Extrapolating Nonstationary Traces by Autoregressive Filters

Jeff Thorson

Introduction

A truncated one-dimensional sample of data can usually be extrapolated with good results by making use of its prediction error filter. That is, the coefficients of the prediction error, or autoregressive (AR) filter are first found from the known data by standard means, such as the Burg algorithm, and then iteratively applied to extrapolate the data. This extrapolation therefore has the property that it has "zero prediction error", or is totally predictable by means of an autoregressive model. Predictability in some sense is the main characteristic that any extrapolation should have. The purpose of this paper is to generalize in a straightforward manner this simple procedure to handle, first, nonstationarity and second, an arbitrary geometry for missing data points.

First, suppose the given data is nonstationary. A prediction error filter estimated from all the known data will be an average of AR filters estimated over smaller, more local segments. It may not be a suitable filter to extrapolate with. Alternately, a suite of AR filters may be calculated over local segments, adapting to the changing statistics of the trace. Suppose we now have a set of AR filters calculated from windowed pieces of a time trace, assigning a filter to each time point. We shall require that the coefficients of each filter be continuous in time. In other words the trace is assumed to have a smoothly varying nonstationarity. In this way an AR filter represents a local spectrum.

Second, consider how gaps among known data points of the trace might be filled. An implicit rather than an explicit scheme of extrapolation is more natural in this case, in order that the interpolated values depend on data on both sides of the gap, rather than having a one-sided dependency. It is easy to see how using our adaptive set of AR filters to "implicitly" extrapolate into a gap can lead to solving a linear system. This is in contrast to using prediction error filters iteratively to solve for points off-end, sequentially one point at a

time. Examples shall be given of adaptive (time-varying) AR filters used to fill in arbitrary gaps in a given trace.

Algorithm

An extrapolation procedure can be divided into three steps: estimation of statistical parameters on the known data, continuation of these parameters into the unknown region, and determining values of points in the unknown region that are consistent with the estimated statistics. To outline these steps in more detail:

A) Estimation. Determine adaptive AR filter coefficients from known points. One way to do this is by an adaptive Burg algorithm using weighted prediction errors (Hale, SEP-28). Another is by least-squares estimation. That is, minimize the mean-square forward and backward prediction errors, weighted by a symmetric window centered at the prediction point (Ulrych and Clayton, 1976).

B) Interpolation. Extend the AR coefficients out to the unknown data region. An assumption is made here: the trace must be locally stationary so that the AR coefficients themselves are continuous and may be easily extrapolated or interpolated. A weight function of sufficient smoothness chosen in step (A) will guarantee that the known AR coefficients are continuous.

C) Once the AR filter coefficients are known "everywhere", a system of equations results from the minimization of the total prediction error by variation of the unknown data points. This is in contrast to step (A) where the same functional is minimized by variation of the AR coefficients. The system is naturally narrow banded because each autoregressive filter provides one relation between a local group of data points.

Let us now go into more detail on each of these steps, and illustrate them with examples. Two examples are used. The first is a synthetic nonstationary trace (figure 1), composed of two "chirps" of a different frequency range. The second is a real time trace (figure 3) which is effectively stationary.

Autoregressive Estimation

In view of what is to be done with the prediction error filters it is not crucial that the filters be minimum-phase. For this reason a simple least-squares algorithm for determining the adaptive AR coefficients is used. Consider a time trace u_t, $t = 0,1,\cdots,n-1$. Define the forward and backward prediction errors to be, respectively,

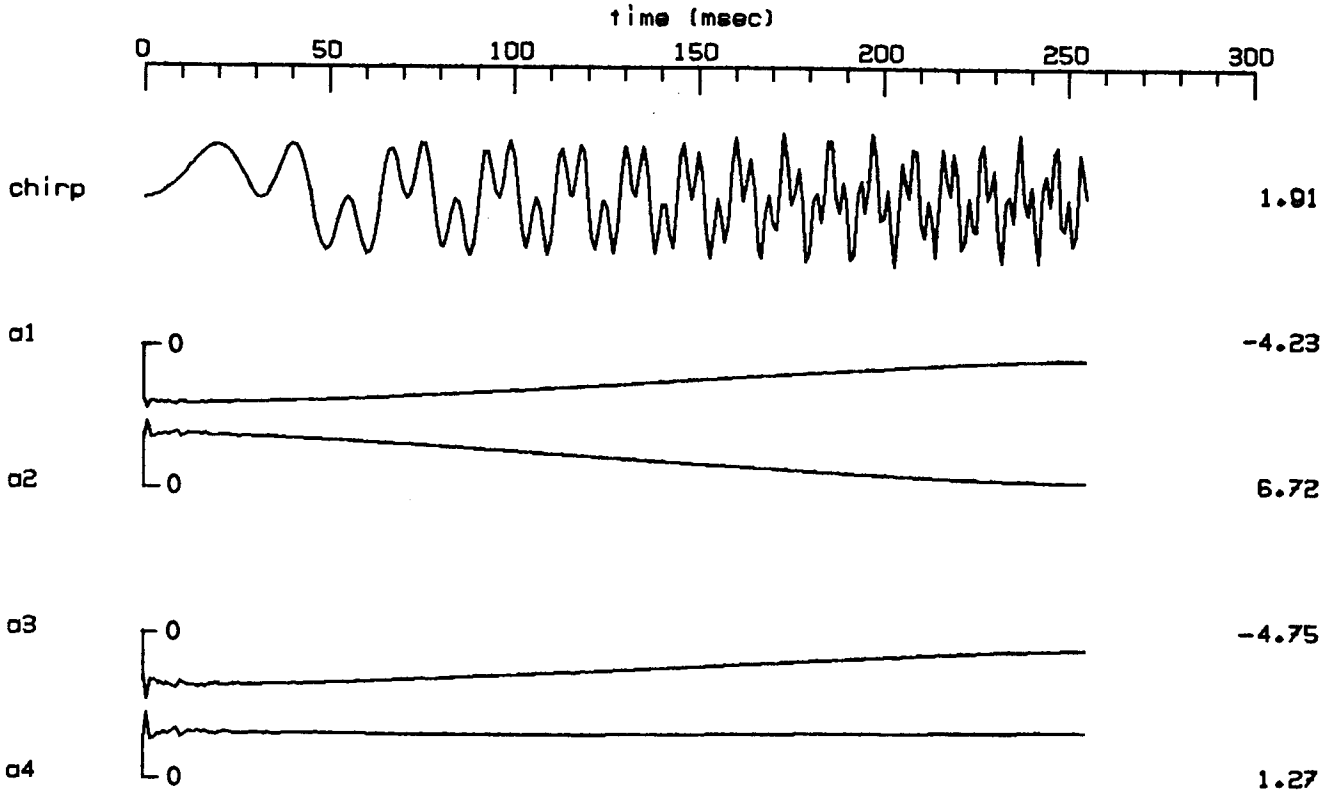

FIG. 1. Double chirp model. The equation for the trace labeled "chirp" is u_k = sin($\omega(k,h_1)$) + sin($\omega(k,h_2)$), where $\omega(k,h) = (2\pi \cdot 60/v)(z^2 + (kh/256)^2)^{1/2}$. Each of the terms is what would be obtained if one took a common midpoint gather with a single event, time transformed the gather and selected the imaginary part of the 60 Hz component of each time trace. The events result from a reflector at depth z = 2500 m, v = 1500 m/sec, and maximum offsets h_1 and h_2 of 1000 m and 1750 m, respectively. The four coefficients $a_1, \cdots, a_4$ of the 4th order adaptive AR filter are labeled a1 to a4 in the figure. Each trace is scaled independently; clip values are to the right. The sample rate is 1 msec.

$$f_t = \sum_{k=0}^{m} a_k u_{t-k} \qquad t = m, \cdots, n-1$$

$$b_t = \sum_{k=0}^{m} a_k u_{t+k} \qquad t = 0, \cdots, n-m-1 \qquad (1)$$

where a_k are the AR coefficients $k = 0, \cdots, m$ and $a_0 = 1$.

The estimation problem is to minimize a weighted energy of the forward and backward error:

$$\mathbf{E}_s = \sum_{t=m}^{n-1} w_{s-t} f_t^2 + \sum_{t=0}^{n-m-1} w_{s-t} b_t^2. \qquad (2)$$

The weight w_s used in the examples is a gaussian envelope of a constant halfwidth, that is, ten points equalling one standard deviation. Thus 30 points away from its center $s = 0$ (with value unity) the envelope has a value of 0.011. The minimization of $\mathbf{E}_s$ yields an AR filter for each $s = 0,1,\cdots,n-1$. The resulting linear system is now derived:

$$0 = \frac{\partial \mathbf{E}_s}{\partial a_i} = 2\sum_{t=m}^{n-1} w_{s-t} f_t u_{t-i} + 2\sum_{t=0}^{n-m-1} w_{s-t} b_t u_{t+i} \qquad i = 1,\cdots,m$$

$$= 2\sum_{t=m}^{n-1} w_{s-t} \sum_{k=0}^{m} a_{ks} u_{t-k} u_{t-i} + 2\sum_{t=0}^{n-m-1} w_{s-t} \sum_{k=0}^{m} a_{ks} u_{t+k} u_{t+i}$$

or,

$$\sum_{k=0}^{m} a_{ks} \left[\sum_{t=m}^{n-1} w_{s-t} u_{t-k} u_{t-i} + \sum_{t=0}^{n-m-1} w_{s-t} u_{t+k} u_{t+i} \right] = 0 \qquad (3)$$

where $a_0 = 1$.

Figure 1 illustrates the estimation of a_{ks} for the signal labeled "chirp". This synthetic signal is a superposition of two chirp sequences whose instantaneous frequencies rise from zero at time $t = 0$ to different maxima at $t = 256$ msec. For each of the 256 time points of the trace ($s = 0,\cdots,255$), linear system (3) was solved for a fourth-order filter ($k = 1,\cdots,4$). The filter coefficients $a_{1s},\cdots,a_{4s}$ are plotted as a function of s in figure 1. As expected, they vary quite smoothly with s, except near $s = 0$ where the instantaneous frequency of the signal is near zero.

Figure 2 shows the same synthetic trace of figure 1 but with a gap of missing data between 128 and 158 msec, and a zero extension from 256 to 316 msec. System (3) is now used to calculate the AR coefficients $a_{1s},\cdots,a_{4s}$ only over the set of points where data is known. Edge effects from the presence of the gap are evident on the plots of $a_1,\cdots,a_4$ in figure 2; there is a flattening of the curves on either side of the gap. This will degrade the performance of the interpolation across the gap which is to be performed next (second half of figure 2).

A second example of a gapped trace is given in figure 3. The leading part of the trace has been zeroed out to 56 msec, as well as a piece from 188 msec to 218 msec, a gap of 30 points. Nevertheless the trace is seen to be stationary when the AR coefficients are computed over the remainder of the trace (figures 3-a1 through a8). Here, an 8th order filter has been computed per time point. The coefficients are nearly constant.

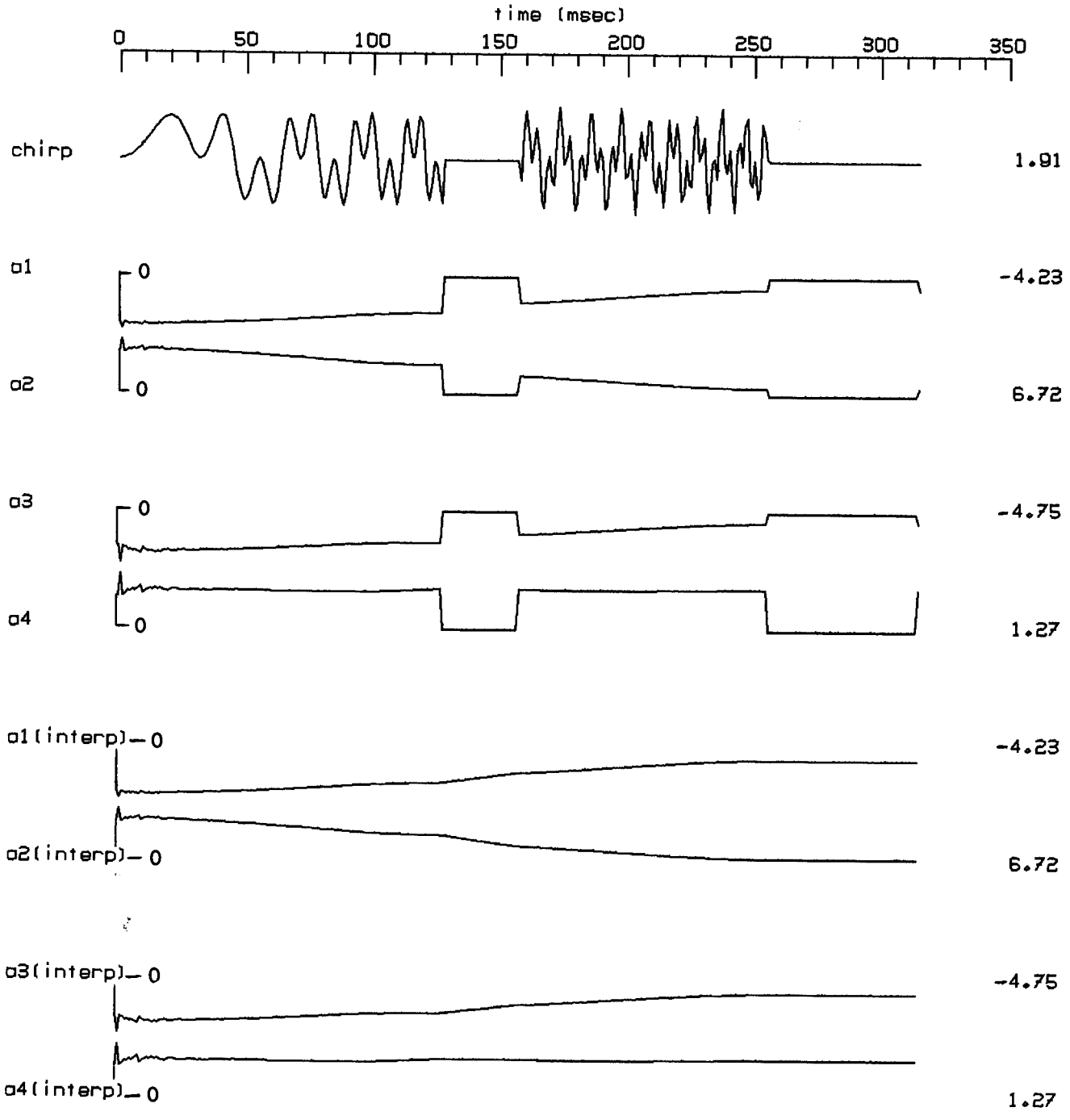

FIG. 2. Truncated chirp signal. "chirp" is now the trace of fig. 1 truncated on points 128-158 and zero-padded from point 256 to 316. The AR coefficients $a_1, \cdots, a_4$ estimated from the remaining nonzero segments of the trace are shown next. The bottom four traces labeled "interp" are linearly interpolated versions of $a_1, \cdots, a_4$.

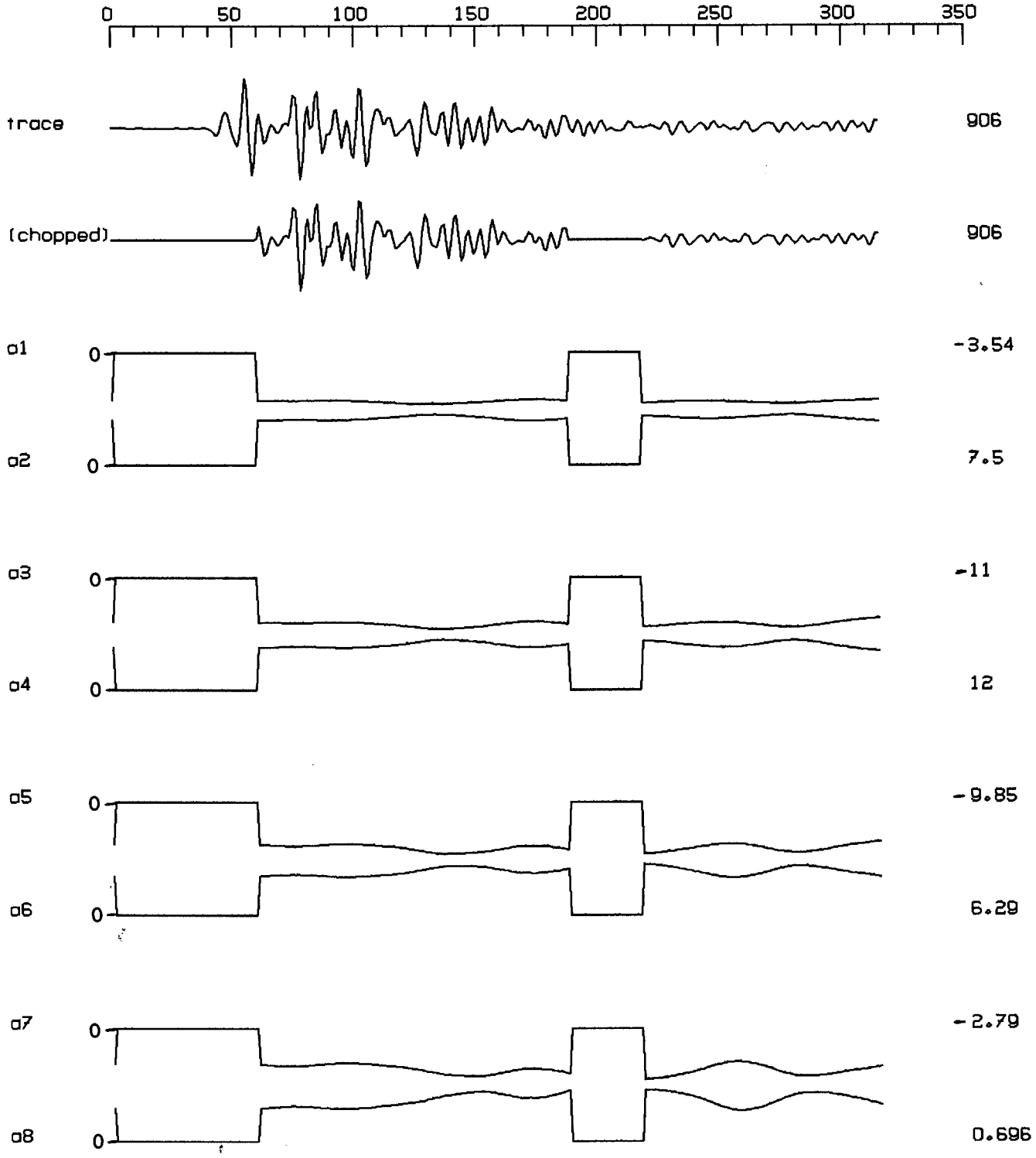

FIG. 3. Real trace, displayed in the same format as figures 1 and 2. The trace is given at the top of the page, with a truncated version of it displayed underneath. Below this are displayed the eight coefficients $a_1, \cdots, a_8$ of an 8th-order AR filter calculated from the nonzero segments of the chopped trace. The coefficients are nearly constant; the trace is stationary. For purposes of this paper the sample rate of the real trace is 1 msec.

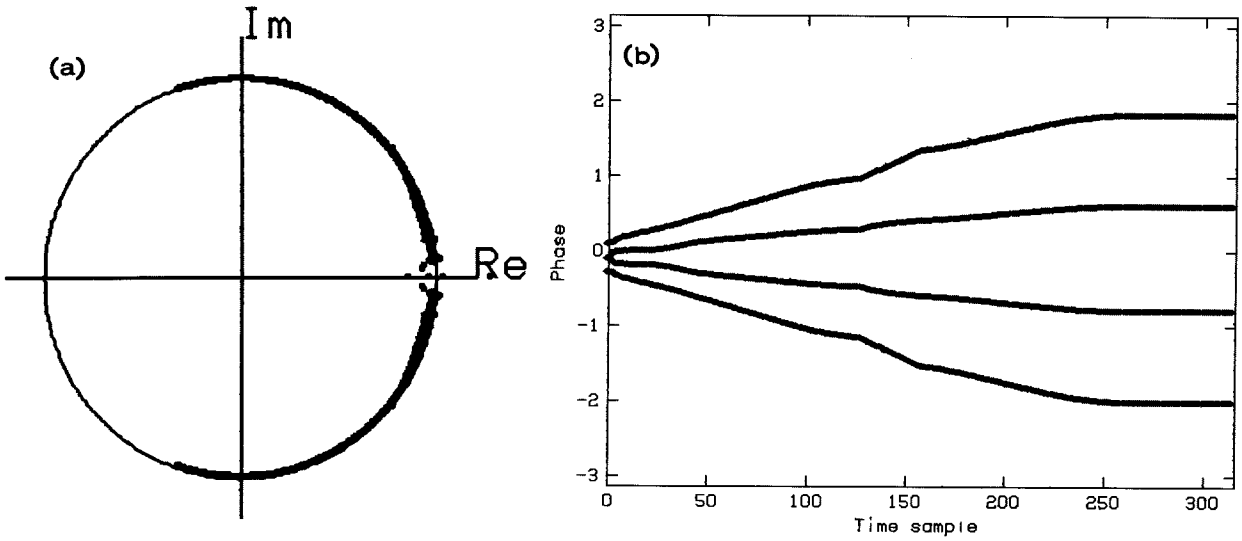

FIG. 4. Plots of roots of the AR filters in figure 2. The interpolated AR coefficients were taken as the coefficients of a fourth order polynomial in z (z-transform of the filter) for each time point, and the roots solved for. Thus for time points 0-311, four complex roots were found. Since the coefficients are real, the roots either come in complex conjugate pairs or lie on the real axis. These roots are plotted in two ways. (a) shows the roots in the complex z plane. (b) plots the phase of the roots as a function of time. The phase plot may be thought of as showing the instantaneous frequencies present in the original trace, which in this case is the double chirp model. Therefore (b) shows two instantaneous frequencies, increasing with time until t = 256 msec when they become constant. The effect of linearly interpolating the AR coefficients in figure 2 is apparent between points 128 and 158 of the phase plot (b), where there is a kink in the smooth phase curve.

Interpolation of Filter Coefficients

The next step is to interpolate or extend the function of each AR coefficient over the unknown points of the trace. All of the examples shown here use a simple linear interpolation to fill the gap within a trace, and a constant-coefficient extrapolation for unknown points off the end of the trace. Figure 2 shows the interpolation applied to the double-chirp model. The error from the end-effects mentioned in the last section is apparent in the interpolation of a_1 and a_2: the restored coefficients are not as smooth as they were in figure 1 for the original trace.

It is enlightening to look at plots of the estimated AR filter's roots in the z-domain. Figure 4(a) depicts in the complex plane the superposition of the roots of all 316 AR filters for the double-chirp model of figure 2. It is not surprising that some roots fall outside the unit circle since the least-squares system (3) does not constrain the solutions to be minimum-phase. However the roots track along the unit circle quite smoothly. Considering each root

to be a function of s, the phase of each root in figure 4(b) is smoothly increasing with s, as it should be for the double-chirp model. Each point of the original trace is modeled precisely by the sum of two sinusoids, so a 4-term (real) AR filter is the optimal length in this case. One conjugate pair of roots, as a function of s, tracks the sinusoid of higher instantaneous frequency, while the other pair tracks the sinusoid of lower frequency.

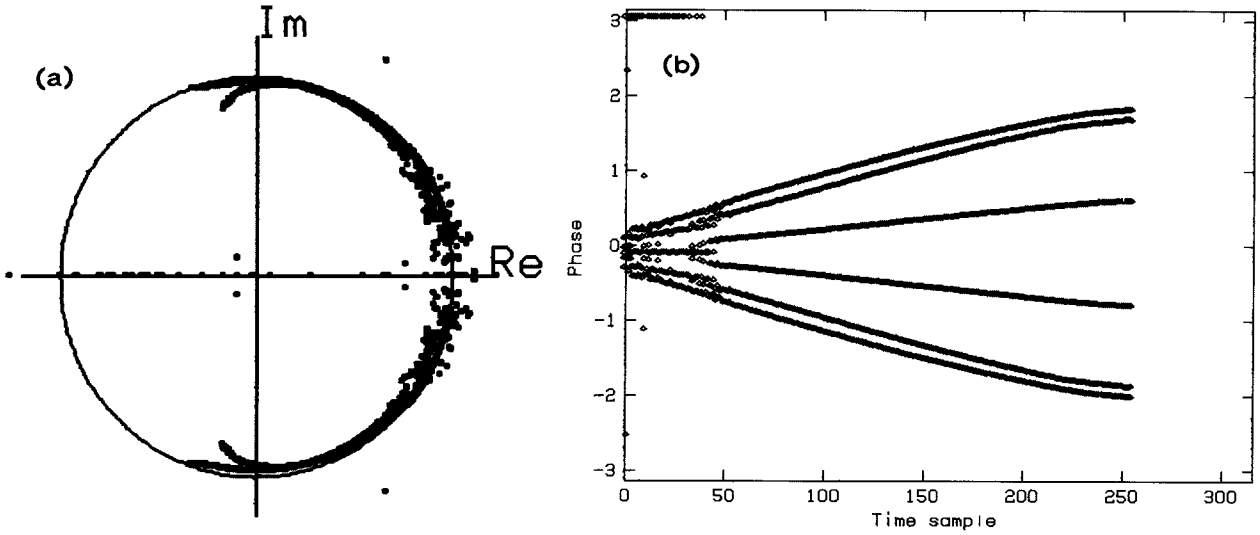

FIG. 5. Root plot as in figure 4 for the double chirp model, but now derived from 6th order AR filters. Comparing this with figure 4, two pairs of roots faithfully track the instantaneous frequencies in figure 4, while the third remains in between, and at late time points peels away from the unit circle. There is much more computational noise at low frequency, where the linear system (3) is ill-conditioned (AR filters do not like predicting DC).

If a 4th order AR filter is the optimal size for the double chirp model, what happens when a 6th order filter is calculated? Figure 5 shows the results of this calculation. First, there is much more scatter at the lower frequencies, many roots falling onto the real axis. Once the lower time samples are left behind, two pairs of roots latch onto the true instantaneous frequency, while the third pair, which essentially has nothing to do, follow along at an intermediate frequency, falling away from the unit circle at the highest frequencies. But as long as two of the root pairs follow the true frequencies of the chirp, the presence of an extraneous root pair does no harm in the extrapolation.

The root plots of the stationary trace of figure 3 are shown in figure 6. The eight roots of each AR filter are quite localized in the z-plane, and the phase plots show a near-

constant behavior of the phase with time.

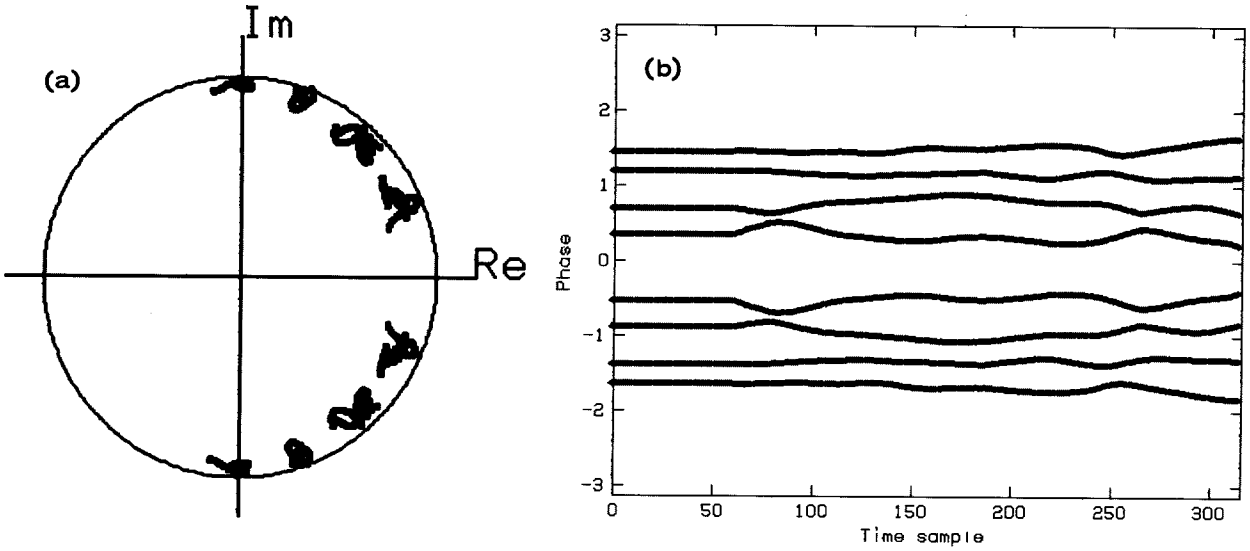

FIG. 6. Root plots as in figure 4, but now on the real trace of figure 3. There, an 8th order AR filter was estimated, and there are eight roots per time point in conjugate pairs. (a) and (b) show that the filters are relatively independent of time.

Extrapolation

The final step is to utilize the interpolated AR filter coefficients to restore missing sections of the trace. The problem is set up by attempting to minimize virtually the same functional that was used in the AR estimation stage: the prediction error energy. This time the energy functional has the form

$$\mathbf{E} = \sum_{t=m}^{n-1} f_t^2 + \sum_{t=0}^{n-m-1} b_t^2 \qquad (4)$$

where f_t and b_t are defined as in (1):

$$f_t = \sum_{k=0}^{m} a_{kt} u_{t-k} \qquad b_t = \sum_{k=0}^{m} a_{kt} u_{t+k}.$$

Each error term, as a function of t, utilizes a seperate AR filter, a_{kt} $(k = 0, \cdots, m)$. One attains a banded linear system in the following manner: minimize **E** by variation of u_s where the index s ranges over the points of the trace where u is to be solved for: the missing data.

$$0 = \frac{\partial \mathbf{E}}{\partial u_s} = 2\sum_{t=m}^{n-1}\left[\sum_{k=0}^{m} a_{kt}u_{t-k} \cdot \frac{\partial}{\partial u_s}\sum_{k=0}^{m} a_{kt}u_{t-k}\right]$$

$$+ 2\sum_{t=0}^{n-m-1}\left[\sum_{k=0}^{m} a_{kt}u_{t+k} \cdot \frac{\partial}{\partial u_s}\sum_{k=0}^{m} a_{kt}u_{t+k}\right]$$

Now

$$\frac{\partial}{\partial u_s}\sum_{k=0}^{m} a_{kt}u_{t-k} = \begin{cases} 0 & \text{if } s < t-m,\ s > t \\ a_{t-s,t} & \text{if } s = (t-m, \cdots, t) \end{cases}$$

and

$$\frac{\partial}{\partial u_s}\sum_{k=0}^{m} a_{kt}u_{t+k} = \begin{cases} 0 & \text{if } s < t,\ s > t+m \\ a_{s-t,t} & \text{if } s = (t, \cdots, t+m) \end{cases}.$$

The sth equation is therefore

$$\sum_{t=s}^{s+m}\sum_{k=0}^{m} a_{kt}a_{t-s,t}u_{t-k} + \sum_{t=s-m}^{s}\sum_{k=0}^{m} a_{kt}a_{s-t,t}u_{t+k} = 0 \qquad (5)$$

where those terms whose indices are out of bounds are defined to be zero. For example, $a_{s-t,t} = 0$ when $s-t < 0$. Now for the first term of (5) perform the substitution $t \to t+k$, and for the second term, $t \to t-k$. After a further adjustment of the indices k, equation (5) becomes

$$\sum_{t=s-m}^{s+m} Q_{st}u_t = 0 \qquad (6)$$

where

$$Q_{st} = \begin{cases} \sum_{k=0}^{m-(s-t)}\left[a_{k,s+k}a_{k+(s-t),s+k} + a_{k,t-k}a_{k+(s-t),t-k}\right] & \text{for } t < s \\ \sum_{k=0}^{m-(t-s)}\left[a_{k,t+k}a_{k+(t-s),t+k} + a_{k,s-k}a_{k+(t-s),s-k}\right] & \text{for } t \geq s \end{cases} \qquad (7)$$

The symmetry $Q_{st} = Q_{ts}$ is easily verified in this form of the system of equations. Again, those elements a_{kt} whose indices fall outside the bounds $k = (0, \cdots, m)$, $t = (0, \cdots, n-1)$ are assumed to be zero.

The known values of u now contribute to the right hand side of the system to be solved. Equation (6) may be rewritten as

$$\sum_{t=s-m,\ t\varepsilon A}^{s+m} Q_{st}u_t = \sum_{t=s-m,\ t\varepsilon B}^{s+m} -Q_{st}u_t \qquad (s\ \varepsilon\ A) \qquad (8)$$

where A is the set of indices for which u is to be found, and B is its complement, the set of indices for which u is given. The system (8) has a maximum bandwidth of $2m+1$, but may easily decouple into smaller systems if the gaps to be filled in on the trace are situated far enough apart from each other.

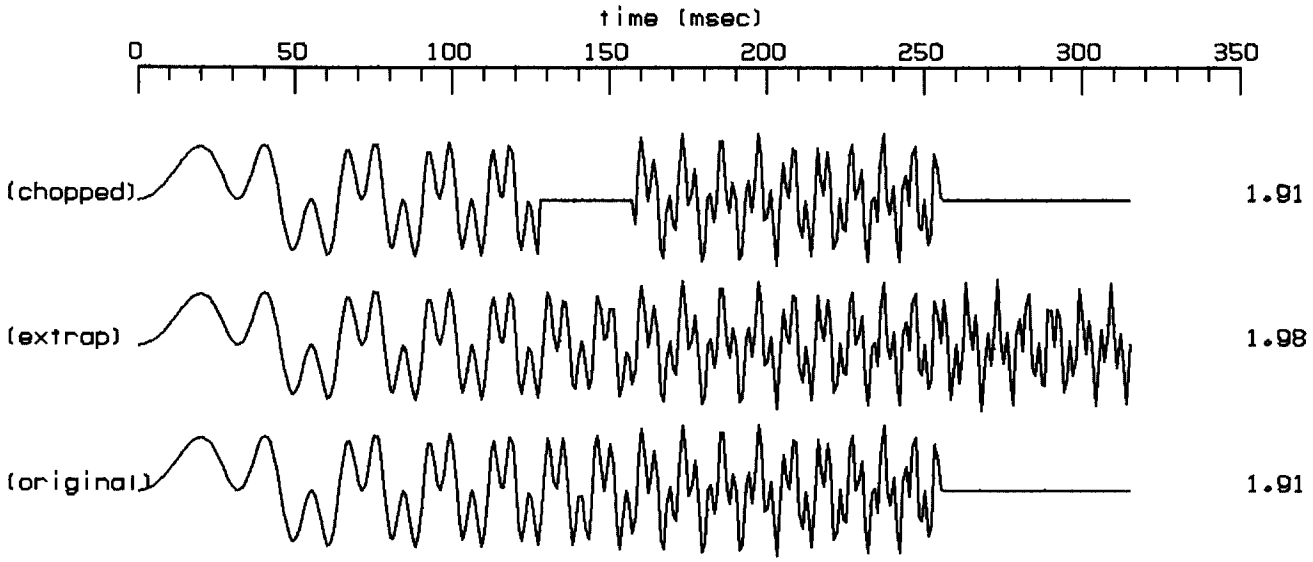

FIG. 7. Extrapolating the example of figure 2. "Chopped" is the trace of figure 2, "extrap" is the result of solving for the missing data points by equation (6), and "original" is the original trace of figure 1. The AR filters supplied to equation (6) are the interpolated set in figure 2.

The system (8) is now applied to the examples described in the previous sections. First, consider the double chirp model of figure 2. Figure 7 shows the results of extrapolation into the gap and off the trace end. The extrapolated trace very closely follows the original trace in the restored gap. The extension off the end is very reasonable too. The examples of figures 8 and 10 are like that of figure 7 (using the double chirp trace) but the extrapolation is performed over wider gaps of 60 msec and 90 msec respectively. The results for the gap infill are incorrect yet interesting, though practically such large gaps may never be chosen. The pole plots of figures 9 (corresponding to figure 8) and 11 (corresponding to figure 10) give a clue to the problem. The plot of figure 11(b) shows the instantaneous frequency of the lower-frequency chirp to be far from linear in the interpolated gap. Correspondingly, the extrapolated trace of figure 10 rapidly moves out of phase

with the original desired answer in the gap. The phase itself is the crucial parameter that must be properly interpolated over a wide gap for the case of the double chirp model.

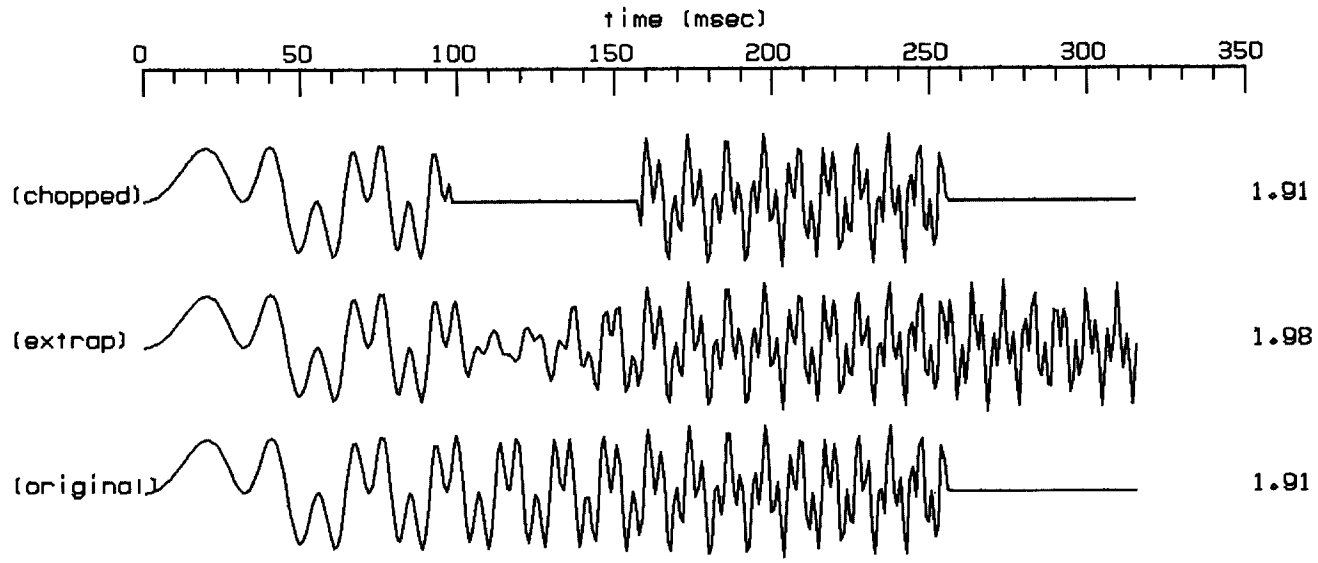

FIG. 8. Extrapolation of the double chirp model, second example. The difference between this and figure 7 is only in the width of the gap: it now being 60 points wide, from 98 msec to 158 msec. The decrease in amplitude in the extrapolated gap is probably due to the interference between frequencies. The phase of the trace in the gap has shifted by a large amount from the desired phase of the original trace.

The last two examples given are extrapolations on the stationary trace of figure 3. Figure 12 is the result of extrapolating the gap geometry of figure 3. It is interesting to note that the extrapolation is able to restore a large arrival present on the original trace, though obviously fails when it gets ahead of the first innovation on the trace. The extrapolation can be considered to be innovation-free: the restored sections have minimal prediction-error energy. In the sense of the prediction error defined in (1), the extrapolation tends to be the most predictable one in *both* directions on the time trace, so the extrapolation will not mirror innovations actually present in the missing gap, such as the first break of the trace on figure 3. Figure 12 also shows how one can control the stability of an extrapolation. The first point of the trace ($t = 0$) is assumed to be known and zero. Consequently to match this, the extrapolation has to taper off to zero at $t = 0$. Figure 13 is an extrapolation with more reasonably sized gaps. The correspondence in the signal phase between the

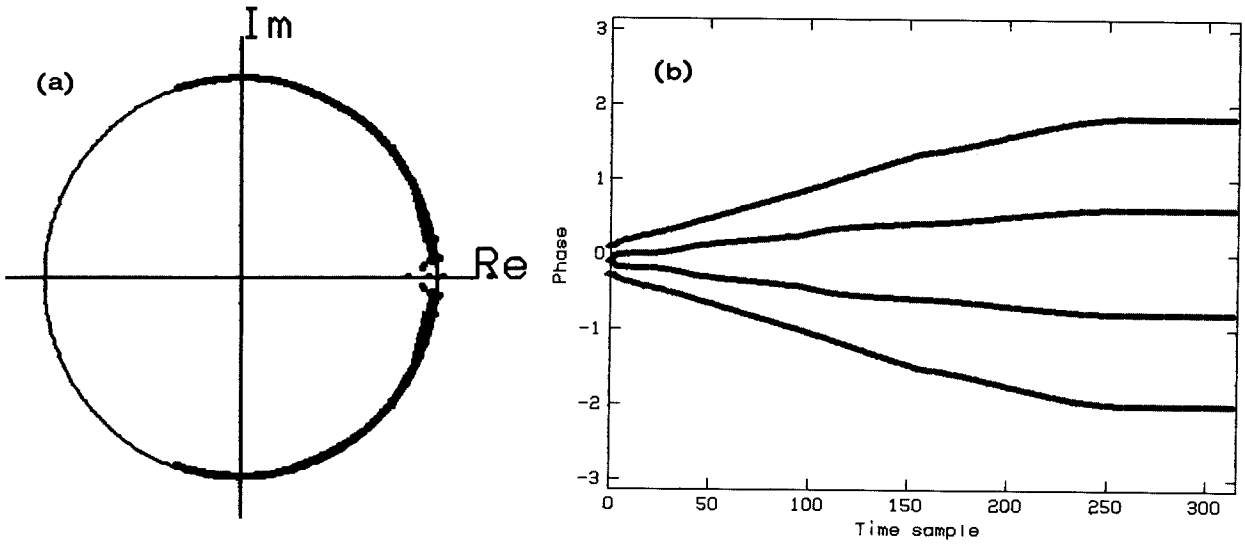

FIG. 9. A root plot of the example of figure 8 (see figure 4 for an explanation of the plots).

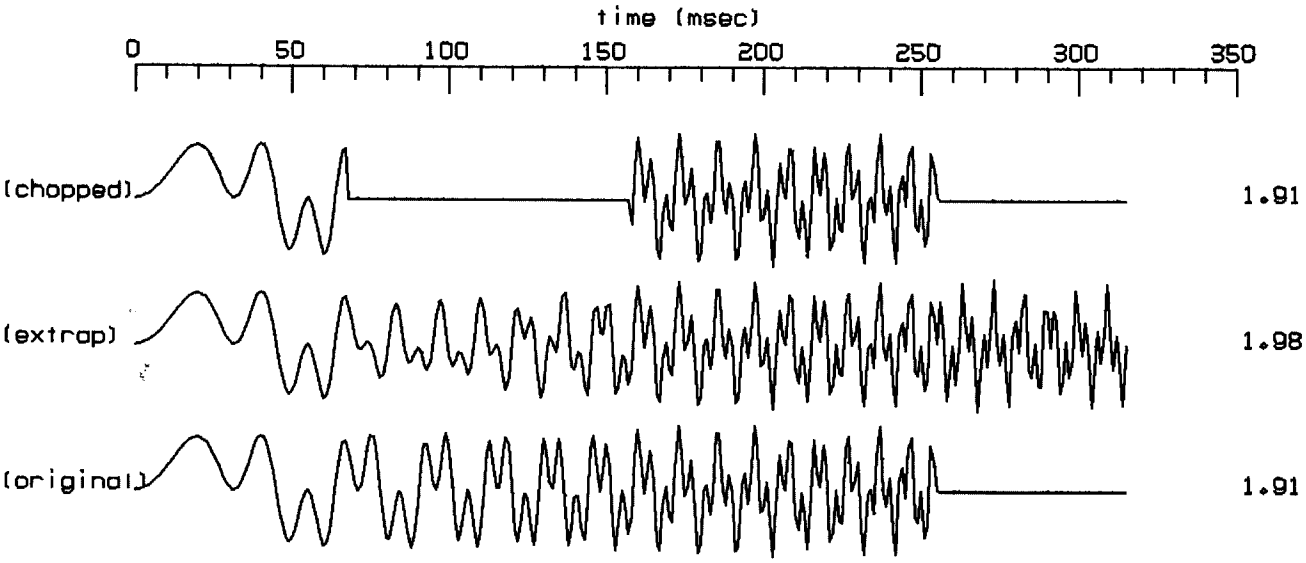

FIG. 10. Extrapolation of the double chirp model, third example. The gap is yet wider in this case, 90 points, as compared to 60 points in figure 8 and 30 points in figure 7. Although the extrapolation seems good, the difference between the phase of the extrapolated trace and the original trace is large.

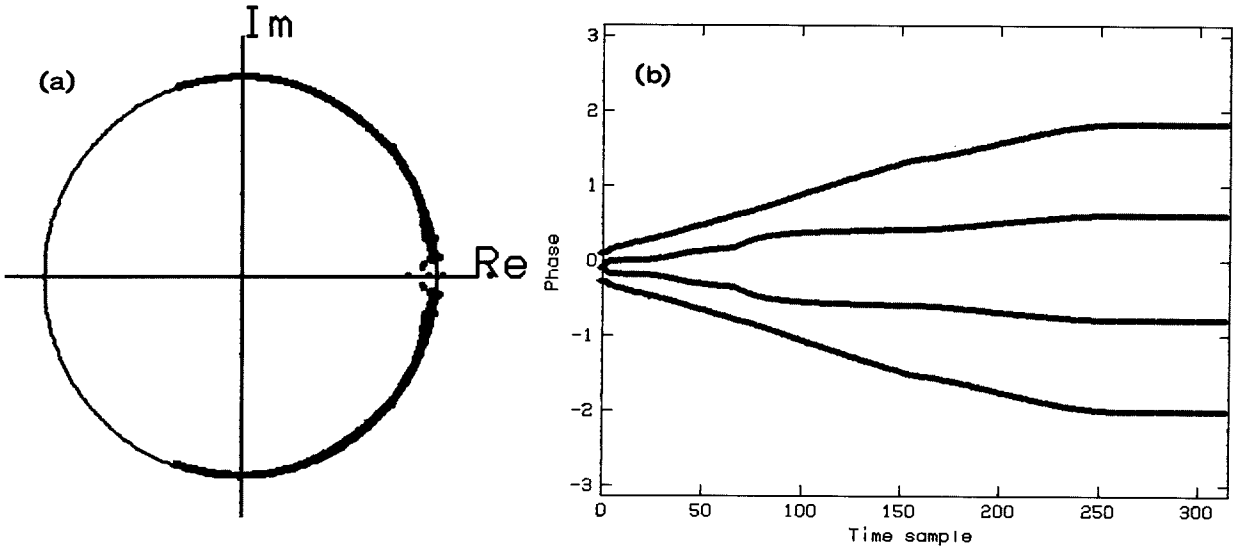

FIG. 11. A root plot of the example of figure 10, the wide gap. An explanation of the observed phase error between the extrapolated and original trace may be seen in (b). The high-frequency root pair display a linear interpolation in the gap, while the low-frequency pair have a characteristic bulge. This narrowing of the range between the two instantaneous frequencies will account for the phase error of figure 10 and the destructive interference of figure 8. Linear interpolation of filter coefficients does not translate to linear interpolation of the phase of its z-transform.

extrapolated trace and the original trace is quite good.

Further Work

The adaptive AR model described in this paper is best suited to work on traces that are innovation-free. No rapid change in the filter coefficients is expected or desired as it varies with time index s. For this reason it may be used to more effect in laterally extrapolating a common-midpoint gather, or other points on a spatial axis rather than a time axis. With least squares there is no conceptual difficulty in extending this to two dimensions; the only difficulties would be those of programming and time considerations.

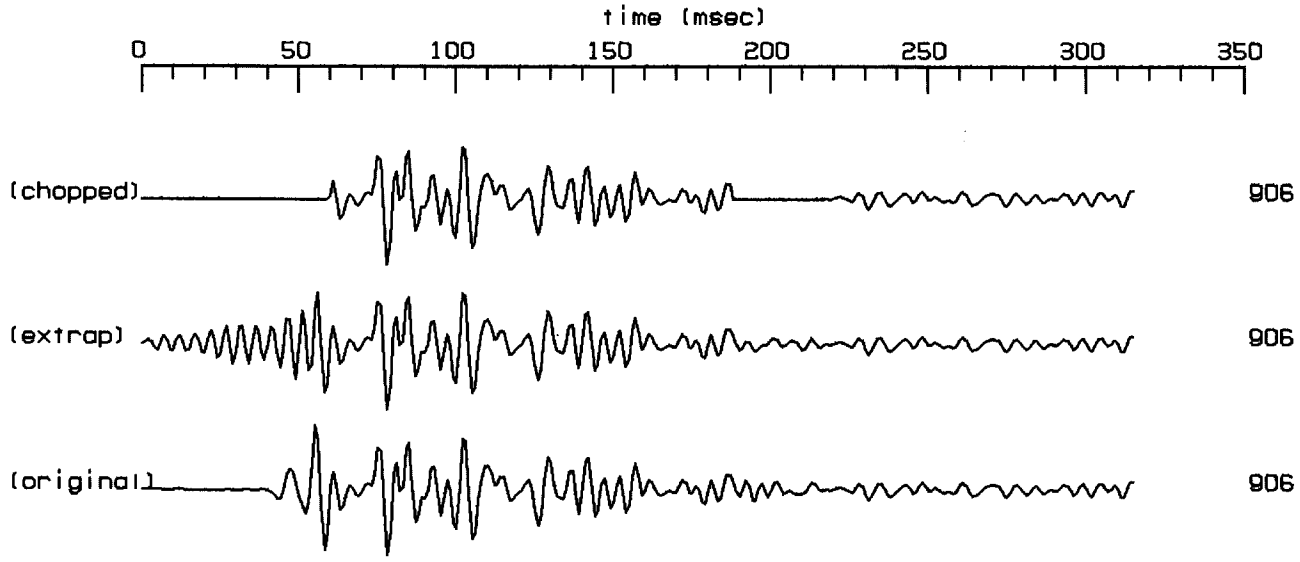

FIG. 12. Extrapolation of the gapped real trace of figure 3, first example. The gapped region is from points 0 to 60 and from points 188 to 218.

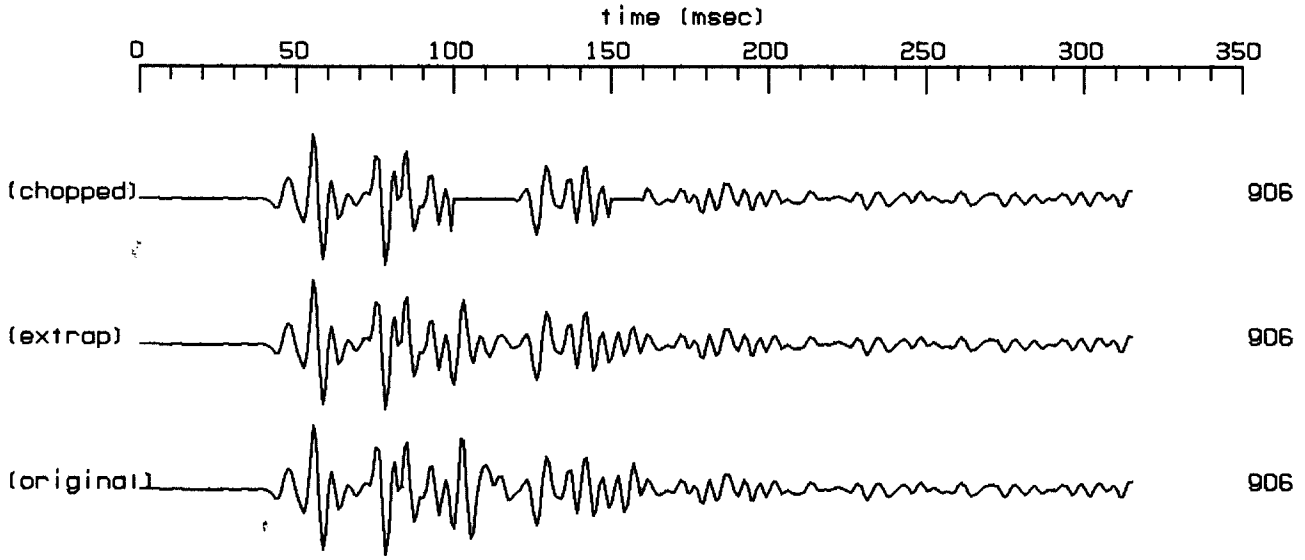

FIG. 13. Second example of extrapolation of the real trace of figure 3. In this case, narrower gaps were chosen: from points 100 to 120, and from points 150 to 160. The phase of the signal in the restored area compares favorably with the true phase.

ACKNOWLEDGMENTS

Bert Jacobs developed the Fortran program for calculating roots of a polynomial of arbitrary order. Many of the illustrations here were created using Dave Hale's interactive trace analysis program "wavelet".

REFERENCES

Hale, Ira David, Q and Adaptive Prediction Error Filters: SEP-28 (this volume).

Ulrych, Tad J. and Rob W. Clayton, Time Series Modelling and Maximum Entropy: Physics of the Earth and Planetary Interiors, 12 (1976) p. 188-200.