Signal Processing 186 (2021) 108118

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Maximum likelihood autoregressive model parameter estimation with noise corrupted independent snapshots

Ömer Çayır*, Çağatay Candan

Department of Electrical and Electronics Engineering, Middle East Technical University (METU), Ankara 06800, Turkey

ARTICLE INFO

Article history:
Received 22 September 2020
Revised 31 March 2021
Accepted 8 April 2021
Available online 20 April 2021

Keywords:
Autoregressive process
Autoregressive model parameter estimation
Multiple snapshots
Expectation-maximization
Parametric spectrum estimation
Spectrum estimation

ABSTRACT

Maximum likelihood autoregressive (AR) model parameter estimation problem with independent snapshots observed under white Gaussian measurement noise is studied. In addition to the AR model parameters, the measurement noise variance is also included among the unknowns of the problem to develop a general solution covering several special cases such as the case of known noise variance, noise-free snapshots, the single snapshot operation etc. The presented solution is based on the expectation-maximization method which is formulated by assigning the noise-free snapshots as the missing data. An approximate version of the suggested method, at a significantly reduced computational load with virtually no loss of performance, has also been developed. Numerical results indicate that the suggested solution brings major performance improvements in terms of estimation accuracy and does not suffer from unstable AR filter estimates unlike some other methods in the literature. The suggested method can be especially useful for small-dimensional multiple-snapshot noisy AR modeling applications such as the clutter power spectrum modeling application in radar signal processing.

1. Introduction

Autoregressive (AR) modeling of random signals is being used in many areas associated with the statistical signal processing [1] such as radar signal processing, speech processing and biomedical signal processing [2–5]. The richness of application venues for AR models can be attributed to their success in representation and also to the availability of efficient methods for model parameter estimation. In this work, we consider the parameter estimation of AR processes observed under white noise. Our main goal is to extend the maximum-likelihood like AR model parameter estimator developed for a single noise-free snapshot in Candan [6] to the operation with multiple snapshots corrupted by white noise. A computationally efficient version of the suggested method is also presented.

Model parameter estimation of an AR process observed under noise (noisy AR parameter estimation problem) is prone to estimator bias and statistical efficiency problems when modeling assumptions are not carefully taken into account. For example, AR parameter estimates obtained from Yule–Walker (YW) equations are typically biased due to the bias of the zero-lag term of the autocorrelation introduced by white noise [7]. Furthermore, when the variance of the noise corrupting the AR process is not known, an asymptotic Cramér–Rao bound (CRB) study by Weruaga et al. reveals that the joint estimation of the autoregressive signal variance and noise variance is not a well-conditioned problem [8,9]. In spite of these setbacks, several practical solutions have been developed for the noisy AR parameter estimation problem in the literature. For example, since the autoregressive moving average (ARMA) model also characterizes the noisy AR processes, it is possible to apply ARMA modeling approaches such as the maximum likelihood [10], the modified YW [11] or the recursive prediction error [12] for the solution of noisy AR parameter estimation problem. In addition, a number of improved least-squares (LS) solutions are suggested to compensate the bias on the parameter estimates due to the measurement noise [13–15]. The main challenge for the bias compensating solutions is the estimation of the measurement noise variance. The solutions based on the eigendecomposition [14] and the inverse filtering coupled with YW equations [15] have been suggested for this purpose. Among other solutions, we can list a subspace based solution [16], an errors-in-variables approach utilizing both low and high order YW equations [17], a nonlinear optimization (for the estimation measurement noise variance) solution [18], a solution with two interacting Kalman filters [19] and some adaptive filtering type solutions [20–22]. A particularly interesting solution is the method presented in Gabrea et al. [23], based on the approach developed by Mehra [24], that avoids the estimation of the process and measurement noise vari-

* Corresponding author.
E-mail addresses: ocayir@metu.edu.tr (Ö. Çayır), ccandan@metu.edu.tr (Ç. Candan).

https://doi.org/10.1016/j.sigpro.2021.108118
0165-1684/

ances. Recently, four novel methods have been proposed in Esfandiari et al. [25]. The first one utilizes the null space of AR parameter vector, the second one solves a constrained LS problem, the third one reduces the parameter estimation problem for an AR(P) process to a problem of estimating two parameters and the fourth one is based on the eigendecomposition of enlarged autocorrelation matrix.

The main aim of this work is to study the noisy AR parameter estimation problem in the presence of multiple independent snapshots. We solely focus on the parameter estimation problem for the scalar AR processes, that is the vector-AR processes, in which the regression (recursion) involves a linear combination previous vector-outputs and current input-vector [26], are not within the scope of this study. We apply the expectation-maximization (EM) method [27] by assigning the noise-free snapshots as the missing data to develop a solution. In the maximization step (M-step) of EM method, we transform the AR parameter estimation problem into a form that can be solved by using an approach similar to the recent work in Candan [6] which is a two-stage method utilized for AR parameter estimation for a single noise-free snapshot. For the calculation of the expectation step (E-step), we describe an approximate, yet highly efficient, method for the computational load reduction.

The EM method has been previously applied for the solution of noisy AR parameter in Deriche [28] and some computational simplifications in the M-step have been suggested. Different from [28], we formulate the M-step such that it is possible to extend the maximum likelihood like estimator given in Candan [6] to the multiple snapshot setting and also describe some novel computational load reduction methods for the E-step. In [29], a related EM-based method utilizing Kalman filters is presented for the colored Gaussian noise. It is well known that the performance of EM algorithm is sensitive to the initialization, that is EM iterations can converge to a local maximum, instead of the global maximum, due to a poor initialization [30]. In this study, we present an initialization method for the suggested method (see Appendix A), consider the cases of known/unknown measurement noise variance individually and present detailed comparisons with the alternative estimators and CRBs derived in Weruaga and Melko [8], Weruaga and Dimitrov [9].

In the literature, there are several works, including [31–33], which are based on the Whittle likelihood [34], a frequency-domain approximation to the exact likelihood function. While Whittle likelihood maximization is computationally easier, the resulting parameter estimates with finite sample sizes are biased especially for short data records [33]. As the sample size increases, time- and frequency-domain solutions yield similar results [35]. Further discussions on time- and frequency-domain approaches can be found in Weruaga and Dimitrov [9]. In this study, we focus on time-domain approach in relation with our main goal of extending the maximum likelihood like estimator in Candan [6] to the noisy, multiple snapshot setting.

The AR modeling has several important applications in speech processing area. In this area, the typical data size can be hundreds of samples and frequency domain methods can be utilized in this large sample-size regime without any performance worries. In some other applications, such as the radar signal processing applications, the data size can be much smaller and data collection mechanism can operate intermittently, in contrast to the continuous data-collection modality in speech processing, leading to multiple, short data-length (snapshot) observations [36–38]. For example, in the clutter (the unwanted echoes received by radar systems [36]) cancellation application of radar signal processing, the clutter power spectral density is estimated from a collection of snapshots [36]. Each entry of the snapshot vector is formed by a radar pulse return from a particular range cell. The number of transmitted pulses, which is the dimension of the snapshot vector, affects all subsequent radar operations and it can be as few as 10-20 pulses due to other constraints [39]. For this application, the structured estimation of the clutter power spectrum from small dimensional multiple snapshots becomes a necessity. Some solutions to this problem, in addition to the examined maximum likelihood solution, are the multiple-snapshot version of the Burg's method [37] or the multiple-snapshot version of any other AR parameter estimation method given in Stoica and Moses [40]. In this study, we consider the AR parameter estimation problem specifically for small-dimensional multiple snapshots and pursue an exact time-domain maximum likelihood parameter estimation solution.

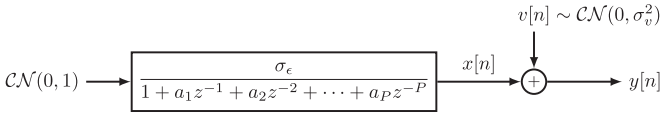


Fig. 1. AR process samples $x[n]$ corrupted by the observation noise $v[n]$ to form the observed sequence $y[n]$.

The main contributions of the study are as follows: 1. Expressing the conventional EM formulation as a multiple-snapshot, noise-free AR parameter estimation problem (Section 2.1); 2. Extension of efficient single snapshot noise-free AR parameter estimation given in Candan [6] to the multiple-snapshot case and its application in the solution of EM problem (Section 2.2); 3. An approximate, but efficient version of the proposed solution by utilizing a matrix inversion free Kalman smoother [41, Sec. 5.2.4] and the Gohberg-Semencul formula [40, Sec. 3.9.4] (Section 2.3).

The notation utilized in this paper is as follows: Scalars, column vectors and matrices are denoted by italic lowercase, boldface lowercase and boldface uppercase letters, respectively. The conjugate, transpose, conjugate transpose and inverse operators are denoted by $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$, respectively. The density of the zero-mean complex-valued (circular symmetric) white Gaussian noise with variance σ^2 is denoted by $\mathcal{CN}(0, \sigma^2)$. $\mathbf{I_N}$ denotes the $N \times N$ identity matrix and $j = \sqrt{-1}$. $\mathbf{0}_N$ and $\mathbf{0}_{N\times N}$ denote the N dimensional column vector and the $N \times N$ matrix, respectively, with all entries being zero. Euclidean norm and trace are denoted by $\|\cdot\|$ and $\mathrm{tr}(\cdot)$, respectively. For a scalar c, $|c|$ is the absolute value of c. For a square matrix $\mathbf{S}$, $|\mathbf{S}|$ is the determinant of $\mathbf{S}$. The n'th entry of the vectors $\mathbf{x}$ and $\mathbf{x}_\ell$ are denoted as x_n and $x_{\ell,n}$, respectively. For a positive integer P, $\mathbf{x}_{\ell,n:n+P} = [\, x_{\ell,n}\ x_{\ell,n+1}\ \ldots\ x_{\ell,n+P}\,]^T$ and $\mathbf{x}_{\ell,n:-1:n-P} = [\, x_{\ell,n}\ x_{\ell,n-1}\ \ldots\ x_{\ell,n-P}\,]^T$. The i'th row and j'th column entry of matrix $\mathbf{M}$ is denoted as $[\mathbf{M}]_{ij}$.

2. Noisy AR parameter estimation problem

Consider the transfer function, given below, for the generation of AR(P) process

$$H(z) = \frac{\sigma_\epsilon}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_P z^{-P}} = \frac{\sigma_\epsilon}{A(z)}, \tag{1}$$

where $a_1, a_2, \ldots, a_P$ can be either the real- or complex-valued constants, and σ_ϵ is a real-valued constant scaling the input.

The filter $H(z)$ is assumed to be excited with zero-mean, unit variance complex-valued (circular symmetric) white Gaussian noise as shown in Fig. 1. The filter $H(z)$ is assumed to be stable so that $x[n]$ in Fig. 1 is a wide-sense stationary process. The output of the filter at steady-state is denoted as $x[n]$. The l'th snapshot vector $\mathbf{x}_l$ with dimension $N \times 1$ is formed by concatenating consecutive $x[n]$ samples. It is assumed that a total number of L snapshot vectors are collected where each snapshot is independent and identically distributed (iid) Gaussian vector.

The autocorrelation matrix of the snapshot $\mathbf{x}_l$ is denoted as $\mathbf{R} = \mathbf{R}_{f,N}\sigma_\epsilon^2$ where $\mathbf{R}_{f,N}$ is an $N \times N$ Hermitian Toeplitz matrix

whose first column entries are $r_f[k] = E\{x[n]x^*[n-k]\}/\sigma_\epsilon^2$ for $k = \{0, 1, \ldots, N-1\}$:

$$\mathbf{R}_{f,N} = \begin{bmatrix} r_f[0] & r_f[-1] & \ldots & r_f[-N+1] \\ r_f[1] & r_f[0] & \ldots & r_f[-N+2] \\ \vdots & \vdots & \ddots & \vdots \\ r_f[N-1] & r_f[N-2] & \ldots & r_f[0] \end{bmatrix}. \tag{2}$$

In this study, it is assumed that snapshot vectors are observed under independent additive white Gaussian noise, i.e., the l'th observation vector is $\mathbf{y}_l = \mathbf{x}_l + \mathbf{v}_l$ as in Fig. 1, where $\mathbf{v}_l$ vectors are an iid circular symmetric Gaussian distributed vector with zero-mean and covariance matrix $\sigma_v^2\mathbf{I_N}$. Hence, the sample signal-to-noise ratio is $\text{SNR} = \sigma_\epsilon^2 r_f[0]/\sigma_v^2$. The general AR parameter estimation problem is the estimation of $\Theta = \{\sigma_v^2, \sigma_\epsilon^2, \mathbf{a}\}$, where $\mathbf{a} = [\, a_1\, a_2\, \ldots\, a_P\,]^T$, given the observation vectors $\mathbf{y}_l$, $l = \{1, 2, \ldots, L\}$. If the noise variance σ_v^2 is known, the unknown parameter set reduces to $\Theta = \{\sigma_\epsilon^2, \mathbf{a}\}$. For this case, the true value for the noise variance can be substituted for its estimate in the formulation given below. If the noise variance is known to be zero, the problem becomes AR parameter estimation with multiple noise-free snapshots. The solution of this problem can be retrieved as the limiting case of the discussion as $\sigma_v^2 \to 0$.

2.1. EM formulation for noisy AR parameter estimation with multiple snapshots

For the solution of the problem, we apply the expectation-maximization method. To do that, we define $N \times L$ matrices $\mathbf{X} = [\mathbf{x}_1 \ldots \mathbf{x}_L]$, $\mathbf{Y} = [\mathbf{y}_1 \ldots \mathbf{y}_L]$ to denote the noise-free AR process snapshots and observations, respectively. The snapshot matrix $\mathbf{X}$ is the latent variable of the problem. The $N \times 2L$ dimensional $\mathbf{Z} = [\mathbf{X}\ \mathbf{Y}]$ is the complete data matrix. The log-likelihood of complete data matrix $\Lambda(\mathbf{Z}, \Theta) = \log f(\mathbf{X}, \mathbf{Y}) = \log f(\mathbf{Y}|\mathbf{X}) + \log f(\mathbf{X})$ can be written as

$$\Lambda(\mathbf{Z}, \Theta) \stackrel{c}{=} -\sum_{l=1}^{L} \left\{ N\log(\sigma_v^2) + \frac{\|\mathbf{y}_l - \mathbf{x}_l\|^2}{\sigma_v^2} + \log(|\sigma_\epsilon^2 \mathbf{R}_{f,N}|) + \frac{\mathbf{x}_l^H \mathbf{R}_{f,N}^{-1} \mathbf{x}_l}{\sigma_\epsilon^2} \right\}, \tag{3}$$

where first two and last two terms of summation correspond to the terms $\log f(\mathbf{Y}|\mathbf{X})$ and $\log f(\mathbf{X})$, respectively. The symbol $\stackrel{c}{=}$ denotes the equality of both sides apart from constant terms.

The EM method has two steps: In the expectation step, the expected value of complete log-likelihood is calculated with respect to the posterior density of the latent variables. The expectation operation can be written as $J(\Theta) = E\{\Lambda(\mathbf{Z}, \Theta)|\mathbf{Y}, \Theta^{\text{old}}\}$. Here $\Theta^{\text{old}} = \{(\sigma_v^2)^{\text{old}}, (\sigma_\epsilon^2)^{\text{old}}, \mathbf{a}^{\text{old}}\}$ are the current estimates for the unknown parameters to be updated. In the second step, the expectation result is maximized with respect to the unknown parameters, $\Theta^{\text{new}} = \text{argmax}_\Theta J(\Theta)$, to update the unknown parameters. The algorithm is initiated with a proper Θ^{old} value for the implementation of the first step (posterior calculation) and iteratively run by using Θ^{new} of an earlier iteration as the Θ^{old} of the next iteration until the convergence of the estimates.

To execute the expectation step, we need the posterior density of latent variables, $f(\mathbf{x}_l|\mathbf{y}_l)$, $l = \{1, 2, \ldots, L\}$. It is well known that the posterior density $f(\mathbf{x}_l|\mathbf{y}_l)$ is the Gaussian vector with mean vector $\widehat{\mathbf{x}}_l$ and covariance matrix $\mathbf{K}$ given as

$$\begin{aligned} \widehat{\mathbf{x}}_l &= (\sigma_\epsilon^2)^{\text{old}} \mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) \left((\sigma_\epsilon^2)^{\text{old}} \mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) + (\sigma_v^2)^{\text{old}} \mathbf{I}_N \right)^{-1} \mathbf{y}_l, \\ \mathbf{K} &= (\sigma_\epsilon^2)^{\text{old}} \mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) - (\sigma_\epsilon^4)^{\text{old}} \mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) \Big((\sigma_\epsilon^2)^{\text{old}} \mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) \\ &\quad + (\sigma_v^2)^{\text{old}} \mathbf{I_N} \Big)^{-1} \mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}). \end{aligned} \tag{4}$$

Taking the expectation of the complete log-likelihood function in (3) with respect to the posterior density results in $J(\Theta) = E\{\Lambda(\mathbf{Z}, \Theta)|\mathbf{Y}, \Theta^{\text{old}}\}$

$$J(\sigma_v^2, \sigma_\epsilon^2, \mathbf{a}) \stackrel{c}{=} -\sum_{l=1}^{L} \left\{ N\log(\sigma_v^2) + \frac{\|\mathbf{y}_l - \widehat{\mathbf{x}}_l\|^2 + \text{tr}(\mathbf{K})}{\sigma_v^2} + \log(|\sigma_\epsilon^2 \mathbf{R}_{f,N}|) + \frac{\widehat{\mathbf{x}}_l^H \mathbf{R}_{f,N}^{-1} \widehat{\mathbf{x}}_l + \text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})}{\sigma_\epsilon^2} \right\}, \tag{5}$$

which completes the first step (expectation step) of EM method.

The maximization step of EM starts with the partial derivative calculation of $J(\sigma_v^2, \sigma_\epsilon^2, \mathbf{a})$, given in (5), with respect to σ_v^2. Setting the result equal to zero, we get the update expression for the measurement noise variance σ_v^2 estimate

$$(\sigma_v^2)^{\text{new}} = \frac{L\,\text{tr}(\mathbf{K}) + \sum_{l=1}^{L} \|\mathbf{y}_l - \widehat{\mathbf{x}}_l\|^2}{LN}. \tag{6}$$

Similarly, by taking the partial derivative of $J(\sigma_v^2, \sigma_\epsilon^2, \mathbf{a})$ with respect to σ_ϵ^2 and equating the result to zero, we get the update expression for the process noise variance σ_ϵ^2 estimate

$$(\sigma_\epsilon^2)^{\text{new}} = \frac{L\,\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K}) + \sum_{l=1}^{L} \widehat{\mathbf{x}}_l^H \mathbf{R}_{f,N}^{-1} \widehat{\mathbf{x}}_l}{LN}. \tag{7}$$

Substituting $(\sigma_v^2)^{\text{new}}$ and $(\sigma_\epsilon^2)^{\text{new}}$ given in (6) and (7), respectively, into (5), we get the compressed expected log-likelihood function as

$$\begin{aligned} J((\sigma_v^2)^{\text{new}}, (\sigma_\epsilon^2)^{\text{new}}, \mathbf{a}) &\stackrel{c}{=} -\log|\mathbf{R}_{f,N}| \\ &\quad -N\log\left(L\,\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K}) + \sum_{l=1}^{L} \widehat{\mathbf{x}}_l^H \mathbf{R}_{f,N}^{-1} \widehat{\mathbf{x}}_l \right), \end{aligned} \tag{8}$$

where $\mathbf{R}_{f,N}$ is a function of unknown parameter vector $\mathbf{a} = [\, a_1\, a_2\, \ldots\, a_P\,]^T$. The maximization of (8) with respect to $\mathbf{a}$ is more challenging than earlier steps and is the main challenge of the problem. Fortunately, by expressing $L\,\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ as a quadratic term, the function (8) can be converted into a form similar to that of single noise-free snapshot case given in Candan [6] and the approach given therein can be utilized to maximize (8) with respect to $\mathbf{a}$.

To express $L\,\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ as a quadratic term, we introduce the eigendecomposition of covariance matrix $\mathbf{K} = \sum_{n=1}^{N} \lambda_n \mathbf{e}_n \mathbf{e}_n^H$ into the problem. Here λ_n and $\mathbf{e}_n$ is an eigenvalue and associated unit norm eigenvector of matrix $\mathbf{K}$. It should be noted that we can find an orthonormal set of the eigenvectors $\mathbf{e}_n$, since the matrix $\mathbf{K}$ is Hermitian. With the eigendecomposition, it is possible to express $\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ as $\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K}) = \sum_{n=1}^{N} \lambda_n \mathbf{e}_n^H \mathbf{R}_{f,N}^{-1} \mathbf{e}_n$. By introducing the scaled versions of eigenvectors $\mathbf{e}_n$ as $\bar{\mathbf{e}}_n = \sqrt{L\lambda_n}\mathbf{e}_n$, we can further simplify this expression to $L\,\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K}) = \sum_{n=1}^{N} \bar{\mathbf{e}}_n^H \mathbf{R}_{f,N}^{-1} \bar{\mathbf{e}}_n$. Hence, the maximization of compressed likelihood relation in (8) is equivalent to the following minimization problem

$$\begin{aligned} \mathbf{a}^{\text{new}} &= \underset{\mathbf{a}}{\text{argmin}}\, \frac{1}{N}\log|\mathbf{R}_{f,N}| + \log\left(\sum_{n=1}^{N} \bar{\mathbf{e}}_n^H \mathbf{R}_{f,N}^{-1} \bar{\mathbf{e}}_n + \sum_{l=1}^{L} \widehat{\mathbf{x}}_l^H \mathbf{R}_{f,N}^{-1} \widehat{\mathbf{x}}_l \right) \\ &= \underset{\mathbf{a}}{\text{argmin}}\, \frac{1}{N}\log|\mathbf{R}_{f,N}| + \log\left(\sum_{\ell=1}^{L_a=L+N} \mathbf{f}_\ell^H \mathbf{R}_{f,N}^{-1}(\mathbf{a}) \mathbf{f}_\ell \right). \end{aligned} \tag{9}$$

In the second line of (9), we combine the sums in the argument of logarithm by augmenting the set of $\widehat{\mathbf{x}}_l$ vectors $l = \{1, 2, \ldots, L\}$ with $\bar{\mathbf{e}}_n$ vectors $n = \{1, 2, \ldots, N\}$ to form a set of vectors with $L_a = L + N$ elements where $\mathbf{f}_{L+n} = \bar{\mathbf{e}}_n$.

Assuming that the solution of optimization problem in (9), that is $\mathbf{a}^{\text{new}}$, is available; the remaining unknown parameter can be

estimated from (7) as

$$(\sigma_\epsilon^2)^{\text{new}} = \frac{L\,\text{tr}(\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\mathbf{K}) + \sum_{l=1}^{L}\widehat{\mathbf{x}}_l^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\widehat{\mathbf{x}}_l}{LN} = \frac{\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{new}})\mathbf{f}_\ell}{LN}. \tag{10}$$

Hence, the crux of the parameter problem is the solution of optimization problem given in (9).

When compared with the minimization problem in Candan [6, Eq. 11] having the cost function of $\frac{1}{N}\log|\mathbf{R}_{f,N}| + \log(\mathbf{x}^H\mathbf{R}_{f,N}^{-1}\mathbf{x})$, we see that cost function of the minimization problem in (9) differs from the earlier one with the inclusion of $L_a = L + N$ snapshots instead of a single one. We present the details for this extension in the following section. We reiterate that some of the snapshots in this formulation are generated from the eigendecomposition of $\mathbf{K}$ matrix and augmented to actual snapshots, called observation vectors, to facilitate a solution similar to the one in Candan [6]. In Section 2.3, we present a reduced complexity implementation alternative for this solution. Different methods of initialization for the EM method are provided in Appendix A.

2.2. AR parameter estimation problem with multiple noise-free snapshots

The method in Candan [6] is an efficient method for AR parameter estimation which is developed for a single snapshot under the noiseless observation scenario. This method can be considered as an alternative for numerical search based maximum likelihood estimator having much higher complexity and it is shown the method performs very similar to the maximum likelihood estimator in many scenarios. In this section, we present the multiple-snapshot extension of this method.

2.2.1. First stage: weighted forward-backward prediction with L_a snapshots

Following [6], we ignore $\log|\mathbf{R}_{f,N}|$ term in (9) and use the weighted forward-backward prediction method to generate the first stage estimate of $\mathbf{a}$, $\mathbf{a}_{\text{FS}}$, as follows:

$$\mathbf{a}_{\text{FS}} = \underset{\mathbf{a}}{\text{argmin}}\sum_{\ell=1}^{L_a}\left(\sum_{n=1}^{N-P} w_b[n]|e_{\ell,b}[n]|^2 + \sum_{n=P+1}^{N} w_f[n]|e_{\ell,f}[n]|^2\right), \tag{11}$$

where $e_{\ell,f}[n] = f_{\ell,n} + \mathbf{a}^T\mathbf{f}_{\ell,n-1:-1:n-P}$ and $e_{\ell,b}[n] = f_{\ell,n} + \mathbf{a}^H\mathbf{f}_{\ell,n+1:n+P}$ are the forward and backward prediction errors with weights $w_f[n] = n - P$ and $w_b[n] = N - P + 1 - n$, respectively. The problem in (11) can be solved by introducing the linear equation systems, $\mathbf{A}_{\ell,f}\mathbf{a} = -\mathbf{b}_{\ell,f}$ and $\mathbf{A}_{\ell,b}\mathbf{a} = -\mathbf{b}_{\ell,b}$ for $\ell = 1,\ldots,L_a$, given below, generating the forward and backward prediction errors:

$$\underbrace{\begin{bmatrix} f_{\ell,P} & f_{\ell,P-1} & \cdots & f_{\ell,1} \\ f_{\ell,P+1} & f_{\ell,P} & \cdots & f_{\ell,2} \\ \vdots & \vdots & \cdots & \vdots \\ f_{\ell,N-1} & f_{\ell,N-2} & \cdots & f_{\ell,N-P} \end{bmatrix}}_{\mathbf{A}_{\ell,f}}\underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}}_{\mathbf{a}} = -\underbrace{\begin{bmatrix} f_{\ell,P+1} \\ f_{\ell,P+2} \\ \vdots \\ f_{\ell,N} \end{bmatrix}}_{\mathbf{b}_{\ell,f}},$$

$$\underbrace{\begin{bmatrix} f^*_{\ell,N-P+1} & f^*_{\ell,N-P+2} & \cdots & f^*_{\ell,N} \\ f^*_{\ell,N-P} & f^*_{\ell,N-P+1} & \cdots & f^*_{\ell,N-1} \\ \vdots & \vdots & \cdots & \vdots \\ f^*_{\ell,2} & f^*_{\ell,3} & \cdots & f^*_{\ell,P+1} \end{bmatrix}}_{\mathbf{A}_{\ell,b}}\underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}}_{\mathbf{a}} = -\underbrace{\begin{bmatrix} f^*_{\ell,N-P} \\ f^*_{\ell,N-P-1} \\ \vdots \\ f^*_{\ell,1} \end{bmatrix}}_{\mathbf{b}_{\ell,b}}.$$

Using the introduced matrices, the final result of the first stage becomes

$$\mathbf{a}_{\text{FS}} = -\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^H\mathbf{W}\mathbf{A}_{\ell,f} + \mathbf{A}_{\ell,b}^H\mathbf{W}\mathbf{A}_{\ell,b}\right)\right)^{-1}\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^H\mathbf{W}\mathbf{b}_{\ell,f} + \mathbf{A}_{\ell,b}^H\mathbf{W}\mathbf{b}_{\ell,b}\right)\right), \tag{12}$$

where $\mathbf{W}$ is the diagonal matrix with the diagonal entries of $w_f[n]$ for $n = \{P+1, P+2, \ldots, N\}$, i.e.,

2.2.2. Second stage: maximizing likelihood around $\mathbf{a}_{FS}$

The second stage takes into account the term $\log|\mathbf{R}_{f,N}|$ in (9), [6]. The nonlinear function is expanded into Taylor series at the operating point of $\mathbf{a} = \mathbf{a}_{FS}$ and a quadratic approximation for both terms of the sum forming the cost function of (9) is formed. We note that both $|\mathbf{R}_{f,N}|$ and $\mathbf{R}_{f,N}^{-1}$ are highly nonlinear functions of $\mathbf{a}$.

For the *quadratic approximation* of $\log|\mathbf{R}_{f,N}|$, we can follow the procedure given in Candan [6, Eq. 11]. The procedure utilizes the expression $|\mathbf{R}_{f,N}| = |\mathbf{R}_{f,P}| = \prod_{i=1}^{P}(1-|k_i|^2)^{-i}$ for $N \geq P$, Hayes [1], Kay [42] that connects the autocorrelation matrix determinant of an AR(P) process to the reflection coefficients $\mathbf{k} = [\,k_1\,k_2\,\ldots\,k_P\,]^T$ of its synthesis filter. The log-determinant $\log|\mathbf{R}_{f,N}| = -\sum_{i=1}^{P} i\log(1-|k_i|^2)$ is approximated via Taylor series at the expansion point of the reflection coefficient vector $\mathbf{k}$ that corresponds to the first stage estimate $\mathbf{a}_{FS}$ and introducing a perturbation vector $\boldsymbol{\delta}_{\mathbf{k}}$ as in Candan [6, Eq. 16].

For the *quadratic approximation* of $\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell$, we can use [6, Eq. 21] directly

$$\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell = \sum_{\ell=1}^{L_a}\left(\|\mathbf{b}_{\ell,1} + \mathbf{M}_{\ell,1}(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_c\boldsymbol{\delta}_{\mathbf{k}}^*)\|^2 - \|\mathbf{b}_{\ell,2} + \mathbf{M}_{\ell,2}(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_c\boldsymbol{\delta}_{\mathbf{k}}^*)\|^2\right), \tag{13}$$

where $\mathbf{M}_{\ell,1}$ and $\mathbf{M}_{\ell,2}$ are Hankel and Toeplitz matrices, respectively, with the definitions of

$$\mathbf{M}_{\ell,1} = \begin{bmatrix} f^*_{\ell,2} & f^*_{\ell,3} & f^*_{\ell,4} & \cdots & f^*_{\ell,P+1} \\ f^*_{\ell,3} & f^*_{\ell,4} & f^*_{\ell,5} & \cdots & f^*_{\ell,P+2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ f^*_{\ell,N-1} & f^*_{\ell,N} & 0 & \cdots & 0 \\ f^*_{\ell,N} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{N\times P},$$

$$\mathbf{M}_{\ell,2} = \begin{bmatrix} f_{\ell,N} & f_{\ell,N-1} & f_{\ell,N-2} & \cdots & f_{\ell,N-P+1} \\ 0 & f_{\ell,N} & f_{\ell,N-1} & \cdots & f_{\ell,N-P+2} \\ 0 & 0 & f_{\ell,N} & \cdots & f_{\ell,N-P+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f_{\ell,N} \end{bmatrix}_{P\times P}, \tag{14}$$

$\mathbf{b}_{\ell,1} = \mathbf{f}^*_{\ell,1:N} + \mathbf{M}_{\ell,1}\mathbf{a}_{\text{FS}}$ and $\mathbf{b}_{\ell,2} = \mathbf{M}_{\ell,2}\mathbf{a}_{\text{FS}}$ are constant vectors. The matrices $\mathbf{G}$ and $\mathbf{G}_c$ are the $P\times P$ Jacobian matrices having entries $[\mathbf{G}]_{ij} = \frac{\partial a_i}{\partial k_j}$ and $[\mathbf{G}_c]_{ij} = \frac{\partial a_i}{\partial k_j^*}$ evaluated at the expansion point of $\mathbf{a}_{\text{FS}}$ and $(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_c\boldsymbol{\delta}_{\mathbf{k}}^*)$ is the perturbation vector for $\mathbf{a}_{\text{FS}}$ such that $\mathbf{k} + \boldsymbol{\delta}_{\mathbf{k}}$ is the reflection coefficients corresponding to $\mathbf{a}_{\text{FS}} + (\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}} + \mathbf{G}_c\boldsymbol{\delta}_{\mathbf{k}}^*)$. These matrices can be efficiently calculated via the inverse Levinson–Durbin recursion, as given in Candan [6, Algorithm 2].

By the Taylor series expansion of $\log(A + Bx) \approx \log(A) + \frac{B}{A}x$ for $|x| \ll 1$, we can approximate the second term of (9) as

$$\log\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell \approx \log(A)$$

$$+\sum_{\ell=1}^{L_a}\frac{\|\mathbf{b}_{\ell,1}+\mathbf{M}_{\ell,1}(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}}+\mathbf{G}_c\boldsymbol{\delta}_{\mathbf{k}}^*)\|^2-\|\mathbf{b}_{\ell,1}\|^2}{A}$$
$$-\sum_{\ell=1}^{L_a}\frac{\|\mathbf{b}_{\ell,2}+\mathbf{M}_{\ell,2}(\mathbf{G}\boldsymbol{\delta}_{\mathbf{k}}+\mathbf{G}_c\boldsymbol{\delta}_{\mathbf{k}}^*)\|^2-\|\mathbf{b}_{\ell,2}\|^2}{A}, \tag{15}$$

where $A=\sum_{\ell=1}^{L_a}\left(\|\mathbf{b}_{\ell,1}\|^2-\|\mathbf{b}_{\ell,2}\|^2\right)$ is the value when $\boldsymbol{\delta}_{\mathbf{k}}$ is replaced by all zeros vector in (13). Hence, $A=\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}_{\mathrm{FS}})\mathbf{f}_\ell$, and $\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}_{\mathrm{FS}})\mathbf{f}_\ell$ can be efficiently calculated by using [6, Algorithm 1], without even constructing $\mathbf{R}_{f,N}$.

Following the step of optimization, we reach the following equation system for the solution of $\boldsymbol{\delta}_{\mathbf{k}}$ and $\boldsymbol{\delta}_{\mathbf{k}}^*$, which are perturbation vectors for the complex-valued reflection coefficients corresponding to the initial reflection coefficients $\mathbf{k}$ generated from $\mathbf{a}_{\mathrm{FS}}$.

$$\begin{bmatrix}\mathbf{Q}_1+\widetilde{\mathbf{Q}}_1 & \mathbf{Q}_2+\widetilde{\mathbf{Q}}_2\\ \mathbf{Q}_2^*+\widetilde{\mathbf{Q}}_2^* & \mathbf{Q}_1^*+\widetilde{\mathbf{Q}}_1^*\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}_{\mathbf{k}}\\ \boldsymbol{\delta}_{\mathbf{k}}^*\end{bmatrix}=-\begin{bmatrix}\mathbf{r}_1+\widetilde{\mathbf{r}}_1\\ \mathbf{r}_1^*+\widetilde{\mathbf{r}}_1^*\end{bmatrix}. \tag{16}$$

In the last equation, we have

$$\widetilde{\mathbf{Q}}_1=\frac{\left(\mathbf{G}^H\mathbf{P}\mathbf{G}_c\right)^*+\mathbf{G}_c^H\mathbf{P}\mathbf{G}}{A},\qquad \widetilde{\mathbf{Q}}_2=\frac{\left(\mathbf{G}^H\mathbf{P}\mathbf{G}\right)^*+\mathbf{G}_c^H\mathbf{P}\mathbf{G}_c}{A},$$
$$\widetilde{\mathbf{r}}_1=\frac{\left(\mathbf{G}^H\mathbf{v}\right)^*+\mathbf{G}_c^H\mathbf{v}}{A}, \tag{17}$$

where $\mathbf{P}=\sum_{\ell=1}^{L_a}\left(\mathbf{M}_{\ell,1}^H\mathbf{M}_{\ell,1}-\mathbf{M}_{\ell,2}^H\mathbf{M}_{\ell,2}\right)$ and $\mathbf{v}=\sum_{\ell=1}^{L_a}\left(\mathbf{M}_{\ell,1}^H\mathbf{b}_{\ell,1}-\mathbf{M}_{\ell,2}^H\mathbf{b}_{\ell,2}\right)$. The matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are diagonal matrices with diagonal entries $[\mathbf{Q}_1]_{ii}=\frac{i(k_i^*)^2}{N(1-|k_i|^2)^2}$ and $[\mathbf{Q}_2]_{ii}=\frac{i}{N(1-|k_i|^2)^2}$, and $\mathbf{r}_1=[\,r_{1,1}\,r_{1,2}\,\ldots\,r_{1,P}\,]^T$ with $r_{1,i}=\frac{ik_i^*}{N(1-|k_i|^2)}$, $i=\{1,2,\ldots,P\}$. In (16), the gradients of the quadratic approximation of $\frac{1}{N}\log|\mathbf{R}_{f,N}|$ and $\log\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell$ with respect to $\boldsymbol{\delta}_{\mathbf{k}}^*$, which can be expressed as $\mathbf{Q}_2^*\boldsymbol{\delta}_{\mathbf{k}}+\mathbf{Q}_1^*\boldsymbol{\delta}_{\mathbf{k}}^*+\mathbf{r}_1^*$ and $\widetilde{\mathbf{Q}}_2^*\boldsymbol{\delta}_{\mathbf{k}}+\widetilde{\mathbf{Q}}_1^*\boldsymbol{\delta}_{\mathbf{k}}^*+\widetilde{\mathbf{r}}_1^*$, respectively, are utilized, as derived in Candan [6].

For the real-valued processes, the reflection coefficients are also real-valued, i.e., $\boldsymbol{\delta}_{\mathbf{k}}=\boldsymbol{\delta}_{\mathbf{k}}^*$, and hence, (16) can be simplified as

$$(\mathbf{Q}_1+\widetilde{\mathbf{Q}}_1+\mathbf{Q}_2+\widetilde{\mathbf{Q}}_2)\boldsymbol{\delta}_{\mathbf{k}}=-(\mathbf{r}_1+\widetilde{\mathbf{r}}_1). \tag{18}$$

The proposed method for AR parameter estimation from multiple noise-free snapshots is outlined in Algorithm 1 . It should be observed from Algorithm 1 listing that the second stage is iteratively applied by using the previous iteration result as the initial point of the following iteration.

Algorithm 1: Suggested AR parameter estimation for multiple noise-free snapshots.

Input : $\mathbf{X}=[\,\mathbf{x}_1\,\mathbf{x}_2\,\ldots\,\mathbf{x}_L\,]$, P
Output: $\widehat{\mathbf{a}}$, $\widehat{\sigma}_\epsilon^2$
// The first stage estimation by using $L_a=L$ and $\mathbf{f}_l=\mathbf{x}_l$, $l=\{1,2,\ldots,L\}$ for (12)
$\mathbf{a}_{\mathrm{FS}}=-\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^H\mathbf{W}\mathbf{A}_{\ell,f}+\mathbf{A}_{\ell,b}^H\mathbf{W}\mathbf{A}_{\ell,b}\right)\right)^{-1}$
$\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^H\mathbf{W}\mathbf{b}_{\ell,f}+\mathbf{A}_{\ell,b}^H\mathbf{W}\mathbf{b}_{\ell,b}\right)\right)$
// The second stage estimation by using $L_a=L$ and $\mathbf{f}_l=\mathbf{x}_l$, $l=\{1,2,\ldots,L\}$
$\mathbf{a}^{\mathrm{old}}=\mathbf{a}_{\mathrm{FS}}$
for iteration $\leftarrow 1$ **to** *10* **do**
  // the loop with 10 iterations yields a good performance, see [6]
  $\mathbf{k}=\mathrm{atog}([\,1\;(\mathbf{a}^{\mathrm{old}})^T\,]^T)$; // atog(·): Step-down recursion, [1, p. 236]
  if X is real-valued **then**
    Solve $(\mathbf{Q}_1+\widetilde{\mathbf{Q}}_1+\mathbf{Q}_2+\widetilde{\mathbf{Q}}_2)\boldsymbol{\delta}_{\mathbf{k}}=-(\mathbf{r}_1+\widetilde{\mathbf{r}}_1)$; // See (18)
  else
    Solve $\begin{bmatrix}\mathbf{Q}_1+\widetilde{\mathbf{Q}}_1 & \mathbf{Q}_2+\widetilde{\mathbf{Q}}_2\\ \mathbf{Q}_2^*+\widetilde{\mathbf{Q}}_2^* & \mathbf{Q}_1^*+\widetilde{\mathbf{Q}}_1^*\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}_{\mathbf{k}}\\ \boldsymbol{\delta}_{\mathbf{k}}^*\end{bmatrix}=-\begin{bmatrix}\mathbf{r}_1+\widetilde{\mathbf{r}}_1\\ \mathbf{r}_1^*+\widetilde{\mathbf{r}}_1^*\end{bmatrix}$; // See (16)
  $\mathbf{k}=\mathbf{k}+\boldsymbol{\delta}_{\mathbf{k}}$; // Reflection coefficients update
  $[\,1\;(\mathbf{a}^{\mathrm{new}})^T\,]^T=\mathrm{gtoa}(\mathbf{k})$; // gtoa(·): Step-up recursion, [1, p. 233]
  $\mathbf{a}^{\mathrm{old}}=\mathbf{a}^{\mathrm{new}}$
$(\sigma_\epsilon^2)^{\mathrm{new}}=\frac{1}{LN}\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\mathrm{new}})\mathbf{f}_\ell$; // $\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\mathrm{new}})\mathbf{f}_\ell$ is calculated via [6, Algorithm 1]
Return: $\widehat{\mathbf{a}}=\mathbf{a}^{\mathrm{new}}$, $\widehat{\sigma}_\epsilon^2=(\sigma_\epsilon^2)^{\mathrm{new}}$

2.3. An efficient implementation for the suggested solution

The expectation step of the suggested method requires the inversion of an $N\times N$ matrix for the calculation of posterior density parameters, i.e., mean vector and error covariance matrix in (4). In general, the observation vector length (N) is much greater than the order of AR process (P), and the implementation cost of the expectation step becomes a computational bottleneck. In this section, we present four approaches to reduce the computational load. We start with the disturbance smoother, a variation of the Kalman smoothers which does not require any matrix inversion for the estimation of the mean vector in the present problem set-up [41, Sec. 5.2.4].

2.3.1. Efficient calculation of the mean vector

It is well known that the Wiener filtering operation, which is the operation implemented with the equation set (4), coincides with the Kalman filtering for the processing of a wide sense stationary input, with rational power spectral density, corrupted by independent measurement noise, Hayes [1]. To facilitate the recursive calculation via the Kalman smoothing, we introduce the following state space model:

$$\underbrace{\begin{bmatrix}x[n]\\ x[n-1]\\ x[n-2]\\ \vdots\\ x[n-P+1]\end{bmatrix}}_{\mathbf{s}_n}=\underbrace{\begin{bmatrix}-a_1^{\mathrm{old}} & -a_2^{\mathrm{old}} & -a_3^{\mathrm{old}} & \ldots & -a_P^{\mathrm{old}}\\ 1 & 0 & 0 & \ldots & 0\\ 0 & 1 & 0 & \ldots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & \ldots & 1 & 0\end{bmatrix}}_{\mathbf{A}}\underbrace{\begin{bmatrix}x[n-1]\\ x[n-2]\\ x[n-3]\\ \vdots\\ x[n-P]\end{bmatrix}}_{\mathbf{s}_{n-1}}+\underbrace{\begin{bmatrix}(\sigma_\epsilon)^{\mathrm{old}}\\ 0\\ 0\\ \vdots\\ 0\end{bmatrix}}_{\mathbf{b}}w_n,$$
$$y_n=\underbrace{\begin{bmatrix}1 & 0 & \ldots & 0\end{bmatrix}}_{\mathbf{c}}\mathbf{s}_n+v_n, \tag{19}$$

where w_n, y_n and v_n denote the process noise $w[n]$, the measurement $y[n]$ and the measurement noise $v[n]$, respectively.

Table 1 gives the disturbance smoother equation for the state space model of (19). The disturbance smoother is a three-pass operation, where in the first pass the Kalman filtering is applied in the forward direction, the disturbance is estimated in the second pass and the smoothed state vector is formed in the last pass,

Table 1
The disturbance smoother implementation of the Kalman smoother (also see [43]).

State Equation:	$\mathbf{s}_n = \mathbf{A}\mathbf{s}_{n-1} + \mathbf{b}w_n, \quad (w_n \sim \mathcal{CN}(0,1))$
	$y_n = \mathbf{c}\mathbf{s}_n + v_n, \quad (v_n \sim \mathcal{CN}(0, (\sigma_v^2)^{\text{old}}))$
Input:	$y_n,\ n = 1, 2, \ldots, N$
Output:	$\mathbf{c}\check{\mathbf{s}}_{n\|N}$ (first element of $\check{\mathbf{s}}_{n\|N}$), $n = 1, 2, \ldots, N$
Kalman Filtering:	$\hat{\mathbf{s}}_0 = [\,0 \ldots 0\,]^T,\ \mathbf{P}_{0\|0} = (\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}),$ (see (2))
	for $n = 0, 1, \ldots, N-1$
	$\hat{\mathbf{s}}_{n+1\|n} = \mathbf{A}\hat{\mathbf{s}}_{n\|n},\ \mathbf{P}_{n+1\|n} = \mathbf{A}\mathbf{P}_{n\|n}\mathbf{A}^H + \mathbf{b}\mathbf{b}^H,$
	$\epsilon_{n+1} = y_{n+1} - \mathbf{c}\hat{\mathbf{s}}_{n+1\|n},$
	$\gamma_{n+1} = \mathbf{c}\mathbf{P}_{n+1\|n}\mathbf{c}^H + (\sigma_v^2)^{\text{old}},$
	$\mathbf{H}_{n+1} = \mathbf{P}_{n+1\|n}\mathbf{c}^H\gamma_{n+1}^{-1},$
	$\hat{\mathbf{s}}_{n+1\|n+1} = \hat{\mathbf{s}}_{n+1\|n} + \mathbf{H}_{n+1}\epsilon_{n+1},$
	$\mathbf{P}_{n+1\|n+1} = \mathbf{P}_{n+1\|n} - \mathbf{H}_{n+1}\mathbf{c}\mathbf{P}_{n+1\|n},$
	$\mathbf{\Lambda}_{n+1} = \mathbf{A} - \mathbf{A}\mathbf{H}_{n+1}\mathbf{c}.$
Backward Smoothing:	$\mathbf{p}_{N-1} = \mathbf{c}^H\epsilon_N\gamma_N^{-1},\ \hat{e}_{N-1\|N} = \mathbf{b}^H\mathbf{p}_{N-1},$
	for $n = N-2, N-3, \ldots, 1$
	$\mathbf{p}_n = \mathbf{c}^H\epsilon_{n+1}\gamma_{n+1}^{-1} + \mathbf{\Lambda}_{n+1}^H\mathbf{p}_{n+1},\ \hat{e}_{n\|N} = \mathbf{b}^H\mathbf{p}_n.$
Smoothed State Vector:	$\check{\mathbf{s}}_{1\|N} = \mathbf{P}_{0\|0}(\mathbf{c}^H\epsilon_1\gamma_1^{-1} + \mathbf{\Lambda}_1^H\mathbf{p}_1),$
	for $n = 2, 3, \ldots, N-1$
	$\check{\mathbf{s}}_{n\|N} = \mathbf{A}\check{\mathbf{s}}_{n-1\|N} + \mathbf{b}\hat{e}_{n-1\|N},$
	$\check{\mathbf{s}}_{N\|N} = \hat{\mathbf{s}}_{N\|N}.$

Cappe et al. [41, Sec. 5.2.4]. The passes do not involve any matrix inversion operation. Hence, the mean vector calculation cost is effectively reduced from the $N \times N$ matrix inversion cost to the order of N complex multiplications. The computational savings becomes very significant when $N \geq 100$.

2.3.2. Approximating error covariance matrix

The matrix $\mathbf{K}$ in (4) corresponds to the error covariance matrix of the Wiener filter estimate. Unfortunately, it is not possible to retrieve an $N \times N$ dimensional error covariance matrix from a Kalman filtering implementation. To avoid the calculation of $\mathbf{K}$ matrix, we examine the limiting case of $N \to \infty$ that corresponds to the infinite impulse response (IIR) noncausal Wiener filtering operation, Hayes [1, Sec. 7.3]. As $N \to \infty$, the error covariance matrix approaches to a Hermitian Toeplitz matrix whose first column is sufficient to characterize the complete matrix, Gray [44]. For the filtering application with the noncausal IIR Wiener filter ($H^{\text{IIR-NC}}(z)$), the error autocorrelation sequence can be expressed as $r_e[k] = (\sigma_v^2)^{\text{old}}h^{\text{IIR-NC}}[k]$, where $h^{\text{IIR-NC}}[k]$ is the impulse response of the noncausal IIR Wiener filter. We suggest using the residue theorem to evaluate the inverse z-transform of $H^{\text{IIR-NC}}(z)$ for the calculation of error correlation sequence:

$$r_e[k] = \frac{(\sigma_v^2)^{\text{old}}}{2\pi j}\oint_{C:|z|=1} H^{\text{IIR-NC}}(z)z^{k-1}dz, \quad k = \{0, 1, \ldots, P\}, \tag{20}$$

where $H^{\text{IIR-NC}}(z) = (\sigma_\epsilon^2)^{\text{old}}/((\sigma_\epsilon^2)^{\text{old}} + (\sigma_v^2)^{\text{old}}A^{\text{old}}(z)(A^{\text{old}}(1/z^*))^*)$, Hayes [1, Sec. 7.3] for $A^{\text{old}}(z) = 1 + a_1^{\text{old}}z^{-1} + a_2^{\text{old}}z^{-2} + \ldots + a_P^{\text{old}}z^{-P}$. We suggest calculating only $P+1$ error correlation lags given in (20) and constructing $\mathbf{K}$ as a banded matrix with P nonzero super/sub-diagonals.

2.3.3. Efficient calculation of $tr(\mathbf{R}_{f,N}^{-1}\mathbf{K})$

The compressed likelihood function in (8) requires the calculation of $\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$, where $\mathbf{R}_{f,N}^{-1}$ depends on the optimization variable $\mathbf{a}$ and $\mathbf{K}$ is a constant matrix. This problem has been suggested to be solved by an eigenvalue decomposition of $\mathbf{K}$ and expressing $\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ as a sum of quadratic terms. We present an efficient method that avoids the computationally costly eigendecomposition step.

We utilize the Gohberg–Semencul (GS) formula from [40, Sec. 3.9.4] for the inverse of $\mathbf{R}_{f,N}$. The GS formula states that $\mathbf{R}_{f,N}^{-1} = \mathbf{A}_1\mathbf{A}_1^H - \mathbf{A}_2\mathbf{A}_2^H$, where

$$\mathbf{A}_1 = \begin{bmatrix} 1 & & & & & & 0 \\ a_1 & 1 & & & & & \\ \vdots & a_1 & 1 & & & & \\ a_P & \ddots & a_1 & 1 & & & \\ 0 & a_P & \ddots & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & 1 & \\ 0 & \ldots & 0 & a_P & \ldots & a_1 & 1 \end{bmatrix}_{N\times N},$$

$$\mathbf{A}_2 = \begin{bmatrix} 0 & & & & & & 0 \\ \vdots & 0 & & & & & \\ 0 & \ddots & 0 & & & & \\ a_P^* & \ddots & \ddots & \ddots & & & \\ a_{P-1}^* & a_P^* & \ddots & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ a_1^* & \ldots & a_{P-1}^* & a_P^* & 0 & \ldots & 0 \end{bmatrix}_{N\times N}. \tag{21}$$

By defining a shift matrix $\mathbf{S}$ as

$$\mathbf{S} = \begin{bmatrix} 0 & & & & 0 \\ 1 & 0 & & & \\ 0 & 1 & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \ldots & 0 & 1 & 0 \end{bmatrix}_{N\times N},$$

$\mathbf{A}_1$ matrix can be expressed as $\mathbf{A}_1 = [\,\mathbf{a}_+ \,|\, \mathbf{S}\mathbf{a}_+ \,|\, \mathbf{S}^2\mathbf{a}_+ \,|\, \ldots \,|\, \mathbf{S}^{N-1}\mathbf{a}_+\,]$, where $\mathbf{a}_+$ denotes the first column of $\mathbf{A}_1$ matrix from (21) and $\mathbf{S}^k\mathbf{a}_+$ is the $(k+1)$'th column formed by shifting $\mathbf{a}_+$ vector k times. We also define $\mathbf{A}_2 = [\,\mathbf{a}_R^* \,|\, \mathbf{S}\mathbf{a}_R^* \,|\, \mathbf{S}^2\mathbf{a}_R^* \,|\, \ldots \,|\, \mathbf{S}^{N-1}\mathbf{a}_R^*\,]$, where $\mathbf{a}_R^*$ is the first column of $\mathbf{A}_2$ matrix. Using these definitions, $\text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ can be expressed as

$$\begin{aligned} \text{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K}) &= \text{tr}((\mathbf{A}_1\mathbf{A}_1^H - \mathbf{A}_2\mathbf{A}_2^H)\mathbf{K}) \\ &= \text{tr}(\mathbf{A}_1^H\mathbf{K}\mathbf{A}_1) - \text{tr}(\mathbf{A}_2^H\mathbf{K}\mathbf{A}_2) \\ &= \mathbf{a}_+^H\left(\sum_{k=0}^{N-1}(\mathbf{S}^k)^H\mathbf{K}\mathbf{S}^k\right)\mathbf{a}_+ - \mathbf{a}_R^T\left(\sum_{k=0}^{N-1}(\mathbf{S}^k)^H\mathbf{K}\mathbf{S}^k\right)\mathbf{a}_R^* \\ &= \mathbf{a}_+^H\mathbf{V}\mathbf{a}_+ - \mathbf{a}_R^T\mathbf{V}\mathbf{a}_R^*, \end{aligned} \tag{22}$$

where $\mathbf{V}=\sum_{k=0}^{N-1}(\mathbf{S}^k)^H\mathbf{K}\mathbf{S}^k$. Hence, $\mathrm{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ is a quadratic product of the unknown parameters.

2.3.4. Efficient calculation of optimal perturbation around $\mathbf{a}_{FS}$

The objective function (8) to be maximized involves $|\mathbf{R}_{f,N}|$ and $L\,\mathrm{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})+\sum_{l=1}^{L}\widehat{\mathbf{x}}_l^H\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l$, where the elements of matrix $\mathbf{R}_{f,N}$ are the functions of vector $\mathbf{a}$. While the determinant of $\mathbf{R}_{f,N}$ can be written as a quadratic product of the reflection coefficients corresponding to the vector $\mathbf{a}$, the other reduces to quadratic product in terms of $\mathbf{a}$. Following the approach given in Candan [6], we convert the problem domain to the reflection coefficient domain and treat $\mathbf{a}$ as a vector-valued function of the reflection coefficients, and then, we expand it into a Taylor series around the reflection coefficients derived from the first stage result ($\mathbf{a}_{\mathrm{FS}}$) and keep only the first order term, i.e., the Jacobian term, in the expansion. In brief, for $\mathrm{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})=\mathbf{a}_+^H\mathbf{V}\mathbf{a}_+-\mathbf{a}_R^T\mathbf{V}\mathbf{a}_R^*$, we express the unknown vectors as $\mathbf{a}_+=\mathbf{a}_{\mathrm{FS},+}+\mathbf{G}_+\boldsymbol{\delta}_{\mathbf{k}}+\mathbf{G}_{+c}\boldsymbol{\delta}_{\mathbf{k}}^*$ and $\mathbf{a}_R=\mathbf{a}_{\mathrm{FS},R}+\mathbf{G}_R\boldsymbol{\delta}_{\mathbf{k}}+\mathbf{G}_{Rc}\boldsymbol{\delta}_{\mathbf{k}}^*$, where $\mathbf{a}_{\mathrm{FS},+}=[\,1\;\mathbf{a}_{\mathrm{FS}}^T\;\mathbf{0}_{N-P-1}^T\,]^T$ and $\mathbf{a}_{\mathrm{FS},R}=[\,\mathbf{0}_{N-P}^T\;\mathbf{a}_{\mathrm{FS},P:-1:1}^T\,]^T$. The vectors $\boldsymbol{\delta}_{\mathbf{k}}$ and its conjugate $\boldsymbol{\delta}_{\mathbf{k}}^*$ denote the unknown reflection coefficient perturbation vectors following the approach in Candan [6]. The matrices $\mathbf{G}_+$, $\mathbf{G}_{+c}$, $\mathbf{G}_R$ and $\mathbf{G}_{Rc}$ are the Jacobian matrices, for the vector-valued functions $\mathbf{a}_{\mathrm{FS},+}$ and $\mathbf{a}_{\mathrm{FS},R}$ in terms of their reflection coefficients, with the definitions of $\mathbf{G}_+=[\,\mathbf{0}_P\;\mathbf{G}^T\;\mathbf{0}_{P\times(N-P-1)}\,]^T$, $\mathbf{G}_{+c}=[\,\mathbf{0}_P\;\mathbf{G}_c^T\;\mathbf{0}_{P\times(N-P-1)}\,]^T$, $\mathbf{G}_R=[\,\mathbf{0}_{P\times(N-P)}\;\widetilde{\mathbf{G}}^T\,]^T$ and $\mathbf{G}_{Rc}=[\,\mathbf{0}_{P\times(N-P)}\;\widetilde{\mathbf{G}}_c^T\,]^T$, where $\mathbf{G}$ and $\mathbf{G}_c$ are the Jacobian matrices corresponding to $\mathbf{a}_{\mathrm{FS}}$, as defined in the description of (13); $[\widetilde{\mathbf{G}}]_{ij}=[\mathbf{G}]_{pj}$ and $[\widetilde{\mathbf{G}}_c]_{ij}=[\mathbf{G}_c]_{pj}$ for $p=P-i+1$ and $i,j\in\{1,2,\ldots,P\}$. With these definitions, the gradient of $\mathbf{a}_+^H\mathbf{V}\mathbf{a}_+$ with respect to $\boldsymbol{\delta}_{\mathbf{k}}^*$ can be given as $\mathbf{Q}_3\boldsymbol{\delta}_{\mathbf{k}}+\mathbf{Q}_{3c}\boldsymbol{\delta}_{\mathbf{k}}^*+\mathbf{r}_3$ with $\mathbf{Q}_3=\mathbf{G}_+^H\mathbf{V}\mathbf{G}_++(\mathbf{G}_{+c}^H\mathbf{V}\mathbf{G}_{+c})^*$, $\mathbf{Q}_{3c}=\mathbf{G}_+^H\mathbf{V}\mathbf{G}_{+c}+(\mathbf{G}_{+c}^H\mathbf{V}\mathbf{G}_+)^*$ and $\mathbf{r}_3=\mathbf{G}_+^H\mathbf{V}\mathbf{a}_{\mathrm{FS},+}+(\mathbf{G}_{+c}^H\mathbf{V}\mathbf{a}_{\mathrm{FS},+})^*$. Similarly, the gradient of $\mathbf{a}_R^T\mathbf{V}\mathbf{a}_R^*$ with respect to $\boldsymbol{\delta}_{\mathbf{k}}^*$ is $\mathbf{Q}_4\boldsymbol{\delta}_{\mathbf{k}}+\mathbf{Q}_{4c}\boldsymbol{\delta}_{\mathbf{k}}^*+\mathbf{r}_4$ with $\mathbf{Q}_4=\mathbf{G}_R^T\mathbf{V}^*\mathbf{G}_{Rc}^*+(\mathbf{G}_{Rc}^T\mathbf{V}^*\mathbf{G}_R^*)^*$, $\mathbf{Q}_{4c}=\mathbf{G}_R^T\mathbf{V}^*\mathbf{G}_R^*+(\mathbf{G}_{Rc}^T\mathbf{V}^*\mathbf{G}_{Rc}^*)^*$ and $\mathbf{r}_4=\mathbf{G}_R^T\mathbf{V}^*\mathbf{a}_{\mathrm{FS},R}^*+(\mathbf{G}_{Rc}^T\mathbf{V}^*\mathbf{a}_{\mathrm{FS},R}^*)^*$. Combining the gradient results of $\mathrm{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ with the gradients of terms used for (16), we get the following equation system:

$$\begin{bmatrix}\mathbf{Q}_1+\widetilde{\mathbf{Q}}_1+\mathbf{Q}_{\mathrm{tr},1} & \mathbf{Q}_2+\widetilde{\mathbf{Q}}_2+\mathbf{Q}_{\mathrm{tr},2}\\ \mathbf{Q}_2^*+\widetilde{\mathbf{Q}}_2^*+\mathbf{Q}_{\mathrm{tr},2}^* & \mathbf{Q}_1^*+\widetilde{\mathbf{Q}}_1^*+\mathbf{Q}_{\mathrm{tr},1}^*\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}_{\mathbf{k}}\\ \boldsymbol{\delta}_{\mathbf{k}}^*\end{bmatrix}=-\begin{bmatrix}\mathbf{r}_1+\widetilde{\mathbf{r}}_1+\mathbf{r}_{\mathrm{tr}}\\ \mathbf{r}_1^*+\widetilde{\mathbf{r}}_1^*+\mathbf{r}_{\mathrm{tr}}^*\end{bmatrix},\tag{23}$$

where $\mathbf{Q}_{\mathrm{tr},1}=\frac{L}{A}(\mathbf{Q}_{3c}-\mathbf{Q}_{4c})^*$, $\mathbf{Q}_{\mathrm{tr},2}=\frac{L}{A}(\mathbf{Q}_3-\mathbf{Q}_4)^*$ and $\mathbf{r}_{\mathrm{tr}}=\frac{L}{A}(\mathbf{r}_3-\mathbf{r}_4)^*$ with $A=L\,\mathrm{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})+\sum_{l=1}^{L}\widehat{\mathbf{x}}_l^H\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l$. In (23), the terms $\mathbf{Q}_1$, $\mathbf{Q}_2$ and $\mathbf{r}_1$ corresponding to $\frac{1}{N}\log|\mathbf{R}_{f,N}|$ are the same as used in (16); the terms $\widetilde{\mathbf{Q}}_1$, $\widetilde{\mathbf{Q}}_2$ and $\widetilde{\mathbf{r}}_1$ corresponding to $\sum_{l=1}^{L}\widehat{\mathbf{x}}_l^H\mathbf{R}_{f,N}^{-1}\widehat{\mathbf{x}}_l$ are calculated as given in (17) by using $\mathbf{P}=\sum_{l=1}^{L}\left(\mathbf{M}_{l,1}^H\mathbf{M}_{l,1}-\mathbf{M}_{l,2}^H\mathbf{M}_{l,2}\right)$ and $\mathbf{v}=\sum_{l=1}^{L}\left(\mathbf{M}_{l,1}^H\mathbf{b}_{l,1}-\mathbf{M}_{l,2}^H\mathbf{b}_{l,2}\right)$, where $\mathbf{b}_{l,1}=\widehat{\mathbf{x}}_{l,1:N}^*+\mathbf{M}_{l,1}\mathbf{a}_{\mathrm{FS}}$, $\mathbf{b}_{l,2}=\mathbf{M}_{l,2}\mathbf{a}_{\mathrm{FS}}$, and the matrices $\mathbf{M}_{l,1}$ and $\mathbf{M}_{l,2}$ are formed by using (14) with $\mathbf{f}_l=\widehat{\mathbf{x}}_l$, $l=\{1,2,\ldots,L\}$.

To determine the optimal perturbation vector for the real-valued processes, by using the condition $\boldsymbol{\delta}_{\mathbf{k}}=\boldsymbol{\delta}_{\mathbf{k}}^*$, (23) can be simplified as

$$(\mathbf{Q}_1+\widetilde{\mathbf{Q}}_1+\mathbf{Q}_{\mathrm{tr},1}+\mathbf{Q}_2+\widetilde{\mathbf{Q}}_2+\mathbf{Q}_{\mathrm{tr},2})\boldsymbol{\delta}_{\mathbf{k}}=-(\mathbf{r}_1+\widetilde{\mathbf{r}}_1+\mathbf{r}_{\mathrm{tr}}).\tag{24}$$

To solve the problem of AR parameter estimation with multiple noisy snapshots, the suggested second stage for the calculation of $\mathrm{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ without an eigendecomposition is outlined in Algorithm 2 . The complete procedure including four cases is given in Algorithm 3 . In Algorithm 3, the first case calculates $\mathbf{K}$ matrix exactly and assumes σ_v^2 is known by following Lines 7, 15 and 19; the second case calculates $\mathbf{K}$ matrix approximately and assumes σ_v^2 is known by following Lines 10, 17 and 19; the third and fourth cases are similar to the first and second cases, respectively, except σ_v^2 is estimated via Line 21.

Algorithm 2: Efficient calculation of optimal perturbation around $\mathbf{a}_{\mathrm{FS}}$.

Input : $\widehat{\mathbf{X}}=[\,\widehat{\mathbf{x}}_1\;\widehat{\mathbf{x}}_2\;\ldots\;\widehat{\mathbf{x}}_L\,]$, $\mathbf{K}$, $\mathbf{a}_{\mathrm{FS}}$
Output: $\mathbf{a}^{\mathrm{new}}$, $(\sigma_\epsilon^2)^{\mathrm{new}}$
1 $\mathbf{a}^{\mathrm{old}}=\mathbf{a}_{\mathrm{FS}}$
2 **for** iteration $\leftarrow 1$ **to** *10* **do**
  // the loop with 10 iterations yields a good performance
3 $\mathbf{k}=\mathrm{atog}([\,1\;(\mathbf{a}^{\mathrm{old}})^T\,]^T)$; // atog(·): Step-down recursion, [1, p. 236]
4 **if** $\widehat{\mathbf{X}}$ is real-valued **then**
5 Solve $(\mathbf{Q}_1+\widetilde{\mathbf{Q}}_1+\mathbf{Q}_{\mathrm{tr},1}+\mathbf{Q}_2+\widetilde{\mathbf{Q}}_2+\mathbf{Q}_{\mathrm{tr},2})\boldsymbol{\delta}_{\mathbf{k}}=-(\mathbf{r}_1+\widetilde{\mathbf{r}}_1+\mathbf{r}_{\mathrm{tr}})$// See (23), where $\mathrm{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ is computed by using the GS formula for $\mathbf{R}_{f,N}^{-1}$
6 **else**
7 Solve $\begin{bmatrix}\mathbf{Q}_1+\widetilde{\mathbf{Q}}_1+\mathbf{Q}_{\mathrm{tr},1} & \mathbf{Q}_2+\widetilde{\mathbf{Q}}_2+\mathbf{Q}_{\mathrm{tr},2}\\ \mathbf{Q}_2^*+\widetilde{\mathbf{Q}}_2^*+\mathbf{Q}_{\mathrm{tr},2}^* & \mathbf{Q}_1^*+\widetilde{\mathbf{Q}}_1^*+\mathbf{Q}_{\mathrm{tr},1}^*\end{bmatrix}\begin{bmatrix}\boldsymbol{\delta}_{\mathbf{k}}\\ \boldsymbol{\delta}_{\mathbf{k}}^*\end{bmatrix}=-\begin{bmatrix}\mathbf{r}_1+\widetilde{\mathbf{r}}_1+\mathbf{r}_{\mathrm{tr}}\\ \mathbf{r}_1^*+\widetilde{\mathbf{r}}_1^*+\mathbf{r}_{\mathrm{tr}}^*\end{bmatrix}$// See (24), where $\mathrm{tr}(\mathbf{R}_{f,N}^{-1}\mathbf{K})$ is computed by using the GS formula for $\mathbf{R}_{f,N}^{-1}$
8 $\mathbf{k}=\mathbf{k}+\boldsymbol{\delta}_{\mathbf{k}}$; // Reflection coefficients update
9 $[\,1\;(\mathbf{a}^{\mathrm{new}})^T\,]^T=\mathrm{gtoa}(\mathbf{k})$; // gtoa(·): Step-up recursion, [1,p. 233]
10 $\mathbf{a}^{\mathrm{old}}=\mathbf{a}^{\mathrm{new}}$
11 $(\sigma_\epsilon^2)^{\mathrm{new}}=\frac{1}{LN}\left(L\,\mathrm{tr}(\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\mathrm{new}})\mathbf{K})+\sum_{l=1}^{L}\widehat{\mathbf{x}}_l^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\mathrm{new}})\widehat{\mathbf{x}}_l\right)$; // $\mathrm{tr}(\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\mathrm{new}})\mathbf{K})$ is computed by using the GS formula for $\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\mathrm{new}})$
Return: $\mathbf{a}^{\mathrm{new}}$, $(\sigma_\epsilon^2)^{\mathrm{new}}$

2.4. Computational complexity considerations

In noise-free multiple-snapshot case given in Algorithm 1, $\mathbf{a}_{\mathrm{FS}}$ is computed with $\mathcal{O}(P^3)$ multiplications in the first stage, and the optimal perturbation vector $\boldsymbol{\delta}_{\mathbf{k}}$ is computed with $\mathcal{O}(P^3)$ and $\mathcal{O}(8P^3)$ multiplications for the real- and complex-valued processes, respectively, in the second stage. In (16), $A=\sum_{\ell=1}^{L_a}\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell$, which is required to get $\widetilde{\mathbf{Q}}_1$, $\widetilde{\mathbf{Q}}_2$ and $\widetilde{\mathbf{r}}_1$, is calculated efficiently by using Candan [6, Algorithm 1], which decreases $\mathcal{O}(N^3)$ operations to $\mathcal{O}(NP)$ multiplications to calculate $\mathbf{f}_\ell^H\mathbf{R}_{f,N}^{-1}\mathbf{f}_\ell$. The noise-free case requires $L_a=L$ snapshots and corresponds to $\mathcal{O}(LNP+P^3)$ and $\mathcal{O}(LNP+8P^3)$ multiplications per EM iteration for the real- and complex-valued processes, respectively.

In noisy multiple-snapshot case with the exact calculation of $\mathbf{K}$ matrix given in Algorithm 3, both of the inversion and eigendecomposition of $N\times N$ matrix required in expectation step of EM and the second part of the maximization step of EM, respectively, are computed with $\mathcal{O}(N^3)$ operations. Owing to Line 15, Algorithm 1 is also utilized for this case with $L_a=L+N$ instead of $L_a=L$ corresponding to the noise-free case. Hence, the overall complexity is $\mathcal{O}(N^3+(L+N)NP+P^3)$ and $\mathcal{O}(N^3+(L+N)NP+8P^3)$ operations for the real- and complex-valued processes, respectively, in each EM iteration.

When $\mathbf{K}$ matrix is calculated approximately in Algorithm 3, the inversion of $N\times N$ matrix is not required owing to the disturbance smoother, and the eigendecomposition of $\mathbf{K}$ matrix is eliminated

Algorithm 3: Proposed AR parameter estimation method, also see [43].

Input : $\mathbf{Y} = [\mathbf{y}_1\, \mathbf{y}_2\, \ldots\, \mathbf{y}_L]$, P, $t_{\max}$: the maximum number of iterations, σ_v^2 (optional)
Output: $\widehat{\mathbf{a}}$, $\widehat{\sigma}_\epsilon^2$, $\widehat{\sigma}_v^2$ (or σ_v^2 if exists)

Initialize parameters $\mathbf{a}^{\text{old}} = \mathbf{a}^{\text{ini}}$, $(\sigma_\epsilon^2)^{\text{old}} = (\sigma_\epsilon^2)^{\text{ini}}$ and $(\sigma_v^2)^{\text{old}} = (\sigma_v^2)^{\text{ini}}$ by following Algorithm 4
for $t \leftarrow 1$ **to** $t_{\max}$ **do**
  `// Expectation step of EM, see (4)`
  if exact $\mathbf{K}$ calculation is desired **then**
    $\mathbf{Q} = ((\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) + (\sigma_v^2)^{\text{old}}\mathbf{I}_N)^{-1}$; `// N × N matrix inversion`
    for $l \leftarrow 1$ **to** L **do**
      $\widehat{\mathbf{x}}_l = (\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})\mathbf{Q}\mathbf{y}_l$; `// Mean vector of the posterior density` $f(\mathbf{x}_l|\mathbf{y}_l)$
    $\mathbf{K} = (\sigma_\epsilon^2)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}}) - (\sigma_\epsilon^4)^{\text{old}}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})\mathbf{Q}\mathbf{R}_{f,N}(\mathbf{a}^{\text{old}})$; `// Covariance matrix`
  else
    Compute $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1\, \widehat{\mathbf{x}}_2\, \ldots\, \widehat{\mathbf{x}}_L]$ by using the disturbance smoother; `// See Table 1`
    Construct $\mathbf{K}$ matrix as a banded matrix with P nonzero super/sub-diagonals by using $P+1$ error correlation lags, $r_e[k] = \frac{(\sigma_v^2)^{\text{old}}}{2\pi j}\oint_{C:|z|=1} H^{\text{IIR-NC}}(z)z^{k-1}dz$, $k = \{0, 1, \ldots, P\}$, where $H^{\text{IIR-NC}}(z) = (\sigma_\epsilon^2)^{\text{old}}/((\sigma_\epsilon^2)^{\text{old}} + (\sigma_v^2)^{\text{old}}A^{\text{old}}(z)(A^{\text{old}}(1/z^*))^*)$ and $A^{\text{old}}(z) = 1 + a_1^{\text{old}}z^{-1} + a_2^{\text{old}}z^{-2} + \ldots + a_P^{\text{old}}z^{-P}$.
  `// Maximization step of EM`
  `// The first stage estimation by using` $L_a = L$ `and` $\mathbf{f}_l = \widehat{\mathbf{x}}_l$, $l = \{1, 2, \ldots, L\}$ `for (12)`
  $\mathbf{a}_{\text{FS}} = -\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^H\mathbf{W}\mathbf{A}_{\ell,f} + \mathbf{A}_{\ell,b}^H\mathbf{W}\mathbf{A}_{\ell,b}\right)\right)^{-1}\left(\sum_{\ell=1}^{L_a}\left(\mathbf{A}_{\ell,f}^H\mathbf{W}\mathbf{b}_{\ell,f} + \mathbf{A}_{\ell,b}^H\mathbf{W}\mathbf{b}_{\ell,b}\right)\right)$ `// The second stage estimation`
  if exact $\mathbf{K}$ calculation is desired **then**
    $\mathbf{K} = \sum_{n=1}^{N}\lambda_n\mathbf{e}_n\mathbf{e}_n^H$; `// the eigendecomposition of` $N \times N$ `matrix`
    Form the set of $L_a = L + N$ vectors $\mathbf{f}_\ell$, $\ell = 1, 2, \ldots, L_a$, where $\mathbf{f}_l = \widehat{\mathbf{x}}_l$, $l = \{1, 2, \ldots, L\}$, and $\mathbf{f}_{L+n} = \sqrt{L\lambda_n}\mathbf{e}_n$, $n = \{1, 2, \ldots, N\}$
    Compute $\mathbf{a}^{\text{new}}$ and $(\sigma_\epsilon^2)^{\text{new}}$ by following the lines from 2 to 12 in Algorithm 1
  else
    Compute $\mathbf{a}^{\text{new}}$ and $(\sigma_\epsilon^2)^{\text{new}}$ by following Algorithm 2
  if σ_v^2 exists **then**
    $(\sigma_v^2)^{\text{new}} = \sigma_v^2$
  else
    $(\sigma_v^2)^{\text{new}} = \frac{1}{LN}\left(L\,\text{tr}(\mathbf{K}) + \sum_{l=1}^{L}\|\mathbf{y}_l - \widehat{\mathbf{x}}_l\|^2\right)$
  $\mathbf{a}^{\text{old}} = \mathbf{a}^{\text{new}}$, $(\sigma_\epsilon^2)^{\text{old}} = (\sigma_\epsilon^2)^{\text{new}}$ and $(\sigma_v^2)^{\text{old}} = (\sigma_v^2)^{\text{new}}$
Return: $\widehat{\mathbf{a}} = \mathbf{a}^{\text{new}}$, $\widehat{\sigma}_\epsilon^2 = (\sigma_\epsilon^2)^{\text{new}}$ and $\widehat{\sigma}_v^2 = (\sigma_v^2)^{\text{new}}$

by using the GS formula for the second stage estimation given in Algorithm 2. Similar to Algorithm 1, and Algorithm 2 has the complexity of $\mathcal{O}\left(LNP + P^3\right)$ and $\mathcal{O}\left(LNP + 8P^3\right)$ multiplications per EM iteration for the real- and complex-valued processes, respectively.

3. Numerical results

We present a performance comparison of the suggested AR parameter estimation method under different noise conditions including noise-free case. The performance comparison of AR parameter estimation methods requires an application specific fidelity criterion such as spectral peak location, spurious peak avoidance, filter coefficient/pole accuracy, etc. as discussed at length in Kay [42]. In this study, we use objective metrics such as total MSE $E\{\|\mathbf{a} - \widehat{\mathbf{a}}\|^2\}$, attained likelihood value and the Hellinger distance between true and estimated Gaussian random vector densities. The Hellinger distance between zero-mean Gaussian vectors $\mathbf{p}$ and $\mathbf{q}$ with covariance matrices Σ_1 and Σ_2 is given as $H(\mathbf{p}, \mathbf{q}) = \sqrt{1 - |\Sigma_1|^{1/4}|\Sigma_2|^{1/4}/|\frac{1}{2}(\Sigma_1 + \Sigma_2)|^{1/2}}$, [45]. The Hellinger distance is a true metric satisfying positivity, symmetry, triangle inequality axioms. The Hellinger distance is utilized to upper bound the detection error of equally likely hypotheses $\mathbf{p}$ and $\mathbf{q}$ with $p_e \leq 1/2(1 - H^2(\mathbf{p}, \mathbf{q}))$, and used as a robust measure for the distance between distributions, [46]. All experiments are conducted by using 100 Monte Carlo runs.

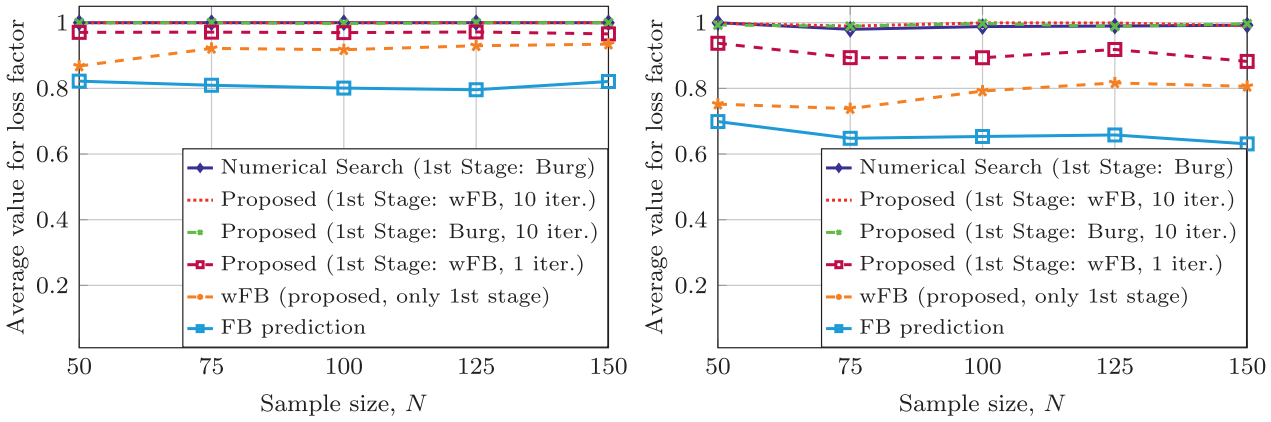
Experiment 1 - Multiple noise-free snapshots: This experiment compares the likelihood values attained by the estimates of different methods including the forward-backward prediction (FB), the weighted forward-backward prediction (wFB), which performs only the first stage estimation presented in Section 2.2.1, and the numerical search method initialized with the Burg's method. For comparison purposes, one of the cases presented in Candan [6] that is the estimation of the real- and complex-valued AR(6) parameters randomly generated from the uniformly distributed reflection coefficients and $\sigma_\epsilon^2 = 0.36$ for a single snapshot scenarios is repeated for $L = 10$ snapshots. The results in Fig. 2 closely follow the earlier conclusions and reveal that the proposed approach yields likelihood values which are almost identical to the ones attained by the numerical search having much higher complexity.

Experiment 2 - Multiple noisy snapshots: This experiment examines the estimation accuracy of the method on an AR(4) system with $A(z) = 1 + 0.1z^{-1} + 0.2z^{-2} + 0.3z^{-3} + 0.4z^{-4}$ under unity measurement noise variance, $\sigma_v^2 = 1$. The σ_ϵ^2 parameter in (1) is varied according to the experiment SNR. Fig. 3 shows the Hellinger distance and total MSE ($E\{\|\mathbf{a} - \widehat{\mathbf{a}}\|^2\}$) comparisons for different number of snapshots L, where each snapshot is a vector of length $N = 50$. Four estimators, for the cases of $\mathbf{K}$ is calculated exactly or approximately and σ_v^2 is known exactly or estimated, are compared. Fig. 3 also includes the estimator performance in the noise-free scenario. The estimator for the noise-free case has $\widehat{\mathbf{x}}_l = \mathbf{y}_l$ and $\mathbf{K} = \mathbf{0}_{N\times N}$. These relations can be also retrieved from (4) as $\sigma_v^2 \to 0$. The asymptotic CRB (ACRB) [8,9] is provided as a performance benchmark for total MSE comparisons. The results of a similar setup constructed to compare the estimation accuracy for the complex-valued AR(4) system with $A(z) = 1 + (0.1 + j0.2)z^{-1} + (0.2 - j0.3)z^{-2} + (0.3 + j0.1)z^{-3} + (0.4 - j0.2)z^{-4}$ and $\sigma_v^2 = 1$ are shown in Fig. 4.

We note from Figs. 3 and 4 that there is a significant performance gap between the cases of known and unknown measurement noise variance. Yet, in both cases the total MSE value for $\mathbf{a}$ vector coincides the ACRB associated with the problem. More detailed numerical results on the real-valued case for SNR of 10 dB is given in Table 2, where the column labeled as "CRB" is the non-asymptotic (exact) bound for the problem by numerical differentiation [47, Eq. (3.31), p. 47].

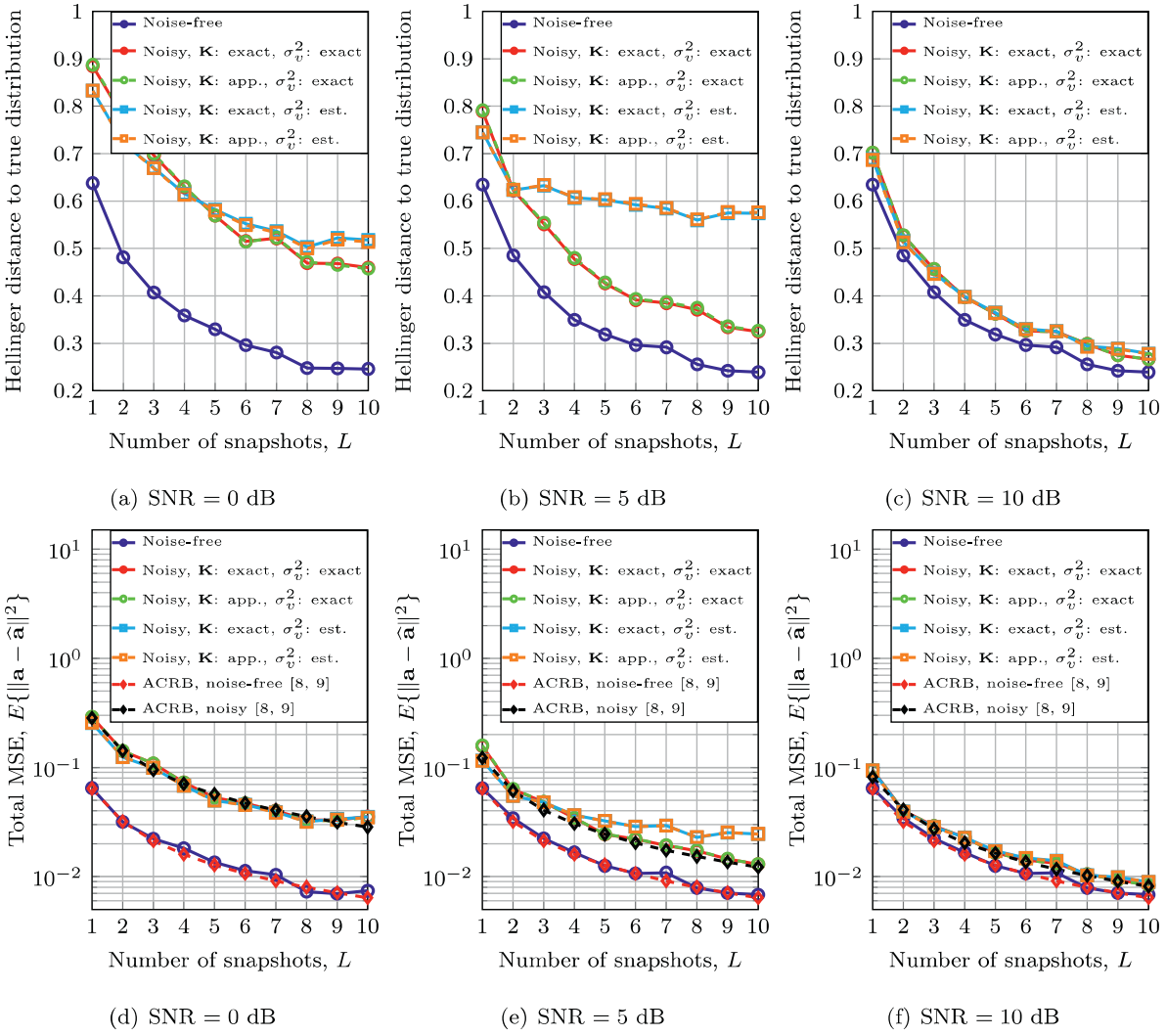
The results shown in Fig. 3 with SNR $= 0$ dB and Fig. 4 with SNR $= 0$ dB and SNR $= 5$ dB are obtained by initializing the suggested method with (A.7), instead of (A.5) and (A.6) used in Algorithm 3. Further details on algorithm initialization is given in Appendix A.

Experiment 3 - Single noisy snapshot: This experiment compares the estimation accuracy of the proposed method for unknown noise variance with the state-of-the-art methods, SS [16], EIV [17], IFILS [15], XZ [18], and four methods (EVK-1, EVK-2, EVK-3, EVK-4) given in Esfandiari et al. [25]. The experiment is con-



(a) Real-valued case (b) Complex-valued case

Fig. 2. Experiment 1 - Multiple noise-free snapshots: Likelihood value comparison for AR(6) parameters randomly generated from the uniformly distributed reflection coefficients and $\sigma_\epsilon^2 = 0.36$.



(a) SNR = 0 dB (b) SNR = 5 dB (c) SNR = 10 dB

(d) SNR = 0 dB (e) SNR = 5 dB (f) SNR = 10 dB

Fig. 3. Experiment 2 - The Hellinger distance metric and total MSE comparisons of proposed method at different number of snapshots for the real-valued AR(4) process.

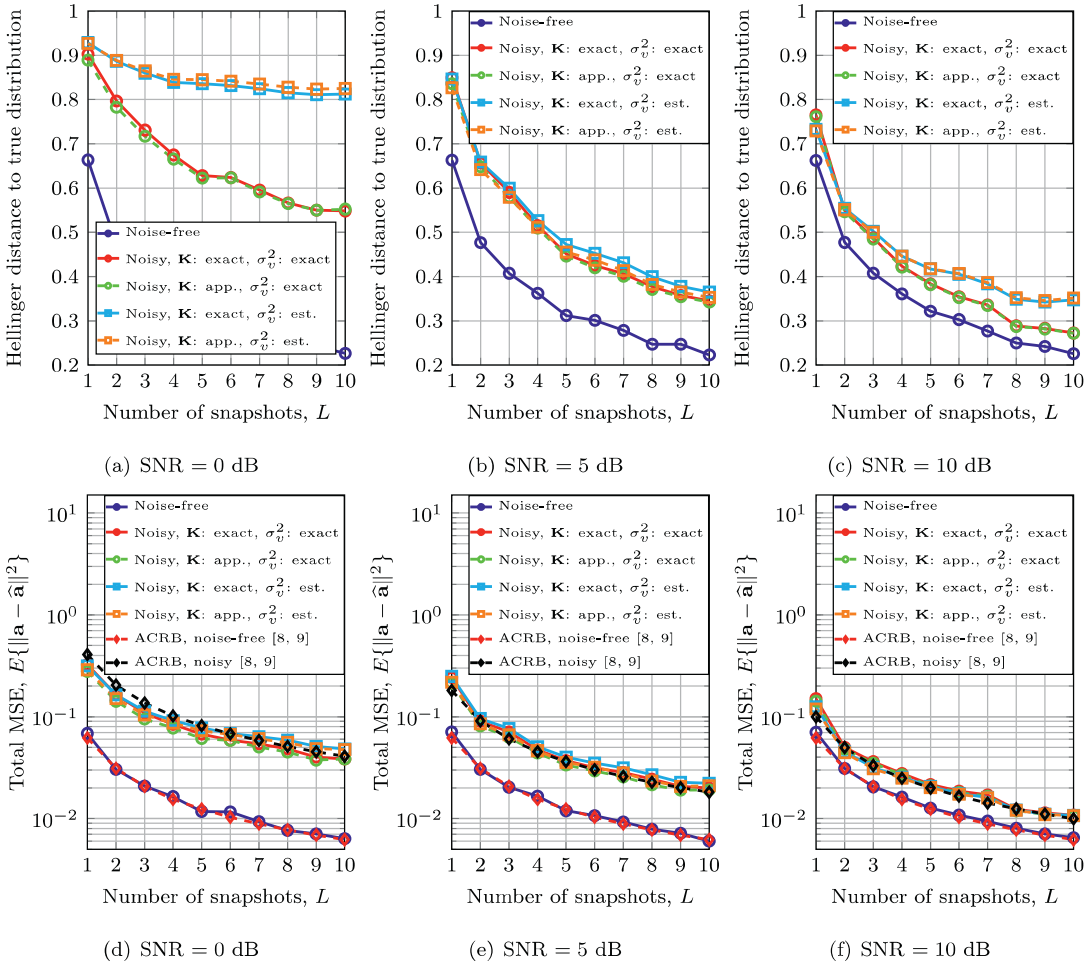


Fig. 4. Experiment 2 - The Hellinger distance metric and total MSE comparisons of proposed method at different number of snapshots for the complex-valued AR(4).

Table 2
The Hellinger distance metric and total MSE comparisons of the proposed method for different sample sizes and number of snapshots on the real-valued AR(4) process given in Experiment 2 (SNR = 10 dB).

		Hellinger distance				$E\{\|\mathbf{a}-\widehat{\mathbf{a}}\|^2\}$					
		K: exact	**K**: app.	**K**: exact	**K**: app.	**K**: exact	**K**: app.	**K**: exact	**K**: app.	ACRB	CRB
N	L	σ_v^2: exact	σ_v^2: exact	σ_v^2: est.	σ_v^2: est.	σ_v^2: exact	σ_v^2: exact	σ_v^2: est.	σ_v^2: est.	[8,9]	σ_v^2: exact
50	1	0.7017	0.7022	0.6884	0.6869	0.0946	0.0948	0.0944	0.0938	0.0817	0.0858
	5	0.3624	0.3627	0.3559	0.3551	0.0163	0.0163	0.0164	0.0164	0.0163	0.0165
	10	0.2602	0.2604	0.2642	0.2643	0.0085	0.0086	0.0090	0.0090	0.0082	0.0082
100	1	0.7041	0.7043	0.6870	0.6866	0.0420	0.0420	0.0424	0.0424	0.0409	0.0418
	5	0.3799	0.3801	0.4270	0.4272	0.0089	0.0089	0.0099	0.0099	0.0082	0.0082
	10	0.2680	0.2681	0.3580	0.3585	0.0043	0.0043	0.0053	0.0053	0.0041	0.0041
150	1	0.6905	0.6906	0.6870	0.6869	0.0280	0.0280	0.0283	0.0283	0.0272	0.0277
	5	0.3794	0.3796	0.4705	0.4706	0.0060	0.0060	0.0078	0.0078	0.0054	0.0055
	10	0.2590	0.2591	0.4288	0.4291	0.0028	0.0028	0.0046	0.0046	0.0027	0.0027
500	1	0.7080	0.7081	0.8247	0.8247	0.0087	0.0087	0.0114	0.0114	0.0082	0.0082
	5	0.3652	0.3653	0.7804	0.7804	0.0017	0.0017	0.0048	0.0048	0.0016	0.0016
	10	0.2645	0.2645	0.7771	0.7771	0.0008	0.0008	0.0041	0.0041	0.0008	0.0008
1000	1	0.6882	0.6882	0.9210	0.9210	0.0043	0.0043	0.0074	0.0074	0.0041	0.0041
	5	0.3733	0.3733	0.9295	0.9295	0.0009	0.0009	0.0046	0.0046	0.0008	0.0008
	10	0.2631	0.2631	0.9222	0.9222	0.0004	0.0004	0.0038	0.0038	0.0004	0.0004

ducted on the single snapshot real-valued AR(4) process defined in Experiment 2. Similar to Example 1 given in Esfandiari et al. [25], the parameter settings are as follows: $q = 2$ and $\delta = 0.001$ for IFILS; $\delta_1 = 0.01$ when SNR = 0 dB and $\delta_1 = 0.001$ when SNR = 5 dB or SNR = 10 dB, $\delta_2 = 0.01$ and $\eta = 0.96$ for XZ; $q = 3$ and $\delta = 0.1$ for EVK-1 and EVK-2; $q = 4$ and $\delta = 0.1$ for EVK-3; $q = 3$ and $m = 8$ for EVK-4. (Readers can consult [25] for the parameter descriptions.)

Table 3 shows that the suggested method using either exact or approximated error covariance matrix **K** provides more accurate AR parameter estimates in terms of Hellinger distance and total MSE in comparison to other methods by denoting the best attained

Table 3
The Hellinger distance metric (first line), total MSE (second line) and the number of unstable system estimates out of 100 runs (third line) comparisons of the proposed method with other methods for the single snapshot real-valued AR(4) process defined in Experiment 2.

N	SNR	**K**: exact σ_v^2: est.	**K**: app. σ_v^2: est.	SS [16]	EIV [17]	IFILS [15]	XZ [18]	EVK-1 [25]	EVK-2 [25]	EVK-3 [25]	EVK-4 [25]
50	0 dB	**0.8312**	0.8315	0.9326	0.9387	0.9402	0.9515	0.9355	0.9441	0.9630	0.8967
		0.2520	0.2534	0.6938	0.3245	0.2554	0.3685	0.5761	0.6164	0.6421	0.3814
		–	–	70	10	13	36	31	35	19	–
	5 dB	0.7451	**0.7446**	0.9299	0.8501	0.8589	0.9472	0.8465	0.9359	0.9034	0.8925
		0.1156	**0.1152**	0.3225	0.1777	0.1691	0.2993	0.2921	0.3943	0.4148	0.2310
		–	–	62	10	8	20	19	22	12	–
	10 dB	0.6884	**0.6869**	0.9226	0.8177	0.8109	0.9149	0.7843	0.9057	0.8659	0.9076
		0.0944	**0.0938**	0.2672	0.1486	0.1456	0.2075	0.2017	0.2301	0.3716	0.1888
		–	–	60	6	11	26	14	14	8	–
100	0 dB	0.8558	**0.8555**	0.9596	0.9535	0.9602	0.9858	0.9606	0.9781	0.9782	0.9259
		0.1265	0.1266	0.9673	0.2867	0.2675	0.4038	0.5159	0.6870	0.5805	0.3014
		–	–	60	10	14	26	15	21	10	–
	5 dB	**0.8297**	0.8299	0.9314	0.8687	0.9177	0.9417	0.8595	0.9487	0.9436	0.9623
		0.0620	**0.0620**	0.2090	0.1213	0.1450	0.1772	0.1587	0.2430	0.3508	0.1452
		–	–	33	4	7	30	10	16	2	–
	10 dB	0.6870	**0.6866**	0.9190	0.8027	0.8405	0.9137	0.7917	0.9048	0.8869	0.9755
		0.0424	**0.0424**	0.1943	0.0867	0.0884	0.1860	0.1572	0.2482	0.3240	0.1193
		–	–	34	1	5	17	8	11	5	–
150	0 dB	0.8703	**0.8698**	0.9594	0.9623	0.9704	0.9703	0.9602	0.9877	0.9752	0.9576
		0.0849	**0.0848**	0.2577	0.1860	0.1711	0.2945	0.4021	0.4964	0.4038	0.2034
		–	–	46	8	8	29	15	19	7	–
	5 dB	0.8893	0.8895	0.9351	**0.8788**	0.9105	0.9286	0.8885	0.9347	0.9432	0.9767
		0.0502	0.0502	0.1090	0.0764	0.1030	0.1918	0.1825	0.2439	0.2787	0.1171
		–	–	19	3	6	16	9	14	7	–
	10 dB	0.6870	**0.6869**	0.9291	0.8233	0.8753	0.8919	0.8122	0.8950	0.8914	0.9855
		0.0283	**0.0283**	0.1235	0.0460	0.0668	0.1647	0.1052	0.1628	0.2125	0.0881
		–	–	23	–	1	15	2	5	3	–

Table 4
The average computational time (in seconds) comparison of the proposed method for different sample sizes and number of snapshots for Experiment 2 (SNR = 10 dB). Rt denotes the ratio of computational times.

		Real-valued AR(4)						Complex-valued AR(4)					
N	L	**K**: exact σ_v^2: exact	**K**: app. σ_v^2: exact	Rt	**K**: exact σ_v^2: est.	**K**: app. σ_v^2: est.	Rt	**K**: exact σ_v^2: exact	**K**: app. σ_v^2: exact	Rt	**K**: exact σ_v^2: est.	**K**: app. σ_v^2: est.	Rt
50	1	0.2810	0.0372	7.55	0.2802	0.0370	7.58	0.5379	0.0562	9.56	0.5351	0.0552	9.69
	5	0.3008	0.0654	4.60	0.3009	0.0654	4.60	0.5829	0.1083	5.38	0.5812	0.1082	5.37
	10	0.3270	0.0941	3.47	0.3270	0.0942	3.47	0.6369	0.1628	3.91	0.6380	0.1638	3.90
100	1	0.5801	0.0427	13.57	0.5806	0.0428	13.55	1.2242	0.0729	16.79	1.2054	0.0725	16.62
	5	0.6071	0.0810	7.49	0.6072	0.0812	7.48	1.2338	0.1328	9.29	1.2373	0.1361	9.09
	10	0.6327	0.1152	5.49	0.6348	0.1151	5.52	1.2879	0.1951	6.60	1.2800	0.1940	6.60
150	1	0.9734	0.0519	18.75	0.9723	0.0518	18.76	1.9246	0.0811	23.72	1.9184	0.0814	23.58
	5	0.9845	0.0976	10.09	0.9863	0.0976	10.10	1.9972	0.1590	12.56	1.9967	0.1607	12.43
	10	1.0143	0.1377	7.36	1.0160	0.1380	7.36	2.0345	0.2273	8.95	2.0287	0.2285	8.88
500	1	5.6314	0.1071	52.58	5.8269	0.1070	54.48	13.7240	0.1844	74.43	13.7681	0.1850	74.42
	5	5.7710	0.2515	22.94	5.9788	0.2525	23.68	14.2745	0.4096	34.85	14.2137	0.4110	34.58
	10	5.9106	0.3766	15.69	6.1155	0.3774	16.20	14.8254	0.6022	24.62	14.7588	0.6093	24.22
1000	1	18.9031	0.2096	90.19	19.0119	0.2093	90.83	48.2300	0.3524	136.88	48.0578	0.3531	136.09
	5	20.2822	0.5896	34.40	20.5374	0.5886	34.89	48.8622	0.8417	58.05	48.3051	0.8422	57.35
	10	21.2807	0.9352	22.75	21.6083	0.9370	23.06	51.6543	1.2907	40.02	50.9763	1.2935	39.41

value of performance metrics (the lowest Hellinger distance and total MSE) with boldface. Table 3 also includes information on the number of unstable AR systems out of 100 trials. (The results for SNR = 0 dB are obtained by initializing with (A.7).)

Experiment 4 - Average computational time: This experiment compares the average computational time (in seconds) of the proposed method for different sample sizes N and the number of snapshots L in Experiment 2 set-up with SNR = 10 dB. Table 4 shows that the suggested method with approximate error covariance matrix **K** (eliminating the inversion of $N \times N$ matrix via the disturbance smoother in expectation step and the eigendecomposition via the GS formula in maximization step of EM) requires significantly less CPU time than the suggested method with exact **K** matrix. The ratios of average computational times are given in the column of Table 4 labeled as "Rt".

4. Conclusion

An expectation-maximization based solution for noisy AR parameter estimation problem and its efficient implementation are given in this study. The heart of the method contains an extension of the formulation given in Candan [6] for a single snapshot likelihood maximization of AR parameter estimation problem to the multiple snapshots. In addition to this, the current formulation examines the problem of AR parameter estimation under the effect of white noise with the unknown variance given multiple independent snapshots. A highly efficient, yet approximate, implementation of the suggested method is also given. The performance of the approximate version is almost identical to the exact version; but the approximate version eliminates $N^3 + N^2P$ multiplications per EM iteration (N is the snapshot vector length) resulting in significant cost savings in both computation and memory. We present

the ready-to-use MATLAB codes of the proposed method reproducing the presented numerical results for further exploration in Çayır and Candan [43].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Initialization of suggested algorithm

The proposed method requires the initial estimates for $\mathbf{a}$, σ_ϵ^2 and σ_v^2, see (4). By applying the weighted forward-backward prediction approach on $\mathbf{y}_l$, $l=\{1,2,\ldots,L\}$, as done for $\mathbf{f}_\ell$ in (11), $\ell=1,\ldots,L_a$, the initial estimate of $\mathbf{a}$, $\mathbf{a}^{\text{ini}}$, is calculated as

$$\mathbf{a}^{\text{ini}} = -\left(\sum_{l=1}^{L}\left(\mathbf{A}_{l,f}^H\mathbf{W}\mathbf{A}_{l,f}+\mathbf{A}_{l,b}^H\mathbf{W}\mathbf{A}_{l,b}\right)\right)^{-1} \left(\sum_{l=1}^{L}\left(\mathbf{A}_{l,f}^H\mathbf{W}\mathbf{b}_{l,f}+\mathbf{A}_{l,b}^H\mathbf{W}\mathbf{b}_{l,b}\right)\right) \tag{A.1}$$

where

$$\mathbf{A}_{l,f}=\begin{bmatrix} y_{l,P} & y_{l,P-1} & \cdots & y_{l,1}\\ y_{l,P+1} & y_{l,P} & \cdots & y_{l,2}\\ \vdots & \vdots & \cdots & \vdots\\ y_{l,N-1} & y_{l,N-2} & \cdots & y_{l,N-P}\end{bmatrix},\quad \mathbf{b}_{l,f}=\begin{bmatrix} y_{l,P+1}\\ y_{l,P+2}\\ \vdots\\ y_{l,N}\end{bmatrix},$$

$$\mathbf{A}_{l,b}=\begin{bmatrix} y_{l,N-P+1}^* & y_{l,N-P+2}^* & \cdots & y_{l,N}^*\\ y_{l,N-P}^* & y_{l,N-P+1}^* & \cdots & y_{l,N-1}^*\\ \vdots & \vdots & \cdots & \vdots\\ y_{l,2}^* & y_{l,3}^* & \cdots & y_{l,P+1}^*\end{bmatrix},\quad \mathbf{b}_{l,b}=\begin{bmatrix} y_{l,N-P}^*\\ y_{l,N-P-1}^*\\ \vdots\\ y_{l,1}^*\end{bmatrix},$$

and $\mathbf{W}=\text{diag}(1,2,\ldots,N-P)$. Then, the initial values of σ_ϵ^2 and σ_v^2 are determined by using YW equations and $\mathbf{a}^{\text{ini}}$.

According to the block diagram shown in Fig. 1, the YW equations for the autocorrelation sequence of the AR(P) process $x[n]$ can be written as

$$r_x[k] = -\sum_{p=1}^{P} a_p r_x[k-p] + \sigma_\epsilon^2\delta[k],\ k\geqslant 0, \tag{A.2}$$

where $r_x[k]=E\{x[n]x^*[n-k]\}$; and the YW equations for the autocorrelation sequence of $y[n]$ can be expressed as

$$r_y[k] = r_x[k] + \sigma_v^2\delta[k] k\geqslant 0, \tag{A.3}$$

which implies that $r_y[k]=r_x[k]$ for $k>0$, [1]. Using (A.2) and (A.3), we get

$$\left(\mathbf{R}_y-\sigma_v^2\mathbf{I_P}\right)\mathbf{a} = -\mathbf{r}_y, \tag{A.4}$$

where

$$\mathbf{R}_y=\begin{bmatrix} r_y[0] & r_y[-1] & \ldots & r_y[-P+1]\\ r_y[1] & r_y[0] & \ldots & r_y[-P+2]\\ \vdots & \vdots & \ddots & \vdots\\ r_y[P-1] & r_y[P-2] & \ldots & r_y[0]\end{bmatrix},\quad \mathbf{r}_y=\begin{bmatrix} r_y[1]\\ r_y[2]\\ \vdots\\ r_y[P]\end{bmatrix}.$$

In (A.4), $\mathbf{R}_y$ and $\mathbf{r}_y$ are replaced with $\widehat{\mathbf{R}}_y$ and $\widehat{\mathbf{r}}_y$ estimates formed by using $\widehat{r}_y[k]=\frac{1}{L}\sum_{l=1}^{L}\widehat{r}_{y_l}[k]$, where $\widehat{r}_{y_l}[k]=\frac{1}{N}\sum_{n=k}^{N-1}y_l[n]y_l^*[n-k]$ for $k=\{0,1,\ldots,P\}$ and $l=\{1,2,\ldots,L\}$, and $\widehat{r}_y[-k]=\widehat{r}_y^*[k]$. Replacing $\mathbf{a}$ with $\mathbf{a}^{\text{ini}}$, the initial value of σ_v^2 is the LS solution of $(\widehat{\mathbf{R}}_y-\sigma_v^2\mathbf{I_P})\mathbf{a}^{\text{ini}}=-\widehat{\mathbf{r}}_\mathbf{y}$ for σ_v^2

$$(\sigma_v^2)^{\text{ini}} = \frac{\Re\left\{(\mathbf{a}^{\text{ini}})^H(\widehat{\mathbf{r}}_y+\widehat{\mathbf{R}}_y\mathbf{a}^{\text{ini}})\right\}}{\|\mathbf{a}^{\text{ini}}\|^2}, \tag{A.5}$$

Algorithm 4: The initialization of the suggested method, also see [43].

Input : $\mathbf{Y}=[\mathbf{y}_1\,\mathbf{y}_2\,\ldots\,\mathbf{y}_L]$, P, σ_v^2 (optional)
Output: $\mathbf{a}^{\text{ini}}$, $(\sigma_\epsilon^2)^{\text{ini}}$, $(\sigma_v^2)^{\text{ini}}$

$\mathbf{a}^{\text{ini}} = -\left(\sum_{l=1}^{L}\left(\mathbf{A}_{l,f}^H\mathbf{W}\mathbf{A}_{l,f}+\mathbf{A}_{l,b}^H\mathbf{W}\mathbf{A}_{l,b}\right)\right)^{-1} \left(\sum_{l=1}^{L}\left(\mathbf{A}_{l,f}^H\mathbf{W}\mathbf{b}_{l,f}+\mathbf{A}_{l,b}^H\mathbf{W}\mathbf{b}_{l,b}\right)\right)$

if σ_v^2 exists **then**

  $(\sigma_v^2)^{\text{ini}}=\sigma_v^2$, and $(\sigma_\epsilon^2)^{\text{ini}}=\frac{1}{LN}\sum_{l=1}^{L}\mathbf{y}_l^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{ini}})\mathbf{y}_l-\sigma_v^2$

else

  $(\sigma_v^2)^{\text{ini}}=\Re\left\{(\mathbf{a}^{\text{ini}})^H(\widehat{\mathbf{r}}_y+\widehat{\mathbf{R}}_y\mathbf{a}^{\text{ini}})\right\}/\|\mathbf{a}^{\text{ini}}\|^2$ and
  $(\sigma_\epsilon^2)^{\text{ini}}=\widehat{r}_y[0]+\widehat{\mathbf{r}}_y^H\mathbf{a}^{\text{ini}}-(\sigma_v^2)^{\text{ini}}$,
  or $(\sigma_v^2)^{\text{ini}}=(\sigma_\epsilon^2)^{\text{ini}}=\frac{1}{2LN}\sum_{l=1}^{L}\mathbf{y}_l^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{ini}})\mathbf{y}_l$

Return: $\mathbf{a}^{\text{ini}}$, $(\sigma_\epsilon^2)^{\text{ini}}$, $(\sigma_v^2)^{\text{ini}}$

where $\Re\{\cdot\}$ denotes the real part of its argument, see [17, Eq. 46] and [25, Eq. 23] for the real-valued AR process. Using YW Eqs. (A.2) and (A.3) for $k=0$, the initial value of σ_ϵ^2 is calculated as

$$(\sigma_\epsilon^2)^{\text{ini}} = \widehat{r}_y[0] + (\mathbf{a}^{\text{ini}})^T\widehat{\mathbf{r}}_y^* - (\sigma_v^2)^{\text{ini}} = \widehat{r}_y[0] + \widehat{\mathbf{r}}_y^H\mathbf{a}^{\text{ini}} - (\sigma_v^2)^{\text{ini}}. \tag{A.6}$$

For the initialization of σ_ϵ^2 and σ_v^2, another suggestion can be made by assuming that $(\sigma_v^2)^{\text{ini}}=(\sigma_\epsilon^2)^{\text{ini}}$ and following the approach for the estimate of $\widehat{\sigma}_\epsilon^2=\frac{1}{N}\mathbf{x}^H\mathbf{R}_{f,N}^{-1}\mathbf{x}$ given in Candan [6] such that $(\sigma_v^2)^{\text{ini}}+(\sigma_\epsilon^2)^{\text{ini}}=\frac{1}{LN}\sum_{l=1}^{L}\mathbf{y}_l^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{ini}})\mathbf{y}_l$. Thus, we have

$$(\sigma_v^2)^{\text{ini}} = (\sigma_\epsilon^2)^{\text{ini}} = \frac{\sum_{l=1}^{L}\mathbf{y}_l^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{ini}})\mathbf{y}_l}{2LN}. \tag{A.7}$$

If the measurement noise variance σ_v^2 is known, say the unknown parameter set reduces to $\Theta=\{\sigma_\epsilon^2,\mathbf{a}\}$ from $\Theta=\{\sigma_v^2,\sigma_\epsilon^2,\mathbf{a}\}$, then the true value σ_v^2 can be substituted for $(\sigma_v^2)^{\text{ini}}$ in $(\sigma_v^2)^{\text{ini}}+(\sigma_\epsilon^2)^{\text{ini}}=\frac{1}{LN}\sum_{l=1}^{L}\mathbf{y}_l^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{ini}})\mathbf{y}_l$ and we get $(\sigma_\epsilon^2)^{\text{ini}}=\frac{\sum_{l=1}^{L}\mathbf{y}_l^H\mathbf{R}_{f,N}^{-1}(\mathbf{a}^{\text{ini}})\mathbf{y}_l}{LN}-\sigma_v^2$.

The initialization methods are summarized in Algorithm 4.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sigpro.2021.108118.

CRediT authorship contribution statement

Ömer Çayır: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - review & editing, Visualization. **Çağatay Candan:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision.

References

[1] M.H. Hayes, Statistical Digital Signal Processing and Modeling, Wiley, 1996.
[2] P. Wang, H. Li, B. Himed, A Bayesian parametric test for multichannel adaptive signal detection in nonhomogeneous environments, IEEE Signal Process. Lett. 17 (4) (2010) 351–354, doi:10.1109/LSP.2009.2039380.
[3] J. Petitjean, R. Diversi, E. Grivel, R. Guidorzi, P. Roussilhe, Recursive errors-in-variables approach for AR parameter estimation from noisy observations. Application to radar sea clutter rejection, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 5, IEEE, 2009, pp. 3401–3404, doi:10.1109/ICASSP.2009.4960355.

[4] Y. Shekofteh, F. Almasganj, Autoregressive modeling of speech trajectory transformed to the reconstructed phase space for ASR purposes, Digit. Signal Process. 23 (6) (2013) 1923–1932, doi:10.1016/j.dsp.2013.06.011.
[5] M. Aktaruzzaman, R. Sassi, Parametric estimation of sample entropy in heart rate variability analysis, Biomed. Signal Process. Control 14 (1) (2014) 141–147, doi:10.1016/j.bspc.2014.07.011.
[6] C. Candan, Making linear prediction perform like maximum likelihood in Gaussian autoregressive model parameter estimation, Signal Process. 166 (2020) 107256, doi:10.1016/j.sigpro.2019.107256. http://www.sciencedirect.com/science/article/pii/S0165168419303081
[7] S.M. Kay, Noise compensation for autoregressive spectral estimates, IEEE Trans. Acoust. Speech Signal Process. 28 (3) (1980) 292–303, doi:10.1109/TASSP.1980.1163406.
[8] L. Weruaga, O.M. Melko, Asymptotic Cramér–Rao bound for noise-compensated autoregressive analysis, IEEE Trans. Circuits Syst. I 59 (9) (2012) 2017–2024, doi:10.1109/TCSI.2012.2185277.
[9] L. Weruaga, L. Dimitrov, The spectral nature of maximum likelihood noise compensated linear prediction, IEEE Trans. Audio Speech Lang. Process. 21 (8) (2013) 1760–1765, doi:10.1109/TASL.2013.2255277.
[10] H. Tong, Autoregressive model fitting with noisy data by Akaike's information criterion (corresp.), IEEE Trans. Inf. Theory 21 (4) (1975) 476–480, doi:10.1109/TIT.1975.1055402.
[11] S.M. Kay, Modern Spectral Estimation: Theory and Application, Prentice Hall Englewood Cliffs, 1988.
[12] A. Nehorai, P. Stoica, Adaptive algorithms for constrained ARMA signals in the presence of noise, IEEE Trans. Acoust. Speech Signal Process. 36 (8) (1988) 1282–1291, doi:10.1109/29.1656.
[13] W.X. Zheng, Autoregressive parameter estimation from noisy data, IEEE Trans. Circuits Syst. II 47 (1) (2000) 71–75, doi:10.1109/82.818897.
[14] W.X. Zheng, On estimation of autoregressive signals in the presence of noise, IEEE Trans. Circuits Syst. II 53 (12) (2006) 1471–1475, doi:10.1109/TCSII.2006.883094.
[15] A. Mahmoudi, M. Karimi, Inverse filtering based method for estimation of noisy autoregressive signals, Signal Process. 91 (7) (2011) 1659–1664, doi:10.1016/j.sigpro.2011.01.008.
[16] C. Davila, A subspace approach to estimation of autoregressive parameters from noisy measurements, IEEE Trans. Signal Process. 46 (2) (1998) 531–534, doi:10.1109/78.655442.
[17] R. Diversi, R. Guidorzi, U. Soverini, Identification of autoregressive models in the presence of additive noise, Int. J. Adapt. Control Signal Process. 22 (5) (2008) 465–481, doi:10.1002/acs.989.
[18] Y. Xia, W.X. Zheng, Novel parameter estimation of autoregressive signals in the presence of noise, Automatica 62 (2015) 98–105, doi:10.1016/j.automatica.2015.09.008.
[19] D. Labarre, E. Grivel, Y. Berthoumieu, E. Todini, M. Najim, Consistent estimation of autoregressive parameters from noisy observations based on two interacting Kalman filters, Signal Process. 86 (10) (2006) 2863–2876, doi:10.1016/j.sigpro.2005.12.001.
[20] J. Treichler, Transient and convergent behavior of the adaptive line enhancer, IEEE Trans. Acoust. Speech Signal Process. 27 (1) (1979) 53–62, doi:10.1109/TASSP.1979.1163200.
[21] W.-R. Wu, P.-C. Chen, Adaptive AR modeling in white Gaussian noise, IEEE Trans. Signal Process. 45 (5) (1997) 1184–1192, doi:10.1109/78.575693.
[22] Y. Zhang, C. Wen, Y.C. Soh, Unbiased LMS filtering in the presence of white measurement noise with unknown power, IEEE Trans. Circuits Syst. II 47 (9) (2000) 968–972, doi:10.1109/82.868469.
[23] M. Gabrea, E. Grivel, M. Najun, A single microphone Kalman filter-based noise canceller, IEEE Signal Process. Lett. 6 (3) (1999) 55–57, doi:10.1109/97.744623.
[24] R. Mehra, On the identification of variances and adaptive Kalman filtering, IEEE Trans. Autom. Control 15 (2) (1970) 175–184, doi:10.1109/TAC.1970.1099422.
[25] M. Esfandiari, S.A. Vorobyov, M. Karimi, New estimation methods for autoregressive process in the presence of white observation noise, Signal Process. 171 (2020) 107480, doi:10.1016/j.sigpro.2020.107480.
[26] H. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, doi:10.1007/978-3-540-27752-1.
[27] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. 39 (1) (1977) 1–22, doi:10.1111/j.2517-6161.1977.tb01600.x.
[28] M. Deriche, AR parameter estimation from noisy data using the EM algorithm, in: Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing, iv, IEEE, 1994, pp. IV/69–IV/72, doi:10.1109/ICASSP.1994.389874.
[29] S. Gannot, D. Burshtein, E. Weinstein, Iterative and sequential Kalman filter-based speech enhancement algorithms, IEEE Trans. Speech Audio Process. 6 (4) (1998) 373–385, doi:10.1109/89.701367.
[30] C.F.J. Wu, On the convergence properties of the EM algorithm, Ann. Stat. 11 (1) (1983) 95–103.
[31] A. El-Jaroudi, J. Makhoul, Discrete all-pole modeling, IEEE Trans. Signal Process. 39 (2) (1991) 411–423, doi:10.1109/78.80824.
[32] L. Weruaga, Frequency-selective noise-compensated autoregressive estimation, IEEE Trans. Circuits Syst. I 58 (10) (2011) 2469–2476, doi:10.1109/TCSI.2011.2142830.
[33] A.M. Sykulski, S.C. Olhede, A.P. Guillaumin, J.M. Lilly, J.J. Early, The debiased Whittle likelihood, Biometrika 106 (2) (2019) 251–266, doi:10.1093/biomet/asy071.
[34] P. Whittle, Estimation and information in stationary time series, Arkiv Mat. 2 (5) (1953) 423–434, doi:10.1007/BF02590998.
[35] P.M.T. Broersen, Automatic Autocorrelation and Spectral Analysis, Springer-Verlag, London, 2006, doi:10.1007/1-84628-329-9.
[36] S. Haykin, B.W. Currie, S.B. Kesler, Maximum-entropy spectral analysis of radar clutter, Proc. IEEE 70 (9) (1982) 953–962, doi:10.1109/PROC.1982.12426.
[37] S. de Waele, P.M.T. Broersen, The Burg algorithm for segments, IEEE Trans. Signal Process. 48 (10) (2000) 2876–2880, doi:10.1109/78.869039.
[38] S. de Waele, P.M.T. Broersen, Spectral analysis of segmented data, in: Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No. 00CH37187), 1, 2000, pp. 189–190, doi:10.1109/CDC.2000.912756.
[39] M.A. Richards, Fundamentals of Radar Signal Processing, McGraw-Hill, New York, 2005.
[40] P. Stoica, R. Moses, Spectral Analysis of Signals, Prentice Hall, 2005.
[41] O. Cappe, E. Moulines, T. Ryden, Inference in Hidden Markov Models, Springer, 2005.
[42] S.M. Kay, Recursive maximum likelihood estimation of autoregressive processes, IEEE Trans. Acoust. Speech Signal Process. 31 (1) (1983) 56–65, doi:10.1109/TASSP.1983.1164050.
[43] Ö. Çayır, Ç. Candan, Maximum likelihood estimator for noisy autoregressive model parameter estimation problem with multiple snapshots, 2020, MATLAB code files. doi:10.24433/CO.7257794.v1.
[44] R.M. Gray, Toeplitz and Circulant Matrices: A Review, Now Publishers, 2006.
[45] A.B. Tsybakov, Introduction to Nonparametric Estimation, Springer, 2008.
[46] D.L. Donoho, X. Huo, Large-sample modulation classification using Hellinger representation, in: First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, 1997, pp. 133–136, doi:10.1109/SPAWC.1997.630175.
[47] S.M. Kay, Fundamentals of Statistical Processing, Volume I: Estimation Theory, Prentice Hall, 1993.