PAPER

Spectrum Estimation by Noise-Compensated Data Extrapolation

Jonah GAMBA[†a)], ***Student Member and*** **Tetsuya SHIMAMURA**[†], ***Member***

SUMMARY High-resolution spectrum estimation techniques have been extensively studied in recent publications. Knowledge of the noise variance is vital for spectrum estimation from noise-corrupted observations. This paper presents the use of noise compensation and data extrapolation for spectrum estimation. We assume that the observed data sequence can be represented by a set of autoregressive parameters. A recently proposed iterative algorithm is then used for noise variance estimation while autoregressive parameters are used for data extrapolation. We also present analytical results to show the exponential decay characteristics of the extrapolated samples and the frequency domain smoothing effect of data extrapolation. Some statistical results are also derived. The proposed noise-compensated data extrapolation approach is applied to both the autoregressive and FFT-based spectrum estimation methods. Finally, simulation results show the superiority of the method in terms of bias reduction and resolution improvement for sinusoids buried in noise.
key words: *spectrum estimation, noise variance, Yule-Walker equations, autoregressive process*

1. Introduction

Traditional methods of spectrum estimation solely rely on the discrete Fourier transform (DFT) to obtain the power spectrum of an observed data sequence [1]–[3]. Popular among these are the periodogram and Blackman-Turkey methods. Their major weakness, however, is that they fail to resolve closely spaced spectral peaks when short-duration data sequences are used. This weakness is the main motivation behind the so-called high-resolution techniques; notably the autoregressive (AR) [4]–[6] and eigen-decomposition techniques [7]. Although these high resolution techniques give better-resolved spectrum estimates, it is only at high signal-to-noise ratios (HSNR's) that their performance can be guaranteed to be superior to the traditional methods [8]. The reason for this could be attributed to the inability to accurately separate the signal subspace from the noise subspace at low signal-to-noise-ratios (LSNR's). The presence of noise is a serious problem for AR parameter estimation [9]–[11], usually involving the use of higher-order Yule-Walker equations (YWE's) [12]. The use of autoregressive moving average (ARMA) models for noisy AR has been recommended in [8]. However, the resulting nonlinear equations cannot be easily solved, thus making the frequency estimates sub-optimal. At LSNR's the distinction between large and small eigenvalues becomes obscure and hence methods like rank reduction fail to produce satisfactory results. In particular, MUSIC [13] performs badly at LSNR's.

Another approach that has been successfully used to enhance the spectral resolution for two-dimensional [14]–[16] and also one-dimensional [17] spectrum estimation is the data extrapolation method. However, the data extrapolation method employed alone has been shown to result in rapid decrease in the magnitude of extrapolated data points [16]. This trend has been intuitively attributed to the increase in prediction error [24]. Nonetheless, this method can become very attractive if it is coupled with noise compensation [18]. If the noise variance can be estimated from the observed data sequence, then the data extrapolation method can improve frequency resolution and bias considerably. The aim of this paper is therefore to analytically characterize the data extrapolation process and to demonstrate the effectiveness of the noise-compensated data extrapolation as a spectrum estimation technique.

To achieve the abovementioned goals we use a recently proposed iterative noise variance estimation (INVE) technique based on low-order YWE's [19]. The noise variance estimate is used in extracting the AR parameters from the observed data sequence. These AR parameters are in turn used in conjunction with data extrapolation for spectrum estimation. We analyze the effect of data extrapolation in both the time domain and the frequency domain to reveal the characteristics of the extrapolated data points. Such an analysis, in relation to noise-compensated data extrapolation, has not been formally presented in the literature. The relationship between data extrapolation, noise variance, bias and resolution is also unveiled. In this paper it is shown that the decrease in value of the extrapolated data points is exponential in nature and has the effect of smoothing the spectrum.

This paper is organized as follows. In Sect. 2 we review the INVE method. Section 3 presents the proposed method of spectrum estimation. Section 4, which is an expansion on Sect. 3, gives a more detailed analysis of the data extrapolation method and implications in both the time and frequency domains. Statistical assessment of the method is also given in this section. In Sect. 5 simulation results are presented. Concluding remarks in Sect. 6 end this paper.

Manuscript received February 17, 2004.
Manuscript revised September 3, 2004.
Final manuscript received December 7, 2004.
[†]The authors are with the Department of Information and Computer Sciences, Saitama University, Saitama-shi, 338-8570 Japan.
a) E-mail: jonah@sie.ics.saitama-u.ac.jp
DOI: 10.1093/ietfec/e88–a.3.702

2. Review of the INVE Method

In this section, we review the INVE method [19] since it will be useful in understanding the analysis of the noise-compensated data extrapolation method. The INVE method is based on low order YWE's. The INVE method is iterative and at each iteration, an estimate of the noise-free Yule-Walker solution is obtained by adjusting the noise variance parameter and constraining the solution to match the autocorrelation sequence of the observed data sequence. The adjustment is achieved through a step-size parameter which depends on the current value of the noise variance parameter. The true noise variance minimizes the difference between the second norms of the noisy Yule-Walker solution and the estimated noise-free Yule-Walker solution.

2.1 Formulation of the Noisy AR Model

For the purpose of noise variance estimation, we define a stationary AR process $y(n)$ of order p by

$$y(n) = -\sum_{i=1}^{p} a(i)y(n-i) + e(n),$$

where $e(n)$ is an uncorrelated driving white noise sequence of variance σ_e^2 and the $a(i)$'s are the noise-free AR parameters. The autocorrelation function (ACF) at lag k for $y(n)$ is defined by $r_{yy}(k) = E[y(n)y(n+k)]$ where $E[\cdot]$ is the expectation operator. The $r_{yy}(k)$ is given in the above case by

$$r_{yy}(k) = -\sum_{i=1}^{p} a(i)r_{yy}(k-i) + \delta(k)\sigma_e^2 \ \ k \geq 0$$

where $\delta(k)$ is the Kronecker delta function. In the presence of noise, the observed data sequence becomes $x(n) = y(n) + w(n)$, where $w(n)$ is assumed to be zero-mean additive white Gaussian noise of variance σ_w^2. The ACF for $x(n)$ is similarly defined like that for $y(n)$ and denoted by $r_{xx}(k)$.

The AR parameters can be estimated by using the YWE's [1] given by

$$\mathbf{R}_{yy}\mathbf{a} = -\mathbf{r}_y \tag{1}$$

$$\mathbf{R}_{xx}\hat{\mathbf{a}} = -\mathbf{r}_x \tag{2}$$

where $\mathbf{R}_{yy}$ and $\mathbf{R}_{xx}$ are $p \times p$ autocorrelation matrices (ACM's) of the sequences $y(n)$ and $x(n)$ respectively. The column vectors on the right hand sides of Eqs. (1) and (2) are $\mathbf{r}_y^T = [r_{yy}(1) \ldots r_{yy}(p)]$ and $\mathbf{r}_x^T = [r_{xx}(1) \ldots r_{xx}(p)]$. The T denotes the transposition operation. The $p \times 1$ vectors $\mathbf{a}$ and $\hat{\mathbf{a}}$ are the noise-free and noisy solutions to the YWE's, respectively. The AR parameter estimates from Eq. (2) are biased since $r_{xx}(k) = r_{yy}(k) + \delta(k)\sigma_w^2$. Using Eqs. (1) and (2), the following relationship between ACM's is validated for noise compensation,

$$\mathbf{R}_{yy} = \mathbf{R}_{xx} - \sigma_w^2\mathbf{I} \tag{3}$$

where $\mathbf{I}$ is a $p \times p$ identity matrix.

2.2 Noise Variance Estimation

Equations (1)–(3) can be utilized in the estimation of the noise variance σ_w^2. It can be shown that it is possible to estimate σ_w^2 from the low-order YWE's by defining a function $f(\alpha)$ such that

$$f(\alpha) = \|\hat{\mathbf{a}}\| - \|\tilde{\mathbf{a}}(\alpha)\| \tag{4}$$

where α is a parameter that gives an estimate of σ_w^2 and $\tilde{\mathbf{a}}(\alpha)$ corresponds to the solution of noise-compensated Yule-Walker equation obtained by combining Eq. (1) with Eq. (3). The value of α that gives the minimum of $f(\alpha)$ results in the noise variance estimate. This property of the function $f(\alpha)$ is described analytically in [19] and forms the basis of the INVE method. We extract the minimum of $f(\alpha)$ by adjusting the α in an iterative fashion.

3. Spectrum Estimation

Given a finite length of data sequence $x(0)$, $x(1)$, …, $x(N-1)$, suppose that based on the INVE method the $f(\alpha)$ is minimized when $\alpha = \alpha_1$. Then the spectrum estimation by noise-compensated data extrapolation can proceed as follows.

1. Estimate the biased noisy autocorrelation sequence $r_{xx}(k)$ from the data sequence as

$$\hat{r}_{xx}(k) = \frac{1}{N}\sum_{n=0}^{N-1-|k|} x(n)x(n+|k|). \tag{5}$$

2. Defining $\hat{\sigma}_w^2$ as the estimate of the noise variance from the INVE method, and substituting it in Eq. (3), obtain the noise-compensated AR parameters from Eq. (1), denoted by $\tilde{\mathbf{a}}(\alpha_1)$, and use these parameters to extrapolate the data sequence $x(n)$ as described in [18]; that is,

$$\hat{x}(n) = -\sum_{k=1}^{p} [\tilde{\mathbf{a}}(\alpha_1)]_k x(n-k) \tag{6}$$

where $\hat{x}(n)$ denotes the extrapolated points, $[\cdot]_k$ is the kth element of $\tilde{\mathbf{a}}(\alpha_1)$, and $n \geq N$. The resulting extrapolated data sequence is denoted by $x'(n) = [x(0), x(1), \ldots, x(N-1), \hat{x}(N), , \ldots, \hat{x}(Z-1)]$, where Z is the total data sequence length and $Z > N$.

3. Obtain new AR parameters from $x'(n)$ by solving Eq. (2) using the Levinson-Durbin algorithm.
4. Estimate the spectrum from the extrapolated data sequence $x'(n)$ by either of the following methods.
   - Through the FFT approach, using the DFT of $x'(n)$ which is defined by

$$X'(k) = \frac{1}{Z}\sum_{n=0}^{Z-1} x(n)e^{-j2\pi kn/Z} \tag{7}$$

   for $k = 0, \ldots, Z-1$. The power spectrum is then given by $|X'(k)|^2/Z$.

- Using the AR approach, with the AR power spectral density $P_{ar}(f)$ which is given by [1]

$$P_{ar}(f) = \frac{g^2}{|A(f)|^2} \tag{8}$$

and $A(f) = \sum_{k=0}^{p} [\hat{\mathbf{a}}']_k e^{-j2\pi fk}$, $(a_0 = 1)$, where the AR parameters $\hat{\mathbf{a}}'$ and gain power g^2 are obtained from Step 3 through the Levinson-Durbin algorithm.

In Step 2, the value of Z can be determined from practical considerations such as resolution requirements. For given length of available data sequence, the spectral resolution limit for the DFT can be readily determined. However, in the case of AR spectrum estimation, the optimum value of Z would depend on both the order and the SNR and in that case the DFT resolution requirements can be used as a guide to the amount of data extrapolation necessary. We also note that instead of performing eigendecomposition as in [18], the noise compensated AR parameters are obtained through the INVE method. As compared to previous methods, the above spectrum estimation method eliminates the need for adjustment parameters suggested in [5] and [18] during the process of noise variance estimation. The *a priori* knowledge of σ_w^2 as in [4] also becomes unnecessary.

4. Analytical Assessment of Data Extrapolation

In this section, we analyze the effect of noise-compensated data extrapolation in both the time and frequency domains. We also present the implications of noise-compensated data extrapolation from a statistical perspective.

4.1 Time Domain Effect

Given the observed data sequence $x(0), x(1), \ldots, x(N-1)$, we can extrapolate the sequence using prediction coefficients to arrive at the following expression for the $(i+1)$th extrapolated data sample:

$$\hat{x}(N+i) \approx \alpha_0 e^{-\beta i}, \tag{9}$$

where β can be taken to be a positive real number corresponding to the rate of decay of the extrapolated data samples. The derivation of Eq. (9) is given in Appendix. Equation (9) could explain the tapering effect of the extrapolated data points observed in [16]; that is, the extrapolated section can be approximated by a decreasing exponential function of the data, with initial value equal to α_0 as shown in Fig. 1.

4.2 Frequency Domain Effect

The effect of the extrapolated points in the frequency domain can be explained with the help of Heaviside's unit step function $u(t)$ and the replicating symbol (function) as defined in [22]. We will represent the replicating symbol (function) by $\mathfrak{R}(t)$ in the sequel. The t here is an independent variable.

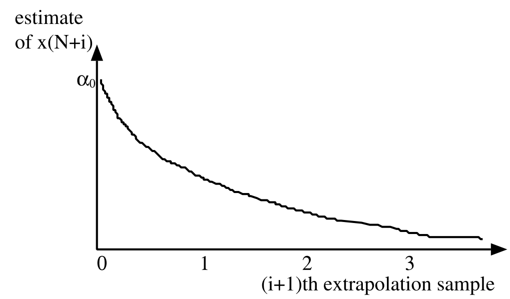

Fig. 1 Extrapolated data points from an observation of length N.

From the observed data sequence $x(0), x(1), \ldots, x(N-1)$, we can obtain the extrapolated sequence as described in the preceding section. Suppose that the extrapolated data sequence becomes $x'(0), x'(1), \ldots, x'(N-1), x'(N), x'(N+1), \ldots, x'(N+N_1-1)$, with N_1 new data points. Let us refer to this process of obtaining $x'(n)$ as forward data extrapolation. We recognize that

$$x'(n) = x(n) \quad n = 0, \ldots, N-1 \tag{10}$$

$$x'(n) = \hat{x}(n) \quad n = N, \ldots, N+N_1-1. \tag{11}$$

If we assume that the observed data sequence is a realization of a continuous time signal $x(t)$, the original data sequence can be arrived at by sampling with the $\mathfrak{R}(t)$ so that we have $x(t)\mathfrak{R}(t)$. The periodicity of N samples can be achieved by convolution to get $r(t)$ which can be expressed as

$$r(t) = [x(t)\mathfrak{R}(t)] * \mathfrak{R}(t)/N, \tag{12}$$

where the asterisk $(*)$ denotes the convolution operation. Assuming that the Fourier transform of $x(t)$ exists, then the Fourier transform of $r(t)$ also exists, and can be written as $R(f) = [X(f)*\mathfrak{R}(f)]\mathfrak{R}(Nf)$, where f denotes the frequency. We see that $R(f)$ is the Fourier transform of $x(t)$ sampled at $1/N$ intervals. Extending similar arguments on the extrapolated data sequence $x'(n)$ as on $x(t)$, the addition theorem of Fourier transforms, can be used to separate $x'(n)$ into two sequences $x'_1(n)$ and $x'_2(n)$ of length N_1 such that

$$x'_1(n) = \begin{cases} x'(n) & n = 0, \ldots, N-1 \\ 0 & n = N, \ldots, N+N_1-1 \end{cases}$$

$$x'_2(n) = \begin{cases} 0 & n = 0, \ldots, N-1 \\ x'(n) & n = N, \ldots, N+N_1-1. \end{cases}$$

This separation of data into two sequences is depicted in Fig. 2. In Fig. 2 the shape of the observed data is arbitrarily chosen. The data sequences $x'_1(n)$ and $x'_2(n)$ can be thought as those resulting from employing the sampling and replicating operations on continuous-time signals $x'_1(t)$ and $x'_2(t)$. Performing the Fourier transformation of the two data sequences and combining the results, we get

$$\begin{aligned} X'(f) = {} & [X'_1(f) * \mathfrak{R}(f)]\mathfrak{R}(N_e f) \\ & + [X'_2(f) * \mathfrak{R}(f)]\mathfrak{R}(N_e f) \end{aligned} \tag{13}$$

where $N_e = N + N_1$. Equation (13) can be simplified to

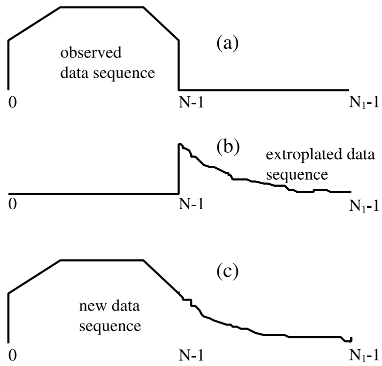

Fig. 2 Splitting of data sequence $x'(n)$ into two: (a) observed data sequence with trailing zeros $x'_1(n)$, (b) extrapolated data sequence with leading zeros $x'_2(n)$ and (c) addition of (a) and (b) to get $x'(n)$.

$$X'(f) = [X'_1(f) * \Re(f) + X'_2(f) * \Re(f)]\Re(N_e f). \tag{14}$$

In Eq. (14) we can interpret $X'_2(f)*\Re(f)$ as the Fourier transform of Eq. (9) multiplied by a delayed Heaviside's unit step function if Eq. (9) is taken into consideration. Since the resulting function is exponential in nature, it has the effect of smoothing the Fourier transform of the original data samples. The smoothing effect can be verified by considering the following. For Heaviside's unit step function $u(t)$, the Fourier transform pairs are defined by [23]

$$Ft\ e^{-\mu t}u(t) = F_p(\omega) = \frac{1}{\mu + j\omega} \tag{15}$$

$$Ft\ e^{\mu t}u(-t) = F_n(\omega) = \frac{1}{\mu - j\omega} \tag{16}$$

where the "Ft" denotes the Fourier transform operator, $\omega = 2\pi f$ is the radian frequency and $\mu > 0$. Equation (15) is similar to the one used for forward data extrapolation; that is, extrapolation into positive time indices to get $\hat{x}(N), \hat{x}(N+1), \ldots$. The Fourier transform of $x'_1(n)$ is the same as that of zero-padded $x(n)$ while that of $x'_2(n)$ is the Fourier transform of time-shifted Eq. (15). According to the shift theorem the magnitude spectrum of $x'_2(n)$ is the same as that of Eq. (15) and is given by

$$|F_p(\omega)| = 1/(\mu^2 + \omega^2)^{1/2}.$$

Data extrapolation, therefore, by the exponential decay nature of the extrapolated data points, produces a smoothing effect in the frequency domain. This smoothing effect distinguishes data extrapolation by prediction from data extrapolation by zero-padding where any oscillatory behavior in the Fourier transform of $x(n)$ is left unaffected. Additionally the spectrum is sampled at $1/N_e$ intervals instead of $1/N$ for the original.

We notice that Eq. (16) could be handy if backward data extrapolation was also performed; that is, extrapolation to negative indices before $n = 0$ to $\hat{x}(-1), \hat{x}(-2), \ldots$. Utilization of both the forward and backward data extrapolation could be postulated to give better results in the frequency domain than either applied separately.

4.3 Statistical Considerations

Extensive analysis of linear prediction in respect of autoregressive processes has been carried in [25]–[28]. Since data extrapolation is based on linear prediction, the results of that analysis could be useful in explaining the effects of data extrapolation on the AR parameters and also on the spectrum estimate.

In [25] it was shown that the error of prediction of the $x(N + s)$-th sample using the most recent p observations assuming that the predictions were made from data samples independent of those used to construct estimators for parameters, is expressed as

$$(\hat{x}(N + s) - x(N + s)) \approx -\sum_{k=0}^{s-1} b_k e_{N+s-k} + (\hat{\mathbf{a}} - \mathbf{a})^T \mathbf{M}_h \mathbf{y}_n, \tag{17}$$

where $\mathbf{M}_h$ is the first $p{\times}p$ sub-matrix of $\sum_{k=0}^{s-1}((\mathbf{A}^T)^k \oplus \mathbf{A}^{s-1-k}$. The $\oplus$ represents the Kronecker delta product and is defined for matrices $\mathbf{A}^{m\times n}$ and $\mathbf{B}^{q\times t}$ by the matrix $\mathbf{D}^{mq\times nt} = \mathbf{A} \oplus \mathbf{B} = [a_{ij}\mathbf{B}]$. If $\{x(n)\}$ is a stationary AR(p) process, $\mathbf{A}$ is a $p \times p$ matrix such that $\mathbf{x}(n) = \mathbf{A}\mathbf{x}(n-1) + \mathbf{e}(n)$, where $\mathbf{x}(n) = [x(n), \ldots x(n-p+1)]^T$ and $\mathbf{e}(n) = [e(n), 0, \ldots, 0]^T$ is a $p \times 1$ zero mean constant variance white noise vector. The b_k's are the MA equivalent parameters for the AR process.

Often, as in our case, the same observation sequence is used for prediction and parameter estimation, hence Eq. (17) can be modified such that the error of the predictor can be written as [26]

$$\hat{\mathbf{X}}(N + s) - \mathbf{X}(N + s) \approx -\sum_{k=0}^{s-1} \mathbf{A}^k e_{N+s-k} - (\hat{\mathbf{A}}^s - \mathbf{A}^s)\mathbf{X}(n), \tag{18}$$

where $\mathbf{X}(n)$ is $[[\mathbf{x}(n)]^T, 1]^T$. Both Eqs. (17) and (18) depend on the bias vector $(\hat{\mathbf{a}}-\mathbf{a})$ of the AR parameters. It was shown in [27] that for the case where the input white noise is uncorrelated with the AR parameters the bias in Yule-Walker AR parameter estimates B_{AR} is

$$B_{AR} = N^{-1}\left[\mathbf{R}^{-1}\xi + \frac{\sigma_e^2 \mathbf{R}^{-1}\mathbf{1}}{\sum_{i=0}^{p} a(i)} + E_{B_{AR}}\right], \tag{19}$$

where p is assumed to be even, $\mathbf{1}$ is a $p \times 1$ vector of ones, $E_{B_{AR}}$ is a residual error of B_{AR} mainly constituted by fourth order cumulant of $e(n)$ and ξ is a $p \times 1$ vector such that $\xi_j = \sum_{k=0}^{p} |j-k| r_{xx}(j-k)a(k)$, $j = 1, \ldots p$. The log-bias of the spectral density function $\log P(\hat{f}) - \log P(f)$ in the least squares sense was shown in [28] to be related to the bias vector of the AR parameter estimates $(\hat{\mathbf{a}} - \mathbf{a})$ as follows:

$$\log P(\hat{f}) - \log P(f) = \frac{\sigma_e^4 - (\hat{\sigma_e}^2 - 2\sigma_e^2)^2}{2\sigma_e^4}$$

$$- \frac{\mathbf{a}'^T H_f(\hat{\mathbf{a}}' - \mathbf{a}')}{G(f,\mathbf{a}')} + (\hat{\mathbf{a}}' - \mathbf{a}')^T \frac{H_f \mathbf{a}' \hat{\mathbf{a}}'^T H_f - G(f,\mathbf{a}')H_f}{2G^2(f,\mathbf{a}')}(\hat{\mathbf{a}}' - \mathbf{a}') + Res(f,\mathbf{a}'), \tag{20}$$

where $\mathbf{a}' = [1, \mathbf{a}^T]^T$, $\hat{\mathbf{a}}' = [1, \hat{\mathbf{a}}^T]^T$, $G(f,\mathbf{a}') = |A(f,\mathbf{a}')|$, $A(f,\mathbf{a}') = \mathbf{a}'^T \mathbf{e}_f$, and $\mathbf{e}_f = [1, e^{jf}, \ldots, e^{jpf}]^T$. The vector H_f is defined as $H_f = \mathbf{e}_f \mathbf{e}_f^H + \mathbf{e}_f^H \mathbf{e}_f$, where $\mathbf{e}_f^H$ is the Hermitian transpose of $\mathbf{e}_f$. The quantity $Res(f,\mathbf{a}')$ is a residual term of order $O(1/N)$. Equation (20) can be simplified to

$$\begin{aligned}\log P(\hat{f}) - \log P(f) &= \mathbf{E}^2_{\sigma_e^2} - \mathbf{a}'^T H_f(\hat{\mathbf{a}}' - \mathbf{a}')/G(f,\mathbf{a}') \\ &+ B'^T_{AR} \frac{H_f \mathbf{a}' \hat{\mathbf{a}}'^T H_f - G(f,\mathbf{a}')H_f}{2G^2(f,\mathbf{a}')} B'_{AR} \\ &+ Res(f,\mathbf{a}'),\end{aligned} \tag{21}$$

where $\mathbf{E}^2_{\sigma_e^2}$ is the prediction error term and $B'_{AR} = \hat{\mathbf{a}}' - \mathbf{a}'^T$. Equations (18)–(21) show that noise-compensated data extrapolation has desired effect of reducing the bias in the spectral estimates since noise compensation minimizes $\mathbf{E}^2_{\sigma_e^2}$ in Eq. (21). As mentioned above, besides increasing the smoothness, and also compactness, of the spectrum, data extrapolation has the inherent effect of reducing the bias because of the $1/N$ term in the Eq. (19). The noise-compensated data extrapolation could, therefore, be expected to invoke effective bias reduction and resolution improvement in the frequency domain.

5. Computer Simulation

The following relationship was used for simulations

$$y(n) = \sum_{i=1}^{q} A_i \exp(j2\pi f_i n) + v(n), \tag{22}$$

where q is the number of frequencies, A_i is the amplitude of the i-th normalized frequency and $v(n)$ is zero-mean white noise of variance σ_v^2. The simulation example is clearly seen to be that of sinusoids in white noise. This signal model choice has been made based on the observation that the AR model interpretation of linear prediction gives high resolution in the case of sinusoids in white noise and can be used as a good benchmark for performance evaluation [29]. The sinusoidal signal model commonly arises in frequency estimation problems where the nature of the signal is known beforehand. However since the poles of AR model for sinusoids are located on the unit circle, the locations of the largest poles can be used to estimate the frequencies. In the noiseless case, poles on the unit circle corresponds sharp peaks in the spectrum while additional spurious peaks arise in the noisy case. Thus, the spectrum of sinusoids can be analyzed by an AR model. Since we are also considering short data sequences with closely spaced spectral peaks at LSNR's, this signal model choice could be useful for illustration purposes. Furthermore, with little information about the nature of the signal under investigation, it could be practical to use the AR model for a given data sequence.

The settings in Eq. (22) were: $q = 2$, $f_1 = 0.20$, $f_2 = 0.22$, and $A_1 = A_2 = 1$. The data sequence of length 40 was generated and used in the simulations. The AR order p for parameter estimation was set to 18. The order was obtained by using the approximation $p = (N/3 + N/2)/2$ [6], where N is the length of the observed data sequence, and p was set to be even. The constant M for the INVE method was set to 1000. For the noise-compensated extrapolation of the proposed method, the data was extended to two times the original length. The SNR, given by the relationship $SNR = 10\log(P_i/\sigma_v^2)$, where $P_i = |A_i|^2$, was varied from -10 dB to 30 dB. A complex-valued sequence has be chosen since it is more general and simpler to work with in the frequency domain. The AR parameters will accordingly be complex although strictly speaking, it would not be a disadvantage.

In Figs. 3–6, the rapid decrease in the amplitude of the extrapolated data points due to the use of noisy AR parameters is illustrated. The effect is more severe at LSNR's. With noise-compensated AR parameters it is possible to obtain a much longer non-zero sequence as shown in Fig. 6.

As shown in Figs. 7 and 8, the effect of noise on the spectrum estimate is more severe with MUSIC than DFT to

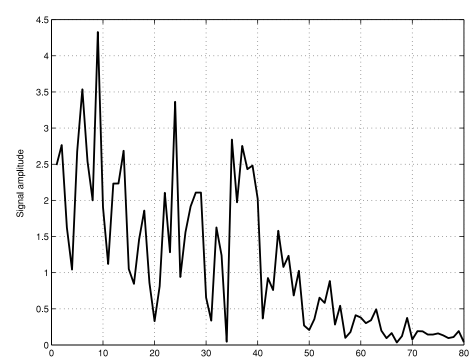

Fig. 3 Extrapolation of a noise-corrupted data sequence with no compensation at SNR of 5 dB. The original data length is 40.

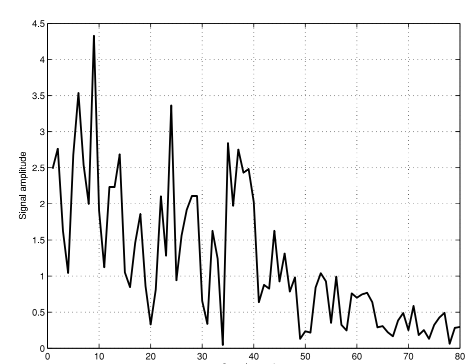

Fig. 4 Extrapolation of a noise-corrupted data sequence with compensation at SNR of 5 dB. The original data length is 40.

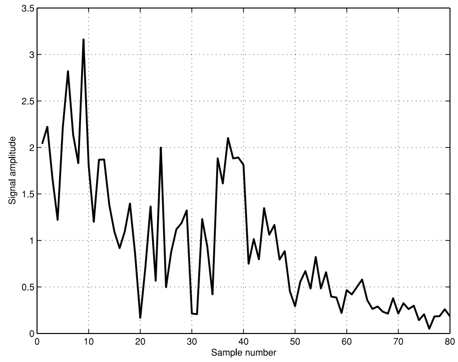

Fig. 5 Extrapolation of a noise-corrupted data sequence with no compensation at SNR of 0 dB. The original data length is 40.

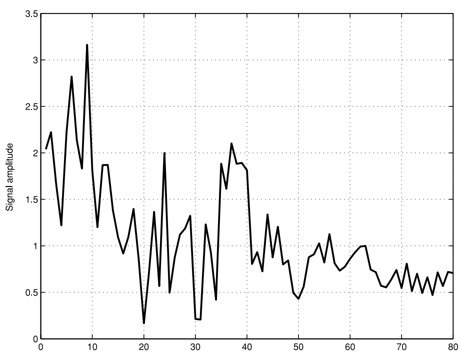

Fig. 6 Extrapolation of a noise-corrupted data sequence with compensation at SNR of 0 dB. The original data length is 40.

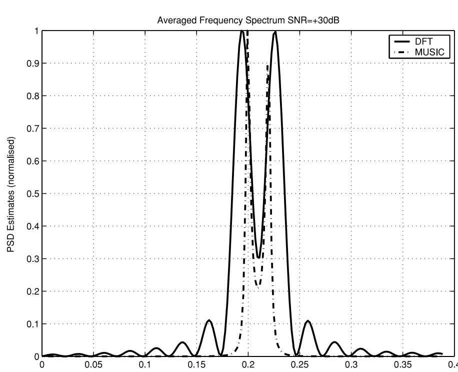

Fig. 7 Spectrum estimate at HSNR (30 dB) for both DFT and MUSIC by averaging 100 independent runs.

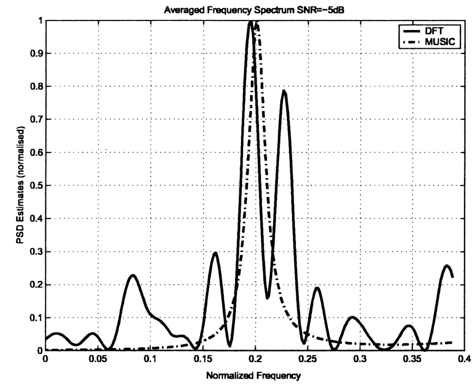

Fig. 8 Spectrum estimate at LSNR (−5 dB) for both DFT and MUSIC by averaging 100 independent runs.

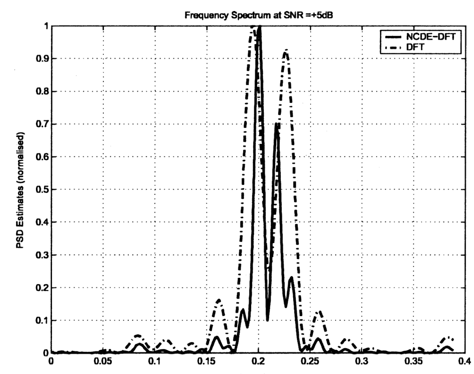

Fig. 9 Spectrum estimate by the proposed method and DFT at relatively HSNR (5 dB) by averaging 100 independent runs. NCDE-DFT denotes the noise-compensated data extrapolation and FFT processing.

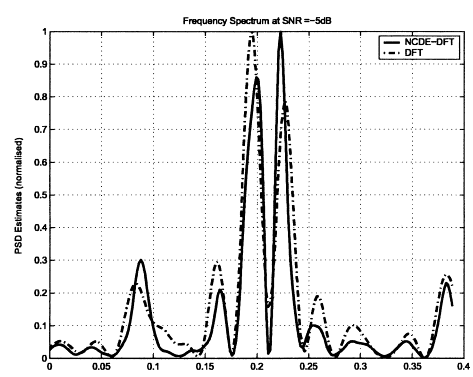

Fig. 10 Spectrum estimate by the proposed method and DFT at LSNR (−5 dB) by averaging 100 independent runs. NCDE-DFT denotes the noise-compensated data extrapolation and FFT processing.

the extent that the two frequencies can be hardly resolved at −5 dB by MUSIC. Both methods give a good estimate as shown by position of peaks. On the other hand, there is a general degradation in the estimates obtained from DFT and an evident increase in side-lobe power. Decreasing the SNR renders the MUSIC method ineffective but the DFT method still shows the presence of two frequency components despite an increased bias.

Figures 9 and 10, and Table 1, show that the noise-compensated data extrapolation method invokes the best estimate for the two frequencies in the given range of SNR. The results also show a considerable decrease in bias by the proposed method.

Figures 11 and 12 compare the use of the proposed noise compensation data extrapolation method to the direct use of YWE's. It can be seen that the proposed method can result in better amplitude response and reduction in bias than using YWE's only. The suppression of the power of spuri-

Table 1 Estimates of the two frequencies at −5 dB, 0 dB, and +5 dB for DFT, NCDE-DFT and MUSIC by averaging 100 independent runs.

SNR	DFT		NCDE-DFT		MUSIC	
dB	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_1$	$\hat{f}_2$
−5	0.1973	0.2207	0.1992	0.2207	0.1992	0.2031
0	0.1973	0.2207	0.1992	0.2207	0.1992	0.2051
+5	0.1973	0.2207	0.2012	0.2207	0.1992	0.2070

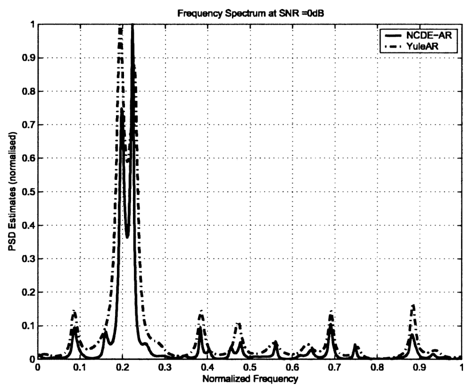

Fig. 11 Spectrum estimate by the proposed method and Yule-Walker AR at 0 dB by averaging 100 independent runs. NCDE-AR denotes the noise-compensated data extrapolation and Yule-Walker AR processing.

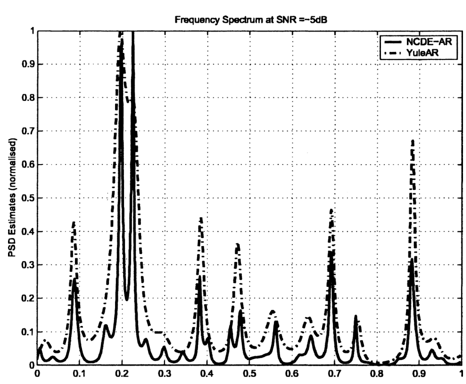

Fig. 12 Spectrum estimate by the proposed method and Yule-Walker AR at 5 dB by averaging 100 independent runs. NCDE-AR denotes the noise-compensated data extrapolation and Yule-Walker AR processing.

Table 2 Estimates of the two frequencies at −5 dB, 0 dB, and +5 dB for Yule-Walker AR and NCDE-AR by averaging 100 independent runs.

SNR	YuleAR		NCDE-AR	
dB	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_1$	$\hat{f}_2$
−5	0.1933	0.2266	0.1973	0.2246
0	0.1934	0.2266	0.1973	0.2226
+5	0.1953	0.2266	0.2012	0.2226

ous peaks by the proposed method is also quite significant. Table 2 shows some results of frequency estimates, which suggest the superiority of the proposed method.

The proposed method also performs better than the DFT and Yule-AR at a very HSNR as shown by peak locations in Figs. 13 and 14. Table 3 shows less bias in the estimate of f_1 by the proposed method.

A comparison of the proposed method with the Burg

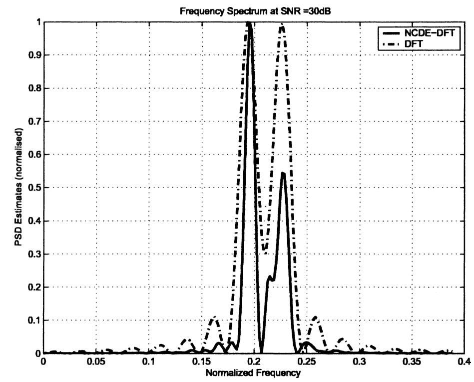

Fig. 13 Spectrum estimate by the proposed method and DFT at very HSNR (30 dB) by averaging 100 independent runs. NCDE-DFT denotes the noise-compensated data extrapolation and FFT processing.

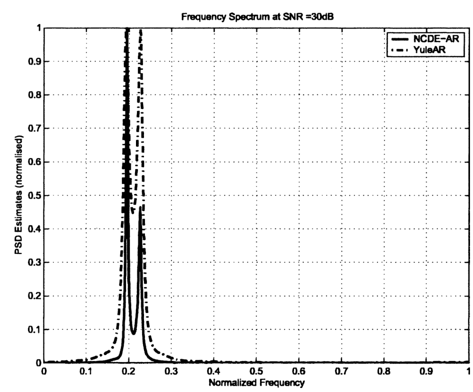

Fig. 14 Spectrum estimate by the proposed method and Yule-Walker AR at very HSNR (30 dB) by averaging 100 independent runs. NCDE-AR denotes the noise-compensated data extrapolation and Yule-Walker AR processing.

Table 3 Estimates of the two frequencies at +30 dB by the DFT, the Yule-Walker AR and NCDE-AR and by averaging 100 independent runs.

SNR	30 dB	
f_i	$\hat{f}_1$	$\hat{f}_2$
DFT	0.1934	0.2266
YuleAR	0.1934	0.2266
NCDE − AR	0.1953	0.2266

method was also done at SNR's of −5 dB and −10 dB as shown in Figs. 15 and 16. For the Burg method, only the original data sequence of length 40 was used. The proposed method appears to give better amplitude response at −5 dB and also still distinguishes two peaks at a SNR of −10 dB. Table 4 gives some estimates from the Burg method and the proposed method. A comparison of the proposed method and the Burg method was also made with the data length made equal to that of the extrapolated data sequence, that is, 80 samples. The frequency estimates are shown in Table 5.

Comparing the performance of the proposed method in Table 4 to that of the Burg method in Table 5, we see that the Burg method is better than the proposed method, getting a longer length of the data sequence. However, if the same length of the data sequence is used for the proposed method,

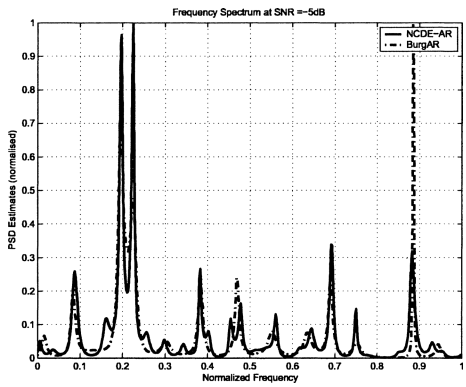

Fig. 15 Spectrum estimate by the proposed method and the Burg method at −5 dB by averaging 100 independent runs.

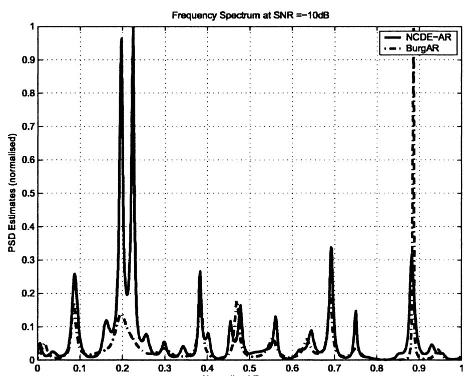

Fig. 16 Spectrum estimate by the proposed method and the Burg method at −10 dB by averaging 100 independent runs.

Table 4 Estimates of the two frequencies at −5 dB, 0 dB, and +5 dB for Burg AR and NCDE-AR using the original data sequence of length 40, by averaging 100 independent runs.

SNR	Burg AR		NCDE-AR	
dB	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_1$	$\hat{f}_2$
−5	0.1933	0.2266	0.1953	0.2246
0	0.1934	0.2266	0.1973	0.2246
+5	0.1953	0.2266	0.1973	0.2246

Table 5 Estimates of the two frequencies at −5 dB, 0 dB, and +5 dB for Burg AR and NCDE-AR with a data sequence of length 80, by averaging 100 independent runs.

SNR	Burg AR		NCDE-AR	
dB	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_1$	$\hat{f}_2$
−5	0.1973	0.2246	0.2148	0.2148
0	0.1973	0.2227	0.2012	0.2207
+5	0.2012	0.2227	0.1992	0.2207

then the proposed method provides an improvement.

6. Concluding Remarks

In this paper we have addressed the problem of spectrum estimation from noisy observations using the noise-compensated data extrapolation method. The noise variance was obtained through an iterative algorithm while extrapolation was effected by linear prediction. An insight into the time and frequency domain implications of noise-compensated data extrapolation and some statistical implications were also given. The proposed method was shown to be more noise-robust compared with popular methods which include plain DFT, Yule-Walker AR and MUSIC. In accordance with analytical results, the proposed method performed better with respect to bias, resolution and spurious peak reduction in the frequency domain.

Acknowledgment

The authors are very grateful to the reviewers of the journal for their kind suggestion and advice.

References

[1] S.M. Kay, Modern Spectral Estimation: Theory and Application, Prentice Hall, Englewood Cliffs, NJ, 1988.
[2] S.V. Vaseghi, Advanced Signal Processing and Digital Noise Reduction, John Wiley and Teubner, England, 1996.
[3] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel, Time Series Analysis: Forecasting and Control, Prentice Hall, Englewood Cliffs, NJ, 1994.
[4] S.M. Kay, "Noise compensation for autoregressive spectral estimates," IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-28, no.3, pp.292–303, June 1980.
[5] L. Marple, "High-resolution autoregressive spectrum analysis using noise power cancellation," Proc. IEEE International Conf. Acoustics, Speech, and Signal Processing, pp.345–348, Oklahoma, USA, April 1978.
[6] T.E. Landers and R.T. Lacoss, "Some geophysical applications of autoregressive spectral estimates," IEEE Trans. Geosci. Electron., vol.GE-15, no.1, pp.26–32, Jan. 1977.
[7] V.F. Pisarenko, "The retrieval of harmonics from a covariance function," Geophys. J. Roy. Astron, vol.33, pp.347–366, Jan. 1973.
[8] T.J. Ulrych and M. Ooe, "Autoregressive and mixed autoregressive-moving average and spectra," in Topics in Applied Physics, vol.34, Nonlinear Methods for Spectral Analysis, ed. S. Haykin, pp.73–125, Springer-Verlag, Berlin, 1983.
[9] C.E. Davila, "A subspace approach to estimation of autoregressive parameters from noisy measurements," IEEE Trans. Signal Process., vol.46, no.2, pp.531–534, Feb. 1998.
[10] H.-T. Hu, "Linear prediction analysis of speech signals in the presence of white Gaussian noise with unknown variance," IEE Proc. Vis. Image Signal Process., vol.145, no.4, pp.303–308, 1998.
[11] W.X. Zheng, "Estimation of autoregressive signals from noisy measurements," IEE Proc. Vis. Image Signal Process, vol.144, no.1, pp.39–45, 1997.
[12] C.E. Davila, "On the noise-compensated Yule-Walker equations," IEEE Trans. Signal Process., vol.49, no.6, pp.1119–1121, June 2001.
[13] R. Schmidt, "Multiple emitter location and signal parameter estimation," Proc. RADC Spectrum Workshop, pp.243–258, 1979.
[14] T. Shimamura, "Two-dimensional spectral estimation method using data extrapolation," Proc. Joint Technical Conf. on Circuits/Systems, Computers and Commun., no.JTC-CSCC'93, pp.189–194, Nara, Japan, July 1993.
[15] T. Shimamura, T. Nagamine, and J. Suzuki, "High-resolution two-dimensional spectral estimation utilizing data extrapolation," Trans. ISCIE (Japanese Edition), vol.7, no.6, pp.213–220, June 1994.
[16] T. Shimamura, W. Miao, and J. Suzuki, "Two-dimensional spectral estimation method with data extension and its improvement," IEICE Trans. Fundamentals (Japanese Edition), vol.J78-A, no.8, pp.965–976, Aug. 1995.

[17] J. Gamba and T. Shimamura, "An iterative approach to spectrum estimation of noisy data by compensation," Proc. International Symposium on Communications and Information Technology, pp.73–76, Pattaya, Thailand, Oct. 2002.

[18] J. Gamba and T. Shimamura, "Two-dimensional spectral estimation with noise-compensated data extrapolation," Proc. International Conf. on Fundamentals of Electronics, Communications and Computer Science, pp.20–17–20-22, Tokyo, Japan, March 2002.

[19] J. Gamba and T. Shimamura, "A new method of noise variance estimation from low-order Yule-Walker equations," IEICE Trans. Fundamentals, vol.E87-A, no.1, pp.270–274, Jan. 2004.

[20] G.H. Golub and C.F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore and London, 1989.

[21] G.M. Phillips and P.J. Taylor, Theory and Applications of Numerical Analysis, Academic Press, London, 1996.

[22] R.N. Bracewell, The Fourier Transform and Its Applications, McGraw-Hill, New York, 1999.

[23] I.S. Gradshteyn and I.M. Ryzhik, Table of Integrals, Series and Products: Corrected and Enlarged Edition, Academic Press, 1992.

[24] P.A.P. Moran, "Some experiments on the prediction of sunspot numbers," J. Royal Statistical Society, Series B (Methodological), vol.16, no.1, pp.112–117, 1954.

[25] T. Yamamoto, "Asymptotic mean square prediction error for an autoregressive model with estimated coefficients," J. Applied Statistics, vol.25, no.2, pp.123–127, 1976.

[26] W.A. Fuller and D.P. Hasza, "Properties of predictors for autoregressive time series," J. American Statistical Association, vol.76, no.373, pp.155–161, March 1981.

[27] P. Shaman and R.A. Stine, "The bias of autoregressive coefficient estimators," J. American Statistical Association, vol.83, no.403, pp.842–848, Sept. 1988.

[28] R.A. Stine and P. Shaman, "Bias of autoregressive spectral estimators," J. American Statistical Association, vol.85, no.412, pp.1091–1098, Dec. 1990.

[29] C.W. Therrien, Discrete Random Signals and Statistical Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1992.

Appendix: Time Domain Effect

Given the observed data sequence $x(0), x(1), \ldots, x(N-1)$, we can extrapolate the sequence using prediction coefficients a_i (with a slight change in notation for simplicity) as follows.

$$\begin{aligned}\hat{x}(N) &= \sum_{i=1}^{p} a_i x(N-i)\\ &= a_1 x(N-1) + \ldots + a_p x(N-p) \qquad \text{(A· 1)}\\ \hat{x}(N+1) &= \sum_{i=1}^{p} a_i x(N+1-i)\\ &= a_1 x(N) + \sum_{i=2}^{p} a_i x(N+1-i)\\ &= a_1 x(N) + \alpha_1\end{aligned}$$

where $\alpha_1 = \sum_{i=2}^{p} a_i x(N+1-i)$. The $\hat{x}(N+2)$ and the $\hat{x}(N+3)$ can be found as

$$\begin{aligned}\hat{x}(N+2) &= \sum_{i=1}^{p} a_i x(N+2-i)\\ &= a_1 x(N+1) + a_2 x(N)\\ &\quad + \sum_{i=3}^{p} a_i x(N+2-i)\\ &= a_1 x(N+1) + a_2 x(N) + \alpha_2\\ \hat{x}(N+3) &= \sum_{i=1}^{p} a_i x(N+3-i)\\ &= a_1 x(N+2) + a_2 x(N+1)\\ &\quad + a_3 x(N) + \sum_{i=4}^{p} a_i x(N+3-i)\\ &= a_1 x(N+2) + a_2 x(N+1) + a_3 x(N) + \alpha_3\end{aligned}$$

In this manner the extrapolation process can be continued. Assuming that a_1 is the dominant coefficient and that cross-products between coefficients are small, Eq. (A· 1) can be expressed as

$$\hat{x}(N) = \alpha_0 + \alpha \qquad \text{(A· 2)}$$

where $\alpha_0 = a_1 x(N-1)$ and $\alpha = a_2 x(N-2) + \ldots + a_p x(N-p)$. By substituting the values of $x(N), x(N+1), \ldots$ by their estimates $\hat{x}(N), \hat{x}(N+1), \ldots$, the approximate values of the extrapolated data samples can be expressed in terms of the original data more compactly as

$$\begin{aligned}&\hat{x}(N) \approx \alpha_0 + Res(0)\\ &\hat{x}(N+1) \approx a_1\alpha_0 + Res(1)\\ &\hat{x}(N+2) \approx a_1^2\alpha_0 + Res(2)\\ &\hat{x}(N+3) \approx a_1^3\alpha_0 + Res(3)\\ &\hat{x}(N+4) \approx a_1^4\alpha_0 + Res(4)\\ &\cdots\\ &\hat{x}(N+i) \approx a_1^i\alpha_0 + Res(i)\\ &\cdots\end{aligned} \qquad \text{(A· 3)}$$

where $Res(\cdot)$ is the residual term at each point in extrapolation. Ignoring the residual terms, the $(i+1)$th extrapolated extrapolated data can be expressed as

$$\hat{x}(N+i) \approx \alpha_0 a_1^i. \qquad \text{(A· 4)}$$

If we assume that $a_1 > 0$ and real, there is a number c for which $a_1^i = e^{ci}$. If furthermore, $0 < a_1 < 1$, then $\hat{x}(N+i)$ can be written as

$$\hat{x}(N+i) \approx \alpha_0 e^{-\beta i}. \qquad \text{(A· 5)}$$

The result confirms Eq. (9). Equation (A· 5) can be further simplified to

$$\hat{x}(N+i) \approx x(N-1)e^{-\beta(i+1)}. \qquad \text{(A· 6)}$$

This show that in this case the extrapolated samples depend strongly on the value of the last sample in the data.

Jonah Gamba received BSc. (Hons) degree in electrical and electronics engineering from the University of Zimbabwe, Harare, Zimbabwe, in 1994 and an MSc. degree in computer science and engineering from Zhejiang University, Hangzhou, China, in 2000. He is currently with Saitama University, Saitama City, Japan, pursuing a doctor degree course in mathematical information systems. His present research interests include spectrum estimation, time delay estimation and adaptive channel equalization.

Tetsuya Shimamura received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1986, 1988, and 1991, respectively. In 1991, he joined Saitama University, Saitama City, Japan, where he is currently an Associate Professor. His interests are in digital signal processing and applications to speech and communication systems. He is a member of IEEE and EURASIP.