SIMILARITIES AND DIFFERENCES BETWEEN ONE-SIDED AND TWO-SIDED LINEAR PREDICTION

Jin-Jen Hsue and Andrew E. Yagle

Dept. of EECS, The University of Michigan, Ann Arbor, MI 48109-2122 aey@eecs.umich.edu

ABSTRACT

We provide a comparison between one-sided linear prediction (OSP) and two-sided linear prediction (TSP) with respect to prediction error, relationships to AR modeling and to two-sided AR modeling, and the application to time series interpolation, linear-phase filter design, and spectral estimation. New contributions of this paper include: (1) proof that TSP produces smaller, non-white residuals than OSP, extending previous results; (2) specification of the frequency-domain error criterion minimized by TSP, and comparison with the analogous OSP criterion; (3) demonstration that TSP and two-sided AR modeling are different problems, unlike OSP; (4) interpretation of performance of TSP interference-rejection filters.

1. INTRODUCTION

The one-sided linear prediction (OSP) problem arises in many signal processing applications, such as spectral estimation, speech coding, time series extrapolation and interpolation, etc. The basic idea of OSP is to estimate the current sample value as a weighted linear combination of past sample values. If the optimal prediction coefficients are determined by minimizing the least-squares prediction error, one must solve a Toeplitz or close-to-Toeplitz system.

Recently, two-sided linear prediction (TSP) has been applied to various signal processing applications, including spectral estimation, speech coding, linear phase filter design, time series interpolation [1], and system identification [2]. OSP is sometimes used on a frame-by-frame basis, in which case all the data samples in an entire frame are available for analysis. A better estimate of a sample would be expected if we "predict" the present sample based on both the past and future samples; this motivates the use of TSP in these problems.

Although OSP and TSP seem to be very similar, there are important fundamental differences between them. In this paper, we study the properties of TSP and their relation to those of OSP, investigating the similarities and differences between OSP and TSP in terms of prediction error, relationship to AR modeling and two-sided AR modeling, time series interpolation, and spectrum estimation.

New contributions of this paper include the following: (1) we prove that TSP produces smaller residuals than OSP for any wide-sense stationary random process (generalizing a result for finite-order AR processes in [1] and [3]); (2) we specify the frequency-domain error criterion minimized by TSP, compare it to the analogous error criterion for OSP, and discuss its implications for TSP spectral estimates; (3) we show that TSP and the two-sided AR modelling problem are different, whereas in OSP they are identical problems; and (4) we provide some numerical examples of TSP applications. We also discuss issues such as why TSP interference-rejection filters have faster rise in their transition bands than OSP filters.

This work was supported by the National Science Foundation under Presidential Young Investigator grant #MIP-8858082.

2. TSP MEAN SQUARE ERROR

2.1. TSP Prediction Error

In this section, we discuss two major properties of the TSP prediction error: (1) The mean-square TSP prediction error is always less than the mean-square OSP prediction error; and (2) the TSP error process is not white. The first result is new for arbitrary wide-sense-stationary (wss) random processes; it has been derived previously for finite-order AR processes in [1]. The second result was shown in [1].

Theorem : (NEW) Let $x(n)$ be a zero-mean w.s.s. process. Then the p^{th}-order TSP mean-square error $\epsilon_2^2 = E[(x(n) + \sum_{i=1}^{p} b_i(x(n-i) + x(n+i)))^2]$ is less than the p^{th}-order OSP mean-square error $\epsilon_1^2 = E[(x(n) + \sum_{i=1}^{p} a_i x(n-i))^2]$ for any specific value of p.

Proof : The mean-square error MSE for OSP and TSP can be written as [2],[3]

$$\epsilon_1^2 = r(0) - r^T T^{-1} r; \quad \epsilon_2^2 = r(0) - 2r^T (T+H)^{-1} r \quad (1)$$

where $r = [r(1) \cdots r(p)]^T$ and $r(n) = E[x(i)x(i+n)]$. T is a symmetric Toeplitz matrix with first row $[r(0) \cdots r(p-1)]$. H is a Hankel matrix whose first column and last row are $[r(2) \cdots r(p+1)]^T$ and $[r(p+1) \cdots r(2p)]$, respectively. Subtracting ϵ_2^2 from ϵ_1^2, we have $\epsilon_1^2 - \epsilon_2^2 = r^T(2(T+H)^{-1} - T^{-1})r$. We want to show that $2(T+H)^{-1} - T^{-1}$ is positive semidefinite; this proves $\epsilon_2^2 \le \epsilon_1^2$.

Consider the covariance matrix of the random vector $[x(0) \cdots x(p-1)\ x(p+1) \cdots x(2p)]^T$, which is

$$\begin{bmatrix} T & JH \\ HJ & JTJ \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & J \end{bmatrix} \begin{bmatrix} I & 0 \\ JHJT^{-1} & I \end{bmatrix}$$

$$\begin{bmatrix} T & 0 \\ 0 & T - JHT^{-1}HJ \end{bmatrix} \times \begin{bmatrix} I & T^{-1}JHJ \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & J \end{bmatrix} \quad (2)$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

Since the above covariance matrix must be positive semidefinite, we conclude that $T - JHT^{-1}HJ =$

$$T^{1/2}[I - (T^{-1/2}JHJT^{-1/2})(T^{-1/2}JHJT^{-1/2})]T^{1/2} \geq 0 \quad (3)$$

which implies $(T^{-1/2}JHJT^{-1/2})^2 \leq I$ or equivalently $0 \leq T + H \leq 2T$. Then one can show $2(T+H)^{-1} \geq T^{-1}$.QED

2.2. Special Cases

We now consider some **SPECIAL CASES**. First, for an AR(p) process, we can show that the p^{th} order one-sided predictor and two-sided predictor are related as follows [1]:

Let $x(n)$ be a 0-mean p^{th}-order AR process. Then [1]

$$\{b_p \cdots b_1\, 1\, b_1 \cdots b_p\} = \frac{1}{1 + \sum_{i=1}^{p} a_i^2}\{1\, a_1 \cdots a_p\} * \{a_p \cdots a_1\, 1\}, \quad (4)$$

$$\varepsilon_2^2 = \frac{\varepsilon_1^2}{1 + \sum_{i=1}^{p} a_i^2}, \quad (5)$$

where a_i, ε_1^2 and b_i, ε_2^2 are prediction coefficients and prediction errors for OSP and TSP.

Let $x(n)$ be a zero-mean w.s.s. process with PSD $P_x(\omega)$. Then [1]

$$\varepsilon_2^2 = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} P_x^{-1}(\omega)d\omega\right]^{-1}, \; B(e^{j\omega}) = \frac{\varepsilon_2^2}{P_x(\omega)}. \quad (6)$$

where ε_2^2 is infinite order TSP MSE and $B(e^{j\omega})$ is the optimal TSP error filter.

Note that the PSD of the infinite-order TSP error $e(n)$ can be written as $P_e(\omega) = |B(e^{j\omega})|^2 P_x(\omega) = \frac{\varepsilon_2^4}{P_x(\omega)}$. The infinite-order TSP error time series is clearly not white.

3. TSP AND TWO-SIDED AR MODELING

It is well-known that the OSP problem is closely related to AR modeling and the AR spectral estimation problem. In fact, the Wiener-Hopf equations for OSP problem and the Yule-Walker equations for AR modeling are identical, provided that they have the same model order.

We will show that, unlike the one-sided case, TSP coefficients differ from two-sided AR (TAR) coefficients.

3.1. TAR vs. OSP: Similarities & Differences

We first present some results for TAR analogous to those for OSP. Consider the LSI system

$$x(n) = -\sum_{k=1}^{p} \beta_k(x(n-k) + x(n+k)) + u(n). \quad (7)$$

If $u(n)$ is a zero-mean wss white process with variance σ^2, then $x(n)$ is a p^{th} order two-sided AR (TAR) process, and the coefficients β_i are termed the TAR parameters.

If $\beta(z) = 1 + \sum_{i=1}^{p} \beta_i(z^i + z^{-i})$ has no zeros on the unit circle, then $x(n)$ is the output of a stable noncausal LSI system $1/\beta(z)$ driven by $u(n)$, since the region of convergence of $1/\beta(z)$ includes the unit circle. The region of convergence is chosen as $\rho < |z| < 1/\rho$ for some $0 < \rho < 1$, so that the inverse Z-transform of $1/\beta(z)$ is noncausal but stable. The PSD of $x(n)$ is related to the PSD of $u(n)$ by

$$P_x(z) = \frac{\sigma^2}{\beta(z)\beta(z^{-1})} = \sum_{k=-\infty}^{\infty} r(k)z^{-k}. \quad (8)$$

Multiplying (8) by $\beta(z)$ and taking the inverse Z-transform yields the Yule-Walker equations for the TAR model (7):

$$\sigma^2 h(-k) = \sum_{i=-p}^{p} \beta_i r(k-i), \; -\infty \leq k \leq \infty \quad (9)$$

where $h(k)$ is the inverse Z-transform of $H(z) = 1/\beta(z)$. Note that $h(k)$ is symmetric, noncausal, and is a function of β_k. Therefore, (9) is a nonlinear system of equations.

The TSP Wiener-Hopf equations can be written as [3]

$$\varepsilon_2^2\delta(k) = \sum_{i=-p}^{p} b_i r(k-i), \; -p \leq k \leq p. \quad (10)$$

It is obvious that (10) and (9) are different. Therefore, given the ACF $\{r(i)\}$, the TSP coefficients b_i obtained by solving (10) will be different from the TAR parameters β_i obtained by solving (9). This differs from the one-sided result.

In summary, for the one-sided model, the Yule-Walker equations which describe the relationship between the AR parameters and the ACF are the same as the Wiener-Hopf equations which describes the relationship between OSP coefficients and ACF. However, for the two-sided model, the Yule-Walker equations (9) and the Wiener-Hopf equations (10) are different. Consequently, the TSP coefficients are not the same as the two-sided AR parameters.

3.2. TSP Spectral Estimation

For this reason, the intuitive generalization of the OSP-based spectral estimator

$$\hat{P}_x(z) = \frac{\varepsilon_2^2}{B(z)B(z^{-1})} = \sum_{k=-\infty}^{\infty} \hat{r}(k)z^{-k} \quad (11)$$

where $B(z)$ is the optimal two-sided predictor and ε_2^2 is TSP MSE, will give a biased spectral estimate. The reason for this is the incorrect assumption implicitly made in (11) that the residual signal of TSP is white. Here we deliberately choose the notation $B(z)$ instead of $\beta(z)$ to distinguish the TSP coefficients b_i from the TAR parameters β_i.

Despite this, TSP is useful in some spectral estimation problems. For example, the Prony line spectrum estimator is based on TSP. We will now discuss another application of TSP, a TSP-based AR spectral estimator.

Define the revised spectral estimator

$$\hat{P}_x(z) = \frac{\varepsilon_2^2}{B(z)} \quad (12)$$

where $B(z)$ is the optimal two-sided predictor and ε_2^2 is the two-sided prediction error power. We now show for a true AR(p) process, (12) yields the true spectrum. We have

$$B(z) = \frac{A(z)A(z^{-1})}{1 + \sum_{i=1}^{p} a_i^2}, \quad \varepsilon_2^2 = \frac{\varepsilon_1^2}{1 + \sum_{i=1}^{p} a_i^2} \quad (13)$$

where $A(z)$ and ε_1^2 are the one-sided prediction error filter and prediction error power, respectively. Then (12) is

$$\hat{P}_x(z) = \frac{\varepsilon_2^2}{B(z)} = \frac{\varepsilon_1^2}{A(z)A(z^{-1})} = P_{True}(z). \quad (14)$$

The last equality comes from the fact that OSP coefficients are the same as the AR parameters and the OSP error power is equal to the excitation noise variance.

Since this only holds for true AR processes, the TSP-based spectral estimator (12) is not suitable for non-AR processes. We can also show that it does not have the correlation matching property of OSP. However, since any wss process can be approximated by a high-order AR process, (12) is valid as $p \to \infty$. This is the result of [1].

3.3. Frequency-Domain Error Criteria

To discuss the frequency domain behavior of (12), we first review some results from OSP. It has been shown that the mean square prediction error criterion in OSP is equivalent to the following frequency domain error criterion:

$$E[e(n)^2] = \frac{\varepsilon_1^2}{2\pi}\int_{-\pi}^{\pi} \frac{P_x(\omega)}{\hat{P}_x(\omega)} d\omega. \quad (15)$$

An important feature of (15) is that $\hat{P}_x(\omega)$ fits $P_x(\omega)$ better where $P_x(\omega)$ is greater than $\hat{P}_x(\omega)$, i.e., $\hat{P}_x(\omega)$ tends to fit peaks better than nulls. This is because the contribution to $E[e(n)^2]$ is more significant when $P_x(\omega)$ is greater than $\hat{P}_x(\omega)$ than when $P_x(\omega)$ is smaller.

We now show that minimizing the TSP mean square error is equivalent to minimizing the following frequency domain error criterion:

$$E[e(n)^2] = \frac{\varepsilon_2^4}{2\pi}\int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)^2} d\omega. \quad (16)$$

Since $e(n)$ is the output of the TSP error filter $B(z)$ driven by $x(n)$, the PSDs of $e(n)$ and $x(n)$ are related by $P_e(\omega) = P_x(\omega)|B(e^{j\omega})|^2$. Therefore, we have

$$E[e(n)^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_e(\omega)d\omega = \frac{\varepsilon_2^4}{2\pi}\int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)^2} d\omega. \quad (17)$$

It implies that $\hat{P}_x(\omega)$ fits $P_x(\omega)$ better where $P_x(\omega)$ is *small*, i.e., $\hat{P}_x(\omega)$ tends to fit *nulls better than peaks* because the contribution to $E[e(n)^2]$ is more significant when $P_x(\omega)/\hat{P}_x(\omega)^2 \approx 1/P_x(\omega)$ is large. This error criterion is quite different from that of the OSP-based spectral estimator.

4. TSP APPLICATIONS

4.1. Linear-Phase FIR Filter Design

A special class of linear-phase FIR filters is the TSP error filters or smoothers

$$x(n) = -\sum_{i=1}^{p} b_i(x(n-i) + x(n+i)) + e(n) \quad (18)$$

where $x(n)$ is the input signal and $e(n)$ is the output signal. It has been used to suppress narrow-band interference in a direct-sequence spread-spectrum system. Although it has been observed that it provides a much faster rise in the transition band, and smaller ripples in the pass band, compared to the combination of OSP error filters and their matched filters, no explanation for this has been given.

From (16), we now know that TSP-based spectral estimators tend to match spectral nulls better than the OSP-based spectral estimator. Thus our treatment of TSP explains a result not understood previously.

4.2. Data Interpolation

To recover a missing point in a time series, either OSP or TSP can be used. According to results above, TSP yields smaller MSE than OSP, and hence will give better estimates with no additional multiplications in the interpolation.

To illustrate this, we used a broadband AR(4) process with coefficients $[-1.352, 1.338, -0.662, 0.24]$. The performances of the one-sided covariance method (MCOV), the two-sided autocorrelation method (TSA), and the two-sided covariance method (TSC) were evaluated. The MSE obtained as the average of 50 realizations of data length 512 each for orders $p = 1, \cdots 6$ are shown in Fig. 1. It is clear that the two-sided covariance method (TSC) yields the smallest MSE.

4.3. Spectral Estimation

To test the ability of the TSP-based spectral estimator to separate close spectral lines, we used

$$x(n) = \cos(1.3n) + \cos(1.5n) + 0.1u(n), \quad 0 \le n \le 255 \quad (19)$$

where $u(n)$ is a white Gaussian sequence with unit variance. The model order used was 4 in OSP and 2 in TSP. 50 estimates of each method were obtained by computer simulation. The OSP covariance method failed to resolve the two sinusoids, while the TSP covariance method successfully resolved them, as shown in Fig. 2.

We now consider the case of a white signal in narrowband interference:

$$x(n) = \sum_{i=0}^{4} \cos((1.3 + 0.05i)n) + 0.1u(n), \quad 0 \le n \le 1023. \quad (20)$$

The model order used was 15 in both OSP and in TSP. The amplitude responses of the TSP error filter and the combination of the OSP error filter and its matched filter were obtained by computer simulation. 50 estimated amplitude responses each of TSP and OSP error filters are shown in Figs. 3 and 4. It is clear that the TSP error filters provide faster rise in the transition band and smaller ripples in the passband. This makes TSP filters suitable for narrowband interference rejection, as noted previously.

5. CONCLUSION

This paper has presented a comparison of the OSP and TSP problems. The results can be summarized as follows. While the OSP error processes are white, the TSP error processes are non-white, and have spectra proportional to the inverse of the signal spectra. The TSP MSE is always smaller than the OSP MSE, and it is smaller by a known factor for finite order AR process. Unlike the one-sided case, the optimal TSP coefficients are different from the TAR parameters. When applied to spectral estimation, frequency-domain error criteria show that OSP tends to match the spectrum peaks while TSP tends to match the spectrum nulls. Furthermore, the correlation matching property holds in OSP, but not in TSP. Various applications of TSP, including FIR filter design, and linear interpolation, were also discussed. These results were published in[4].

6. REFERENCES

[1] S. M. Kay, "Some Results in Linear Interpolation Theory," IEEE Trans. Acoust., Speech, and Sig. Proc., vol. ASSP-31, pp. 746-749, June 1983.

[2] S. L. Marple, "Fast Algorithms for Linear Prediction and System Identification Filters with Linear Phase," IEEE Trans. Acoust., Speech, and Sig. Proc., vol. ASSP-30, pp. 942-953, 1982.

[3] J.-J. Hsue and A. E. Yagle, "Fast Algorithms for Close-to-Toeplitz-Plus-Hankel Systems and Two-Sided Linear Prediction," IEEE Trans. Sig. Proc., vol. 41, no. 7, pp. 2349-2361, July 1993.

[4] J.-J. Hsue and A. E. Yagle, "Similarities and Differences between One-Sided and Two-Sided Linear Prediction," to appear in IEEE Trans. Sig. Proc., 1995.