Kybernetika

Wladyslaw Magiera; Urszula Libal; Agnieszka Wielgus

Efficient measurement of higher-order statistics of stochastic processes

Kybernetika, Vol. 54 (2018), No. 5, 865–887

Persistent URL: `http://dml.cz/dmlcz/147531`

KYBERNETIKA — VOLUME *54* (2018), NUMBER 5, PAGES 865–887

EFFICIENT MEASUREMENT OF HIGHER-ORDER STATISTICS OF STOCHASTIC PROCESSES

Wladyslaw Magiera, Urszula Libal and Agnieszka Wielgus

This paper is devoted to analysis of block multi-indexed higher-order covariance matrices, which can be used for the least-squares estimation problem. The formulation of linear and nonlinear least squares estimation problems is proposed, showing that their statements and solutions lead to generalized 'normal equations', employing covariance matrices of the underlying processes. Then, we provide a class of efficient algorithms to estimate higher-order statistics (generalized multi-indexed covariance matrices), which are necessary taking in mind practical aspects of the nonlinear treatment of the least-squares estimation problem. The algorithms are examined for different higher-order and non-Gaussian processes (time-series) and an impact of signal properties on covariance matrices is analysed.

Keywords: covariance matrix, higher-order statistics, adaptive, nonlinear

Classification: 15B51, 93E24, 15B05, 60G10, 60G15

1. INTRODUCTION

The least-squares estimation problems [3, 5] can be stated and solved using covariance data of the underlying stochastic processes. If the considered problem is linear, the 'sufficient statistics' are second-order (two-dimensional or two-indexed) covariance matrices, being actually positive-definite Hermitian matrices. If the underlying process is wide-sense stationary (in a weak second-order sense), those matrices become Toeplitz [11].

If the least-squares estimation problem considered becomes nonlinear, the 'sufficient statistics' are higher-order (multi-dimensional or multi-indexed) covariance (precisely – generalized, block multi-indexed) matrices, being generalized positive-definite Hermitian. If a higher-order process is wide-sense stationary (in a weak, higher-order sense), those matrices become generalized block-Toeplitz matrices.

If the underlying process is Gaussian, its second-order covariance matrix contains 'all' statistical information [11] (as in this case higher-order covariances are expressible in terms of sums of products of second-order covariances, and do not contain any 'new' statistical information about the process). Thus, in the Gaussian case, second-order statistics are sufficient statistics and it is enough to consider the linear estimation case for this class of stochastic processes.

DOI: 10.14736/kyb-2018-5-0865

In a non-Gaussian case higher-order statistics [6] become non trivial, and one has to employ generalized covariance matrices to enhance the estimation accuracy. Consequently, the linear approach has to be replaced by a nonlinear treatment of the problem.

In this paper we wish to propose a class of efficient algorithms for measurement/estimation of higher-order statistics (generalized block multi-indexed covariance matrices of Gaussian and non-Gaussian stochastic processes), taking in mind practical aspects of the nonlinear treatment of the least-squares estimation problem [7, 8]. We begin with formulation of the linear/nonlinear least squares estimation problems, showing that their statements/solutions lead to the generalized 'normal equations', employing covariance matrices of the underlying processes (two-indexed in the linear case and/or block multi-indexed in the nonlinear situation).

Then we briefly consider properties of higher-order generalized covariance matrices in the nonstationary versus stationary cases.

Next, we propose a class of efficient higher-order statistics estimation/measurement, including a new class of adaptive algorithms based on a generalized isomorphism between the spaces of random variables (and their products), and (linear and nonlinear) sample-observation vectors.

The results obtained are then confirmed by simulations for different classes of higher-order and non-Gaussian processes (actually – time-series).

2. LEAST-SQUARES ESTIMATION PROBLEM PROBLEM FOR SECOND- AND HIGHER-ORDER STOCHASTIC PROCESSES

Let $\{\Omega, \mathcal{B}, \mu\}$ be a probability space where Ω is an abstract set of elements $\omega \in \Omega$, $\mathcal{B}$ – σ-algebra of Borel subsets, and μ – a probability measure on $\mathcal{B}$. Via $L_2\{\Omega, \mathcal{B}, \mu\}$ we denote a separable Hilbert space of σ-measurable mappings $w : \Omega \to \mathcal{C}$, satisfying

$$\int_\Omega |w(\omega)|^2 \mu(\mathrm{d}\omega) < \infty.$$

We introduce in $L_2\{\Omega, \mathcal{B}, \mu\}$ the inner-product as

$$(w, v)_\Omega \triangleq \int_\Omega w(\omega)\bar{v}(\omega)\mu(\mathrm{d}\omega) = \mathbf{E}w\bar{v}$$

where $\bar{}$ denotes complex conjugate and $\mathbf{E}$ stands for the expectation operator. This inner-product induces the norm

$$\|w\|^2_\Omega = \int_\Omega |w(\omega)|^2 \mu(\mathrm{d}\omega) = \mathbf{E}|w|^2$$

and metric

$$d_\Omega(w, v) = \|w - v\|_\Omega.$$

With completeness of $L_2\{\Omega, \mathcal{B}, \mu\}$, that space will be a Hilbert space. Let T denote the set of natural numbers and let $t \in T$. The mapping

$$t \mapsto y_t = y_t(\omega) \in L_2\{\Omega, \mathcal{B}, \mu\}$$

will be called a discrete-time Hilbert stochastic process $\mathbf{y}$ if $\forall_{t\in T} \|y_t\|^2_\Omega < \infty$. Let $\mathbf{y}$ denote a discrete-time centred process.

2.1. Linear least-squares estimation problem

Let us consider a finite set of random variables

$$\{y_t, y_{t-1}, \dots, y_{t-n}\} \tag{2.1}$$

and take $t = 0$ for simplicity. This set can equivalently be rewritten as

$$Y = [y_{-i_1}]_{i_1=0,\dots,n}.$$

Observe that

$$^1Y = [y_{-i_1}]_{i_1=1,\dots,n} \tag{2.2}$$

will be the nth order past w.r. to the offset $t = 0$. Assuming that the random variables (2.1) form a linearly independent set, we can introduce the space spanned by the linear past

$$S^1 = \vee\{^1Y\}$$

where $\vee$ stands for 'the span of'. In the linear prediction problem we consider the space

$$S = \vee\{y_0\} \dot{+} S^1 \tag{2.3}$$

(where $\dot{+}$ denotes direct sum of subspaces) and define the nth order linear estimate as

$$\hat{y}_0^1 \triangleq P(S^1)y_0 \in S^1$$

where $P(S)$ denotes the orthogonal projection operator on S. We consider the nth order linear prediction error

$$\begin{aligned} \varepsilon_0^1 &\triangleq P(S \ominus S^1)y_0 \perp S^1 \\ &= y_0 + \sum_{i_1=1}^{n} a_{i_1} y_{-i_1} \end{aligned} \tag{2.4}$$

where $P(S \ominus S^1)$ denotes the orthogonal projection operator on the orthogonal complement of S^1 w.r. to S. Orthogonality of the error (2.4) w.r. to the subspace S^1 implies the following set of optimality conditions for the linear least-squares estimation problem

$$(\varepsilon_0^1, y_{-k_1}) = 0\ ,\ k_1 = 1, \dots, n.$$

Equivalently,

$$\begin{aligned} (\varepsilon_0^1, y_{-k_1}) &= (y_0, y_{-k_1}) + \sum_{i_1=1}^{n} a_{i_1}(y_{-i_1}, y_{-k_1}) \\ &= \mathbf{E}y_0 y_{-k_1} + \sum_{i_1=1}^{n} a_{n;i_1}\mathbf{E}y_{-i_1}y_{-k_1} \\ &= h_{0;k_1} + \sum_{i_1=1}^{n} a_{i_1} h_{i_1;k_1} = 0\ ,\quad k = 1, \dots, n \end{aligned}$$

where

$$h_{i_1;k_1} \triangleq \mathbf{E} y_{-i_1} y_{-k_1}$$

indicates (two-indexed) covariance of the random variables y_{-i_1} and y_{-k_1}. Observing that

$$\|\varepsilon_0^1\|^2 = h_{0;0} + \sum_{i_1=1}^{n} a_{i_1} h_{i_1;0}$$

we obtain the following 'normal equations'

$$\begin{bmatrix} h_{0;0} & h_{0;1} & \dots & h_{0;n} \\ h_{1;0} & h_{1;1} & \dots & h_{1;n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n;0} & h_{n;1} & \dots & h_{n;n} \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \|\varepsilon_n\|^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

associated with the linear least-squares prediction problem. We notice that the second-order statistics of the underlying process (actually, its two-indexed covariance matrix)

$$^{1\oplus 1}H = \begin{bmatrix} h_{0;0} & h_{0;1} & \dots & h_{0;n} \\ h_{1;0} & h_{1;1} & \dots & h_{1;n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n;0} & h_{n;1} & \dots & h_{n;n} \end{bmatrix} \tag{2.5}$$

is sufficient in order to state and solve the linear estimation problem (see Levinson [4]).

2.2. Nonlinear least-squares estimation problem

Let us introduce for $m = 2, \dots, M$ the following m-variate sets of nonlinear observations

$$^{m}Y \triangleq [y_{-i_1} \dots y_{-i_m}]_{i_1=1,\dots n;\; i_2=i_1,\dots,n;\; i_m=i_{m-1},\dots,n} \tag{2.6}$$

and notice that (2.6) reduces to ^{1}Y (2.2) if $m = 1$. Let for $m = 2, \dots, M$

$$S^m \triangleq \vee\{^{m}Y\}$$

denote the subspace spanned by the mth degree nonlinear past of the process $\mathbf{y}$. Now introduce the entire Mth degree nonlinear estimation subspace

$$^{M}S = S^1 \dot{+} S^2 \dot{+} \dots \dot{+} S^M$$

so that

$$S \triangleq \vee\{y_0\} \;\dot{+}\; {}^{M}S \tag{2.7}$$

and observe that (2.7) immediately reduces to (2.3) if $M = 1$. This means that the linear estimation problem, corresponding to $M = 1$, is actually the 'simplest' nonlinear estimation problem. Let us introduce the nth order, Mth degree nonlinear estimate

$$\hat{y}_0^M \triangleq P(^{M}S) y_0 \in {}^{M}S$$

and the associated nth order, Mth degree nonlinear prediction error

$$\begin{aligned}\varepsilon_0^M &\triangleq P(S \ominus^M S)y_0 \perp {}^M S \\ &= y_0 + \sum_{m=1}^{M} \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \dots \sum_{i_m=i_{m-1}}^{n} a_{i_1,i_2,\dots,i_m} y_{-i_1} y_{-i_2} \dots y_{-i_m}. \end{aligned} \tag{2.8}$$

Orthogonality (2.8) of the error w.r. to the subspace ${}^M S$ implies the following set of optimality conditions for the nth order, Mth degree nonlinear least-squares estimation problem

$$(\varepsilon_0^M, y_{-k_1} \dots y_{-k_u}) = 0$$

for $u = 1, \dots, M$ and $k_1 = 1, \dots, n;\ k_2 = k_1, \dots, n;\ \dots k_u = k_{u-1}, \dots, n$ or, equivalently,

$$\begin{aligned}(\varepsilon_0^M, y_{-k_1} \dots y_{-k_u}) &= (y_0, y_{-k_1} \dots y_{-k_u}) \\ &\quad + \sum_{m=1}^{M} \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \dots \sum_{i_m=i_{m-1}}^{n} a_{i_1,i_2,\dots,i_m} \ (y_{-i_1} y_{-i_2} \dots y_{-i_m}, y_{-k_1} \dots y_{-k_u}) \\ &= \mathbf{E} y_0 y_{-k_1} \dots y_{-k_u} \\ &\quad + \sum_{m=1}^{M} \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \dots \sum_{i_m=i_{m-1}}^{n} a_{i_1,i_2,\dots,i_m} \ \mathbf{E} y_{-i_1} y_{-i_2} \dots y_{-i_m} y_{-k_1} \dots y_{-k_u} \\ &= h_{0;k_1 k_2 \dots k_u} \\ &\quad + \sum_{m=1}^{M} \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \dots \sum_{i_m=i_{m-1}}^{n} a_{i_1,i_2,\dots,i_m} \ h_{i_1,i_2\dots,i_m;k_1,\dots,k_u}\end{aligned}$$

where

$$h_{i_1,i_2\dots,i_m;k_1,\dots,k_u} \triangleq \mathbf{E} y_{-i_1} y_{-i_2} \dots y_{-i_m} y_{-k_1} \dots y_{-k_u}$$

indicates the $(m \oplus u)$-indexed covariance of the process $\mathbf{y}$. The norm of the Mth degree nonlinear error is then given by

$$\|\varepsilon_0^M\|^2 = h_{0;0} + \sum_{m=1}^{M} \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \dots \sum_{i_m=i_{m-1}}^{n} a_{i_1,i_2,\dots,i_m} \ h_{i_1,i_2\dots,i_m;0}.$$

Let us introduce the following generalized matrices (the notation similar as in [12])

$${}^m A = [a_{i_1,i_2,\dots,i_m}]_{i_1=1,\dots,n;\ i_2=i_1,\dots,n;\ i_m=i_{m-1},\dots,n} \ , \ m=1,\dots,M$$

and

$${}^{\{M\}} A = col[{}^m A]_{m=1,\dots,M}.$$

Moreover, let

$${}^{m \oplus u} H = [h_{i_1,\dots,i_m;k_1,\dots,k_u}]_{i_1=1,\dots,n;\dots\ ;i_m=i_{m-1},\dots,n;k_1=1,\dots,n;\dots\ ;k_u=k_{u-1},\dots,n}$$

and

$${}^{\{M \times M\}} H = [{}^{m \oplus u} H]_{m=1,\dots,M;\ u=1,\dots,M}. \tag{2.9}$$

Then the generalized 'normal equations', associated with the Mth degree nonlinear estimation problem can be compactly expressed as

$$^{\{M\times M\}}H\ ^{\{M\}}A = col[\ ^1P\ ^20\ \ldots\ ^M0] \tag{2.10}$$

where $^1P = [\|\varepsilon_0^M\|^2\ 0\ \ldots\ 0]'$ and $^m0$ is an m-indexed zero-matrix.

From the above it follows that in order to state and solve, via computation of the m-indexed coefficient-matrices ^mA (actually – the multidimensional impulse responses – or the Volterra kernels in the Regular Volterra Functional Polynomials, RVFPs, – of the optimal nonlinear innovations filter [12]), the generalized (block, multi-indexed) covariance matrix (2.9) is sufficient. In other words, for an Mth degree nonlinear estimation problem the $2M$th order statistics of the underlying stochastic process $\mathbf{y}$ are the 'sufficient statistics', much like the second-order statistics are the 'sufficient statistics' in the linear case, which is immediately obtained from (2.10) if $M = 1$.

Remark. The multidimensional impulse responses of the nonlinear optimal filter can efficiently be computed employing the generalized nonlinear Levinson [4] and/or Schur [9] algorithms.

From the above it clearly follows that higher-order statistics of stochastic processes are of crucial importance in nonlinear (Volterra-like) estimation problems. Moreover, efficient measurement/estimation algorithms are required and desired for application and implementation of the nonlinear processing schemes.

3. COVARIANCE MATRICES OF STOCHASTIC PROCESSES

In this section we consider higher-order statistics (covariance matrices) of stochastic processes, as an introduction to efficient measurement/estimation algorithms, presented in Section 4. Starting with second-order statistics, we focus on higher-order statistics, being block, multi-indexed generalized matrices. We present their properties and show consequences of weak stationarity in the higher-order sense.

3.1. Second-order processes

The covariance matrix of a stochastic process (2.5)

$$^{1\oplus 1}H = [h_{i_1;k_1}]_{i_1,k_1=0,\ldots,n} = \begin{bmatrix} h_{0;0} & h_{0;1} & \ldots & h_{0;n} \\ h_{1;0} & h_{1;1} & \ldots & h_{1;n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n;0} & h_{n;1} & \ldots & h_{n;n} \end{bmatrix}$$

is Hermitian (symmetric) as

$$h_{i_1;k_1} = (y_{-i_1}, \bar{y}_{-k_1}) = \mathbf{E}y_{-i_1}\bar{y}_{-k_1} = \mathbf{E}\bar{y}_{-i_1}y_{-k_1} = (\bar{y}_{-i_1}, y_{-k_1}) = \bar{h}_{k_1;i_1}$$

and positive-definite (if the observations form a linearly independent set). It proves to be useful to 'normalize' this matrix as

$$h_{i_1;k_1} \leftarrow \frac{h_{i_1;k_1}}{\sqrt{h_{i_1;i_1}h_{k_1;k_1}}} \tag{3.1}$$

so that

$$h_{i_1;k_1} = 1 \text{ if } i_1 = k_1.$$

3.2. Higher-order processes

If we consider a $2M$th order stochastic process, its higher-order statistics (2.9) are inluded into a generalized (block, multi-indexed) matrix (notation as in [12])

$$^{\{M\times M\}}H = [^{m\oplus u}H]_{m=1,\dots,M;\ u=1,\dots,M} = \begin{bmatrix} ^{1\oplus 1}H & ^{1\oplus 2}H & \dots & ^{1\oplus M}H \\ ^{2\oplus 1}H & ^{2\oplus 2}H & \dots & ^{2\oplus M}H \\ \vdots & \vdots & \ddots & \vdots \\ ^{M\oplus 1}H & ^{M\oplus 2}H & \dots & ^{M\oplus M}H \end{bmatrix}.$$

Let us observe that (2.9) is a Hermitian (symmetric) generalized matrix as we have

$$\begin{aligned} h_{i_1,\dots,i_m;k_1,\dots,k_u} &= (y_{-i_1}\cdots y_{-i_m}, \bar{y}_{-k_1}\cdots \bar{y}_{-k_u}) = \mathbf{E} y_{-i_1}\cdots y_{-i_m}\bar{y}_{-k_1}\cdots \bar{y}_{-k_u} \\ &= \mathbf{E}\bar{y}_{-i_1}\cdots \bar{y}_{-i_m} y_{-k_1}\cdots y_{-k_u} = (\bar{y}_{-i_1}\cdots \bar{y}_{-i_m}, y_{-k_1}\cdots y_{-k_u}) \\ &= h_{k_1,\dots,k_u;i_1,\dots,i_m} \end{aligned}$$

so that (2.9) is a generalized Hermitian, block-Hankel matrix. Generalizing (3.1), we can normalize (2.9) as follows

$$h_{i_1,\dots,i_m;k_1,\dots,k_u} \leftarrow \frac{h_{i_1,\dots,i_m;k_1,\dots,k_u}}{\sqrt{h_{i_1,\dots,i_m;i_1,\dots,i_m} h_{k_1,\dots,k_u;k_1,\dots,k_u}}}$$

so that

$$h_{i_1,\dots,i_n;k_1,\dots,k_n} = 1$$

if $i_p = k_p$ for $p = 1,\dots,n$. Let us observe that each odd-indexed block entry $^{m\oplus u}H$ can be rewritten as a generalized block-row (or block-column) 'flat' matrix while each even-indexed block entry – as a generalized block-square 'flat' matrix, according to some ordering of its rows or columns.

3.3. Nonstationary versus stationary

If the underlying process is wide-sense stationary, its higher-order statistics become generalized Toeplitz matrices. For the second-order statistics $^{1\oplus 1}H$ we obtain in this case

$$h_{i_1;k_1} = \mathbf{E} y_{-i_1}\bar{y}_{-k_1} = \mathbf{E} y_{-i_1+\sigma}\bar{y}_{-k_1+\sigma} \overset{\sigma=i_1}{=} \mathbf{E} y_0 \bar{y}_{-(k_1+i_1)} = h_{k_1-i_1}.$$

For higher-order covariances $^{m\oplus u}H$ we get

$$h_{i_1,\dots,i_m;k_1,\dots,k_u} = h_{i_1+\sigma,\dots,i_m+\sigma;k_1+\sigma,\dots,k_u+\sigma} \overset{\sigma=i_1}{=} h_{i_2-i_1,\dots,i_m-i_1;k_1-i_1,\dots,k_u-i_1}.$$

This means that in the stationary case the generalized covariance matrix $^{\{M\times M\}}H$ becomes a generalized block-Toeplitz matrix (see Figure 1 in Section 4.1). Hence, the measurement (estimation) is much simpler in this case.

4. EFFICIENT ESTIMATION ALGORITHMS OF HIGHER-ORDER STATISTICS

Let us consider in some detail fourth-order statistics of stochastic processes, associated with the 2-nd degree nonlinear least-squares estimation problem. In further considerations we assume that time-series has only real values.

4.1. Properties of fourth-order statistics

The generalized (block, multi-indexed) covariance matrix (2.9) for $M = 2$ is as follows

$$ {}^{\{2\times 2\}}H = \begin{bmatrix} {}^{1\oplus 1}H & {}^{1\oplus 2}H \\ {}^{2\oplus 1}H & {}^{2\oplus 2}H \end{bmatrix} \tag{4.1} $$

where the second-order statistics of the process are given by

$$ {}^{1\oplus 1}H = [h_{i;k}]_{i,k=0,\ldots,n}. $$

The third-order statistics are

$$ {}^{1\oplus 2}H = [h_{i;k,l}]_{i=0,\ldots,n\ ;\ k=0,\ldots,n\ ;\ l=k,\ldots,n} $$

or

$$ {}^{2\oplus 1}H = [h_{i,j;k}]_{i=0,\ldots,n\ ;\ j=i,\ldots,n\ ;\ k=0,\ldots,n} $$

due to (Hermitian) symmetry of (4.1). The fourth-order statistics are then

$$ {}^{2\oplus 2}H = [h_{i,j;k,l}]_{i=0,\ldots,n\ ;\ j=i,\ldots,n\ ;\ k=0,\ldots,n\ ;\ l=k,\ldots,n}. $$

Let us observe that the third-order statistics can be described as a generalized block-row 'flat' matrix

$$ {}^{1\oplus 2}H = [{}^{1\oplus 2}H_k]_{k=0,\ldots,n} $$

or – equivalently – generalized block-column 'flat' matrix

$$ {}^{2\oplus 1}H = col[{}^{2\oplus 1}H_i]_{i=0,\ldots,n}, $$

while the fourth-order statistics as a block-square 'flat' matrix

$$ {}^{2\oplus 2}H = [{}^{2\oplus 2}H_{i;k}]_{i,k=0,\ldots,n}. \tag{4.2} $$

Hence, the 'flat' form of the matrix (4.1) will be

$$ {}^{\{2\times 2\}}H = \begin{bmatrix} {}^{1\oplus 1}H & {}^{1\oplus 2}H_0 & \ldots & {}^{1\oplus 2}H_n \\ {}^{2\oplus 1}H_0 & {}^{2\oplus 2}H_{0,0} & \ldots & {}^{2\oplus 2}H_{0,n} \\ \vdots & \vdots & \ddots & \vdots \\ {}^{2\oplus 1}H_n & {}^{2\oplus 2}H_{n,0} & \ldots & {}^{2\oplus 2}H_{n,n} \end{bmatrix}. $$

Example. If $M = 2$ and $n = 2$, we get

$$
{}^{\{2\times 2\}}H = \begin{bmatrix}
\begin{matrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{matrix} &
\begin{matrix} h_{000} & h_{001} & h_{002} \\ h_{100} & h_{101} & h_{102} \\ h_{200} & h_{201} & h_{202} \end{matrix} &
\begin{matrix} h_{010} & h_{011} & h_{012} \\ h_{110} & h_{111} & h_{112} \\ h_{210} & h_{211} & h_{212} \end{matrix} &
\begin{matrix} h_{020} & h_{021} & h_{022} \\ h_{120} & h_{121} & h_{122} \\ h_{220} & h_{221} & h_{222} \end{matrix} \\ \\
\begin{matrix} h_{000} & h_{001} & h_{002} \\ h_{010} & h_{011} & h_{012} \\ h_{020} & h_{021} & h_{022} \end{matrix} &
\begin{matrix} h_{0000} & h_{0001} & h_{0002} \\ h_{0100} & h_{0101} & h_{0102} \\ h_{0200} & h_{0201} & h_{0202} \end{matrix} &
\begin{matrix} h_{0010} & h_{0011} & h_{0012} \\ h_{0110} & h_{0111} & h_{0112} \\ h_{0210} & h_{0211} & h_{0212} \end{matrix} &
\begin{matrix} h_{0020} & h_{0021} & h_{0022} \\ h_{0120} & h_{0121} & h_{0122} \\ h_{0220} & h_{0221} & h_{0222} \end{matrix} \\ \\
\begin{matrix} h_{100} & h_{101} & h_{102} \\ h_{110} & h_{111} & h_{112} \\ h_{120} & h_{121} & h_{122} \end{matrix} &
\begin{matrix} h_{1000} & h_{1001} & h_{1002} \\ h_{1100} & h_{1101} & h_{1102} \\ h_{1200} & h_{1201} & h_{1202} \end{matrix} &
\begin{matrix} h_{1010} & h_{1011} & h_{1012} \\ h_{1110} & h_{1111} & h_{1112} \\ h_{1210} & h_{1211} & h_{1212} \end{matrix} &
\begin{matrix} h_{1020} & h_{1021} & h_{1022} \\ h_{1120} & h_{1121} & h_{1122} \\ h_{1220} & h_{1221} & h_{1222} \end{matrix} \\ \\
\begin{matrix} h_{200} & h_{201} & h_{202} \\ h_{210} & h_{211} & h_{212} \\ h_{220} & h_{221} & h_{222} \end{matrix} &
\begin{matrix} h_{2000} & h_{2001} & h_{2002} \\ h_{2100} & h_{2101} & h_{2102} \\ h_{2200} & h_{2201} & h_{2202} \end{matrix} &
\begin{matrix} h_{2010} & h_{2011} & h_{2012} \\ h_{2110} & h_{2111} & h_{2112} \\ h_{2210} & h_{2211} & h_{2212} \end{matrix} &
\begin{matrix} h_{2020} & h_{2021} & h_{2022} \\ h_{2120} & h_{2121} & h_{2122} \\ h_{2220} & h_{2221} & h_{2222} \end{matrix}
\end{bmatrix} .
$$

For the 'symmetric domains', corresponding to 2-nd degree RVFPs, we get

$$
{}^{\{2\times 2\}}H_{sym} = \begin{bmatrix}
\begin{matrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{matrix} &
\begin{matrix} h_{000} & h_{001} & h_{002} \\ h_{100} & h_{101} & h_{102} \\ h_{200} & h_{201} & h_{202} \end{matrix} &
\begin{matrix} h_{011} & h_{012} \\ h_{111} & h_{112} \\ h_{211} & h_{212} \end{matrix} &
\begin{matrix} h_{022} \\ h_{122} \\ h_{222} \end{matrix} \\ \\
\begin{matrix} h_{000} & h_{001} & h_{002} \\ h_{010} & h_{011} & h_{012} \\ h_{020} & h_{021} & h_{022} \end{matrix} &
\begin{matrix} h_{0000} & h_{0001} & h_{0002} \\ h_{0100} & h_{0101} & h_{0102} \\ h_{0200} & h_{0201} & h_{0202} \end{matrix} &
\begin{matrix} h_{0011} & h_{0012} \\ h_{0111} & h_{0112} \\ h_{0211} & h_{0212} \end{matrix} &
\begin{matrix} h_{0022} \\ h_{0122} \\ h_{0222} \end{matrix} \\ \\
\begin{matrix} h_{110} & h_{111} & h_{112} \\ h_{120} & h_{121} & h_{122} \end{matrix} &
\begin{matrix} h_{1100} & h_{1101} & h_{1102} \\ h_{1200} & h_{1201} & h_{1202} \end{matrix} &
\begin{matrix} h_{1111} & h_{1112} \\ h_{1211} & h_{1212} \end{matrix} &
\begin{matrix} h_{1122} \\ h_{1222} \end{matrix} \\ \\
\begin{matrix} h_{220} & h_{221} & h_{222} \end{matrix} &
\begin{matrix} h_{2200} & h_{2201} & h_{2202} \end{matrix} &
\begin{matrix} h_{2211} & h_{2212} \end{matrix} &
\begin{matrix} h_{2222} \end{matrix}
\end{bmatrix} . \tag{4.3}
$$

Due to symmetry we have

$$
\begin{aligned}
h_{i;k} &= h_{k;i} \\
h_{i;k,l} &= h_{k,l;i} \\
h_{i,j;k} &= h_{k;i,j} \\
h_{i,j;k,l} &= h_{k,l;i,j}
\end{aligned} \tag{4.4}
$$

so (4.3) became

$$
{}^{\{2\times 2\}}H_{sym} = \left[\begin{array}{ccc ccc cc c}
h_{00} & h_{01} & h_{02} & h_{000} & h_{001} & h_{002} & h_{011} & h_{012} & h_{022} \\
h_{10} & h_{11} & h_{12} & h_{100} & h_{101} & h_{102} & h_{111} & h_{112} & h_{122} \\
h_{20} & h_{12} & h_{22} & h_{200} & h_{201} & h_{202} & h_{211} & h_{212} & h_{222} \\[2ex]
h_{000} & h_{100} & h_{200} & h_{0000} & h_{0001} & h_{0002} & h_{0011} & h_{0012} & h_{0022} \\
h_{001} & h_{101} & h_{201} & h_{0001} & h_{0101} & h_{0102} & h_{0111} & h_{0112} & h_{0122} \\
h_{002} & h_{102} & h_{202} & h_{0002} & h_{0102} & h_{0202} & h_{0211} & h_{0212} & h_{0222} \\[2ex]
h_{011} & h_{111} & h_{211} & h_{0011} & h_{0111} & h_{0211} & h_{1111} & h_{1112} & h_{1122} \\
h_{012} & h_{112} & h_{212} & h_{0012} & h_{0112} & h_{0212} & h_{1112} & h_{1212} & h_{1222} \\[2ex]
h_{022} & h_{122} & h_{222} & h_{0022} & h_{0122} & h_{0222} & h_{1122} & h_{1222} & h_{2222}
\end{array}\right]. \tag{4.5}
$$

If the process is stationary (in a weak fourth-order sense), we obtain (for index i – the same properties are valid for other indices)

$$
\begin{aligned}
h_{i;k} &= h_{k-i} \\
h_{i;k,l} &= h_{k-i,l-i} \\
h_{i,j;k} &= h_{j-i,k-i} \\
h_{i,j;k,l} &= h_{j-i,k-i,l-i}
\end{aligned} \tag{4.6}
$$

so that (4.1) becomes a block-Toeplitz multi-indexed matrix; i. e.,

$$
{}^{\{2\times 2\}}H_{st} = \left[\begin{array}{ccc cccccc}
h_0 & h_1 & h_2 & h_{0,0} & h_{0,1} & h_{0,2} & h_{1,1} & h_{1,2} & h_{2,2} \\
h_1 & h_0 & h_1 & h_{1,0} & h_{1,1} & h_{1,2} & h_{0,0} & h_{0,1} & h_{1,1} \\
h_2 & h_1 & h_0 & h_{2,0} & h_{2,1} & h_{2,2} & h_{1,0} & h_{1,1} & h_{0,0} \\[2ex]
h_{0,0} & h_{1,0} & h_{2,0} & h_{0,0,0} & h_{0,0,1} & h_{0,0,2} & h_{0,1,1} & h_{0,1,2} & h_{0,2,2} \\
h_{0,1} & h_{1,1} & h_{2,1} & h_{0,0,1} & h_{1,0,1} & h_{1,0,2} & h_{1,1,1} & h_{1,1,2} & h_{1,2,2} \\
h_{0,2} & h_{1,2} & h_{2,2} & h_{0,0,2} & h_{1,0,2} & h_{2,0,2} & h_{2,1,1} & h_{2,1,2} & h_{2,2,2} \\[2ex]
h_{1,1} & h_{0,0} & h_{1,0} & h_{0,1,1} & h_{1,1,1} & h_{2,1,1} & h_{0,0,0} & h_{0,0,1} & h_{0,1,1} \\
h_{1,2} & h_{0,1} & h_{1,1} & h_{0,1,2} & h_{1,1,2} & h_{2,1,2} & h_{0,0,1} & h_{1,0,1} & h_{1,1,1} \\[2ex]
h_{2,2} & h_{1,1} & h_{0,0} & h_{0,2,2} & h_{1,2,2} & h_{2,2,2} & h_{0,1,1} & h_{1,1,1} & h_{0,0,0}
\end{array}\right].
$$

In Figure 1, we present example of block-Toeplitz fourth-order covariance matrix ${}^{2\oplus 2}H$. The matrix was obtained according to eq. (4.2) for $n = 2$, after simplifications which came from symmetry (4.4) and stationarity (4.6).

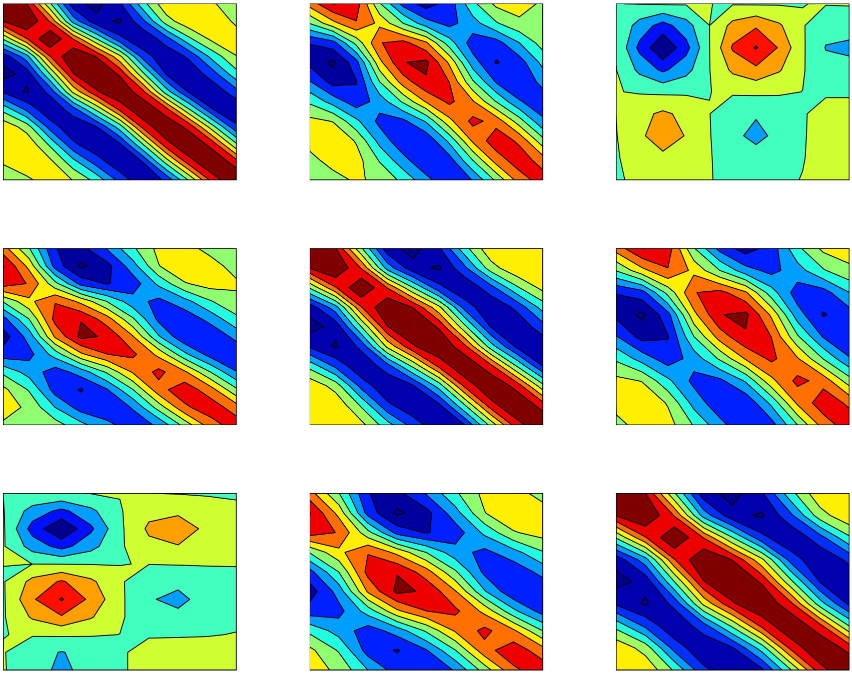

Fig. 1. Example of block-Toeplitz fourth-order covariance matrix ${}^{2\oplus 2}H$ of narrow band time-series.

Gaussian case. If the process was Gaussian, we would have [11]

$$h_{i_1,\ldots i_m;k_1,\ldots,k_u} = \begin{cases} \Sigma\Pi h_{i_r;k_s} & \text{if} \quad m+u \quad \text{is} \quad \text{even} \\ 0 & \text{if} \quad m+u \quad \text{is} \quad \text{odd} \end{cases} \tag{4.7}$$

where the symbol $\Sigma\Pi$ stands for the summation over all distinct ways of partitioning the $m+u$ random variables into products of averages of their pairs. This means that in the Gaussian case, each higher-order covariance of even-order can be expressed in terms of the second-order covariance values while all odd-order covariances vanish. Therefore, it is sufficient to treat a Gaussian stochastic sequence as the second-order sequence (as its higher-order statistics do not contain any 'new' statistical information). If the process was white Gaussian, (4.7) would become

$$h_{i_1,\ldots i_m;k_1,\ldots,k_u} = \begin{cases} \Sigma\Pi \delta_{i_r;k_s} & \text{if} \quad m+u \quad \text{is} \quad \text{even} \\ 0 & \text{if} \quad m+u \quad \text{is} \quad \text{odd} \end{cases} \tag{4.8}$$

where $\delta_{i_r;k_s}$ denotes the Kronecker delta (see Figure 2).

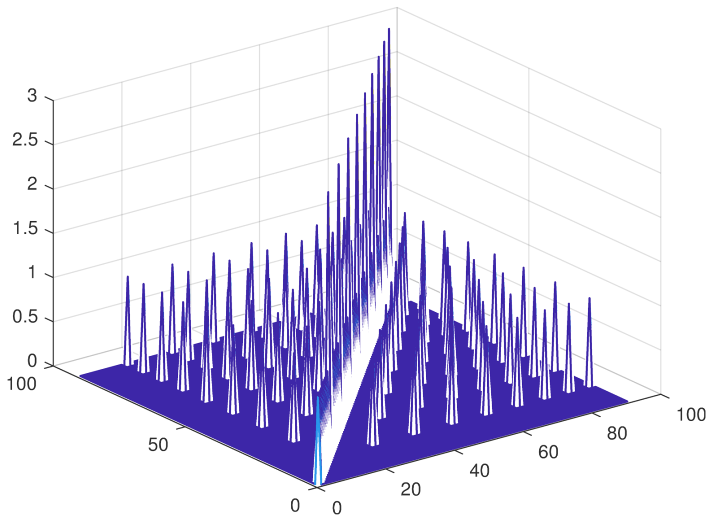

Fig. 2. Multi-indexed covariance matrix ${}^{\{2\times 2\}}H$ of Gaussian white noise.

4.2. Isomorphism of the space of nonlinear observations and the space of sample-product coefficient-vectors

Consider a set of samples of a fourth-order time-series

$$\{\mathsf{y}_0, \ldots, \mathsf{y}_T\}$$

and employ the $\langle bra|ket\rangle$ notation, following [2]. Define a (column) ket-vector

$$|\mathsf{y}\rangle_T = [\mathsf{y}_0, \ldots, \mathsf{y}_T]' \tag{4.9}$$

where $'$ stands for transpose. According to Kolmogorov isomorphism [1], we have

$$\begin{aligned} y_{-i} &\leftrightarrow |z^i\mathsf{y}\rangle_T \\ y_{-i}y_{-j} &\leftrightarrow |z^i\mathsf{y}\cdot z^j\mathsf{y}\rangle_T \end{aligned}$$

where the delay operator is defined as

$$|z^i\mathsf{y}\rangle_T = [\underbrace{0\ldots0}_{i}\ \mathsf{y}_0, \ldots, \mathsf{y}_{T-i}]'.$$

Then we can introduce the entries of the Gram-matrix, being the estimates of higher-order covariances of the process $\mathbf{y}$, as

$$\begin{aligned} \hat{h}_{i;k} &= \langle z^i\mathsf{y}|z^k\mathsf{y}\rangle_T \\ \hat{h}_{i;k,l} &= \langle z^i\mathsf{y}|z^k\mathsf{y}\cdot z^l\mathsf{y}\rangle_T \\ \hat{h}_{i,j;k} &= \langle z^i\mathsf{y}\cdot z^j\mathsf{y}|z^k\mathsf{y}\rangle_T \\ \hat{h}_{i,j;k,l} &= \langle z^i\mathsf{y}\cdot z^j\mathsf{y}|z^k\mathsf{y}\cdot z^l\mathsf{y}\rangle_T \end{aligned} \tag{4.10}$$

We assume that the process **y** is ergodic. That is why the length T of a sample vector (4.9) should be long enough.

5. IMPLEMENTATION OF HIGHER-ORDER STATISTICS ESTIMATION

5.1. Algorithm for fourth-order covariance matrices

Below we present the Matlab implementation of the proposed earlier algorithms for measurement/estimation algorithms for second-, third- and fourth-order covariance matrices estimates. The function *cov_reduced* returns the estimate of the multi-indexed covariance matrix $^{\{2\times 2\}}H$, in accordance with (4.3).

```
function [H] = cov_reduced(n, y)
% The function returns 4-order covariance matrix H
H11=zeros(n+1,n+1);
for i=1:n+1
    for k=1:n+1
        H11(i,k)=h11(y,i-1,k-1);
    end
end

H12=zeros(n+1,(n+1)*(n+2)/2);
for i=1:n+1
    ind=1;
    for k=1:n+1
        for l=k:n+1
            H12(i,ind)=h12(y,i-1,k-1, l-1);
            ind=ind+1;
        end
    end
end

H21=H12';

H22=zeros((n+1)*(n+2)/2,(n+1)*(n+2)/2);
ind1=0;
for i=1:n+1
    for j=i:n+1
        ind1=ind1+1;
        ind2=1;
        for k=1:n+1
            for l=k:n+1
                H22(ind1,ind2)=h22(y,i-1,j-1,k-1, l-1);
                ind2=ind2+1;
            end
        end
    end
end
% 4-order covariance matrix
H=[[H11, H12]; [H21, H22]];
end
```

5.2. Simulations results

We present simulation results for higher-order time-series. In Figures 3, 4 and 5 we show the results for pseudo-random realizations (Matlab *randn* source) with band-pass spectral density obtained from band-pass filtering of a 'white' time-series of full range PSD (with the normalized frequencies from $0.00-1.00$). For example, time-series used in Figure 4 was filtered by a FIR pass-band filter with the lower cut-off frequencies 0.1 and 0.2, and the upper cut-off frequencies 0.3 and 0.4. In Figure 6 we used pseudo-random narrow-band time-series $|y\rangle_T$ from Figure 3, but squared (i. e. $|y \cdot y\rangle_T$). Sine-type time-series were illustrated in Figures 7 and 8. The 'length' of time-series was $T = 1000$ samples while the dimension of second, third and fourth-order covariance matrix was set to $n = 10$.

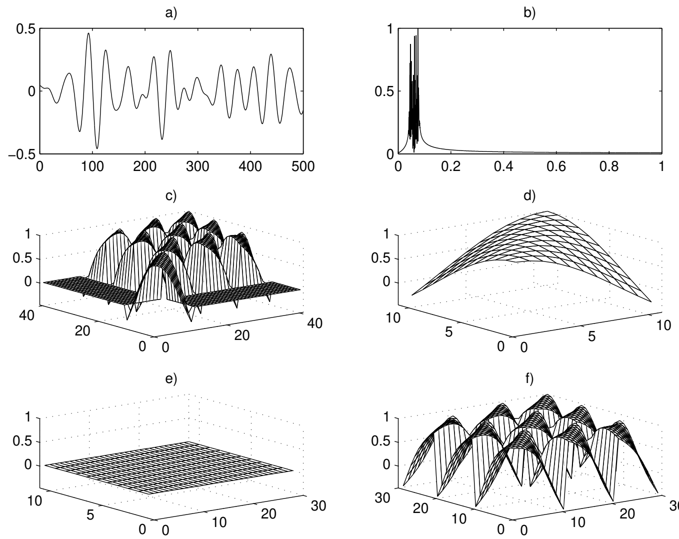

Fig. 3. Time-series $|y\rangle_T$ with frequency band $(0.03, 0.09)$: a) trajectory, b) power spectral density, c) ${}^{\{2\times2\}}H$ matrix, d) ${}^{1\oplus1}H$ matrix, e) ${}^{1\oplus2}H$ matrix, f) ${}^{2\oplus2}H$ matrix.

In the Gaussian case, presented in Figures 3e), 4e), 5e), third-order statistics ${}^{1\oplus2}H$ and ${}^{2\oplus1}H$ disappears – see eq. (4.8). Narrowing the band we obtain stronger correlation – compare Figure 4 with Figure 3. The band shift to from lower-frequencies (Figure 3) to higher-frequencies (Figure 5) causes weaker correlation. Taking time-series of squared samples produces non-zero third-order statistics – see Figures 3e) and 6e).

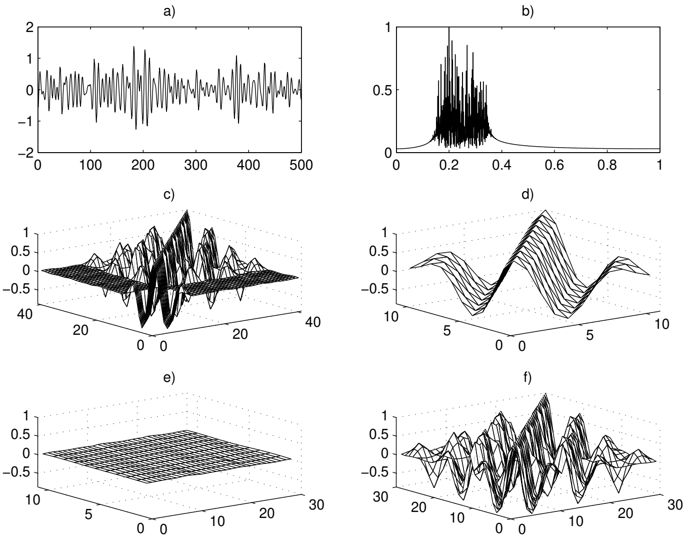

Fig. 4. Time-series $|y\rangle_T$ with frequency band $(0.1, 0.4)$: a) trajectory, b) power spectral density, c) ${}^{\{2\times 2\}}H$ matrix, d) ${}^{1\oplus 1}H$ matrix, e) ${}^{1\oplus 2}H$ matrix, f) ${}^{2\oplus 2}H$ matrix.

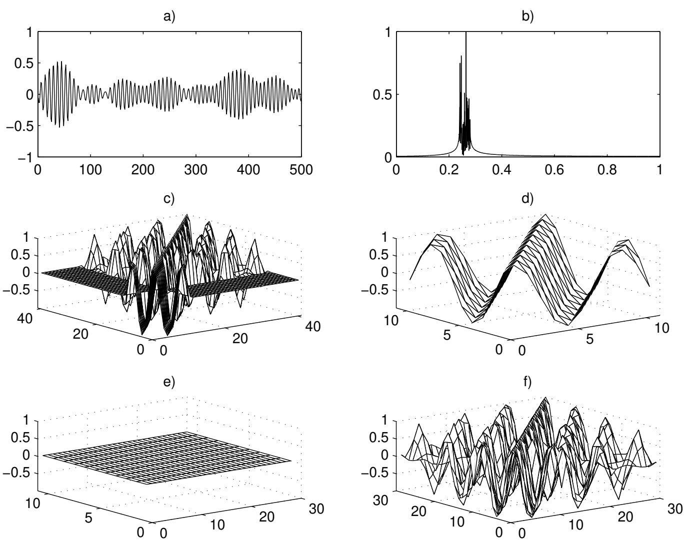

Fig. 5. Time-series $|y\rangle_T$ with frequency band $(0.23, 0.29)$: a) trajectory, b) power spectral density, c) ${}^{\{2\times 2\}}H$ matrix, d) ${}^{1\oplus 1}H$ matrix, e) ${}^{1\oplus 2}H$ matrix, f) ${}^{2\oplus 2}H$ matrix.

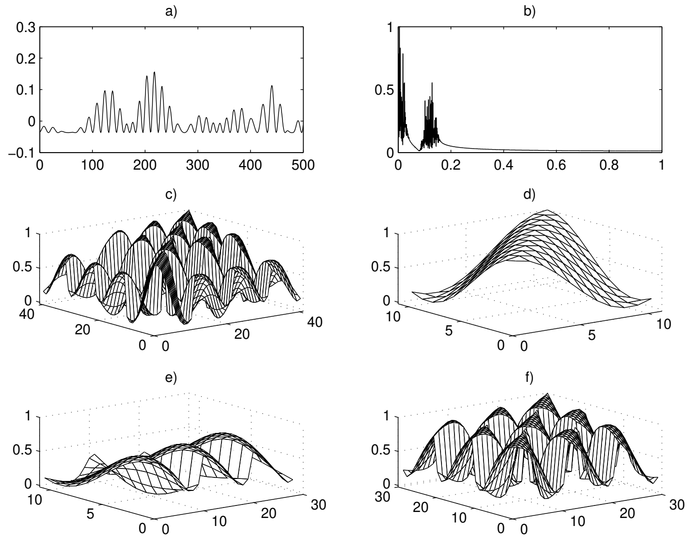

Fig. 6. Time-series $|y \cdot y\rangle_T$ (with frequency band $(0.03, 0.09)$): a) trajectory, b) power spectral density, c) ${}^{\{2\times2\}}H$ matrix, d) ${}^{1\oplus1}H$ matrix, e) ${}^{1\oplus2}H$ matrix, f) ${}^{2\oplus2}H$ matrix.

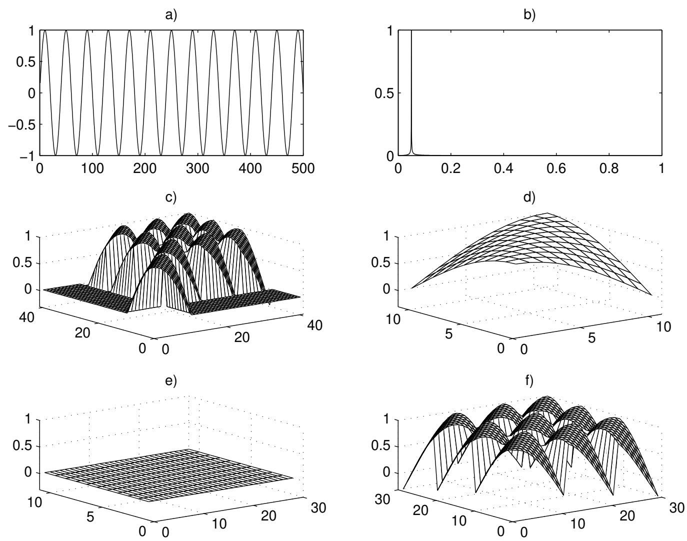

Fig. 7. Time-series $|y\rangle_T = \sin(2\pi 0.05t)$: a) trajectory, b) power spectral density, c) ${}^{\{2\times2\}}H$ matrix, d) ${}^{1\oplus1}H$ matrix, e) ${}^{1\oplus2}H$ matrix, f) ${}^{2\oplus2}H$ matrix.

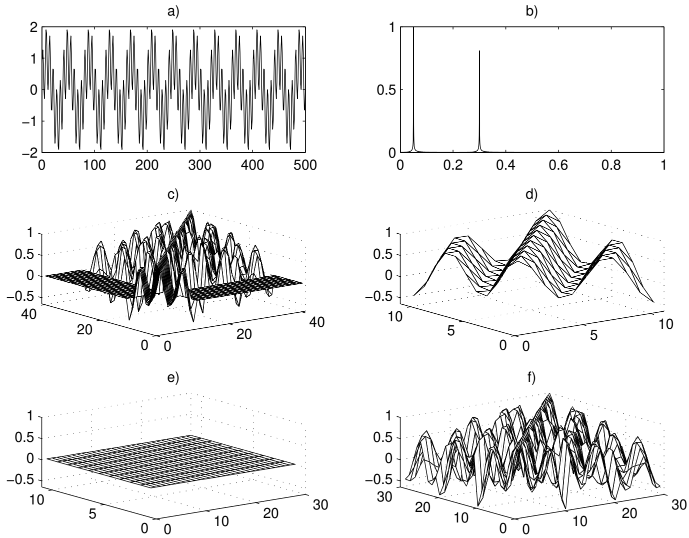

Fig. 8. Time-series $|y\rangle_T = \sin(2\pi 0.05t) + \sin(2\pi 0.3t)$: a) trajectory, b) power spectral density, c) ${}^{\{2\times 2\}}H$ matrix, d) ${}^{1\oplus 1}H$ matrix, e) ${}^{1\oplus 2}H$ matrix, f) ${}^{2\oplus 2}H$ matrix.

6. ADAPTIVE ESTIMATION

In the Section, we present adaptive estimation algorithm, used to find minimal time-series length T, necessary for proper representation of time-series statistics. Another approach, for stationary case, can be found in [10]. At the end of the Section, we show simulation results.

6.1. Algorithm

Let us rewrite the estimates (4.10) in a way indicating that the estimates employ the time-series of the 'length' T which may be from now on treated as a 'current time'; i. e., $T = 1, 2, \ldots$

$$
\begin{aligned}
\hat{h}^T_{i;k} &= < z^i \mathsf{y} | z^j \mathsf{y} \rangle_T \\
\hat{h}^T_{i;k,l} &= < z^i \mathsf{y} | z^k \mathsf{y} \cdot z^l \mathsf{y}_T \\
\hat{h}^T_{i,j;k} &= < z^i \mathsf{y} \cdot z^j \mathsf{y} | z^k \mathsf{y} \rangle_T \\
\hat{h}^T_{i,j;k,l} &= < z^i \mathsf{y} \cdot z^j \mathsf{y} | z^k \mathsf{y} \cdot z^l \mathsf{y} \rangle_T
\end{aligned}
$$

This gives the estimates of the covariance matrices

$$\begin{aligned} {}^{1\oplus 1}H^T &= [\hat{h}^T_{i;k}]_{i,k=0,\ldots,n} \\ {}^{1\oplus 2}H^T &= [\hat{h}^T_{i;k,l}]_{i=0,\ldots,n\ ;\ k=0,\ldots,n\ ;\ l=k,\ldots,n} \\ {}^{2\oplus 1}H^T &= [\hat{h}^T_{i,j;k}]_{i=0,\ldots,n\ ;\ j=i,\ldots,n\ ;\ k=0,\ldots,n} \\ {}^{2\oplus 2}H^T &= [\hat{h}^T_{i,j;k,l}]_{i=0,\ldots,n\ ;\ j=i,\ldots,n\ ;\ k=0,\ldots,n\ ;\ l=k,\ldots,n} \end{aligned}$$

yielding the overall higher-order statistics estimate (actually the generalized Gram matrix)

$${}^{\{2\times 2\}}H^T = \begin{bmatrix} {}^{1\oplus 1}H^T & {}^{1\oplus 2}H^T \\ {}^{2\oplus 1}H^T & {}^{2\oplus 2}H^T \end{bmatrix}.$$

Computational complexity of the higher-order statistics estimation crucially depends on the 'length' T of the time-series realization. When the estimation procedure is started, we can hardly predict which value of the parameter T should be chosen to obtain the resulting estimates of satisfactory accuracy. We can propose, instead, an adaptive estimation algorithm, based on the observation that

$$\langle x|y\rangle_{T+1} = \sum_{t=0}^{T+1} x_t y_t = \langle x|y\rangle_T + x_{T+1}y_{T+1}$$

following from (2.28). Hence,

$$\begin{aligned} \hat{h}^{T+1}_{i;k} &= \hat{h}^T_{i;k} + \Delta^{T+1}_{i;k} \\ \hat{h}^{T+1}_{i;k,l} &= \hat{h}^T_{i;k,l} + \Delta^{T+1}_{i;k,l} \\ \hat{h}^{T+1}_{i,j;k} &= \hat{h}^T_{i,j;k} + \Delta^{T+1}_{i,j;k} \\ \hat{h}^{T+1}_{i,j;k,l} &= \hat{h}^T_{i,j;k,l} + \Delta^{T+1}_{i,j;k,l} \end{aligned}$$

and consequently,

$$\begin{aligned} {}^{1\oplus 1}H^{T+1} &= {}^{1\oplus 1}H^T + {}^{1\oplus 1}\Delta^{T+1} \\ {}^{1\oplus 2}H^{T+1} &= {}^{1\oplus 2}H^T + {}^{1\oplus 2}\Delta^{T+1} \\ {}^{2\oplus 1}H^{T+1} &= {}^{2\oplus 1}H^T + {}^{2\oplus 1}\Delta^{T+1} \\ {}^{2\oplus 2}H^{T+1} &= {}^{2\oplus 2}H^T + {}^{2\oplus 2}\Delta^{T+1} \end{aligned}$$

resulting in

$${}^{\{2\times 2\}}H^{T+1} = {}^{\{2\times 2\}}H^T + {}^{2\times 2}\Delta^{T+1}.$$

Hence, at each step, we can use a criterion based – for example – on the Frobenius distance

$$\frac{\|{}^{\{2\times 2\}}H^{T+1} - {}^{\{2\times 2\}}H^T\|_{Frobenius}}{\|{}^{\{2\times 2\}}H^T\|_{Frobenius}} < \delta \tag{6.1}$$

and decide to terminate the estimation algorithm for a given value of the parameter δ. The Frobenius norm for a matrix $A = [a_{ij}]_{i,j}$ was calculated in the following way

$$\|A\|_{Frobenius} = \sqrt{\sum_i \sum_j |a_{ij}|^2}. \tag{6.2}$$

We chose the Frobenius distance, because it can be interpreted as least squares error for matrices. It gives us possibility to stop the adaptive procedure for time-series length T. This means that we can find minimal value of parameter T, instead of the larger, fixed and presumed value in 'classical' procedure.

The Matlab implementation of the adaptive algorithm implementation is presented below.

```
1  function [h, frobnorm] = cov_adaptive(u,n,wx1,wx2,wy1,wy2,threshold)
2  % The function returns adaptive covariance matrix with given order {(wx1
       ,...,wx2) x (wy1,...,wy2)} for given delta
3      h_old = [];
4      T = size(u);
5      for t = 1 : T %
6          h = [];
7          for w = wx1 : wx2
8              hy = [];
9              for k = wy1 : wy2
10                 hy = [hy cov_adaptive_update(u(1:t),n,w,k)];
11             end
12             h = [h; hy];
13         end
14         frobnorm(t) = frob(h-h_old)/frob(h_old);
15         if frobnorm(t) < threshold
16             return
17         end
18         h_old = h;
19     end
20 end
21
22 function h = cov_adaptive_update(u,n,wx,wy)
23 % The function returns covariance matrix of update values for time t+1 w.r
       .t t with given order {wx x wy}
24
25 function y = frob(h)
26 % The function returns frobenius norm of matrix h
27     y = 0;
28     for w = 1 : size(h,1)
29         for k = 1 : size(h,2)
30             y = y + h(w,k)^2;
31         end
32     end
33     y = sqrt(y);
34 end
```

6.2. Simulations results

The Frobenius norm can be used as a stop criterion for the procedure listed above in *cov_adaptive* Matlab function (line 15). In the performed simulations we take three threshold values of $\delta = 0.5,\ 0.15,\ 0.01$. Lower δ value provides better estimation, due to lower relative difference between correlation matrices obtained for T and $T+1$.

The Frobenius distance calculated on each step (for each $T = 1, 2, \ldots$) according to equation (6.1) is on left-hand side of Figures 9, 10, 11 and 12. The corresponding generalized covariance matrices are placed on right-hand side.

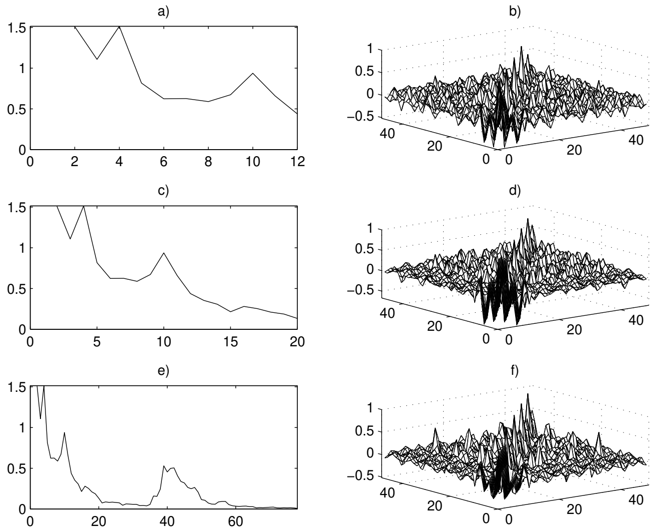

Fig. 9. Narrow-band time-series $|y\rangle_T$ with frequency-band $(0.1, 0.6)$. Adaptive estimation of the ${}^{\{2\times2\}}H$ matrix (right side) and Frobenius-norm (left side) for $\delta = 0.5$ a)-b), $\delta = 0.15$ c)-d), $\delta = 0.01$ e)-f).

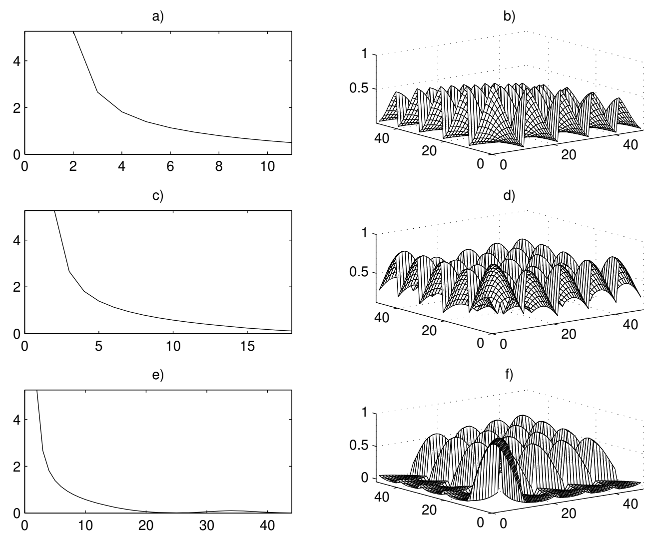

Fig. 10. Time-series $|y\rangle_T = \sin(2\pi 0.05t)$. Adaptive estimation of the ${}^{\{2\times2\}}H$ matrix (right side) and Frobenius-norm (left side) for $\delta = 0.5$ a)-b), $\delta = 0.15$ c)-d), $\delta = 0.01$ e)-f).

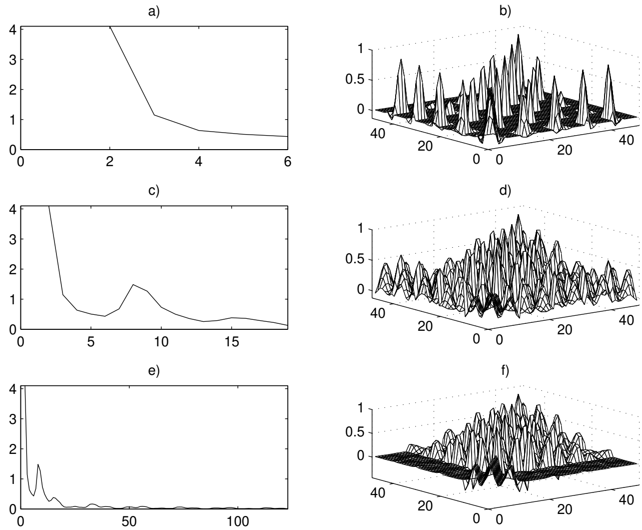

Fig. 11. Time-series $|y\rangle_T = \sin(2\pi 0.05t) + \sin(2\pi 0.3t)$. Adaptive estimation of the ${}^{\{2\times 2\}}H$ matrix (right side) and Frobenius-norm (left side) for $\delta = 0.5$ a)-b), $\delta = 0.15$ c)-d), $\delta = 0.01$ e)-f).

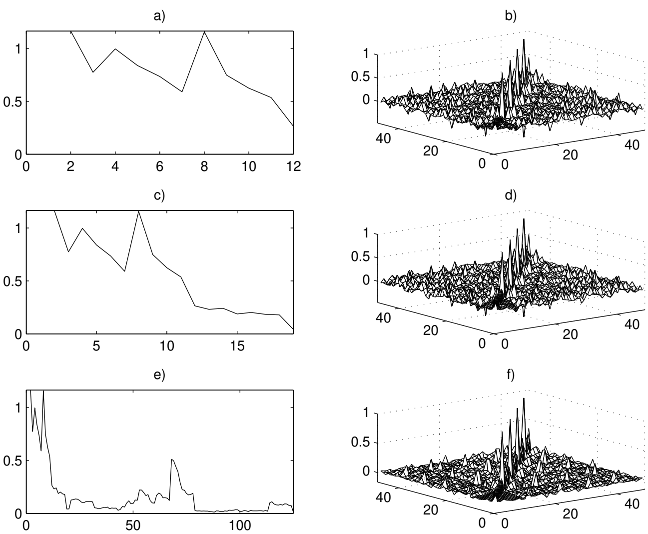

Fig. 12. Time-series $|y\rangle_T = randn(1, T)$. Adaptive estimation of the ${}^{\{2\times 2\}}H$ matrix (right side) and Frobenius-norm (left side) for $\delta = 0.5$ a)-b), $\delta = 0.15$ c)-d), $\delta = 0.01$ e)-f).

In Figure 11 for sum of two sines, we see that for the value of $\delta = 0.5$ the length of time-series obtained from adaptive procedure was very short $T = 6$ (see Figure 11a). In this case covariance estimation (see Figure 11b) did not reproduce the real covariance matrix, presented in Figure 8c (for 'classical' procedure and $T = 1000$). For much smaller $\delta = 0.01$, the covariance estimation was performed with better results, shown in Figure 11f. In that case, time-series length was $T = 120$ (see Figure 11f), which is longer than T taken for $\delta = 0.5$, but much smaller than $T = 1000$. Similar observations can be made for simpler example – one sine. Compare Figure 10 to Figure 7.

For Gaussian white noise, the estimated covariance matrix for $\delta = 0.01$, shown in Figure 12f, is quite similar to the theoretical covariance matrix, illustrated in Figure 2. We observe peaks on diagonal and smaller peaks for fourth-order statistics, however there are still some non-zero values for third-order statistics.

We can observe that there is a global decreasing trend for Frobenius-norm (left-hand side in Figures $9-12$), but there are some local fluctuations – the decrease is not monotonic.

7. FINAL REMARKS

At the beginning, we defined linear and nonlinear least-squares estimation problem using second- and higher-order statistics, respectively eq. (2.5) and (2.9). We defined generalized multi-indexed matrix ${}^{\{M\times M\}}H$ and we used it to analyse time-series features such as stationarity (block-Toeplitz for fourth-order statistics), gaussianity (zeroed third-order statistics), frequency band width and shift. We also verified the convergence speed of proposed method, with respect to relative Frobenius norm (6.1).

We proposed new efficient algorithm for estimation of higher-order statistics – see Sec. 6. Higher-order statistics allow us to perform a deeper analysis of time-series.

Our future research will focus on applying fourth-order statistics in nonlinear least-squares estimation problem. Also problem of determining proper value of δ for a given time-series is not clear and should be investigated in the future.

(Received August 18, 2017)

REFERENCES

[1] P. Dewilde: Stochastic modelling with orthogonal filters. In: Outils et modeles mathematiques pour l'automatique, l'analyse de systemes et le traitement du signal, CNRS (ed.), Paris 1982, pp. 331–398.

[2] D. T. L. Lee, M. Morf, and B. Friedlander: Recursive least-squares ladder estimation algorithms. IEEE Trans. Circuit Systems CAS *28* (1981), 467–481. DOI:10.1109/tcs.1981.1085020

[3] J. Jurečková: Regression quantiles and trimmed least squares estimator under a general design. Kybernetika *20* (1984), 5, 345–357.

[4] N. Levinson: The Wiener RMS error criterion in filter design and prediction. J. Math. Physics *25* (1947), 261–278. DOI:10.1002/sapm1946251261

[5] P. Mandl, T. E. Duncan, and B. Pasik-Duncan: On the consistency of a least squares identification procedure. Kybernetika *24* (1988), 5, 340–346.

[6] J. M. Mendel: Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. Proc. IEEE *79* (1991), 3, 278–305. DOI:10.1109/5.75086

[7] A. Pázman: Probability distribution of the multivariate nonlinear least squares estimates. Kybernetika *20* (1984), 3, 209–230.

[8] L. Pronzato and A. Pázman: Second-order approximation of the entropy in nonlinear least-squares estimation. Kybernetika *30* (1994), 2, 187–198.

[9] I. Schur: Methods in 0perator Theory and Signal Processing. Operator Theory: Advances and Applications *18*, Springer-Verlag 1086. DOI:10.1007/978-3-0348-5483-2

[10] H. M. Stellakis and E. M. Manolakos: Adaptive computation of higher order moments and its systolic realization. Int. J. Adaptive Control Signal Process. *10* (1996), 283–302. DOI:10.1002/(sici)1099-1115(199603)10:2/3¡283::aid-acs351¿3.3.co;2-2

[11] N. Wiener: Nonlinear Problems in Random Theory. MIT Press, 1958.

[12] J. Zarzycki: Multidimensional nonlinear Schur parametrization of non-gaussian stochastic signals — Part one: Statement of the problem. MDSSP J. *15* (2004), 3, 217–241. DOI:10.1023/b:mult.0000028007.05748.48

Wladyslaw Magiera, Wroclaw University of Science and Technology. Poland.
e-mail: wladyslaw.magiera@pwr.edu.pl

Urszula Libal, Wroclaw University of Science and Technology. Poland.
e-mail: urszula.libal@pwr.edu.pl

Agnieszka Wielgus, Wroclaw University of Science and Technology. Poland.
e-mail: agnieszka.wielgus@pwr.edu.pl