PACS numbers: 02.50.+s, 06.50.dc

Error-propagation in weakly nonlinear inverse problems

H.J.S. Dorren[1], R.K. Snieder[1] and R.D. Gill[2]

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

2) Department of Mathematics, Utrecht University, P.O. Box 80.125, 3508 TC Utrecht, The Netherlands

Abstract

In applications of inversion methods to real data, nonlinear inverse problems are often simplified to more easily solvable linearized inverse problems. By doing so one introduces an error made by the linearization. Nonlinear inverse methods are more accurate because the methods that are used are more correct from a physical point of view. However, if data are used that have a statistical error, nonlinear inversion methods lead to a bias in the retrieved model parameters, caused the by nonlinear propagation of errors. If the bias in the estimated model parameters is larger than the linearization error, a linearized inverse problem leads to better estimation of the model parameter. In this paper the error-propagation is investigated for inversion methods that account the nonlinearity quadratically.

1 Introduction

Inverse problems are widely used in many fields of science to relate measured data to physically relevant model parameters. In applications of inversion methods to real data, inverse problems are often simplified to more easily solvable linearized inverse problems. However, by doing so an error in the simplified model is introduced due to the incorrect theory that is used.

In this paper we focus on the situation that the data that are used are contaminated with a statistical error described by a certain density function. If the density function of the data is Gaussian, then in the situation of a linearized inversion method, the density function of the estimated model parameter is also Gaussian. In this situation the mean of the distribution of the model estimator equals the mode of the distribution of the model estimator. However, due to the physical incorrect theory that is used, nonlinear effects are neglected and a linearization error in the estimated model parameter is introduced. We remark that if a nonlinear inversion method is linearized, the linearization is carried out implicitly around a reference model. The quality of the prior information (reference model) is therefore a measure for the linearization error.

In situations where a nonlinear inversion method is used, such a linearization error is obviously absent. However, due to the nonlinear relation between the data and the model parameter, a data set with a Gaussian error law in general is mapped onto a estimator of the model having a non-symmetric density function. This leads to the situation that the mean and the mode of the estimated model parameter are not equal, and a noise-bias in the model estimator is introduced [1].

An experimentalist often has the choice between using a linearized inversion method or using a nonlinear inversion method. If the linearization error due to the incorrect theory that is used is larger than the noise-bias introduced by nonlinear error propagation, then the nonlinear inversion method leads to better estimation of the model parameter. However, if the noise-bias due to nonlinear error propagation is larger then the linearization error, using a linearized inversion method leads to better estimation of the model parameter. This may appear to be a surprising result. Which situation arises is dependent on the degree of nonlinearity and the variance of the data errors and implicitly the quality of the prior information. In this paper it is assumed that all the data are uncorrelated and have equal variance. This situation can always be realized by a linear transformation of the data, if the covariance is known.

This paper has the following structure: in Section 2 we focus on the mathematical principles concerning the various statistical properties in the estimated model parameters. In Section 3, we focus on a geophysical example in order to discuss the principles of Section 2. We distinguish two special situations. In the first case we consider one data-point that is mapped on one model parameter. In the second case we consider a large number of data that are mapped on one model parameter.

2 The direct and inverse problem

Suppose a direct problem G_i relates a model function $\overline{m}$ to a set of discrete data $\overline{d}_i$:

$$\overline{d}_i = G_i(\overline{m}), \qquad \overline{d}_i^0 = G_i(\overline{m}_0). \tag{1}$$

In equation (1), the hypothetical data $\overline{d}_i^0$ corresponds to a reference model $\overline{m}_0$. Defining a new set of data by $d_i \equiv \overline{d}_i - \overline{d}_i^0$, we assume that the relation between the data d_i and the model $m = \overline{m} - \overline{m}_0$ can be expanded in a regular perturbation series (see Appendix A):

$$d_i = G_i(\overline{m}) - G_i(\overline{m}_0) = G_i^{(1)}(m) + G_i^{(2)}(m^2) + \cdots. \tag{2}$$

The objects $G_i(\overline{m})$ are defined in Appendix A by equation (A-1) and equation (A-2). In the following we refer to the data d_i as the error free data, and to m as the error free model parameter. We consider an inverse problem with a finite number of error free data d_i that depend on discrete error free model parameters m_j (if the model space is continuous, we assume for simplicity that the model estimator is restricted to a sub-space spanned by a finite number of basis functions. In that case a finite set of model parameters m_j results, see Appendix A). We will consider weakly nonlinear problems where in the case of error free data, up to second order, the forward problem (2) can be described by a regular perturbation series [2]:

$$d_i = G^{(1)}_{ij} m_j + G^{(2)}_{ijk} m_j m_k + \text{h.o.t.} \tag{3}$$

In equation (3) and in the following the Einstein summation convention is used which implies that summation is implied over all repeated indices. The abbreviation h.o.t. in equation (3) stands for higher order terms. In a real experiment the data d_i are measured, and one wants to retrieve the model parameter m_j. In this study it is assumed that the estimated error free model parameters $\hat{m}_i$ can be expressed as a regular perturbation series in the data d_i:

$$m_i = a^{(1)}_{ij} d_j + a^{(2)}_{ijk} d_j d_k + \text{h.o.t.} \quad (\equiv \hat{m}_i + \text{h.o.t.})\,. \tag{4}$$

The inverse problem is solved if the coefficients $a^{(1)}_{ij}$ and $a^{(2)}_{ijk}$ are known. In Appendix B we derive the coefficients $a^{(1)}_{ij}$ and $a^{(2)}_{ijk}$ using a least-squares technique.

Once the coefficients $a^{(1)}_{ij}$ and $a^{(2)}_{ijk}$ are known, the error made by linearizing the inverse problem can easily be calculated. If the inverse problem is linearized, but if the measured data d_i are generated by a nonlinear direct problem we find (see Appendix B) that the estimated model parameters $\hat{m}^L_i$ are given by:

$$\hat{m}^L_i = a^{(1)}_{ij} d_j. \tag{5}$$

In equation (5), the data d_j are generated by equation (3). The linearized model estimator $\hat{m}^L_i$ should be compared with the model parameter inferred from the nonlinear estimation defined by equation (4). The error $\Delta\hat{m}_i^L$ made by incorrectly assuming that the inverse problem is linearized, is up to second order equal to:

$$\Delta\hat{m}_i^L = \hat{m}_i - \hat{m}_i^L = a_{ijk}^{(2)} d_j d_k. \tag{6}$$

This quantity will be referred to as the linearization error. The linearization error is a systematic error made by incorrectly assuming that the inverse problem is linear. We want to remark that due to the fact that the data have a random error, the linearization error is also contaminated with a random error. Since we are interested in the systematic part of the linearization error we treat it as a non-stochastic variable. In the following, quantities contaminated with a random error are underlined. In this paper it is assumed that all the data $\underline{d}_i$ are uncorrelated and have an equal variance $\sigma_{\underline{d}}^2$. If the data $\underline{d}_i$ have a variance $\sigma_{\underline{d}}^2$ and are uncorrelated, then the variance $\sigma_{\underline{\hat{m}}_i}^2$ of a model estimator $\underline{\hat{m}}_i$ can be approximately calculated from the variance in the data:

$$\sigma_{\underline{\hat{m}}_i}^2 = \sum_{j=1}^{N} \left(\frac{\partial \hat{m}_i}{\partial d_j} \sigma_{\underline{d}} \right)^2 . \tag{7}$$

From equation (7) and equation (4) it follows that:

$$\sigma_{\underline{\hat{m}}_i}^2 = \sum_{q=1}^{N} \left(\left\{ a_{iq}^{(1)} + a_{iqk}^{(2)} d_k + a_{ijq}^{(2)} d_j \right\} \sigma_{\underline{d}} \right)^2 . \tag{8}$$

This implies that to lowest order the variance in the model parameter estimator is given by:

$$\sigma_{\underline{\hat{m}}_i}^2 = \sum_{q=1}^{N} \left(a_{iq}^{(1)} \sigma_{\underline{d}} \right)^2 . \tag{9}$$

Lastly, we calculate the noise-bias in the estimated model parameter. Suppose that the data $\underline{d}_i$ are contaminated with a random error $\underline{\eta}_i$ which has no noise-bias

($\langle \underline{\eta}_i \rangle = 0$). Then the difference $\Delta \underline{\hat{m}}_i$ between the contaminated model and the model obtained from error-free data is given by:

$$\Delta \underline{\hat{m}}_i = a^{(1)}_{ij} \underline{\eta}_j + a^{(2)}_{ijk} \{ \underline{d}_j \underline{\eta}_k + \underline{\eta}_j \underline{d}_k \} + a^{(2)}_{ijk} \underline{\eta}_j \underline{\eta}_k . \tag{10}$$

Taking the value of this expression, and taking into account that the data covariance matrix C_{ij} is given by $\langle \underline{\eta}_i \underline{\eta}_j \rangle$, then we find using $\langle \underline{\eta}_i \rangle = 0$ that:

$$\langle \Delta \underline{\hat{m}}_i \rangle = a^{(2)}_{ijk} C_{jk} . \tag{11}$$

If the data are uncorrelated and have equal variance $\sigma^2_{\underline{d}}$ ($C_{ij} = \delta_{ij} \sigma^2_{\underline{d}}$), this reduces to:

$$\langle \Delta \underline{\hat{m}}_i \rangle = a^{(2)}_{ijj} \sigma^2_{\underline{d}} . \tag{12}$$

This implies that even when the data errors are free of a noise-bias ($\langle \underline{\eta}_i \rangle = 0$), the nonlinearity leads to a noise-bias in the model estimators.

For practical applications of inverse problems it is interesting to know the ratio of the noise-bias in the estimated model parameters and the linearization error. If the noise-bias in a model estimator is larger than the linearization error, a linearized inverse problem leads paradoxically to better estimation of the model parameter. From equation (6) and equation (12), it follows that the ratio of the noise-bias and the linearization error is equal to:

$$\frac{\langle \Delta \underline{\hat{m}}_i \rangle}{\Delta \hat{m}^L_i} = \frac{a^{(2)}_{ijj} \sigma^2_{\underline{d}}}{a^{(2)}_{ijk} d_j d_k} . \tag{13}$$

From equation (13) it can be concluded that the ratio of the noise-bias and the linearization error is roughly proportional to $\sigma^2_{\underline{d}} / \|d\|^2$. This implies that the signal to noise ratio (S/N) is an indication for the ratio of the linearization error and the noise-bias. If we refer in the following to the S/N ratio, we mean the ratio of the rms value of the data $S \equiv \sqrt{\sum_i d_i^2}$, and the noise N which is equal to $\sigma_{\underline{d}}$. From equation (13) it follows that we can use as a rule of thumb, that the noise-bias is larger than the linearization error if $S/N < 1$. In the next section, we verify equation (13) explicitly in a geophysical example in the case of one single model parameter. Alternatively, the error in the linearized model estimators can be compared to the error in the nonlinear model estimators by computing the mean squared error, being equal to the noise-bias squared plus the variance [3]. The fact that the variances of the linearized and nonlinear density functions are equal (to leading order), the mean squared error does not have to be discussed in this paper and it is sufficient to limit oneself to the noise-bias only. Therefore, our conclusions on noise-bias carry over to conclusions on the mean squared error. It representatively indicates how far on average is the point estimate away from the truth. If a confidence interval around a point estimator would be given, this would be an interval with a length roughly proportional to the root of the mean squared error.

The ratio of the noise-bias in the model estimator obtained using a nonlinear inversion method and the dispersion of the estimated model parameter is given by:

$$\frac{\langle \Delta \hat{\underline{m}}_i \rangle}{\sigma_{\hat{\underline{m}}_i}} = \frac{a_{ijj}^{(2)} \sigma_{\underline{d}}^2}{\sqrt{\sum_q \left(a_{iq}^{(1)} \sigma_{\underline{d}} \right)^2}}. \tag{14}$$

From equation (14), it is concluded that if the inverse problem is nearly linear ($a_{ijk}^{(2)} \approx 0$), the noise-bias in the estimated model parameter is smaller than the variance $\sigma_{\hat{\underline{m}}_i}$ of the estimated model parameters. More generally, the ratio (14) depends on the nonlinearity over the confidence interval being equal to $\sigma_{\underline{d}}$. From equation (14), it can be concluded that if the nonlinearity is strong over the range of the variation of d, a linearized inversion method leads to better estimation of the model parameters. Lastly, the ratio of the linearization error and the dispersion is given by:

$$\frac{\Delta \hat{m}_i^L}{\sigma_{\underline{\hat{m}}_i}} = \frac{a_{jkl}^{(2)} d_k d_l}{\sqrt{\sum_q \left(a_{iq}^{(1)} \sigma_{\underline{d}_q} \right)^2}}. \tag{15}$$

In the following section we will illustrate this for a geophysical example in the simple case of only one model parameter.

3 A geophysical example

In this section we give a simplified numerical illustration of nonlinear error propagation. We do not intend to give an example of a realistic experiment, but we want to illustrate the principles of the previous section. It is shown in ref. [4] that in a medium having a constant velocity gradient:

$$c(z) = c_0 + \gamma z, \tag{16}$$

the position of a ray traveling through this medium is given by a circle segment. In equation (16) the velocity field $c(z)$ has a constant velocity gradient γ and a velocity c_0 at $z = 0$. Furthermore, it is shown in ref. [4] that the traveltime T of a wavefront traveling along the ray is given by:

$$\cosh[\gamma(T - T_0)] = \frac{(x_r - x_0)^2 + (z_r + \frac{c_0}{\gamma})^2 + \frac{c_0^2}{\gamma^2}}{2\frac{c_0}{\gamma}(z_r + \frac{c_0}{\gamma})}. \tag{17}$$

In equation (17) the position of the receiver is represented by the coordinates x_r and z_r, the position of the source is given by x_0, $z_0 = 0$. The reference time T_0 gives the time that the ray leaves the source in $(x_0, z_0 = 0)$.

In the numerical experiments that follow we consider a very simple earth model in which the trend of the P-velocity is based upon the iasp91 model [5]. Above and below a depth of 660 km a constant velocity gradient (16) is assumed. At a depth of 660 km a discontinuity is present. The jump of the velocity across this discontinuity is in fact the model parameter that we want to resolve. In Figure 1, an example of six velocity models for five different values of the discontinuity, increasing from a discontinuity that is equal to zero to a maximum discontinuity that is equal to 2 km/sec is given.

For the velocity models of Figure 1, the rays and the traveltimes in both the media above and below the discontinuity are circle segments. If the distance between the source and the receiver is smaller than 3000 km, all the rays turn above the 660 km discontinuity. If the source receiver distance is larger, then the rays penetrate below the discontinuity. If a ray crosses the discontinuity, the boundary conditions are given by Snell's law. In Figure 2 an example of the rays is given for 100 different velocity models. The distance between the source and the receiver is 4000 km. For all the rays that are plotted in Figure 2, the traveltimes can be calculated using equation (17). In Figure 3 the traveltimes curves are given for source-receiver distances between 2000 km and 9000 km as a function of the discontinuity. We see from Figure 3 that for source-receiver distances between 4000 and 7000 km the traveltime curve is a nonlinear function of the model parameter. For source receiver distances larger than 3000 km, the relation between the traveltimes and the discontinuity is nonlinear.

In the experiments that follow, the measured traveltimes are the data. The single model parameter is the velocity jump across the discontinuity. In the most simple illustration of nonlinear error propagation only one data-point and one model parameter are involved. Suppose a traveltime is measured for only one single source-receiver distance, then, if the relation between the traveltime and the discontinuity is bijective, the corresponding discontinuity can be estimated using the traveltime curves of Figure 3. In the following the error-propagation is discussed in the case of one data-point and one model parameter.

3.1 Case 1: One model parameter and one data-point

If only one single data-point and only one single model parameter is present, equation (3) reduces to:

$$d = G^{(1)}m + G^{(2)}m^2 + \text{h.o.t.} \tag{18}$$

where $G^{(1)}$ and $G^{(2)}$ are constants. Similarly the corresponding error free inverse problem (4) reduces to:

$$m = a^{(1)}d + a^{(2)}d^2 + \text{h.o.t.} \quad (\equiv \hat{m} + \text{h.o.t.})\,. \tag{19}$$

The inverse problem (19) is solved if the coefficients $a^{(1)}$ and $a^{(2)}$ are known. The coefficients (B-8) in Appendix B simplify in this case to:

$$a^{(1)} = \frac{1}{G^{(1)}}. \tag{20}$$

Similarly, the coefficient $a^{(2)}$ can be derived from equation (B-9) in Appendix B;

$$a^{(2)} = -\frac{G^{(2)}}{\{G^{(1)}\}^3}. \tag{21}$$

Following Section 2 the expression for the linearization error (6) reduces to:

$$\Delta\hat{m}^L = \hat{m}^L - \hat{m} = a^{(2)}d^2 = -\frac{G^{(2)}}{\{G^{(1)}\}^3}d^2. \tag{22}$$

From Section 2 and equation (19) it can be concluded that the linearization error is large compared to the linear model update ($a^{(1)}d$) if $a^{(2)}d^2 \gg a^{(1)}d$. We can calculate the variance in the model estimator by simplifying the formula of Section 2. The dispersion of the estimated model parameter is equal to:

$$\sigma_{\underline{\hat{m}}} = \frac{1}{|G^{(1)}|}\sigma_{\underline{d}}. \tag{23}$$

Lastly, following Section 2 we find that the noise-bias in the model which is represented by equation (12) reduces to:

$$\langle\Delta\underline{\hat{m}}\rangle = a^{(2)}\sigma_{\underline{d}}^2 = -\frac{G^{(2)}}{\{G^{(1)}\}^3}\sigma_{\underline{d}}^2. \tag{24}$$

From equation (13), we get the simple result for the ratio of the noise-bias and the linearization error:

$$\frac{\langle\Delta\underline{\hat{m}}\rangle}{\Delta\hat{m}^L} = -\left(\frac{\sigma_{\underline{d}}}{d}\right)^2. \tag{25}$$

From equation (25), it follows that the linearization error is much smaller than the noise-bias in the estimated model parameter if $\sigma_{\underline{d}} \gg \|d\|$. It is remarkable that this result does not depend on the coefficients $G^{(1)}$ and $G^{(2)}$ that characterize the direct and inverse problem but only the of the noise $N \equiv \sigma_{\underline{d}}$ and the signal $S \equiv d$. Equation (25) implies that if the linearization is carried out closely around the true model ($d \equiv \overline{d} - \overline{d}_0 \rightarrow 0$), the linearization error is relatively small with compared to the noise-bias. In the following numerical example we illustrate this principle.

In Figure 4 the full curve represents the nonlinear traveltime curve at a source receiver distance of 4000 km. The broken curve represents the linearization of the relation between the discontinuity and the the traveltime around 1.4 km/sec. We construct the density function of the data numerically using a random number generator that generates an ensemble of data consistent with the properties of the data density function. The density function of the data set is assumed to be Gaussian with variance $\sigma_{\underline{d}}$. It was shown for I.S.C. traveltimes that the density function can be approximated well by a Gaussian density function [6]. The histogram of the density function of the model parameter estimator is constructed by mapping every randomly generated traveltime on its corresponding value of the discontinuity.

We distinguish two situations. In the first situation, we choose $\sigma_{\underline{d}} > \|d\|$. From equation (25), it follows that in this case the noise-bias is larger than the linearization error, which implies that a linearized inversion method leads to the best estimation of the model parameters. This situation is realized for a density function of the data having an expectation value at $t = 432.5$ sec and a variance $\sigma_{\underline{d}} = 0.5$ sec. The full curve in Figure 5 represents the density function of the model estimator if a nonlinear inversion method is used. It is observed from the full curve in Figure 5 that the density function of the estimated model parameter is non-symmetric due to the nonlinearity in the traveltime curve that is used. The mean of the nonlinear density function in Figure 5 is equal to 1.01 km/sec (this is indicated by the thick solid vertical line), whereas the true model value is equal to 1.1 km/sec (the thin solid vertical line). This under-estimation is partly due to the long tail of the density function of the model estimator introduced by the flattening of the nonlinear traveltime curve in Figure 4. The distance between the thick and the vertical solid lines corresponds to the noise-bias $\langle \Delta \underline{\hat{m}} \rangle$. If the noise-bias is computed using equation (24), we find that $\langle \Delta \underline{\hat{m}} \rangle$ is equal to -1.73 km/sec. This is in disagreement with Figure 5 because of the fact that the data variance is large with respect to the scale of the nonlinearity in traveltime curve in Figure 4 and because of the non-bijective mapping between the data and the model parameter. The broken curve in Figure 5 represents the density function of the model parameter if the relation between the data set and the model-parameter is linearized around $\hat{m} = 1.4$ km/sec. The mode of the density function obtained from the linearized inversion is equal to $\underline{\hat{m}} = 1.13$ km/sec. The distance between the dashed and the thin vertical solid lines corresponds to the linearization error $\Delta \hat{m}^L$.

In this experiment, it is illustrated that if $\sigma_{\underline{d}} > \|d\|$, the noise-bias in the estimated model parameter is larger than the linearization error. As a result of this, one should conclude that if $\sigma_{\underline{d}} > \|d\|$, a linearized inversion method leads to better estimation of the model parameters. This observation is in accordance with equation (25), in which it is shown that the ratio of the noise-bias and the linearization error depends only on the signal to noise ratio.

The second case that we distinguish is $\sigma_{\underline{d}} < \|d\|$. It follows from formula (25), that if $\sigma_{\underline{d}} < \|d\|$, the noise-bias is smaller than the linearization error. Consequently, a nonlinear inversion method leads to better estimation of the model parameters. In Figure 6 the experiments that are carried out for $\sigma_{\underline{d}} = 0.5$ sec, are repeated for $\sigma_{\underline{d}} = 0.1$ sec. The full curve represents the density function of the model estimator in case of a nonlinear inversion. The density function of the model estimator is nearly Gaussian because the nonlinearity of the traveltime curve is small over the range of the variation of d. The mean of the nonlinear density function estimates a discontinuity of 1.09 km/sec, whereas the true value of the discontinuity is 1.1 km/sec. If the noise-bias is computed using equation (24), we find that $\langle \Delta \underline{\hat{m}} \rangle$ is equal to 0.07 km/sec. This is in agreement with Figure 6 because of the fact that the data variance is small with respect to the scale of the nonlinearity in traveltime curve in Figure 4.

The broken curve in Figure 6 represents the density function of the model parameter which is the result of the linearized inversion method. The mean of the density function obtained after a linearized inversion is equal to 1.13 km/sec. This implies that if $\sigma_{\underline{d}} < \|d\|$, using a nonlinear inversion method leads to better estimation of the model parameter. Note also that an error analysis based on a linearized inversion theory would give a false impression of the accuracy of the inversion because the true model parameter (the thin vertical line), lies completely in the tail of the density function computed with a linearized theory (the dashed density function).

The numerical experiments that are presented above indicate that in the special case of one model parameter and one data-point a one-step linearized inversion method leads to the best estimation of the model parameter if $\sigma_{\underline{d}} > \|d\|$ and a nonlinear inversion method leads to the best estimation of the model parameter if $\sigma_{\underline{d}} < \|d\|$. This implies that if the linearization is carried out closely around the true model parameter (good quality prior information), a linearized inversion method leads to best estimation of the model parameter. In the following subsection, we repeat these experiments in the case of N data and one single model parameter.

It is shown that using more data can lead to a smaller noise-bias in the estimated model parameters.

3.2 Case 2: One model parameter and N data

In the case of one single model parameter and N data, the direct problem (3), takes the following form:

$$d_i = G_i^{(1)} m + G_i^{(2)} m^2 + \text{h.o.t.} \tag{26}$$

The inverse problem that corresponds with equation (4) is given by:

$$m = a_i^{(1)} d_i + a_{ij}^{(2)} d_i d_j + \text{h.o.t.} \quad (\equiv \hat{m} + \text{h.o.t.})\,. \tag{27}$$

The inverse problem (27) is solved if both the coefficients $a_i^{(1)}$ and $a_{ij}^{(2)}$ are known. Due to the Gaussian error-law the least-squares solution of the coefficients $a_i^{(1)}$ and $a_{ij}^{(2)}$ is computed. From equation (B-8) in Appendix B, it follows that the least-squares solution of the coefficients $a_i^{(1)}$ is equal to:

$$a_j^{(1)} = \frac{G_j^{(1)}}{\sum_i (G_i^{(1)})^2}. \tag{28}$$

In a similar fashion it can be derived from equation (B-9) in Appendix B, that the coefficients $a_{ij}^{(2)}$ are equal to:

$$a_{iv}^{(2)} = \frac{2G_i^{(2)} a_v^{(1)} - 3\sum_q \{G_q^{(1)} G_q^{(2)}\} a_i^{(1)} a_v^{(1)}}{\{\sum_p G_p^{(1)}\}^2}. \tag{29}$$

From equation (5) it follows that if it is incorrectly assumed that the inverse problem is linear while the direct problem is nonlinear, a linearization error which is equal to:

$$\Delta \hat{m}^L = \hat{m}^L - \hat{m} = a_{ij}^{(2)} d_i d_j. \tag{30}$$

is introduced.

We can calculate the dispersion of the model estimator similarly as in Section 2. We find that the dispersion of the estimated model parameter is given by:

$$\sigma_{\underline{\hat{m}}} = \sqrt{\sum_{j=1}^{N}\left(a_j^{(1)}\sigma_{\underline{d}_j}\right)^2} \equiv \sqrt{\sum_{j=1}^{N}\left(a_j^{(1)}\sigma_{\underline{d}}\right)^2}. \tag{31}$$

Lastly, equation (12) for the noise-bias in the estimated model parameter reduces to:

$$\langle\Delta\underline{\hat{m}}\rangle = a_{ij}^{(2)}\sigma_{\underline{d}_i} \equiv a_{ii}^{(2)}\sigma_{\underline{d}}^2. \tag{32}$$

From equation (30) and equation (32), it follows that the ratio of the noise-bias and the linearization error is given by:

$$\frac{\langle\Delta\underline{\hat{m}}\rangle}{\Delta\hat{m}^L} = \frac{a_{ii}^{(2)}\sigma_{\underline{d}}^2}{a_{rs}^{(2)}d_r d_s}. \tag{33}$$

It is observed from equation (33) that the ratio of the noise-bias and the linearization still depends on signal to noise ratio S/N, but in contrast to equation (25), equation (33) also depends on the on the coefficients $a_{ij}^{(2)}$. This implies that depending on the coefficients $a_{ij}^{(2)}$, the sum $\sum_{i=1}^{N} a_{ii}^{(2)}\sigma_{\underline{d}}^2$ in the numerator of equation (33) or the sum $\sum_{r,s=1}^{N} a_{rs}^{(2)}d_r d_s$ in the denominator of equation (33) may vanish.

From the expressions for the noise-bias (32), the dispersion (31) and the linearization error (30), one can conclude that, due to the summation over the data $\underline{d}_i$ or the variance $\sigma_{\underline{d}}^2$, using more data can lead to a smaller variance and noise-bias, but not to smaller linearization error. This can be shown explicitly if the direct problem (26) has identical data-kernels *i.e.*: the same measurement is carried out repeatedly ($G_i^{(1)} = G^{(1)}$ and $G_i^{(2)} = G^{(2)}$). It can easily be checked that in this situation the coefficients $a_i^{(1)}$ and $a_{ij}^{(2)}$ are equal to:

$$a_i^{(1)} = \frac{1}{N}\frac{1}{G^{(1)}}, \qquad a_{ij}^{(2)} = \frac{-1}{N^2}\frac{G^{(2)}}{\{G^{(1)}\}^3}, \tag{34}$$

hence the dispersion of the model estimator $\sigma_{\underline{\hat{m}}}$ is equal to:

$$\sigma_{\underline{\hat{m}}} = \frac{1}{\sqrt{N}}\frac{\sigma_{\underline{d}}}{G^{(1)}}. \tag{35}$$

From equation (35), it follows that in case of a direct problem having identical data kernels, the variance $\sigma_{\underline{\hat{m}}}$ of the model estimator becomes smaller if more data are added to the data set. In this special case, the noise-bias of the model estimator is equal to:

$$\langle \Delta \underline{\hat{m}} \rangle = -\frac{1}{N}\frac{G^{(2)}}{\{G^{(1)}\}^3}\sigma_{\underline{d}}^2. \tag{36}$$

Therefore, we can conclude from equation (36) that if the number of data is increased, the noise-bias in the estimated model parameter becomes smaller and that the noise-bias decreases faster with the number of measurements than the standard error. Finally, using equation (34) the linearization error is equal to:

$$\Delta \hat{m}^L = \sum_{i,j=1}^{N} a_{ij}^{(2)} d_i d_j = -\frac{G^{(2)}}{\{G^{(1)}\}^3} d^2. \tag{37}$$

Note that the linearization error is independent on the number of measurements. From equation (35) and (36), it can be concluded in the special case of equal data kernels, more data leads to a smaller variance and a smaller noise-bias of the model estimator. Ultimately, for large values of N, the linearization error will be larger than both the variance and the noise-bias. This implies that in that situation a nonlinear inversion is needed.

The principle that using more data must lead to a potentially better estimation of the model parameter is shown in an example in which no equal data kernels are present. In Figure 7, the density function of the model parameter estimator is presented, if traveltime curves at distances of 2000, 3000 , ..., 9000 km are used (Figure 3). The model parameter is estimated by minimization of the norm $\sum_i \|\underline{d}_i - \hat{\underline{d}}_i(\underline{m})\|^2$, where the data $\hat{\underline{d}}_i(\underline{m})$ are given by equation (26). The data that are used are uncorrelated and have an equal variance $\sigma_{\underline{d}} = 0.5$ sec. The mean of retrieved density function estimates a discontinuity of 1.09 km/sec, whereas the true value of the discontinuity is equal to 1.1 km/sec. If Figure 7 is compared to Figure 4, it is concluded that using more data leads to a more accurate model estimation with a smaller noise-bias and variance. The reason for this result lies in the fact that measurement errors are averaged and the $\sqrt{N}$-law applies.

4 Conclusions

In this paper the error propagation for weakly nonlinear inverse problems is discussed. In applications of inversion methods to real data, nonlinear inversion methods often are simplified to more easily solvable linearized inversion methods. If the data set is contaminated with a statistical error having a Gaussian density function, a linearized inverse problem, leads to a model estimator that is also contaminated with a statistical error having a Gaussian density function. However due to the physical incorrect theory that is used, a linearization error is introduced. On the other hand if a nonlinear inversion method is used, the applied theory is more correct from physical point of view, but as a result of the nonlinear propagation of errors, the density function of the model parameter estimator is non-Gaussian. This implies that the mean and the mode are no longer equal. As a result a noise-bias in the estimated model parameter is introduced. It depends on the choice of the model estimator whether a linearized inversion method or a nonlinear inversion method leads to the best estimation of the model parameters. If the model parameter is estimated by the mean, a nonlinear inversion method leads to the best result if the noise-bias is smaller than the linearization error and conversely, a linearized inversion method leads to the best result if the noise-bias is larger than the linearization error.

For the simple case of only one model parameter and one data-point it is concluded from Section 3 that it depends completely on the ratio $\sigma_{\underline{d}}/d$ whether the linearization error is larger than the noise-bias. If $\sigma_{\underline{d}} > \|d\|$, the noise-bias is larger than the linearization error, hence a linearized inversion method paradoxically leads to the best estimation of the model parameters. Conversely, if $\sigma_{\underline{d}} < \|d\|$, the linearization error is larger than the noise-bias and a nonlinear inversion method leads to the best estimation of the model parameter. We remark that since the data d depend on the initial guess of the model parameter, that if a linearization is carried out around the true model parameter, a linearized inversion method always leads to the best estimation of the model parameter.

It is shown in section 3 that using more data to estimate the model parameter can lead to a more accurate estimation of the model parameter. This is made explicit in the case of equal data kernels.

Acknowledgments

This research was supported by the Netherlands Organization for Scientific Research through the Pioneer project PGS 76-144. This is the Geodynamics Research Institute (Utrecht University) publication 96.024

Appendix A

The direct problem having discrete data

We consider the situation that the model function f is related to a set of discrete data d_i. This relation may be linear or it may be nonlinear. If one has N discrete data, then the data are the values of N, generally nonlinear continuous functionals of the unknown model function f. Following ref. [2], we assume that the relation between the discrete data d_i and the model function f can be expressed in a regular perturbation series:

$$d_i = \int F_i^{(1)}(x_1)f(x_1)dx_1 + \int F_i^{(2)}(x_1,x_2)f(x_1)f(x_2)dx_1dx_2 + \cdots \quad i = 1, 2\cdots N. \tag{A-1}$$

Expressions for the kernels $F_i^{(1)}(x_1)$ and $F_i^{(2)}(x_1,x_2)$ can be obtained from ref. [2]. We search for solutions of the function f in a sub-space spanned by a finite set of basis functions $\phi_i(x)$ $(i = 1,\cdots M)$:

$$f(x) = \sum_{j=1}^{M} m_j\phi_j(x). \tag{A-2}$$

If we substitute equation (A-2) is equation (A-1) we find:

$$\begin{aligned} d_i &= \sum_{j=1}^{M}\int F_i^{(1)}(x_1)\phi_j(x_1)m_j dx_1 \\ &+ \sum_{j=1}^{M}\sum_{k=1}^{M}\int\int F_i^{(2)}(x_1,x_2)\phi_j(x_1)\phi_k(x_2)m_j m_k dx_1 dx_2 + \cdots . \end{aligned} \tag{A-3}$$

If we identify:

$$G_{ij}^{(1)} = \int F_i^{(1)}(x_1)\phi_j(x_1)dx_1, \tag{A-4}$$

$$G_{ijk}^{(2)} = \int\int F_i^{(2)}(x_1,x_2)\phi_j(x_1)\phi_k(x_2)dx_1dx_2, \tag{A-5}$$

then equation (A-3) takes the following form:

$$d_i = \sum_{j=1}^{M} G_{ij}^{(1)} m_j + \sum_{j=1}^{M} \sum_{k=1}^{M} G_{ijk}^{(2)} m_j m_k + \cdots . \tag{A-6}$$

Equation (A-6) is the starting point of Section 5.2.

Appendix B

The coefficients $a_{ij}^{(1)}$ and $a_{ijk}^{(2)}$ for the least-squares solution

The inverse problem is solved if the coefficients $a_{ij}^{(1)}$ and $a_{ijk}^{(2)}$ are known. In this Appendix we formulate a perturbation method to derive these coefficients using a criterion based on a least-squares data fit. The least-squares solution of equation (4) is the estimator $\hat{m}_j$ that minimizes:

$$\min \sum_{i=1}^{N} \left\| d_i - G_{ij}^{(1)} \hat{m}_j - G_{ijk}^{(2)} \hat{m}_j \hat{m}_k + \cdots \right\|^2 . \tag{B-1}$$

This minimum is reached if the following gradient vector g_q is equal to zero:

$$g_q \equiv \frac{\partial}{\partial \hat{m}_q} \sum_i \left(d_i - G_{ij}^{(1)} \hat{m}_j - G_{ijk}^{(2)} \hat{m}_j \hat{m}_k \cdots \right)^2 = 0. \tag{B-2}$$

The gradient vector g_q of equation (B-2) has components which are equal to:

$$g_q = 2 \sum_i \left(d_i - G_{ij}^{(1)} \hat{m}_j - G_{ijk}^{(2)} \hat{m}_j \hat{m}_k \right) \left(G_{iq}^{(1)} + G_{iqk}^{(2)} \hat{m}_k + G_{ijq}^{(2)} \hat{m}_j \right) . \tag{B-3}$$

Expanding expression (B-3) to order $\hat{m}^2$ we find that the gradient vector g_q is zero if the following relation is satisfied:

$$d_i G_{iq}^{(1)} = G_{ij}^{(1)} G_{iq}^{(1)} \hat{m}_j + \left(G_{ij}^{(1)} G_{iqk}^{(2)} + G_{ij}^{(1)} G_{ikq}^{(2)} + G_{ijk}^{(2)} G_{iq}^{(1)} \right) \hat{m}_j \hat{m}_k -$$

$$\left(G_{iqk}^{(2)} + G_{ikq}^{(2)} \right) \hat{m}_k d_i. \tag{B-4}$$

In order to determine the coefficient $a_{ij}^{(1)}$ and $a_{ijk}^{(2)}$, we insert the estimator (4) in equation (B-4) and expand to second order in the data d_i: This leads to the following perturbation series in d_i:

$$G_{iq}^{(1)} d_i = \left\{ G_{ij}^{(1)} G_{iq}^{(1)} \right\} a_{jr}^{(1)} d_r + \left\{ G_{ij}^{(1)} G_{iq}^{(1)} a_{jrs}^{(2)} + \left(G_{ij}^{(1)} G_{iqk}^{(2)} + G_{ik}^{(1)} G_{ikq}^{(2)} + \right.\right.$$

$$G^{(2)}_{ijk}G^{(1)}_{iq}\Big)\, a^{(1)}_{jr}a^{(1)}_{ks} - \left(G^{(2)}_{rqj} + G^{(2)}_{rjq}\right) a^{(1)}_{js}\Big\}\, d_r s_s + \mathcal{O}(d^3). \tag{B-5}$$

This expression must hold for all data d_i, hence the coefficients of the $\mathcal{O}(d)$ and $\mathcal{O}(d^2)$ contribution at both sides can be equalized: The $\mathcal{O}(d)$ contributions are:

$$d_i G^{(1)}_{iq} = \left\{G^{(1)}_{ij}G^{(1)}_{iq}\right\} a^{(1)}_{jr} d_r. \tag{B-6}$$

From this we can solve the linear term of the inverse problem easily. For notational convenience we rewrite equation (B-6) in a matrix notation:

$$\left(\mathbf{G}^{(1)}\right)^T \mathbf{d} = \left\{\mathbf{G}^{(1)}\right\}^T \mathbf{G}^{(1)}\mathbf{a}^{(1)}\mathbf{d}. \tag{B-7}$$

We then see that immediately follows that the matrix $\mathbf{a}^{(1)}$ whose entries are $a^{(1)}_{ij}$ lead to the standard linear least squares estimator [7, 8]:

$$\mathbf{a}^{(1)} = \left(\left\{\mathbf{G}^{(1)}\right\}^T \mathbf{G}^{(1)}\right)^{-1} \left\{\mathbf{G}^{(1)}\right\}^T. \tag{B-8}$$

We can find an expression for the tensor $a^{(2)}_{jrs}$ in a similar fashion. If all the terms in equation (B-5) of order d^2 are collected we find for $a^{(2)}_{jrs}$:

$$a^{(2)}_{jrs} = \left(G^{(1)}_{iq}G^{(1)}_{ij}\right)^{-1} \Big\{G^{(2)}_{rqk}a^{(1)}_{ks} + G^{(2)}_{rkq}a^{(1)}_{ks} - G^{(1)}_{lm}G^{(2)}_{lqk}a^{(1)}_{mr}a^{(1)}_{ks} -$$

$$G^{(1)}_{lm}G^{(2)}_{lkq}a^{(1)}_{mr}a^{(1)}_{ks} - G^{(1)}_{lq}G^{(2)}_{lmk}a^{(1)}_{mr}a^{(1)}_{ks}\Big\}. \tag{B-9}$$

Once the coefficients $a^{(1)}_{ij}$ and $a^{(2)}_{ijk}$ the least-squares solution of (4) is known.

References

[1] D.D. Jackson and M. Matsu 'ura, *A Bayesian approach to nonlinear inversion*, J. Geophys. Res, **90**, 581-591, 1985.

[2] R.K. Snieder, *An extension of Backus-Gilbert theory to nonlinear inverse problems*, Inverse Problems **7**, 409-433, 1991.

[3] J.A. Rice, *Mathematical statistics and data analysis*, Brooks/Cole, Pacific Grove, 1988.

[4] V. Červený, *Ray tracing algorithms in three dimensional laterally varying layered structures*, in Seismic Tomography, With Applications in Global Seismology and Exploration Geophysics, G. Nolet (ed.) Reidel, Dordrecht, 1987.

[5] B.L.N. Kennett and E.R. Engdahl, *Traveltimes for global earthquake location and phase identification*, Geophys. J. Int. **105**, 429-465, 1991.

[6] R.D. van der Hilst and E.R. Engdahl, *On ISC PP and pP data and their use in delay-time tomography of the Caribbean region*, Geophys. J. Int. **106**, 169-188, 1991.

[7] R.V. Hogg and A.T. Craig, *Introduction to mathematical statistics*, The Macmillian, New York, 1958.

[8] G.A.F. Seber and C.J. Wild, *Nonlinear Regression*, Wiley, New York, 1989.

Captions for Figures

Figure 1: Example of the velocity models. Each model is indicated with a different line thickness.

Figure 2: Example of the rays at a source receiver distance of 4000 km within the range of the model parameter.

Figure 3: The traveltime curves for source-receiver distances between 1000 km and 9000 km (thick solid curve) as a function of the discontinuity and the linearized traveltime curves around a discontinuity of 0.5 km/sec (thin solid curve).

Figure 4: Traveltime curve for a source-receiver distance of 4000 km (full curve), and the linearization around a discontinuity of 1.4 km/sec (broken curve).

Figure 5: Probability density function of the retrieved model if the distance between the source and the receiver is 4000 km for a data variance $\sigma_d = 0.5$ km/sec. The P.D.F. for the nonlinear inversion is given by the full curve, the P.D.F. of the linearized inversion is given by the broken curve (the vertical lines in the same line-style indicate the mean of both curves). The thin vertical line indicates the true model estimator.

Figure 6: Same as Figure 5, but for $\sigma_d = 0.1$ km/sec.

Figure 7: Probability Density Function of the retrieved model parameter if the nonlinear traveltime curves for source receiver distances of $2000, 3000, \cdots, 9000$ km are used. The data variance is $\sigma_d = 0.5$ km/sec.