Estimation of Autoregressive Parameters from Noisy Observations Using Iterated Covariance Updates

Todd K. Moon * and **Jacob H. Gunther**

Electrical and Computer Engineering Department, Utah State University, Logan, UT 84332, USA; jake.gunther@usu.edu
* Correspondence: todd.moon@usu.edu

Abstract: Estimating the parameters of the autoregressive (AR) random process is a problem that has been well-studied. In many applications, only noisy measurements of AR process are available. The effect of the additive noise is that the system can be modeled as an AR model with colored noise, even when the measurement noise is white, where the correlation matrix depends on the AR parameters. Because of the correlation, it is expedient to compute using multiple stacked observations. Performing a weighted least-squares estimation of the AR parameters using an inverse covariance weighting can provide significantly better parameter estimates, with improvement increasing with the stack depth. The estimation algorithm is essentially a vector RLS adaptive filter, with time-varying covariance matrix. Different ways of estimating the unknown covariance are presented, as well as a method to estimate the variances of the AR and observation noise. The notation is extended to vector autoregressive (VAR) processes. Simulation results demonstrate performance improvements in coefficient error and in spectrum estimation.

Keywords: autoregressive model estimation; spectrum estimation; vector AR model; RLS algorithm

1. Introduction

The problem of estimating the parameters $\{b_i, i = 1, \ldots, p\}$ of an autoregressive (AR) process

$$\xi(m) = \sum_{i=1}^{p} b_i \xi(m-i) + \eta(m), \tag{1}$$

where the input $\eta(n)$ is a white noise process, is important in many aspects of signal processing. It plays roles a variety of applications, such as speech coding and analysis (see References [1–7]). Autoregressive modeling is instrumental in many spectrum estimation algorithms, and algorithms for noise-free measurements have been developed from that perspective [8–14]; see also Reference [5] for a survey and perspective. Vector autoregressive modeling is also important in econometric modeling [15]. Among many other applications, AR models have also been used in biomedical signal processing (see, e.g., Reference [16]); communication (see, e.g., Reference [17]); and financial modeling (see, e.g., Reference [15]).

Because of its importance, the AR parameter estimation problem has been well-studied from a realization of noise-free process $\{\xi(m)\}$ When autocorrelations $r_k = E[\xi(m)\xi^*(m - k)]$ are known (or estimated), the Yule-Walker equations describe the solutions [18,19]. This problem may also be considered as the problem of finding optimal predictor coefficients, and solutions invoking adaptive filters are well known [20]. The parameters can also be estimated using a Kalman filter applied to a statespace

formulation ([21], Section 3.3), in which the unknown parameters are taken to be the state, and the observation matrix is from the measurements. The Kalman filter method bears resemblance to the approach presented here, but as will be shown there are differences between the two. First, in the case of noisy observations, the previous AR values $\xi(m-i), i = 1, \ldots, p$ are not known, so the elements of observation matrix are only approximately known in the Kalman filter approach. No such approximation is used in the method here. Second, the approach presented here does not require the introduction of the system noise with the question of what the variance should be. And thirdly, instead of a scalar observation equation, vector observations are used, allowing the covariance structure to be used with a time-varying covariance.

It is well known ([5], Section 6.7), [22,23] that additive noise broadens spectral peaks in autoregressive spectral estimation and can result in loss of resolution of closely-spaced sinusoids. Dealing with noise added to AR processes has been addressed many times in the literature, as we summarize below. None of these previous methods use the covariance structure employed here. Some of these techniques are related to finding rank p representations of a data matrix (using, for example, an SVD or total least squares) [24–30]. From another perspective, it is known that white noise added to an $AR(p)$ process produces an $ARMA(p, p)$ process. This has resulted in algorithms for dealing with the noisy AR process using an ARMA estimator. For example, modified Yule-Walker (YW) equations [31] can be used to find the ARMA model. However, this does not take advantage of the relationship between the AR and the MA parameters in this noisy AR case. Another approach is to estimate the ARMA process using a large-order AR model. Another approach to handling noisy AR observations is total least squares (TLS) [32–35]. None of these approaches is equivalent to the approach described here. Noisy AR estimation has been studied over many years by Zheng [36–55], who estimates the variance of the observation noise to reduce bias. An approach that is somewhat similar to our approach is bias-corrected RLC (BCRLS) [56–61], but the BRCLS does not employ the covariance matrix-weighted least squares used here. Other approaches to noisy AR parameter estimation have been investigated in References [4,62–92]

The approach described here uses the fact that the observation noise produces an AR-like process with correlated noise, where the noise correlation depends on the AR coefficients. The correlation is exploited by dealing with vectors of stacked observations, rather than scalar observations. A maximum likelihood approach results in a weighted least-squares problem, in which previous estimates are used to update the covariance matrix in an iterative fashion. We refer to this general algorithm as iterative covariance-weighted autoregressive estimation (ICWARE). While maximum likelihood has been used for AR parameter estimation [18] ([93], Chapter 5), [94], in previous estimators the noise is assumed white, so the autocorrelation structure derived here is not exploited. Iterative parameter estimation has been used for system parameter estimation. An early iterative method of fitting a given impulse response to a transfer function, without an explicit noise model, is in Reference [95]. Iterative quadratic ML (IQML) [96] is another iterative technique. IQML is similar to the algorithm of Reference [97] , which requires knowledge of the input signal. IQML is also similar to the method of Reference [98]. An IQML approach [99] has been used for noisy AR estimation. None of these make use of the covariance weighting of ICWARE.

There is some similarity of this approach to the measurement error models studied in the statistical literature [100], in which parameters of regression models such as $Y_t = \beta_0 + \beta_1 x_t + e_t$ are estimated when only noisy observations of the regressors $X_t = x_t + u_t$ are available. These noisy observations induce changes to conventional minimum mean-square error (MMSE) estimators which are somewhat related to errors-in-variables approaches to AR estimation [101–103].

Another approach to this problem makes use of coupled Kalman filters [104] or cross-coupled H_∞ filters [105,106] has been developed. In essence these move beyond the Kalman filter approach referenced above ([21], Section 3.3) and operate as follows. One Kalman filter estimates the true autoregressive value $\hat{\xi}(m)$, assuming knowledge of the AR parameters $\{b_i\}$, while the other Kalman filter estimates the $\{b_i\}$

using the estimates $\hat{\zeta}(m)$ as if they were the true values. These two filters operate in conjunction to jointly converge to the AR parameter estimate. To evaluate our method in comparison with the dual Kalman filter method, the algorithm here is compare to models also seen in Reference [104].

Recently, Monte Carlo methods, such as particle filtering has been used for AR and ARMA estimation [107–110]. Particle filtering methods are quite general and can track dynamical variables and jointly estimate static parameters. These methods offer the possibility of convergence to good estimates in situations when the noise is not Gaussian (which has been assumed in this paper), but offer the usual drawbacks of potentially high computational complexity if many particles are used. These represent an alternative to the methods of this paper, and a thorough comparison, while valuable, lies beyond the scope of this paper.

This topic is relevant to its journal of publication (*Entropy*), since spectral analysis under a maximum entropy criterion has (famously) been shown to be equivalent to AR modeling [11]. As this work shows, however, when there are noisy observations particular attention must be devoted to the information present in the autocorrelation matrix of the observations, beyond the first order equations that result from the historical maximum entropy approach.

We present the method first for the scalar AR model. For generality, noise is assumed to be circular complex Gaussian noise; modifications for real signals is straightforward. In addition to the coefficient estimation problem, we also address the problem of variance estimation. Following development for scalar AR processes, the notation is developed for the vector AR (VAR) processes. Several simulation results demonstrate that the ICWARE method provide significant improvements over classical YW-type methods.

2. Scalar Parameter Estimation in Noise

The noisy observation equation, represented by the system in Figure 1, is

$$y(m) = \xi(m) + v(m) = \sum_{i=1}^{p} b_i \xi(m-i) + \eta(m) + v(m), \tag{2}$$

where $\xi(m) \in \mathbb{C}$ and each $b_i \in \mathbb{C}$, and where $v(m)$ is a white noise process. The process $\{\xi(m)\}$ is said to be the noise-free AR process, and the process $\{y(m)\}$ is said to be the noisy AR process.
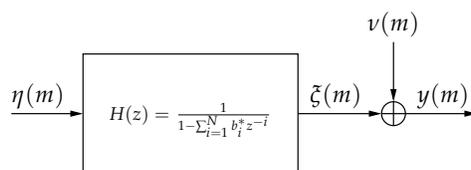
Figure 1. Noisy Autoregressive Model.

The input and observation noises are assumed to be complex circular Gaussian, so that

$$\eta(m) = \sigma_\eta(\mathcal{N}(0,1) + j\mathcal{N}(0,1)) \quad \text{and} \quad v(m) = \sigma_v(\mathcal{N}(0,1) + j\mathcal{N}(0,1)),$$

where the real and imaginary parts are uncorrelated, and draws at different times are independent. Furthermore, these noises are assumed to be uncorrelated with each other, so

$$E[v(m)v(m-l)^*] = 2\sigma_v^2 \delta_l \qquad E[\eta(m)\eta(m-l)^*] = 2\sigma_\eta^2 \delta_l \qquad E[v(m)\eta(n^*)] = 0.$$

Let $\mathbf{b} = \begin{bmatrix} b_1 & b_2 & \cdots & b_p \end{bmatrix}^T$ and $\mathbf{y}(i) = \begin{bmatrix} y(i) & y(i-1) & \cdots & y(i-p+1) \end{bmatrix}^T$ (this vector is re-defined below to have a different length). The conventional least-squares approach to estimating $\mathbf{b}$ (in the forward prediction error sense) would be to compute

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}} \sum_i \left| y(i) - \mathbf{b}^T \mathbf{y}(i-1) \right|^2, \tag{3}$$

which results in the conventional estimate

$$\hat{\mathbf{b}}^* = \left(\sum_i \mathbf{y}(i-1)\mathbf{y}(i-1)^H \right)^{-1} \left(\sum_i y^*(i)\mathbf{y}(i-1) \right).$$

This is essentially the covariance method [19]. The estimate (3) can be computed using recursive formulation, resulting in a recursive least-squares (RLS) adaptive filter. This approach essentially neglects the noise $v(m)$. It is known that the noise will bias the estimate results [22].

Note: The error could also be computed in a backward prediction error sense, or in a combined forward-backward sense as is done in some spectrum estimation algorithms such as the modified covariance method [14], ([5], Section 75). The ICWARE method we describe below can be extended to these generalizations as well, but for brevity only forward prediction error is used.

Substituting the AR signal in terms of the observation $\xi(m) = y(m) - v(m)$ into the AR model (1) and using the observation model (2) we obtain

$$y(m) = \sum_{i=1}^{p} b_i y(m-i) + \left(\eta(m) + v(m) - \sum_{i=1}^{p} b_i v(m-i) \right) \tag{4}$$

$$= \sum_{i=1}^{p} b_i y(m-i) + w(m), \tag{5}$$

where

$$w(m) = \left(\eta(m) + v(m) - \sum_{i=1}^{p} b_i v(m-i) \right)$$

denotes the noise. The expression (4) is an $ARMA(p, p)$ process and is the basis for some ARMA-based approaches to AR estimation in noise. The expression (5) has the appearance of an $AR(p)$ process, except that the noise $w(m)$ is not white, having correlation

$$E[w(m)w^*(m)] = E\left[\left(\eta(m) + v(m) - \sum_{i=1}^{p} b_i v(m-i) \right) \left(\eta(m) + v(m) - \sum_{i=1}^{p} b_i v(m-i) \right)^* \right]$$

$$= 2(\sigma_\eta^2 + \sigma_v^2 + \sum_{i=1}^{p} |b_i|^2 \sigma_v^2) \triangleq r_w(0)$$

$$E[w(m)w(m+\ell)^*] = E\left[\left(\eta(m) + v(m) - \sum_{i=1}^{p} b_i v(m-i) \right) \left(\eta(m+\ell) + v(m+\ell) - \sum_{i=1}^{p} b_i v(m+\ell-i) \right)^* \right]$$

$$= 2\sigma_v^2 (\sum_{i=1}^{p-\ell} b_i b_{i+\ell}^* - b_\ell^*) \triangleq r_w(-\ell). \tag{6}$$

If the noise $w(n)$ were uncorrelated, the solution of (3) would be least-squares optimal. However, since $w(m)$ is correlated, the sample-by-sample approach of (3) is suboptimal since it does not take the correlation into account. As we now show, there is information in the covariance structure of the signal

that can be used to improve the estimate. To take the correlation into account, stack the observations of (5) into vectors of length d (that is, the depth), as

$$
\begin{bmatrix} y(m) \\ y(m-1) \\ \vdots \\ y(m-d+1) \end{bmatrix} = \begin{bmatrix} y(m-1) & \cdots & y(m-p) \\ y(m-2) & \cdots & y(m-1-p) \\ \vdots & & \\ y(m-d) & \cdots & y(m-d-p+1) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} + \begin{bmatrix} w(m) \\ w(m-1) \\ \vdots \\ w(m-d+1) \end{bmatrix}. \tag{7}
$$

Write this as

$$
\mathbf{y}(m) = Y(m)\mathbf{b} + \mathbf{w}(m)
$$

The correlation matrix of the vector noise is

$$
E[\mathbf{w}(m)\mathbf{w}(m)^H] =
$$
$$
\begin{bmatrix} r_w(0) & r_w(-1) & \cdots & r_w(-(d-1)) \\ r_w(1) & r_w(0) & \cdots & r_w(-(d-2)) \\ \vdots & & & \\ r_w(d-1) & r_w(d-2) & \cdots & r_w(0) \end{bmatrix} \triangleq R_w(\mathbf{b}). \tag{8}
$$

In moving through a sequence of data, the data can be advanced by a skip s to form a sequence of vectors $\mathbf{y}(m), \mathbf{y}(m+s), \mathbf{y}(m+2s), \ldots, \mathbf{y}(m+s(k-1))$ and matrices $Y(m), Y(m+s), Y(m+2s), \ldots, Y(m+s(k-1))$, which we write for convenience as $\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1}$, and $Y_0, Y_1, \ldots, Y_{k-1}$, respectively. The index m is chosen to make it possible to use only causal data, $m \geq d + p - 1$.

Assuming independence (such as when $s \geq d$) and assuming that the correlation function $R_w(\mathbf{b})$ is known, the likelihood function can be written as a circular complex Gaussian as

$$
f(\mathbf{y}_0, \ldots, \mathbf{y}_{k-1} | \mathbf{b}, R_w(\mathbf{b})) \propto \frac{1}{|\det(R_w(\mathbf{b}))|^k} \exp[-\sum_{i=0}^{k-1} (\mathbf{y}_i - Y_i \mathbf{b})^H R_w(\mathbf{b})^{-1} (\mathbf{y}_i - Y_i \mathbf{b})]. \tag{9}
$$

Finding the true maximum likelihood solution by directly maximizing (9) with respect to $\mathbf{b}$ is difficult, since $\mathbf{b}$ enters nonlinearly in $R_w(\mathbf{b})$. Instead, an iterative approach is employed. An estimated value $R_w(\mathbf{b})_i$ is used at the ith step. Assuming still that each $R_w(\mathbf{b})_i$ is known, a maximum likelihood solution may be obtained from

$$
\hat{\mathbf{b}}_k = \arg\min_{\mathbf{b}} \sum_{i=1}^{k} \lambda^{k-i} \|\mathbf{y}_i - Y_i \mathbf{b}\|^2_{R_w(\mathbf{b})_i^{-1}}. \tag{10}
$$

Here, a forgetting factor λ^{k-i}, with $\lambda < 1$, has been introduced to allow for tracking a time-varying $\mathbf{b}$. Significantly, the norm here is weighted by the inverse covariance $R_w(\mathbf{b})_i^{-1}$. If we take $d = 1$, we obtain the regular least-squares problem, and the additional correlation structure does not appear. Taking the gradient of the cost functional (10) results in the normal equation

$$
\left(\sum_{i=1}^{k} \lambda^{k-i} Y_i^H R_w(\mathbf{b})_i^{-1} Y_i \right) \hat{\mathbf{b}}_k = \left(\sum_{i=1}^{k} \lambda^{k-i} Y_i^H R_w(\mathbf{b})_i^{-1} \mathbf{y}_i \right). \tag{11}
$$

Write this as $\Phi_k^{-1} \hat{\mathbf{b}}_k = \boldsymbol{\phi}_k$. This can be recursively updated using RLS recursions (see, e.g., References ([111], Section 8.7), [112]), starting from an initial $\Phi_{-1} = \frac{1}{\delta} I$ for a small scalar δ and propagating Φ_k.

Computing (11) requires knowledge of $R_w(\mathbf{b})_i$, which is not available since $\mathbf{b}$ is to be found. In combination with a recursively updated solution to (11), $R_w(\mathbf{b})_i$ is estimated using previously computed

values of $\hat{\mathbf{b}}$ and used as a time-varying covariance matrix. That is, we write $R_w(\mathbf{b})_i = R_w(\hat{\mathbf{b}}_{i-1})_i$. The result is the framework shown in Algorithm 1, similar to a vector RLS adaptive filter.

Algorithm 1: AR Parameter estimation with noisy data.

Input: $\mathbf{y}_k$, Y_k, $R_w(\hat{\mathbf{b}}_{k-1})_k$
Previous Conditions: Φ_{k-1}, $\hat{\mathbf{b}}_{k-1}$
Compute

$$K_k = \lambda^{-1}\Phi_{k-1}Y_k^H(\lambda^{-1}Y_k\Phi_{k-1}Y_k^H + R_w(\hat{\mathbf{b}}_{k-1}))^{-1}$$
$$\hat{\mathbf{b}}_k = \hat{\mathbf{b}}_{k-1} + K_k(\mathbf{y}_k - Y_k\hat{\mathbf{b}}_{k-1})$$
$$\Phi_k = \lambda^{-1}(I - \lambda^{-1}K_kY_k)\Phi_{k-1}.$$

Update the covariance to obtain $R_w(\hat{\mathbf{b}}_k)_{k+1}$ and compute $R_w(\hat{\mathbf{b}}_k)_{k+1}^{-1}$.
Return $\hat{\mathbf{b}}_k$, Φ_k, $R(\hat{\mathbf{b}}_k)_{k+1}$.

The "Update the covariance" step, detailed below, is how the structure of $R_b(\mathbf{b})$ can be used to improve the parameter estimates.

We have explored several different approaches to computing and using $R_w(\hat{\mathbf{b}})$:

1. Ignore $R_w(\mathbf{b})$: Neglect the correlation structure and simply assume that $R_w(\hat{\mathbf{b}}_k)_i = I$. This gives the equivalent of taking a scalar measurement and is used as a sort of worst-case basis for comparison among the different algorithms.
2. Use the correct value of $R_w(\mathbf{b})$: That is, assume that $\mathbf{b}$ and σ_v^2 and σ_η^2 are known and compute $R_w(\hat{\mathbf{b}})_k$ according to (6). This provides a limit on best-case performance against which other methods can be compared.
3. Use the estimate of $\mathbf{b}$: Using the correct values of σ_v^2 and σ_η^2, compute the autocorrelation matrix using $\hat{\mathbf{b}}_k$ in (6).
4. Estimate $\hat{\mathbf{b}}$, fix σ_v^2 and σ_η^2: With assumed values of σ_v^2 and σ_η^2, compute the autocorrelation matrix using $\hat{\mathbf{b}}_k$ in (6).
5. Estimate everything: Estimate the values of σ_v^2 and σ_η^2, then use them with $\hat{\mathbf{b}}_k$ in (6).

In the early stages while $\hat{\mathbf{b}}_k$ is poorly converged, it is best to use option 1 (assuming identity covariance matrix) until $\hat{\mathbf{b}}$ has settled near its final value, then switch to option 4 using estimated $\hat{\mathbf{b}}_k$ until the moments in $\Sigma(\mathbf{b})$ have converged sufficiently well that reliable estimates of the variances can be obtained and option 5 can be used. As described in Section 3, the information necessary to estimate the variances can be accumulated without yet having a decent estimate for $\hat{\mathbf{b}}_k$.

The covariance update/inverse does not necessarily need to be done at every step. Particularly as $\hat{\mathbf{b}}_k$ settles towards its final value, there is little to be gained by updating $R_w(\hat{\mathbf{b}})$ at every step.

3. Estimating the Variances

As the results below will indicate, fixed values of σ_v^2 and σ_η^2 can be used in the estimate of $R_w(\mathbf{b})$ instead of estimated values, so for the purposes of estimating $\mathbf{b}$, it is not strictly necessary to estimate these variances. But for other purposes it may be necessary to have an estimate of σ_v^2 and σ_η^2. A maximum likelihood estimation approach is thus described in this section.

For a given estimate of the coefficients $\hat{\mathbf{b}}$, write $R_w(\hat{\mathbf{b}})$ as

$$R_w(\hat{\mathbf{b}}) = 2\sigma_\eta^2 I + 2\sigma_v^2 B(\hat{\mathbf{b}}).$$

For a sequence of k vectors $\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1}$, assumed to be nonoverlapping, encompassing N samples, the joint log likelihood function is a complex (circularly symmetric) Gaussian

$$\log f(\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{k-1} | \mathbf{b}, R_w(\mathbf{b}), \sigma_v^2, \sigma_\eta^2) = -N \log \pi - k \log \det R_w(\mathbf{b}) - \sum_{i=0}^{k-1} (\mathbf{y}_i - Y_i \mathbf{b})^H R_w(\mathbf{b})^{-1} (\mathbf{y}_i - Y_i \mathbf{b})]. \quad (12)$$

Let

$$\Sigma(\mathbf{b}) = \frac{1}{k} \sum_{i=0}^{k-1} (\mathbf{y}_i - Y_i \mathbf{b})(\mathbf{y}_i - Y_i \mathbf{b})^H \quad (13)$$

denote the sample covariance. It can be shown that (see Appendix A)

$$\frac{\partial}{\partial \sigma_\eta^2} \log f(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k | \mathbf{b}, \sigma_v^2, \sigma_\eta^2) = -k \operatorname{tr}(R_w(\mathbf{b})^{-1}) + k \operatorname{tr}(R_w(\mathbf{b})^{-1} \Sigma(\mathbf{b}) R_w(\mathbf{b})^{-1}). \quad (14)$$

and that

$$\frac{\partial}{\partial \sigma_v^2} \log f(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k | \mathbf{b}, \sigma_v^2, \sigma_\eta^2) = -k \sum_{i,j} (R_w(\mathbf{b})^{-1})_{i,j} (B(\mathbf{b}))_{ij} + k \sum_{i,j} (R_w(\mathbf{b})^{-1} \Sigma(\mathbf{b}) R_w(\mathbf{b})^{-1})_{i,j} (B(\mathbf{b}))_{i,j}. \quad (15)$$

The gradients (14) and (15) are equal to zero when

$$R_w(\mathbf{b}) = \Sigma(\mathbf{b}). \quad (16)$$

The variance estimates are chosen to satisfy the condition (16) using an estimate of $\mathbf{b}$ from previous estimates, that is

$$2\sigma_\eta^2 I + 2\sigma_v^2 B(\hat{\mathbf{b}}) = \Sigma(\hat{\mathbf{b}}). \quad (17)$$

We define the offset trace as

$$\operatorname{tr}(B, \ell) = \sum_i B_{i,i+\ell},$$

where the usual trace is obtained when $\ell = 0$, and for $\ell > 0$, the sum is taken on the ℓth superdiagonal. Let $\beta_\ell = \operatorname{tr}(B(\hat{\mathbf{b}}), \ell)$ for $\ell = 0, 1, \ldots, d-1$, where that is, the trace on the ℓth superdiagonal. Taking the offset trace of (17) gives

$$2d\sigma_\eta^2 + 2\sigma_v^2 \beta_0 = \operatorname{tr}(\Sigma(\hat{\mathbf{b}}), 0)$$
$$2\sigma_v^2 \beta_\ell = \operatorname{tr}(\Sigma(\hat{\mathbf{b}}), \ell), \qquad \ell = 1, 2, \ldots, d-1.$$

The ML estimates of the variances are the solutions to

$$2 \begin{bmatrix} d & \beta_0 \\ 0 & \beta_1 \\ \vdots & \\ 0 & \beta_{d-1} \end{bmatrix} \begin{bmatrix} \hat{\sigma}_\eta^2 \\ \hat{\sigma}_v^2 \end{bmatrix} = \begin{bmatrix} \operatorname{tr}(\Sigma(\hat{\mathbf{b}}), 0) \\ \operatorname{tr}(\Sigma(\hat{\mathbf{b}}), 1) \\ \vdots \\ \operatorname{tr}(\Sigma(\hat{\mathbf{b}}), d-1). \end{bmatrix} \quad (18)$$

A significant number of terms of $\Sigma(\hat{\mathbf{b}})$ must be accumulated in order to estimate the variances well. Initially, before reasonably accurate estimates of $\mathbf{b}$ are available, using inaccurate $\hat{\mathbf{b}}$s can result in a highly

inaccurate $\Sigma(\hat{\mathbf{b}})$. As a result, it is desirable to accumulate moments of $\mathbf{y}_i$ and Y_i without using $\hat{\mathbf{b}}$. The trace of the sample covariance is

$$\text{tr}(\Sigma(\mathbf{b}), \ell) = \frac{1}{k}\sum_{i=1}^{k} \text{tr}(\mathbf{y}_i \mathbf{y}_i^H, \ell) - \text{tr}(\mathbf{y}_i \mathbf{b}^H Y_i^H, \ell) - \text{tr}(Y_i \mathbf{b} \mathbf{y}_i^H, \ell) + \text{tr}(Y_i \mathbf{b} \mathbf{b}^H Y_i^H, \ell).$$

The different terms can be written as

$$\sum_i \text{tr}(\mathbf{y}_i \mathbf{b}^H Y_i^H, \ell) = \sum_j \sum_k \sum_i \mathbf{y}_{i,j} \mathbf{b}_k^* Y_{i,j+\ell,k}^* = \sum_k b_k^* \sum_i [Y_{i,\ell+1:\text{end},k}]^H \mathbf{y}_{i,1:\text{end}-\ell}$$

$$\sum_i \text{tr}(Y_i \mathbf{b} \mathbf{y}_i^H) = \sum_i \sum_j \sum_k Y_{i,k,j} \mathbf{b}_j \mathbf{y}_{i,k+\ell}^* = \sum_j b_j \sum_i [\mathbf{y}_{i,1+\ell:\text{end}}]^H Y_{i,1:\text{end}-\ell,j}.$$

(It is noted that, except when $\ell = 0$, these two terms are not conjugates of each other, so that both of these terms must be retained.)

$$\sum_i \text{tr}(Y_i \mathbf{b} \mathbf{b}^H Y_i^H, \ell) = \sum_j \sum_k \sum_l \sum_i Y_{i,k,j} (\mathbf{b} \mathbf{b}^H)_{j,l} [Y_i^H]_{l,k+\ell}$$

$$= \sum_j \sum_l (\mathbf{b} \mathbf{b}^H)_{j,l} \sum_i [Y_{i,1+\ell:\text{end},:}]^H Y_{i,1:\text{end}-\ell,:}$$

These terms can be computed by recursively propagating the following quantities, for $k = 1, 2, \ldots, p$ and $\ell = 0, 1, \ldots, p$:

$$a^0(\ell, k) = \frac{1}{k}\sum_{i=0}^{k-1} \mathbf{y}_{i,1+\ell,\text{end}}^H \mathbf{y}_{i,1:\text{end}-\ell}$$

$$= \frac{1}{k}(\mathbf{y}_{k,1+\ell:\text{end}}^H \mathbf{y}_{k,1:\text{end}-\ell} + (k-1)a^0(\ell, k-1))$$

$$\mathbf{a}_k^1(\ell, k) = \frac{1}{k}\sum_{i=0}^{k-1} [Y_{i,\ell+1:\text{end},k}]^H \mathbf{y}_{i,1:\text{end}-\ell}$$

$$= \frac{1}{k}([Y_{k,\ell+1:\text{end},k}]^H \mathbf{y}_{k,1:\text{end}-\ell} + (k-1)\mathbf{a}_k^1(\ell, k-1))$$

$$\mathbf{a}_k^2(\ell, k) = \frac{1}{k}\sum_{i=0}^{k-1} \mathbf{y}_{i,\ell+1:\text{end}}^H Y_{i,1:\text{end}-\ell,k}$$

$$= \frac{1}{k}(\mathbf{y}_{k,\ell+1:\text{end}}^H Y_{k,1:\text{end}-\ell,k} + (k-1)\mathbf{a}_k^2(\ell, k-1))$$

$$A(\ell, k) = \frac{1}{k}\sum_{i=0}^{k-1} Y_{i,\ell+1:\text{end},:}^H Y_{i,1:\text{end}-\ell,:}$$

$$= \frac{1}{k}(Y_{k,\ell+1:\text{end},:}^H Y_{k,1:\text{end}-\ell,:} + (k-1)A(\ell, k-1)).$$

Then

$$\text{tr}(\Sigma(\hat{\mathbf{b}}), \ell) = a^0(\ell, k) - \hat{\mathbf{b}}^H \mathbf{a}^1(\ell, k) - \hat{\mathbf{b}}^T \mathbf{a}^2(\ell, k) + \text{tr}((\hat{\mathbf{b}}\hat{\mathbf{b}}^H)A(\ell, k)).$$

Figure 2 shows the result of this estimation in an example with $p = 3$ and $d = 5$. The mean and standard deviation over 50 independent runs is shown, for up 2000 samples averaged to produce $\Sigma(\mathbf{b})$. The true value of $\mathbf{b}$ is used. The least-squares solution to (18) can produce variance estimates for σ_η^2 which are negative, which is nonphysical. Solutions constrained so that the variances are constrained to the range

$[0.1, 5]$ were also computed using `CVX` [113]. From these results, about 500 samples are needed before the variance estimate converges to a value somewhat close to the true value.
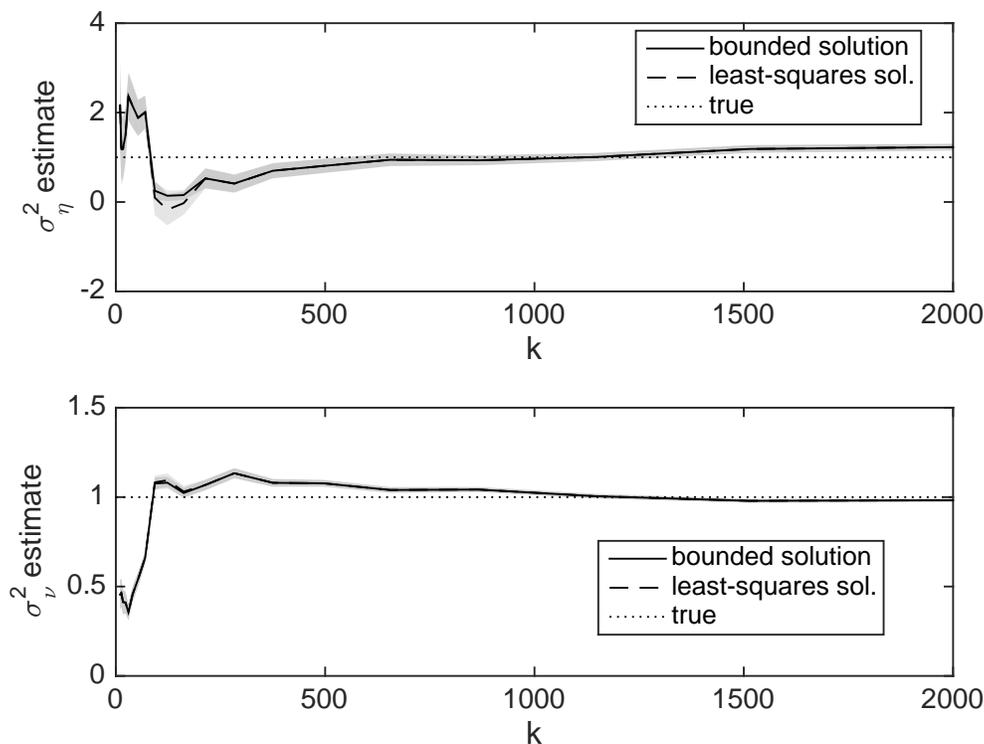
Figure 2. Example of estimating the variances σ_ν^2 (**top plot**) and σ_η^2 (**bottom plot**), using true **b**. k indicates the number of points used in the estimates. Shading indicates standard deviation of the estimates over 50 iterations. Bounded solution is computed using CVX to avoid negative estimates.

4. Vector Autoregressive Formulation

In this section we extend the results of the previous section to vector autoregressive random processes in noise, where at each time vector of length K is produced, and the coefficients are $K \times K$ matrices. In the interest of generality, a constant offset **v** is also included. The noisy observations can be written as

$$\mathbf{y}(m) = \boldsymbol{\xi}(m) + \boldsymbol{\nu}(m) = \mathbf{v} + \sum_{i=1}^{p} B_i \boldsymbol{\xi}(m-i) + \boldsymbol{\eta}(m) + \boldsymbol{\nu}(m).$$

As in the scalar case, the noisy observations can be written as

$$\mathbf{y}(m) = \mathbf{v} + \sum_{i=1}^{p} B_i \mathbf{y}(m-i) + \mathbf{w}(m),$$

where

$$\mathbf{w}(m) = \boldsymbol{\eta}(m) + \boldsymbol{\nu}(m) - \sum_{i=1}^{p} B_i \boldsymbol{\nu}(m-i).$$

Following the usual practice [15], this is vectorized to obtain a vector of unknown parameters as follows.

$$\mathbf{y}(m) = \begin{bmatrix} \mathbf{v} & B_1 & \cdots & B_p \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{y}(m-1) \\ \vdots \\ \mathbf{y}(m-p) \end{bmatrix} + \mathbf{w}(m).$$

Applying the vec operator ([114], Chapter 9) we obtain

$$\mathbf{y}(m) = \left(\begin{bmatrix} 1 \\ \mathbf{y}(m-1) \\ \vdots \\ \mathbf{y}(m-p) \end{bmatrix}^T \otimes I_K \right) \text{vec}\left(\begin{bmatrix} \mathbf{v} & B_1 & \cdots & B_p \end{bmatrix} \right) + \mathbf{w}(m).$$

Let

$$\mathbf{b} = \text{vec}\left(\begin{bmatrix} \mathbf{v} & B_1 & \cdots & B_p \end{bmatrix} \right) \in \mathbb{C}^{(K^2 p + K) \times 1}$$

and

$$Y(m) = \left(\begin{bmatrix} 1 \\ \mathbf{y}(m-1) \\ \vdots \\ \mathbf{y}(t-p) \end{bmatrix}^T \otimes I_K \right) \in \mathbb{C}^{K \times (K^2 p + K)}.$$

Then

$$\mathbf{y}(m) = Y(m)\mathbf{b} + \mathbf{w}(m).$$

The noise structure for a single vector can be written as

$$\mathbf{w}(m) = \begin{bmatrix} I & I & -B_1 & -B_2 & \cdots & -B_p \end{bmatrix} \begin{bmatrix} \eta(m) \\ \nu(m) \\ \nu(m-1) \\ \nu(m-2) \\ \vdots \\ \nu(m-p) \end{bmatrix}.$$

Then

$$E[\mathbf{w}(m)\mathbf{w}^H(m)] = 2(\sigma_\eta^2 I + \sigma_\nu^2 (I + \sum_{i=1}^{p} B_i B_i^H) \triangleq R_w(0).$$

We also find

$$E[\mathbf{w}(m)\mathbf{w}^H(m+\ell)] = 2\sigma_\nu^2 (-B_\ell^H + \sum_{i=1}^{p-\ell} B_i B_{i+\ell}^H) \triangleq R_w(\ell).$$

As in the scalar case, stack up multiple observations so that the correlation structure may be exploited:

$$\begin{bmatrix} \mathbf{y}(m) \\ \mathbf{y}(m-1) \\ \vdots \\ \mathbf{y}(m-d+1) \end{bmatrix} = \begin{bmatrix} Y(m) \\ Y(m-1) \\ \vdots \\ Y(m-d+1) \end{bmatrix} \mathbf{b} + \begin{bmatrix} \mathbf{w}(m) \\ \mathbf{w}(m-1) \\ \vdots \\ \mathbf{w}(m-d+1) \end{bmatrix}.$$

Write this as

$$\mathfrak{y}(m) = \mathfrak{Y}(m)\mathbf{b} + \mathfrak{w}(m).$$

The correlation structure of the stacked noise vector is

$$E[\mathfrak{w}(m)\mathfrak{w}(m)^H] =$$

$$\begin{bmatrix} R_w(0) & R_w(-1) & \cdots & R_w(-d+1) \\ R_w(1) & R_w(0) & \cdots & R_w(-d+2) \\ \vdots & & & \\ R_w(d-1) & R_w(d-2) & \cdots & R_w(0) \end{bmatrix}$$

$$\stackrel{\triangle}{=} \mathfrak{R}(\mathbf{b}).$$

Let $\mathfrak{y}(m), \mathfrak{y}(m+s), \ldots, \mathfrak{y}(m+sk)$ and $\mathfrak{Y}(m), \mathfrak{Y}(m+s), \ldots, \mathfrak{Y}(m+sk)$ be a sequence of vectors and matrices where the vector samples are skipped by s samples at each step, and for convenience denote these as $\mathfrak{y}_1, \mathfrak{y}_2, \ldots, \mathfrak{y}_k$ and $\mathfrak{Y}_1, \mathfrak{Y}_2, \ldots, \mathfrak{Y}_k$. Following the method of section 2, the estimate

$$\hat{\mathbf{b}}_k = \arg\max_{\mathbf{b}} \sum_{i=1}^{k} \lambda^{k-i} \|\mathfrak{y}_i - \mathfrak{Y}_i \mathbf{b}\|_{\mathfrak{R}(\mathbf{b})^{-1}}$$

is determined by the solution to the normal equations

$$\left(\sum_{i=0}^{k} \lambda^{k-i} \mathfrak{Y}_i^H \mathfrak{R}(\hat{\mathbf{b}})^{-1} \mathfrak{Y}_i \right)^{-1} \mathbf{b} = \left(\sum_{i=0}^{k} \lambda^{k-i} \mathfrak{Y}_i^H \mathfrak{R}_w(\mathbf{b})^{-1} \cdot \mathfrak{y}_i \right)$$

With the appropriate use of vectorized data, Algorithm 1 can be used for the VAR as well.

5. Some Results

Several test cases were examined to determine the performance of the ICWARE approach; these are summarized in Table 1, where the pole locations are given in polar form $\rho e^{j\theta}$.

Table 1. Parameters for test cases.

Case	Order	Pole Locations
1	3	$0.95e^{j2\pi0.65}, 0.95e^{j2\pi0.7}, 0.95e^{j2\pi0.75}$
2	3	$0.9e^{j2\pi0.65}, 0.9e^{j2\pi0.7}, 0.9e^{j2\pi0.75}$
3	3	$e^{j2\pi0.65}, e^{j2\pi0.7}, e^{j2\pi0.75}$
4	6	$0.75e^{\pm j2\pi0.1}, 0.8e^{\pm j2\pi0.2}, 0.85e^{\pm j2\pi0.35}$
5	6	$0.98e^{\pm j2\pi0.05}, 0.97e^{\pm j2\pi0.15}, 0.8e^{\pm j2\pi0.35}$
6	6	$0.98e^{\pm j2\pi0.1}, 0.97e^{\pm j2\pi0.1}, 0.98e^{\pm j2\pi0.15}$

The first example, designated Case 1, is a system with $p = 3$ having poles at $\rho e^{j\theta_1}, \rho e^{j\theta_2}, \rho e^{j\theta_3}$, with $\rho = 0.95$ and $(\theta_1, \theta_2, \theta_3) = (0.65, 0.7, 0.75)$. resulting in a fairly narrowband complex AR signal. The noise variances are $\sigma_\eta^2 = 1$, and $\sigma_\nu^2 = 1$. This case is used to explore some of variations in performance as parameters of the algorithm are varied. Figure 3 shows $\|\mathbf{b} - \hat{\mathbf{b}}\|_2^2$ as a function of iteration using various forms of $R_b(\mathbf{b})$ and values of d (the stack height) from 1 to 7. The "skip" is set to $s = 3$. The results are obtained by averaging the results of 50 independent runs. In these plots, "Iteration number" refers to the number of blocks of data used in an online scheme, each iteration of the algorithm corresponding to

one data block. The results are compared with $\|\mathbf{b} - \hat{\mathbf{b}}\|_2^2$ for YW with noisy measurements and noise-free measurements and scalar AR estimation. The autocorrelation values for the YW method are computed using 30000 points. The Burg method was also used, but with this many points there is little difference between Burg and conventional YW. The scalar AR estimation (black dotted line) converges to the YW performance, as does the ICWARE with $d = 1$, as expected.

In Figure 3, three different ways of determining $R_b(\mathbf{b})$ are used. The solid lines (subplot (a)) use the true $R_b(\mathbf{b})$. This is, of course, unavailable in practice, but these plots serve as a basis for comparison against real algorithms. The dashed lines (subplot (b)) use an identity matrix for $R_b(\mathbf{b})$ for the first 30 iterations (to establish an estimate of $\mathbf{b}$), following which $R_b(\hat{\mathbf{b}})$ is computed using estimated $\mathbf{b}$ and the correct variances σ_v^2 and σ_ϵ^2. The dotted lines (subplot (c)) are similar, except that fixed (but wrong) values of $\sigma_v^2 = 5$ and $\sigma_\epsilon^2 = 5$ are used in the computation of $R_b(\hat{\mathbf{b}})$. It is clear that as d increases, the performance significantly improves, especially for the first few values of d. Interestingly, for $d = 7$, the error performance is on the order of the error in the noise-free YW.
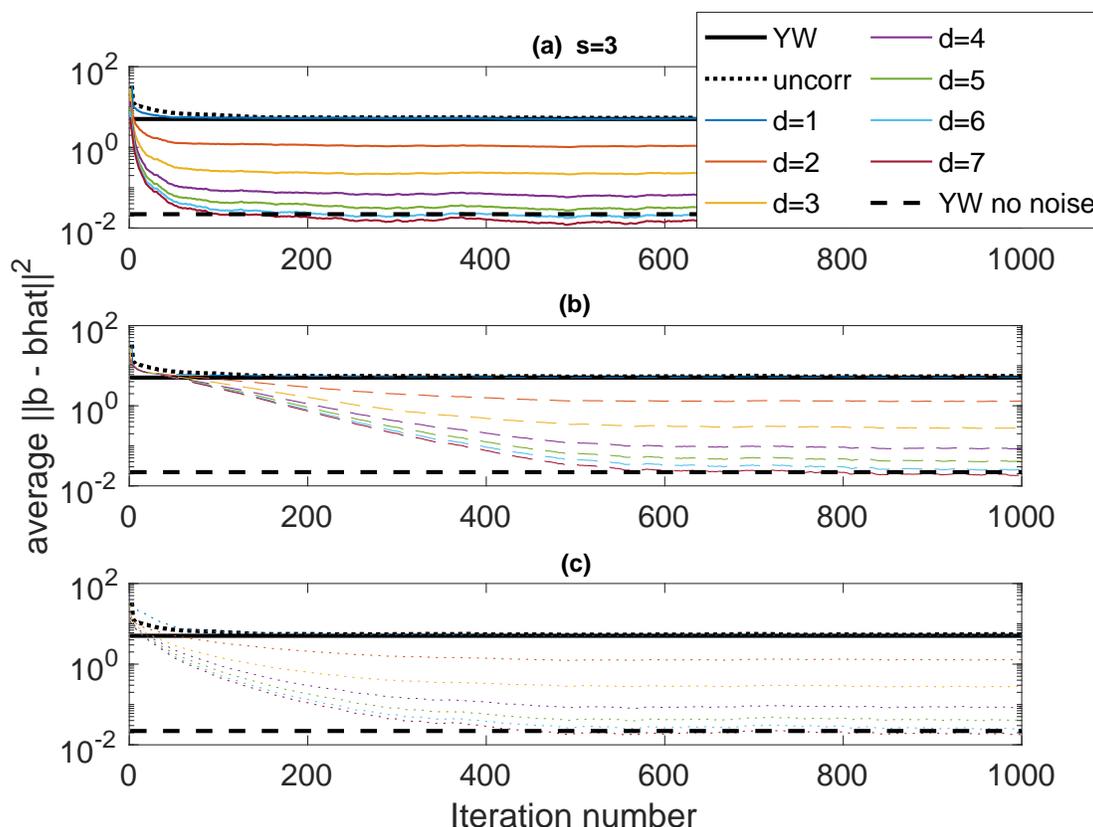
Figure 3. Case 1: Error performance fo different values of d with $s = 3$. (**a**) Solid: True $R_b(\mathbf{b})$; (**b**) Dashed: $R_b(\hat{\mathbf{b}})$ estimated from $\hat{\mathbf{b}}$ and correct variances; (**c**) Dotted: $R_b(\mathbf{b})$ estimated from $\hat{\mathbf{b}}$ and fixed incorrect variances. Comparison: Black dotted: conventional least-squares; Solid black: Yule-Walker (YW) with noisy observations; Dashed black: NW with noise-free observations.

The improvements of the new method after 1000 iterations (denoted by $\|\mathbf{b} - \hat{\mathbf{b}}_k\|_d^2$) compared with the YW result $\|\mathbf{b} - \hat{\mathbf{b}}_{YW}\|_2^2$ tabulated in Table 2 for $s = 2, 3$ and 4. There is little difference between $s = 2$ and $s = 3$, but when $s = 4$, that is, $s > p$, the performance declines.

Table 2. Comparison of $\|\mathbf{b} - \hat{\mathbf{b}}_k\|^2$ computed for $d = 1, \ldots, 7$ with $\|\mathbf{b} - \hat{\mathbf{b}}_{YW}\|^2$ for YW.

d	$10\log_{10}\|\mathbf{b} - \hat{\mathbf{b}}_k\|_2^2/\|\mathbf{b} - \hat{\mathbf{b}}_{YW}\|^2$		
	$s = 2$	$s = 3$	$s = 4$
2	-5.9	-5.8	11.1
3	-12.4	-12.5	4.5
4	-17.5	-17.5	-0.6
5	-20.6	-20.6	-3.9
6	-22.6	-22.7	-5.9
7	-23.9	-24.0	-7.2

Figure 4 shows the influence of the skip s on the performance, with the same methods of estimating $R_b(\mathbf{b})$ as shown in Figure 3. For a fixed value of $d = 7$, the error curves for different values of s are shown. Larger s for $s \leq p$ does improve the performance, but only slightly.

(a) Solid: True $R_b(\mathbf{b})$; (b) Dashed: $R_b(\hat{\mathbf{b}})$ estimated from $\hat{\mathbf{b}}$ and correct variances; (c) Dotted: $R_b(\mathbf{b})$ estimated from $\hat{\mathbf{b}}$ and fixed incorrect variances.
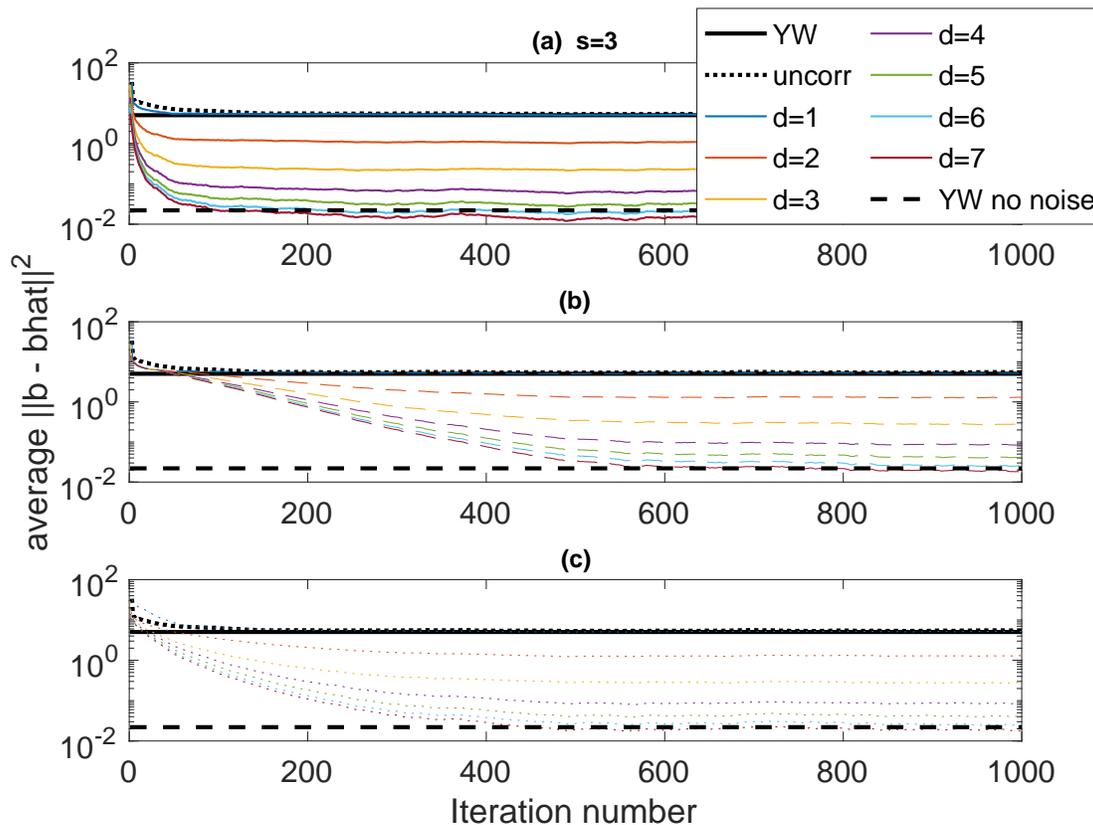
Figure 4. Case 1: Error performance for different values of s with $d = 7$. (**a**) Solid: True $R_b(\mathbf{b})$; (**b**) Dashed: $R_b(\mathbf{b})$ estimated from $\hat{\mathbf{b}}$ and variances; (**c**) Dotted: $R_b(\mathbf{b})$ estimated from $\hat{\mathbf{b}}$ and fixed incorrect variances. Comparison: Black dotted: conventional least-squares; Solid black: Yule-Walker (YW) with noisy observations; Dashed black: NW with noise-free observations.

Figure 5 shows estimates of the variances σ_v^2 and σ_η^2 for different values of d. Obviously we need $d > 2$ (in order to have two equations). However, when $d = 2$ there is a very large variance. By the point $d \geq 4$, the variance estimates are very close to the true variances.
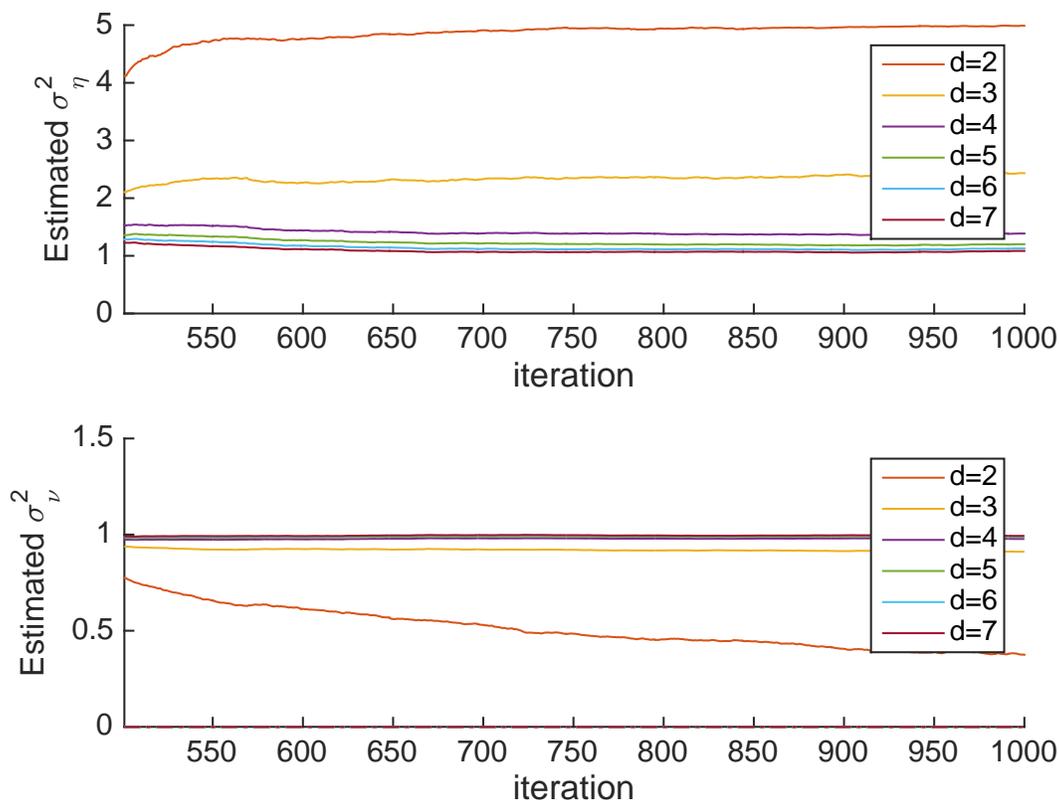
Figure 5. Case 1: Estimated variances for different values of d. **Top**: Estimated σ_η^2; **Bottom**: Estimated σ_ν^2.

Figure 6 shows the magnitude frequency response for the true spectrum, the YW spectrum, and the spectrum computed from $\hat{\mathbf{b}}$ obtained using $d = 7$ and $d = 3$ after convergence. (The spectrum is not an average, but only a single outcome.) The ICWARE estimated spectrum is very close to the true spectrum while the YW spectrum has considerable bias, exhibiting strong peaks not present in the true spectrum.
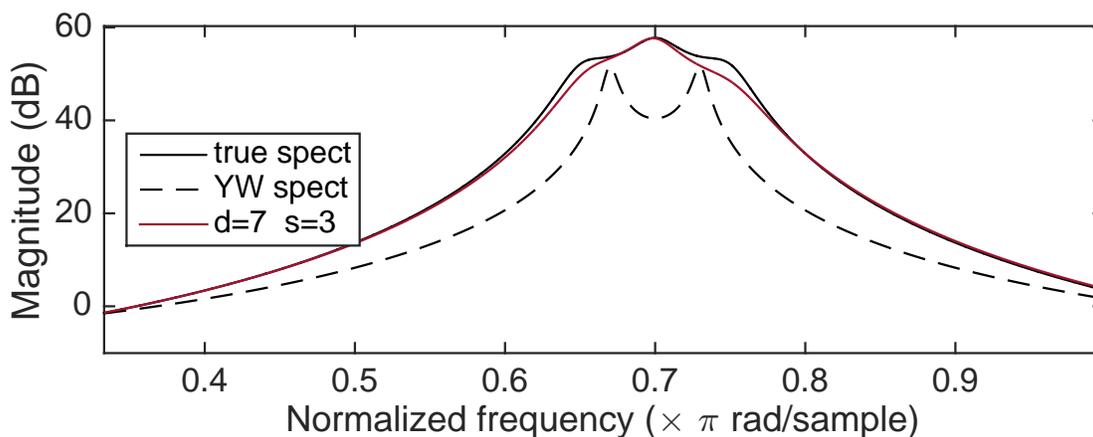
Figure 6. Case 1: Comparison of ICWARE estimated spectrum with true spectrum, and the YW estimated spectrum.

As a point of comparison, TLS is used to estimate the AR parameters. In TLS, the equation

$$
\begin{bmatrix}
y(m) & y(m-1) & \cdots & y(m-p+1) \\
y(m+1) & y(m) & \cdots & y(m-p+2) \\
\vdots & & & \\
y(m+M) & y(m+M-1) & \cdots & y(m+M-p)
\end{bmatrix}
\begin{bmatrix}
b_1 \\ b_2 \\ \vdots \\ b_p
\end{bmatrix}
=
\begin{bmatrix}
y(m+1) \\ y(m+2) \\ \vdots \\ y(m+M+1)
\end{bmatrix}
+
\begin{bmatrix}
v(m+1) \\ v(m+2) \\ \vdots \\ v(m+M+1)
\end{bmatrix}
\tag{19}
$$

(for some m and M) is perturbed in both the matrix on the left and the vector of observations on the right. TLS can be computed using the singular value decomposition (SVD), making it computationally complex. It also makes it more difficult to create algorithms which track changing parameters, and, as shown below, gives inferior parameter estimates than the method presented here. Figure 7 shows the error for $\hat{b}$ computed using TLS. The parameter M is the height of the linear system that is solved using the TLS method described in Reference [114]. The values $M \in \{20, 60, 100, 140\}$ are examined. At each iteration, a different set of points are used in the TLS equations, each of which is computed using a (relatively complicated) singular value decomposition. Even with 20 rows in the TLS equations, performance only roughly comparable to the YW equations is obtained. For 60 or more equations, the improvement of the performance quickly saturates, so that 60 or 140 perform rather comparably. The computational complexity is rather high. For each solution (at each iteration), the SVD of a $M \times 3$ matrix is computed. Despite the computational complexity, the ICWARE method performs superior to the TLS for $d \geq 4$. (The method of estimating $R_b(\mathbf{b})$ in is the same as used for Figure 3b.)
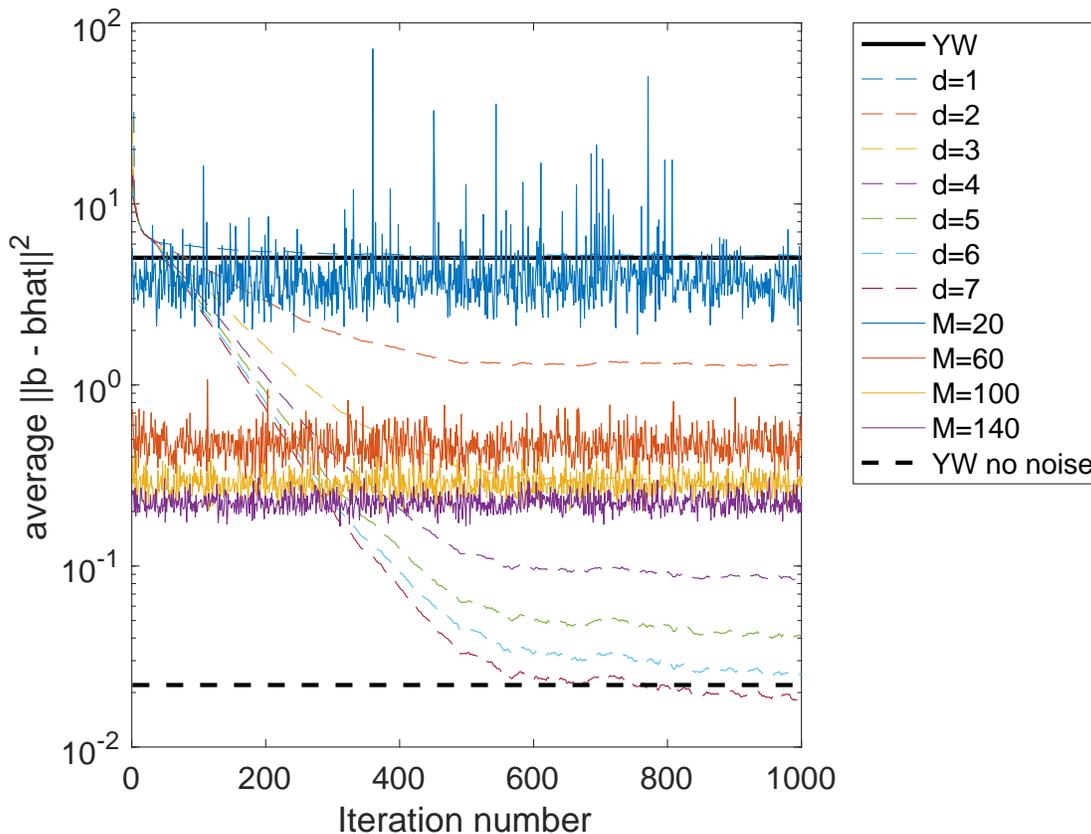
Figure 7. Case 1: Comparison of new method with total least squares (TLS) solutions for different TLS sizes. Dashed: ICWARE for different values of d; Solid: TLS for different matrix sizes.

The results presented to this point take 1000 blocks of data. A consideration is whether it is possible to iteratively re-use data so that less data is required, but the reduced error from the ICWARE algorithm can be obtained. In Figure 8, $k = 100$ blocks of data are employed, and the processing is repeated 10 times on each block. Figure 9 shows the error as a result of each iteration. Each line in the plot corresponds to one pass through the data. The top plot uses $R_b(\hat{\mathbf{b}})$ estimated using $\hat{\mathbf{b}}$ and true variances; the bottom plot uses $R_b(\hat{\mathbf{b}})$ using $\hat{\mathbf{b}}$ and fixed variances. These figures show that iterating over the data provides essentially the same performance as longer data.
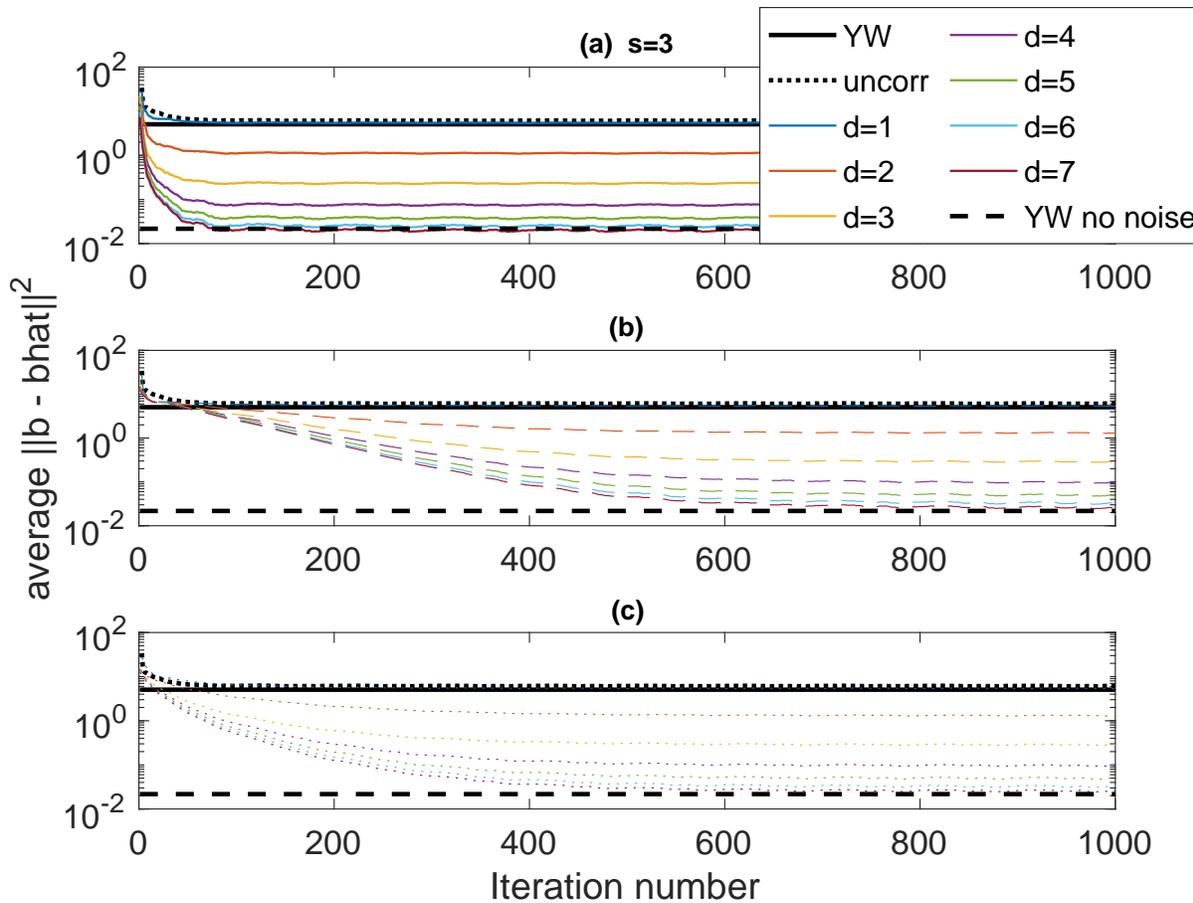
Figure 8. Case 1: Error performance of 10 repetitions on blocks of length 100. (**a**) Solid: True $R_b(\mathbf{b})$; (**b**) Dashed: $R_b(\mathbf{b})$ estimated from $\hat{\mathbf{b}}$ and variances; (**c**) Dotted: $R_b(\mathbf{b})$ estimated from $\hat{\mathbf{b}}$ and fixed incorrect variances. Comparison: Black dotted: conventional least-squares; Solid black: Yule-Walker (YW) with noisy observations; Dashed black: NW with noise-free observations.

The next example, designated Case 2, is also a third order system with poles at the same angles as Case 1, but with $\rho = 0.9$. The following figures show examples of the same sort of results shown for Case 1. In Figure 10 the various methods are compared. In this case, however, the minimum error does not reach as low as the noise-free YW case. Figure 11 shows the resulting spectrum, again showing that the ICWARE spectrum is closer than the YW spectrum.
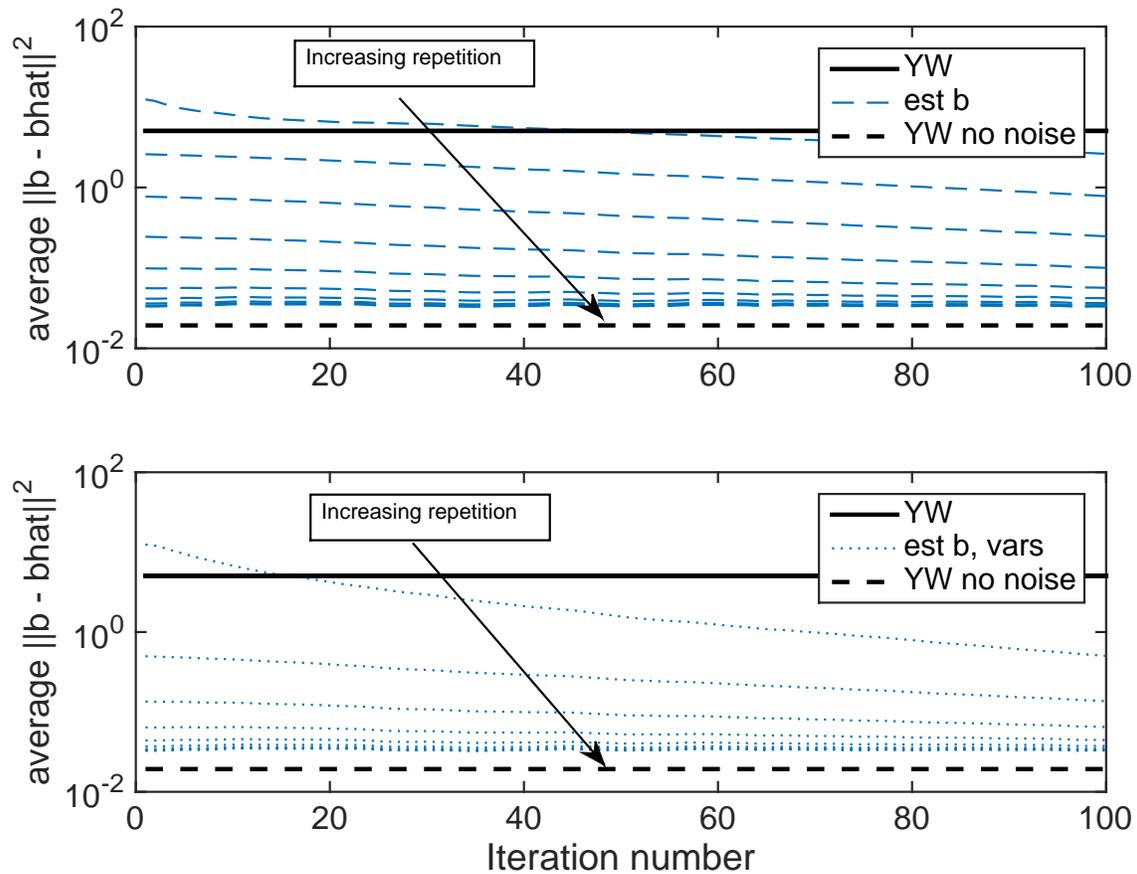
Figure 9. Performance on 10 repetitions on blocks of length 100, with the results folded for each iteration. **Top**: Estimated $R_b(\hat{\mathbf{b}})$ using $\hat{\mathbf{b}}$ and true variances; **Bottom**: Estimated $R_b(\hat{\mathbf{b}})$ using $\hat{\mathbf{b}}$ and fixed variances.

Case 3 involves a system having poles at $\rho = 1$—that is, pure sinusoidal signals—and angles $(\theta_1, \theta_2, \theta_3) = (0.65, 0.655, 0.66)$. The estimated spectra are shown in Figure 12. The spectral peaks are clearly evident in the ICWARE estimate.

As another comparison, ICWARE spectral estimates are compared to the spectra from dual Kalman filters using the examples from Reference [104], shown as Cases 4,5, and 6 in Table 1. Figure 13 shows the results. Spectra for twenty realizations are plotted, along with the true spectrum and the mean. The values of $d = p + 3$ and $s = 3$ were selected, since simulations for Case 1 (above) suggest these are reasonable values. Comparison of these results with figures 3, 4, and 6 of Reference [104] shows fairly comparable performance. When the poles are not near the unit circle (Case 4), the ICWARE estimate tend to have smoother spectra (estimated poles with smaller magnitude). When the poles are nearer the units circle (Case 5, Case 6), ICWARE seems to do a comparable job at capturing the peaks, and does somewhat better at representing the high-frequency dropoff of the spectrum. In all of these simulations, an estimated $R_b(\mathbf{b})$ was employed.
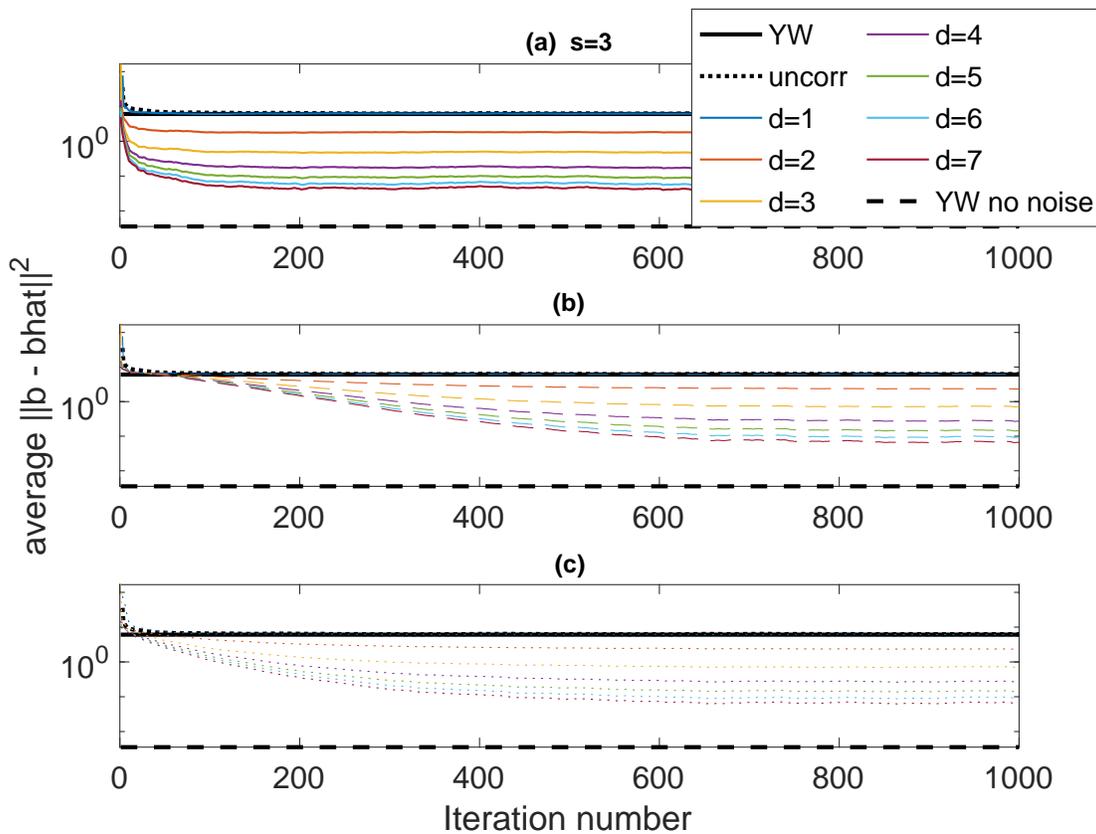
Figure 10. Case 2: Error performance for different values of d with $s = 3$. (**a**) Solid: True $R_b(\mathbf{b})$; (**b**) Dashed: $R_b(\mathbf{b})$ estimated from $\hat{\mathbf{b}}$ and variances; (**c**) Dotted: $R_b(\mathbf{b})$ estimated from $\hat{\mathbf{b}}$ and fixed incorrect variances. Comparison: Black dotted: conventional least-squares; Solid black: Yule-Walker (YW) with noisy observations; Dashed black: NW with noise-free observations.
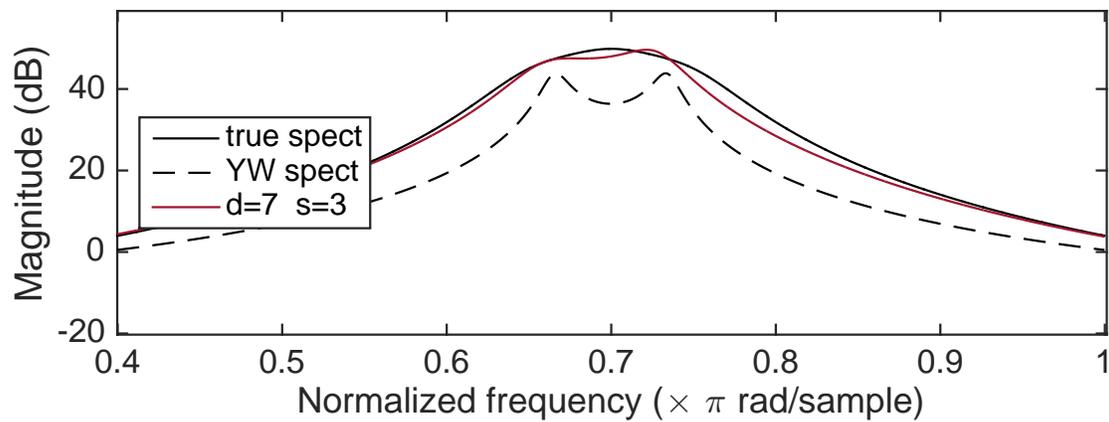
Figure 11. Case 2: Comparison of ICWARE estimated spectrum with true spectrum, and the YW estimated spectrum.
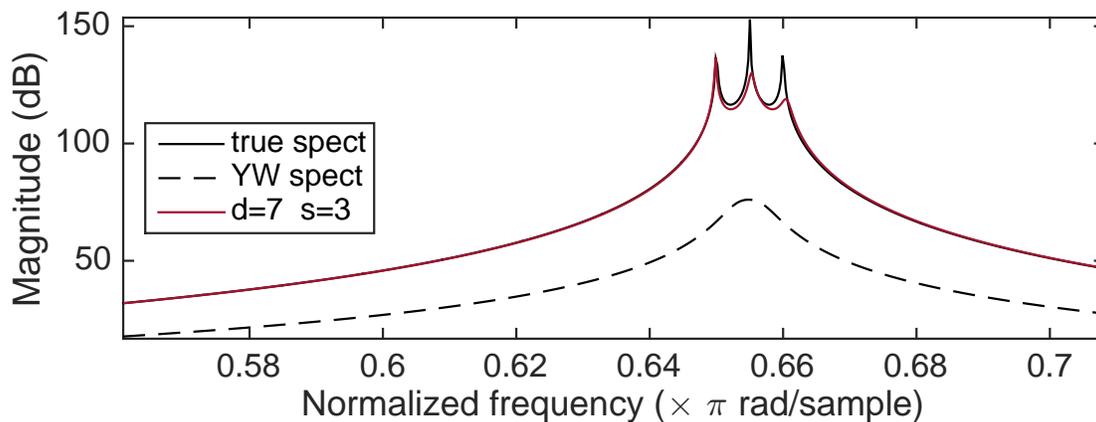
Figure 12. Case 3: Comparison of ICWARE estimated spectrum with true spectrum, and the YW estimated spectrum.
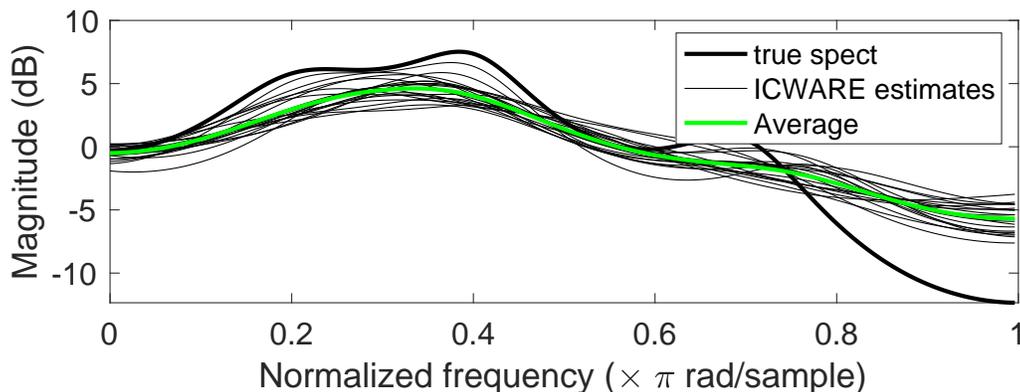
(**a**) Spectrum for Case 4
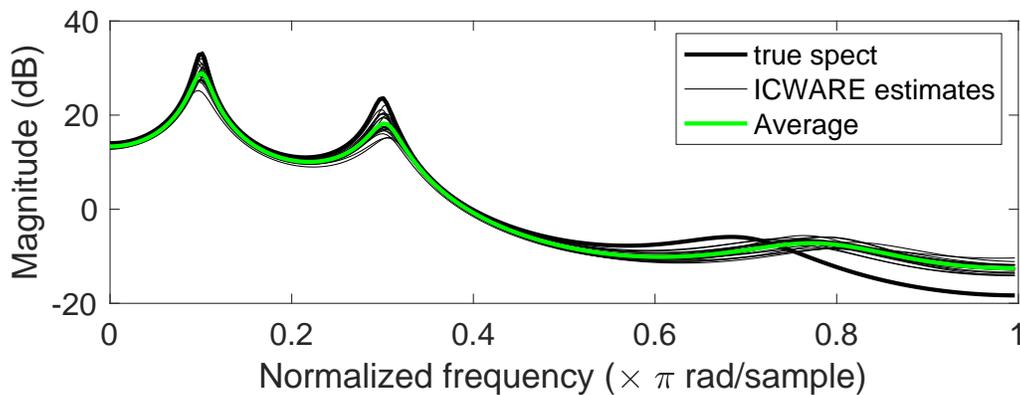
(**b**) Spectrum for Case 5

Figure 13. *Cont.*
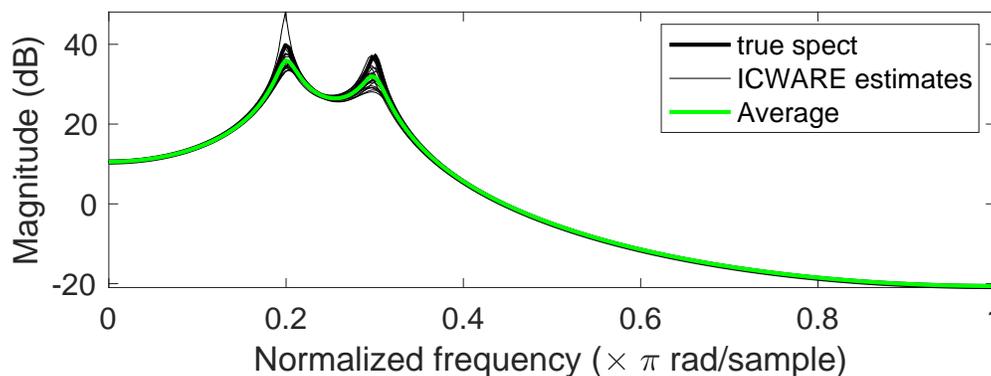
(**c**) Spectrum for Case 6

Figure 13. Spectral estimation results for examples from Reference [104]. (**a**) Case 4; (**b**) Case 5; (**c**) Case 6. $\sigma_\eta^2 = 1$,

6. Summary and Conclusions

In this paper, we have shown that accounting for observation noise added to a pure AR(p) process results in noise which is correlated across lags. This makes it expedient to employ stacked observation vectors, and estimating the parameters in a vector sense. This results in an algorithm that is essentially a vector RLS adaptive filter. Several different methods were described to obtain information about the correlation matrix $R_w(\mathbf{b})$. Also, estimating the variances of the AR and observation noise was described.

It was shown by simulation that the method can be applied repetitively to the same block of data, providing accurate results with data of moderate length.

The improvement of the technique compared with Yule-Walker is a function of the depth d to which the observations are stacked. Values even greater than p continue to yield improvement. The improvement relative to YW also is system dependent. Based on simulations, a depth $d = p + 3$, where p is the order of the system, seems a reasonable choice.

Comparisons were made against a dual Kalman filter approach, with comparable or slightly superior results.

Author Contributions: Visualization, J.H.G.; Writing—original draft, T.K.M. All authors have read and agreed to the published version of the manuscript.

Appendix A. Maximizing The Log Likelihood Function for Estimating Variances

In this section we present details on the computation of the derivative of the likelihood (12) with respect to σ_v^2. We note that

$$\frac{\partial}{\partial \sigma_\eta^2} R_w(\mathbf{b}) = 2I \qquad \frac{\partial}{\partial \sigma_v^2} R_w(\mathbf{b}) = 2\sigma_v^2 B(\mathbf{b}).$$

The derivative of the second term of (12) is

$$\frac{\partial}{\partial \sigma_\eta^2} \log \det R_w(\mathbf{b}) = \sum_{i,j} \left(\frac{\partial}{\partial R_w(\mathbf{b})} \log \det R_w(\mathbf{b}) \right)_{i,j} \left(\frac{\partial}{\partial \sigma_\eta^2} R_w(\mathbf{b}) \right)_{i,j}$$

$$= \sum_{i,j} (R_w(\mathbf{b})^{-1})_{i,j} (I)_{i,j} = \mathrm{tr}(R_w(\mathbf{b})^{-1}). \tag{A1}$$

The summation term of (12) can be written as

$$\sum_{i=1}^{k}(\mathbf{y}_i - Y_i\mathbf{b})^H R_w(b)^{-1}(\mathbf{y}_i - Y_i\mathbf{b}) = \text{tr}\left(\sum_{i=1}^{k}(\mathbf{y}_i - Y_i\mathbf{b})^H R_w(b)^{-1}(\mathbf{y}_i - Y_i\mathbf{b})\right)$$

$$= k\,\text{tr}\left(R_w(\mathbf{b})^{-1}\Sigma(\mathbf{b})\right)$$

so that the derivative is

$$\frac{\partial}{\partial\sigma_\eta^2}\sum_{i=1}^{k}(\mathbf{y}_i - Y_i\mathbf{b})^H R_w(b)^{-1}(\mathbf{y}_i - Y_i\mathbf{b}) =$$

$$k\sum_{i,j}\left(\frac{\partial}{\partial R_w(\mathbf{b})}\,\text{tr}(R_w(\mathbf{b})^{-1}\Sigma(\mathbf{b}))\right)_{i,j}\left(\frac{\partial}{\partial\sigma_\eta^2}R_w(\mathbf{b})\right)_{i,j} \tag{A2}$$

$$= k\sum_{i,j}(-R_w(\mathbf{b})^{-1}\Sigma(\mathbf{b})R_w(\mathbf{b})^{-1})^H(I)_{i,j}$$

$$= -\text{tr}(R_w(\mathbf{b})^{-1}\Sigma(\mathbf{b})R_w(\mathbf{b})^{-1}).$$

Combining (A1) and (A2) gives (14).

References

1. Makhoul, J. Spectral analysis of speech by linear prediction. *IEEE Trans. Audio Electroacoust.* **1973**, *21*, 140–148. [CrossRef]
2. Lim, J.; Oppenheim, A. All-pole modeling of degraded speech. *IEEE Trans. Signal Process.* **1978**, *26*, 197–210. [CrossRef]
3. Lim, J.; Oppenheim, A. Enhancement and bandwidth compression of noisy speech. *Proc. IEEE Inst. Electr. Electron. Eng.* **1979**, *67*, 1586–1604. [CrossRef]
4. Wrench, A.; Cowan, C. A new approach to noise-robust LPC. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '87, New York, NY, USA, 6–9 April 1987; Volume 12, pp. 305–307. [CrossRef]
5. Kay, S. *Modern Spectral Estimation: Theory and Application*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1988.
6. Deller, J.R.; Proakis, J.G.; Hansen, J.H.L. *Discrete–Time Processing of Speech Signals*; Macmillan: New York, NY, USA, 1993.
7. Ramamurthy, K.N.; Spanias, A.S. *Matlab Software for The Code Excited Linear Prediction Algorithm: The Federal Standard 1016*; Morgan & Claypool: San Rafael, CA, USA, 2009.
8. de Prony, B.R. Essai Expérimental et Analytique: Sur les Lois de la Dilatabilité de Fluides Élastiques et sur Celles de la Force Exapnsive de la Vapeur de l'eau et de la Vapeur de l'alcool, à Différentes Températures. *J. l'École Polytech.* **1795**, *1*, 24–76.
9. Ulrich, T.; Clayton, R. Time series modeling and maximum entropy. *Phys. Earth Planetary Interiors* **1976**, *12*, 188–200.
10. Capon, J. High-Resolution Frequency-Wavenumber Spectrum Analysis. *Proc. IEEE* **1969**, *57*, 1408–1418. [CrossRef]
11. Burg, J. Maximum entropy spectral analysis. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 1975.
12. Marple, L. High Resolution Autoregressive Spectrum Analysis Using Noise Power Cancellation. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '78, Tulsa, OK, USA, 10–12 April 1978; Volume 3, pp. 345–348. [CrossRef]
13. Marple, L. A New Autoregressive Spectrum Analysis Algorithm. *IEEE Trans. ASSP* **1980**, *28*, 441–454. [CrossRef]

14. Nuttall, A. *Spectral Analysis of a Univariate Process with Bad Data Points, Via Maximum Entropy and Linear Predictive Techniques*; Technical Report NUSC Tech. Rep. 5303; Naval Underwater Systems Center: New London, CT, USA, 1976.

15. Helmut Lütkepohl. *New Introduction fo Multiple Time Series Analysis*; Springer: Berlin/Heidelberg, Germany, 2005.

16. Guler, I.; Kiymik, M.; Akin, M.; Alkan, A. AR spectral analysis of EEG signals by using maximum likelihood estimation. *Comput. Biol. Med.* **2001**, *31*, 441–450. [CrossRef]

17. Baddour, K.; Beaulieu, N. Autoregressive modeling for fading channel simulation. *IEEE Trans. Wireless Comm.* **2005**, *4*, 1650–1662. [CrossRef]

18. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1978.

19. Makhoul, J. Linear Prediction: A Tutorial Review. *Proc. IEEE* **1975**, *63*, 561–580. [CrossRef]

20. Sayed, A.H. *Fundamentals of Adaptive Filtering*; Wiley Interscience: Hoboken, NJ, USA, 2003.

21. Anderson, B.D.O.; Moore, J.B. *Optimal Filtering*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1979.

22. Kay, S. The Effects of Noise on the Autoregressive Spectral Estimator. *IEEE Trans. Acoust. Speech Signal Process.* **1979**, *27*, 478–485. [CrossRef]

23. Lacoss, R.T. Data adaptive spectral analysis methods. *Geophysics* **1971**, *36*, 661–675. [CrossRef]

24. Tufts, D.; Kumaresan, R. Improved spectral resolution. *Proc. IEEE* **1980**, *68*, 419–420. [CrossRef]

25. Tufts, D.; Kumaresan, R. Improved spectral resolution II. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '80, Denver, CO, USA, 9–11 April 1980; Volume 5, pp. 592–597.

26. Kumaresan, R.; Tufts, D. Improved spectral resolution III: Efficient realization. *Proc. IEEE* **1980**, *68*, 1354–1355. [CrossRef]

27. Kumaresan, R.; Tufts, D. Accurate parameter estimation of noisy speech-like signals. In Proceedings of the Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82, Paris, France, 3–5 May 1982; Volume 7; pp. 1357–1361.

28. Kumaresan, R.; Tufts, D. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Trans. Acoust. Speech Signal Process.* **1982**, *30*, 833–840. [CrossRef]

29. Tufts, D.; Kumaresan, R. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. *Proc. IEEE* **1982**, *70*, 975–989. [CrossRef]

30. Kumaresan, R.; Tufts, D.; Scharf, L. A Prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models. *Proc. IEEE* **1984**, *72*, 230–233. [CrossRef]

31. Cadzow, J. High Performance Spectral Estimation: A New ARMA Method. *IEEE Trans. Acoust. Speech Signal Process.* **1980**, *28*, 524–529. [CrossRef]

32. Zheng, W.X. On TLS estimation of autoregressive signals with noisy measurements. In Proceedings of the SSeventh International Symposium on Signal Processing and Its Applications, Paris, France, 4 July 2003; Volume 2, pp. 287–290. [CrossRef]

33. Huffel, S.V.; Vandewalle, J. The Use of Total Linear Least Squares Technique for Identification and Parameter Estimation. In Proceedings of the IFAC/IFORS Symposium on Identification and Parameter Estimation, York, UK, 3–7 July 1985; pp. 1167–1172.

34. van Huffel, S.J.; Vandewalle, S.J.; Haegemans, A. An Efficient and Reliable Algorithm for Computing the Singular Subspace of a Matrix, Associated with Its Smallest Singular Values. *J. Comput. Appl. Math.* **1987**, *21*, 313–320. [CrossRef]

35. Huffel, S.V.; Vandewalle, J. The Partial Total Least Squares Algorithm. *J. Comput. Appl. Math.* **1988**, *21*, 333–341. [CrossRef]

36. Zheng, W. Identification of Autoregressive Signals Observed in Noise. In Proceedings of the American Control Conference, San Francisco, CA, USA, 2–4 June 1993; pp. 1229–1230.

37. Zheng, W.X. An efficient algorithm for parameter estimation of noisy AR processes. In Proceedings of the 1997 IEEE International Symposium on Circuits and Systems, ISCAS '97, Hong Kong, China, 12 June 1997; Volume 4, pp. 2509–2512.

38. Zheng, W. Estimation of autoregressive signals from noisy measurements. *IEEE Proc. Vis. Image Signal Process.* **1997**, *144*, 39–45. [CrossRef]

39. Zheng, W.X. Unbiased identification of autoregressive signals observed in colored noise. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Seattle, WA, USA, 15 May 1998; Volume 4, pp. 2329–2332.

40. Zheng, W.X. Adaptive parameter estimation of autoregressive signals from noisy observations. In Proceedings of the ICSP '98. 1998 Fourth International Conference on Signal Processing Proceedings, Beijing, China, 12–16 October 1998; Volume 1, pp. 449–452.

41. Zheng, W.X. On implementation of a least-squares based algorithm for noisy autoregressive signals. In Proceedings of the ISCAS '98, 1998 IEEE International Symposium on Circuits and Systems, Monterey, CA, USA, 31 May–3 June 1998; Volume 5, pp. 21–24.

42. Zheng, W.X. A least-squares based method for autoregressive signals in the presence of noise. *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.* **1999**, *46*, 81–85. [CrossRef]

43. Zheng, W.X. Adaptive linear prediction of autoregressive models in the presence of noise. In Proceedings of the WCCC-ICSP 2000, 5th International Conference on Signal Processing Proceedings, Beijing, China, 21–25 August 2000; Volume 1, pp. 555–558.

44. Zheng, W.X. Autoregressive parameter estimation from noisy data. *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.* **2000**, *47*, 71–75. [CrossRef]

45. Zheng, W.X. Estimation of the parameters of autoregressive signals from colored noise-corrupted measurements. *IEEE Signal Process. Lett.* **2000**, *7*, 201–204. [CrossRef]

46. Zheng, W.X. A fast convergent algorithm for identification of noisy autoregressive signals. In Proceedings of the ISCAS 2000 Geneva. The 2000 IEEE International Symposium on Circuits and Systems, Geneva, Switzerland, 28–31 May 2000; Volume 4, pp. 497–500.

47. Zheng, W.X. A new estimation algorithm for AR signals measured in noise. In Proceedings of the 2002 6th International Conference on Signal Processing, Beijing, China, 26–30 August 2002; Volume 1, pp. 186–189.

48. Zheng, W.X. An alternative method for noisy autoregressive signal estimation. In Proceedings of the ISCAS 2002. IEEE International Symposium on Circuits and Systems, Phoenix-Scottsdale, AZ, USA, 26–29 May 2002; Volume 5, pp. V–349–V–352.

49. Zheng, W.X. On unbiased parameter estimation of autoregressive signals observed in noise. In Proceedings of the 2003 International Symposium on Circuits and Systems, ISCAS '03, Bangkok, Thailand, 25–28 May 2003; Volume 4, pp. IV–261–IV–264.

50. Zheng, W.X. Fast adaptive identification of autoregressive signals subject to noise. In Proceedings of the 2004 International Symposium on Circuits and Systems, ISCAS '04, Vancouver, BC, Canada, 23–26 May 2004; Volume 3, pp. III–313–16.

51. Zheng, W.X. An efficient method for estimation of autoregressive signals in noise. In Proceedings of the 2005 IEEE International Symposium on Circuits and Systems, ISCAS 2005, Kobe, Japan, 23–26 May 2005; Volume 2, pp. 1433–1436.

52. Zheng, W.X. A new look at parameter estimation of autoregressive signals from noisy observations. In Proceedings of the 2006 IEEE International Symposium on Circuits and Systems, ISCAS 2006, Island of Kos, Greece, 21–24 May 2006; p. 4.

53. Zheng, W.X. On Estimation of Autoregressive Signals in the Presence of Noise. *IEEE Trans. Circuits Syst. II Express Briefs* **2006**, *53*, 1471–1475. [CrossRef]

54. Zheng, W.X. An Efficient Method for Estimation of Autoregressive Signals Subject to Colored Noise. In Proceedings of the 2007 IEEE International Symposium on Circuits and Systems, ISCAS 2007, New Orleans, LA, USA, 27–20 May 2007; pp. 2291–2294.

55. Xia, Y.; Zheng, W.X. On unbiased identification of autoregressive signals with noisy measurements. In Proceedings of the 2015 IEEE International Symposium on Circuits and Systems (ISCAS), Lisbon, Portugal, 24–27 May 2015; pp. 2157–2160.

56. Arablouei, R.; Dogancay, K.; Adali, T. Unbiased RLS identification of errors-in-variables models in the presence of correlated noise. In Proceedings of the 2014 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 1–5 September 2014; pp. 261–265.

57. Ai-guo, W.; Fan, Y.; Yang-Yang, Q. Bias compensation based recursive least squares identification for equation error models with colored noises. In Proceedings of the 2014 33rd Chinese Control Conference (CCC), Nanjing, China, 28–30 July 2014; pp. 6715–6720.

58. Arablouei, R.; Dogancay, K.; Adali, T. Unbiased Recursive Least-Squares Estimation Utilizing Dichotomous Coordinate-Descent Iterations. *IEEE Trans. Signal Process.* **2014**, *62*, 2973–2983.

59. Yong, Z.; Hai, T.; Zhaojing, Z. A modified bias compensation method for output error systems with colored noises. In Proceedings of the 2011 30th Chinese Control Conference (CCC), Yantai, China, 1–30 April 2011; pp. 1565–1569.

60. Wu, A.G.; Qian, Y.Y.; Wu, W.J. Bias compensation-based recursive least-squares estimation with forgetting factors for output error moving average systems. *IET Signal Process.* **2014**, *8*, 483–494. [CrossRef]

61. Arablouei, R.; Dogancay, K.; Werner, S. Recursive Total Least-Squares Algorithm Based on Inverse Power Method and Dichotomous Coordinate-Descent Iterations. *IEEE Trans. Signal Process.* **2015**, *63*, 1941–1949. [CrossRef]

62. Tomcik, J.; Melsa, J. Least squares estimation of predictor coefficients from noisy observations. In Proceedings of the 1977 IEEE Conference on Decision and Control including the 16th Symposium on Adaptive Processes and A Special Symposium on Fuzzy Set Theory and Applications, New Orleans, LA, USA, 7–9 December 1977; pp. 3–6.

63. Lee, T.S. Identification and spectral estimation of noisy multivariate autoregressive processes. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '81, Atlanta, GA, USA, 30 March–1 April 1981; Volume 6, pp. 503–507.

64. Cadzow, J.A. Autoregressive Moving Average Spectral Estimation: A Model Equation Error Procedure. *IEEE Trans. Geosci. Remote Sens.* **1981**, *GE-19*, 24–28. [CrossRef]

65. Gingras, D. Estimation of the autoregressive parameters from observations of a noise corrupted autoregressive time series. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '82, Paris, France, 3–5 May 1982; Volume 7, pp. 228–231.

66. Ahmed, M. Estimating the parameters of a noisy AR-process by using a bootstrap estimator. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '82, Paris, France, 3–5 May 1982; Volume 7, pp. 152–155.

67. Tugnait, J. Recursive parameter estimation for noisy autoregressive signals. *IEEE Trans. Inform. Theory* **1986**, *32*, 426–430. [CrossRef]

68. Paliwal, K. A noise-compensated long correlation matching method for AR spectral estimation of noisy signals. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '86, Tokyo, Japan, 7–11 April 1986; Volume 11, pp. 1369–1372.

69. Cernuschi-Frias, B.; Rogers, J. On the exact maximum likelihood estimation of Gaussian autoregressive processes. *IEEE Trans. Acoust. Speech Signal Process.* **1988**, *36*, 922–924. [CrossRef]

70. Gingras, D.; Masry, E. Autoregressive spectral estimation in additive noise. *IEEE Trans. Acoust. Speech Signal Process.* **1988**, *36*, 490–501. [CrossRef]

71. Masry, E. Almost sure convergence analysis of autoregressive spectral estimation in additive noise. *IEEE Trans. Inform. Theory* **1991**, *37*, 36–42. [CrossRef]

72. Deriche, M. AR parameter estimation from noisy data using the EM algorithm. In Proceedings of the 1994 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-94, South Australia, Australia, 19–22 April 1994; Volume iv, pp. IV/69–IV/72.

73. Lee, L.M.; Wang, H.C. An extended Levinson-Durbin algorithm for the analysis of noisy autoregressive process. *IEEE Signal Process. Lett.* **1996**, *3*, 13–15.

74. Hasan, T.; Yahagi, T. Accurate noise compensation technique for the identification of multichannel AR processes with noise. In Proceedings of the 1996 IEEE International Symposium on Circuits and Systems Connecting the World, ISCAS '96, Atlanta, GA, USA, 15 May 1996; Volume 2, pp. 401–404. [CrossRef]

75. Doblinger, G. An adaptive Kalman filter for the enhancement of noisy AR signals. In Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, ISCAS '98, Monterey, CA, USA, 31 May–3 June 1998; Volume 5, pp. 305–308.

76. Hasan, T.; Hossain, J. Multichannel autoregressive spectral estimation from noisy observations. In Proceedings of the TENCON 2000, Kuala Lumpur, Malaysia, 24–27 September 2000; Volume 1, pp. 327–332.

77. Hasan, T.; Ahmed, K. A joint technique for autoregressive spectral estimation from noisy observations. In Proceedings of the TENCON 2000, Kuala Lumpur, Malaysia, 24–27 September 2000; Volume 2, pp. 120–125.

78. Davila, C. On the noise-compensated Yule-Walker equations. *IEEE Trans. Signal Process.* **2001**, *49*, 1119–1121. [CrossRef]

79. So, H. LMS algorithm for unbiased parameter estimation of noisy autoregressive signals. *Electron. Lett.* **2001**, *37*, 536–537. [CrossRef]

80. Jin, C.Z.; Jia, L.J.; Yang, Z.J.; Wada, K. On convergence of a BCLS algorithm for noisy autoregressive process estimation. In Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, NV, USA, 10–13 December 2002; Volume 4, pp. 4252–4257.

81. Hasan, M.; Rahim Chowdhury, A.; Adnan, R.; Rahman Bhuiyan, M.; Khan, M. A new method for parameter estimation of autoregressive signals in colored noise. In Proceedings of the 2002 11th European Signal Processing Conference, Toulouse, France, 3–6 September 2002; pp. 1–4.

82. Hasan, T.; Fattah, S.; Khan, M. Identification of noisy AR systems using damped sinusoidal model of autocorrelation function. *IEEE Signal Process. Lett.* **2003**, *10*, 157–160. [CrossRef]

83. Hasan, M.; Chowdhury, A.; Khan, M. Identification of autoregressive signals in colored noise using damped sinusoidal model. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2003**, *50*, 966–969. [CrossRef]

84. Diversi, R.; Guidorzi, R.; Soverini, U. Identification of ARX models with noisy input and output. In Proceedings of the 2007 European Control Conference (ECC), Kos, Greece 2–5 July 2007; pp. 4073–4078.

85. Fattah, S.; Zhu, W.; Ahmad, M. Identification of autoregressive moving average systems from noise-corrupted observations. In Proceedings of the 2008 Joint 6th International IEEE Northeast Workshop on Circuits and Systems and TAISA Conference, NEWCAS-TAISA 2008, Montreal, QC, Canada, 22–25 June 2008; pp. 69–72.

86. Fattah, S.; Zhu, W.; Ahmad, M. A correlation domain algorithm for autoregressive system identification from noisy observations. In Proceedings of the 51st Midwest Symposium on Circuits and Systems, MWSCAS 2008, Knoxville, TN, USA, 10–13 August 2008; pp. 934–937.

87. Babu, P.; Stoica, P.; Marzetta, T. An IQML type algorithm for AR parameter estimation from noisy covariance sequences. In Proceedings of the 2009 17th European Signal Processing Conference, Glasgow, UK, 24–28 August 2009; pp. 1022–1026.

88. Qu, X.; Zhou, J.; Luo, Y. An Advanced Parameter Estimator of Multichannel Autoregressive Signals from Noisy Observations. In Proceedings of the 2009 International Conference on Information Engineering and Computer Science, ICIECS 2009, Wuhan, China, 19–20 December 2009; pp. 1–4.

89. Weruaga, L.; Al-Ahmad, H. Frequency-selective autoregressive estimation in noise. In Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, TX, USA, 15–19 March 2010; pp. 3854–3857.

90. Mahmoudi, A.; Karimi, M. Parameter estimation of noisy autoregressive signals. In Proceedings of the 2010 18th Iranian Conference on Electrical Engineering (ICEE), Isfahan, Iran, 11–13 May 2010; pp. 145–149.

91. Weruaga, L. Two methods for autoregressive estimation in noise. In Proceedings of the 2011 IEEE GCC Conference and Exhibition (GCC), Dubai, United Arab Emirates, 19–22 February 2011; pp. 501–504. [CrossRef]

92. Youcef, A.; Diversi, R.; Grivel, E. Errors-in-variables identification of noisy moving average models. In Proceedings of the 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 31 August–4 September 2015; pp. 968–972.

93. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 1994.

94. Kay, S. Recursive Maximum Likelihood Estimation of Autoregressive Processes. *IEEE Trans. Acoust. Speech Signal Process.* **1983**, *31*, 292–303. [CrossRef]

95. Evans, A.; Fischl, R. Optimal least squares time-domain synthesis of recursive digital filters. *IEEE Trans. Audio Electroacoust.* **1973**. [CrossRef]

96. Bresler, Y.; Macovski, A. Exact maximum likelihood parameter estimation of superimposed exponentials in noise. *IEEE Trans. Acoust. Speech Signal Process.* **1986**, *34*, 1081–1089. [CrossRef]

97. Steiglitz, K.; McBride, L. A technique for the identification of linear systems. *IEEE Trans. Autom. Control* **1965**, *10*, 461–464. [CrossRef]

98. Kumaresan, R.; Scharf, L.; Shaw, A. An Algorithm for pole-zero modeling and spectral analysis. *IEEE Trans. Acoust. Speech Signal Process.* **1986**, *34*, 637–640. [CrossRef]

99. Anderson, J.; Giannakis, G. Noisy input/output system identification using cumulants and the Steiglitz-McBride algorithm. In Proceedings of the 1991 Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 4–6 November 1991; Volume 1, pp. 608–612.

100. Fuller, W.A. *Measurement Error Models*; Wiley: Hoboken, NJ, USA, 1987.

101. Petitjean, J.; Grivel, E.; Bobillet, W.; Roussilhe, P. Multichannel AR parameter estimation from noisy observations as an errors-in-variables issue. In Proceedings of the 2008 16th European Signal Processing Conference, Lausanne, Switzerland, 25–29 August 2008; pp. 1–5.

102. Petitjean, J.; Diversi, R.; Grivel, E.; Guidorzi, R.; Roussilhe, P. Recursive errors-in-variables approach for AR parameter estimation from noisy observations. Application to radar sea clutter rejection. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009, Taipei, Taiwan, 19–24 April 2009; pp. 3401–3404.

103. Petitjean, J.; Grivel, E.; Diversi, R.; Guidorzi, R. A recursive errors-in-variables method for tracking time varying autoregressive parameters from noisy observations. In Proceedings of the 2010 18th European Signal Processing Conference, Aalborg, Denmark, 23–27 August 2010; pp. 840–844.

104. Labarre, D.; Grivel, E.; Berthoumieu, Y.; Todini, E.; Najim, M. Consistent estimation of autoregressive parameters from noisy observations based on two interacting Kalman filters. *Signal Process.* **2006**, *86*, 2863–2876. [CrossRef]

105. Jamoos, A.; Grivel, E.; Christov, N.; Najim, M. Estimation of autoregressive fading channels based on two cross-coupled H_∞ filters. *Signal Image Video Process. J. Springer* **2009**, *3*, 209–216. [CrossRef]

106. Jamoos, A.; Grivel, E.; Shakarneh, N.; Abdel-Nour, H. Dual Optimal filters for parameter estimation of a multivariate autoregressive process from noisy observations. *IET Signal Processing* **2011**, *5*, 471–479. [CrossRef]

107. Bugallo, M.F.; Martino, L.; Corander, J. Adaptive importance sampling in signal procesing. *Digit. Signal Process.* **2015**, *47*, 36–49. [CrossRef]

108. Urteaga, I.; Djuric, P.M. Sequential estimation of Hidden ARMA Processes by Particle Filtering—Part I. *IEEE Trans. Signal Process.* **2017**, *65*, 482–493. [CrossRef]

109. Urteaga, I.; Djuric, P.M. Sequential estimation of Hidden ARMA Processes by Particle Filtering—Part II. *IEEE Trans. Signal Process.* **2017**, *65*, 494–504. [CrossRef]

110. Martino, L.; Elvira, V.; Camps-Valls, G. Distributed Particle Metropolis-Hastings Schemes. In *IEEE Statistical Signal Processing Workshop*; IEEE: Piscataway, NJ, USA, 2018; pp. 553–557.

111. Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*; Prentice-Hall: Upper Saddle River, NJ, 1993.

112. Farahmand, S.; Giannakis, G.B. Robust RLS in the Presence of Correlated Noise Using Outlier Sparsity. *IEEE Trans. Signal Process.* **2012**, *60*, 3308–3313. [CrossRef]

113. Grant, M.; Boyd, S. CVX: Matlab Software for Disciplined Convex Programming, Version 1.21. 2011. Available online: http://cvxr.com/cvx (accessed on 19 May 2020).

114. Moon, T.K.; Stirling, W.C. *Mathematical Methods and Algorithms for Signal Processing*; Prentice-Hall: Upper Saddle River, NJ, USA, 2000.