Maximum-likelihood Q estimation

Dave Hale

Abstract

Assuming a noise-free, autoregressive, non-stationary model, the maximum-likelihood estimate of Q may be computed from a single seismogram, using an iterative algorithm similar to the prediction-error algorithm described in SEP-30 (Hale, 1982). The only difference in the two algorithms is that the maximum-likelihood algorithm is sensitive to the time-varying amplitude as well as color of a seismogram, whereas the prediction-error algorithm is sensitive to color variations only.

Maximum-likelihood estimation of Q for a seismogram contaminated with ambient noise is more difficult. The maximum-likelihood formulation of the estimation problem leads to a set of equations which is highly non-linear in the unknown parameters, including Q; and no reasonably efficient algorithm has yet been found which solves this estimation problem.

Introduction

In SEP-30 (Hale, 1982), I described a method for deconvolving attenuated seismograms. I called this method "Q-adaptive deconvolution" (QAD) to distinguish it from more conventional adaptive or time-varying deconvolutions. The latter methods typically assume no model for the time-varying color of seismograms, whereas QAD is based on a model which attributes this non-stationarity to inelastic attenuation. In addition to estimating autoregressive (AR) coefficients, as in conventional predictive deconvolution, QAD estimates the quality factor Q.

The criterion used by QAD to estimate Q and AR coefficients is that of minimizing a sum of squared prediction errors, the same criterion used in predictive deconvolution. In this paper, I describe an algorithm for computing the maximum-likelihood (ML) estimate of Q from a noise-free seismogram; and I compare this estimator with the least-squares algorithm described in SEP-30. The ML algorithm is derived assuming a Gaussian distribution for reflection coefficients, but experience with synthetic examples has shown that it yields good estimates even when this assumption is not satisfied. In practice, the poorer assumption is that of a noiseless seismogram. In the last section of the paper, I describe a ML estimator for noisy seismograms, and discuss some of its computational aspects.

The likelihood function for a noise-free model

Let us first assume that a seismogram may be represented by the following linear model:

$$\mathbf{y} = \mathbf{FQDr} \tag{1}$$

$\mathbf{y}$ denotes the $N\times 1$ column vector $\mathbf{y} \equiv (y_1\ y_2\ \cdots\ y_N)^T$, where y_t is the t'th sample of a noise-free seismogram. $\mathbf{r}$ denotes an $N\times 1$ column vector containing the sampled response of a stratified, non-attenuating earth to an impulsive plane wave. Neglecting multiple reflections and transmission losses, $\mathbf{r}$ is a vector of reflection coefficients. Multiplication by the $N\times N$ "divergence" matrix $\mathbf{D}$ converts this plane wave response to a point source response. For a constant velocity earth, the elements of $\mathbf{D}$ are $D_{ts} = \delta_{t-s}/t$. Multiplication by the $N\times N$ "attenuation" matrix $\mathbf{Q}$ converts the response of a non-attenuating earth to that of an attenuating earth. The elements of $\mathbf{Q}$ depend only on the quality factor Q or, equivalently, Q^{-1}. $\mathbf{F}$ is an $N\times N$ Toeplitz matrix with coefficients of the source waveform f_t on its diagonals; $F_{ts} = f_{t-s}$. I use the term "source waveform" loosely to include near surface reverberations at source and receiver locations, as well as distortions introduced by recording instruments. Multiplication by $\mathbf{F}$ is equivalent to convolution with this composite source waveform.

Assuming that the reflection coefficients are independently and normally distributed with constant variance σ_r^2 and zero mean, the probability density function (PDF) of **r** is

$$p_{\mathbf{R}}(\mathbf{r}) = \frac{1}{(2\pi\sigma_r^2)^{N/2}} e^{-\frac{\mathbf{r}^T\mathbf{r}}{2\sigma_r^2}}$$

Using the model of equation (1), the PDF of the seismogram **y** is

$$p_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi\sigma_r^2)^{N/2}|\mathbf{FQD}|} e^{-\frac{1}{2\sigma_r^2}\mathbf{y}^T\mathbf{F}^{-T}\mathbf{Q}^{-T}\mathbf{D}^{-T}\mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{F}^{-1}\mathbf{y}} \qquad (2)$$

where $|\mathbf{FQD}|$ denotes the determinant of the matrix **FQD**. The PDF of **y** depends on the unknown σ_r^2, Q^{-1}, and source waveform, and is called the likelihood function of these unknown parameters. Our goal in ML estimation is to find the PDF which is most likely to yield a given seismogram **y**. Define **f** to be a vector containing the source waveform coefficients; $\mathbf{f} = (f_0 \; f_1 \; \cdots \; f_{l_f})$. The ML estimates of σ_r^2, Q^{-1}, and **f** are those values for which the likelihood function $p_{\mathbf{Y}}(\mathbf{y};\sigma_r^2,Q^{-1},\mathbf{f})$ attains its maximum.

Defining $\mathbf{E} \equiv \mathbf{D}^{-1}$, $\mathbf{P} \equiv \mathbf{Q}^{-1}$, and $\mathbf{A} \equiv \mathbf{F}^{-1}$ so that $\mathbf{r} = \mathbf{EPAy}$, the likelihood function may be rewritten as

$$p_{\mathbf{Y}}(\mathbf{y};\sigma_r^2,Q^{-1},\mathbf{a}) = \frac{|\mathbf{EPA}|}{(2\pi\sigma_r^2)^{N/2}} e^{-\frac{1}{2\sigma_r^2}\mathbf{y}^T\mathbf{A}^T\mathbf{P}^T\mathbf{E}^T\mathbf{EPAy}} \qquad (3)$$

Instead of estimating the source waveform **f**, we estimate its inverse **a**. To simplify the estimation of **a**, we should assume that **a** is causal, which is equivalent to assuming that **f** is minimum-phase. This assumption is justified by the fact that the likelihood function of equation (2) is independent of the phase of **f**; so we may as well choose the phase which is most convenient. Even with this assumption, equation (3) implies that **a** and σ_r^2 cannot be uniquely determined since multiplication of both by a scale factor leaves the likelihood function unchanged. We may remove this ambiquity by constraining the first coefficient of **a** to be unity; i.e., $\mathbf{a} = (1 \; a_1 \; a_2 \; \cdots \; a_{l_a})$. In the next section, I describe an algorithm for determining the parameters σ_r^2, Q^{-1}, and **a** which maximize $p_{\mathbf{Y}}(\mathbf{y};\sigma_r^2,Q^{-1},\mathbf{a})$.

Parameter estimation and deconvolution

The problem of maximizing the likelihood $p_{\mathbf{Y}}(\mathbf{y};\sigma_r^2,Q^{-1},\mathbf{a})$ of equation (3) is equivalent to that of minimizing $L(\sigma_r^2,Q^{-1},\mathbf{a})$ defined by

$$L(\sigma_r^2,Q^{-1},\mathbf{a}) \equiv -2\ln p_{\mathbf{Y}}(\mathbf{y};\sigma_r^2,Q^{-1},\mathbf{a}) \tag{4}$$

$$= \frac{\mathbf{y}^T\mathbf{A}^T\mathbf{P}^T\mathbf{E}^T\mathbf{EPAy}}{\sigma_r^2} + N\ln\sigma_r^2 - 2\ln|\mathbf{EPA}| + constant$$

$$= \frac{\mathbf{r}^T\mathbf{r}}{\sigma_r^2} + N\ln\sigma_r^2 - 2\ln|\mathbf{E}| - 2\ln|\mathbf{P}| - 2\ln|\mathbf{A}| + constant$$

Remember that $\mathbf{r} = \mathbf{EPAy}$ is a function of Q^{-1}, $\mathbf{a}$, and $\mathbf{y}$. The determinant of $\mathbf{E}$ does not depend on the unknown parameters, so it may be treated as constant. And the assumption that $\mathbf{a} = (1\ a_1\ \cdots\ a_{la})$ results in $|\mathbf{A}| = 1$.

The only determinant which depends on the parameters is that of $\mathbf{P}$ which depends on Q^{-1}. In SEP-30, I showed that

$$\mathbf{P} = \mathbf{I} + \pi Q^{-1}\mathbf{TG} + \frac{1}{2}\pi^2 Q^{-2}\mathbf{T}^2\mathbf{G}^2 + \cdots \tag{5}$$

$$= \sum_{j=0}^{\infty}\frac{1}{j!}\pi^j Q^{-j}\mathbf{T}^j\mathbf{G}^j$$

where $\mathbf{T}$ is an $N\times N$ diagonal matrix with elements $T_{ts} = t\,\delta_{t-s}$, and $\mathbf{G}$ is an $N\times N$ Toeplitz matrix with elements $G_{ts} = g_{t-s}$ given by

$$g_t = \begin{cases} 1/4 & , \quad t = 0 \\ -2/(\pi t)^2 & , \quad t = 1,3,5,\cdots \\ 0 & , \quad otherwise \end{cases}$$

Note that $\mathbf{P}$ is lower triangular with diagonal elements $P_{tt} = \exp(\pi t/4Q)$. The determinant of $\mathbf{P}$ is the product of the elements on the diagonal so that $L(\sigma_r^2,Q^{-1},\mathbf{a})$ of equation (4) may be written as

$$L(\sigma_r^2,Q^{-1},\mathbf{a}) = \frac{\mathbf{r}^T\mathbf{r}}{\sigma_r^2} + N\ln\sigma_r^2 - 2\ln\prod_{t=1}^{N} e^{\frac{\pi t}{4Q}}$$

$$= \frac{\mathbf{r}^T\mathbf{r}}{\sigma_r^2} + N\ln\sigma_r^2 - \frac{N(N+1)\pi}{4Q}$$

where I have dropped the constant terms.

Differentiating L with respect to each of the unknown parameters, one finds that the ML estimates must satisfy the following equations:

$$0 = \hat{\sigma}_r^2 - \frac{\hat{\mathbf{r}}^T\hat{\mathbf{r}}}{N} \tag{6a}$$

$$0 = \hat{\mathbf{r}}^T \frac{\partial \hat{\mathbf{r}}}{\partial Q^{-1}} - \frac{N(N+1)\pi\hat{\sigma}_r^2}{8} \tag{6b}$$

$$0 = \hat{\mathbf{r}}^T \frac{\partial \hat{\mathbf{r}}}{\partial a_k} \quad ; \quad k = 1,2, \cdots ,la \tag{6c}$$

where $\hat{\mathbf{r}}$ is the estimate of $\mathbf{r}$ obtained by using the ML estimates of Q^{-1} and $\mathbf{a}$ in deconvolving the seismogram $\mathbf{y}$. $\partial\hat{\mathbf{r}}/\partial Q^{-1}$ should be interpreted as $\partial\mathbf{r}/\partial Q^{-1}$ evaluated at the ML estimates $(\hat{\sigma}_r^2, \hat{Q}^{-1}, \hat{\mathbf{a}})$. A similar interpretation applies to $\partial\hat{\mathbf{r}}/\partial a_k$.

Equations (6) may be simplified somewhat by using equation (6a) to eliminate $\hat{\sigma}_r^2$ in equation (6b), yielding

$$0 = \hat{\mathbf{r}}^T \left[\frac{\partial \hat{\mathbf{r}}}{\partial Q^{-1}} - \frac{(N+1)\pi}{8}\hat{\mathbf{r}} \right] \tag{7a}$$

$$0 = \hat{\mathbf{r}}^T \frac{\partial \hat{\mathbf{r}}}{\partial a_k} \quad ; \quad k = 1,2, \cdots ,la \tag{7b}$$

Unfortunately, because $\hat{\mathbf{r}}$ is a non-linear function of Q^{-1}, these equations are non-linear in the parameters Q^{-1} and $\mathbf{a}$. Hence, their simultaneous solution must be found iteratively. Notice, however, that equation (7b) is linear in the unknown $\mathbf{a}$ and is, in fact, exactly that which we solve in conventional unit-lag predictive deconvolution, where we assume $Q^{-1} = 0$. This linearity greatly simplifies the iterative solution of equations (7).

In general, time-variable filters do not commute. To a good approximation, however, the matrices $\mathbf{E}$, $\mathbf{P}$, and $\mathbf{A}$ do commute. Correction for spherical divergence (multiplication by $\mathbf{E}$), for example, is commonly performed prior to predictive deconvolution (multiplication by $\mathbf{A}$). We therefore assume $\mathbf{r} = \mathbf{EPAy} \approx \mathbf{APEy}$, and compute $\mathbf{Ey}$ prior to estimating Q^{-1} and $\mathbf{a}$.

Assuming that we have a guess for Q^{-1}, we can compute $\mathbf{PEy}$ and then solve equations (7b) for $\mathbf{a}$ using the standard predictive deconvolution algorithm. However, unless our guess of Q^{-1} was very good, equation (7a) will not be satisfied, so we must perturb the estimate $\hat{Q}^{-1}$ by an amount proportional to the size of the right-hand-side of equation (7a). The iterative algorithm is discussed further in SEP-30 (Hale, 1982) and is given below using the notation of this paper:

Initially $\hat{Q}^{-1} = guess$

(1) compute **PEy**

(2) compute $\hat{\mathbf{a}}$ and $\hat{\mathbf{r}}$ from **PEy** to satisfy equations (7b)

(3) compute Δ, a perturbation to $\hat{Q}^{-1}$

(4) $\hat{Q}^{-1} = \hat{Q}^{-1} - \Delta$

(5) if $|\Delta| > small$ go to (1)

Converged $\hat{\mathbf{r}}$ is the deconvolved seismogram

Δ is computed using equation (7a) and Newton's method to be

$$\Delta = \frac{\hat{\mathbf{r}}^T \left[\frac{\partial \hat{\mathbf{r}}}{\partial Q^{-1}} - \frac{(N+1)\pi}{8} \hat{\mathbf{r}} \right]}{\frac{\partial \hat{\mathbf{r}}^T}{\partial Q^{-1}} \frac{\partial \hat{\mathbf{r}}}{\partial Q^{-1}} + \hat{\mathbf{r}}^T \frac{\partial^2 \hat{\mathbf{r}}}{\partial (Q^{-1})^2} - \frac{(N+1)\pi}{4} \hat{\mathbf{r}}^T \frac{\partial \hat{\mathbf{r}}}{\partial Q^{-1}}}$$

$$\approx \frac{1}{\pi} \frac{\hat{\mathbf{r}}^T \mathbf{T}\mathbf{G}\hat{\mathbf{r}} - \frac{(N+1)}{8} \hat{\mathbf{r}}^T \hat{\mathbf{r}}}{(\mathbf{T}\mathbf{G}\hat{\mathbf{r}})^T (\mathbf{T}\mathbf{G}\hat{\mathbf{r}})} \tag{8}$$

where the last approximation follows from keeping only the first term in the denominator. Although including the second and third terms should improve the convergence rate of the algorithm, I have not found the expense of computing these terms worthwhile. I have also approximated the derivative of **r** as $\partial \mathbf{r} / \partial Q^{-1} \approx \pi \mathbf{T}\mathbf{G}\mathbf{r}$. From equation (5), this approximation is valid insofar as $\mathbf{T}\mathbf{G} \approx \mathbf{G}\mathbf{T}$. Although not absolutely necessary, this approximation is desirable because it permits Δ to be expressed in terms of the deconvolved seismogram estimate $\hat{\mathbf{r}}$, which we have already computed in step (2) of the above algorithm.

The algorithm iterates until $\Delta \approx 0$, that is, until equation (7a) is satisfied. Remember that equation (7b), the predictive deconvolution equation, is satisfied at each iteration. The only difference between the ML algorithm and the prediction-error algorithm of SEP-30 is in the computation of Δ. In the notation of this paper, the latter algorithm computes Δ according to

$$\Delta = \frac{1}{\pi} \frac{\hat{\mathbf{r}}^T \mathbf{T}(\mathbf{G} - \tfrac{1}{4}\mathbf{I})\hat{\mathbf{r}}}{[\mathbf{T}(\mathbf{G} - \tfrac{1}{4}\mathbf{I})\hat{\mathbf{r}}]^T [\mathbf{T}(\mathbf{G} - \tfrac{1}{4}\mathbf{I})\hat{\mathbf{r}}]} \tag{9}$$

Close comparison of the numerators of equations (8) and (9) reveals that the ML and prediction-error algorithms will converge to the same estimates if

$$\sum_{t=1}^{N} t \hat{r}_t^2 = \frac{(N+1)}{2} \sum_{t=1}^{N} \hat{r}_t^2$$

This equation depends only on the amplitude of the r_t, not the color. And the equation is quite reasonable, for taking expectations of both sides yields

$$\sigma_r^2 \sum_{t=1}^{N} t \ = \ \sigma_r^2 \frac{N(N+1)}{2}$$

which is certainly true. The presence of the amplitude terms in equation (8) implies that the ML algorithm estimates Q^{-1} from temporal variations in amplitude as well as color.

The likelihood function for a noisy model

My motive for deriving a ML estimator for Q was not merely to show the similarity between prediction-error and ML estimates, but rather to find an "optimal" Q estimation algorithm for noisy seismograms. Unfortunately, ML estimation becomes much more complicated when ambient noise is introduced into the model as in

$$\mathbf{z} \ = \ \mathbf{y} + \mathbf{n}$$

where

$$\mathbf{y} \ = \ \mathbf{FQDr}$$

as before, and the ambient noise **n** is given by

$$\mathbf{n} \ = \ \mathbf{Cw}$$

I assume that **w** is white noise, normally distributed with covariance matrix $\sigma_w^2\mathbf{I}$, and that **C** is a Toeplitz matrix with elements $C_{ts} = c_{t-s}$. $\mathbf{c} \equiv (1 \ c_1 \ c_2 \ \cdots \ c_{lc})$ is a filter which colors the white noise **w**, and is assumed to be unknown. If one expects to have sharp peaks in the noise spectrum, as is the case, for example, with 60 Hz "cultural" noise, then an autoregressive model for the noise would be preferred. For definiteness, we assume the moving average model above.

Without the presence of the non-Toeplitz matrices **D** and **Q** in this model, we would have no hope of distinguishing between signal and noise from a single seismogram. The non-stationarity of a seismogram gives us a clue as to what is signal and what is noise.

The PDF of **z** is

$$p_{\mathbf{Z}}(\mathbf{z};\sigma_r^2,\sigma_w^2,Q^{-1},\mathbf{f},\mathbf{c}) \ = \ \frac{1}{(2\pi)^{N/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\mathbf{V}^{-1}\mathbf{z}}$$

where the dependence on the unknown parameters is buried in the covariance matrix **V** of **z** given by

$$\mathbf{V} = \mathbf{FQD}\sigma_r^2\mathbf{D}^T\mathbf{Q}^T\mathbf{F}^T + \mathbf{C}\sigma_w^2\mathbf{C}^T$$

The ML estimates of the parameters may, in principle, be found by minimizing

$$L(\sigma_r^2,\sigma_w^2,Q^{-1},\mathbf{f},\mathbf{c}) \equiv -2\ln p_{\mathbf{Z}}(\mathbf{z};\sigma_r^2,\sigma_w^2,Q^{-1},\mathbf{f},\mathbf{c})$$

$$= \ln|\mathbf{V}| + \mathbf{z}^T\mathbf{V}^{-1}\mathbf{z}$$

In practice, however, the minimization of L is made difficult by the fact that no simple expression exists for $|\mathbf{V}|$ and $\mathbf{V}^{-1}$ in terms of the parameters. Going through the matrix calculus, the partial derivative of L with respect to an unknown parameter, say α, is

$$\frac{\partial L}{\partial \alpha} = Trace\left[\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{z}\mathbf{z}^T\mathbf{V}^{-1}\right)\frac{\partial \mathbf{V}}{\partial \alpha}\right]$$

The ML estimates are found by setting this derivative to zero for each of the unknown parameters and then solving the resulting set of non-linear equations.

The solution of these equations must, as in the noise-free case, be obtained by iteration. Now, however, none of the equations appears to be linear in the unknowns; and we cannot easily eliminate any of the equations as we did, for example, equation (6a). I have not found an iterative algorithm which makes the minimization of L any easier than, say, the method of steepest descent; and computation of the gradient of L for even one iteration of steepest descent is expensive. Furthermore, the cost remains high even if one is willing to assume knowledge of some of the unknown parameters.

One way to simplify the problem is to assume that all of the matrices which make up the covariance $\mathbf{V}$ commute with each other. This approximation is justifiable for typical earth values of Q^{-1} and short $\mathbf{f}$ and $\mathbf{c}$; and it permits the determinant of $\mathbf{V}$ to be expressed analytically in terms of σ_r^2, σ_w^2, and Q^{-1}, which in turn greatly simplifies the computation of the gradient of L. The details will be left to a later report in which, optimistically, a more complete discussion of the noisy case will be possible.

Conclusions

For a noise-free seismogram, the ML estimation of autoregressive coefficients and Q is easily accomplished using an iterative algorithm which closely resembles that discussed in SEP-30 (Hale, 1982). ML estimation of these parameters for noise-contaminated seismograms appears to be more difficult and a numerical algorithm has, to my knowledge, not yet been developed to solve this particular problem.

REFERENCE

Hale, D., 1982, Q-adaptive deconvolution: SEP Report 30, p.133-158.