mathbf{k}_1, \dots, \mathbf{k}_n)$ is nothing but the Fourier transform of $f(\mathbf{x}_1, \dots, \mathbf{x}_n)$. The higher-order data kernels are therefore contained in the function space of the first-order data kernels. The two requirements of section 4 are therefore satisfied, and all higher-order resolution kernels vanish, so that there is no spurious nonlinear mapping from the true potential on the estimated potential.

The data set analysed in this study is therefore sufficient both for the reconstruction of the potential from the linear part of the data and for the proper subtraction of all multiple-scattering components in the data. This means that the Newton–Marchenko algorithm does indeed require a redundant data set. The conclusion of redundancy is therefore not affected by the nonlinearity of the inverse problem.

7. Discussion

The formalism of this paper can be used for the determination of nonlinear model functionals that describe the properties of the true model in an optimal fashion. Since

the theory relies on a perturbation treatment and since the number of required coefficients $a^{(n)}$ increases with the order n at a staggering rate, it is only possible to apply the method to inverse problems that are weakly nonlinear or in problems where it is possible to compute the $a^{(n)}$ and the model functional (4) analytically to all orders. The resolution kernels provide useful information on the resolution and degree of nonlinear contamination of models obtained in inversions. As with the theory of Backus and Gilbert (1968, 1970), the theory may be more valuable for the determination of the resolution properties (including the nonlinear effects) of the inverse problem than for the actual estimation of the model. The analysis of both the linear and the nonlinear resolution kernels can be helpful in the optimal design of experiments since it provides the means to determine the properties of the inverse problem as a function of the (nonlinear) data kernels.

In the ideal case, the nonlinear model functionals are good local estimates of the true model. One should be cautious in the interpretation of the resolution kernels. The first-order resolution kernel describes to what degree some local averaging is present in the inversion. For a given first-order resolution kernel, the model functional is a more informative estimate of the true model for a smoothly varying model than for a rapidly oscillating model (Parker 1977b). For the nonlinear resolution kernels a similar effect occurs. The higher-order resolution kernels describe to what extent the nonlinear components are removed from the data functionals. For a given nonlinear resolution kernel, the nonlinear mapping from the true model to the model functional is smaller for a weakly perturbed model than for a model that is strongly perturbed and nonlinear.

A statistical analysis of the model functionals reveals that the nonlinearity alters the statistics of the inverse problem. The most important result is that for unbiased data the nonlinearity of the problem may lead to a bias in the inversion. In many practical experiments the data have significant errors. A large amount of data is frequently used to eliminate the effect of these random errors. The results of section 3 imply that this procedure for suppressing the effect of data errors does not work for nonlinear inverse problems, and that the resulting model may suffer from a bias. It is of interest to note here that one frequently linearizes a nonlinear problem and that one applies a statistical analysis to the linearized inversion. This procedure may give the false impression that there is no bias in the model obtained from the linearized inversion.

The analysis of this paper relies on a perturbative treatment of both the forward and the inverse problem. Note that the derivation of section 2 breaks down when all first-order kernels $G_i^{(1)}$ vanish. It should be kept in mind that the theory provides a local analysis of the nonlinear inverse problem. A caveat should be made here that the global properties of the nonlinear inverse problem may be different from the local properties. As an example, consider the inverse problem (39) with $f(\xi) = \sin \xi$. This inverse problem is ill-posed; for $|d| \leq 1$ there are infinitely many solutions, whereas for $|d| > 1$ there is no solution. Now consider the truncated third-order expansion for this problem: $d = \int H(x)m(x)dx - \frac{1}{6}(\int H(x)m(x)dx)^3$. This equation leads to three possible values of $\int H(x)m(x)dx$ for $|d| \leq \frac{2}{3}\sqrt{2} \approx 0.9428$, and for one possible value for $|d| > \frac{2}{3}\sqrt{2}$. This example shows that truncation of the perturbation series for the forward problem changes the global properties of the inverse problem. It should be kept in mind that the theory of this paper only leads to a local analysis of the properties of the inverse problem. In the example shown here this implies that the theory can only be applied for models close to $m = 0$.

The theory of this paper relies completely on the perturbation expansions (2) and (4) of the forward and the inverse problem. It is in general extremely difficult to prove the

convergence of these expressions. It should be noted that convergence of the perturbation series (2) of the forward problem does not guarantee the convergence of the series (4) of the inverse problem. To see this, consider the forward problem (39) with $\mu = f(\xi) = \sin \xi$, the perturbation expansion

$$\mu = \xi - \frac{1}{3!}\xi^3 + \frac{1}{5!}\xi^5 + \dots$$

converges for all values of ξ . However, the perturbation series of the inverse

$$\xi = f^{-1}(\mu) = \arcsin \mu = \mu + \frac{1}{2}\frac{\mu^3}{3} + \frac{1.3}{2.4}\frac{\mu^5}{5} + \dots$$

converges only for $|\mu| < 1$. The divergence of the perturbation series is related to the ill-posedness of the inverse problem; for $|\mu| > 1$ the inverse problem has no solution and the derivative $\partial f^{-1}/\partial \mu$ is singular for $|\mu| = 1$. The issue of the convergence of the perturbation series for the inverse problem is therefore closely related to the well-posedness of the inverse problem. Since the theory of this paper assumes the existence of the perturbation series *a priori*, one ideally should supplement each problem with a proof of convergence of the forward and inverse series (2) and (4).

A proof of convergence usually entails certain restrictions to be imposed on the model. For example, Jost and Kohn (1952) establish the convergence of perturbation series similar to (2) and (4) for the Schrödinger equation for spherically symmetric potentials when the potential satisfies.

$$\int_0^{\infty} r^n |V(r)| dr < \infty \quad \text{for} \quad n = 1, 2 \quad \text{and} \quad r|V(r)| < M < \infty. \quad (80)$$

From a practical point of view this is not very satisfactory. Before performing the inversion one does not know what the model is, or whether the relation (80) is satisfied. However, one does know the data and one may know certain characteristics of the model. (For example, a mass density should be positive.) A convergence proof of the series (2) and (4) should therefore be based either on the data (instead of the model), or on physical *a priori* notions of the model (rather than convenient mathematical inequalities).

This implies that for the purpose of estimating a model, the series solution (4) of the inverse problem should be regarded as an asymptotic solution. The numerical example of figure 4 shows that the resulting model estimates may be substantially more accurate than the result from a linearized inversion. The usefulness of the theory of this paper for practical problems therefore depends on the strength of the nonlinearity. As a check on the consistency of the model obtained from a truncated solution of the series (4) one can compute synthetic data for the determined model using a fully nonlinear theory for the forward problem and verify whether the computed synthetic data agree with the used data within the measurement error. Truncating the inverse series (4) may thus provide useful nonlinear estimators of the model.

With these remarks in mind, the algorithm of section 2 can be used to estimate models from a finite data set. The example of section 6 for the inverse problem for the 3D Schrödinger equation shows how the theory can be used to determine whether a given data set is insufficient, sufficient, or redundant for the reconstruction of the model. With this theory, the redundancy of the data set required for the Newton–Marchenko method (Newton, 1980, 1981) is established, even in the case where the nonlinearity is strong and there is a large discrepancy between the scattering amplitude and its Born

approximation. The theory is therefore useful both for model estimation using a finite data set, and for idealized studies for an infinite data set.

Acknowledgments

The critical and constructive comments of T J Moser, Kathy Whaler and two anonymous reviewers are greatly appreciated.

References

- Agranovich Z S and Marchenko V A 1963 *The Inverse Problem of Scattering Theory* (New York: Gordon and Breach)
- Aki K and Richards P G 1980 *Quantitative Seismology* vol II (San Francisco: Freeman)
- 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
- 1970 Uniqueness in the inversion of inaccurate gross earth data *Phil. Trans. R. Soc. A* **266** 123–92
- Bender C M and Orszag S A 1978 *Advanced Mathematical Methods for Scientists and Engineers* (New York: McGraw-Hill)
- Borg G 1946 Eine Umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe. Bestimmung der Differentialgleichung durch die Eigenwerte *Acta Math.* **78** 1–96
- Chadan K and Sabatier P C 1989 *Inverse Problems in Quantum Scattering Theory* (Berlin: Springer)
- Franklin J N 1970 Well-posed stochastic extensions of ill-posed linear problems *J. Math. Anal. Appl.* **31** 682–716
- Jackson D D 1972 Interpretation of inaccurate, insufficient and inconsistent data *Geophys. J. R. Astron. Soc.* **28** 97–109
- Jost R and Kohn W 1957 Construction of a potential from a phase shift *Phys. Rev.* **87** 977–92
- Krappe H J and Lipperheide R 1985 *Advanced Methods for the Evaluation of Nuclear Scattering Data (Lecture Notes in Physics 236)* (Berlin: Springer)
- Morse P M and Feshbach H 1953 *Methods of Theoretical Physics* (New York: McGraw-Hill)
- Newton R G 1980 Inverse Scattering II: three dimensions *J. Math. Phys.* **21** 1698–715
- 1981 Inverse Scattering III: three dimensions continued *J. Math. Phys.* **22** 2191–200
- Parker R L 1977a Linear inference and underparametrized models *Rev. Geophys. Space Phys.* **15** 446–56
- 1977b Understanding inverse theory *Ann. Rev. Earth. Planet. Sci.* **5** 35–64
- Prosser R T 1968 Formal solutions of inverse scattering problems *J. Math. Phys.* **10** 1819–22
- Rodberg L S and Thaler R M 1967 *Introduction to the Quantum Theory of Scattering* (New York: Academic)
- Snieder R 1990a The role of the Born approximation in nonlinear inversion *Inverse Problems* **6** 247–66
- 1990b A perturbative analysis of non-linear inversion *Geophys. J. Int.* **101** 545–56
- Snieder R and Tarantola T 1989 Imaging of quantum-mechanical potentials *Phys. Rev. A* **39** 3303–9
- Takahashi T 1955 Analysis of the dispersion curves of Love-waves *Bull. Earthquakes Res. Inst.* **33** 287–96
- Tarantola A 1987 *Inverse Problem Theory* (Amsterdam: Elsevier)
- Wiggins R A 1972 The general linear inverse problem: implication of surface waves and free oscillations for Earth structure *Rev. Geophys. Space Phys.* **10** 251–85
- Zeeman E C 1977 *Catastrophe Theory* (Reading MA: Addison-Wesley).