Asymptotic Cramér–Rao Bound for Noise-Compensated Autoregressive Analysis

Luis Weruaga, Senior Member, IEEE and O. Michael Melko

Abstract—Noise-compensated autoregressive (AR) analysis is a problem insufficiently explored with regard to the accuracy of the estimate. This paper studies comprehensively the lower limit problem insufficiently explored with regard to the accuracy of the estimation variance, presenting the asymptotic Cramér–Rao bound (CRB) for Gaussian processes and additive Gaussian noise. This novel result is obtained by using a frequency-domain perspective of the problem as well as an unusual parametrization of an AR model. The Wiener filter rule appears as the distinctive building element in the Fisher information matrix. The theoretical analysis is validated numerically, showing that the proposed CRB is attained by competitive ad hoc estimation methods under a variety of Gaussian color noise and realistic scenarios.

Index Terms—Autoregressive analysis, noise compensation, additive Gaussian color noise, Cramér–Rao bound.

I. INTRODUCTION

NOISE compensation in autoregressive (AR) estimation [1]–[3] is a problem that has aroused great interest in several applied fields, such as speech, biomedical signal processing, control systems, and time series forecasting. This problem was first addressed by Walker [4], who evaluated the asymptotic variance for the estimate of a first-order AR model. Two decades later [5]–[7] the accuracy of the estimation was analytically studied to a deeper extent: Friedlander and Porat [5] derived the asymptotic Cramér–Rao bound for an AR model corrupted by additive white Gaussian noise (AWGN); in [6], [7] the asymptotic characteristics of the biased least squares AR estimate in the presence of AWGN were explored. In the last decade several methods [8]–[11] (to name a few) for compensating the effects of noise in the AR estimate have been proposed. A common point shared by all methods regards the strategy used in the evaluation: the estimation accuracy is measured in comparative terms, versus previous techniques. It is not only that no noise-compensated AR estimation method has yet been accepted as a gold standard (as every new method seems to prove better than the previous ones), but that, despite the existence of seminal works [5]–[7], an objective bound, such as the Cramér–Rao bound, has not been sufficiently explored or understood.

In this paper the asymptotic CRB for noise-compensated AR estimation is comprehensively discussed. The second order statistics of the noise are assumed to be known. The methodology that we used to deduce the CRB differs from previous related works essentially in three points: 1) the analysis is conducted in the frequency domain (but the final result is given in the time domain), 2) the parametrization required for this analysis is unusual in AR analysis, and 3) the corrupting noise is not restricted to the AWGN type. The organization of the paper is as follows: Sec. II serves to present the problem statement and basic notation, Sec. III contains the derivation of the Fisher information matrix, Sec. IV presents the asymptotic CRB as well as a comparative analysis between noiseless and noise-compensated AR estimation, a discussion on the implication of this result is included in Sec. V, the numerical validation of the theoretical results is brought into Sec. VI, finally the conclusions close the paper.

II. PROBLEM STATEMENT

Let the signal \( y[n] \) be a realization of an autoregressive model corrupted with additive stationary noise, that is,

\[
y[n] = x[n] + v[n]
\]

where \( x[n] \) is a pure (uncorrupted) signal, and \( v[n] \) is the realization of the noise, a Gaussian stationary random process statistically uncorrelated with the AR process \( x[n] \). The process can be described by a difference equation of the form

\[
x[n] = \sum_{\ell=1}^{P} a_\ell x[n - \ell] + \sigma e[n]
\]

where \( a_\ell \) are the AR coefficients, \( P \) is the order, \( \sigma > 0 \) is the amplitude level, and \( e[n] \) is the excitation, assumed here to be Gaussian random with zero mean and variance equal to one. A less frequent expression for an AR process follows

\[
\sum_{\ell=0}^{P} a_\ell x[n - \ell] = e[n].
\]

Models (2) and (3) are equivalent, as can be seen from the following choice of coefficients in the latter: \( a_0 = 1/\sigma \), and \( a_\ell = -a_\ell/\sigma \) for \( \ell > 0 \). The expression in (3) has the advantage of being more symmetric in form than (2), hence we will use (3) as the basis of the ensuing discussion. The transfer function of the prediction residue filter is thus

\[
A(e^{j\omega}) = \sum_{\ell=0}^{P} a_\ell e^{-j\omega \ell}
\]

where \( \omega \) denotes frequency.

The problem of noise-compensated AR analysis (NCAR) deals with the estimation of the AR parameters \( \alpha = [\alpha_0, \alpha_1, \cdots, \alpha_P] \) from a finite segment \( y \) of signal \( y[n] \). The estimate must correspond to a stable solution, that is, the roots

---

L. Weruaga and O. M. Melko are with Khalifa University of Science, Technology & Research, P.O. 573 Sharjah, UAE. Tel. +971 6 5043552, Fax: +971 6 5611789, E-mail: weruaga@ieee.org, mike.melko@gmail.com

1In contrast, both the asymptotic and the exact Cramér–Rao bound (CRB) for a clean (noiseless) AR model are known [14].
of the polynomial $A(z) = 0$ must lie within the unit circle $\{z : |z| \leq 1\}$. Lastly, the second order statistics of the noise are assumed to be known. The present paper explores the accuracy bound for any unbiased estimator to this problem. Estimation accuracy is commonly assessed in terms of the square error (SE) between estimate $\hat{\alpha}$ and true values $\alpha$. The Cramér–Rao bound (CRB) provides the lower bound to the expectation of that measure [12]

$$E \{ (\alpha - \hat{\alpha})^T (\alpha - \hat{\alpha}) \} \geq CRB(\alpha|\mathcal{V}) \triangleq J(\alpha|\mathcal{V})^{-1}$$

(5)

where $E\{\}$ denotes expectation, $\mathcal{V}$ denotes the second-order statistics of the noise, $CRB(\alpha|\mathcal{V})$ is the Cramér–Rao bound, and $J(\alpha|\mathcal{V})$ corresponds to the Fisher information matrix (FIM).

### III. ASYMPTOTIC FISHER INFORMATION

We assume here that the information of the signal is found in the frequency domain, that is, the sample set $y$ is composed of $N$ spectral samples, $y = \{Y_0, \cdots, Y_{N-1}\}$, equidistant in frequency. From a frequency-domain perspective, a priori knowledge of the second order statistics of the noise implies that the power spectral density (PSD) of the noise, $\mathcal{V}(e^{j\omega})$, is known in advance. Since the AR process and the noise are uncorrelated, the following equality holds

$$\mathcal{Y}(e^{j\omega}) = \mathcal{X}(e^{j\omega}) + \mathcal{V}(e^{j\omega})$$

(6)

where $\mathcal{Y}(e^{j\omega})$ and $\mathcal{X}(e^{j\omega})$ correspond to the PSD of the noisy and noiseless signals respectively. From (3), the power spectral density of the AR process results in

$$\mathcal{X}(e^{j\omega}) = |A(e^{j\omega})|^2 \cdot$$

(7)

Throughout the paper, we use the following notation $Z_k = Z(w_N^k)$, where $Z(e^{j\omega})$ corresponds to a discrete-time Fourier transform, and $w_N^k = e^{2\pi jk/N}$.

The entry in the $\ell$-th row and $m$-th column of the Fisher information matrix (FIM) is given by the following expression

$$\tilde{J}_{\ell m}(\alpha|\mathcal{V}) \triangleq E \left\{ \frac{\partial \ln f(y|\mathcal{V})}{\partial \alpha_{\ell}} \frac{\partial \ln f(y|\mathcal{V})}{\partial \alpha_{m}} \right\}$$

(8)

where $f(y)$ refers to the probability density that the true model $\alpha$ produced the sample set $y$. Based on its definition, the FIM (8) is clearly symmetric and can be assumed without loss of generality to be positive definite. Given the statistical independence between the samples $Y_k$, as claimed in [13], the likelihood of the set $y$ becomes $f(y|\mathcal{V}) = \prod_{k=0}^{N-1} f(Y_k|\mathcal{V}_k)$. As is also claimed in [13], we may take the sample $Y_k$ to be an independent complex (bivariate) Gaussian variable of zero mean and variance equal to $\mathcal{X}_k + \mathcal{V}_k$. The probability density function of each sample $Y_k$ can be thus written as

$$f(Y_k|\mathcal{V}_k) = \frac{1}{\pi(\mathcal{X}_k + \mathcal{V}_k)} \exp \left( - \frac{|Y_k|^2}{\mathcal{X}_k + \mathcal{V}_k} \right).$$

(9)

Note that, as equations (7) and (4) show, $\mathcal{X}_k$ holds the parametric dependence with the AR parameters $\alpha$. The log-likelihood of the set $y$, present in (8), is thus

$$\ln f(y|\mathcal{V}) = \sum_{k=0}^{N-1} \ln f(Y_k|\mathcal{V}_k).$$

(10)

The gradient of the log-likelihood for the $k$-th sample is equal to the following expression

$$\frac{\partial \ln f(Y_k|\mathcal{V}_k)}{\partial \alpha_{\ell}} = \left( \frac{|Y_k|^2}{(\mathcal{X}_k + \mathcal{V}_k)^2} - \frac{1}{\mathcal{X}_k + \mathcal{V}_k} \right) \frac{\partial \mathcal{X}_k}{\partial \alpha_{\ell}}$$

$$= W_k^2 (\mathcal{X}_k + \mathcal{V}_k - |Y_k|^2) \rho_{k\ell}$$

(11)

where $W_k$ corresponds to the Wiener filter

$$W_k = \frac{\mathcal{X}_k}{\mathcal{X}_k + \mathcal{V}_k}$$

(12)

and

$$\rho_{k\ell} = A_k w_N^{k\ell} + A_k w_N^{-k\ell}.$$  

(13)

The objective of this section is to deduce the asymptotic FIM, which is in [6] defined as follows

$$J_{\ell m}(\alpha|\mathcal{V}) \triangleq N \lim_{K \to \infty} I_{\ell m}(K)$$

(14)

where

$$I_{\ell m}(N) = \frac{1}{N} \sum_{k=0}^{N-1} E \{ \frac{\partial \ln f(Y_k|\mathcal{V}_k)}{\partial \alpha_{\ell}} \frac{\partial \ln f(Y_k|\mathcal{V}_k)}{\partial \alpha_{m}} \}.$$  

(15)

Based on the expansion (10), and given that the expected value of the information gradient (11) vanishes, that is, $E \{ \partial \ln f_{\alpha_k}(Y_k|\mathcal{V}_k)/\partial \alpha_{\ell} \} = 0$, the previous term (15) can be expanded and simplified to

$$I_{\ell m}(N) = \frac{1}{N} \sum_{k=0}^{N-1} E \{ \frac{\partial \ln f(Y_k|\mathcal{V}_k)}{\partial \alpha_{\ell}} \frac{\partial \ln f(Y_k|\mathcal{V}_k)}{\partial \alpha_{m}} \}.$$  

(16)

By substituting the information gradient (11) into (16), this term becomes

$$I_{\ell m}(N) = \frac{1}{N} \sum_{k=0}^{N-1} W_k^4 \left( \mathcal{V}_k - |Y_k|^2 \right) \rho_{k\ell} \rho_{km}.$$  

(17)

The complex Gaussian random variable $Y_k$ fulfills

$$E \left\{ |Y_k|^4 \right\} = 2E \left\{ |Y_k|^2 \right\}^2 = 2Y_k^2$$

(18)

which allows us to simplify (17) to

$$I_{\ell m}(N) = \frac{1}{N} \sum_{k=0}^{N-1} W_k^4 \left( \mathcal{V}_k - |Y_k|^2 \right) \rho_{k\ell} \rho_{km}.$$  

(19)

On the other hand, the factor $\rho_{k\ell} \rho_{km}$ can be expanded to

$$\rho_{k\ell} \rho_{km} = |A_k|^2 \left( w_N^{k(\ell-m)} + w_N^{-k(\ell-m)} \right)$$

$$+ (A_k)^2 w_N^{k(\ell+m)} + (A_k)^2 w_N^{-k(\ell+m)}.$$  

(20)

In defining the Fisher information matrix, we have not imposed any assumption on the autoregressive process. Therefore, it can be any complex process such that $\alpha_i \in \mathbb{C}$, and as a consequence all $N$ spectral samples are meaningful. We are however primarily interested in real-valued autoregressive processes $\alpha_i \in \mathbb{R}$, in which case $Y_k = Y_N^{* -k}$. The previous condition implies that in every complex conjugate pair $\{Y_k, Y_N^{* -k}\}$ one component does not provide information, its
contribution to the information gradient thus vanishing. Based on this reasoning, it is simple to see that the Fisher information for a real-valued autoregressive process is actually half the value defined in the general (complex-valued) case (16). In addition to that, the case of a real-valued process yields the following simplification: when replacing (20) into (19), the first two terms in (20) yield the same summations, while the last two terms result in equal summations as well. The resulting factor of 2 cancels with the previous factor 1/2 from a real-valued AR model. We can thus finally write the term (15) as

\[
I_{tm}(N) = \frac{1}{N} \sum_{k=0}^{N-1} \frac{W_k^2}{|A_k|^2} w_N^{k(\ell-m)} + \frac{W_k^2}{A_k} w_N^{-k(\ell+m)}. \tag{21}
\]

We are now in a position to present a major result of the paper.

**Proposition 3.1:** Let \( h[n] \) be the impulse response of the AR model (3), \( w[n] \) the impulse response of the Wiener filter, and \( h_w[n] \) the Wiener-filtered impulse response of the AR model, that is, \( h_w[n] = h[n] * w[n] \), where * denotes convolution. The **asymptotic** Fisher information matrix (14) for Gaussian noise-compensated AR analysis is then

\[
J_{tm}(\alpha|\mathcal{V}) = N h_w[n] * h_w[-n] \big|_{n=\ell-m} + N h_w[n] * h_w[n] \big|_{n=-\ell-m}. \tag{22}
\]

**Proof:** It is immediate that when \( N \) goes to infinity, the expression in (21) converges to

\[
\lim_{N \to \infty} I_{tm}(N) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{W(\epsilon^{j\omega})^2}{|A(\epsilon^{j\omega})|^2} \epsilon^{j\omega(\ell-m)} d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{W(\epsilon^{j\omega})^2}{A(\epsilon^{j\omega})^2} \epsilon^{-j\omega(\ell+m)} d\omega. \tag{23}
\]

where \( W(\epsilon^{j\omega}) \) is the Wiener filter

\[
W(\epsilon^{j\omega}) = \frac{\mathcal{X}(\epsilon^{j\omega})}{\mathcal{X}(\epsilon^{j\omega}) + \mathcal{V}(\epsilon^{j\omega})}. \tag{24}
\]

Given that the time-domain counterpart of the Wiener filter fulfills \( w[n] = w[-n] \), the time-domain counterpart of (23), scaled by \( N \), as observed in (14), yields immediately (22). This concludes the proof.

\[ \text{In order to stress the simplicity of the result in Proposition 3.1, we bring the asymptotic FIM in matrix form in (25). The asymptotic FIM (25) is composed of a positive definite symmetric Toeplitz matrix plus a Hankel matrix. Toeplitz-plus-Hankel matrices occur in several engineering and statistical applications, such as in the design of linear-phase digital filters [15]. The Hankel part appears here as result of the cross-correlation (20) between the forward \( A_k \) and backward \( A_{-k} = A_k^* \) prediction filters. The underlying (anti-causal) Wiener filtering in (22) makes both predictors share information from the signal \( y[n] \) (and hence from the process \( x[n] \)). The Hankel matrix cannot be proven to be positive semidefinite. Note, however, that the whole FIM is, according to its definition, positive definite.}

A required exercise arising from the previous result regards the structure of the asymptotic Fisher information matrix for the particular case of noise absence, and how this result relates to the existing well-known lower bound there. The next proposition addresses this issue.

**Proposition 3.2:** In the absence of noise, the asymptotic FIM (22) reduces to the asymptotic FIM of a (noiseless) Gaussian real-valued AR model.

**Proof:** Given that \( \mathcal{V}(\epsilon^{j\omega}) = 0 \), then \( h_w[n] = h[n] \), and

\[
J_{tm}(\alpha|0) = N h[n] * h[-n] \big|_{n=\ell-m} + N h[n] * h[n] \big|_{n=-\ell-m}. \tag{26}
\]

The first term in (26) corresponds to the autocorrelation of the impulse response \( h[n] \), which we denote here as \( r_{\ell} \), that is, \( r_{\ell} = h[n] * h[-n] \). Since the AR model is causal, \( h[n] = 0 \) for \( n < 0 \), the auto-correlation in the second term of (26) reduces to a single non-zero value equal to \( N\sigma_0^{-2} \) at \( \ell = m = 0 \). Therefore, the asymptotic FIM of the AR model becomes

\[
J(\alpha|0) = N \begin{bmatrix} r_0 + \sigma_0^{-2} & r_1 & \cdots & r_p \\
1 & r_0 & \cdots & r_{p-1} \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & r_0 \end{bmatrix}. \tag{27}
\]

A familiar expression for the asymptotic FIM of an AR model, related to the parameter set \( \theta = [\sigma^2, a_1, \ldots, a_p] \), is [14]

\[
J(\theta) = N \begin{bmatrix} \frac{1}{\sigma^2} & 0 & \cdots & 0 \\
0 & r_0 & \cdots & r_{p-1} \\
\vdots & \vdots & \ddots & \vdots \\
0 & r_{p-1} & \cdots & r_0 \end{bmatrix}. \tag{28}
\]

Formally, while the \( k \)-th sample real part \( Y_{k,k}^R \) and the \( N-k \)-th sample imaginary part \( Y_{k,k}^I \) (for \( k \neq 0 \)) are Gaussian of zero mean and variance \((\epsilon_k + \epsilon_k^*)/2\), the PDF of \( Y_{k,k}^R \) and \( Y_{k,k}^I \) can be considered fully deterministic, that is, \( p_{k,k}^R(x) = \delta(x + Y_{k,k}^R) \) and \( p_{N-k,k}^I(x) = \delta(x - Y_{k,k}^R) \), respectively, where \( \delta(x) \) is the Dirac delta; on the other hand, \( p_0^R(x) = \delta(x) \), and \( p_0^{N-k,k}(x) = \delta(x) \) for \( N \) even; these PDFs are non-parametric and thus vanish when computing the gradient of the log-likelihood.

\[ ^2 \text{Yule–Walker equations, and lattice methods, such as Burg’s, aim either implicitly or explicitly to minimize both the forward and backward errors.} \]
It is tedious but straightforward to verify that both matrices fulfill the following required relation

\[
J(\theta) = \left[ \frac{\partial \alpha}{\partial \theta} \right] J(\alpha) \left[ \frac{\partial \alpha}{\partial \theta} \right]^T
\]

where the Jacobian matrix results in

\[
\left[ \frac{\partial \alpha}{\partial \theta} \right] = \begin{bmatrix}
\frac{1}{(2\sigma^2)} & -a/(2\sigma^2) \\
0_{P \times 1} & I_{P \times P}
\end{bmatrix}.
\]

Here \(a = [a_1 \cdots a_P] \), \(0_{P \times 1} \) is the P-zero column vector, and \(I_{P \times P} \) is the \(P \times P \) identity matrix. This completes the proof.

The rule (29) illustrates the way to evaluate the lower bound for the classical parametrization \( \theta \) (or any other parametrization provided that the corresponding Jacobian matrix is obtained). It is worth pointing out that, unlike the noiseless case (28), the first column and first row of the resulting matrix \( J(\theta|V) \) in general does not contain zeros. This nonzero information covariance between the power level \( \sigma^2 \) and the \( P \) AR coefficients \( a_p \) has important implications in any estimation method, namely, all \( P+1 \) parameters must (should) be obtained simultaneously.

\section{IV. Asymptotic Cramér–Rao Bound}

Although it is common practice in noiseless AR analysis to appeal to a theoretical lower bound, no such bound is being used for the noise-compensated case. In the most relevant works published in the last decade [8]–[11] (among others, not included here for the sake of brevity), the performance evaluation is carried out with a comparison analysis of previous methods. Although uncommon, in other works [16] the use of the noiseless bound (28) is suggested as a valid lower bound to this problem. The noiseless bound is a valid lower bound, but it is not optimal. It therefore consists of an optimistic estimate in so far as the actual lower bound will generally be significantly higher.

We introduce the following notation convention required here: we write \( C > 0 \) if the matrix \( C \) is positive definite, respectively \( C \geq 0 \) if it is positive semidefinite; if \( A \) and \( B \) are two matrices such that \( A - B \geq 0 \), we write \( A \geq B \).

\textit{Proposition 4.1:} The noise-compensated AR estimation is a more difficult task than the (noiseless) AR estimation, i.e.,

\[
\text{CRB}(\alpha|V) \geq \text{CRB}(\alpha|0).
\]

Intuitive as this fact may be, one could still be inclined to assert that, given that the power spectral density \( V \) of the noise is available, this knowledge would make the problem resemble the classical noiseless AR estimation or one of similar complexity. In order to clarify and settle this issue, Proposition 4.1 is proven in the following.

\textit{Proof:} The asymptotic FIM matrix of the noise case and that of the clean case relate as follows

\[
J(\alpha|0) = J(\alpha|V) + S
\]

where the matrix \( S \) is given by

\[
S_{\ell m} = N \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ 1 - W(e^{j\omega})^2 \right] e^{j\omega(\ell - m)} d\omega
\]

\[
+ N \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ 1 - W(e^{j\omega})^2 \right] c^{-j\omega(\ell + m)} d\omega.
\]

Since the matrix \( S \) has an equivalent structure to the positive definite matrix (23), and given that \( 0 \leq W(e^{j\omega})^2 \leq 1 \), \( S \) is necessarily symmetric and positive semidefinite. Therefore

\[
J(\alpha|V)^{-1} \geq J(\alpha|0)^{-1}
\]

which is equivalent to (31). This completes the proof.

\section{V. Further Analysis}

The parametrization \( \alpha \) in (3) has a symmetric form, which in turn yields a compact result (22). A comparable analysis based on the classical formulation \( \theta \) in (2) and for AWGN noise was explored in [5]. The asymptotic FIM there is stated with the Whittle formula [17] for the \( \theta \) parametrization as

\[
J_{\ell m}(\theta|V) = N \frac{1}{4\pi} \int \frac{d\gamma(z)}{d\theta} \frac{[\gamma(z)]^{-1} z}{z} \frac{d\gamma(z)}{d\theta_m} \gamma(z)^{-1} d\gamma(z) d\omega.
\]

After simplifications, the FIM submatrix that corresponds to the autoregressive coefficients \( a_\ell \) results in

\[
J_{\ell m}(\theta|V) = N \frac{1}{2\pi} \int \frac{\sigma^4 z^{(\ell - m)}}{(c(z)c(z^{-1}))^2 a(z)a(z^{-1})} dz z
\]

\[
+ N \frac{1}{2\pi} \int \frac{\sigma^4 z^{(\ell - m)}}{(c(z)c(z^{-1}))^2 a(z)^2} dz z
\]

for \( \ell, m = 1, \cdots, P \), where \( a(z) = \sigma A(z) \), the polynomial \( c(z) = c(z^{-1}) = \sigma^2 + \sigma_a^2 A(z^2) \), and \( \sigma_a^2 \) is the AWGN power. Given that the Wiener filter in the AWGN case is \( W(z) = 1/(1 + \sigma_a^2 A(z^2) \) and \( z = e^{j\omega} \), hence \( dz = jz\omega dz \omega \), we can simplify the term (36) as

\[
J_{\ell m}(\theta|V) = N \frac{1}{\sigma_a^2} \int_{-\pi}^{\pi} \left[ 1 - W(e^{j\omega})^2 \right] e^{j\omega(\ell - m)} d\omega
\]

\[
+ N \frac{1}{\sigma_a^2} \int_{-\pi}^{\pi} \left[ 1 - W(e^{j\omega})^2 \right] c^{-j\omega(\ell + m)} d\omega
\]

which resembles one of our results (23). From this result it is clear that \( J_{\ell m}(\alpha|V) = \sigma_a^2 J_{\ell m}(\theta|V) \) for \( \ell, m = 1, \cdots, P \). On the other hand, we can simplify accordingly the FIM cross terms between \( a_\ell \) and the power \( \sigma^2 \) given in [5] as

\[
J_{\ell 0}(\theta|V) = N \frac{1}{4\pi} \int \frac{d\gamma(z)}{d\theta} \frac{[\gamma(z)]^{-1} z}{z} \frac{[\gamma(z)]^{-1} z}{z} d\gamma(z) dz
\]

\[
+ N \frac{1}{\sigma^2} \int_{-\pi}^{\pi} \frac{W(e^{j\omega})^2}{A(e^{j\omega})} c^{-j\omega(\ell + m)} d\omega
\]

\[
= N \int_{-\pi}^{\pi} \left[ 1 - W(e^{j\omega})^2 \right] c^{-j\omega(\ell + m)} d\omega.
\]
for \( \ell = 1, \ldots, P \). Finally, we can simplify the FIM term related to the power \( \sigma^2 \) as
\[
J_{00}(\theta|\mathcal{V}) = \frac{N}{4\pi} \int \left( \frac{\partial \mathcal{Y}(z)}{\partial \sigma^2} \right)^2 \frac{dz}{z} = \frac{N}{\sigma^4} \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{j\omega})^2 d\omega. \tag{39}
\]

Even though we have been able to obtain simplified results (37)-(39) from the previous work [5], the authors there did not acknowledge that possibility and degree of simplicity. Instead, the expansion of the FIM terms was achieved with a long non-trivial analysis that resulted in large and not very insightful expressions scattered over several pages. Equally important, the Wiener filter, a key element for building the FIM (26), is never mentioned in that work. The FIM terms for the usual AR parametrization \( \theta \), summarized in (37)-(39), do not result in such a simple and compact expression as the one derived in the present paper (25) for the parametrization \( \alpha \).

Finally, the work [11], published recently by one of the authors, claims to have deduced an approximate asymptotic FIM \( F(\theta|\mathcal{V}) \) for this problem with the simple expression related to the Wiener filter and the noiseless FIM
\[
F(\theta|\mathcal{V}) = \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} W(e^{j\omega})^2 d\omega \right) J(\theta|0). \tag{40}
\]

This result is actually a by-product in that work, which proposes a novel NCAR estimator founded on the maximum likelihood (ML) criterion. This approach leads to a frequency-selective scenario, in which the spectral samples are selectively used based on the local SNR. From this “missing sample” perspective, the mentioned approximate FIM (40) was intuitively deduced. In order for (40) to yield a valid lower bound in this problem, it must fulfill
\[
F(\theta|\mathcal{V}) \geq J(\theta|\mathcal{V}). \tag{41}
\]

Unfortunately, this condition (41) does not always hold, and therefore the tentative approximate bound resulting from (40) is not a valid lower bound to the NCAR problem.

VI. NUMERICAL VALIDATION

While the proposed asymptotic FIM reduces to a well-known expression under noiseless conditions as Proposition 3.2 analytically shows, the main findings of the paper will be validated empirically over different noise scenarios.

The first experiment corresponds to the estimation of the fourth-order AR process of coefficients \( a_1 = 1.65, a_2 = -1.6, a_3 = 1.05, \) and \( a_4 = -0.38 \). Two types of noise have been considered, namely, a noise in band with the AR process, and white noise. The in-band noise follows an AR model as well, whose poles are those of the AR process pulled to the center of the unit circle (by multiplying each pole by 0.7). In each noise scenario, two levels of signal-to-noise ratio (SNR) were considered, namely, 5 dB and 0 dB. Fig. 1 illustrates the simulation scenario considered.

The aim of this experiment is to verify that the asymptotic CRB is attained by state-of-art estimation methods such as the errors-in-variable (EIV) method [9]\(^5\) and the maximum-likelihood (ML) estimate [11]. The accuracy of the estimate corresponds to the square error (SE) of the AR coefficients of the classical parametrization (2),
\[
\text{SE} \triangleq ||\hat{a} - \hat{a}||^2 = \text{tr}\left\{ (\hat{a} - \hat{a})^T (\hat{a} - \hat{a}) \right\}. \tag{42}
\]

where \( ||e||^2 \) denotes norm-2 of vector e, \( \text{tr}\{A\} \) is the trace of matrix A, and \( \hat{a} \) and \( \hat{a} \) are the true AR coefficients and estimate respectively (as row vectors). For each simulation scenario the asymptotic bound \( \text{CRB}(\theta|\mathcal{V}) \), as disclosed in this paper, was computed. The trace of its lower \( P \times P \) submatrix (i.e., disregarding the power level \( \sigma^2 \)) represents the lower bound for the expected value of the square error (42). The empirical accuracy was obtained by averaging the square error from 1000 Monte Carlo simulations for both AR model and noise. The results are shown in Fig. 2.

In all situations, as the number of samples \( N \) increases, the mean square error delivered by the ML estimator, marked with the “+” sign, nearly attains the proposed CRB. The accuracy of the EIV method, marked with “×”, nearly attains the CRB in the first noise type as well; this method falters however with the white noise especially at low SNR levels. The performance of the EIV method assuming AWGN for actual color noise is shown in Fig. 2.a) with “□”: this discrepancy between assumed noise PSD and actual one introduces a bias in this estimator, such that the proposed CRB is not applicable therein; moreover, the empirical estimation accuracy does not improve with the segment length \( N \). This behavior, observed in [11] as well, stresses the importance of having a priori knowledge of the noise PSD. On the other hand, comparison between the proposed CRB with the noiseless CRB reveals the different corrupting effect resulting from each noise type: the white noise yields a larger loss of accuracy essentially because one of the resonances is nearly hidden, while the in-band noise happens to disturb the AR process in a balanced way.

\(^5\)Originally, the EIV method aims to estimate all AR parameters plus the noise power. In order to match the complexity of the problem defined by the CRB, the EIV variant used here assumes that the ratio \( \sigma^2/E\{[\hat{w}[n]]^2\} \) is known, thus resulting in a problem with \( P + 1 \) degrees of freedom. Note however that in general this ratio cannot be obtained in the practice.
While the previous experiment validates the analytical results on a simple scenario, the following experiment brings a more realistic case into view, namely, the estimation of a 10th-order AR model from a noise-corrupted speech signal. The noiseless digital signal, resampled at 8 kHz, was chopped into segments of 160 samples (equivalent to 20 ms), where the 10-th order AR model (with Burg’s method) was computed for each segment. Each segment becomes thus an independent estimation problem, where a Monte Carlo statistical analysis was carried out. Regarding the way the noisy signal was generated, two different sub-experiments were considered:

1) The noiseless component corresponds to a realization of the true AR process with a white Gaussian excitation; synthetic noise is added to the segment. This realistic case serves to validate $E\{||\hat{a} - a||^2\} \geq \text{tr}\{\text{CRB}(a|\mathcal{V})\}$. 

2) The clean signal corresponds to actual speech segment; synthetic noise is added to the segment. Since the clean part is fixed, this case serves to validate $E\{||\hat{a} - a||^2\} \geq \text{tr}\{\text{CRB}(a|\mathcal{V}) - \text{CRB}(a|0)\}$.

In all cases, the synthetic noise in the segment is in-band.

---

**Fig. 2.** Cramér-Rao bound in noise-compensated autoregressive estimation (solid), and noiseless CRB (dashed). Estimation accuracy (averaged over 1000 realizations) of the EIV estimator with AWGN assumptions [9] , EIV with actual noise PSD knowledge ×, maximum-likelihood estimator [11] + and that of the spectral subtraction method • [3].

<table>
<thead>
<tr>
<th>SNR = 5 dB</th>
<th>SNR = 5 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNR = 0 dB</th>
<th>SNR = 0 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Fig. 3.** Empirical distribution of $E\{||a - \hat{a}||^2\}/\text{tr}\{\text{CRB}(a|\mathcal{V})\}$ (white) and $E\{||a - \hat{a}||^2\}/\text{tr}\{\text{CRB}(a|0)\}$ (gray).

<table>
<thead>
<tr>
<th>dB</th>
<th>dB</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Fig. 4.** Empirical distribution of $E\{||a - \hat{a}||^2\}/\text{tr}\{\text{CRB}(a|\mathcal{V}) - \text{CRB}(a|0)\}$ for speech (white) and Gaussian synthetic segment (gray).
with the AR model, according to the procedure detailed in the first experiment. This choice assures that the estimation of the AR coefficients is well conditioned. The SNR was set to 5 dB. The estimation variance delivered by the ML method was averaged from 1000 Monte Carlo simulations. The number of total analyzed speech segments was 750. Fig. 3 and Fig. 4 contain the results of these experiments. Each figure illustrates the statistical distribution of the normalized estimation variance in logarithmic (dB) scale for both sub-scenarios 1) and 2) respectively.

The results in Fig. 3 indicate that the proposed CRB is attained, as the normalized estimation variance for all segments falls in the positive dB axis. On the other hand, the results corresponding to the normalization by the noiseless CRB (in gray) confirm this lower bound as “optimistic”, which means that this bound should not be used as a performance indicator in this problem. Regarding the results of the second sub-scenario, the normalized variance shown in Fig. 4 (in white) falls mostly on the positive dB axis, while less than 2% thereof are on the negative side, close to the 0 mark. This anomaly is probably due to the fact that the speech spectrum is better described with super-Gaussians [19]. The previous experiment was repeated this time with synthetic speech segments built with a Gaussian excitation: the distribution of the variance (in gray) falls entirely to the right of the 0 mark. Nonetheless, the results on Gaussian-synthetic or on real speech segments do not differ significantly, which suggests that the Gaussian model is fairly representative of real speech segments.

VII. CONCLUSIONS

A frequency-domain perspective along with an unusual parametrization of the autoregressive (AR) model brings notable advantages for deducing the lower accuracy bound of noise-compensated AR estimation. The deduced asymptotic Fisher information matrix (FIM) results in a compact and simple formula (related to the Wiener-filtered impulse response of the AR filter) that is easy to use in practice. The resulting Cramér–Rao bound (CRB) is shown empirically to be an optimal lower bound. The proposed asymptotic CRB represents a significant contribution for assessing the performance of any existing and future ad hoc estimator.

REFERENCES


If the pole’s local SNR is very low, the estimation problem becomes ill posed, yielding statistical results difficult to interpret.