YULE-WALKER EQUATIONS APPLIED TO HESSIANS OF THE CHARACTERISTIC FUNCTION FOR IMPROVED AR ESTIMATION

Arie Yeredor

School of Electrical Engineering
Tel-Aviv University, Tel-Aviv, 69978, ISRAEL
e-mail: arie@eng.tau.ac.il

ABSTRACT

Estimation of the autoregressive (AR) parameters of an AR process often involves applying Yule-Walker (YW) equations to the estimated correlations. When the process is Gaussian, the resulting estimate is asymptotically optimal, coinciding with the Maximum-Likelihood (ML) estimate. However, for non-Gaussian processes, applying the YW equations to the estimated correlations may be significantly sub-optimal, whereas computation of the exact ML estimate may be prohibitively cumbersome. In this paper we show how the YW equations may be applied to an alternative statistic, namely to off-origin Hessians of the second characteristic function. Although still not optimal, we show in simulation that the resulting estimate can significantly outperform the classical correlation-based estimate, as well as a cumulants-based estimate.

Index Terms— autoregressive, Yule-Walker, characteristic function, Hessian, charrelation matrix.

1. INTRODUCTION

Estimation of the autoregressive (AR) parameters of an observed AR process is a fundamental problem in time-series analysis, with diverse applications, such as spectral estimation, blind channel identification, signal detection, adaptive filtering, speech analysis, speech coding and many more. A classical tools for this estimation problem are the Yule-Walker equations (e.g., [6]), which are commonly applied to the estimated autocorrelation of the observed process. More specifically, let the AR process $x[n]$ (of known order p) be given by

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + w[n] \quad \forall n \tag{1}$$

where $\boldsymbol{a} = [a_1\ a_2\ \cdots\ a_p]^T$ are the unknown AR parameters, to be estimated from N observations $x[0], \ldots, x[N-1]$. $w[n]$ denotes some independent, identically distributed (iid) zero-mean sequence with variance σ_w^2, often termed the "driving noise". It is assumed that the polynomial $A(z) \stackrel{\triangle}{=} 1 + a_1 z^{-1} + \cdots + a_p z^{-p}$ has all its roots inside the unit-circle, hence $x[n]$ is stationary. Using some straightforward manipulations, mainly accounting for statistical independence (hence no correlation) between $w[n]$ and past values $x[n-\ell]$ (for $\ell > 0$), the "full" $((p+1)\times(p+1))$ form of the Yule-Walker equations can be easily obtained as

$$\underbrace{\begin{bmatrix} R[0] & R[-1] & \cdots & R[-p] \\ R[1] & R[0] & \cdots & R[-p+1] \\ \vdots & \ddots & \ddots & \vdots \\ R[p] & R[p-1] & \cdots & R[0] \end{bmatrix}}_{\boldsymbol{R}_{p+1}} \underbrace{\begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}}_{\bar{\boldsymbol{a}}} = \underbrace{\begin{bmatrix} \sigma_w^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\sigma_w^2 \boldsymbol{e}_1}, \tag{2}$$

where $R[\ell] \stackrel{\triangle}{=} E[x[n]x[n-\ell]]$ denotes the autocorrelation sequence of $x[n]$, $\bar{\boldsymbol{a}} \stackrel{\triangle}{=} [1\ \boldsymbol{a}^T]^T$, and $\boldsymbol{e}_1$ is the first column of the $(p+1)\times(p+1)$ identity matrix. When the correlation values $R[0], \ldots, R[p]$ are known, the coefficients $\boldsymbol{a}$ can be extracted from this set (along with σ_w^2), e.g., by solving the system with the right-hand side (RHS) replaced by $\boldsymbol{e}_1$, followed by scaling the solution (along with the RHS) such that its leading element becomes 1. An alternative, "reduced" $(p\times p)$ form can be obtained by maintaining only the last p equations, as in

$$\boldsymbol{R}_p \boldsymbol{a} = -\boldsymbol{r}_p, \tag{3}$$

where $\boldsymbol{R}_p$ is the lower-right $p\times p$ block of $\boldsymbol{R}_{p+1}$, and $\boldsymbol{r}_p \stackrel{\triangle}{=} [R[1]\ R[2]\ \cdots\ R[p]]^T$ is the lower-left $p\times 1$ vector of the same matrix.

When the correlation values are unknown, $\boldsymbol{R}_{p+1}$ may usually be consistently estimated, e.g., using

$$\hat{\boldsymbol{R}}_{p+1} = \frac{1}{N-p}\sum_{n=p}^{N-1} \boldsymbol{x}_n \boldsymbol{x}_n^T, \tag{4}$$

where $\boldsymbol{x}_n \stackrel{\triangle}{=} [x[n]\ x[n-1]\ \cdots\ x[n-p]]^T$.

When the driving noise $w[n]$ is Gaussian, the estimate of $\boldsymbol{a}$ resulting from solving the YW equations with these estimated correlations coincides (asymptotically in N, when end-effects are negligible) with the Maximum-Likelihood (ML) estimate. Consequently, this estimate is asymptotically unbiased and optimal (among all unbiased estimates) in the sense of mean square estimation error, attaining (asymptotically) the associated Cramér-Rao lower bound (CRLB).

However, when the driving noise is non-Gaussian, the resulting estimate is no longer ML, and may therefore be far from optimal (even asymptotically). The derivation and computation of the ML estimate may then become computationally cumbersome in some cases (see, e.g., [7] for the case of a Gaussian-Mixture), and intractable in other cases. It is therefore of interest, in such cases, to look for other, more simple estimates, which, although not optimal, may still offer significant improvement over the correlations-based estimate. In this paper we show how such an estimate can be based on a similar form of YW equations, applied to off-origin Hessians of the log-characteristic function (LCF), rather than to correlations. We demonstrate (in simulation) the potential for significant improvement in the estimation accuracy for non-Gaussian processes.

2. OFF-ORIGIN HESSIANS OF THE LCF

Off-origin Hessians of the LCF are a relatively new emerging tool, offering a "hybrid" statistic, conceptually reconciling second-order statistics with (classical) higher-order statistics. Cumulants of any order are well-known to be the derivatives[1] (of respective orders) of the LCF at the origin. When second-order cumulants (correlations) are insufficient to extract sufficient statistical information, a classical approach is to resort to higher-order derivatives (cumulants) at the origin. As an appealing alternative, it is also possible to remain at the more comfortable second-order differentiation, but to move away from the origin. These second-order derivatives maintain the convenient form of matrices (called Hessians), rather than tensors (multi-way arrays), which represent higher-order derivatives. In the sequel we shall refer to these matrices as "charrelation" (pronounced "car-relation") matrices (substituting correlation matrices), reflecting their link to the characteristic function.

The use of off-origin derivatives (of arbitrary order) of the LCF seems to have been first proposed by Gürelli and Nikias in [4] in the context of various array-processing applications, but has not been further pursued by these authors in open literature since. More recently, the use of second-order off-origin derivatives has been proposed by Yeredor *et al.* in various contexts, such as blind source separation [8], blind Finite Impulse Response (FIR) channel identification [9, 2], and Direction of Arrival (DOA) estimation [10]. Off-origin derivatives have also been used by Kawanabe and Theis [5] and by Comon and Rajih [1].

Let $\boldsymbol{y} \in \mathbb{R}^K$ denote a random vector and let $\boldsymbol{\tau} \in \mathbb{C}^K$ denote an arbitrary (deterministic) vector, to which we shall refer as a "processing point". The (generalized[2]) characteristic function (CF) and the LCF are defined, respectively, as

$$\phi_{\boldsymbol{y}}(\boldsymbol{\tau}) \stackrel{\triangle}{=} E[e^{\boldsymbol{\tau}^T \boldsymbol{y}}] \quad , \quad \psi_{\boldsymbol{y}}(\boldsymbol{\tau}) \stackrel{\triangle}{=} \log(\phi_{\boldsymbol{y}}(\boldsymbol{\tau})), \qquad (5)$$

whenever these means exist. Note that if the support of the probability distribution function of $\boldsymbol{y}$ is finite, then $\phi_{\boldsymbol{y}}(\boldsymbol{\tau})$ exists for all $\boldsymbol{\tau}$. Otherwise, it exists for all imaginary-valued $\boldsymbol{\tau}$, and may or may not exist for general complex-valued $\boldsymbol{\tau}$. The $K \times 1$ gradients and $K \times K$ Hessians are defined, respectively, as:

$$\boldsymbol{\phi}_{\boldsymbol{y}}(\boldsymbol{\tau}) \stackrel{\triangle}{=} \left.\frac{\partial^T \phi_{\boldsymbol{y}}(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\right|_{\boldsymbol{\tau}} \quad , \quad \boldsymbol{\psi}_{\boldsymbol{y}}(\boldsymbol{\tau}) \stackrel{\triangle}{=} \left.\frac{\partial^T \psi_{\boldsymbol{y}}(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\right|_{\boldsymbol{\tau}}, \quad (6)$$

$$\boldsymbol{\Phi}_{\boldsymbol{y}}(\boldsymbol{\tau}) \stackrel{\triangle}{=} \left.\frac{\partial^2 \phi_{\boldsymbol{y}}(\tau)}{\partial \boldsymbol{\tau}^2}\right|_{\boldsymbol{\tau}} \quad , \quad \boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{\tau}) \stackrel{\triangle}{=} \left.\frac{\partial^2 \psi_{\boldsymbol{y}}(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}^2}\right|_{\boldsymbol{\tau}}. \quad (7)$$

$\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{\tau})$ will serve as our alternative "charrelation" matrix. The following general properties of charrelation matrices $\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{\tau})$ would be useful in our derivations:

Property 1. *If $\boldsymbol{y}$ can be partitioned into statistically independent groups, then $\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{\tau})$ is block-diagonal (with the respective partition) for all $\boldsymbol{\tau}$ (at which it exists). Namely, two statistically independent random vectors are not only uncorrelated, but also "uncharrelated".*

Proof. Assume that $\boldsymbol{y} = [\boldsymbol{y}_1^T \ \boldsymbol{y}_2^T]^T$, where $\boldsymbol{y}_1 \in \mathbb{R}^{K_1}$ and $\boldsymbol{y}_2 \in \mathbb{R}^{K_2}$ are statistically independent (with $K_1 + K_2 = K$). Then, defining a similar partition for $\boldsymbol{\tau} = [\boldsymbol{\tau}_1^T \ \boldsymbol{\tau}_2^T]^T$, we have, due to the statistical independence,

$$\phi_{\boldsymbol{y}}(\boldsymbol{\tau}) = E[e^{\boldsymbol{\tau}_1^T \boldsymbol{y}_1 + \boldsymbol{\tau}_2^T \boldsymbol{y}_2}] = \phi_{\boldsymbol{y}_1}(\boldsymbol{\tau}_1)\phi_{\boldsymbol{y}_2}(\boldsymbol{\tau}_2), \qquad (8)$$

thus

$$\psi_{\boldsymbol{y}}(\boldsymbol{\tau}) = \psi_{\boldsymbol{y}_1}(\boldsymbol{\tau}_1) + \psi_{\boldsymbol{y}_2}(\boldsymbol{\tau}_2), \qquad (9)$$

hence $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are decoupled in $\psi_{\boldsymbol{y}}(\boldsymbol{\tau})$, therefore the cross-derivatives (the respective off-diagonal blocks of $\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{\tau})$) vanish, and $\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{\tau})$ is block-diagonal. □

[1] Up to irrelevant multiplications by powers of $j = \sqrt{-1}$.

[2] This definition slightly differs from the classical definition, by allowing a complex-valued argument rather than a real-valued argument multiplied by j.

Property 2. *If $\boldsymbol{y}$ can be expressed as a linear transformation of another random vector $\boldsymbol{z} \in \mathbb{R}^L$, namely $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}$ where $\boldsymbol{A}$ is any $K \times L$ matrix, then*

$$\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{\tau}) = \boldsymbol{A}\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{A}^T\boldsymbol{\tau})\boldsymbol{A}^T, \tag{10}$$

where $\boldsymbol{\Psi}_{\boldsymbol{z}}(\boldsymbol{A}^T\boldsymbol{\tau})$ is the charrelation matrix of $\boldsymbol{z}$ at $\boldsymbol{A}^T\boldsymbol{\tau}$. Thus, the effect of a linear transformation on the charrelation matrix resembles its effect on the correlation matrix.

Proof. Note that $\phi_{\boldsymbol{y}}(\boldsymbol{\tau}) = E[e^{\boldsymbol{\tau}^T\boldsymbol{A}\boldsymbol{z}}] = \phi_{\boldsymbol{z}}(\boldsymbol{A}^T\boldsymbol{\tau})$, so $\psi_{\boldsymbol{y}}(\boldsymbol{\tau}) = \psi_{\boldsymbol{z}}(\boldsymbol{A}^T\boldsymbol{\tau})$, and the result follows immediately by applying the chain-rule in the differentiation. □

3. CHARRELATION-BASED YW EQNS.

We now return to the AR process $x[n]$. Define a vector $\boldsymbol{y}_n \stackrel{\triangle}{=} [w[n]\ x[n-1]\ x[n-2]\ \cdots\ x[n-p]]^T$, which can be expressed as a linear transformation of the vector $\boldsymbol{x}_n$ (of (4)), as $\boldsymbol{y}_n = \boldsymbol{A}\boldsymbol{x}_n$:

$$\underbrace{\begin{bmatrix} w[n] \\ x[n-1] \\ \vdots \\ \vdots \\ x[n-p] \end{bmatrix}}_{\boldsymbol{y}_n} = \underbrace{\begin{bmatrix} 1 & a_1 & a_2 & \cdots & a_p \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ \vdots \\ x[n-p] \end{bmatrix}}_{\boldsymbol{x}_n}. \tag{11}$$

It then follows from Property 2 that (for convenience,we drop the index n from charrelations of $\boldsymbol{x}_n$, $\boldsymbol{y}_n$, due to stationarity)

$$\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{A}^{-1}\boldsymbol{\tau}) = \boldsymbol{A}\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})\boldsymbol{A}^T, \tag{12}$$

where, due to the statistical independence between $w[n]$ and past-values of $x[n]$, and by virtue of Property 1, $\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{A}^{-1}\boldsymbol{\tau})$ is block-diagonal: It consists of an upper-left 1×1 block and a lower-right $p\times p$ block. Therefore, multiplying (12) by $\boldsymbol{e}_1$ on the right, we obtain

$$v_w(\boldsymbol{\tau})\boldsymbol{e}_1 = \boldsymbol{A}\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})\boldsymbol{A}^T\boldsymbol{e}_1, \tag{13}$$

where $v_w(\boldsymbol{\tau})$ is some constant. In fact, $v_w(\boldsymbol{\tau})$ is the upper-left 1×1 "block" of $\boldsymbol{\Psi}_{\boldsymbol{y}}(\boldsymbol{A}^{-1}\boldsymbol{\tau})$, which is merely the second derivative of the LCF of the driving noise $w[n]$ at $(\boldsymbol{A}^{-1}\boldsymbol{\tau})_1$ (the first element of the vector $(\boldsymbol{A}^{-1}\boldsymbol{\tau})$). It is a statistical property of the driving-noise, depending on the "processing-point" $\boldsymbol{\tau}$ and, generally, on the AR parameters. For $\boldsymbol{\tau} = \boldsymbol{0}$, $v_w(\boldsymbol{0})$ is the variance σ_w^2 of $w[n]$.

Since $\boldsymbol{A}^{-1}\boldsymbol{e}_1 = \boldsymbol{e}_1$ and $\boldsymbol{A}^T\boldsymbol{e}_1 = \bar{\boldsymbol{a}}$, we can obtain the full-form of the YW equations by left-multiplying (13) with $\boldsymbol{A}^{-1}$ (and switching sides, for convenience):

$$\boxed{\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})\bar{\boldsymbol{a}} = v_w(\boldsymbol{\tau})\boldsymbol{e}_1}. \tag{14}$$

This is our new "full-form" of the YW equations, applied to the charrelation matrix $\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})$, rather than to the correlation matrix $\boldsymbol{R}$ as in (2). In fact, for $\boldsymbol{\tau} = \boldsymbol{0}$, $\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})$ coincides with $\boldsymbol{R}$, and (14) reduces to (2). To obtain the "reduced-form" (as in (3)), we only consider the last p rows,

$$\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})_{(2:p+1,2:p+1)}\boldsymbol{a} = \boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})_{(2:p+1,1)}, \tag{15}$$

where the indices in the subscripts are in Matlab® notations. With any selected "processing-point" $\boldsymbol{\tau}$ for which $\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})$ exists, these equations can be used for extracting AR parameters from the charrelation matrix.

4. CHARRELATION MATRIX ESTIMATES

In order to exploit these YW equations in estimating the AR parameters from the available data $x[n]$, the charrelation matrix $\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})$ has to be estimated first. To this end, we now present a simple, intuitively appealing estimation scheme. Observe first, by straightforward differentiation, that the Hessian of the LCF can be expressed in terms of the CF, its gradient and its Hessian (defined in (5)-(7)) as follows:

$$\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau}) = \frac{\boldsymbol{\Phi}_{\boldsymbol{x}}(\boldsymbol{\tau})\phi_{\boldsymbol{x}}(\boldsymbol{\tau}) - \boldsymbol{\phi}_{\boldsymbol{x}}(\boldsymbol{\tau})\boldsymbol{\phi}_{\boldsymbol{x}}^T(\boldsymbol{\tau})}{\phi_{\boldsymbol{x}}^2(\boldsymbol{\tau})}. \tag{16}$$

These quantities can in turn be consistently[3] estimated from the observed data using straightforward empirical averaging and differentiation: Defining $\alpha_n \stackrel{\triangle}{=} e^{\boldsymbol{\tau}^T\boldsymbol{x}_n}$,

$$\hat{\phi}_{\boldsymbol{x}}(\boldsymbol{\tau}) = \frac{1}{N}\sum_n \alpha_n\ , \quad \hat{\boldsymbol{\phi}}_{\boldsymbol{x}}(\boldsymbol{\tau}) = \frac{1}{N}\sum_n \boldsymbol{x}_n\alpha_n,$$
$$\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}(\boldsymbol{\tau}) = \frac{1}{N}\sum_n \boldsymbol{x}_n\boldsymbol{x}_n^T\alpha_n. \tag{17}$$

With slight manipulations, an estimate of the charrelation matrix, resulting from substitution of (4) in (16), can be conveniently expressed as

$$\hat{\boldsymbol{\Psi}}_{\boldsymbol{x}}(\boldsymbol{\tau}) = \frac{1}{\sum_{n=p}^{N-1}\alpha_n}\sum_{n=p}^{N-p}\alpha_n(\boldsymbol{x}_n - \tilde{\boldsymbol{x}})(\boldsymbol{x}_n - \tilde{\boldsymbol{x}})^T, \tag{18}$$

which can be regarded as a "specially weighted" covariance matrix, where the "weights" are α_n, and $\tilde{\boldsymbol{x}}$ is a similarly weighted mean, $\tilde{\boldsymbol{x}} \stackrel{\triangle}{=} \sum_{n=p}^{N-1}\alpha_n\boldsymbol{x}_n / \sum_{n=p}^{N-1}\alpha_n$. Note that when $\boldsymbol{\tau}$ is real-valued, the α_n are all real-positive, and the interpretation of "weights" is valid. However, when $\boldsymbol{\tau}$ is complex-valued, so are (in general)

[3] Under commonly met regularity conditions; Discussion omitted for lack of space.

the α_n, which can no longer be intuitively interpreted as weights - nevertheless, the expression (18) still forms a valid, usually consistent estimate of $\boldsymbol{\Psi}_{\boldsymbol{x}}(\boldsymbol{\tau})$. Obviously, when $\boldsymbol{\tau} = \mathbf{0}$, all α_n are 1, and this "weighted covariance" estimate reduces to the standard covariance estimate, as expected.

Once the charrelation matrix is consistently estimated at some pre-selected "processing-point" $\boldsymbol{\tau}$, the YW equations (15) may be solved for a consistent estimate of $\boldsymbol{a}$.

5. SIMULATION RESULTS

To demonstrate the potential improvement, we present the results of two experiments involving non-Gaussian AR processes, where the "driving-noise" is a Gaussian-mixture composed of two equiprobable Gaussians with means ± 1 and equal variances 0.2^2.

In the first experiment we consider a first-order process, and demonstrate the dependence of the performance (MSE) on selection of the processing-point $\boldsymbol{\tau}$ (two-dimensional in this case). It is evident (Fig.1) that the point $\boldsymbol{\tau} = \mathbf{0}$, corresponding to the classical correlation-based estimate, is far from optimal.

In the second experiment we consider a third-order process, comparing the performance of the proposed estimator (at one selected processing-point) to the correlation based estimate, as well as to a fourth-cumulant-slice based estimate (e.g., [3]) and to the CRLB, vs. the number of samples N. The results (Fig.2) show the MSE in estimation of each of the coefficients, clearly demonstrating the potential superiority of the proposed method, which, although still sub-optimal, significantly reduces the gap to the bound.

6. CONCLUSION

We introduced the use of the Hessian of the LCF (termed "charrelation" in here), evaluated at a pre-selected "processing point" $\boldsymbol{\tau}$, in the framework of YW equations for AR parameters estimation. When the process is non-Gaussian, the point $\boldsymbol{\tau} = \mathbf{0}$, leading to the standard correlation-based equations, is generally sub-optimal, and better selection of $\boldsymbol{\tau}$ can lead to improved performance. A remaining key question is how to select a "good" $\boldsymbol{\tau}$ (or a combination of equations based on several $\boldsymbol{\tau}$-s) - which is the subject of continuing research.

7. REFERENCES

[1] Comon P. and Rajih M., "Blind Identification of Under-Determined Mixtures based on the Characteristic Function," *Signal Processing*, vol.86 no.9, pp.2271-2281, 2006.

[2] Eidinger E. and Yeredor A., "Blind mimo identification using the second characteristic function," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4067–4079, 2005.

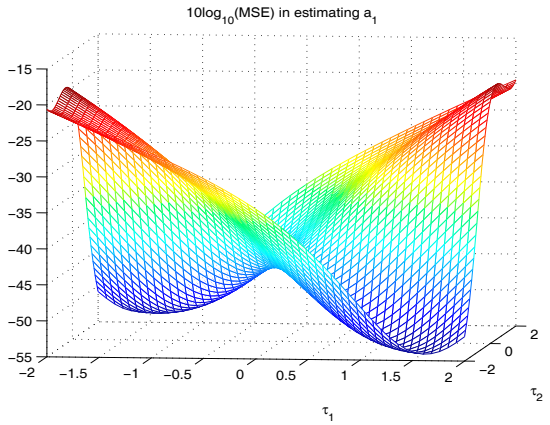

Figure 1: MSE in estimating a_1 for each "processing-point" $\boldsymbol{\tau} = [\tau_1\ \tau_2]^T$. True value: $a_1 = -0.8$, $N = 10000$. Each point on the mesh represents the average of 400 trials.

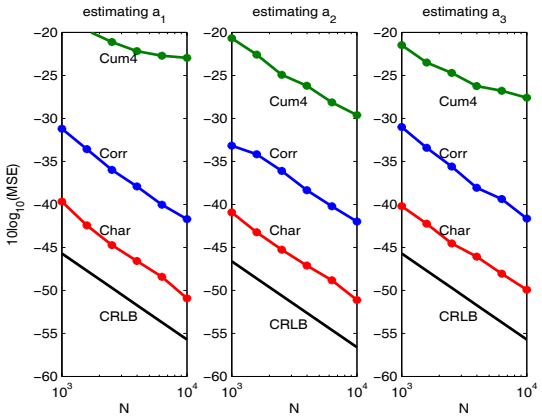

Figure 2: MSE vs. N in estimating a_1, a_2, a_3. 'Char' - proposed approach with $\boldsymbol{\tau} = [-1.6\ 0.4\ 0.8\ -0.8]^T$, 'Corr' - correlation-based, 'Cum4' - 4th cumulants-based. True value: $\boldsymbol{a} = [-0.3\ -0.47\ 0.549]^T$ (poles at $0.9, 0.6 \pm 0.5j$). Each point averaged over 400 trials.

[3] Giannakis G.B., Mendel J.M. and Wang W, "ARMA Modeling using cumulant and autocorrelation statistics," *Proc. ICASSP'87*, pp.61-64, 1987.

[4] Gürelli M. I. and Nikias C. L., "Blind identification and array processing applications of generalized higher-order statistics," *Proc. MILCOM*, pp. 838–842, 1996.

[5] Kawanabe M. and Theis F.J. "Estimating non-gaussian subspaces by characteristic functions," *Proc. ICA'2006* pp.157-164, 2006.

[6] Kay S.M. *"Fundamentals of Statistical Signal Processing - Estimation Theory,"* Prentice-Hall, 1993.

[7] Verbout S.M., Jeffrey J.M.O., Ludwig J.T. and Oppenheim, A.V., "Parameter estimation for autoregressive Gaussian-mixture processes: the EMAX algorithm ," *IEEE Trans. on Signal Processing*, vol.46, no.10, pp.2744-2756, Oct. 1998.

[8] Yeredor A. "Blind Source Separation via the Second Characteristic Function," *Signal Processing*, vol. 80, no. 5, pp. 897–902, 2000.

[9] Yeredor A., "Blind channel estimation using first and second derivatives of the characteristic function," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 100–103, 2002.

[10] Yeredor A., "MUSIC using Off-Origin Hessians of the Second Characteristic Function," *Proc. SAM2006*, 2006.