AN FIR PREDICTOR INTERPRETATION OF LS ESTIMATION OF SINUSOIDAL AMPLITUDES FOLLOWED BY EXTRAPOLATION

Torbjörn Ekman

UniK - University Graduate Center, PO Box 70, N-2027 Kjeller, Norway
phone: +47 6484 4733, fax: +47 6381 8146, email: torbjorn@unik.no
web: www.unik.no

ABSTRACT

Prediction of time series consisting of a weighted sum of complex sinusoids corrupted by additive noise can be used for e.g. mobile radio channels. Least squares estimation of the complex amplitudes of sinusoids with known frequencies followed by extrapolation to the desired prediction range, is here interpreted as a linear FIR predictor using a sinusoidal filter. This interpretation enables an analysis of the properties of the predictor. An improved method utilizing a Wiener predictor, designed under the assumption that all the complex sinusoids have equal magnitudes, is proposed. The performance of the different prediction methods are evaluated in simulations and the proposed predictor is seen to have a performance close to the optimal Wiener predictor.

1. INTRODUCTION

There has been recent attention to the estimation of the parameters of a sum of complex sinusoids from noisy data with applications in prediction of mobile radio channels [1, 2, 3, 4]. The time varying mobile radio channel is commonly modelled as a weighted sum of complex sinusoids [5]. As this model is valid only over a limited measurement interval, prediction methods relying on parametric models have to use quite short data windows for the estimation of the parameters. The prediction is performed in three steps: First the frequencies of the contributing sinusoids are estimated using some subspace based method like ESPRIT or MUSIC. Then the corresponding amplitudes are estimated using a least squares fit of the sinusoids to the data. Finally the complex sinusoids are extrapolated to the desired prediction range. This has been demonstrated to work on simulated data but as soon as the number of sinusoids grow and there is additional noise, the length of the estimation window has to be rather long to obtain feasible prediction results.

In Section 2 of this paper the signal model is introduced and in Sections 3 and 4 it is shown that the least squares estimation of the complex amplitudes and the following extrapolation of the signal can be interpreted as a linear FIR predictor. This predictor is a sinusoidal filter that only depends on the estimated frequencies. It minimize the noise gain under the constraint of perfect sinusoidal prediction in the absence of noise. The resulting sinusoidal FIR predictor is the complex equivalent to the predictor derived in [6] for real valued signals.

Under the assumption of known frequencies, the FIR predictor interpretation of the last two steps enables an analysis of the performance of the parameter based prediction method. As we will see in Section 6 the performance of the sinusoidal filter is limited when the estimation interval is short. It is well known that such filters can have a high gain peak in their passband and that the stopband attenuation is poor. In [7] an IIR extension of an FIR predictor is proposed to reduce the noise gain for a similarly designed polynomial filter. This approach could be applied to sinusoidal filters to [6] but IIR filters are not suited for short windows of data, which is the case here. In Section 5 a method of designing an FIR predictor that makes a compromise between the perfect prediction of sinusoids and the noise gain is proposed. Using this method the length of the estimation interval that has to be used to obtain reasonable prediction results can be decreased.

2. SIGNAL MODEL

Denote the measured quantity by $y(t)$ where t is the discrete time instants. We assume an additive signal model, that is

$$y(t) = x(t) + n(t) \tag{1}$$

where $x(t)$ is the target signal to be extracted and $n(t)$ describes the measurement noise, modelling errors, etc. The target signal is modelled as a sum of M weighted complex sinusoids

$$x(t) = \sum_{k=1}^{M} \alpha_k f_k(t) \tag{2}$$

where

$$f_k(t) = e^{i\omega_k t} \tag{3}$$

denotes the complex sinusoidal basis function with frequency ω_k and α_k is the corresponding complex weight. For mobile radio channels the frequencies are given by

$$\omega_k = 2\pi f_D \cos\theta_k / f_s \tag{4}$$

where θ_k is the angle towards the kth wavefront, f_D is the Doppler frequency [5] and f_s is the channel sampling frequency. The signal is thus bandlimited to $\pm f_D/f_s$ with an over sampling ratio (OSR) of $f_s/2f_D$. The small scale fading is mainly a spatial effect and it is thus natural to measure the length of the estimation interval in travelled wavelengths. The number of samples in the estimation interval relates to the physical length r (measured in wavelengths λ), as $N = rf_s/f_D$. The fastest oscillating component experience at most one full cycle in an interval of length one wavelength.

The model (2) can be expressed as a vector multiplication

$$x(t) = [f_1(t) \ldots f_M(t)] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_M \end{bmatrix} = \mathbf{f}^T(t)\boldsymbol{\alpha}. \tag{5}$$

This model may be valid only locally in time, that is the parameters $\{\alpha_k\}$ and frequencies $\{\omega_k\}$ may be slowly time varying or might even change abruptly. In this work, the parameters are assumed to be constant within a window of length N samples. A window of N noise free data points can be expressed as

$$\begin{bmatrix} x(t) \\ x(t-1) \\ \vdots \\ x(t-N+1) \end{bmatrix} = \begin{bmatrix} \mathbf{f}^T(t) \\ \mathbf{f}^T(t-1) \\ \vdots \\ \mathbf{f}^T(t-N+1) \end{bmatrix} \boldsymbol{\alpha}, \tag{6}$$

which for short is written as

$$\mathbf{x}(t) = \mathbf{F}(t)\boldsymbol{\alpha}. \tag{7}$$

Including the noise, the N measured data points can be expressed as

$$\mathbf{y}(t) = \mathbf{F}(t)\boldsymbol{\alpha} + \mathbf{n}(t), \tag{8}$$

where $\mathbf{y}(t)$ and $\mathbf{n}(t)$ are vectors formed like $\mathbf{x}(t)$ but consisting of measured data points and noise samples respectively.

3. ESTIMATION AND EXTRAPOLATION

The aim is to predict the target signal $x(t)$ in (2), L steps ahead. Under the assumption that the frequencies are known, a straight forward solution is to estimate $\boldsymbol{\alpha}$ and use this estimate in the parametric model (5) as [1, 3, 4]

$$\hat{x}(t+L) = \mathbf{f}^T(t+L)\,\hat{\boldsymbol{\alpha}}. \tag{9}$$

The maximum likelihood estimate of $\boldsymbol{\alpha}$ is the least squares estimate obtained as

$$\hat{\boldsymbol{\alpha}} = \mathbf{F}^{\dagger}(t)\mathbf{y}(t), \tag{10}$$

where $\mathbf{F}^{\dagger}(t) = (\mathbf{F}^H(t)\mathbf{F}(t))^{-1}\mathbf{F}^H(t)$ denotes the Moore-Penrose pseudoinverse. A predictor for $x(t+L)$ then comes through

$$\hat{x}(t+L) = \mathbf{f}^T(t+L)\,\mathbf{F}^{\dagger}(t)\,\mathbf{y}(t), \tag{11}$$

which is an extrapolation of the estimated waveforms.

4. FIR PREDICTOR INTERPRETATION OF ESTIMATION AND EXTRAPOLATION

4.1 Derivation of the Equivalent FIR Filter

For sinusoidal basis functions $\mathbf{f}(t+L)$ can be decomposed in a time invariant vector and a time varying invertible matrix

$$\mathbf{f}^T(t+L) = \mathbf{f}^T(L)\mathbf{B}(t) \tag{12}$$

$$= [e^{i\omega_1 L} \dots e^{i\omega_M L}] \begin{bmatrix} e^{i\omega_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{i\omega_M t} \end{bmatrix}. \tag{13}$$

This leads to the following result:

Theorem 4.1 *If the basis functions can be decomposed as*

$$\mathbf{f}^T(t+L) = \mathbf{f}^T(L)\mathbf{B}(t), \tag{14}$$

where $\mathbf{B}(t)$ is an invertible quadratic matrix, then the predictor (11) can be expressed as the filter operation

$$\hat{x}(t+L) = \mathbf{f}^T(t+L)\mathbf{F}^{\dagger}(t)\mathbf{y}(t) = \mathbf{c}^T\mathbf{y}(t), \tag{15}$$

where the time invariant FIR filter $\mathbf{c}$ is

$$\mathbf{c} = \mathbf{F}^{\dagger T}\mathbf{f}(L). \tag{16}$$

Here the time index 0 for $\mathbf{F}(0)$ is dropped for notational convenience, that is $\mathbf{F} = \mathbf{F}(0)$.

Proof 4.1 *With the condition (14) fulfilled the matrix of basis functions can be decomposed as*

$$\mathbf{F}(t+L) = \mathbf{F}(L)\mathbf{B}(t), \tag{17}$$

and the relation (8), with $L = 0$, can be written as

$$\mathbf{y}(t) = \mathbf{F}\mathbf{B}(t)\boldsymbol{\alpha} + \mathbf{n}(t). \tag{18}$$

The LS estimate (10) of $\boldsymbol{\alpha}$ using the N latest samples is then

$$\hat{\boldsymbol{\alpha}} = (\mathbf{F}\mathbf{B}(t))^{\dagger}\mathbf{y}(t) = \left(\mathbf{B}^H(t)\mathbf{F}^H\mathbf{F}\mathbf{B}(t)\right)^{-1}\mathbf{B}^H(t)\mathbf{F}^H\mathbf{y}(t)$$
$$= \mathbf{B}^{-1}(t)\mathbf{F}^{\dagger}\mathbf{y}(t). \tag{19}$$

Using (14) and (19) in (9) the L step ahead predictor can be written as

$$\hat{x}(t+L) = \mathbf{f}^T(t+L)\hat{\boldsymbol{\alpha}} = \mathbf{f}^T(L)\mathbf{B}(t)\mathbf{B}^{-1}(t)\mathbf{F}^{\dagger}\mathbf{y}(t)$$

where the time dependent part cancels out so we obtain the time invariant filter expression

$$\hat{x}(t+L) = \mathbf{f}^T(L)\mathbf{F}^{\dagger}\mathbf{y}(t) = \mathbf{c}^T\mathbf{y}(t). \tag{20}$$

The filter $\mathbf{c}$ is a time invariant linear FIR filter with N taps, whose design equation is given by (16). ■

FIR prediction using the coefficients in (16) can be interpreted as first making an LS estimate of the weights of the basis functions and then evaluating the weighted basis functions at time $t+L$. Damped complex exponentials, as in [2], also fulfills (14) and thus result in linear FIR predictors.

The time series is not limited to be stationary since a length-N data window is utilized. The optimization criterion used in the design is not to minimize the prediction error but to minimize the model error for the basis functions on a window of data. The method optimizes the wrong property and result in a *suboptimal* linear predictor for the signal.

4.2 Properties of the FIR Predictor

The sensitivity of the filter to broadband noise is given by the noise gain (NG). For an FIR filter the noise gain is the sum of the squared amplitude of the filter coefficients. In particular, for the filter in equation (16)

$$\mathrm{NG} = \mathbf{c}^T\mathbf{c}^* = \mathbf{f}^T(L)\mathbf{F}^{\dagger}\,\mathbf{F}^{\dagger H}\mathbf{f}^*(L) = \mathbf{f}^T(L)\,(\mathbf{F}^H\mathbf{F})^{-1}\mathbf{f}^*(L). \tag{21}$$

Here $\cdot^*$ denotes complex conjugate. In the presence of additive white noise with variance σ_n^2 the variance of the prediction error will be $\sigma_\varepsilon^2 = \mathrm{NG}\,\sigma_n^2$.

The elements of the matrix $\mathbf{F}^H\mathbf{F}$ are

$$\left[\mathbf{F}^H\mathbf{F}\right]_{k,l} = \sum_{n=0}^{N-1} e^{i(\omega_k-\omega_l)n} = \begin{cases} N & k = l \\ \frac{1-e^{i(\omega_k-\omega_l)N}}{1-e^{i(\omega_k-\omega_l)}} & k \neq l \end{cases}. \tag{22}$$

For basis functions that are orthogonal over the estimation interval, that is $1 = e^{i(\omega_k-\omega_l)N}\ \forall\, k,l$, we obtain $\mathbf{F}^H\mathbf{F} = N\mathbf{I}$. As $\mathbf{f}^T(L)\mathbf{f}^*(L) = M$ the noise gain then becomes M/N.

If we let the length of the estimation interval approach infinity in (22) we obtain

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{F}^H\mathbf{F} = \mathbf{I}. \tag{23}$$

The noise gain thus approaches the value M/N for all choices of frequencies when N grows.

5. IMPROVED FIR PREDICTOR

5.1 Correlations

The autocorrelation for a complex sinusoidal signal is [8]

$$r_x(n) = \mathrm{E}\{x(t)x^*(t-n)\} = \sum_{k=1}^{M} |\alpha_k|^2 e^{j\omega_k n} = \mathbf{f}^H(0)\mathbf{A}\mathbf{f}(n), \tag{24}$$

where $\mathbf{A}$ is the M diagonal matrix with squared amplitudes

$$\mathbf{A} = \mathrm{diag}(|\alpha_1|^2, \ldots, |\alpha_M|^2). \tag{25}$$

The $N \times N$ correlation matrix can be obtained from (24) as

$$\mathbf{R}_x = \mathrm{E}\{\mathbf{x}^*(t)\mathbf{x}^T(t)\} = \mathbf{F}^*\mathbf{A}\mathbf{F}^T. \tag{26}$$

Under the assumption that the signal $x(t)$ is uncorrelated with the additive noise, the correlation matrix for $\mathbf{y}(t)$ is

$$\mathbf{R}_y = \mathrm{E}\{\mathbf{y}^*(t)\mathbf{y}^T(t)\} = \mathbf{R}_x + \mathbf{R}_n, \tag{27}$$

where $\mathbf{R}_n$ is the correlation matrix for the additive noise. The correlation between the target for prediction $x(t+L)$ and the vector of measurements $\mathbf{y}(t)$ is obtained as

$$\mathbf{r}_{xy} = \mathrm{E}\{x(t+L)\mathbf{y}^*(t)\} = \mathbf{F}^*\mathbf{A}\mathbf{f}(L). \tag{28}$$

5.2 Wiener Filter

The optimal linear predictor is obtained from the Wiener filter equation

$$\mathbf{R}_y\mathbf{c} = \mathbf{r}_{xy}, \tag{29}$$

which for sinusoidal signals using (26), (27) and (28) is

$$(\mathbf{F}^*\mathbf{A}\mathbf{F}^T + \mathbf{R}_n)\mathbf{c} = \mathbf{F}^*\mathbf{A}\mathbf{f}(L). \tag{30}$$

This Wiener predictor minimize the variance of the prediction error. If there is no noise we see that the optimal filter $\mathbf{c}$ is obtained as in (16). That filter is thus the minimum norm filter, which means that it is the predictor with the lowest noise gain, that fulfills the property of perfect prediction in the case of no additive noise. To reduce the noise gain we therefore have to relax the constraint of perfect prediction.

5.3 Improved Prediction Method

The frequencies of the sinusoids are assumed to be known but the amplitudes are unknown and we can hence not use the optimal Wiener predictor. A common assumption about the amplitudes are that they are stochastic variables with zero mean and equal variance, drawn from a complex Gaussian distribution. Instead of the true signal correlation we can use the average correlation given a particular set of frequencies. To obtain a signal with the variance σ_x^2 the squared magnitudes of the weights should equal σ_x^2/M. We thus use $\mathbf{A} = (\sigma_x^2/M)\,\mathbf{I}$ in (30).

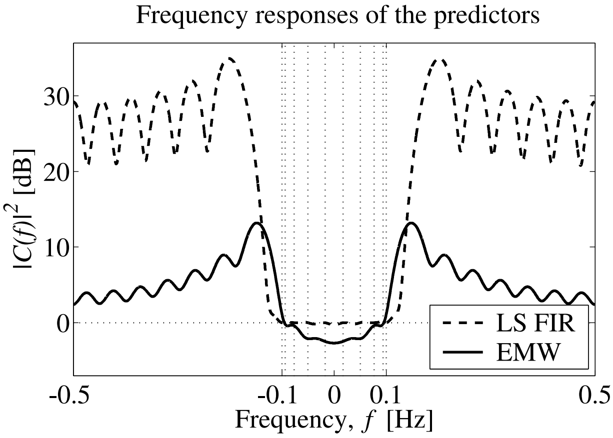

Figure 1: Frequency response for the LS FIR predictor in (16) and EMW (31). The predictors are designed for 10 complex sinusoids, an estimation interval of 2 wavelengths consisting of $N = 20$ samples (OSR=5) and a prediction range of 0.2 wavelengths. The white additive noise give an SNR of 20 dB. The vertical lines give the positions of the frequencies.

We can assume that the properties of the noise change more slowly than the signal and it is feasible to assume the correlation matrix, or at least the variance of the noise, to be known. The improved FIR predictor is then obtained as

$$\mathbf{c} = \left(\mathbf{F}^*\mathbf{F}^T + \frac{M}{\sigma_x^2}\mathbf{R}_n\right)^{-1} \mathbf{F}^*\mathbf{f}(L). \tag{31}$$

This is an equal magnitude Wiener (EMW) predictor for complex sinusoidal signals. When the full correlation matrix of the noise is unknown we can assume white additive noise and set $\mathbf{R}_n = \sigma_n^2\mathbf{I}$. The noise variance σ_n^2 can be used as a tuning parameter to obtain a desired noise gain. This predictor makes the necessary trade off between the noise gain and a relaxed condition of perfect prediction in the noiseless case.

5.4 A Mobile Radio Channel Example

The frequency responses for the predictors for 10 complex sinusoids, obtained as in (16) and (31) are plotted in Fig. 1. The angles $\{\theta_k\}$ are equally spaced in the interval $[0\ \pi]$ and the weights $\{\alpha_k\}$ have equal magnitudes. The frequency response of the predictor obtained as in (16) reassembles that of a predictor of a bandlimited noiseless signal [9]. High gain outside the bandpass region result in high amplification of additive broadband noise and consequently poor prediction. The EMW-predictor designed for an SNR of 20 dB has a much lower gain outside the bandpass region but does not have the desired 0 dB amplification at the frequencies of the sinusoids.

6. SIMULATIONS

6.1 Signal Model

The signal is modelled as in (2)-(4). All directions for the contributing wavefronts are assumed equally probable and the angles θ_k are thus drawn from a uniform distribution between 0 and 2π. The amplitudes are drawn from a complex

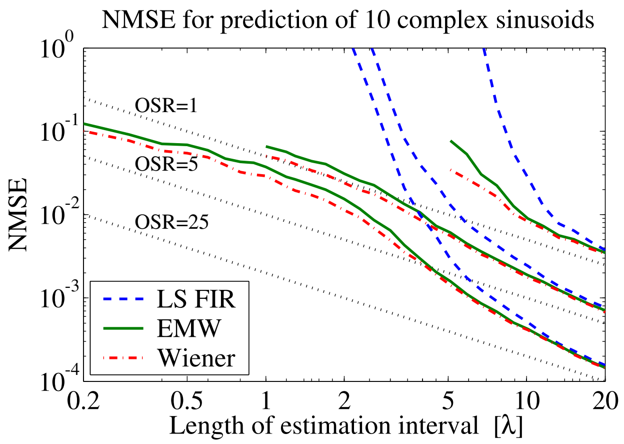

Figure 2: The median prediction NMSE for a signal consisting of 10 complex sinusoids. The prediction range is 0.2 wavelengths and the over sampling ratios 1, 5 and 25. The straight dotted line is $\sigma_n^2 M/N$.

circular Gaussian distribution and normalized to give the signal unit variance. The noise is white with variance $\sigma_n^2 = 0.01$, resulting in an SNR of 20 dB. The OSRs are 1, 5 and 25 corresponding to 2, 10 and 50 samples per wavelength, respectively. To obtain the MSE of the prediction error we only need to calculate the correlations (27), (28) and the different predictors (16) and (31). The prediction MSE is obtained as

$$\sigma_\varepsilon^2 = \sigma_x^2 - \mathbf{r}_{xy}^H \mathbf{c} - \mathbf{c}^H \mathbf{r}_{xy} + \mathbf{c}^H \mathbf{R}_y \mathbf{c}. \quad (32)$$

For comparison the optimal Wiener predictor (30) is also evaluated. 200 Monte Carlo trials are performed and the prediction NMSE, $\sigma_\varepsilon^2/\sigma_x^2$, is calculated for estimation intervals of lengths from 0.2 up to 20 wavelengths.

6.2 Results

In Fig. 2-3 the median NMSE for a prediction range of 0.2 wavelengths, for the Monte Carlo trials are shown. The difference in performance between the optimal Wiener predictor and the EMW-predictor is small. The predictors approach the performance obtained for the LS-method with orthogonal basis functions, when the length of the estimation interval increases. With an OSR of 25 and $M = 20$ the estimation interval has to be at least 6 wavelengths, $N = 300$ samples, for the LS-basis function approach to perform better than using the average as prediction. As comparison, the EMW predictor obtains an NMSE below 0.01 at the same length.

7. CONCLUSION

Least squares estimation of the complex amplitudes of sinusoids with known frequencies, followed by extrapolation of the sinusoids to the desired prediction range can be interpreted as linear FIR prediction. The noise gain of this FIR predictor can be so high that prediction becomes unfeasible. This occurs when the number of complex sinusoids is high and the estimation interval, measured in wavelengths, is short. Much shorter estimation intervals at a given prediction performance can be obtained using the proposed EMW-predictor, which is a Wiener predictor designed under the assumption that all the complex sinusoids have equal magnitudes.

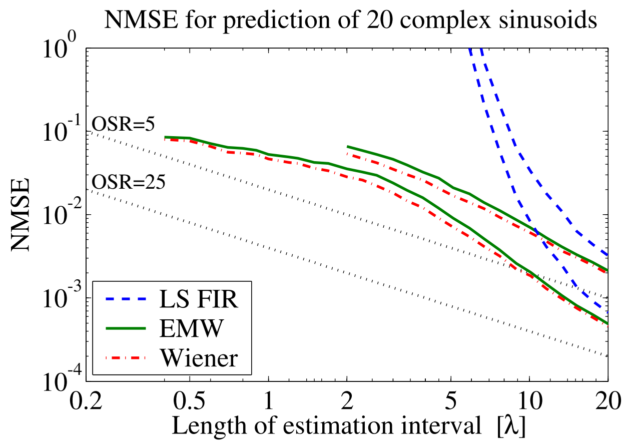

Figure 3: As Fig. 2 but for 20 complex sinusoids.

For white additive noise the NMSE of the predictors approach $\sigma_n^2 M/\sigma_x^2 N$ when N increases. High oversampling ratios are thus beneficial, as then there are more samples available (high N) within a given estimation interval.

REFERENCES

[1] J.-K. Hwang and J. H. Winters, "Sinusoidal modeling and prediction of fast fading processes," in *Proc. IEEE Globecom'98*, Sidney, Australia, Nov. 1998, pp. 892–897.

[2] J. B. Andersen, J. Jensen, S. H. Jensen and F. Fredriksen, "Prediction of future fading based on past measurements," in *Proc. IEEE VTC Fall 1999*, Amsterdam, Netherlands, Sept. 1999, pp. 151–155.

[3] R. Vaughan, P. Teal and R. Raich, "Short-term mobile channel prediction using discrete scatterer propagation model and subspace signal processing algorithms," in *Proc. VTC Fall 2000*, Boston, MA, USA, Sept. 2000, pp.751–758.

[4] Liang Dong, Guanghan Xu, Hao Ling, "Prediction of fast fading mobile radio channels in wideband communication systems," in *Proc. IEEE Globecom'01*, San Antonio, TX, USA, Nov. 2001, pp. 3287–3291.

[5] W.C. Jakes, Jr, *Microwave Mobile Communications*. Address: IEEE Press, Piscataway, NY, 1974.

[6] P. Händel, "Predictive digital filtering of sinusoidal signals," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 364–374, Feb. 1998.

[7] S. J. Ovaska, O. Vainio and T. I. Laakso , "Design of predictive IIR filters via feedback extension of FIR forward predictors," *IEEE Trans. Instrumentation and Measurement*, vol. 46, no. 5, pp. 1196–1208, Feb. 1997.

[8] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Address: Prentice-Hall, Upper Saddle River, New Jersy, 1997.

[9] A. Papoulis, "A note on the predictability of bandlimited processes," *Proc. IEEE*, vol. 73, no. 8, pp. 1332–1333, Aug. 1985.