Yule-Walker Equations Using Higher Order Statistics for Nonlinear Autoregressive Model

Urszula Libal*† and Karl H. Johansson*

*School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology
Stockholm, Sweden
urli@kth.se, kallej@kth.se

†Signal Processing Systems Department
Wroclaw University of Science and Technology
Wroclaw, Poland

***Abstract*—The paper presents the derivation of Yule-Walker equations for the nonlinear autoregressive model NAR(p) of a time-series. The proposed method allows for easy calculation of parameters of the model. In the determined equations, higher order statistics are used, instead of autocovariances. For the linear autoregressive model AR(p), the standard Yule-Walker equations are directly based on autocovariances of time-series, or equivalently autocorrelations if the equations are rescaled. Unfortunately, it does not apply for nonlinear model. The authors show a compact matrix notation of Yule-Walker equations for the nonlinear autoregressive model with the nonlinearity of polynomial type of degree 2, with the use of higher order statistics up to 4th order, and numerical examples for electromyography signals for different hand movements.**

***Index Terms*—nonlinear model, autoregressive model, Yule-Walker equations, time-series, higher order statistics**

I. Introduction

The autoregressive (AR) model, all together with the moving average (MA) process, are the most basic model types, very well-known and widely used in statistics and signal processing. The linear prediction problem is overall the estimation of the linear model parameters by minimizing mean square error. Generally, the Yule-Walker equations [7], [21] leads to parameter estimation problem using the covariance matrix of the time-series. If the time-series is stationary in wide sense, then the linear prediction is based only on covariance function of the signal, ie. the covariance matrix is Toeplitz now, and can be efficiently solved by the Schur algorithm [3], [4] and the Levinson algorithm [1], [10].

The mostly known, specific, nonlinear autoregressive models are the following two: the threshold autoregressive (TAR) model [5], [19] and the smooth transition autoregressive (STAR) model [18], [20]. Mainly, they are used for financial forecasting to predict the dynamics of various market indices. The first one, TAR model is piecewise linear with switching of model parameter values. The switching behaviour depends on achieving the threshold value by a time-series. The second – STAR model – can be explained as a 'smooth' version of TAR model, with switch in parameter values controlled by a known distribution function. In fact, the TAR and the STAR models are nonlinear only because of the use of the parameter values switching.

In scientific literature, we can also find other attempts to derive nonlinear autoregressive models for particular cases, as in [13] for a very specific and simple nonlinear case, but the paper focuses rather on increasing the accuracy of minimizing a model error. The separate group provide kernel autoregressive models [8], [9], for which we have the freedom for choosing a mapping function, but but at the same time we do not get a simple interpretation of the dependency based on second- and higher-order statistics [14], as it is for the proposed nonlinear method.

The nonlinear autoregressive (NAR) model, proposed in the paper, drops the standard assumptions of Gaussian innovations [2] for linear autoregressive (AR) model. The assumption that holds in our paper is zero-mean independent innovations, ie. independence of innovation at a given time instance from its previous lags. In literature, eg. [6], we can find also considerations about the first-order nonlinear autoregressive model with dependent innovations, but the present paper focus strictly on independent innovations. The formal details about necessary assumptions can be found in the following Sections I-A, II and III.

A. Assumptions and denotations

Let $\{y_t\}_t$ be a discrete time-series. Without loss of generalization, we can fix the time instance t and operate only on the delays. Then, we assume that the time-series has mean value $\mathbb{E}y_{t-i} = h_i$ and finite variance $\mathbb{E}(y_{t-i} - \mathbb{E}y_{t-i})^2 = c_{i;i} < \infty$ for each $i \leq t$. We also introduce the following denotations for second- and higher-order statistics (HOS) [14]:

$$h_{ik} = \mathbb{E}y_{t-i}y_{t-k}, \tag{1}$$

$$h_{ijk} = \mathbb{E}y_{t-i}y_{t-j}y_{t-k}, \tag{2}$$

$$h_{ijkl} = \mathbb{E}y_{t-i}y_{t-j}y_{t-k}y_{t-l}. \tag{3}$$

978-1-7281-1715-7/19/$31.00 ©2019 IEEE

The order of indices i, j, k, l is not important due to the alternation of multiplication of first- and second-degree expressions as y_{t-i} and $y_{t-i}y_{t-j}$. On the other hand, the autocovariances can be expressed by those statistics:

$$\begin{aligned} c_{i;k} &= Cov(y_{t-i}; y_{t-k}) \\ &= \mathbb{E}y_{t-i}y_{t-k} - \mathbb{E}y_{t-i}\mathbb{E}y_{t-k} \\ &= h_{ik} - h_i h_k, \end{aligned} \tag{4}$$

$$\begin{aligned} c_{ij;k} &= Cov(y_{t-i}y_{t-j}; y_{t-k}) \\ &= \mathbb{E}y_{t-i}y_{t-j}y_{t-k} - \mathbb{E}y_{t-i}y_{t-j}\mathbb{E}y_{t-k} \\ &= h_{ijk} - h_{ij}h_k, \end{aligned} \tag{5}$$

$$\begin{aligned} c_{ij;kl} &= Cov(y_{t-i}y_{t-j}; y_{t-k}y_{t-l}) \\ &= \mathbb{E}y_{t-i}y_{t-j}y_{t-k}y_{t-l} - \mathbb{E}y_{t-i}y_{t-j}\mathbb{E}y_{t-k}y_{t-l} \\ &= h_{ijkl} - h_{ij}h_{kl}. \end{aligned} \tag{6}$$

The autocovariances for different permutations of indices are also equal, but with respect to the division for the initial expressions. And so, the symmetry for second-order autocovariance is $c_{i;k} = c_{k;i}$, third-order $c_{ij;k} = c_{ji;k} = c_{k;ij} = c_{k;ji}$ and fourth-order $c_{ij;kl} = c_{ji;kl} = c_{kl;ij} = c_{kl;ji} = c_{ij;lk} = c_{ji;lk} = c_{lk;ij} = c_{lk;ji}$.

Let $\{\epsilon_t\}_t$ be discrete-time white noise, i.e. the random variables $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots$ are independent and identically distributed (i.i.d.) through time, with zero mean $\mathbb{E}\epsilon_{t-i} = 0$ for each i, and autocovariance is $\mathbb{E}(\epsilon_{t-i}\epsilon_{t-k}) = \sigma^2\delta_{ik}$ for each i, k, where δ_{ik} is Kronecker delta and $\sigma < \infty$.

If we additionally assume that time-series $\{y_t\}_t$ is weak-sense stationary, then of course it is possible to simplify the proposed denotations. We resign intentionally from simplifying the indices in order to maintain the transparent structure of the 'flat' matrix ${}^{\{2\times 2\}}H$ of higher order statistics, which will be presented in the Section III. In spite of all, for stationary time-series we need to remember that mean value is constant $h_i = h_{i+\Delta}$ and we have the following types of equalities for second-order autocovariances $c_{i;k} = c_{i+\Delta;k+\Delta}$, third-order $c_{ij;k} = c_{i+\Delta,j+\Delta;k+\Delta}$, and fourth-order $c_{ij;kl} = c_{i+\Delta,j+\Delta;k+\Delta,l+\Delta}$. The same apply for second and higher order statistics: $h_{ik} = h_{i+\Delta,k+\Delta}$, third-order $h_{ijk} = h_{i+\Delta,j+\Delta,k+\Delta}$, and fourth-order $h_{ijkl} = h_{i+\Delta,j+\Delta,k+\Delta,l+\Delta}$.

II. Linear autoregressive model

Assume that discrete time-series $\{y_t\}_t$ is weakly stationary. We consider the linear autoregressive model, denoted AR(p), of order p:

$$y_t = \sum_{k=1}^{p} a_k y_{t-k} + \epsilon_t, \tag{7}$$

where innovations $\{\epsilon_t\}_t$ are zero mean white noise. One of the methods of finding the model parameters $a_1, a_2, \ldots, a_p$ is based on Yule-Walker equations. For p unknown parameters $a_k,\ k = 1, 2, \ldots, p$, we construct p equations in the following way. The first equation is obtained by multiplying the equation (7) by term y_{t-1}, and next by taking the expected value of both sides of the equation. The second – by multiplying the equation (7) by term y_{t-2} and taking the expected value. And so on, till the last equation received by multiplying (7) by y_{t-p}, and as previously by taking the expected value. The Yule-Walker equations are as follows:

$$\begin{cases} \mathbb{E}y_t y_{t-1} = a_1\mathbb{E}y_{t-1}y_{t-1} + \cdots + a_p\mathbb{E}y_{t-p}y_{t-1} \\ \mathbb{E}y_t y_{t-2} = a_1\mathbb{E}y_{t-1}y_{t-2} + \cdots + a_p\mathbb{E}y_{t-p}y_{t-2} \\ \mathbb{E}y_t y_{t-3} = a_1\mathbb{E}y_{t-1}y_{t-3} + \cdots + a_p\mathbb{E}y_{t-p}y_{t-3} \\ \qquad\qquad\vdots \\ \mathbb{E}y_t y_{t-p} = a_1\mathbb{E}y_{t-1}y_{t-p} + \cdots + a_p\mathbb{E}y_{t-p}y_{t-p} \end{cases} \tag{8}$$

The expressions $\mathbb{E}\epsilon_t y_{t-k} = 0$ for all $k = 1, \ldots, p$, because of independence of innovation ϵ_t from previous ones $\epsilon_{t-1}, \epsilon_{t-2} \ldots$ and zero-mean assumption. The additional equation can be formed

$$h_{00} = a_1 h_{10} + \cdots + a_p h_{p0} + \sigma^2. \tag{9}$$

Denotating the second-order statistics h_{ik} as in (1), we obtain the following matrix form of Yule-Walker equations:

$$\begin{bmatrix} h_{01} \\ h_{02} \\ h_{03} \\ \vdots \\ h_{0p} \end{bmatrix} = \begin{bmatrix} h_{11} \ldots h_{p1} \\ h_{12} \ldots h_{p2} \\ h_{13} \ldots h_{p3} \\ \vdots \ \ddots \ \vdots \\ h_{1p} \ldots h_{pp} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix}. \tag{10}$$

The problem of finding model parameters $\mathbf{a} = [a_1, \ldots, a_p]^T$ is completely defined by second-order statistics ${}^{1\oplus 1}H = [h_{ik}]_{i,k=1,\ldots,p}$:

$${}^{2}\mathbf{h} = {}^{1\oplus 1}H \cdot \mathbf{a}. \tag{11}$$

By replacing the theoretical expected value in the term $h_{ik} = \mathbb{E}y_{t-i}y_{t-k}$ with its empirical analog – sample second-order statistic (in signal processing also called sample autocorrelation), e.g. $\widehat{h}_{ik} = (\sum_{t=0}^{N} y_{t-i}y_{t-k})/(N+1)$, the problem (11) transforms to

$${}^{2}\widehat{\mathbf{h}} = {}^{1\oplus 1}\widehat{H} \cdot \widehat{\mathbf{a}}, \tag{12}$$

where all elements of vector ${}^{2}\widehat{\mathbf{h}}$ and matrix ${}^{1\oplus 1}\widehat{H}$ can be easily estimated using the measured data.

III. Nonlinear autoregressive model

Now, we will consider the nonlinear autoregressive model, denoted NAR(p), of order p with polynomial nonlinearity type of degree 2:

$$y_t = \sum_{k=1}^{p} a_k y_{t-k} + \sum_{k_1=1}^{p}\sum_{k_2 \leqslant k_1} a_{k_1k_2} y_{t-k_1}y_{t-k_2} + \epsilon_t. \tag{13}$$

This NAR(p) is the simplest case of more general nonlinear autoregressive model of polynomial degree n of the form:

$$\begin{aligned} y_t = &\sum_{k=1}^{p} a_k y_{t-k} + \sum_{k_1=1}^{p}\sum_{k_2 \leqslant k_1} a_{k_1k_2} y_{t-k_1}y_{t-k_2} + \cdots + \\ &+ \sum_{k_1=1}^{p}\sum_{k_2 \leqslant k_1} \cdots \sum_{k_n \leqslant k_{n-1}} a_{k_1k_2\ldots k_n} y_{t-k_1}y_{t-k_2}\cdots y_{t-k_n} + \epsilon_t. \end{aligned} \tag{14}$$

All previously introduced assumptions stay.

The NAR(p) model uses p parameters $\{a_k\}$, $k = 1, \dots, p$ in the linear part and additional $p(p+1)/2$ parameters $\{a_{k_1 k_2}\}$ in nonlinear part, where $k_1 = 1, \dots, p$ and $k_2 = 1, \dots, k_1$, due to reduce the repeating nonlinear terms. This gives in total

$$N_p = p + p(p+1)/2 \tag{15}$$

unknown coefficients of the nonlinear model.

To construct the Yule-Walker equations for NAR(p) model, we act in the same way as in the linear case. Firstly, we multiply the equation (13) from both sides by linear terms y_{t-k} for $k = 1, \dots, p$ and then by nonlinear $y_{t-k_1} y_{t-k_2}$ for $k_1 = 1, \dots, p$ and $k_2 = 1, \dots, k_1$. And next, we put expected value on all $p + p(p+1)/2$ equations, what gives:

$$\begin{cases} \text{linear terms:} \\ \mathbb{E} y_t y_{t-1} = \sum a_k \mathbb{E} y_{t-k} y_{t-1} + \sum\sum a_{ij} \mathbb{E} y_{t-i} y_{t-j} y_{t-1} \\ \mathbb{E} y_t y_{t-2} = \sum a_k \mathbb{E} y_{t-k} y_{t-2} + \sum\sum a_{ij} \mathbb{E} y_{t-i} y_{t-j} y_{t-2} \\ \mathbb{E} y_t y_{t-3} = \sum a_k \mathbb{E} y_{t-k} y_{t-3} + \sum\sum a_{ij} \mathbb{E} y_{t-i} y_{t-j} y_{t-3} \\ \vdots \\ \mathbb{E} y_t y_{t-p} = \sum a_k \mathbb{E} y_{t-k} y_{t-p} + \sum\sum a_{ij} \mathbb{E} y_{t-i} y_{t-j} y_{t-p} \\ \hline \text{nonlinear terms:} \\ \mathbb{E} y_t y_{t-1} y_{t-1} = \\ \sum a_k \mathbb{E} y_{t-k} y_{t-1} y_{t-1} + \sum\sum a_{ij} \mathbb{E} y_{t-i} y_{t-j} y_{t-1} y_{t-1} \\ \mathbb{E} y_t y_{t-2} y_{t-1} = \\ \sum a_k \mathbb{E} y_{t-k} y_{t-2} y_{t-1} + \sum\sum a_{ij} \mathbb{E} y_{t-i} y_{t-j} y_{t-2} y_{t-1} \\ \mathbb{E} y_t y_{t-2} y_{t-2} = \\ \sum a_k \mathbb{E} y_{t-k} y_{t-2} y_{t-2} + \sum\sum a_{ij} \mathbb{E} y_{t-i} y_{t-j} y_{t-2} y_{t-2} \\ \vdots \\ \mathbb{E} y_t y_{t-p} y_{t-p} = \\ \sum a_k \mathbb{E} y_{t-k} y_{t-p} y_{t-p} + \sum\sum a_{ij} \mathbb{E} y_{t-i} y_{t-j} y_{t-p} y_{t-p} \end{cases} \tag{16}$$

Analogously to linear model, the expressions $\mathbb{E}\epsilon_t y_{t-k}$ disappear and also $\mathbb{E}\epsilon_t y_{t-i} y_{t-j} = 0$ for all $i, j = 1, \dots, p$, due to independence of innovation ϵ_t from random variables $z_{t,i,j} = y_{t-i} y_{t-j}$. The set of equations (16) can be extended by the additional equation

$$h_{00} = \sum a_k h_{k0} + \sum\sum a_{ij} h_{ij0} + \sigma^2, \tag{17}$$

obtained by multiplying (13) by y_t and taking expected value.

Using a matrix notation of the Yule-Walker equations (16) for NAR(p) model, we get:

$$\begin{bmatrix} h_{01} \\ h_{02} \\ h_{03} \\ \vdots \\ h_{0p} \\ \hline h_{011} \\ h_{021} \\ h_{022} \\ h_{031} \\ h_{032} \\ h_{033} \\ \vdots \\ h_{0pp} \end{bmatrix} = \left[\begin{array}{c|c} h_{11} \dots h_{p1} & h_{111} h_{211} \dots h_{pp1} \\ h_{12} \dots h_{p2} & h_{112} h_{212} \dots h_{pp2} \\ h_{13} \dots h_{p3} & h_{113} h_{213} \dots h_{pp3} \\ \vdots \ \ddots \ \vdots & \vdots \quad \vdots \ \ddots \ \vdots \\ h_{1p} \dots h_{pp} & h_{11p} h_{21p} \dots h_{ppp} \\ \hline h_{111} \dots h_{p11} & h_{1111} h_{2111} \dots h_{pp11} \\ h_{121} \dots h_{p21} & h_{1121} h_{2121} \dots h_{pp21} \\ h_{122} \dots h_{p22} & h_{1122} h_{2122} \dots h_{pp22} \\ h_{131} \dots h_{p31} & h_{1131} h_{2131} \dots h_{pp31} \\ h_{132} \dots h_{p32} & h_{1132} h_{2132} \dots h_{pp32} \\ h_{133} \dots h_{p33} & h_{1133} h_{2133} \dots h_{pp33} \\ \vdots \ \ddots \ \vdots & \vdots \quad \vdots \ \ddots \ \vdots \\ h_{1pp} \dots h_{ppp} & h_{11pp} h_{21pp} \dots h_{pppp} \end{array}\right] \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \\ \hline a_{11} \\ a_{21} \\ a_{22} \\ a_{31} \\ a_{32} \\ a_{33} \\ \vdots \\ a_{pp} \end{bmatrix} \tag{18}$$

The generalized (block, multi-indexed) higher order statistics matrix ${}^{\{2\times2\}}H$, shown in equation (18), consist of 4 main submatrices

$${}^{\{2\times2\}}H = \begin{bmatrix} {}^{1\oplus1}H & {}^{1\oplus2}H \\ {}^{2\oplus1}H & {}^{2\oplus2}H \end{bmatrix}. \tag{19}$$

The generalized multidimensional matrix ${}^{\{2\times2\}}H$ was modified to be a N_pxN_p 'flat' matrix in the following way. The second-order statistics matrix

$${}^{1\oplus1}H = [h_{ik}]_{i,k=1,\dots,p}. \tag{20}$$

remains as it is. The three-dimensional matrices, containing of third-order statistics $[h_{ijk}]_{i,j,k=1,\dots,p}$, are rewritten slice by slice in the 'flat' way in a row

$${}^{1\oplus2}H = [{}^{1\oplus2}H_k]_{k=1,\dots,p} \tag{21}$$

or in a column

$${}^{2\oplus1}H = col[{}^{2\oplus1}H_i]_{i=1,\dots,p}. \tag{22}$$

The four-dimensional matrix

$${}^{2\oplus2}H = [h_{ijkl}]_{i,j,k,l=1,\dots,p} \tag{23}$$

with fourth-order statistics is transformed to block-matrix of matrices, and hence, the 'flat' form is as follows

$${}^{\{2\times2\}}H = \begin{bmatrix} {}^{1\oplus1}H & {}^{1\oplus2}H_1 & \dots & {}^{1\oplus2}H_p \\ {}^{2\oplus1}H_1 & {}^{2\oplus2}H_{1,1} & \dots & {}^{2\oplus2}H_{p,1} \\ \vdots & \vdots & \ddots & \vdots \\ {}^{2\oplus1}H_p & {}^{2\oplus2}H_{1,p} & \dots & {}^{2\oplus2}H_{p,p} \end{bmatrix}. \tag{24}$$

Thereafter, the parameters $\mathbf{a} = [a_1, \dots, a_p, a_{11}, \dots, a_{pp}]^T$ of the nonlinear model must fulfill the matrix equation

$${}^{2,3}\mathbf{h} = {}^{\{2\times2\}} H \cdot \mathbf{a}. \tag{25}$$

The parameter estimates $\widehat{\mathbf{a}}$ can be found by solving the equation

$${}^{2,3}\widehat{\mathbf{h}} = {}^{\{2\times2\}}\widehat{H} \cdot \widehat{\mathbf{a}}, \tag{26}$$

where theoretical statistics of second-, third- and fourth-order are replaced by the sample statistics.

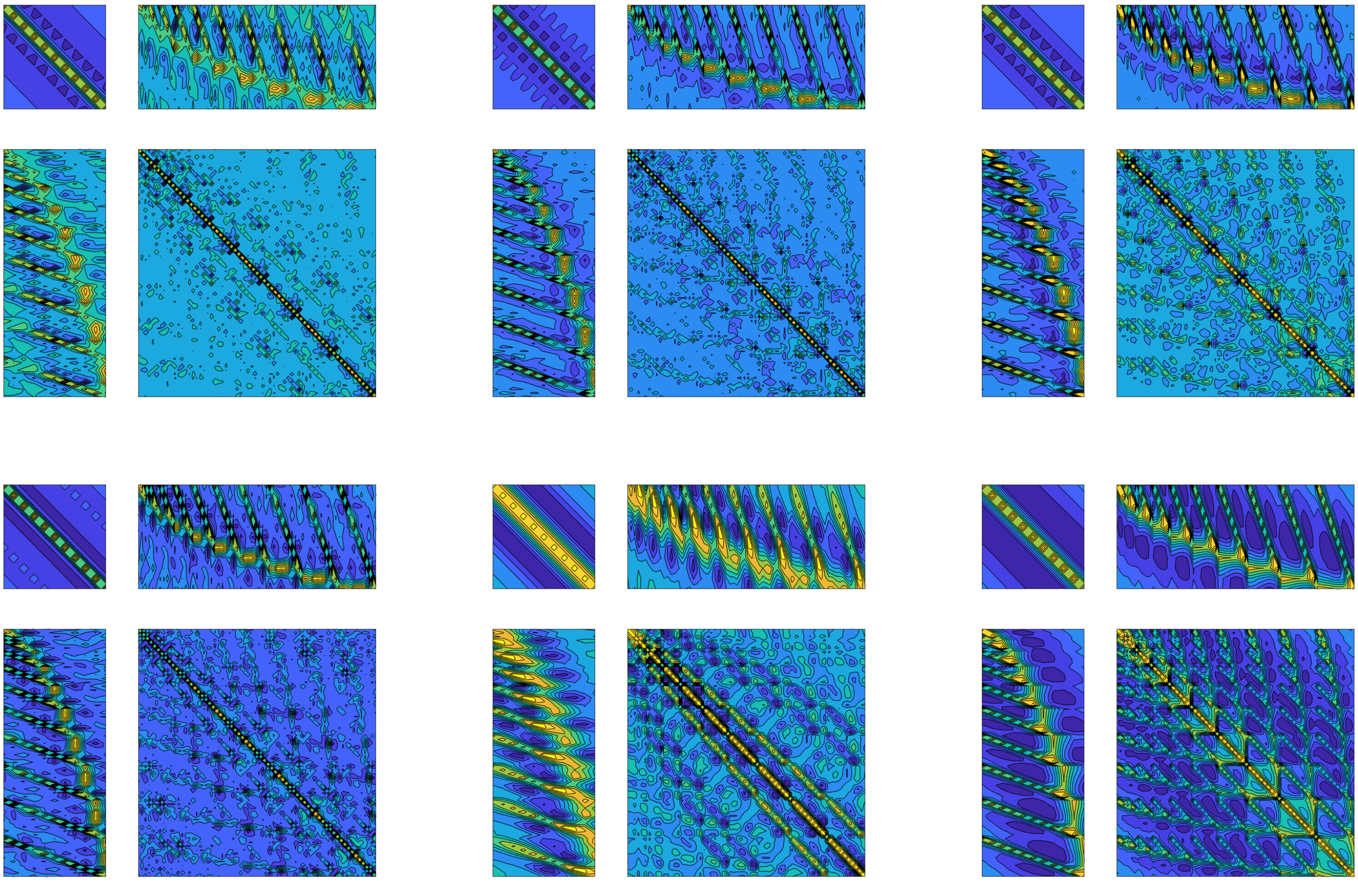

Fig. 1. Generalized higher order statistics matrices ${}^{\{2\times2\}}\widehat{H}$ for exemplary electromyographic signals from database [15] for 6 different hand movements.

IV. Practical example

In the following section, we will show that extending the basic linear autoregressive model to the nonlinear one, presented in Section III, allows to calculate model estimates based on not only second-order covariances, but on higher-order statistics ${}^{\{2\times2\}}H$. The numerical example was chosen to portray that nonlinear versions of Yule-Walker equations, or NAR models, can be more useful comparing to linear ones, thanks to information hidden in and represented by third- and fourth-order statistics.

The database used to illustrate higher-order statistics is a data set of (EMG) electromyographic signals [15]–[17], openly available at UCI Machine Learning Repository [12]. The data set consist of 6 basic hand movements, such as daily hand grasps: a) Spherical: for holding spherical tools, b) Tip: for holding small tools, c) Palmar: for grasping with palm facing the object, d) Lateral: for holding thin, flat objects, e) Cylindrical: for holding cylindrical tools, f) Hook: for supporting a heavy load.

The higher order statistics for EMG signals for 6 different hand movements show that second-order covariance matrices ${}^{1\oplus1}H$ are almost always nearly identity matrices (see Fig. 1), ie. there is no significant distinction between movement classes if only second-order statistics are considered. On the other hand, the third- ${}^{1\oplus2}H = \left[{}^{2\oplus1}H\right]^T$ and fourth-order ${}^{2\oplus2}H$ statistics are more discriminative, what for example can have crucial impact on movement recognition based on AR(p) estimates $\hat{\mathbf{a}}$.

V. Conclusion

The proposed methodology of constructing Yule-Walker equations for nonlinear autoregressive model NAR(p) was inspired by linear case AR(p) using directly autocovariance matrix for stationary zero-mean time-series. The Yule-Walker equations for NAR(p) model with polynomial type nonlinearity of degree 2, defined in equation (13), brings the problem to solving the matrix equation (26).

The main advantage of the proposed approach using higher-order statistics, comparing to kernel methods [8], [9], is clear and easy to interpret correspondence between signal properties and significant values in generalised block-matrix ${}^{\{2\times2\}}H$ – for details see [14]. On the other hand, this method is redundant for example for stationary Gaussian signals, for which there is no additional information hidden in higher-order statistics, therefore for such signals it is justified to use the linear AR model.

There are possible further generalizations for NAR(p) model with polynomial type nonlinearity of degree n, because each particular matrix ${}^{i\oplus k}H$ can be decomposed to row or column of its own slices and therefor the generalized matrix ${}^{\{n\times n\}}H$ could be also presented in 'flat' way. However, more interesting would be investigation on the possible calculations improvements as the famous LevinsonDurbin recursion for linear autoregressive model, using the Toeplitz property of autocovariance matrix.

REFERENCES

[1] M. Basseville, A. Benveniste and A.S. Willsky, "Multiscale autoregressive processes, Part I: Schur-Levinson parametrizations", IEEE Trans. Signal Processing, Vol. 40, No. 8, pp.1915–1934, 1992.

[2] F. Blasques, S.J. Koopman and A. Lucas, "Optimal formulations for nonlinear autoregressive processes", Tinbergen Institute Discussion Paper 14-103/III, 2014.

[3] P. Dewilde, H. Dym, "Schur recursions, error formulas and convergence of rational estimators for stationary stochastic sequences", IEEE Trans. on Information Theory, Vol. IT-27(4), July 1981, pp.446–461.

[4] P. Dewilde, E.F.A. Deprettere, "The generalized Schur algorithm: Approximation and hierarchy", in: Operator Theory: Advances and Applications, Vol. 29, Birkhäuser Verlag, Basel, 1988, pp.97–116.

[5] D. Gibson and D. Nur, "Threshold Autoregressive Models in Finance: A Comparative Approach", Proceedings of the 4th Annual ASEARC Conference, 17-18 February 2011, University of Western Sydney, Paramatta, Australia, 2011.

[6] A. Hajrajabi, A.R. Yazdanian and R. Farnoosh, "Nonlinear autoregressive model with stochastic volatility innovations: Semiparametric and Bayesian approach," Journal of Computational and Applied Mathematics, Vol. 344, pp.37–46, 2018.

[7] M.J.L. de Hoon, T.H.J.J. van der Hagen, H. Schoonewelle and H. van Dam, "Why Yule-Walker should not be used for autoregressive modelling", Annals of Nuclear Energy, Vol. 23, No. 15, pp.1219–1228, 1996.

[8] M. Kallas, C. Francis, P. Honeine, H. Amoud and C. Richard, "Modeling electrocardiogram using Yule-Walker equations and kernel machines," 2012 19th International Conference on Telecommunications (ICT), Jounieh, pp.1–5, 2012.

[9] M. Kallas, P. Honeine, C. Francis and H. Amoud, "Kernel autoregressive models using Yule-Walker equations", Signal Processing, Vol. 93, No. 11, pp.3053–3061, 2013.

[10] N. Levinson, "The Wiener RMS error criterium in filter design and prediction", Journal of Mathematical Physics, Vol. 25, pp.261–278, 1947.

[11] U. Libal, A. Wielgus and W. Magiera, "Nonlinear orthogonal parametrization and modeling of higher-order non-gaussian time-series", 2017 Signal Processing Symposium (SPSympo), Jachranka, pp.1–6, 2017.

[12] M. Lichman, "UCI Machine Learning Repository", University of California, Irvine, School of Information and Computer Sciences, 2013, http://archive.ics.uci.edu/ml.

[13] S. Lu, K.H. Chon, "Nonlinear Autoregressive and Nonlinear Autoregressive Moving Average Model Parameter Estimation by Minimizing Hypersurface Distance", IEEE Transactions on Signal Processing, Vol. 51, pp.3020–3026, 2003.

[14] W. Magiera, U. Libal, A. Wielgus, "Efficient measurement of higher-order statistics of stochastic processes", Kybernetika, Vol. 54, No. 5, pp.865–887, 2018.

[15] C. Sapsanis, A. Tzes and G. Georgoulas, "sEMG for basic hand movements data set", UCI ML Repository, https://archive.ics.uci.edu/ml/datasets/sEMG+for+Basic+Hand+movements, 2014.

[16] C. Sapsanis, "Recognition of basic hand movements using electromyography", (Diploma Thesis), 2013.

[17] C. Sapsanis, G. Georgoulas and A. Tzes, "EMG based classification of basic hand movements based on time-frequency features", 21st Mediterranean Conference on Control and Automation, Chania, pp.716–722, 2013.

[18] T. Teräsvirta and H.M. Anderson, "Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models", John Wiley and Sons, Vol. 7, 1992.

[19] H. Tong, "Threshold Models in Nonlinear Time Series Analysis", Springer-Verlag New York Inc., 1983.

[20] U.M. Umer, T. Sevil and G. Sevil, "Forecasting performance of smooth transition autoregressive (STAR) model on travel and leisure stock index", The Journal of Finance and Data Science, Vol. 4, No. 2, pp.90–100, 2018.

[21] U. Yule, "On a method for investigating period-icities in disturbed series, with special reference to Wolferssunspot numbers", Philosophical Transactions of the Royal Society of London, Vol. 226, pp.267–298, 1927.

[22] J. Zarzycki, Orthogonal Digital Filtering of Stochastic Signals (in polish), WNT, Warsaw, 1998.

[23] J. Zarzycki, P. Dewilde, "The nonlinear nonstationary Schur algorithm", Proc. Workshop on Advanced Algorithms and Their Realizations, Chateaux de Bonas, (M. Verhaegen, Ed.), Paper V3, 1991.

[24] J. Zarzycki, A. Wielgus and U. Libal, "Nonlinear Schur-type orthogonal transformations of higher-order stochastic processes: An overview of current topics", 2017 Signal Processing Symposium (SPSympo), Jachranka, pp.1–5, 2017.