

Identification of seismic reflections using singular value decomposition†

BJØRN URSIN‡ and YUYING ZHENG§

Keywords: Seismic reflections, singular value decomposition, least-squares estimation

Singular value decomposition (SVD) is applied to the identification of seismic reflections by using two different models: the impulse response model where a seismic trace is assumed to consist of a known signal pulse convolved with a reflection coefficient series plus noise, and the delayed pulse model where the seismic signal is assumed to consist of a small number of delayed pulses of known shape and with unknown amplitudes and arrival times.

SVD clearly shows how least-squares estimation of the reflection coefficients may become unstable, since a division by the singular values is required. Two methods for stabilizing this procedure are investigated. The inverse of the singular values may be replaced by zeroes when they are less than a given threshold. This is called the SVD cut-off method. Alternatively, we may use ridge regression which in filter design corresponds to assuming white noise. Statistical methods are used to compute an optimal SVD cut-off level, and also to compute an optimal weighting parameter in ridge regression. Numerical studies indicate that the use of SVD cut-off or ridge regression stabilizes the least-squares procedure, but that the results are inferior to maximum-likelihood estimation where the noise is assumed to be filtered white noise.

For the delayed pulse model, we use a linearization procedure to iteratively update the estimates of both the reflection amplitudes and the arrival times. In each step, the optimal SVD cut-off method is used. Confidence regions for the estimated reflection amplitudes and arrival times are also computed. Synthetic data examples demonstrate the effectiveness of this method. In a real data example, the maximum-likelihood method assuming an impulse response model is first used to obtain initial estimates of the number of reflections and their amplitudes and traveltimes. Then the iterative procedure is used to obtain improved estimates of the reflection amplitudes and traveltimes.

1. Introduction

Reflection coefficient estimation plays an important role in seismic data processing. Knowledge of the seismic wavelet may be used in the design of least-squares wavelet inverse filters (Berkhout, 1977). Ursin and Holberg (1985) estimated the reflection coefficients by the method of maximum likelihood. In this case the noise is assumed to be filtered white noise, and the discrete filter coefficients are estimated in addition to the reflection coefficients. Another method which has been used to improve the least-squares procedure is singular value decomposition (SVD) (Levy

Received 1 February 1985.

† Paper presented at the 46th meeting of the European Association of Exploration Geophysicists, London, June 1984, revised August 1984. Also published in *Geophysical Prospecting*, October 1985, 33, 773-779.

‡ Seismic Research and Development a/s, P.O. Box 1965 Moholtan, N-7001 Trondheim, Norway.

§ Petroleum Technology Research Institute, N-7034 Trondheim-NTH, Norway. On leave of absence from: Geophysical Research Institute of the Petroleum Ministry, P.O. Box 11-3 Zhuoxian, Hebei Province, The People's Republic of China.

and Clowes, 1980; Shim and Cho, 1981; Van Riel, 1982; Tufts and Kumaresan, 1982; Van Riel and Berkhout, 1985; Lines and Treitel, 1984). Two methods may be used to stabilize the least-squares procedure (Lawson and Hanson, 1974). One is the SVD cut-off method which has been discussed by Shim and Cho (1981) and Van Riel (1982). The inverse of the singular values may be replaced by zeroes when they are less than some threshold. Another method is ridge regression which has been discussed by Van Riel (1982) and Lines and Treitel (1984). The weighting parameter used in ridge regression is similar to the white noise factor used in filter design. The linear parameter estimation problem is discussed in Appendix A, the optimal choice of SVD cut-off level in Appendix B, and the optimal choice of weighting factor in ridge regression is discussed in Appendix C. For the seismic impulse response model, in which the seismic signal is assumed to consist of a reflection coefficient series convolved with a known seismic pulse, these different methods will be compared on synthetic data.

An alternative model of the seismic signal is a sum of delayed pulses of known shape but with unknown amplitudes and arrival times (Van Riel, 1982, Van Riel and Berkhout, 1985). The amplitudes and arrival times are estimated using non-linear least-squares model fitting and an iterative Gauss-Newton procedure as discussed in Appendix D (see also Gjøystdal and Ursin, 1981; Lines and Treitel, 1984). Assuming that the model errors are Gaussian and that the errors made in the linearization of the model are small, it is possible to compute confidence regions for the estimated parameters (Ursin 1981). The effectiveness of this method is illustrated by synthetic and real data examples. In the real data example the number of reflections and the initial estimates of the amplitudes and traveltimes are obtained from interpretation of the results obtained with ML estimation by Ursin and Holberg (1984). Then the iterative procedure is used to obtain improved estimates of the reflection amplitudes and traveltimes.

2. The seismic impulse response model

The measured seismogram is assumed to consist of a known source wavelet convolved with a reflection coefficient series plus a random noise series. The discrete impulse response model used here is then

$$\begin{aligned} y_k &= r_k * p_k + w_k \\ &= \sum_{i=0}^N r_i p_{k-i} + w_k \end{aligned} \quad (1)$$

where y_k is the measured seismogram ($k = 0, 1, \dots, L + N$), p_k is the known source wavelet ($k = 0, 1, \dots, L$), r_k is the estimated reflection coefficients ($k = 0, 1, \dots, N$) and w_k is the unknown random noise ($k = 0, 1, \dots, L + N$).

Equation (1) can be rewritten in matrix-vector form (Claerbout 1976) as

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \\ \vdots \\ y_{L+N} \end{bmatrix} = \begin{bmatrix} p_0 & 0 & \dots & 0 \\ p_1 & p_0 & \dots & \vdots \\ \vdots & p_1 & \dots & \vdots \\ p_L & \vdots & \dots & 0 \\ 0 & p_L & \dots & p_0 \\ \vdots & \vdots & \dots & p_1 \\ \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & p_L \end{bmatrix} \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_N \\ \vdots \end{bmatrix} + \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_N \\ \vdots \\ w_{L+N} \end{bmatrix} \quad (2)$$

or

$$\mathbf{y} = \mathbf{P}\mathbf{r} + \mathbf{w} \quad (3)$$

where \mathbf{y} and \mathbf{w} are $(L + N + 1) \times 1$ vectors, \mathbf{r} is a $(N + 1) \times 1$ vector and \mathbf{P} is a $(L + N + 1) \times (N + 1)$ matrix.

The least-squares estimate of \mathbf{r} is

$$\hat{\mathbf{r}} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{y} \quad (4)$$

if $(\mathbf{P}^T \mathbf{P})^{-1}$ exists. The matrix inversion may be numerically unstable as some eigenvalues of the matrix $(\mathbf{P}^T \mathbf{P})$ may be close to zero. Hence, in the conventional LS inverse filtering, an extra term $\lambda^2 \mathbf{I}$ is often added to $(\mathbf{P}^T \mathbf{P})$ to make the inversion stable. λ^2 is a white noise factor, and \mathbf{I} is an identity matrix of dimension $(N + 1) \times (N + 1)$. This gives the stabilized LS estimate

$$\hat{\mathbf{r}} = (\mathbf{P}^T \mathbf{P} + \lambda^2 \mathbf{I})^{-1} \mathbf{P}^T \mathbf{y} \quad (5)$$

To analyse the stability of the LS estimation problem we may use singular value decomposition (SVD). Then the matrix \mathbf{P} can be decomposed as (Golub and Reinsch 1970; Lawson and Hanson 1974; Dongarra *et al.* 1979)

$$\mathbf{P} = \mathbf{U} \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^T \quad (6)$$

where the square matrices \mathbf{U} with dimension $(L + N + 1) \times (L + N + 1)$ and \mathbf{V} with dimension $(N + 1) \times (N + 1)$ are orthogonal matrices. The columns of \mathbf{U} and \mathbf{V} are the left and right singular vectors of the matrix \mathbf{P} respectively. (T denotes transpose). The matrix

$$\begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix}$$

is a rectangular diagonal matrix of the same dimension as \mathbf{P} with the real non-negative singular value series arranged in descending order of magnitude, that is

$$\mathbf{S} = \text{diag} [s_0, s_1, \dots, s_N] \quad (7)$$

where $s_0 \geq s_1 \geq s_N \geq 0$. Substituting (6) into (3) and multiplying by \mathbf{U}^T gives

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \tilde{\mathbf{r}} + \tilde{\mathbf{w}} \quad (8)$$

where $\tilde{\mathbf{y}} = \mathbf{U}^T \mathbf{y}$, $\tilde{\mathbf{w}} = \mathbf{U}^T \mathbf{w}$, and $\tilde{\mathbf{r}} = \mathbf{V}^T \mathbf{r}$ are the transformed vectors. The least-squares estimate corresponding to equation (4) is

$$\hat{\tilde{\mathbf{r}}} = [\mathbf{S}^{-1} \mathbf{0}] \tilde{\mathbf{y}} \quad (9)$$

In appendix A we consider the linear estimate

$$\hat{\tilde{\mathbf{r}}} = [\mathbf{D} \mathbf{0}] \tilde{\mathbf{y}} \quad (10)$$

where \mathbf{D} is a diagonal matrix approximating \mathbf{S}^{-1} .

In the following it will be assumed that \mathbf{r} and \mathbf{w} are uncorrelated stochastic vectors with zero mean values and covariance matrices $\sigma_r^2 \mathbf{I}$ and $\sigma_w^2 \mathbf{I}$ respectively. If the expected value of \mathbf{r} is known to be $\mathbf{r}_0 \neq 0$, we apply the estimation procedure to $\Delta \mathbf{r} = \mathbf{r} - \mathbf{r}_0$ and $\Delta \mathbf{y} = \mathbf{y} - \mathbf{P} \mathbf{r}_0$ (see Appendix A). The estimate of \mathbf{r} is now $\hat{\mathbf{r}} = \mathbf{r}_0 + \Delta \hat{\mathbf{r}}$. However, if we assume that \mathbf{r} has zero mean value, when it is in fact \mathbf{r}_0 , we see from equation (A3) that the expected value of the estimate is

$$E\{\hat{\mathbf{r}}\} = \mathbf{R} \mathbf{r}_0 \quad (11)$$

The resolution matrix (Jackson, 1972) $\mathbf{R} = \mathbf{H}\mathbf{P}$ should be close to the identity matrix for the estimate to be unbiased.

The stabilized least-squares estimate in equation (5) corresponds to equation (10) with

$$\mathbf{D} = \text{diag}[s_0/(s_0^2 + \lambda^2), \dots, s_N/(s_N^2 + \lambda^2)] \quad (12)$$

From equations (8) and (9) we see that if some of the singular values s_j are close to zero, the least-squares solution will become unstable. This stabilized LS procedure is also called ridge regression (Lawson and Hanson 1974), and in Appendix B it is shown that an optimal value of λ^2 is given by

$$\lambda^2 = \sigma_w^2/\sigma_r^2 \quad (13)$$

An alternative method for stabilizing the LS procedure is the SVD cut-off method (Van Riel 1982; Shim and Cho 1981; Lawson and Hanson 1974). The singular value series is truncated at point K (this means that the reciprocal elements of s_j ($j > K$) are set to zero), so that the estimate of the transformed reflection coefficient series is again given by equation (10) with a matrix \mathbf{D} having elements

$$d_j = \begin{cases} 1/s_j & j = 0, 1, \dots, K \\ 0 & j = K + 1, \dots, N \end{cases} \quad (14)$$

Combining equations (8) and (10) with (14) we obtain

$$\hat{\mathbf{r}}_j = \begin{cases} \tilde{r}_j + \tilde{w}_j/s_j & j = 0, 1, \dots, K \\ 0 & j = K + 1, \dots, N \end{cases} \quad (15)$$

the noise may become large. Hence, the singular value series should be cut-off to keep the error term at a proper level. In Appendix C it is discussed how to choose the optimal cut-off point K such that the expected value of $\|\hat{\mathbf{r}} - \mathbf{r}\|^2$ is minimized. After making the same assumptions as in ridge regression, it is shown that the cut-off level should be chosen such that

$$s_K \geq \sigma_w/\sigma_r \quad (16a)$$

and

$$s_{K+1} < \sigma_w/\sigma_r \quad (16b)$$

This result has also been obtained by Van Riel (1982) using signal-to-noise ratio considerations.

When we compare SVD methods with Fourier transform methods for wavelet deconvolution (Berkhout 1977), some simple analogies are apparent (see also Van Riel and Berkhout 1985). From equation (9) it is seen that the simple LS procedure corresponds to spectral division. From equations (14) and (10) we see that the SVD cut-off method corresponds to a spectral division in the frequency band where the signal spectrum is larger than the noise spectrum while the estimates are replaced by zeros outside this frequency band. Finally it is seen that ridge regression (or Bayes estimation) corresponds to a two-sided wavelet deconvolution filter which takes the noise spectrum into account. The analogy between the transformed quantities using SVD and the Fourier spectrum is further strengthened by results given by Ekstrom (1973). It is shown that the singular values of the matrix \mathbf{P} are approaching the absolute values of the Fourier spectrum of the discrete pulse p_k , given by

$$|P(f)| = \left| \sum_{k=0}^L p_k \exp(-2\pi i f k \Delta t) \right| \quad (17)$$

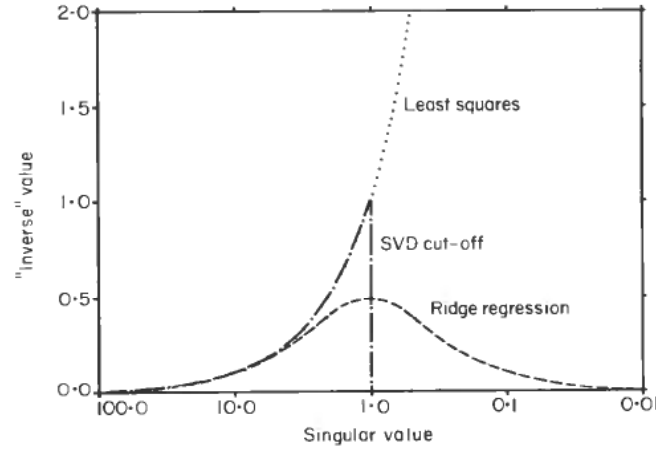


Figure 1. Characteristic curves for the inverse of the normalized singular values by different methods. Dotted line: Conventional least-squares. Dashed-dotted line: SVD cut-off. Dashed line: Ridge regression.

where Δt is the sampling interval. The least squares procedure may be stabilized by SVD cut-off or by ridge regression. In the transformed domain it is seen that the difference between these two methods lies in the choice of matrix \mathbf{D} for the linear estimator in equation (10) as given in equations (12) and (14). For the optimal choice of cut-off level (given in equation (16)) and for the optimal weighting factor in ridge regression (given in equation (13)) the two methods give similar results for large and small values of the singular values s_k . For $s_k \gg \lambda = \sigma_w/\sigma_r$ both methods give $d_k \approx s_k^{-1}$ and for $s_k \ll \lambda$ both methods give $d_k \approx 0$. Thus the main difference between the methods consists in different weights d_k for $s \approx \lambda$. This is illustrated in Fig. 1 where the diagonal elements of the matrix \mathbf{D} are plotted as a function of the normalized singular values (s_k/λ) with decreasing value. The dotted line shows the least-squares solution $d_k = 1/s_k$, which becomes unstable when s_k goes to zero. The dashed-dotted line shows the SVD cut-off solution which is equal to the LS solution for $s_k/\lambda \geq 1$ and zero otherwise. The dashed line shows the ridge regression solution which gives a smoother variation of the elements of the matrix \mathbf{D} as a function of the normalized singular values.

In order to compute the optimal weighting factor in ridge regression of the optimal cut-off value in the SVD cut-off method we need to know σ_r^2 and σ_w^2 . The former must be specified *a priori* while the latter may be estimated from the data as indicated in equation (A17). Then the residuals are used to estimate σ_w^2 , and the estimated value depends on the value of λ^2 or the cut-off level. Thus the formulas (13) and (16) may be used, but the expected squared error norm must be minimized numerically. We have used the SVD cut-off method and the expression (see equation (C-8)):

$$E\{\|\mathbf{r} - \hat{\mathbf{f}}\|^2\} = \sigma_r^2(N - K) + \frac{1}{N + L + 1} \sum_{L=k+1}^{N+L} \tilde{y}_k^2 \sum_{k=0}^K 1/s_k^2 \quad (18)$$

has been minimized with respect to K .

3. The seismic delayed pulse model

A different seismic signal model which has been studied by Van Riel (1982) and Van Riel and Berkhout (1985) is the delayed pulse model

$$y_i = \sum_{j=1}^N r_j p(i\Delta t - \tau_j) + w_i, \quad i = 1, 2, \dots, M \quad (19)$$

where r_j is the reflection amplitude and τ_j is the traveltime of the j th reflected wave, Δt is the sampling interval and N is the number of the reflections. In Appendix D a general non-linear parameter estimation problem is studied. The measurement data vector is

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \mathbf{w} \quad (20)$$

where $\mathbf{f}(\mathbf{x})$ is a non-linear function of the parameter vector \mathbf{x} , and \mathbf{w} is the noise or model error vector. It is assumed that \mathbf{x} and \mathbf{w} are uncorrelated stochastic variables. It is also assumed that the second-order error terms in the first-order Taylor-expansion of $\mathbf{f}(\mathbf{x})$ are small compared to the noise terms, so that we may write

$$\Delta \mathbf{y}_k = \mathbf{F}_k \Delta \mathbf{x} + \mathbf{w} \quad (21)$$

where $\Delta \mathbf{y} = \mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}_k)$, $\Delta \mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}_k$, and the elements of the Jacobian matrix \mathbf{F}_k are $f_{k,ij} = \partial f_i / \partial x_j$ evaluated at $\mathbf{x} = \hat{\mathbf{x}}_k$, $\hat{\mathbf{x}}_k$ is the estimate of \mathbf{x} at iteration step number k . Equation (19) is a special case of equation (20) with the unknown parameter vector $\mathbf{x}^T = (\mathbf{r}^T, \boldsymbol{\tau}^T)$ consisting of the reflection amplitudes and traveltimes. Linearizing equation (19) gives

$$\Delta \mathbf{y}_k = \mathbf{P}_k \Delta \mathbf{r} + \mathbf{Q}_k \Delta \boldsymbol{\tau} + \mathbf{w} \quad (22)$$

where

$$P_{k,ij} = p(i\Delta t - \hat{\tau}_{k,j}) \quad (23a)$$

and

$$Q_{k,ij} = -\hat{r}_{k,j} p'(i\Delta t - \hat{\tau}_{k,j}) \quad (23b)$$

for $i = 1, 2, \dots, M$, and $j = 1, 2, \dots, N$. The parameters are $\mathbf{r} = \hat{\mathbf{r}}_k + \Delta \mathbf{r}$, $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}_k + \Delta \boldsymbol{\tau}$. In equations (19) and (23) the seismic pulse $p(t)$ and its derivative $p'(t)$ are zero for $t < 0$ and $t > L\Delta t$.

The iterative parameter estimation scheme is illustrated in Fig. 2. The inputs to the algorithm are the seismic pulse $p(t)$, the data y_i , $i = 1, 2, \dots, M$, the initial estimate of the parameter vector, $\mathbf{x}_0^T = [\mathbf{r}_0^T, \boldsymbol{\tau}_0^T]$, and the covariance matrix of the parameters. This is assumed to be

$$\mathbf{C}_x = \begin{bmatrix} \sigma_{\Delta r}^2 \mathbf{I} & 0 \\ 0 & \sigma_{\Delta \tau}^2 \mathbf{I} \end{bmatrix} \quad (24)$$

At each step of the algorithm the problem is linearized according to equation (22), and the SVD cut-off method is applied to the linearized problem.

The new parameter estimate

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \Delta \mathbf{x}_{k+1} = \hat{\mathbf{x}}_k + \mathbf{H}_k \Delta \mathbf{y}_k \quad (25)$$

has bias (as computed in Appendix D)

$$E\{\hat{\mathbf{x}}_{k+1}\} - E\{\mathbf{x}\} = (\mathbf{I} - \mathbf{R}_k)(\mathbf{I} - \mathbf{R}_{k-1}) \dots (\mathbf{I} - \mathbf{R}_0)(\mathbf{x}_0 - E\{\mathbf{x}\}) \quad (26)$$

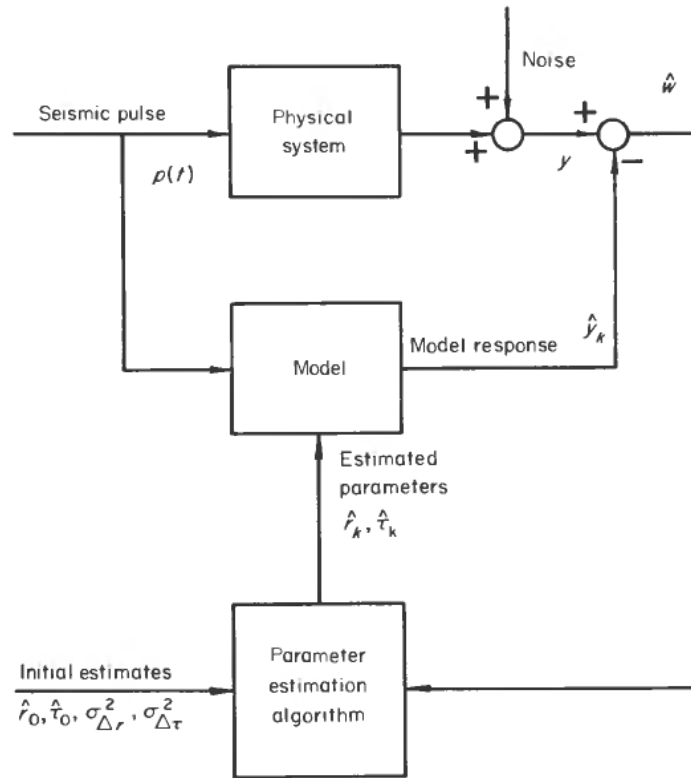


Figure 2. Block diagram of the parameter estimation scheme.

where $\mathbf{R}_k = \mathbf{H}_k \mathbf{F}_k$ is the resolution matrix at step number k , and \mathbf{x}_0 is the initial value of the parameter vector. The covariance matrix of the estimated parameter vector is computed from

$$\mathbf{C}_{\hat{\mathbf{x}}_{k+1}} = \mathbf{C}_{1,k} \mathbf{C}_x \mathbf{C}_{1,k}^T + \mathbf{C}_{2,k} \mathbf{C}_w \mathbf{C}_{2,k}^T \quad (27)$$

where $\mathbf{C}_{1,k}$ and $\mathbf{C}_{2,k}$ are given by

$$\mathbf{C}_{1,k} = (\mathbf{I} - \mathbf{R}_k) \mathbf{C}_{1,k-1} + \mathbf{R}_k \quad (28a)$$

and

$$\mathbf{C}_{2,k} = (\mathbf{I} - \mathbf{R}_k) \mathbf{C}_{2,k-1} + \mathbf{H}_k \quad (28b)$$

with starting values $\mathbf{C}_{1,-1} = \mathbf{C}_{2,-1} = \mathbf{0}$ for $k = 0$. Since $\mathbf{C}_w = \sigma_w^2 \mathbf{I}$ is unknown, we compute approximately in each step (for $k = 0, 1, \dots$):

$$\begin{aligned} \hat{\sigma}_w^2 &= \frac{1}{M} \|\Delta \mathbf{y}_k - \mathbf{F}_k \Delta \hat{\mathbf{x}}_{k+1}\|^2 \\ &= \frac{1}{M} \|\Delta \tilde{\mathbf{y}}_k - \mathbf{D}_k \mathbf{S}_k \Delta \tilde{\mathbf{y}}_k\|^2 \end{aligned} \quad (29)$$

From equation (26) it is seen that the final estimate is unbiased only if $E\{\mathbf{x}\} = \mathbf{x}_0$ or if $\mathbf{R}_k = \mathbf{I}$ for any value of k .

Assuming that the linearized analysis is valid and with the statistical assumptions made, the estimate of the parameters is asymptotically Gaussian with covariance $C_{\hat{x}}$ given by equation (27) (dropping the index k). Assuming $E\{\mathbf{x}\} = \mathbf{x}_0$ it follows (Silvey 1975, p. 91; Ursin, 1981, Appendix C) that $(\hat{\mathbf{x}} - \mathbf{x})C_{\hat{x}}(\hat{\mathbf{x}} - \mathbf{x})$ is distributed approximately as χ^2 with $2N$ degrees of freedom (N is the number of reflected pulses, so that $2N$ is the dimension of the parameter vector in this case). Consequently we can find a number χ_a^2 (see Lindley and Miller 1968, Table 5) so that for all \mathbf{x}

$$(\hat{\mathbf{x}} - \mathbf{x})^T C_{\hat{x}}^{-1} (\hat{\mathbf{x}} - \mathbf{x}) \leq \chi_a^2 \quad (30)$$

in $100(1 - \alpha)$ per cent of the cases. The region of the parameter space defined by equation (30) is an $2N$ dimensional ellipsoid and it is a $100(1 - \alpha)$ per cent confidence region.

In order to produce graphical results we shall only consider pairs of variables $(\hat{r}_j, \hat{\tau}_j)$ for a single reflection. With $\mathbf{M}_x = C_x^{-1}$ the ellipse

$$(\hat{r}_j - r_j)^2 M_{x,j,j} + 2(\hat{r}_j - r_j)(\hat{\tau}_j - \tau_j) M_{x,j,j+N} + (\hat{\tau}_j - \tau_j)^2 M_{x,j+N,j+N} \leq \chi_a^2 \quad (31)$$

is a $100(1 - \alpha)$ per cent confidence region (the number of degrees of freedom of the χ^2 variable is 2 in this case). Equation (31) has been used in the numerical examples with $\alpha = 0.05$ producing 95 per cent confidence regions. Note that the size of the confidence region depends largely on the initial covariance of the parameters, C_x , as given by the interpreter according to equation (24).

The proposed scheme for computing confidence regions for the estimated parameters is based on the assumption that the linearization made in each step is a good approximation, and that a good initial estimate of the covariance matrix of the parameters is available. Another approach would be to compute the second-derivative matrix (Hessian matrix), and use the negative of this as an estimate of the Fisher's information matrix (Bard 1974; Ursin 1981). This is again an estimate of the inverse of the covariance matrix which is used in the same way as above to compute confidence regions for the estimated parameters. A third approach is the random search method used by Silva and Hohmann (1983). Different initial values are used in the search algorithm to compute possibly different parameter estimates. The final estimate is the mean value of these estimates, and the covariance matrix of the estimated parameter vector is computed from the spreading of the estimates. This is a computationally expensive method since a large number of Gauss-Newton searches has to be done.

4. Numerical results

4.1. Comparison of different estimation methods

The model consists of a single layer between two homogeneous half-spaces, resulting in two reflection coefficients of opposite sign (0.20, -0.20). Synthetic seismograms were generated with an airgun array pulse shown in Fig. 3. Low-pass filtered (0-62.5 Hz) white noise was added to the synthetic seismogram. The seismic pulse was also digitally filtered with same low-pass filter which was applied to the noise. The data were sampled at a 1 msec interval. The synthetic seismograms are shown in Fig. 4(a) for a signal-to-noise ratio of 12 db, and the true reflection coeffi-

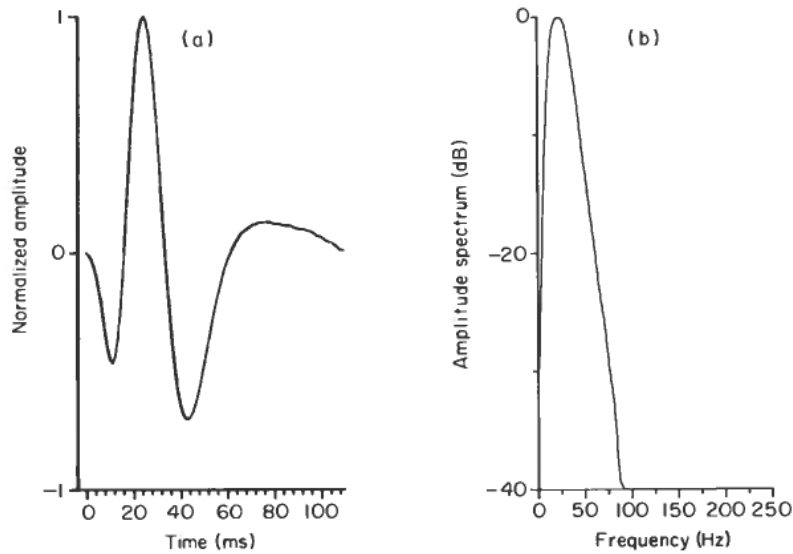


Figure 3. Pulse used to generate the synthetic seismograms. (a) Pulse form. (b) Amplitude spectrum.

cients are shown in Fig. 4(b). The result of applying different estimation algorithms to these synthetic data are shown in Figs. 4(c)–4(g). Figs. 4(c)–4(e) show the results of least-squares estimation, SVD cut-off, and ridge regression applied to the seismic impulse response model. It is seen that both SVD cut-off and ridge regression improve the results of least-squares estimation, but that all these results are inferior to the results of maximum likelihood estimation shown in Fig. 4(g) (from Ursin and Holberg 1985). These last results were used to pick starting values for the non-linear search method in the delayed pulse model. At each step in the search SVD cut-off was used, and the final results are shown in Fig. 4(f). These results are superior to the other results due to an improved signal model and good starting values. In Fig. 5 similar results are shown for a signal-to-noise ratio of 2 db, and it is seen that the same conclusion can be drawn from these data as for the case of a higher signal-to-noise ratio.

4.2. Estimation with the delayed pulse model.

The zero-phase pulse shown in Fig. 6 was used to generate the synthetic seismograms in this example. Fig. 7(a) shows the synthetic data for a signal-to-noise ratio of 13 db. The starting values (marked with \times) of the estimation algorithm were picked from the maxima of these data. The results of the non-linear search is given by the bars, and the ellipses mark the 95% confidence regions. The true values are indicated by the symbol $+$. From Fig. 7(b) it is seen that for almost ideal data, there is only a slight improvement in the estimates. Figs 8 and 9 show similar results for a signal-to-noise ratio of 4 db and 0.05 db respectively. It is seen that improved estimates have been obtained by the estimation algorithm, and that in most cases the true parameter values are within the 95% confidence region.

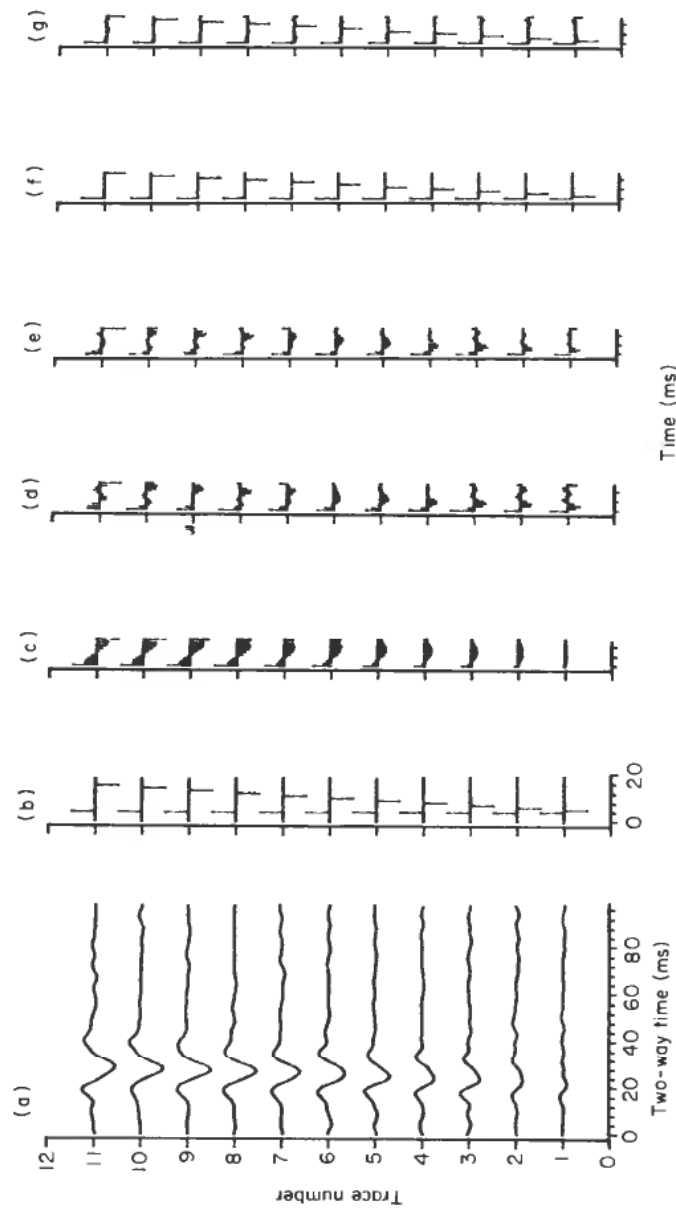


Figure 4. Comparison of different estimation methods applied to seismograms with SNR = 12 db. (a) Synthetic seismograms. (b) Theoretical reflection coefficients. (c) Estimates using conventional LS method. (d) Estimates using SVD cut-off method. (e) Estimates using ridge regression. (f) Estimates using iterative estimation scheme for delayed pulse model. (g) Estimates using ML method.

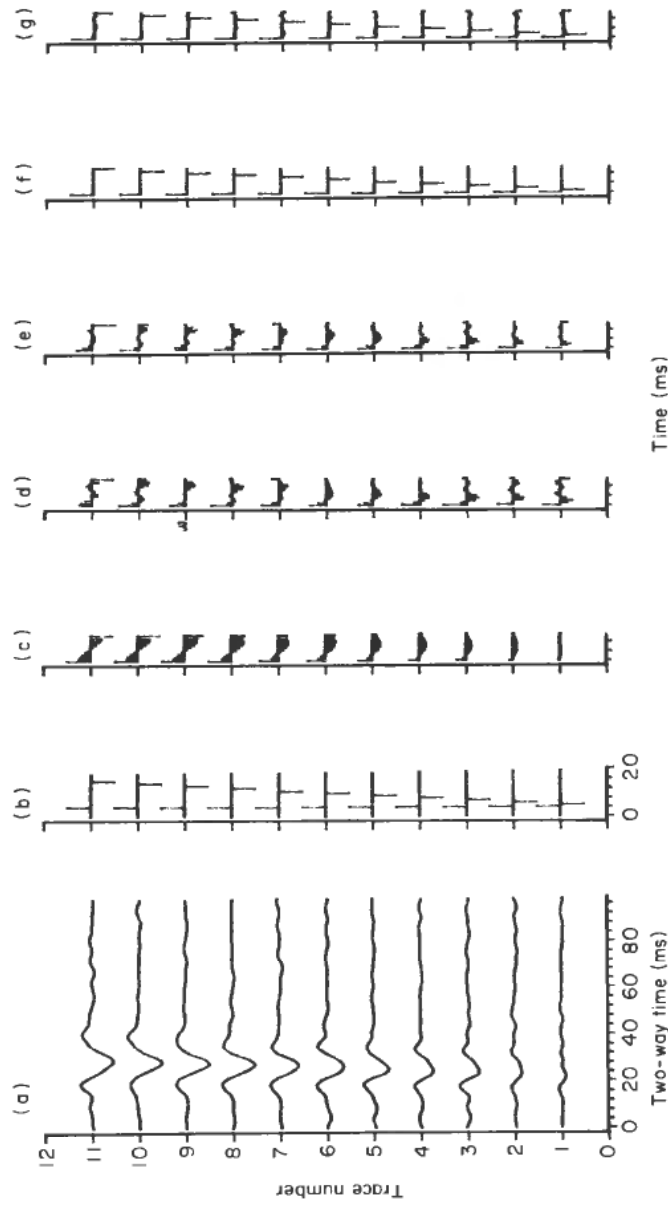


Figure 5. Comparison of different estimation methods applied to seismograms with SNR = 2 db. (a) Synthetic seismograms. (b) Theoretical reflection coefficients. (c) Estimates using conventional LS method. (d) Estimates using SVD cut-off method. (e) Estimates using ridge regression. (f) Estimates using iterative estimation scheme for delayed pulse model. (g) Estimates using ML method.

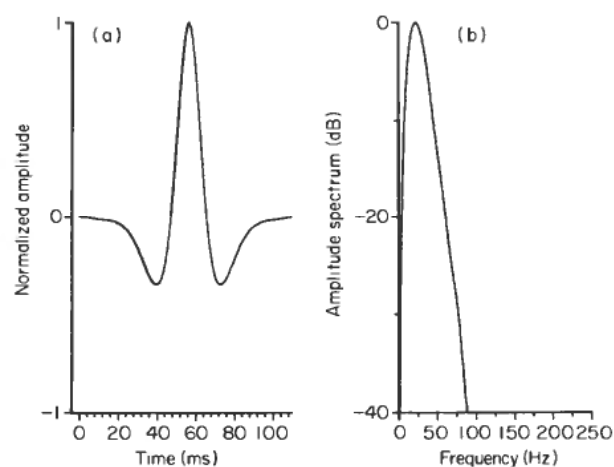


Figure 6. Zero-phase pulse used to generate the synthetic seismograms. (a) Pulse form. (b) Amplitude spectrum.

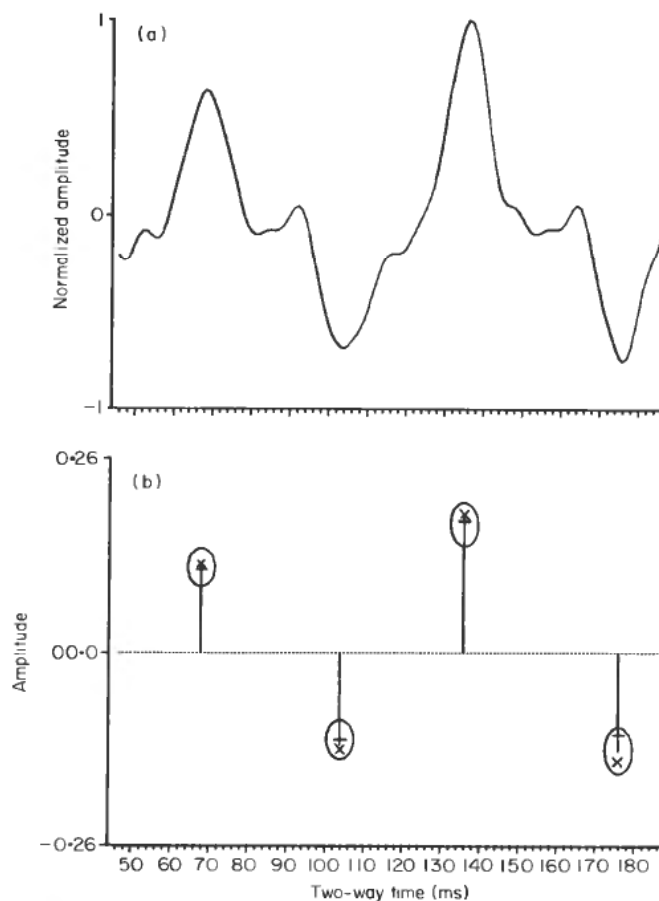


Figure 7. Synthetic seismogram with $\text{SNR} = 13$ db generated by the zero-phase pulse in Fig. 6 and the corresponding estimates of the reflections for the delayed pulse model. (a) Synthetic seismogram. (b) Estimate of the reflections (bars) with confidence regions compared to the true values (+) and the initial values (x) which are picked up from the maxima of (a).

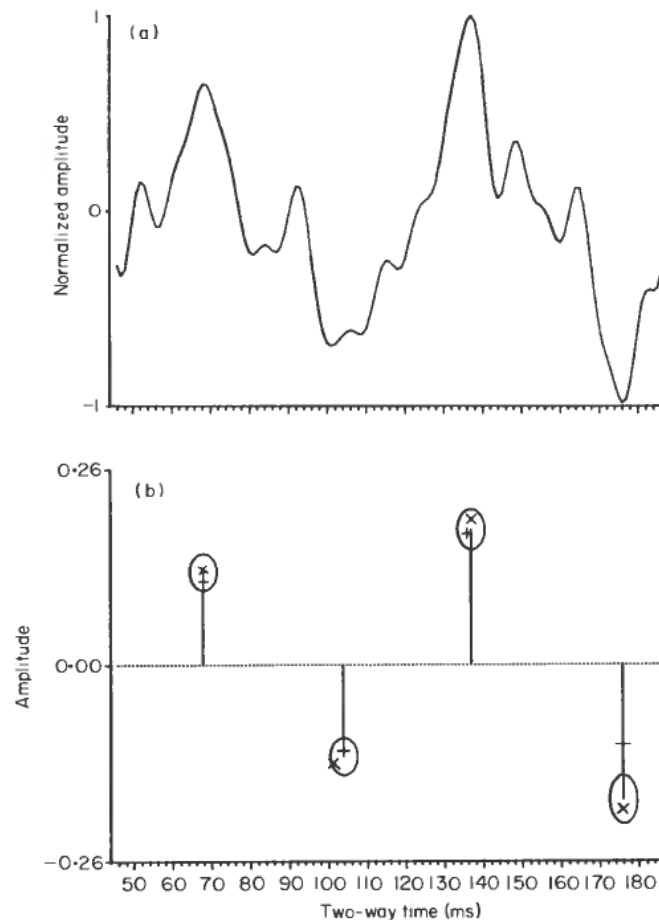


Figure 8. Synthetic seismogram with $\text{SNR} = 4$ db generated by the zero-phase pulse in Fig. 6 and the corresponding estimates of the reflections for the delayed pulse model. (a) Synthetic seismogram. (b) Estimate of the reflections (bars) with confidence regions compared to the true values (+) and the initial values (x) which are picked up from the maxima of (a).

4.3. Real seismic data

The iterative estimation scheme has also been applied to real data. The trace from CDP 278 (Fig. 10(a)) was used with the reflection coefficients derived from well logs (Fig. 10(b)) to obtain an estimate of the seismic pulse (Fig. 10(c) and 10(d)) (Ursin and Holberg 1985). The stacked seismic data from CDP 268 to CDP 288 are shown in Fig. 11(a). Using this estimated pulse and the ML estimation scheme the reflection coefficients (Fig. 11(c)) were obtained. The three main reflectors were interpreted from the ML results and used as initial values in the iterative estimation scheme for the delayed pulse model. The results are illustrated in Fig. 11(b). In Fig. 12 the estimated values are indicated by bars together with the 95% confidence regions for different CDP locations. The initial values obtained from the ML estimates are indicated by the symbol +. It is seen that mainly the reflection amplitude estimates have been changed compared to the ML estimates.

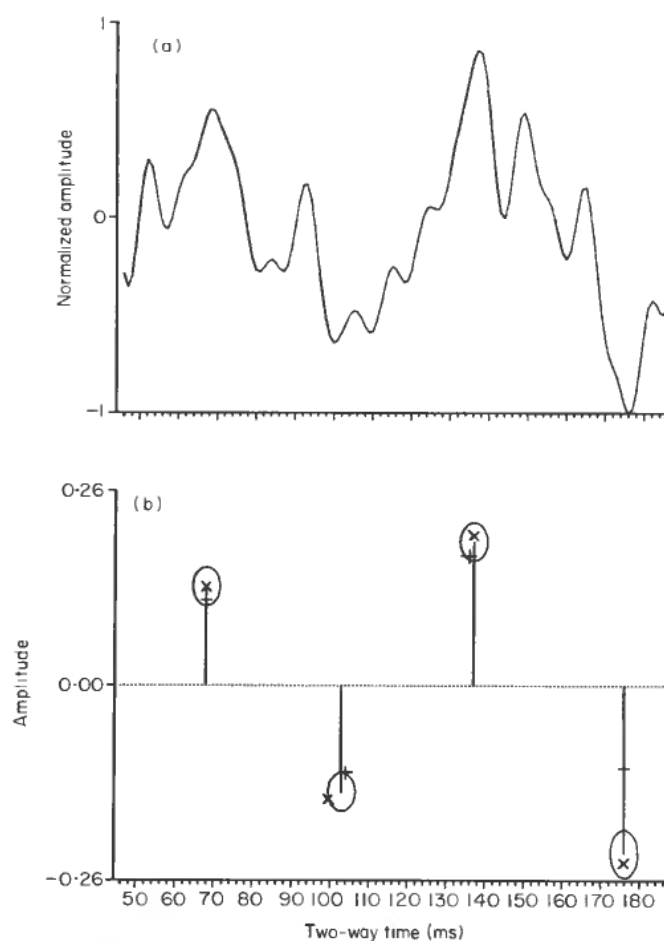


Figure 9. Synthetic seismogram with $\text{SNR} = 0.05$ db generated by the zero-phase pulse in Fig. 6 and the corresponding estimates of the reflections for the delayed pulse model. (a) Synthetic seismogram. (b) Estimate of the reflections (bars) with confidence regions compared to the true values (+) and the initial values (x) which are picked up from the maxima of (a).

5. Discussion and conclusions

Two different models have been used to identify seismic reflections. For the impulse response model it was shown that SVD cut-off and ridge regression stabilized the least-squares procedure, but that the results were inferior to the maximum-likelihood estimation assuming colored noise. The least-squares estimates could possibly have been improved by using an *a priori* estimate of the reflection series which is different from zero. This is only practical if the signal-to-noise ratio is high, and the seismic pulse is zero phase. In this case the other methods give excellent results, so that the possibility of using a non-zero *a priori* estimate of the reflection series was not investigated. The maximum likelihood algorithm was used to obtain initial estimates of the number of reflections and their amplitudes and traveltimes in the delayed pulse model. Then the iterative Gauss-Newton procedure was used to compute refined estimates of the traveltimes and reflection amplitudes. The number

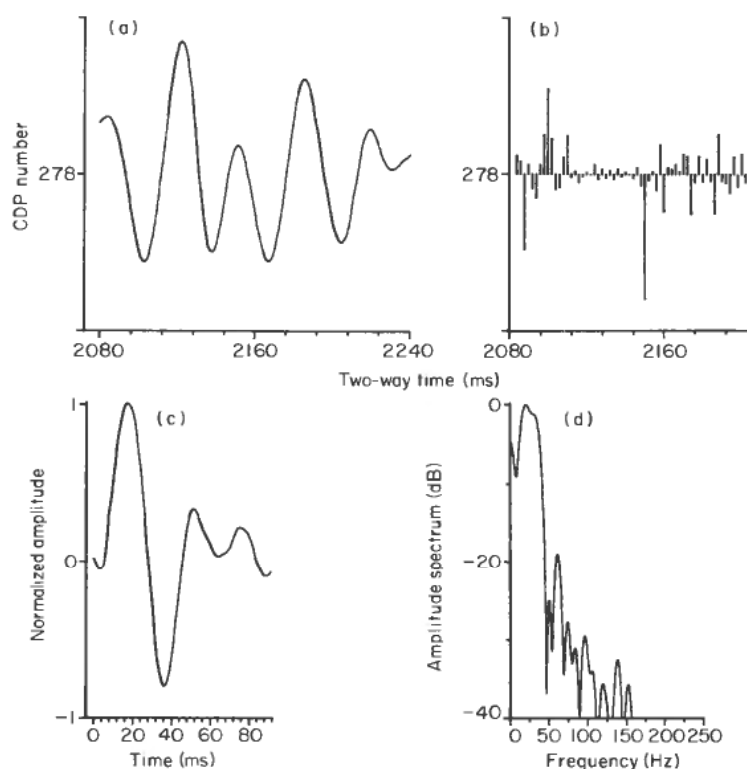


Figure 10. (Adapted from Ursin and Holberg, 1985) Stacked trace and well log data at CDP 278 and pulse estimated from these data. (a) Stacked trace from 2080 ms to 2240 ms at CDP 278. (c) Pulse estimated from the stacked data in (a) and well log data in (b) by the maximum-likelihood method. (d) Amplitude spectrum corresponding to the pulse in (c).

of iterations in the Gauss-Newton procedure varied from 10 to 30 iterations for poor signal-to-noise ratio and poor initial values and from 1 to 10 iterations for good signal-to-noise ratio and good initial values. For the practical data example the number of iterations varied from 1 to 5. The number of singular values which were included at each step varied from all at iteration number one to only one at the last iteration (in the extreme cases). In general, the number of singular values, which were used at each iteration, decreased with the iteration number.

The automatic method of choosing the SVD cut-off level seemed to work properly for the examples shown.

Appendix A

Linear parameter estimation

We consider the estimation problem

$$\mathbf{y} = \mathbf{P}\mathbf{r} + \mathbf{w} \quad (\text{A1})$$

where \mathbf{y} is the $(L + N + 1) \times 1$ data vector, \mathbf{P} is the $(L + N + 1) \times (N + 1)$ given matrix, \mathbf{r} is the $(N + 1) \times 1$ unknown vector and \mathbf{w} is a $(L + N + 1) \times 1$ noise vector.

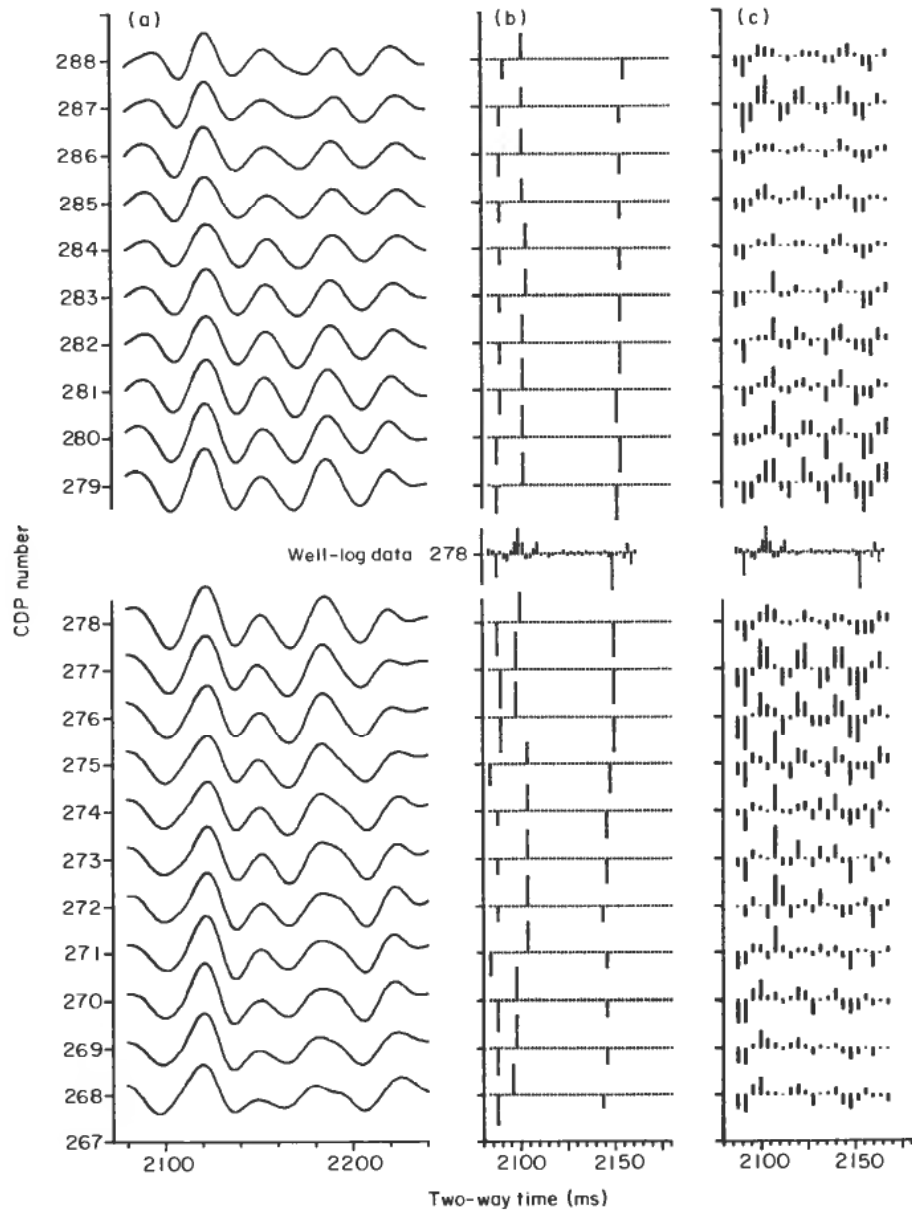


Figure 11. Real data example. (a) Stacked traces. (b) Estimated reflections using iterative estimation scheme for the delayed pulse model. (c) Estimated reflection coefficients by the ML method (from Ursin and Holberg 1985).

We assume that \mathbf{r} and \mathbf{w} are uncorrelated stochastic vectors with mean values

$$E\{\mathbf{r}\} = \mathbf{r}_0 \quad E\{\mathbf{w}\} = \mathbf{0}$$

and covariance matrices

$$\text{cov } \mathbf{r} = \mathbf{C}_r \quad \text{cov } \mathbf{w} = \mathbf{C}_w$$

We shall consider linear estimators

$$\hat{\mathbf{r}} = \mathbf{H}\mathbf{y} \quad (\text{A2})$$

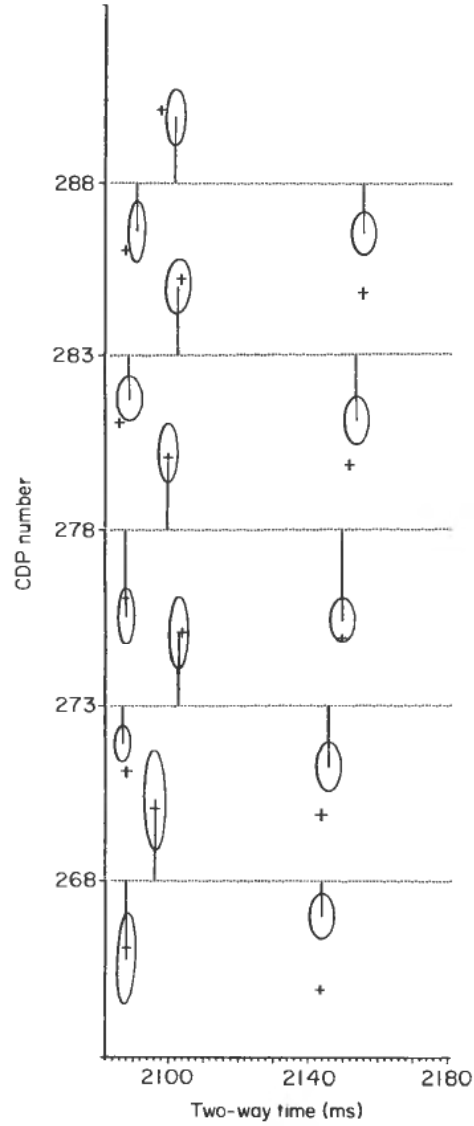


Figure 12. Estimated reflections (bars) with confidence regions compared to the initial values (+) taken from ML estimations for different CDP location.

where \mathbf{H} will be given later. This estimate has mean value

$$E\{\hat{\mathbf{r}}\} = \mathbf{H}\mathbf{P}\mathbf{r}_0 = \mathbf{R}\mathbf{r}_0 \quad (\text{A3})$$

where $\mathbf{R} = \mathbf{H}\mathbf{P}$ is the resolution matrix (Jackson 1972). The covariance of the data vector is

$$\mathbf{C}_y = \mathbf{P}\mathbf{C}_r\mathbf{P}^T + \mathbf{C}_w \quad (\text{A4})$$

so that the covariance of the estimate is

$$\mathbf{C}_{\hat{\mathbf{r}}} = \mathbf{H}\mathbf{C}_y\mathbf{H}^T = \mathbf{R}\mathbf{C}_r\mathbf{R}^T + \mathbf{H}\mathbf{C}_w\mathbf{H}^T \quad (\text{A5})$$

If \mathbf{r}_0 is known, we may compute the estimate

$$\hat{\mathbf{r}} = \mathbf{r}_0 + \mathbf{H}(\mathbf{y} - \mathbf{P}\mathbf{r}_0) \quad (\text{A6})$$

with $E\{\hat{\mathbf{r}}\} = \mathbf{r}_0$ and covariance matrix as above.

The singular value decomposition (SVD) of \mathbf{P} is

$$\mathbf{P} = \mathbf{U} \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \mathbf{C}^T \quad (\text{A7})$$

where \mathbf{U} and \mathbf{V} are orthogonal matrices of dimension $(L + N + 1) * (L + N + 1)$ and $(N + 1) * (N + 1)$ respectively, and

$$\mathbf{S} = \text{diag} [s_0, s_1, \dots, s_N] \quad (\text{A8})$$

of dimension $(N + 1) * (N + 1)$ with $s_0 \geq s_1 \geq \dots \geq s_N \geq 0$. The transformed vectors $\tilde{\mathbf{y}} = \mathbf{U}^T \mathbf{y}$, $\tilde{\mathbf{w}} = \mathbf{U}^T \mathbf{w}$, and $\tilde{\mathbf{r}} = \mathbf{C}^T \mathbf{r}$ satisfy the equation

$$\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \tilde{\mathbf{r}} + \tilde{\mathbf{w}} \quad (\text{A9})$$

We shall consider the linear estimate

$$\hat{\tilde{\mathbf{r}}} = [\mathbf{D} \quad \mathbf{0}] \tilde{\mathbf{y}} \quad (\text{A10})$$

where \mathbf{D} is a diagonal matrix approximating \mathbf{S}^{-1} . This gives $\mathbf{H} = \mathbf{V}[\mathbf{D} \quad \mathbf{0}]\mathbf{U}^T$ and we obtain the resolution matrix

$$\mathbf{R} = \mathbf{VDSV}^T \quad (\text{A11})$$

Assuming that the mean value of \mathbf{r} is known to be \mathbf{r}_0 we use the estimate in equation (A6) and the error vector is

$$\mathbf{e} = \hat{\mathbf{r}} - \mathbf{r} = (\mathbf{HP} - \mathbf{I})(\mathbf{r} - \mathbf{r}_0) + \mathbf{H}\mathbf{w} \quad (\text{A12})$$

which gives the transformed error vector

$$\tilde{\mathbf{e}} = \hat{\tilde{\mathbf{r}}} - \tilde{\mathbf{r}} = [\mathbf{DS} - \mathbf{I}](\tilde{\mathbf{r}} - \tilde{\mathbf{r}}_0) + [\mathbf{D} \quad \mathbf{0}]\tilde{\mathbf{w}} \quad (\text{A13})$$

Then

$$E\{\tilde{\mathbf{e}}\tilde{\mathbf{e}}^T\} = [\mathbf{DS} - \mathbf{I}]\mathbf{C}_{\tilde{\mathbf{r}}}[\mathbf{DS} - \mathbf{I}] + [\mathbf{D} \quad \mathbf{0}]\mathbf{C}_{\tilde{\mathbf{w}}} \begin{bmatrix} \mathbf{D} \\ \mathbf{0} \end{bmatrix} \quad (\text{A14})$$

where $\mathbf{C}_{\tilde{\mathbf{r}}} = \mathbf{V}^T \mathbf{C}_{\mathbf{r}} \mathbf{V}$ is the covariance matrix of $\tilde{\mathbf{r}}$ and $\mathbf{C}_{\tilde{\mathbf{w}}}$ is the covariance matrix of $\tilde{\mathbf{w}}$.

An optimal choice of \mathbf{D} is now obtained by minimizing the expected squared error:

$$E\{\|\mathbf{e}\|^2\} = E\{\|\tilde{\mathbf{e}}\|^2\} = \text{tr } E\{\tilde{\mathbf{e}}\tilde{\mathbf{e}}^T\} = \sum_{k=0}^N [(d_k s_k - 1)^2 \sigma_{\tilde{\mathbf{r}}_k}^2 + d_k^2 \sigma_{\tilde{\mathbf{w}}_k}^2] \quad (\text{A15})$$

where tr denotes trace, and $\sigma_{\tilde{\mathbf{r}}_k}^2$ and $\sigma_{\tilde{\mathbf{w}}_k}^2$ are the variances of $\tilde{\mathbf{r}}_k$ and $\tilde{\mathbf{w}}_k$ respectively. In the case that $\mathbf{C}_{\mathbf{r}} = \sigma_r^2 \mathbf{I}$ and $\mathbf{C}_{\mathbf{w}} = \sigma_w^2 \mathbf{I}$, we obtain

$$E\{\|\mathbf{e}\|^2\} = \sum_{k=0}^N [(d_k s_k - 1)^2 \sigma_r^2 + d_k^2 \sigma_w^2] \quad (\text{A16})$$

In some cases we may use the following estimate of the noise variance

$$\begin{aligned}\hat{\sigma}_w^2 &= \frac{1}{L+N+1} \|\mathbf{y} - \mathbf{P}\hat{\mathbf{r}}\|^2 \\ &= \frac{1}{L+N+1} \left\| \mathbf{y} - \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} [\mathbf{D} \quad \mathbf{0}] \hat{\mathbf{y}} \right\|^2 \\ &= \frac{1}{L+N+1} \left[\sum_{k=0}^N (1 - s_k d_k)^2 \tilde{y}_k^2 + \sum_{k=N+1}^{N+L} \tilde{y}_k^2 \right] \quad (\text{A17})\end{aligned}$$

while the variance of the reflection coefficients, σ_r^2 , must be known *a priori*.

Appendix B

Optimal weighting factor in ridge regression

The estimation problem discussed here is the same as in Appendix A. In ridge regression (Lawson and Hanson 1974) the quadratic form

$$J = \|\mathbf{P}\mathbf{r} - \mathbf{y}\|^2 + \lambda^2 \|\mathbf{r}\|^2 \quad (\text{B1})$$

is minimized. This leads to a linear estimator with a matrix \mathbf{D} in equation (A10) given by

$$\mathbf{D} = \text{diag} [s_0/(s_0^2 + \lambda^2), \dots, s_N/(s_N^2 + \lambda^2)] \quad (\text{B2})$$

Assuming that $\mathbf{C}_r = \sigma_r^2 \mathbf{I}$ and $\mathbf{C}_w = \sigma_w^2 \mathbf{I}$, the expected value of the squared error norm is given in equation (A16) which gives

$$E\{\|\mathbf{e}\|^2\} = \sum_{k=0}^N \frac{\lambda^4 \sigma_r^2 + s_k^2 \sigma_w^2}{(s_k^2 + \lambda^2)^2} \quad (\text{B3})$$

In the case that σ_r^2 and σ_w^2 are assumed to be known constants, the optimal λ^2 (which gives least expected squared error) is found to be

$$\lambda^2 = \sigma_w^2 / \sigma_r^2 \quad (\text{B4})$$

This value of λ is equal to the optimal SVD cut-off level with the same assumption (see Appendix C). It is also interesting to note that for this last case, ridge regression (with the optimal value of λ) and Bayes estimation give the same result (Tarantola and Valette 1982).

The optimal value of λ^2 in equation (B4) is used in equation (B3) giving

$$E\{\|\mathbf{e}\|^2\} = \sum_{k=0}^N \frac{\sigma_w^2}{s_k^2 + \lambda^2} \quad (\text{B5})$$

When σ_w^2 is estimated from equation (A17), we obtain using equation (B2):

$$\hat{\sigma}_w^2 = \frac{1}{N+L+1} \left[\sum_{k=0}^N \frac{\lambda^4 \tilde{y}_k^2}{(\lambda^2 + s_k^2)^2} + \sum_{k=N+1}^{N+L} \tilde{y}_k^2 \right] \quad (\text{B6})$$

The result in equation (B4) is no longer valid since σ_w^2 depends on λ^2 . Therefore the expression for $\hat{\sigma}_w^2$ in equation (B6) must be used in equation (B3) and the expected squared error norm must be minimized numerically with respect to λ^2 to give the optimal value of λ^2 .

Appendix C

Optimal SVD cut-off level

With the SVD cut-off method, we use the matrix **D** in Appendix A with elements

$$d_k = \begin{cases} 1/s_k & k = 0, 1, \dots, K \\ 0 & k = K + 1, \dots, N \end{cases} \quad (C1)$$

The expected squared error in equation (A15) now is

$$E\{\|\mathbf{e}\|^2\} = \sum_{k=K+1}^N \sigma_{\tilde{r}_k}^2 + \sum_{k=0}^K (\sigma_{\tilde{w}_k}^2/s_k^2) \quad (C2)$$

which may be minimized with respect to K to give the optimal SVD cut-off level. This idea has been used by Shim and Cho (1981), but our results differ from theirs. In the case that $\mathbf{C}_r = \sigma_r^2 \mathbf{I}$ and $\mathbf{C}_w = \sigma_w^2 \mathbf{I}$, we may use equation (A16) which gives

$$E\{\|\mathbf{e}\|^2\} = \sigma_r^2(N - K) + \sigma_w^2 \sum_{k=0}^K (1/s_k^2) \quad (C3)$$

Then the expected squared error is minimum if we choose K such that

$$s_K \geq \sigma_w/\sigma_r \quad (C4a)$$

and

$$s_{K+1} < \sigma_w/\sigma_r \quad (C4b)$$

This result has also been derived by Van Riel (1982) using signal-to-noise ratio considerations. Assuming that \mathbf{r} and \mathbf{w} are white, we may normalize the transformed variables $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{w}}$, so that they have unit variance. This gives

$$\frac{\tilde{y}_i}{\sigma_w} = \frac{\sigma_r}{\sigma_w} s_i \frac{\tilde{r}_i}{\sigma_r} + \frac{\tilde{w}_i}{\sigma_w} \quad (C5)$$

In this equation, the contribution from the signal dominates as long as

$$\frac{\sigma_r}{\sigma_w} s_i \geq 1 \quad (C6)$$

Using this criterion to select the cut-off level, we again obtain equation (C4).

In practice σ_w^2 has to be estimated and σ_r^2 has to be given *a priori*. Using equation (A17) and (C1) we obtain

$$\hat{\sigma}_w^2 = \frac{1}{N + L + 1} \sum_{k=K+1}^{N+L} \tilde{y}_k^2 \quad (C7)$$

which gives

$$E\{\|\mathbf{e}\|^2\} = \sigma_r^2(N - K) + \frac{1}{N + L + 1} \sum_{k=K+1}^{N+L} \tilde{y}_k^2 \sum_{k=0}^K 1/s_k^2 \quad (C8)$$

This expression has to be minimized numerically with respect to K to find the optimal cut-off value.

Appendix D

Non-linear parameter estimation

We consider the estimation problem

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \mathbf{w} \quad (\text{D1})$$

where the data vector \mathbf{y} , the model vector $\mathbf{f}(\mathbf{x})$ and the noise (or error) vector \mathbf{w} are of dimension $M \times 1$, \mathbf{x} is an unknown parameter vector of dimension $N \times 1$. We shall assume that $M \geq N$, that \mathbf{x} and \mathbf{w} are uncorrelated stochastic variables with mean values $E\{\mathbf{x}\} \neq \mathbf{0}$ and $E\{\mathbf{w}\} = \mathbf{0}$, and covariance matrices:

$$\text{cov } \mathbf{x} = \mathbf{C}_x \quad \text{cov } \mathbf{w} = \mathbf{C}_w = \sigma_w^2 \mathbf{I}$$

The non-linear problem in equation (D1) is often solved iteratively by linearization in each step and using the method of least-squares (Jackson 1972; Wiggins 1972). It is assumed that the second-order error terms in the first-order Taylor expansion of $\mathbf{f}(\mathbf{x})$ are small compared to the noise terms, so that we may write

$$\Delta \mathbf{y}_k = \mathbf{F}_k \Delta \mathbf{x} + \mathbf{w} \quad (\text{D2})$$

where $\Delta \mathbf{y}_k = \mathbf{y} - \mathbf{f}(\hat{\mathbf{x}}_k)$, $\Delta \mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}_k$, and the elements of the Jacobian matrix \mathbf{F}_k are $f_{k,ij} = \partial f_i / \partial x_j$ evaluated at $\mathbf{x} = \hat{\mathbf{x}}_k$, the estimate of \mathbf{x} at iteration step number k . The next estimate of the parameter vector is $\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \Delta \hat{\mathbf{x}}_{k+1}$, and $\Delta \hat{\mathbf{x}}_{k+1}$ is found by minimizing $\|\Delta \mathbf{y}_k - \mathbf{F}_k \Delta \mathbf{x}\|^2$ which gives

$$\Delta \hat{\mathbf{x}}_{k+1} = (\mathbf{F}_k^T \mathbf{F}_k)^{-1} \mathbf{F}_k^T \Delta \mathbf{y}_k \quad (\text{D3})$$

This least-squares solution may be stabilized in each step using SVD cut-off or ridge regression as discussed in the previous appendices. We may write

$$\Delta \hat{\mathbf{x}}_{k+1} = \mathbf{H}_k \Delta \mathbf{y}_k = \mathbf{V}_k [\mathbf{D}_k \quad \mathbf{0}] \mathbf{U}_k^T \Delta \mathbf{y}_k \quad (\text{D4})$$

where the SVD of \mathbf{F}_k is

$$\mathbf{F}_k = \mathbf{U}_k \begin{bmatrix} \mathbf{S}_k \\ \mathbf{0} \end{bmatrix} \mathbf{V}_k^T \quad (\text{D5})$$

We now obtain, with $\mathbf{R}_k = \mathbf{H}_k \mathbf{F}_k$,

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \mathbf{R}_k (\mathbf{x} - \hat{\mathbf{x}}_k) + \mathbf{H}_k \mathbf{w} \quad (\text{D6})$$

and then

$$E\{\hat{\mathbf{x}}_{k+1}\} = E\{\hat{\mathbf{x}}_k\} + \mathbf{R}_k (E\{\mathbf{x}\} - E\{\hat{\mathbf{x}}_k\}) \quad (\text{D7})$$

which gives

$$E\{\hat{\mathbf{x}}_{k+1}\} - E\{\mathbf{x}\} = (\mathbf{I} - \mathbf{R}_k)(E\{\hat{\mathbf{x}}_k\} - E\{\mathbf{x}\}) \quad (\text{D8})$$

With $\hat{\mathbf{x}}_0 = \mathbf{x}_0$ as initial value, this results in

$$E\{\hat{\mathbf{x}}_{k+1}\} - E\{\mathbf{x}\} = (\mathbf{I} - \mathbf{R}_k)(\mathbf{I} - \mathbf{R}_{k-1}) \cdots (\mathbf{I} - \mathbf{R}_0)(\mathbf{x}_0 - E\{\mathbf{x}\}) \quad (\text{D9})$$

From equations (D6) and (D7) we obtain

$$\hat{\mathbf{x}}_{k+1} - E\{\hat{\mathbf{x}}_{k+1}\} = (\mathbf{I} - \mathbf{R}_k)(\hat{\mathbf{x}}_k - E\{\hat{\mathbf{x}}_k\}) + \mathbf{R}_k(\mathbf{x} - E\{\mathbf{x}\}) + \mathbf{H}_k \mathbf{w} \quad (\text{D10})$$

The first steps are

$$\hat{\mathbf{x}}_1 - E\{\hat{\mathbf{x}}_1\} = \mathbf{R}_0(\mathbf{x} - E\{\mathbf{x}\}) + \mathbf{H}_0 \mathbf{w}$$

$$\hat{\mathbf{x}}_2 - E\{\hat{\mathbf{x}}_2\} = [(\mathbf{I} - \mathbf{R}_1)\mathbf{R}_0 + \mathbf{R}_1](\mathbf{x} - E\{\mathbf{x}\}) + [(\mathbf{I} - \mathbf{R}_1)\mathbf{H}_0 + \mathbf{H}_1]\mathbf{w}$$

....

Since \mathbf{x} and \mathbf{w} are assumed to be uncorrelated, we obtain for the covariance matrix of the estimate at step $k + 1$

$$\mathbf{C}_{\hat{\mathbf{x}}_{k+1}} = \mathbf{C}_{1,k} \mathbf{C}_x \mathbf{C}_{1,k}^T + \mathbf{C}_{2,k} \mathbf{C}_w \mathbf{C}_{2,k}^T \quad (\text{D11})$$

where

$$\mathbf{C}_{1,k} = (\mathbf{I} - \mathbf{R}_k) \mathbf{C}_{1,k-1} + \mathbf{R}_k \quad (\text{D12a})$$

and

$$\mathbf{C}_{2,k} = (\mathbf{I} - \mathbf{R}_k) \mathbf{C}_{2,k-1} + \mathbf{H}_k \quad (\text{D12b})$$

with starting values $\mathbf{C}_{1,-1} = \mathbf{C}_{2,-1} = \mathbf{0}$ for $k = 0$. Since $\mathbf{C}_w = \sigma_w^2 \mathbf{I}$ is unknown, we use the residual in each step (for $k = 0, 1, \dots$) to compute approximately:

$$\hat{\sigma}_w^2 = \frac{1}{M} \|\Delta \mathbf{y}_k - \mathbf{F}_k \Delta \hat{\mathbf{x}}_{k+1}\|^2 = \frac{1}{M} \|\Delta \tilde{\mathbf{y}}_k - \mathbf{D}_k \mathbf{S}_k \Delta \tilde{\mathbf{y}}_k\|^2 \quad (\text{D13})$$

From equation (D9) it is seen that the final estimate is unbiased only if $E\{\mathbf{x}\} = \mathbf{x}_0$ or if $\mathbf{R}_k = \mathbf{I}$ for any value of k (assuming that the linearized analysis is valid).

The analysis leading to the optimal SVD cut-off value or the optimal ridge regression weighting factor must also be modified for this iterative scheme. With $\mathbf{e}_k = \hat{\mathbf{x}}_k - \mathbf{x}$ equation (D6) gives

$$\mathbf{e}_{k+1} = (\mathbf{I} - \mathbf{R}_k) \mathbf{e}_k + \mathbf{H}_k \mathbf{w} \quad (\text{D14})$$

Using the initial values the result is

$$\mathbf{e}_{k+1} = \mathbf{B}_k \mathbf{e}_0 + \mathbf{C}_{2,k} \mathbf{w} \quad (\text{D15})$$

where \mathbf{B}_k is given by

$$\mathbf{B}_k = (\mathbf{I} - \mathbf{R}_k)(\mathbf{I} - \mathbf{R}_{k-1}) \cdots (\mathbf{I} - \mathbf{R}_0) \quad (\text{D16})$$

and $\mathbf{C}_{2,k}$ is given in equation (D12b). We obtain

$$E\{\mathbf{e}_{k+1} \mathbf{e}_{k+1}^T\} = \mathbf{B}_k \mathbf{C}_x \mathbf{B}_k^T + \mathbf{C}_{2,k} \mathbf{C}_w \mathbf{C}_{2,k}^T \quad (\text{D17})$$

with the assumption $\mathbf{x}_0 = E\{\mathbf{x}\}$. For computational purposes we note that

$$\begin{aligned} E\{\mathbf{e}_{k+1} \mathbf{e}_{k+1}^T\} &= (\mathbf{I} - \mathbf{R}_k) \mathbf{B}_{k-1} \mathbf{C}_x \mathbf{B}_{k-1}^T (\mathbf{I} - \mathbf{R}_k) \\ &\quad + [(\mathbf{I} - \mathbf{R}_k) \mathbf{C}_{2,k-1} + \mathbf{H}_k] \mathbf{C}_w [\mathbf{C}_{2,k-1}^T (\mathbf{I} - \mathbf{R}_k) + \mathbf{H}_k^T] \end{aligned} \quad (\text{D18})$$

which gives

$$\begin{aligned} E\{\tilde{\mathbf{e}}_{k+1} \tilde{\mathbf{e}}_{k+1}^T\} &= (\mathbf{I} - \mathbf{D}_k \mathbf{S}_k) \mathbf{V}_k^T \mathbf{B}_{k-1} \mathbf{C}_x \mathbf{B}_{k-1}^T \mathbf{V}_k (\mathbf{I} - \mathbf{D}_k \mathbf{S}_k) \\ &\quad + \hat{\sigma}_w^2 [(\mathbf{I} - \mathbf{D}_k \mathbf{S}_k) \mathbf{V}_k^T \mathbf{C}_{2,k-1} \mathbf{U}_k + [\mathbf{D}_k \quad \mathbf{0}]] \\ &\quad \times \left[\mathbf{U}_k^T \mathbf{C}_{2,k-1}^T \mathbf{V}_k (\mathbf{I} - \mathbf{D}_k \mathbf{S}_k) + \begin{bmatrix} \mathbf{D}_k \\ \mathbf{0} \end{bmatrix} \right] \end{aligned} \quad (\text{D19})$$

The trace of this matrix is the norm of the error vector which may be minimized in each step to give the optimal matrix \mathbf{D}_k . $\hat{\sigma}_w^2$ also depends on \mathbf{D}_k as given in equation (D13), and this expression is used in equation (D19) when the norm of the error vector is minimized. We have used the SVD cut-off method (as described in Appendix C), but the ridge regression method described in Appendix B may also be used.

REFERENCES

- BARD, Y. (1974). *Non-linear parameter estimation* (Academic Press, New York).
- BERKHOUT, A. J. (1977). Least-squares inverse filtering and wavelet deconvolution. *Geophysics*, **42**, 1369–1383.
- CLAERBOUT, J. F. (1976). *Fundamentals of geophysical data processing: with application to petroleum prospecting* (McGraw-Hill, New York).
- DONGARRA, J. J., MOLER, C. B., BUNCH, J. R., and STEWART, G. W. (1979). *LINPACK Users' guide* (SIAM, Philadelphia).
- EKSTROM, M. P. (1973). A spectral characterization of the ill-conditioning in deconvolution. *IEEE Trans. Audio and Electroacoustics*, **AU-21**, 344–348.
- GJØYSTDAL, H., and URSIN, B. (1981). Inversion of reflection times in three dimensions. *Geophysics*, **46**, 972–983.
- GOLUB, G. H., and REINSCH, C. (1970). Singular value decomposition and least squares solution, *Numerical Math.*, **14**, 403–420.
- JACKSON, D. D. (1972). Interpretation of inaccurate, insufficient and inconsistent data. *Geophys. J. R. Astron. Soc.*, **28**, 97–110.
- LAWSON, C. L., and HANSON, R. J. (1974). *Solving least squares problems* (Prentice-Hall, Englewood Cliffs, New Jersey).
- LEVY, S., and CLOWES, R. M. (1980). Debubbling: A generalized linear inverse approach, *Geophysical Prospecting*, **28**, 840–858.
- LINDLEY, D. V., and MILLER, J. C. P. (1968). *Cambridge elementary statistical tables* (Cambridge University Press, Cambridge).
- LINES, L. R., and TREITEL, S. (1984). Tutorial: A review of least-squares inversion and its application to geophysical problems. *Geophysical Prospecting*, **32**, 159–186.
- SHIM, Y. S., and CHO, Z. H. (1981). SVD pseudoinversion image reconstruction. *IEEE Trans. ASSP*, **29**, 904–909.
- SILVA, J. B. C., and HOHMANN, G. W., 1983, Nonlinear magnetic inversion using a random search method. *Geophysics*, **48**, 1645–1658.
- SILVEY, S. D. (1975). *Statistical Inference* (Chapman and Hall, London).
- TARANTOLA, A., and VALETTE, B. (1982). Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys. Space Phys.* **20**, 219–232.
- TUFTS, D. W., and KUMARESAN, R. (1982). Singular value decomposition and improved frequency estimation using linear prediction. *IEEE Trans. ASSP*, **30**, 671–675.
- URSIN, B. (1981). Error estimates for inverse modeling schemes using seismic traveltimes. *Geophysics*, **46**, 1227–1234.
- URSIN, B., and HOLBERG, O. (1985). Maximum likelihood estimation of seismic impulse responses. *Geophysical Prospecting*, **33**, 233–251.
- VAN RIEL, P. (1982). Seismic trace inversion, MSc thesis, Delft University of Technology.
- VAN RIEL, P. and BERKHOUT, A. J. (1985). Resolution in seismic trace inversion by parameter estimation, *Geophysics*, **50**, 1440–1455.
- WIGGINS, R. A. (1972). The general linear inverse problem: Implications of surface waves and free oscillations for earth structure. *Rev. Geophys. Space Phys.*, **10**, 251–285.