Q and Kalman Filtering

Dave Hale

Abstract

An important first step in applying the Kalman filter to seismic data is the interpretation of the so called "state equations" on which the filter is based. We propose two new interpretations, both of which include the time-varying effects of attenuation and spherical divergence. We derive the first interpretation from a time-varying, moving-average (MA) model of the seismogram, the second from a time-varying, auto-regressive, moving-average (ARMA) model.

The primary difference between the MA and the ARMA approaches is the way in which estimates of reflectivity are computed. With the MA model, fixed-lag smoothed, linear, minimum-error-variance estimates are obtained directly from the most basic Kalman filter algorithm. With the ARMA model, such estimates require costly extensions to the basic algorithm, implying that the ARMA approach may be preferred only when the MA model cannot parsimoniously represent a seismogram. We illustrate both the MA and ARMA interpretations by Kalman filtering synthetic seismograms.

Introduction

Kalman filtering theory provides a means of estimating, in an "optimum" way, random inputs to a known, time-varying, linear system, given noise-contaminated outputs of that system. To apply this theory to seismic data, we may think of the random inputs as reflection coefficients of a layered earth and the outputs as samples of a recorded seismogram. Several papers, including those by Bayless and Brigham (1970), Ott and Meder (1972), Crump (1974), and Mendel and Kormylo (1978), have demonstrated this approach. In all of these papers, the authors assume, as did Kalman (1960), that parameters describing the system are available. For seismic data, these parameters include the source waveform, Q, velocities, etc. The usefulness of Kalman filtering depends on our ability to estimate the system parameters.

We too assume that estimates of the required parameters are available and then address the problem of how to use these estimates in the model represented by the following equations:

$$\mathbf{x}_t = \mathbf{\Phi}_t \mathbf{x}_{t-1} + \boldsymbol{\gamma}_t r_t \qquad (1a)$$

$$z_t = \mathbf{h}_t^T \mathbf{x}_t + n_t \qquad (1b)$$

These are the so called "state equations", central to Kalman filtering; equation (1a) defines the "message model" and equation (1b) defines the "observation model."[1] r_t represents a scalar, zero-mean, white, input process; n_t represents a scalar, zero-mean, white, additive noise process; and z_t represents the recorded data as a discrete function of time t. $\mathbf{x}_t$ is called the "state vector". If $\mathbf{x}_t$ is an $m \times 1$ column vector (m being the "order" of the state equations), then $\boldsymbol{\gamma}_t$ and $\mathbf{h}_t$ are $m \times 1$ column vectors, and $\mathbf{\Phi}_t$ is an $m \times m$ matrix.

If we let r_t represent the reflectivity series, then the three matrices $\boldsymbol{\gamma}_t$, $\mathbf{\Phi}_t$, and $\mathbf{h}_t$ stand between the reflectivity and the noise-contaminated seismogram z_t. These matrices must depend on the seismic system parameters; but one has considerable freedom in specifying which parameters go where, as is evident from the different approaches taken in the four above mentioned papers. Interpretation of the state vector $\mathbf{x}_t$ is also an important step in the application of Kalman filtering to seismic data; and here, too, interpretations differ widely in published geophysical applications.

Perhaps adding to the confusion over which is most suitable, we develop two new interpretations of equations (1) for seismic data, both of which include the time-varying effects of attenuation and spherical divergence. Although non-stationarity provides one of

Actually, equations (1) are a particular form of more general state equations which may include vector (multichannel) processes as well as colored noise. See, for example, Sage and Melsa (1979).

our strongest motives for Kalman filtering, attenuation effects have been ignored in previously published applications.

A seismogram model

We begin by specifying a model for a seismic trace, one which includes the effects of attenuation and spherical divergence but which is easier (perhaps only for geophysicists) to understand than the model of equations (1). In matrix form, our model is

$$\mathbf{z} = \mathbf{FQDr} + \mathbf{n} \tag{2}$$

where

- **z** is a column vector composed of the samples z_t of a seismic trace,
- **F** is a lower-triangular, Toeplitz matrix with samples of the source waveform f_t on its diagonals. The first column of **F** is $\mathbf{f} = [f_0 \, f_1 \, f_2 \, \cdots \, f_l \, 0 \, 0 \, \cdots \, 0]^T$.
- **Q** is a lower-triangular matrix described by Hale (1981) which transforms the impulse response for a non-attenuating medium to that for an attenuating medium. The elements of this matrix depend only on Q, the quality factor of the medium (assumed constant).
- **D** is a diagonal matrix which transforms the impulsive, vertically propagating plane-wave response for a layered medium to the point source response for such a medium. In other words, multiplying by **D** applies spherical divergence. For a constant velocity medium, the diagonal, non-zero elements of **D**, D_{tt}, are proportional to $1/t$.
- **r** is a column vector containing the sampled response r_t of a layered, non-attenuating medium to a vertically propagating, impulsive plane-wave. **r** should include all multiple reflections and transmission losses.
- **n** is a column vector containing samples n_t of additive noise.

Figure 1 illustrates the creation of a synthetic trace based on equation (2). **c** is a column vector (here plotted on its side) 500 samples long which represents a reflection coefficient sequence; the coefficients were derived from a Gaussian distribution of variance 0.05 with a 10% probability of having a non-zero value at any given sample. **r** was computed from **c** via the algorithm given by Claerbout (1976,p.160). **Dr** is the result of applying spherical divergence; **QDr** is the result of applying attenuation for $Q = 100$. **f** is the source waveform with Z-transform $F(Z) = (1 - 0.9Z^3)(1 - 0.9Z^4)$, possibly representing the combined effects of source and receiver ghost reflections in marine data. The signal **s** is then

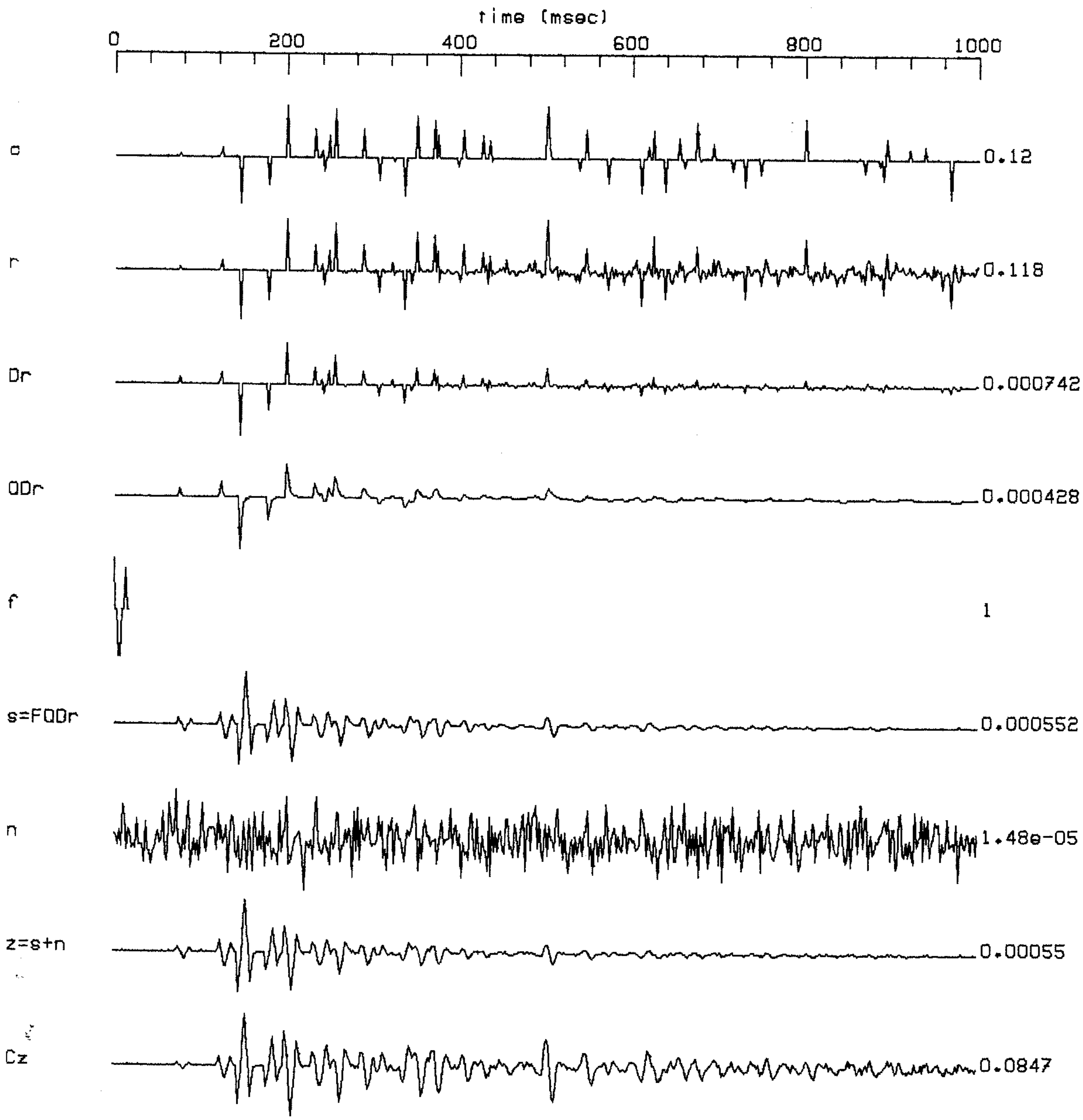

FIG. 1. Creation of a synthetic trace via the model of equation (2). Each trace is scaled independently for plotting; the maximum amplitude is given to the right of each trace. **c** is the reflection coefficient sequence; **r** includes all multiples and transmission losses; **Dr** includes spherical divergence; **QDr** includes attenuation for $Q = 100$; **f** is the source waveform; **s** is the signal including all of the above; **n** is the noise; **z** is the synthetic trace; and **Cz** is **z** corrected for spherical divergence. The sampling interval is 2 msec.

obtained by convolving **f** with the earth's impulse response or, equivalently, **s** = **FQDr**. **n** is a Gaussian random noise sequence of variance 5.0×10^{-6}. The synthetic trace **z** is the sum of the signal **s** and the noise **n**. For display only, we correct for spherical divergence (i.e., multiply z_t by t) to obtain the trace labeled **Cz**.

Because the transformation from **c** to **r** is linear, it could be included as an additional matrix multiplication in equation (2). However, the elements of the added matrix, unlike those of **F**, **Q**, and **D**, cannot be estimated prior to estimating elements of **c**. We, therefore, choose to estimate **r** instead of **c**, though **c** is certainly the more desirable vector.

Linear, minimum-error-variance estimation

Define a linear estimate of **r** by $\hat{\mathbf{r}} = \mathbf{Gz}$ or, equivalently,

$$\hat{r}_t = \sum_s G_{ts} z_s \qquad (3)$$

The linear, minimum-error-variance estimate of **r** is that obtained by minimizing $E[(\hat{r}_t - r_t)^2]$ with respect to **G** where E denotes ensemble, *not time*, averaging. Booton (1952) showed that the following condition is necessary and sufficient for $\hat{\mathbf{r}}$ to be the linear, minimum-error-variance estimate of **r**:

$$E(r_t z_v) = \sum_s G_{ts} E(z_s z_v) \qquad (4)$$

If we assume that both r_t and n_t are white, uncorrelated, stationary random processes, with variances σ_r^2 and σ_n^2, respectively, then equations (2), (3), and (4) can be combined to show that $\hat{\mathbf{r}}$ must satisfy

$$[\mathbf{I} + \alpha(\mathbf{FQD})^{-1}(\mathbf{FQD})^{-T}]\hat{\mathbf{r}} = (\mathbf{FQD})^{-1}\mathbf{z} \qquad (5)$$

where $\alpha = \sigma_n^2 / \sigma_r^2$. An alternative form, particularly useful if **FQD** is singular, is

$$[\alpha\mathbf{I} + (\mathbf{FQD})^T(\mathbf{FQD})]\hat{\mathbf{r}} = (\mathbf{FQD})^T\mathbf{z}$$

Verify that Equation (5) yields the expected results in the limits of very small and very large α. As $\alpha \to 0$ (no noise), $\hat{\mathbf{r}} \to \mathbf{r}$; and as $\alpha \to big$, $\hat{\mathbf{r}}$ becomes a scaled, matched filtered version of **z**. [Multiplication of **z** by $(\mathbf{FQD})^T$ is equivalent to matched filtering of z_t.]

If the source waveform f_t is minimum-phase, then $(\mathbf{FQD})^{-1}$ will be lower-triangular. In general, for any non-zero α, the matrix on the left-hand side of equation (5) will contain both upper and lower-triangular non-zero elements. Hence, the estimate $\hat{\mathbf{r}}$ cannot be found by back-substitution (i.e., $\hat{r}_t$ cannot be found by feedback-filtering). This matrix is, however, symmetric and positive-definite; and these properties can be exploited in the computation of $\hat{\mathbf{r}}$. We can also assume that the matrix is effectively banded, that non-zero elements far from the main diagonal are small enough to be neglected. Bandedness may make numerical solution of equation (5) quite feasible for seismic data, particularly if the same filter is to be applied to many consecutive traces (likely since we have assumed **FQD** is a known matrix).

Numerical solution of equation (5), however, is not Kalman filtering. The reader should find the intuitive handles offered by equation (5) useful in interpreting the equations and results of Kalman filtering, for, like the solution of equation (5), Kalman filtering provides a linear, minimum-error-variance estimate of **r**.

KMA -- Kalman filtering for a MA model

The Kalman filter is summarized by the following five equations (after Mendel and Kormylo, 1978):

Prediction:

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{\Phi}_t \hat{\mathbf{x}}_{t-1|t-1} \tag{6a}$$

$$\mathbf{P}_{t|t-1} = \mathbf{\Phi}_t \mathbf{P}_{t-1|t-1} \mathbf{\Phi}_t^T + \boldsymbol{\gamma}_t \sigma_r^2 \boldsymbol{\gamma}_t^T \tag{6b}$$

Correction:

$$\mathbf{k}_t = \mathbf{P}_{t|t-1} \mathbf{h}_t [\mathbf{h}_t^T \mathbf{P}_{t|t-1} \mathbf{h}_t + \sigma_n^2]^{-1} \tag{6c}$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{k}_t (z_t - \mathbf{h}_t^T \hat{\mathbf{x}}_{t|t-1}) \tag{6d}$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{k}_t \mathbf{h}_t^T) \mathbf{P}_{t|t-1} \tag{6e}$$

$\hat{\mathbf{x}}_{t|t-1}$ denotes the linear, minimum-error-variance (LMEV) estimate of the state vector $\mathbf{x}_t$ given the past recorded data $z_{t-1}, z_{t-2}, z_{t-3}, \cdots$; $\hat{\mathbf{x}}_{t|t}$ denotes the LMEV estimate of $\mathbf{x}_t$ given the present and past data $z_t, z_{t-1}, z_{t-2}, \cdots$. If we define the error $\tilde{\mathbf{x}}_{t|t-1} \equiv \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}$, then $\mathbf{P}_{t|t-1}$ denotes the ($m \times m$ matrix) covariance of $\tilde{\mathbf{x}}_{t|t-1}$, often referred to as the "prior error covariance". Similarly, $\mathbf{P}_{t|t}$ represents the "posterior error covariance". $\mathbf{k}_t$ is usually called the "Kalman gain" vector. $\boldsymbol{\gamma}_t$, $\mathbf{\Phi}_t$, and $\mathbf{h}_t$ are the, as yet, undefined matrices of equations (1).

Kalman filtering is just the sequential estimation of $\mathbf{x}_t$ using equations (6) for $t = 1, 2, 3, \cdots$. The required inputs are $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$, σ_r^2, σ_n^2, and z_t.

Before we can use equations (6), we must define $\mathbf{x}_t$, $\boldsymbol{\gamma}_t$, $\boldsymbol{\Phi}_t$, and $\mathbf{h}_t$ in equations (1), and these definitions should be consistent with our seismogram model of equation (2). Because r_t is the quantity we wish to estimate and because equations (6) yield estimates of $\mathbf{x}_t$, we are motivated to make equation (1a) as simple as possible. For $m = 4$, our proposed form is

$$\begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ x_{t-3} \\ x_{t-4} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} r_t \tag{7}$$

The state vector $\mathbf{x}_t$ is simply

$$\mathbf{x}_t = \left[r_t \; r_{t-1} \; r_{t-2} \; r_{t-3} \right]^T$$

The form of equation (7) for arbitrary m should be evident. $\boldsymbol{\Phi}_t$ and $\boldsymbol{\gamma}_t$ are constant, sparse matrices which contain none of the seismic system parameters. For equations (1) to be consistent with our seismogram model, $\mathbf{h}_t$ must do the work of **FQD** of equation (2). To determine $\mathbf{h}_t$, first rewrite equation (2) as

$$z_t = \sum_s F_{ts} \sum_u Q_{su} \frac{r_u}{u} + n_t$$

where we have assumed a constant velocity medium to replace the spherical divergence matrix **D**. Using the definition of **F** from the discussion of equation (2) and the definition of **Q** in Hale (1981), we have

$$z_t = \sum_s f_{t-s} \sum_u q_{s-u}^u \frac{r_u}{u} + n_t$$

The superscript u on the Q-filter q reflects the fact that q is a time-variable filter. After manipulation of summation indices, we obtain

$$z_t = \sum_{s=0}^{l} \sum_{u=0}^{k} f_s q_u^{t-s-u} \frac{r_{t-s-u}}{t-s-u} + n_t$$

where we choose l and k so that f_s and q_u^{t-u} are effectively zero for $s > l$ and $u > k$, respectively. A change of index from s to $s-u$ results in

$$z_t = \sum_{s=0}^{l+k} \left[\sum_{u=0}^{k} \frac{1}{t-s} q_u^{t-s} f_{s-u} \right] r_{t-s} + n_t$$

If we define the quantity in parentheses to be the coefficients h_s^t of a time-variable filter, then $\mathbf{h}_t$ of equation (1b) must be

$$\mathbf{h}_t \equiv \left[h_0^t \; h_1^t \; h_2^t \; \cdots \; h_{m-1}^t \right]^T$$

where $m = l+k+1$ is the order of the state equations.

Having defined the variables in the state equations (1), we can now apply the Kalman filtering equations (6) to estimate r_t given z_t. But recall that equations (6) yield estimates of the state vector $\mathbf{x}_t$ based only on $z_t, z_{t-1}, z_{t-2}, \cdots$. Each estimate $\hat{\mathbf{x}}_{t|t}$ contains the m estimates $\hat{r}_{t|t}, \hat{r}_{t-1|t}, \hat{r}_{t-2|t}, \cdots, \hat{r}_{t-m+1|t}$. Thus, our proposed interpretation of equations (1) provides m distinct estimates of any r_t. Intuition, perhaps based on equation (5), and the work of Mendel and Kormylo (1978) suggest that the best of the m estimates is $\hat{r}_{t|t+m-1}$, an estimate based on future as well as past observations z_t. $\hat{r}_{t|t+\tau}$ is called a "fixed-lag smoothed" estimate of r_t; and, for the examples that follow, we chose $\tau = m-1$.

Figure 2 contains the results of Kalman filtering both the pure signal **s** and the noise-contaminated trace **z** of Figure 1. The correct answer **r** is replotted for comparison. **Cs** and **Cz** are the divergence-corrected (for display only) **s** and **z**, respectively. The traces labeled **r1** and **r2** are estimates of **r** based on **s**. **r1** was obtained with σ_n^2 set to zero (the correct value) in equations (6); **r2** was obtained with $\sigma_n^2 = 1.0\times10^{-9}$. Recall that, aside from providing the source waveform, Q, etc., we must also provide σ_r^2, σ_n^2, $\mathbf{P}_{0|0}$, and $\mathbf{x}_{0|0}$ to the Kalman filtering algorithm. In all of our examples, we used the initial conditions $\mathbf{P}_{0|0} = \sigma_r^2\mathbf{I}$ and $\mathbf{x}_{0|0} = 0$. **r1** is the result of providing the correct σ_r^2 and σ_n^2 in filtering **s**; because **s** is noise-free, **r1** is a good estimate of **r**. The source of the error at early times is unknown.

r2, on the other hand, is the result of "telling" the algorithm that the input **s** is noisy when it is not. Notice that the filter progressively refuses to back out the effects of **FQD** at the later times so as not to amplify (non-existent) noise.

The danger of underestimating the noise variance σ_n^2 is illustrated by the estimate **r3**, obtained by Kalman filtering the noisy trace **z** with σ_n^2 set to zero in equations (6). In practice, we would never be so confident in our data (or our model) as to assume $\sigma_n^2 = 0$.

Like **r3**, **r4** was obtained by filtering **z**, but with $\sigma_n^2 = 1.0\times10^{-9}$ in equations (6). Note that, at early times, **r4** closely resembles the correct **r**; at later times, **r4** is less noisy than the input **z**, as seen by comparing **r4** with **Cz**. This increase in "signal-to-noise ratio" is expected because the fixed-lag smoothed estimate obtained by Kalman filtering is roughly equivalent to the matched-filter estimate (for large α) of equation (5). Remember that both Kalman filtering and equation (5) represent solutions to the same problem -- linear, minimum-error-variance estimation of **r**. The basic difference between the two methods is that

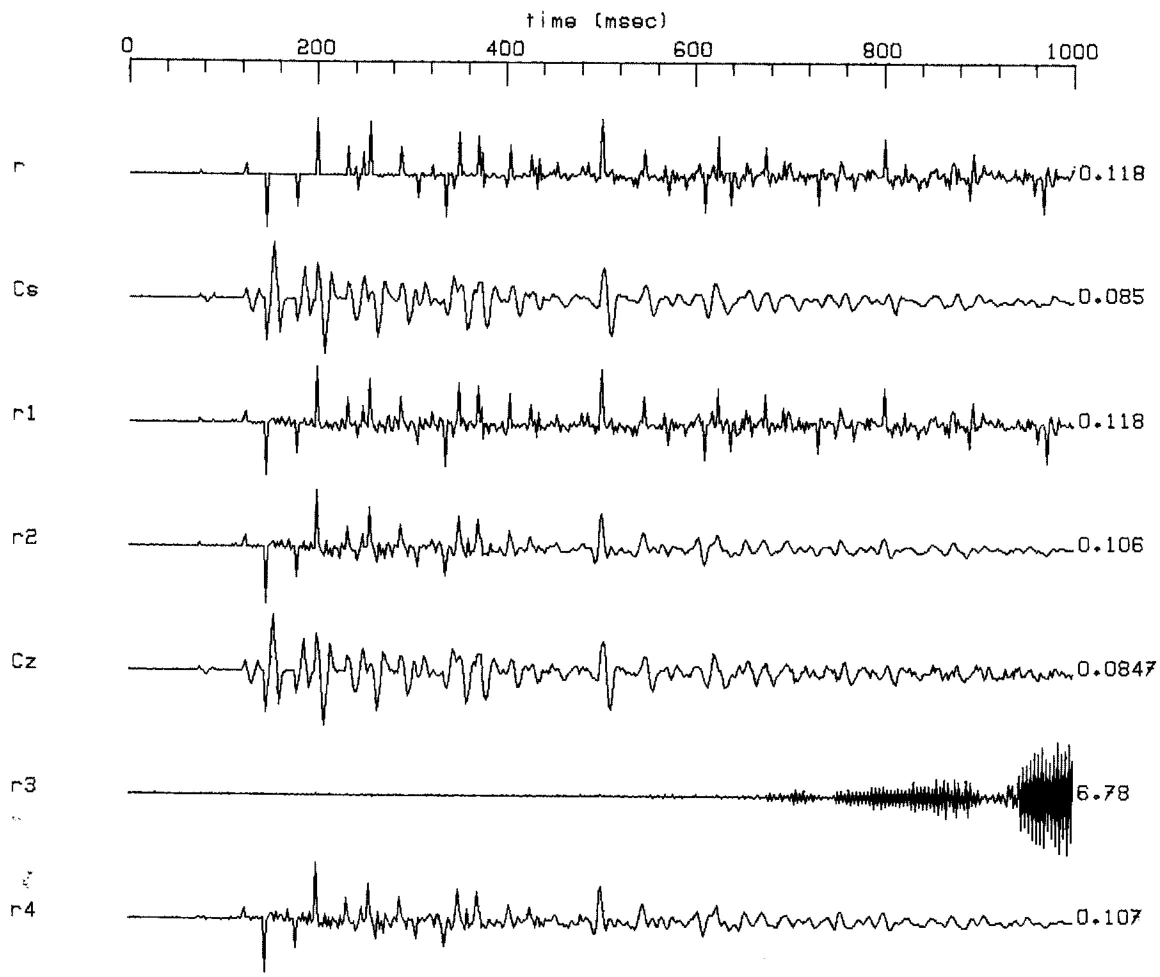

FIG. 2. KMA -- Kalman filtering for a MA model. **r** is the "correct answer" replotted from Figure 1 for comparison. **Cs** and **Cz** are the noise-free and noise-contaminated traces, respectively, of Figure 1, divergence-corrected for display only. **r1** is the fixed-lag smoothed estimate of **r** obtained by Kalman filtering **s** with the correct parameters. In particular, $\sigma_n^2 = 0$ in equations (6). **r2** was also obtained from the noise-free **s**, except that a Kalman filter with noise variance $\sigma_n^2 = 1.0\times10^{-9}$ was used. The estimates **r3** and **r4** were computed using Kalman filters identical to those used to compute **r1** and **r2**, respectively, the differences being that the noisy **z** was input rather than the noise-free **s**.

equation (5) implies that all past and future z_t are used to estimate the present r_t, whereas our Kalman filter estimate is derived from all past but only $m-1$ future z_t.

The results in Figure 2 indicate that our interpretation of equations (1) enables reasonable estimates of r_t when we provide the Kalman filter with the correct parameters. One disadvantage of this interpretation, however, is that the order m of the state equations may become quite large if, for example, the source waveform is long or Q is small. $m = 35$ in the above examples. Larger m imply greater computation time; for example, the cost of computing $\mathbf{P}_{t|t}$ in equation (6e) is proportional to m^2. Clearly, we want to keep m as small as possible.

Mendel and Kormylo (1978) obtained small m by interpreting the state equations (1) in a manner quite different from ours. In their interpretation, the source waveform f_t was specified analytically to obtain "a continuous-time state space model" which was subsequently discretized to obtain equations (1). Analytical functions for a real source waveform may be difficult to obtain.

A more practical way to decrease m may be to permit poles as well as zeros in our discrete specification of the source waveform. As written, equation (2) implies that z_t is the output of a time-varying, moving-average process. In the next section, we develop a new interpretation of equations (1) by including auto-regressive (AR) as well as moving-average (MA) components in our model.

KARMA -- Kalman filtering for an ARMA model

First rewrite equation (2) as

$$\mathbf{z} = \mathbf{B}\mathbf{A}^{-1}\mathbf{Q}\mathbf{D}\mathbf{r} + \mathbf{n}$$

We simply replace $\mathbf{F}$ with $\mathbf{B}\mathbf{A}^{-1}$; $\mathbf{B}$ is a lower-triangular, Toeplitz matrix with MA coefficients on its diagonals. The first column of $\mathbf{B}$ is $\mathbf{b} = [b_0\ b_1\ b_2\ \cdots\ b_{lb}\ 0\ 0\ \cdots\ 0]^T$. $\mathbf{A}$ is also lower-triangular and Toeplitz but with AR coefficients on its diagonals. The first column of $\mathbf{A}$ is $\mathbf{a} = [1\ a_1^0\ a_2^0\ \cdots\ a_{la}^0\ 0\ 0\ \cdots\ 0]^T$. (The meaning of the superscript will become apparent.) With these definitions, we rewrite our model as

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

where $\mathbf{y}$ is defined by

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{y} \equiv \mathbf{D}\mathbf{r}$$

Using the definition of the inverse Q-filter in Hale (1981),

$$\sum_s p_{t-s}^t \sum_u a_{s-u}^0 y_u = \frac{r_t}{t}$$

Manipulate summation indices to obtain

$$\sum_{s=0}^{k} \sum_{u=0}^{la} p_s^t a_u^0 y_{t-s-u} = \frac{r_t}{t} \tag{8}$$

where k is chosen so that $p_s^t \approx 0$ for $s > k$. Next, define a normalized inverse Q-filter with coefficients $g_s^t \equiv p_s^t / p_0^t$, change summation index u to $u-s$, and interchange summations to obtain

$$\sum_{u=0}^{k+la} \left[\sum_{s=0}^{k} g_s^t a_{u-s}^0 \right] y_{t-u} = \frac{r_t}{t p_0^t}$$

In parentheses are the coefficients of a time-variable filter which we define to be a_u^t. Because of our normalization, $a_0^t = 1$; the above equation can, therefore, be rewritten as

$$y_t = -\sum_{u=1}^{k+la} a_u^t y_{t-u} + \frac{r_t}{t p_0^t}$$

Recall that the purpose of this lengthy derivation is to provide a new interpretation of the state equations (1) as follows (for $m = 4$):

$$\mathbf{x}_t \equiv \left[y_t \; y_{t-1} \; y_{t-2} \; y_{t-3} \right]^T \tag{9a}$$

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{bmatrix} = \begin{bmatrix} -a_1^t & -a_2^t & -a_3^t & -a_4^t \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ y_{t-4} \end{bmatrix} + \begin{bmatrix} 1/(t p_0^t) \\ 0 \\ 0 \\ 0 \end{bmatrix} r_t \tag{9b}$$

$$\mathbf{h}_t \equiv \left[b_0 \; b_1 \; b_2 \; b_3 \right]^T \tag{9c}$$

Definition of the quantities in equations (1) is now complete. In practice, m must be chosen to be the greater of $lb+1$ and $k+la$. If k is much greater that either la or lb, then we may wish to decompose the matrix **Q** into AR and MA components, just as we decomposed the matrix **F**.

An unfortunate consequence of introducing AR as well as MA components in our seismogram model is that equations (6) no longer yield direct estimates of r_t. They instead provide estimates of the state vector $\mathbf{x}_t$ or, equivalently, y_t [equation (9a)]. Ott and Meder (1972) proposed the following "prediction error" estimate of r_t

$$\boldsymbol{\gamma}_t \hat{r}_t = \hat{\mathbf{x}}_{t|t} - \boldsymbol{\Phi}_t \hat{\mathbf{x}}_{t-1|t-1} = \mathbf{k}_t (z_t - \mathbf{h}_t^T \hat{\mathbf{x}}_{t|t-1}) \quad , \tag{10}$$

an estimate suggested by equations (1a), (6a), and (6b). Because, in our interpretation, γ_t has only one non-zero element, equation (10) provides a unique estimate of r_t. Mendel and Kormylo (1978) point out that this estimate is "ad hoc", that it is not a LMEV estimate, and then develop methods for computing LMEV, smoothed (fixed-lag, fixed-point, and fixed-interval) estimates of r_t, essentially extensions of the basic Kalman filter equations (6). [See also Mendel and Kormylo (1977).]

We propose yet another estimate of r_t. Recalling that the Kalman filter equations (6) provide fixed-lag smoothed estimates $\hat{y}_{t|t+\tau}$ of y_t, equation (8) motivates

$$\hat{r}_t = t \sum_{s=0}^{k} \sum_{u=0}^{la} p_s^t a_u^0 \hat{y}_{t-s-u|t-s-u+\tau} \quad (11)$$

However, the $\hat{r}_t$ so obtained are not fixed-lag smoothed estimates; i.e.,

$$\begin{aligned} \hat{r}_t &= t p_0^t \hat{y}_{t|t+\tau} + t(p_0^t a_1^0 + p_1^t)\hat{y}_{t-1|t-1+\tau} + \cdots \\ &\neq t p_0^t \hat{y}_{t|t+\tau} + t(p_0^t a_1^0 + p_1^t)\hat{y}_{t-1|t+\tau} + \cdots = \hat{r}_{t|t+\tau} \end{aligned}$$

The estimate $\hat{r}_t$ in equation (11) (unlike Ott and Meder's estimate) is derived from future as well as past z_t, but because every term in equation (11) is not based on $z_{t+\tau}$, $\hat{r}_t \neq \hat{r}_{t|t+\tau}$. Nevertheless, we expect equation (11) to provide reasonable estimates, particularly since p^t and a^0 are minimum-phase so that the dominant terms in the sum are those for small s and u. Estimates based on equation (11) were computed from the synthetic **s** and **z** and are plotted in Figure 3 in the same format as those plotted in Figure 2. For our synthetic traces, the source waveform is best represented by an all-zero (MA) form; therefore, the AR component of our model consisted only of the inverse Q-filter. Again, we chose $m = 35$.

r1 and **r4** were obtained, as in Figure 2, by providing the Kalman filter with reasonable noise variances. **r2** and **r3** again illustrate the result of supplying incorrect noise variances to the Kalman filter equations (6). Comparison of corresponding estimates in Figures 2 and 3 confirms our expectation that little practical advantage is to be gained in using a truly fixed-lag smoothed estimate rather than the "quick and dirty" estimate of equation (11).

Conclusions

In the words of Crump (1974), "the real potential of the [Kalman filtering] method lies in its capability for handling continually time-varying values for all parameters which are required. The accurate estimation of these time-varying parameters is the major problem in taking full advantage of the potential of the method. Further investigation of this problem is needed."

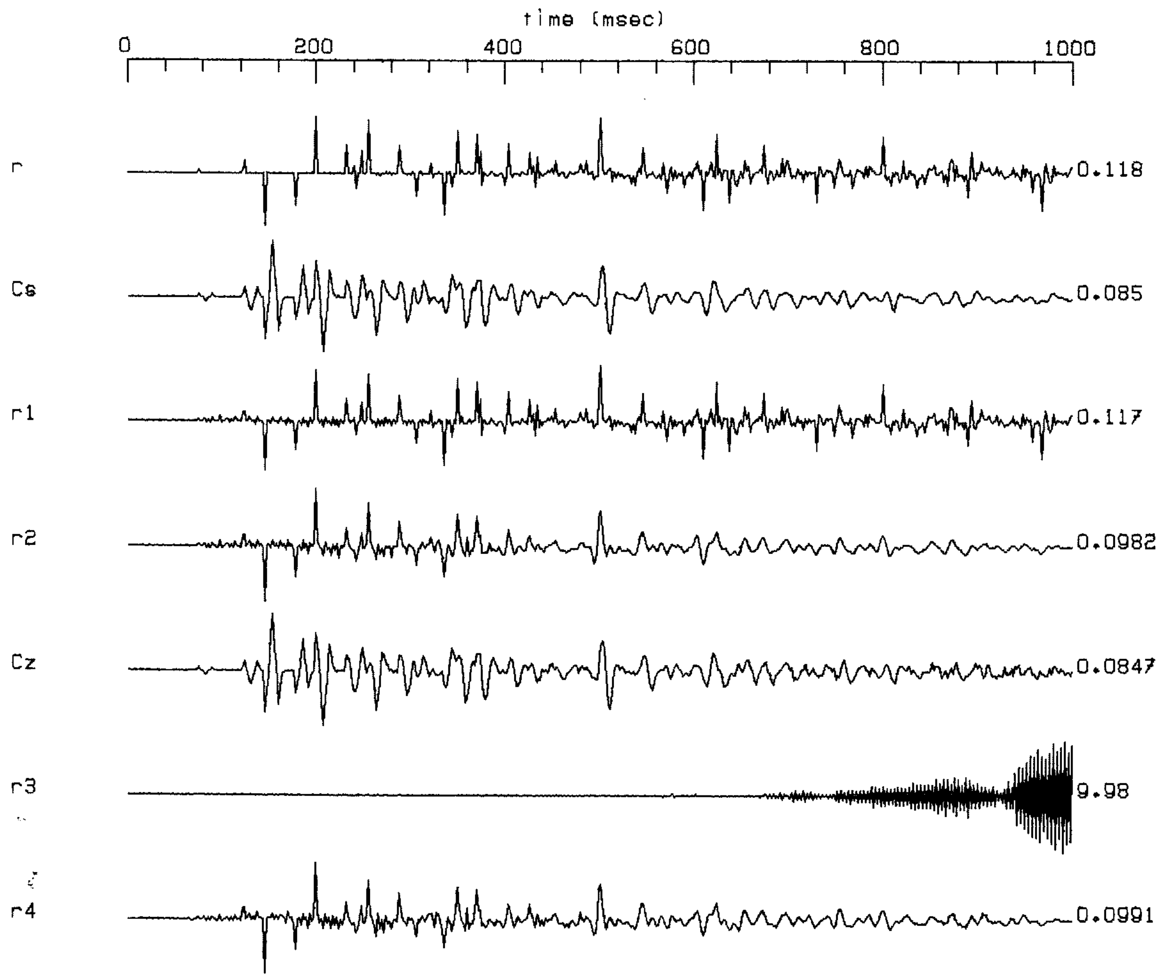

FIG. 3. KARMA -- Kalman filtering for an ARMA model. **r** is the "correct answer" replotted from Figure 1 for comparison. **Cs** and **Cz** are the noise-free and noise-contaminated traces, respectively, of Figure 1, divergence-corrected for display only. **r1** is the estimate [via equation (11)] of **r** obtained by Kalman filtering **s** using the correct parameters. In particular, $\sigma_n^2 = 0$ in equations (6). **r2** was also obtained from the noise-free **s**, except that a Kalman filter with noise variance $\sigma_n^2 = 1.0 \times 10^{-9}$ was used. The estimates **r3** and **r4** were computed using Kalman filters identical to those used to compute **r1** and **r2**, respectively, the differences being that the noisy **z** was input rather than the noise-free **s**.

But, assuming that this major problem can be solved, we must still decide how to best use the estimated parameters in the context of Kalman filtering; we must interpret the state equations (1). Of the two interpretations presented in this paper, both of which include the time-varying effects of attenuation and spherical-divergence, the first (KMA) is perhaps most appealing due to its smoothed estimate of r_t. With KMA, a fixed-lag smoothed estimate of r_t is available for the same computational cost and algorithmic complexity as for an unsmoothed estimate. This statement cannot be made about the method of Mendel and Kormylo (1978).

An alternative to KMA, motivated by a desire to decrease the order m (and, hence, the computational cost) of the Kalman filter, is to interpret equations (1) with AR as well as MA components (KARMA). Results for synthetic data indicate that KARMA provides estimates comparable to those obtained with KMA, even though the estimates are not strictly fixed-lag smoothed. Highly resolved estimates of r_t are obtained when the noise level is low, matched filtering is performed when the noise level is high.

Finally, we should not too quickly disregard the direct, numerical solution of equation (5). The symmetric, positive-definite, and, in particular, the banded structure of this system of equations may make its direct solution less costly than any practical Kalman filtering solution.

REFERENCES

Bayless, J.W., and Brigham, E.D., 1970, Application of the Kalman filter to continuous signal restoration: Geophysics, v.35, p.2-23.

Booton, R.C., 1952, An optimization theory for time-varying linear systems with nonstationary statistical inputs: Proc IRE, v.40, p.977-981.

Claerbout, J.C., 1976, Fundamentals of geophysical data processing with applications to petroleum prospecting: New York, McGraw-Hill Book Company.

Crump, N.D., 1974, A Kalman filter approach to the deconvolution of seismic signals: Geophysics, v.39, p.1-13.

Hale, D., 1981, An inverse Q-filter: Stanford Exploration Project Report 26, p.231-243.

Kalman, R.E., 1960, A new approach to linear filtering and prediction problems: Trans. ASME. J. Basic Eng., ser. D, v.82, p.34-45.

Mendel, J.M., and Kormylo, J., 1977, New fast optimal white-noise estimators for deconvolution: Spec. iss., IEEE Trans. on Geosci. Electron, no. GE-15, p.32-41.

——— 1978, Single-channel white-noise estimators for deconvolution: Geophysics, v.43, p.102-124.

Ott N., and Meder, H.G., 1972, The Kalman filter as a prediction error filter: Geophys. Prosp., v.20, p.549-560.

Sage, A.P., and Melsa J.L., 1979, Estimation theory with applications to communications and control: New York, Robert E. Krieger Publishing Company.