Vector Time-Frequency AR Models for Nonstationary Multivariate Random Processes

Michael Jachan, Gerald Matz, *Senior Member, IEEE*, and Franz Hlawatsch, *Senior Member, IEEE*

***Abstract*—We introduce the *vector time-frequency autoregressive (VTFAR) model* for a parsimonious parametric description of nonstationary vector random processes. The VTFAR model generalizes the recently proposed scalar TFAR model to the multivariate case. It is physically meaningful because nonstationarity and spectral correlation are represented in terms of frequency shifts, and it is parsimonious for the practically relevant class of underspread vector processes (i.e., nonstationary vector processes with rapidly decaying correlation in time and frequency). For vector processes with decaying correlation across the signals, we introduce a variant of the VTFAR model with banded parameter matrices. Furthermore, we present a VTFAR parameter estimator that is based on a system of linear equations with two-level block-Toeplitz structure, and we develop an efficient order-recursive algorithm for solving these equations. We also present information criteria for estimating the VTFAR model order and the matrix bandwidth of the banded VTFAR model. The performance of the proposed VTFAR parameter and order estimators is assessed through numerical simulations. Finally, an application to nonstationary multivariate spectral analysis is presented.**

***Index Terms*—Nonstationary multivariate random processes, nonstationary spectral estimation, order estimation, parametric modeling, time-frequency analysis, time-varying AR models, time-varying systems, vector processes, Yule–Walker equations.**

I. Introduction

RANDOM processes occurring in applications like mobile communications, machine monitoring, biomedical signal processing, geophysics, etc., are often most suitably described by parametric models [1], [2]. Usually, such models involve a parametric representation of an innovations system driven by white innovations noise. The statistics of the output process are then characterized by the parameters of the innovations system.

In this paper, we develop parametric models for *nonstationary vector (multivariate) processes*. Our models can be applied to the analysis of, e.g., the impulse response of a mobile radio channel [3], multichannel EEG data [4], and geophysical sensor data [5]. The ideas and methods we present generalize the time-frequency autoregressive (TFAR) model for nonstationary *scalar* (univariate) processes and corresponding parameter estimators [6] to the vector/multivariate case. Our contributions can be summarized as follows.

- We introduce the vector time-frequency autoregressive (VTFAR) model as a highly parsimonious parametric model for underspread nonstationary vector processes (i.e., nonstationary vector processes with rapidly decaying correlation in time and frequency). This model involves a time-varying multiple-input multiple-output (MIMO) innovations system that is underspread, i.e., it effectively produces only limited time shifts (delays) and frequency (Doppler) shifts.
- For vector processes with decaying correlation across the component signals, we introduce an even more parsimonious variant of the VTFAR model that uses banded parameter matrices.
- We exploit the underspread property to obtain simple approximate time-frequency (TF) representations of the innovations system and of the second-order statistics of the VTFAR process (corresponding to a "time-varying VTFAR spectrum"), thus avoiding expensive and potentially unstable MIMO operator inversions.
- We propose an estimator of the VTFAR parameters that constitutes a TF extension of the multichannel Yule-Walker method [2]. We also propose an approximation to this estimator that is based on the underspread assumption. Exploiting the two-level block-Toeplitz structure of the resulting system of equations, we derive an efficient solution algorithm that extends the Wax–Kailath algorithm [7] to the vector case. Furthermore, we present model order estimators that are based on classical information criteria.
- We assess the performance of our estimation methods via numerical simulations, and we demonstrate the application of the VTFAR model and estimators to nonstationary multivariate spectral analysis.

This paper is organized as follows. In Section II, the VTFAR model and a variant involving banded parameter matrices are introduced. In Section III, the innovations system of a VTFAR process as well as TF representations of the innovations system and of the second-order statistics are discussed, and simplifying approximations valid in the underspread case are developed. Parameter estimators based on TF extensions of the multichannel Yule-Walker equations are derived in Section IV, and order estimators are provided in Section V. Section VI presents numerical results of the proposed estimators and demonstrates their application to nonstationary multivariate spectral analysis. Finally, for efficient implementation of the "underspread" VTFAR

Manuscript received April 07, 2008; accepted March 07, 2009. First published July 14, 2009; current version published November 18, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arie Yeredor. This work was supported by the research grants "Statistical Inference" (S10603-N13) and "Information Networks" (S10606-N13) within the National Research Network SISE funded by the Austrian Science Fund (FWF). This work in part was previously presented at the IEEE SPAWC, New York, June 2005.

M. Jachan is with Brain Products GmbH, D-79098 Freiburg i. Br., Germany (e-mail: michael.jachan@gmx.net).

G. Matz and F. Hlawatsch are with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, A-1040 Vienna, Austria.

Digital Object Identifier 10.1109/TSP.2009.2026600

parameter estimator, a novel multichannel version of the Wax-Kailath algorithm is developed in an Appendix.

II. The VTFAR Model

In this section, we present the VTFAR model and a variant with banded parameter matrices. By way of introduction and motivation, we start with a brief review of the time-varying vector AR model.

A. Review of the Time-Varying Vector AR Model

By extension of both the time-invariant vector AR model [8] and the time-varying scalar AR model [9]–[13], the *time-varying vector AR model* for a nonstationary, zero-mean, discrete-time, D-dimensional vector process $\boldsymbol{x}[n] = [x_1[n] \ \cdots \ x_D[n]]^T$ has been defined as [8]

$$\boldsymbol{x}[n] = -\sum_{m=1}^{M} \boldsymbol{A}_m[n](\mathbb{T}^m\boldsymbol{x})[n] + \boldsymbol{e}[n], \quad n = 0, \ldots, N-1. \tag{1}$$

Here, the M matrices $\boldsymbol{A}_m[n], m = 1, \ldots, M$ of size $D \times D$ contain the time-varying AR parameters; M is the AR model order; $\mathbb{T}$ is the time-shift operator acting as[1] $(\mathbb{T}\boldsymbol{x})[n] = \boldsymbol{x}[n-1]$; and $\boldsymbol{e}[n] = [e_1[n] \ \cdots \ e_D[n]]^T$ is a zero-mean, temporally uncorrelated, generally nonstationary innovations noise vector process with correlation matrix $\mathbb{E}\{\boldsymbol{e}[n]\boldsymbol{e}^H[n']\} = \boldsymbol{C}[n]\delta[n-n']$, where $\boldsymbol{C}[n]$ is nonnegative definite (and, thus, Hermitian) for all $n, \mathbb{E}$ is the expectation operator, the superscript ^{H} denotes Hermitian transposition, and $\delta[\cdot]$ is the unit sample.

For the time-varying vector AR model in (1), the number of scalar AR parameters (elements of all matrices $\boldsymbol{A}_m[n], m = 1, \ldots, M, n = 0, \ldots, N-1$) equals NMD^2, and the time-varying innovations correlation $\boldsymbol{C}[n], n = 0, \ldots, N-1$ adds another $ND(D+1)/2$ scalar parameters. For practical values of N, M, and D, the resulting total number of parameters is usually unacceptably large (note that these parameters have to be estimated from the observed vector signal). However, the fact that the number of parameters depends on the signal duration N can be avoided by using a basis expansion of the time-varying parameters (see, e.g., [9]–[11] for the scalar case). A special basis expansion will be considered at a later point.

B. Formulation of the VTFAR Model

The time-varying vector AR model in (1) allows for arbitrary time variations of the parameter matrices, which was seen to result in a significant loss of parsimony. Here, we pursue a different approach: we propose to extend the time-invariant vector AR model to the nonstationary case by including a limited number of frequency shifts in addition to time shifts. The frequency shifts (Doppler shifts) provide an intuitive and physically motivated way of capturing the spectral correlation of nonstationary vector processes without a severe loss in parsimony.

[1]We consider all signals on the finite time interval $\{0, \ldots, N-1\}$, with N assumed even. Equivalently, the signals may be considered periodic with period N. This means that $\mathbb{T}$ is actually a cyclic time-shift operator (for simplicity of notation, we avoid writing more explicitly $(\mathbb{T}\boldsymbol{x})[n] = \boldsymbol{x}[(n-1) \bmod N]$). The cyclic/periodic time structure implies a discrete-frequency framework that allows for efficient FFT-based computations.

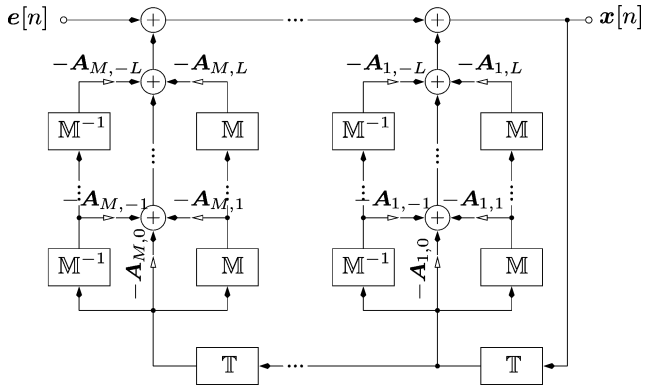

Fig. 1. Block diagram of the proposed VTFAR model (white arrows represent matrix multiplication).

They will be expressed by the frequency-shift (modulation) operator $\mathbb{M}$ that acts as $(\mathbb{M}\boldsymbol{x})[n] = e^{j2\pi n/N}\boldsymbol{x}[n]$. It will furthermore be convenient to combine multiple time shifts (operator $\mathbb{T}^m$) and multiple frequency shifts (operator $\mathbb{M}^l$) into the *joint TF shift operator* $\mathbb{S}_{m,l} \triangleq \mathbb{M}^l\mathbb{T}^m$ acting as

$$(\mathbb{S}_{m,l}\boldsymbol{x})[n] = (\mathbb{M}^l\mathbb{T}^m\boldsymbol{x})[n] = e^{j\frac{2\pi}{N}ln}\boldsymbol{x}[n-m]$$

where both shift indexes m and l are constrained to the range $[-N/2, N/2-1]$.

The proposed VTFAR model is now defined by the relation

$$\boldsymbol{x}[n] = -\sum_{m=1}^{M}\sum_{l=-L}^{L} \boldsymbol{A}_{m,l}(\mathbb{S}_{m,l}\boldsymbol{x})[n] + \boldsymbol{e}[n] \tag{2a}$$

$$= -\sum_{m=1}^{M}\sum_{l=-L}^{L} \boldsymbol{A}_{m,l}e^{j\frac{2\pi}{N}ln}\boldsymbol{x}[n-m] + \boldsymbol{e}[n] \tag{2b}$$

for $n = 0, \ldots, N-1$. Here, the $\boldsymbol{A}_{m,l}$ are parameter matrices of size $D \times D$; $\boldsymbol{e}[n]$ is a zero-mean, temporally uncorrelated, circularly symmetric complex, generally nonstationary innovations noise vector with correlation matrix $\mathbb{E}\{\boldsymbol{e}[n]\boldsymbol{e}^H[n']\} = \boldsymbol{C}[n]\,\delta[n-n']$ (where $\boldsymbol{C}[n]$ will be further specified presently); $M \leq N/2-1$ is the delay order (determining the extent of temporal correlation in $\boldsymbol{x}[n]$); and $L \leq N/2-1$ is the Doppler order (determining the extent of spectral correlation and, thus, the degree of nonstationarity in $\boldsymbol{x}[n]$). In scalar (elementwise) notation, (2a) reads

$$x_d[n] = -\sum_{m=1}^{M}\sum_{l=-L}^{L}\sum_{d'=1}^{D} a_{m,l}^{(d,d')}(\mathbb{S}_{m,l}x_{d'})[n] + e_d[n], \quad d = 1, \ldots, D \tag{3}$$

where $a_{m,l}^{(d,d')}$ is the element of the matrix $\boldsymbol{A}_{m,l}$ located in the dth row and d'th column. We will briefly refer to the model defined by (2) as the VTFAR(M, L) model. Note that if $\boldsymbol{e}[n]$ is stationary and $L = 0$, the time-invariant vector AR model [2], [8] is reobtained.

The VTFAR(M, L) input-output relation (2) in the form $\boldsymbol{x}[n] = \sum_{m=1}^{M}[\sum_{l=-L}^{L}(-\boldsymbol{A}_{m,l})\mathbb{M}^l]\mathbb{T}^m\boldsymbol{x}[n] + \boldsymbol{e}[n]$ is depicted in Fig. 1. This is a generalized recursive tapped delay line in which the taps are replaced by modulation circuits. A different though mathematically equivalent

block diagram can be obtained by writing (2) in the form $\boldsymbol{x}[n] = \sum_{l=-L}^{L}[\sum_{m=1}^{M}(-\boldsymbol{A}_{m,l})e^{j2\pi lm/N}\mathbb{T}^m]\mathbb{M}^l\boldsymbol{x}[n] + \boldsymbol{e}[n]$.

Comparing (2b) with (1), it is seen that the VTFAR model can be viewed as a time-varying vector AR model (1) in which the time-varying AR parameter matrices $\boldsymbol{A}_m[n]$ are given by the *basis expansions*

$$\boldsymbol{A}_m[n] = \sum_{l=-L}^{L} \boldsymbol{A}_{m,l}e^{j\frac{2\pi}{N}ln}.$$

According to this expression, the elements of the matrices $\boldsymbol{A}_m[n]$ are constrained to lie in the subspace spanned by the basis functions (complex exponentials) $f_l[n] = e^{j2\pi ln/N}, l = -L, \ldots, L$. To complete the definition of the VTFAR(M, L) model, we impose a similar subspace restriction (basis expansion) on the innovations correlation matrix $\boldsymbol{C}[n]$, i.e.,

$$\boldsymbol{C}[n] = \sum_{l=-2L}^{2L} \boldsymbol{C}_l e^{j\frac{2\pi}{N}ln}. \tag{4}$$

Here, the Doppler order is chosen as $2L$, so that one can find a matrix square root $\boldsymbol{C}^{1/2}[n]$ with Doppler order L (if $\boldsymbol{C}^{1/2}[n]$ has Doppler order L, then $\boldsymbol{C}[n] = \boldsymbol{C}^{1/2}[n]\boldsymbol{C}^{H/2}[n]$ has Doppler order $2L$). The Doppler orders for $\boldsymbol{A}_m[n]$ and $\boldsymbol{C}^{1/2}[n]$ are chosen equal for simplicity but could be different in general. We note that the VTFAR model is a vector (multivariate) extension of the TFAR model[2] proposed in [6], which is reobtained for $D = 1$.

The number of scalar parameters describing the VTFAR(M, L) model is

$$\begin{aligned}\mathcal{N}_{\mathrm{VTFAR}}(M, L) &= M(2L+1)D^2 + (2L+1)D^2 \\ &= (M+1)(2L+1)D^2\end{aligned} \tag{5}$$

where the first term in (5) is the total number of elements of all matrices $\boldsymbol{A}_{m,l}$ and the second term is the number of elements of all matrices $\boldsymbol{C}_l$ that can be chosen independently (note that $\boldsymbol{C}_{-l}^H = \boldsymbol{C}_l$). For the practically relevant case $L \ll N$, the VTFAR(M, L) model thus requires significantly fewer parameters than the general time-varying vector AR model in (1), which was seen in Section II-A to require $MND^2 + N(D^2 + D)/2$ parameters. The parsimony of the VTFAR model is better for smaller orders M and L. We shall argue in Section III-B that small M and L imply that the VTFAR process $\boldsymbol{x}[n]$ in (2) is *underspread*, i.e., a process whose correlation in time and frequency decays quickly [14]–[16]. Thus, the VTFAR model is particularly suited for underspread vector processes. The underspread property will be shown to play an important role in the analysis of the VTFAR model and for the development of computationally efficient parameter estimators.

C. Banded VTFAR Model

Whereas the number of parameters in the VTFAR model scales linearly with the delay and Doppler model orders M and L, it scales quadratically with the signal dimension D. The quadratic scaling can be avoided by assuming that the parameter matrices $\boldsymbol{A}_{m,l}$ and the innovations correlation matrices $\boldsymbol{C}_l$ are banded, i.e., the entries of these matrices satisfy $a_{m,l}^{(d,d')} = 0$ and $c_l^{(d,d')} = 0$ for $|d - d'| > B$, where $B \in \{0, \ldots, D-1\}$ denotes the (one-sided) matrix bandwidth. Thus, we define the *banded VTFAR(M, L, B) model* elementwise as [cf. (3)]

$$x_d[n] = -\sum_{m=1}^{M}\sum_{l=-L}^{L}\sum_{d'\in\mathcal{B}_d} a_{m,l}^{(d,d')}(\mathbb{S}_{m,l}x_{d'})[n] + e_d[n], \quad d = 1, \ldots, D \tag{6}$$

with $\mathcal{B}_d \triangleq \{d' \in \{1, \ldots, D\} \mid |d' - d| \leq B\}$. The banded VTFAR model yields an even more parsimonious representation than the VTFAR model. It is suited to situations where the correlation between two components $x_d[n]$ and $x_{d'}[n]$ of the vector signal $\boldsymbol{x}[n]$ decreases with increasing "index distance" $|d - d'|$. This frequently happens, e.g., when d corresponds to space or time (see [3] for an example in mobile communications). A smaller matrix bandwidth B presupposes that the correlation of the components of $\boldsymbol{x}[n]$ decays more quickly. In the extreme case where $B = 0$, the vector elements $x_d[n]$ are uncorrelated, which means that the banded VTFAR$(M, L, 0)$ model degenerates to D individual scalar TFAR(M, L) models. The other extreme, $B = D-1$, corresponds to no band restriction at all, i.e., to the VTFAR(M, L) model introduced in Section II-B. In the more general case where only a (known) subset of the elements of the matrices $\boldsymbol{A}_{m,l}$ and $\boldsymbol{C}_l$ is nonzero but the matrices do not necessarily have a banded structure, a reordering of the elements of $\boldsymbol{x}[n]$ may result in a banded structure.

The number of scalar parameters characterizing the banded VTFAR(M, L, B) model can be shown to be

$$\mathcal{N}_{\text{B-VTFAR}}(M, L, B) = (M+1)(2L+1)D'(B) \tag{7}$$

where

$$D'(B) \triangleq \sum_{d=1}^{D}|\mathcal{B}_d| = D(2B+1) - B(B+1) \tag{8}$$

is the number of nonzero elements of a banded $D \times D$ matrix with bandwidth B. Note that $\mathcal{N}_{\text{B-VTFAR}}(M, L, B)$ is upper bounded by $(M+1)(2L+1)D(2B+1)$. Thus, the number of parameters scales linearly with the delay order M, Doppler order L, vector dimension D, and matrix bandwidth B.

III. Analysis of the VTFAR Model

Next, we discuss some basic aspects of the VTFAR model. We characterize the innovations system of a VTFAR process and consider TF representations of the innovations system and of the second-order statistics. Simplifications are obtained through approximations valid in the underspread case $ML \ll N$.

A. Innovations System

With the convention that $\boldsymbol{A}_{0,l} \triangleq \mathbf{I}_D\delta[l]$, where $\mathbf{I}_D$ is the $D \times D$ identity matrix, the VTFAR input-output relation (2a) can be rewritten as

$$\sum_{m=0}^{M}\sum_{l=-L}^{L}\boldsymbol{A}_{m,l}(\mathbb{S}_{m,l}\boldsymbol{x})[n] = \boldsymbol{e}[n] \tag{9}$$

[2]In [6], we assumed the innovations noise to be a *stationary* temporally uncorrelated process and we explicitly included a degenerate zero-lag TF moving-average (TFMA) model. In our setup, this corresponds to setting $\mathbf{C}[n] = \mathbf{B}[n]\mathbf{B}^H[n]$ where $\mathbf{B}[n] = \sum_{l=-L}^{L}\mathbf{B}_l e^{j2\pi ln/N}$ is a time-varying instantaneous mixture matrix representing the zero-lag VTFMA part.

or elementwise as

$$\sum_{m=0}^{M}\sum_{l=-L}^{L}\sum_{d'=1}^{D} a_{m,l}^{(d,d')}(\mathbb{S}_{m,l}x_{d'})[n] = e_d[n], \quad d=1,\ldots,D.$$

Introducing the MIMO operator[3] (i.e., matrix-valued operator or operator-valued matrix of size $D\times D$)

$$\underline{\mathbb{A}} \triangleq \sum_{m=0}^{M}\sum_{l=-L}^{L} \boldsymbol{A}_{m,l}\mathbb{S}_{m,l} = \begin{bmatrix} \mathbb{A}^{(1,1)} & \cdots & \mathbb{A}^{(1,D)} \\ \vdots & \ddots & \vdots \\ \mathbb{A}^{(D,1)} & \cdots & \mathbb{A}^{(D,D)} \end{bmatrix}$$

$$\text{with } \mathbb{A}^{(d,d')} = \sum_{m=0}^{M}\sum_{l=-L}^{L} a_{m,l}^{(d,d')}\mathbb{S}_{m,l} \tag{10}$$

we can express (9) more compactly as

$$(\underline{\mathbb{A}}\boldsymbol{x})[n] = \boldsymbol{e}[n]. \tag{11}$$

The inverse of the MIMO operator $\underline{\mathbb{A}}$ is defined via the equivalence

$$\underline{\mathbb{H}} \triangleq \underline{\mathbb{A}}^{-1} \Longleftrightarrow (\underline{\mathbb{H}}\underline{\mathbb{A}})^{(d,d')} = \mathbb{I}\delta_{d,d'} \tag{12}$$

where MIMO operator composition is defined by $(\underline{\mathbb{H}}\underline{\mathbb{A}})^{(d,d')} = \sum_{d''=1}^{D}\mathbb{H}^{(d,d'')}\mathbb{A}^{(d'',d')}$. The inverse $\underline{\mathbb{H}} = \underline{\mathbb{A}}^{-1}$ then allows us to rewrite (11) as the *innovations representation*

$$\boldsymbol{x}[n] = (\underline{\mathbb{H}}\boldsymbol{e})[n] = \sum_{m=0}^{N-1}\boldsymbol{H}[n,m]\boldsymbol{e}[n-m] \tag{13}$$

where $\underline{\mathbb{H}}$ is the MIMO *innovations system* of the VTFAR model and the $D\times D$ matrix $\boldsymbol{H}[n,m]$ is the impulse response matrix of $\underline{\mathbb{H}}$. Note that $\underline{\mathbb{H}}$ transforms the "innovations noise" $\boldsymbol{e}[n]$ into the VTFAR process $\boldsymbol{x}[n]$. The innovations representation can be written elementwise as

$$\begin{aligned} x_d[n] &= \sum_{d'=1}^{D}(\mathbb{H}^{(d,d')}e_{d'})[n] \\ &= \sum_{d'=1}^{D}\sum_{m=0}^{N-1} h^{(d,d')}[n,m]e_{d'}[n-m], \qquad d=1,\ldots,D \end{aligned} \tag{14}$$

where $h^{(d,d')}[n,m]$ is the impulse response of the operator $\mathbb{H}^{(d,d')}$. Note that $h^{(d,d')}[n,m] = [\boldsymbol{H}[n,m]]_{d,d'}$ is the element of the MIMO impulse response matrix $\boldsymbol{H}[n,m]$ located in the dth row and d'th column. According to (14), $h^{(d,d')}[n,m]$ is the tap coefficient describing the contribution of $e_{d'}[n-m]$, i.e., the input signal $e_{d'}$ at time instant $n-m$, to the output signal x_d at time instant n.

For small M and L (more precisely $ML \ll N$), the operator $\underline{\mathbb{A}} = \sum_{m=0}^{M}\sum_{l=-L}^{L}\boldsymbol{A}_{m,l}\mathbb{S}_{m,l}$ introduces only a limited amount of TF shifts, and thus it is a special case of an operator with a rapidly decaying TF shift expansion (*underspread* operator [17]). For scalar operators, it has been shown in [18] that the inverse of an operator with rapidly decaying TF shift expansion has a rapidly decaying TF shift expansion. Under some technical assumptions, this result can be extended to the case of MIMO operators [19]. Thus, if $ML \ll N$, the innovations system $\underline{\mathbb{H}} = \underline{\mathbb{A}}^{-1}$ is an underspread operator; specifically, its impulse response $\boldsymbol{H}[n,m]$ has exponential decay in m. For N sufficiently large, it then follows that $\boldsymbol{H}[n,m]$ is approximately zero for $m \in \{N/2,\ldots,N-1\}$, which means that $\underline{\mathbb{H}}$ is approximately causal (recall that all signals and systems are N-periodic). Hereafter, we will assume $\underline{\mathbb{H}}$ to be exactly causal, which means that the summation range for m in (13) and (14) is $\{0,\ldots,N/2-1\}$ instead of $\{0,\ldots,N-1\}$. This assumption will be exploited in our development of VTFAR parameter estimators in Section IV.

B. Time-Frequency Representations and Underspread Approximations

We will now study the innovations system $\underline{\mathbb{H}}$ and the second-order statistics of $\boldsymbol{x}[n]$ using tools from TF analysis. To obtain expressions that are simple to compute and easy to interpret, we will use underspread approximations that extend those developed in [14] and [17] to the vector case.

Innovations System: By extension of the scalar case [20], [21], we define the *TF transfer function matrix* of the innovations system $\underline{\mathbb{H}}$ as

$$\mathbf{H}[n,k] \triangleq \sum_{m=0}^{N/2-1}\boldsymbol{H}[n,m]e^{-j\frac{2\pi}{N}km} \tag{15}$$

where $k\in\{-N/2,\ldots,N/2-1\}$ denotes discrete frequency. We furthermore define the *delay-Doppler spreading function matrix* of $\underline{\mathbb{H}}$ as

$$\tilde{\mathbf{H}}[m,l] \triangleq \sum_{n=0}^{N-1}\boldsymbol{H}[n,m]e^{-j\frac{2\pi}{N}ln} \tag{16}$$

where $l\in\{-N/2,\ldots,N/2-1\}$ denotes discrete frequency lag (Doppler). Both $\mathbf{H}[n,k]$ and $\tilde{\mathbf{H}}[m,l]$ have size $D\times D$. The delay-Doppler spreading function matrix $\tilde{\mathbf{H}}[m,l]$ can be shown to be (to within a factor of $1/N$) the coefficient function in an expansion of $\underline{\mathbb{H}}$ into TF shift operators $\mathbb{S}_{m,l}$, i.e.,

$$\underline{\mathbb{H}} = \frac{1}{N}\sum_{m=0}^{N/2-1}\sum_{l=-N/2}^{N/2-1}\tilde{\mathbf{H}}[m,l]\mathbb{S}_{m,l}. \tag{17}$$

Combining (15) and (16) yields the relation

$$\mathbf{H}[n,k] = \frac{1}{N}\sum_{m=0}^{N/2-1}\sum_{l=-N/2}^{N/2-1}\tilde{\mathbf{H}}[m,l]e^{-j\frac{2\pi}{N}(km-nl)}. \tag{18}$$

Let $\mathbf{A}[n,k]$ and $\tilde{\mathbf{A}}[m,l]$ denote the TF transfer function matrix and the delay-Doppler spreading function matrix, respectively, of $\underline{\mathbb{A}}$. Both of these matrices have size $D\times D$. We assume that the delay-Doppler order product satisfies $ML \ll N$, i.e., $\underline{\mathbb{A}} = \sum_{m=0}^{M}\sum_{l=-L}^{L}\boldsymbol{A}_{m,l}\mathbb{S}_{m,l}$ is an underspread operator. As explained in Section III-A, $\underline{\mathbb{H}} = \underline{\mathbb{A}}^{-1}$ is then also an underspread operator; this means that the delay-Doppler spreading function matrix $\tilde{\mathbf{H}}[m,l]$ in (17) is rapidly decaying. Furthermore, based on a matrix extension of the TF operator inversion result in [22, Sec. 2.3.7], the relation $\underline{\mathbb{H}} = \underline{\mathbb{A}}^{-1}$ can then be shown [23, App. A] to imply the approximation

$$\mathbf{H}[n,k] \approx \mathbf{A}^{-1}[n,k], \tag{19}$$

[3]MIMO operators will be denoted using underscored symbols, e.g., $\underline{\mathbb{A}}$.

where $\mathbf{A}^{-1}[n,k]$ denotes the inverse of the matrix $\mathbf{A}[n,k]$, for each (n,k). The accuracy of the approximation (19) improves with decreasing ML.

Finally, comparing $\underline{\mathbb{A}} = \sum_{m=0}^{M}\sum_{l=-L}^{L} \boldsymbol{A}_{m,l}\mathbb{S}_{m,l}$ in (10) with $\underline{\mathbb{A}} = (1/N)\sum_{m=0}^{N/2-1}\sum_{l=-N/2}^{N/2-1} \tilde{\mathbf{A}}[m,l]\mathbb{S}_{m,l}$ [cf. (17)], we see that

$$\tilde{\mathbf{A}}[m,l] = N\boldsymbol{A}_{m,l}. \tag{20}$$

Here, we used the convention that $\boldsymbol{A}_{m,l} = \mathbf{0}$ for $(m,l) \notin \{0,\ldots,M\}\times\{-L,\ldots,L\}$. Inserting (20) into (18) (reformulated for $\underline{\mathbb{A}}$), we see that $\mathbf{A}[n,k]$ can be computed from the VTFAR parameter matrices $\boldsymbol{A}_{m,l}$ by means of the two-dimensional (2-D) discrete Fourier transform (DFT)

$$\mathbf{A}[n,k] = \sum_{m=0}^{M}\sum_{l=-L}^{L} \boldsymbol{A}_{m,l} e^{-j\frac{2\pi}{N}(km-nl)}. \tag{21}$$

From (21) and (19), we conclude that an approximation to $\mathbf{H}[n,k]$ can be computed by a 2-D DFT of the VTFAR parameter matrix sequence $\boldsymbol{A}_{m,l}$ (with complexity $\mathcal{O}(D^2N^2\log N)$) followed by inversion of the matrices $\mathbf{A}[n,k]$ of size $D\times D$ for all N^2 points (n,k) (with complexity $\mathcal{O}(N^2D^3)$). This is much more efficient than a direct calculation of $\underline{\mathbb{H}}$ through inversion of the $ND\times ND$ matrix representation of $\underline{\mathbb{A}}$ (with complexity $\mathcal{O}(N^3D^3)$).

In some situations (especially with estimated VTFAR parameters), $\mathbf{A}[n,k]$ has a high condition number and thus its inversion requires a regularization. This can be achieved by a pseudo-inverse algorithm that thresholds the singular values of $\mathbf{A}[n,k]$ or by diagonal loading techniques [24].

Second-Order Statistics and VTFAR Spectrum: We will now characterize the second-order statistics of a VTFAR process in the TF domain. We first introduce the *cross Rihaczek spectrum* of two nonstationary vector (multivariate) processes $\boldsymbol{x}[n]$ and $\boldsymbol{y}[n]$ as the matrix-valued function (cf. [25])

$$\mathbf{P}_{\boldsymbol{x},\boldsymbol{y}}[n,k] \triangleq \sum_{m=-N/2}^{N/2-1} \boldsymbol{R}_{\boldsymbol{x},\boldsymbol{y}}[n,m]e^{-j\frac{2\pi}{N}km} \tag{22}$$

where $\boldsymbol{R}_{\boldsymbol{x},\boldsymbol{y}}[n,m] = \mathbb{E}\{\boldsymbol{x}[n]\boldsymbol{y}^H[n-m]\}$ is the cross-correlation function of $\boldsymbol{x}[n]$ and $\boldsymbol{y}[n]$. Furthermore, we define the *expected cross ambiguity function* of $\boldsymbol{x}[n]$ and $\boldsymbol{y}[n]$ as

$$\begin{aligned}\mathbf{F}_{\boldsymbol{x},\boldsymbol{y}}[m,l] &\triangleq \sum_{n=0}^{N-1} \boldsymbol{R}_{\boldsymbol{x},\boldsymbol{y}}[n,m]e^{-j\frac{2\pi}{N}ln}\\ &= \mathbb{E}\{\langle \boldsymbol{x}, \mathbb{S}_{m,l}\boldsymbol{y}\rangle\}\end{aligned} \tag{23}$$

with the matrix-valued inner product $\langle \boldsymbol{x},\boldsymbol{y}\rangle \triangleq \sum_{n=0}^{N-1}\boldsymbol{x}[n]\boldsymbol{y}^H[n]$. We will briefly write $\boldsymbol{R}_{\boldsymbol{x}}[n,m]$ for $\boldsymbol{R}_{\boldsymbol{x},\boldsymbol{x}}[n,m]$, and similarly for the other second-order statistics. Finally, we define the *correlation operator* $\underline{\mathbb{R}}_{\boldsymbol{x}}$ as the MIMO operator (operator matrix) whose impulse response is $\boldsymbol{R}_{\boldsymbol{x}}[n,m]$.

From (13) and (12), it follows that the correlation operator of a VTFAR process $\boldsymbol{x}[n]$ is given by[4]

$$\underline{\mathbb{R}}_{\boldsymbol{x}} = \underline{\mathbb{H}}\underline{\mathbb{R}}_{\boldsymbol{e}}\underline{\mathbb{H}}^H = \underline{\mathbb{A}}^{-1}\underline{\mathbb{R}}_{\boldsymbol{e}}\underline{\mathbb{A}}^{-H}. \tag{24}$$

[4]Here, $\underline{\mathbb{H}}^H$ denotes the adjoint of $\underline{\mathbb{H}}$, i.e., the MIMO operator with impulse response matrix $\boldsymbol{H}^H[n-m,-m]$.

Here, $\underline{\mathbb{R}}_{\boldsymbol{e}}$ is the correlation operator of the innovations noise $\boldsymbol{e}[n]$, whose impulse response is given by $\boldsymbol{R}_{\boldsymbol{e}}[n,m] = \boldsymbol{C}[n]\delta[m]$. The Rihaczek spectrum of $\boldsymbol{e}[n]$ follows as $\mathbf{P}_{\boldsymbol{e}}[n,k] = \boldsymbol{C}[n]$. Furthermore, the expected ambiguity function is obtained as $\mathbf{F}_{\boldsymbol{e}}[m,l] = \boldsymbol{C}_l\delta[m]$, where the $\boldsymbol{C}_l$ were defined in (4) (we use the convention that $\boldsymbol{C}_l = \mathbf{0}$ for $|l| > 2L$).

For $ML \ll N$, and assuming that L is not too large, $\underline{\mathbb{H}}$ and $\underline{\mathbb{R}}_{\boldsymbol{e}}$ are jointly underspread [17] [recall that $\underline{\mathbb{R}}_{\boldsymbol{e}}$ corresponds to $\boldsymbol{C}[n]$, whose Doppler expansion is $\boldsymbol{C}[n] = \sum_{l=-2L}^{2L}\boldsymbol{C}_l e^{j2\pi ln/N}$ according to (4)]. Thus, it follows from (24) that $\underline{\mathbb{R}}_{\boldsymbol{x}}$ is an underspread operator or, equivalently, $\boldsymbol{x}[n]$ is an *underspread process*, i.e., a process with rapidly decaying correlation in time and frequency. The Rihaczek spectrum of $\boldsymbol{x}[n]$ can then be approximated as

$$\begin{aligned}\mathbf{P}_{\boldsymbol{x}}[n,k] &\approx \mathbf{H}[n,k]\mathbf{P}_{\boldsymbol{e}}[n,k]\mathbf{H}^H[n,k]\\ &= \mathbf{H}[n,k]\boldsymbol{C}[n]\mathbf{H}^H[n,k].\end{aligned} \tag{25}$$

This is an extension of a (scalar) underspread approximation in [22, Sec. 3.6] to the vector case. Indeed, using the triangle inequality, the magnitude of the elementwise approximation error can be bounded as

$$\begin{aligned}&\left|(\mathbf{P}_{\boldsymbol{x}}[n,k])^{(d,d')} - (\mathbf{H}[n,k]\boldsymbol{C}[n]\mathbf{H}^H[n,k])^{(d,d')}\right| \qquad (26)\\ &= \left|\sum_{d_1,d_2}\Big[Z_{\mathbb{H}^{(d,d_1)}\mathbb{R}_{\boldsymbol{e}}^{(d_1,d_2)}\mathbb{H}^{(d',d_2)H}}[n,k]\right.\\ &\qquad \left. - H^{(d,d_1)}[n,k]C^{(d_1,d_2)}[n]H^{(d',d_2)*}[n,k]\Big]\right|\\ &\le \sum_{d_1,d_2}\left|Z_{\mathbb{H}^{(d,d_1)}\mathbb{R}_{\boldsymbol{e}}^{(d_1,d_2)}\mathbb{H}^{(d',d_2)H}}[n,k]\right.\\ &\qquad \left. - H^{(d,d_1)}[n,k]C^{(d_1,d_2)}[n]H^{(d',d_2)*}[n,k]\right| \qquad (27)\end{aligned}$$

where $Z_{\mathbb{G}}[n,k] \triangleq \sum_{m=0}^{N/2-1} g[n,m]e^{-j2\pi km/N}$ denotes the TF transfer function of a scalar operator $\mathbb{G}$ with impulse response $g[n,m]$ [cf. (15)]. An underspread bound on the error (26) then follows since each of the individual error terms in (27) obeys a (scalar) underspread bound (see [22, Sec. 3.6]).

Using (19) in (25), we obtain further

$$\mathbf{P}_{\boldsymbol{x}}[n,k] \approx \mathbf{A}^{-1}[n,k]\boldsymbol{C}[n]\mathbf{A}^{-H}[n,k]. \tag{28}$$

This "VTFAR spectrum" is a nonstationary vector extension of the well-known expression $P_x(\theta) = \sigma^2/|A(\theta)|^2$ for the power spectrum of a stationary scalar AR process [1], [2] (here, σ^2 is the variance of the stationary innovations noise and $1/A(\theta)$ is the frequency-domain transfer function of the time-invariant innovations system). The approximations (25) and (28) again presuppose the underspread case $ML \ll N$, and they become more accurate for a smaller delay-Doppler order product ML. Finally, the expected ambiguity function $\mathbf{F}_{\boldsymbol{x}}[m,l]$ can be calculated from $\mathbf{P}_{\boldsymbol{x}}[n,k]$ by a 2-D DFT [this is seen by combining (23) and (22)].

IV. VTFAR Parameter Estimation

The statistics of the VTFAR model (2), (4) are determined by the matrices $\boldsymbol{A}_{m,l}$ and $\boldsymbol{C}_l$. In this section, we consider the estimation of these matrices from a single observed realization of the vector process $\boldsymbol{x}[n]$. For estimation of the $\boldsymbol{A}_{m,l}$, we will develop a method that amounts to the solution of a system of linear equations, referred to as *multichannel time-frequency Yule–Walker (TFYW) equations* because of their similarity with the classical Yule-Walker equations [1], [2]. We will also derive an "underspread" approximation to the multichannel TFYW equations, which will serve as the basis for a fast order-recursive solution algorithm (to be developed in the Appendix). The case of the banded VTFAR model will be given special attention.

A. Multichannel TFYW Equations

The derivation of the multichannel TFYW equations extends that of the classical Yule-Walker equations [1], [2]. Forming the matrix-valued inner product of both sides of (9) with $(\mathbb{S}_{m,l}\boldsymbol{x})[n]$ and taking expectations on both sides yields

$$\sum_{m'=0}^{M}\sum_{l'=-L}^{L}\boldsymbol{A}_{m',l'}\mathbf{F}_{\boldsymbol{x}}[m-m',l-l']e^{-j\frac{2\pi}{N}m'(l-l')}=\mathbf{F}_{\boldsymbol{e},\boldsymbol{x}}[m,l]. \tag{29}$$

Applying (23) to $\underline{\mathbb{R}}_{\boldsymbol{e},\boldsymbol{x}}=\underline{\mathbb{R}}_{\boldsymbol{e}}\underline{\mathbb{H}}^H$ and representing both $\underline{\mathbb{R}}_{\boldsymbol{e}}$ and $\underline{\mathbb{H}}^H$ in terms of their delay-Doppler spreading function according to (17), one can show that

$$\mathbf{F}_{\boldsymbol{e},\boldsymbol{x}}[m,l]=\sum_{l'=-2L}^{2L}\boldsymbol{C}_{l'}\tilde{\mathbf{H}}^H[-m,l-l']e^{-j\frac{2\pi}{N}m(l-l')}. \tag{30}$$

Due to the causality of $\underline{\mathbb{H}}$, we have $\tilde{\mathbf{H}}[m,l]=\mathbf{0}$ for $m=-N/2,\ldots,-1$ and thus $\mathbf{F}_{\boldsymbol{e},\boldsymbol{x}}[m,l]$ vanishes for $m=1,\ldots,N/2$. We take advantage of this fact and consider (29) for $m=1,\ldots,M$ (recall that $M\leq N/2-1$). Moving the $m'=0$ term of the sum in (29) to the right-hand side while recalling that $\boldsymbol{A}_{0,l}=\mathbf{I}_D\delta[l]$, we obtain

$$\sum_{m'=1}^{M}\sum_{l'=-L}^{L}\boldsymbol{A}_{m',l'}\mathbf{F}_{\boldsymbol{x}}[m-m',l-l']e^{-j\frac{2\pi}{N}m'(l-l')}=-\mathbf{F}_{\boldsymbol{x}}[m,l],\quad m=1,\ldots,M,\quad l=-L,\ldots,L. \tag{31}$$

This system of $M(2L+1)D^2$ linear equations in the $M(2L+1)D^2$ unknown elements of the VTFAR parameter matrices $\boldsymbol{A}_{m,l}$ will be termed the multichannel TFYW equations. Calculating the $\boldsymbol{A}_{m,l}$ via solution of the multichannel TFYW equations has a complexity of order $\mathcal{O}(M^3(2L+1)^3D^3)$.

The multichannel TFYW equations (31) involve the expected ambiguity function $\mathbf{F}_{\boldsymbol{x}}[m,l]$ for $m=-M+1,\ldots,M-1$ and $l=-2L,\ldots,2L$. In practice, $\mathbf{F}_{\boldsymbol{x}}[m,l]$ is usually unknown and has to be estimated from a given observation of $\boldsymbol{x}[n]$. We recall from (23) that the expected ambiguity function can be written as $\mathbf{F}_{\boldsymbol{x}}[m,l]=\mathbb{E}\{\langle\boldsymbol{x},\mathbb{S}_{m,l}\boldsymbol{x}\rangle\}$. An unbiased estimate of $\mathbf{F}_{\boldsymbol{x}}[m,l]$ is hence obtained simply by omitting the expectation, i.e.,

$$\begin{aligned}\hat{\mathbf{F}}_{\boldsymbol{x}}[m,l]&\triangleq\langle\boldsymbol{x},\mathbb{S}_{m,l}\boldsymbol{x}\rangle\\&=\sum_{n=0}^{N-1}\boldsymbol{x}[n]\boldsymbol{x}^H[n-m]e^{-j\frac{2\pi}{N}ln}\end{aligned} \tag{32}$$

which is the *ambiguity function* of $\boldsymbol{x}[n]$. The VTFAR parameter estimates obtained by solving (31) with $\mathbf{F}_{\boldsymbol{x}}[m,l]$ replaced by $\hat{\mathbf{F}}_{\boldsymbol{x}}[m,l]$ will be denoted as $\hat{\boldsymbol{A}}_{m,l}$. When multiple observations $\boldsymbol{x}^{(i)}[n]$ are available, an improved estimate $\hat{\mathbf{F}}_{\boldsymbol{x}}[m,l]$ is given by the arithmetic average of the ambiguity functions $\hat{\mathbf{F}}_{\boldsymbol{x}^{(i)}}[m,l]$ of all observations $\boldsymbol{x}^{(i)}[n]$ calculated according to (32).

B. Underspread Multichannel TFYW Equations

The complexity of calculating the matrices $\boldsymbol{A}_{m,l}$ via the multichannel TFYW equations (31) can be reduced by an "underspread approximation" that will be derived next. In the practically prevalent case where $\underline{\mathbb{A}}$ is underspread [17], i.e., $ML\ll N$, the phase factor in (31) can be approximated by 1. Indeed, the approximation error $|e^{-j2\pi m'(l-l')/N}-1|$ can be bounded as

$$\max_{\substack{m'\in\{1,\ldots,M\}\\ l,l'\in\{-L,\ldots,L\}}}\left|e^{-j\frac{2\pi}{N}m'(l-l')}-1\right|=2\left|\sin\left(\frac{\pi M\cdot 2L}{N}\right)\right|\leq 4\pi\frac{ML}{N}$$

and this bound will be small for $ML\ll N$. The resulting underspread approximation to the multichannel TFYW equations (31) then reads

$$\sum_{m'=1}^{M}\sum_{l'=-L}^{L}\boldsymbol{A}_{m',l'}\mathbf{F}_{\boldsymbol{x}}[m-m',l-l']\approx-\mathbf{F}_{\boldsymbol{x}}[m,l],\quad m=1,\ldots,M,\qquad l=-L,\ldots,L. \tag{33}$$

This can be written elementwise as

$$\sum_{m'=1}^{M}\sum_{l'=-L}^{L}\sum_{d''=1}^{D}a_{m',l'}^{(d,d'')}f_{\boldsymbol{x}}^{(d'',d')}[m-m',l-l']\approx-f_{\boldsymbol{x}}^{(d,d')}[m,l] \tag{34}$$

for $m=1,\ldots,M, l=-L,\ldots,L$, and $d,d'=1,\ldots,D$, where $f_{\boldsymbol{x}}^{(d,d')}[m,l]$ is the (d,d')th element of the matrix $\mathbf{F}_{\boldsymbol{x}}[m,l]$. These equations have a 2-D convolution structure, which can be exploited for an efficient solution.

The first step towards an efficient solution method is to use a suitable stacking to rewrite the $M(2L+1)$ matrix equations (33) as a single matrix equation involving a two-level block-Toeplitz (2LBT) matrix. We start out by considering the Hermitian transpose of (33),

$$\sum_{m'=1}^{M}\sum_{l'=-L}^{L}\mathbf{F}_{\boldsymbol{x}}^T[m'-m,l'-l]\boldsymbol{A}_{m',l'}^H=-\mathbf{F}_{\boldsymbol{x}}^T[-m,-l],\quad m=1,\ldots,M,\qquad l=-L,\ldots,L \tag{35}$$

where the symmetry property $\mathbf{F}_{\boldsymbol{x}}^H[m,l]=\mathbf{F}_{\boldsymbol{x}}^T[-m,-l]$ [cf. (23)] has been used. We first perform a stacking with respect to l and l' that allows us to write (35) as

$$\sum_{m'=1}^{M}\boldsymbol{\mathcal{F}}_{m'-m}\boldsymbol{\mathcal{A}}_{m'}=-\mathbf{F}_m,\qquad m=1,\ldots,M. \tag{36}$$

The matrices involved in these equations are the $2M-1$ block-Toeplitz matrices of size $(2L+1)D\times(2L+1)D$

$$\begin{aligned}\boldsymbol{\mathcal{F}}_m&\triangleq\operatorname{toep}\left\{\mathbf{F}_{\boldsymbol{x}}^T[m,-2L],\ldots,\mathbf{F}_{\boldsymbol{x}}^T[m,2L]\right\}\\&=\begin{bmatrix}\mathbf{F}_{\boldsymbol{x}}^T[m,0]&\cdots&\mathbf{F}_{\boldsymbol{x}}^T[m,2L]\\\vdots&\ddots&\vdots\\\mathbf{F}_{\boldsymbol{x}}^T[m,-2L]&\cdots&\mathbf{F}_{\boldsymbol{x}}^T[m,0]\end{bmatrix},\\&\qquad m=-M+1,\ldots,M-1\end{aligned} \tag{37}$$

the M right-hand side matrices of size $(2L+1)D \times D$

$$\mathbf{F}_m \triangleq \left[\mathbf{F}_{\boldsymbol{x}}^T[-m,L]\cdots\mathbf{F}_{\boldsymbol{x}}^T[-m,-L]\right]^T, \qquad m=1,\ldots,M \tag{38}$$

and the M VTFAR parameter matrices of size $(2L+1)D \times D$

$$\boldsymbol{\mathcal{A}}_m \triangleq [\boldsymbol{A}_{m,-L}\cdots\boldsymbol{A}_{m,L}]^H, \qquad m=1,\ldots,M.$$

A second stacking, with respect to m and m', allows us to combine the M matrix equations (36) into the single matrix equation

$$\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{A}} = -\mathbf{F} \tag{39}$$

with the 2LBT matrix of size $M(2L+1)D \times M(2L+1)D$

$$\begin{aligned}\boldsymbol{\mathcal{F}} &\triangleq \operatorname{toep}\{\boldsymbol{\mathcal{F}}_{-M+1},\ldots,\boldsymbol{\mathcal{F}}_{M-1}\}\\ &= \begin{bmatrix} \boldsymbol{\mathcal{F}}_0 & \cdots & \boldsymbol{\mathcal{F}}_{M-1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{F}}_{-M+1} & \cdots & \boldsymbol{\mathcal{F}}_0 \end{bmatrix}\end{aligned}$$

and the matrices of size $M(2L+1)D \times D$

$$\mathbf{F} \triangleq [\mathbf{F}_1^T\cdots\mathbf{F}_M^T]^T, \quad \boldsymbol{\mathcal{A}} \triangleq \left[\boldsymbol{\mathcal{A}}_1^T\cdots\boldsymbol{\mathcal{A}}_M^T\right]^T.$$

The 2LBT structure of the matrix equations (39) is the basis for a fast solution algorithm with complexity $\mathcal{O}(M^2(2L+1)^3D^3)$. This algorithm (which we call the *multichannel Wax-Kailath algorithm*) is developed in the Appendix.

C. Underspread Multichannel TFYW Equations for the Banded VTFAR Model

In the case of the banded VTFAR(M,L,B) model (6), we cannot solve (33) or (34) directly since these equations do not explicitly account for the band structure of the $\boldsymbol{A}_{m,l}$. Keeping in (34) only the terms involving nonzero elements of $\boldsymbol{A}_{m,l}$ (i.e., elements $a_{m,l}^{(d,d')}$ within the matrix band) yields

$$\sum_{m'=1}^{M}\sum_{l'=-L}^{L}\sum_{d''\in\mathcal{B}_d} a_{m',l'}^{(d,d'')} f_{\boldsymbol{x}}^{(d'',d')}[m-m',l-l'] \approx -f_{\boldsymbol{x}}^{(d,d')}[m,l] \tag{40}$$

for $m=1,\ldots,M, l=-L,\ldots,L$, and $d,d'=1,\ldots,D$, where as before $\mathcal{B}_d = \{d' \in \{1,\ldots,D\} |\ |d'-d| \le B\}$. These are $M(2L+1)D^2$ linearly dependent equations in the $M(2L+1)D'$ nonzero VTFAR parameters $a_{m,l}^{(d,d')}$ (we recall that $D' = D(2B+1) - B(B+1)$).

We shall now derive a system of $M(2L+1)D'$ linearly independent equations in the $M(2L+1)D'$ unknowns. In what follows, let $d_{\min}(d)$ and $d_{\max}(d)$ denote, respectively, the minimum and maximum element of the index set $\mathcal{B}_d$. Stacking (40) with respect to d' and d'', we obtain

$$\sum_{m'=1}^{M}\sum_{l'=-L}^{L} \tilde{\mathbf{F}}_{m-m',l-l',d}\boldsymbol{a}_{m',l',d} = -\mathbf{f}_{m,l,d},$$
$$m=1,\ldots,M, \qquad l=-L,\ldots,L, \qquad d=1,\ldots,D \tag{41}$$

with the matrices of size $|\mathcal{B}_d| \times |\mathcal{B}_d|$

$$\begin{aligned}&\tilde{\mathbf{F}}_{m,l,d}\\ &\quad\triangleq \begin{bmatrix} f_{\boldsymbol{x}}^{(d_{\min}(d),d_{\min}(d))}[m,l] & \cdots & f_{\boldsymbol{x}}^{(d_{\max}(d),d_{\min}(d))}[m,l] \\ \vdots & & \vdots \\ f_{\boldsymbol{x}}^{(d_{\min}(d),d_{\max}(d))}[m,l] & \cdots & f_{\boldsymbol{x}}^{(d_{\max}(d),d_{\max}(d))}[m,l] \end{bmatrix}\end{aligned}$$

and the vectors of length $|\mathcal{B}_d|$

$$\begin{aligned}\mathbf{f}_{m,l,d} &\triangleq \left[f_{\boldsymbol{x}}^{(d,d_{\min}(d))}[m,l]\cdots f_{\boldsymbol{x}}^{(d,d_{\max}(d))}[m,l]\right]^T\\ \boldsymbol{a}_{m,l,d} &\triangleq \left[a_{m,l}^{(d,d_{\min}(d))}\cdots a_{m,l}^{(d,d_{\max}(d))}\right]^T.\end{aligned}$$

It should be noted that the size of $\tilde{\mathbf{F}}_{m,l,d}, \mathbf{f}_{m,l,d}$, and $\boldsymbol{a}_{m,l,d}$ depends on d. We next use a stacking with respect to d to represent (41) as

$$\sum_{m'=1}^{M}\sum_{l'=-L}^{L} \tilde{\mathbf{F}}_{m-m',l-l'}\boldsymbol{a}_{m',l'} = -\mathbf{f}_{m,l},$$
$$m=1,\ldots,M, \qquad l=-L,\ldots,L \tag{42}$$

with the block-diagonal matrices of size $D' \times D'$

$$\tilde{\mathbf{F}}_{m,l} \triangleq \operatorname{diag}\{\tilde{\mathbf{F}}_{m,l,d}\}_{d=1,\ldots,D}$$

and the vectors of length D'

$$\begin{aligned}\mathbf{f}_{m,l} &\triangleq [\mathbf{f}_{m,l,1}^T\cdots\mathbf{f}_{m,l,D}^T]^T\\ \boldsymbol{a}_{m,l} &\triangleq \left[\boldsymbol{a}_{m,l,1}^T\cdots\boldsymbol{a}_{m,l,D}^T\right]^T.\end{aligned}$$

Finally, we use two more stackings: first a stacking with respect to l and l', to obtain the block-Toeplitz matrices of size $(2L+1)D' \times (2L+1)D'$

$$\tilde{\boldsymbol{\mathcal{F}}}_m \triangleq \operatorname{toep}\{\tilde{\mathbf{F}}_{m,-2L},\ldots,\tilde{\mathbf{F}}_{m,2L}\}$$

and the vectors of length $(2L+1)D'$

$$\begin{aligned}\mathbf{f}_m &\triangleq [\mathbf{f}_{m,-L}^T\cdots\mathbf{f}_{m,L}^T]^T\\ \boldsymbol{a}_m &\triangleq \left[\boldsymbol{a}_{m,-L}^T\cdots\boldsymbol{a}_{m,L}^T\right]^T;\end{aligned}$$

and then a stacking with respect to m and m', yielding the single 2LBT matrix of size $M(2L+1)D' \times M(2L+1)D'$

$$\tilde{\boldsymbol{\mathcal{F}}} \triangleq \operatorname{toep}\{\tilde{\boldsymbol{\mathcal{F}}}_{-M+1},\ldots,\tilde{\boldsymbol{\mathcal{F}}}_{M-1}\}$$

and the two vectors of length $M(2L+1)D'$

$$\mathbf{f} \triangleq [\mathbf{f}_1^T\cdots\mathbf{f}_M^T]^T, \quad \boldsymbol{a} \triangleq \left[\boldsymbol{a}_1^T\cdots\boldsymbol{a}_M^T\right]^T.$$

With these definitions, we can finally write (42) as

$$\tilde{\boldsymbol{\mathcal{F}}}\boldsymbol{a} = -\mathbf{f}.$$

This is a system of $M(2L+1)D'$ generally linearly independent linear equations in the $M(2L+1)D'$ unknowns (i.e., the nonzero VTFAR parameters $a_{m,l}^{(d,d')}, d' \in \mathcal{B}_d$). Since $\tilde{\boldsymbol{\mathcal{F}}}$ has

2LBT structure, this system of equations can again be efficiently solved with complexity $\mathcal{O}(M^2(2L+1)^3D'^3)$ by means of the multichannel Wax-Kailath algorithm derived in the Appendix.

D. Estimation of Innovations Correlation

Once estimates $\hat{\boldsymbol{A}}_{m,l}$ have been calculated, estimates of the innovations correlation matrices $\boldsymbol{C}_l$ can be obtained by evaluating (29) at $m=0$, i.e.,

$$\sum_{m'=0}^{M}\sum_{l'=-L}^{L}\boldsymbol{A}_{m',l'}\mathsf{F}_{\boldsymbol{x}}[-m',l-l']e^{-j\frac{2\pi}{N}m'(l-l')}=\mathsf{F}_{\boldsymbol{e},\boldsymbol{x}}[0,l]. \tag{43}$$

Combining $\boldsymbol{x}[n]=(\underline{\mathbb{H}}\boldsymbol{e})[n]$ and (17) and (2b), we have

$$\boldsymbol{x}[n]=\frac{1}{N}\sum_{m=0}^{N/2-1}\sum_{l=-N/2}^{N/2-1}\tilde{\mathsf{H}}[m,l]e^{j\frac{2\pi}{N}ln}\boldsymbol{e}[n-m] \tag{44}$$

$$=-\sum_{m=1}^{M}\sum_{l=-L}^{L}\boldsymbol{A}_{m,l}e^{j\frac{2\pi}{N}ln}\boldsymbol{x}[n-m]+\boldsymbol{e}[n]. \tag{45}$$

Due to causality, $\boldsymbol{x}[n-m], m>0$ in (45) can only depend on $\boldsymbol{e}[n-m'], m'\geq m$. Thus, for the term $\boldsymbol{e}[n]$ in (45) to appear also in (44), it is necessary and sufficient that

$$\tilde{\mathsf{H}}[0,l]=N\mathbf{I}_D\delta[l]. \tag{46}$$

From (30), $\mathsf{F}_{\boldsymbol{e},\boldsymbol{x}}[0,l]=\sum_{l'=-2L}^{2L}\boldsymbol{C}_{l'}\tilde{\mathsf{H}}^H[0,l-l']$. Inserting (46), this simplifies to $\mathsf{F}_{\boldsymbol{e},\boldsymbol{x}}[0,l]=\boldsymbol{C}_l$, and hence (43) can be rewritten as

$$\boldsymbol{C}_l=\sum_{m'=0}^{M}\sum_{l'=-L}^{L}\boldsymbol{A}_{m',l'}\mathsf{F}_{\boldsymbol{x}}^H[m',l'-l],\qquad l=-2L,\ldots,2L \tag{47}$$

where we used the symmetry relation $\mathsf{F}_{\boldsymbol{x}}^H[-m,-l]\times e^{j2\pi ml/N}=\mathsf{F}_{\boldsymbol{x}}[m,l]$. Estimates $\hat{\boldsymbol{C}}_l$ of the $\boldsymbol{C}_l$ are then obtained by inserting the previously computed estimates $\hat{\boldsymbol{A}}_{m,l}$ and estimates of the expected ambiguity function $\mathsf{F}_{\boldsymbol{x}}[m,l]$ into (47).

Finally, an estimate of the time-varying innovations correlation $\boldsymbol{C}[n]$ can be obtained via (4) as $\hat{\boldsymbol{C}}[n]=\sum_{l=-2L}^{2L}\hat{\boldsymbol{C}}_l e^{j2\pi ln/N}$. It is important to note that the matrices $\hat{\boldsymbol{C}}[n]$ are not guaranteed to be positive definite. Positive definiteness can be enforced by calculating the eigendecomposition $\hat{\boldsymbol{C}}[n]=\boldsymbol{U}[n]\boldsymbol{\Lambda}[n]\boldsymbol{U}^H[n]$ and replacing the diagonal eigenvalue matrix $\boldsymbol{\Lambda}[n]$ with $\boldsymbol{\Lambda}_+[n]+\alpha\mathbf{I}_D$, where $\boldsymbol{\Lambda}_+[n]$ denotes the positive semidefinite part of $\boldsymbol{\Lambda}[n]$ (in which all negative eigenvalues are replaced by zero) and α is a small positive constant. Thus, the new estimate of $\boldsymbol{C}[n]$ is

$$\hat{\boldsymbol{C}}'[n]\triangleq\boldsymbol{U}[n](\boldsymbol{\Lambda}_+[n]+\alpha\mathbf{I}_D)\boldsymbol{U}^H[n].$$

Unfortunately, this method of enforcing positive definiteness will generally destroy the Doppler band-limitation of $\hat{\boldsymbol{C}}[n]$, i.e., $\hat{\boldsymbol{C}}'[n]$ no longer admits a basis expansion $\hat{\boldsymbol{C}}'[n]=\sum_l\hat{\boldsymbol{C}}'_l e^{j2\pi ln/N}$ with $\hat{\boldsymbol{C}}'_l=\mathbf{0}$ for $|l|>2L$. In our experiments, however, we observed that an iterative scheme of alternately enforcing positive definiteness and Doppler band-limitation produced satisfactory results after a few iterations. In the case of a banded model, the iterations need to include a further step in which the bandedness of the matrices $\hat{\boldsymbol{C}}'[n]$ is enforced by setting all out-of-band elements to zero. As observed in [3], this iterative procedure is actually a POCS (projection onto convex sets) scheme [26].

V. Model Order Estimation

Along with the VTFAR parameter matrices $\boldsymbol{A}_{m,l}$ and $\boldsymbol{C}_l$, it is necessary to estimate the VTFAR model order (M,L) and, for the banded model, the matrix bandwidth B. In this section, therefore, we propose information criteria (IC) for order and bandwidth estimation.

A. Review of Information Criteria for the TFAR and VAR Models

Our IC are motivated by IC that were previously proposed for the scalar TFAR model of nonstationary processes and for the vector AR (VAR) model of stationary processes. A general form of IC for the scalar TFAR(M,L) model [27] with a temporally uncorrelated innovations process $e[n]$ of time-varying variance $\sigma_e^2[n]$ is given by

$$\mathrm{IC}_{\mathrm{TFAR}}(M,L)=\log\widehat{\sigma_{e_0}^2}(M,L)+p\frac{(M+1)(2L+1)}{N} \tag{48}$$

where $\widehat{\sigma_{e_0}^2}(M,L)$ is an estimate of the variance of the (stationarized) innovations process $e_0[n]\triangleq e[n]/\sigma_e[n], p$ is a penalty factor, and $(M+1)(2L+1)$ is the number of parameters (the sum of $M(2L+1)$, the number of pure TFAR parameters and $2L+1$, the number of freely adjustable parameters of the time-varying innovations variance[5] $\sigma_e^2[n]$). The estimate of the innovations variance is calculated as $\widehat{\sigma_{e_0}^2}(M,L)=(1/N)\sum_{n=0}^{N-1}|\hat{e}[n]|^2/\widehat{\sigma_e^2}[n]$, where $\hat{e}[n]$ is an estimate of the innovations signal that is obtained by filtering the observation $x[n]$ using the inverse of the innovations system of the estimated (pure) TFAR(M,L) model [27] and $\widehat{\sigma_e^2}[n]$ is an estimate of the time-varying innovations variance $\sigma_e^2[n]$. The order estimate $(\hat{M},\hat{L})$ then is defined as the order (M,L) for which $\mathrm{IC}_{\mathrm{TFAR}}(M,L)$ is minimum. Important special cases of $\mathrm{IC}_{\mathrm{TFAR}}(M,L)$ are the TFAR versions of Akaike's IC (AIC) [28] and of the minimum description length (MDL) criterion [29], which are obtained for $p=2$ and $p=\log N$, respectively.

For the stationary VAR(M) model, the following general form of IC has been proposed [8]:

$$\mathrm{IC}_{\mathrm{VAR}}(M)=\log\det\hat{\boldsymbol{C}}_{\boldsymbol{e}}(M)+p\frac{MD^2}{N} \tag{49}$$

where D is the dimension of the vector process $\boldsymbol{x}[n]$, the $D\times D$ matrix $\hat{\boldsymbol{C}}_{\boldsymbol{e}}(M)$ is an estimate of the correlation matrix of the stationary innovations process $\boldsymbol{e}[n]$, and p is a penalty factor as above.

B. Order Estimation for the (Banded) VTFAR Model

Next, we extend the TFAR IC and VAR IC reviewed above to the VTFAR model. We directly consider the banded

[5]We note that $\mathrm{IC}_{\mathrm{TFAR}}(M,L)$ is slightly different from the expression obtained by appropriately specializing the IC proposed for the TFARMA$(M,L;M',L')$ model in [27]; this is due to the different parameterization of the nonstationary innovations noise used in [27].

VTFAR(M, L, B) model and also address estimation of the matrix bandwidth B. For the special case of the full (non-banded) VTFAR model, we just have to set $B = D - 1$ and omit the optimization with respect to B.

The form of IC we propose for VTFAR order and bandwidth estimation is

$$\mathrm{IC}_{\mathrm{VTFAR}}(M, L, B) \triangleq \log\det \hat{\boldsymbol{C}}_{\boldsymbol{e}_0}(M, L, B) + p\frac{\mathcal{N}_{\text{B-VTFAR}}(M, L, B)}{N} \quad (50)$$

where $\hat{\boldsymbol{C}}_{\boldsymbol{e}_0}(M, L, B)$ is an estimate of the correlation matrix of the stationarized innovations process $\boldsymbol{e}_0[n] \triangleq \boldsymbol{C}^{-1/2}[n]\boldsymbol{e}[n]$; p is a penalty factor as before (AIC: $p = 2$, MDL: $p = \log N$); and $\mathcal{N}_{\text{B-VTFAR}}(M, L, B) = (M + 1)(2L + 1)D'(B)$ with $D'(B) = D(2B+1) - B(B+1)$ is the total number of freely adjustable parameters of the banded VTFAR(M, L, B) model as given by (7), (8). We note that (48) is reobtained from (50) by setting $D = 1$ and $B = 0$, while (49) is reobtained by setting $L = 0$ and $B = D - 1$ and replacing $M + 1$ with M (this replacement is necessary because the stationary VAR model has a time-independent innovations correlation $\boldsymbol{C}$). The innovations correlation matrix is estimated by the sample correlation matrix, i.e.,

$$\hat{\boldsymbol{C}}_{\boldsymbol{e}_0}(M, L, B) = \frac{1}{N}\sum_{n=0}^{N-1} \hat{\boldsymbol{e}}_0[n]\hat{\boldsymbol{e}}_0^H[n]$$

where $\hat{\boldsymbol{e}}_0[n] = \hat{\boldsymbol{C}}^{-1/2}[n]\hat{\boldsymbol{e}}[n]$. Here, $\hat{\boldsymbol{e}}[n]$ is an estimate of the innovations signal that is obtained by filtering the observation $\boldsymbol{x}[n]$ using the inverse of the innovations system of the estimated banded VTFAR(M, L, B) model, and $\hat{\boldsymbol{C}}[n]$ is an estimate of the time-varying correlation matrix of $\boldsymbol{e}[n]$ (see Section IV-D).

Let us assume that $\mathrm{IC}_{\mathrm{VTFAR}}(M, L, B)$ has been calculated for all M and L up to some maximum orders $M_{\max}$ and $L_{\max}$, respectively, as well as for all B up to $D-1$. Then, estimates of delay order M, Doppler order L, and matrix bandwidth B are obtained by minimizing $\mathrm{IC}_{\mathrm{VTFAR}}(M, L, B)$:

$$(\hat{M}, \hat{L}, \hat{B}) = \arg \min_{\substack{M\in\{1,\ldots,M_{\max}\}\\ L\in\{0,\ldots,L_{\max}\}\\ B\in\{0,\ldots,D-1\}}} \mathrm{IC}_{\mathrm{VTFAR}}(M, L, B).$$

VI. Numerical Results

We next present numerical results assessing the performance of the proposed VTFAR parameter and order estimators. We also demonstrate the application of the VTFAR model and estimators to nonstationary multivariate spectral analysis within an array processing setting. Additional simulation results concerning the application of VTFAR methods to the modeling of wireless channels can be found in [3].

A. VTFAR Parameter Estimation

First, we analyze the accuracy of the VTFAR parameter estimator based on the underspread multichannel TFYW equations (see Sections IV-B and IV-D). We generated 100 realizations of a Gaussian bivariate $(D = 2)$ VTFAR(M, L) process with orders $M = 3, L = 2$ and signal duration $N = 512$. This model is described by $\mathcal{N}_{\mathrm{VTFAR}}(3, 2) = 80$ parameters. We then estimated the VTFAR parameters from each single realization. As mean performance measures, we computed the normalized mean-square error (MSE), normalized variance, and normalized squared bias of the estimated parameters. The normalized MSE was calculated as

$$\frac{\sum_{m=1}^{M}\sum_{l=-L}^{L}\mathrm{MSE}\{\hat{\boldsymbol{A}}_{m,l}\} + \sum_{l=-2L}^{2L}\mathrm{MSE}\{\hat{\boldsymbol{C}}_l\}}{\sum_{m=1}^{M}\sum_{l=-L}^{L}\|\boldsymbol{A}_{m,l}\|_{\mathrm{F}}^2 + \sum_{l=-2L}^{2L}\|\boldsymbol{C}_l\|_{\mathrm{F}}^2}$$

where, e.g., $\mathrm{MSE}\{\hat{\boldsymbol{A}}_{m,l}\}$ denotes the result of averaging $\|\hat{\boldsymbol{A}}_{m,l} - \boldsymbol{A}_{m,l}\|_{\mathrm{F}}^2$ over all 100 realizations (here, $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, i.e., the Euclidean matrix norm). This definition of the normalized MSE can be justified by the fact that the magnitudes of the parameters $\hat{a}_{m,l}^{(d,d')}$ and $\hat{c}_l^{(d,d')}$ were comparable in our simulations. The normalized variance and normalized squared bias were calculated in an analogous manner.

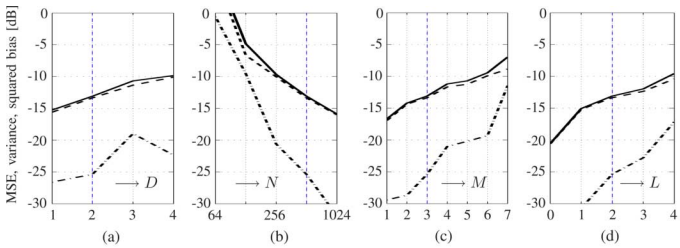

Fig. 2. Normalized MSE (solid lines), normalized variance (dashed lines), and normalized squared bias (dash-dotted lines) of the underspread multichannel TFYW parameter estimator for a VTFAR(M, L) process of dimension D and duration N: (a) D variable, $N = 512, M = 3, L = 2$; (b) $D = 2, N$ variable, $M = 3, L = 2$; (c) $D = 2, N = 512, M$ variable, $L = 2$; and (d) $D = 2, N = 512, M = 3, L$ variable. The dash-dotted vertical lines indicate the VTFAR(3, 2) process with $D = 2$ and $N = 512$.

This experiment was then repeated for varying signal dimension $D = 1, \ldots, 4$ or observation length $N = 64, \ldots, 1024$ or delay order $M = 1, \ldots, 7$ or Doppler order $L = 0, \ldots, 4$ while leaving the respective other parameters unchanged. The results are shown in Fig. 2. One can see that the MSE increases with increasing D, M, and L but decreases with increasing N. The results in parts (a), (c), and (d) of Fig. 2 are quite good for most choices of the parameters D, M, and L because the signal duration N (number of observed data) is sufficiently large compared to the number of parameters to be estimated or, in other words, because the process is well underspread (i.e., $ML \ll N$). It is also seen that the squared bias tends to be much smaller than the variance.

B. VTFAR Order Estimation

Next, we present numerical results for the order estimators proposed in Section V. We generated 100 realizations of a Gaussian banded VTFAR(M, L, B) process with dimension $D = 2$, orders $M = 3$ and $L = 2$, matrix bandwidth $B = 0$, and duration $N = 512$. Note that $B = 0$ means that we are actually considering a diagonal model, i.e., each of the two scalar signal components corresponds to a scalar TFAR model [6], without any correlation between the two components. We estimated the orders M, L and matrix bandwidth B for each realization, using the IC in (50) with penalty factors $p = 2$ (AIC)

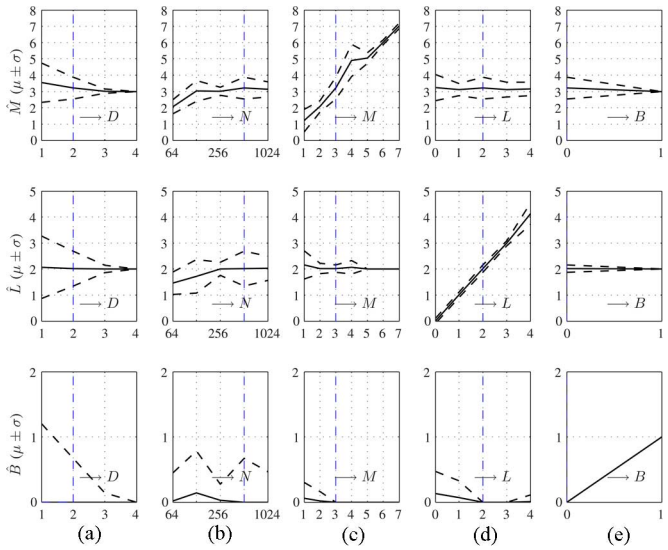

Fig. 3. Performance of AIC-type banded VTFAR order estimators (solid lines: mean, dashed lines: mean $\pm$ standard deviation) for various banded VTFAR(M, L, B) models: (a) D variable, $N = 512, M = 3, L = 2, B = 0$; (b) $D = 2, N$ variable, $M = 3, L = 2, B = 0$; (c) $D = 2, N = 512, M$ variable, $L = 2, B = 0$; (d) $D = 2, N = 512, M = 3, L$ variable, $B = 0$; and (e) $D = 2, N = 512, M = 3, L = 2, B$ variable. Top: estimator $\hat{M}$, center: estimator $\hat{L}$, bottom: estimator $\hat{B}$. The dash-dotted vertical lines indicate the banded VTFAR(3, 2, 0) process with $D = 2$ and $N = 512$.

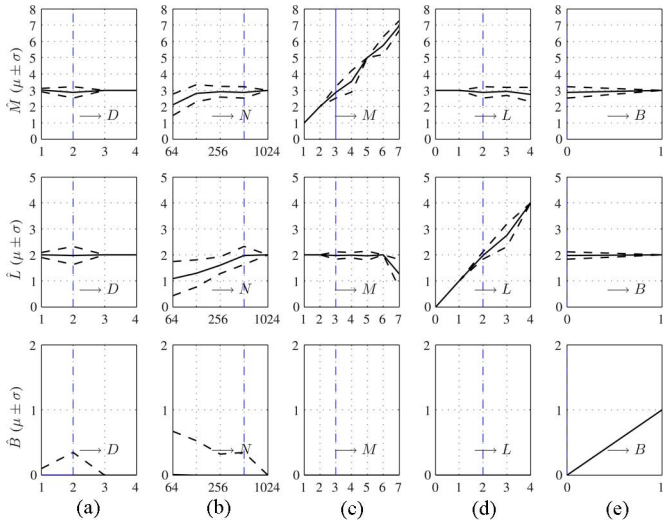

Fig. 4. Performance of MDL-type banded VTFAR order estimators (see caption of Fig. 3 for details).

and $p = \log N$ (MDL). This experiment was then repeated for varying D or N or M or L or B while leaving all other parameters unchanged. The means and standard deviations of the estimates $\hat{M}, \hat{L}$, and $\hat{B}$ are shown for the AIC in Fig. 3 and for the MDL in Fig. 4, as a function of D, N, M, L, and B. It is seen that both IC produce very good results; significant deviations from the true model orders occur only for small N, large M, or large L, i.e., if the underspread condition $ML \ll N$ is not well satisfied. Here, the AIC tends to overestimate the model order whereas the MDL tends to underestimate it.

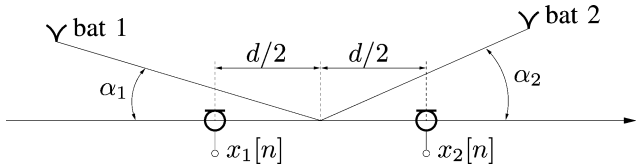

Fig. 5. Geometric setup of the microphone array.

C. Application to Nonstationary Multivariate Spectral Analysis

Our final simulation results demonstrate the application of the VTFAR model and estimators to nonstationary multivariate spectral analysis within the array processing setup illustrated in Fig. 5. Two microphones separated by distance d receive acoustic signals $s_1[n]$ and $s_2[n]$ from two sources located at angles α_1 and α_2, respectively. The source signals impinge on the microphones with a phase difference depending on d, α_1, and α_2. We assume that the sources are sufficiently distant from the microphones so that the signal attenuation for a given source is effectively the same at both microphones (without loss of generality, the attenuation factor will be set equal to 1). All signals are represented by their discrete-time, complex (analytic) versions.

The microphone output signals, denoted by $x_1[n]$ and $x_2[n]$, are noisy mixtures of the two source signals $s_1[n], s_2[n]$:

$$\boldsymbol{x}[n] \triangleq \begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} s_1[n] \\ s_2[n] \end{bmatrix} + \begin{bmatrix} w_1[n] \\ w_2[n] \end{bmatrix}$$

with the mixing matrix [30]

$$\boldsymbol{A} = \begin{bmatrix} 1 & e^{j\frac{2\pi}{\lambda} d \cos \alpha_2} \\ e^{j\frac{2\pi}{\lambda} d \cos \alpha_1} & 1 \end{bmatrix}$$

(here, λ is the wavelength) and with independent Gaussian white noise sequences $w_1[n]$ and $w_2[n]$. In our simulation, we used bat chirp signals (taken from [31]) for the source signals $s_1[n], s_2[n]$. We emphasize that these signals do not conform to our VTFAR model. The signal duration is $N = 512$. The signals have a relatively smooth attack and decay behavior. (For signals with a sudden attack and/or decay, windowing should be used to reduce leakage artifacts.) The physical parameters are $d = 10$ m, $\alpha_1 = 20°, \alpha_2 = 10°$, and $\lambda = 11.4$ mm; the signal-to-noise ratio is 15 dB.

For nonstationary spectral analysis of the bivariate signal $\boldsymbol{x}[n] = [x_1[n]\ x_2[n]]^T$, we compared the parametric VTFAR approach with a nonparametric method. The nonparametric analysis was performed by the auto and cross versions of the smoothed pseudo Wigner distribution (SPWD) [32], [33]; the results are shown in Fig. 6. For the parametric analysis, we used the VTFAR spectrum as given by the right-hand side of (28), based on a VTFAR(4, 2) model. The $\mathcal{N}_{\mathrm{VTFAR}}(4, 2) = 100$ parameters of this model were estimated from $\boldsymbol{x}[n]$ by means of the underspread multichannel TFYW estimator of Sections IV-B and IV-D, and the VTFAR model order of (4, 2) was obtained by the MDL order estimator of Section V-B. The VTFAR spectra are shown in Fig. 7. It can be seen from Figs. 6 and 7 that the source signals are clearly represented by both the nonparametric SPWD analysis and the parametric VTFAR

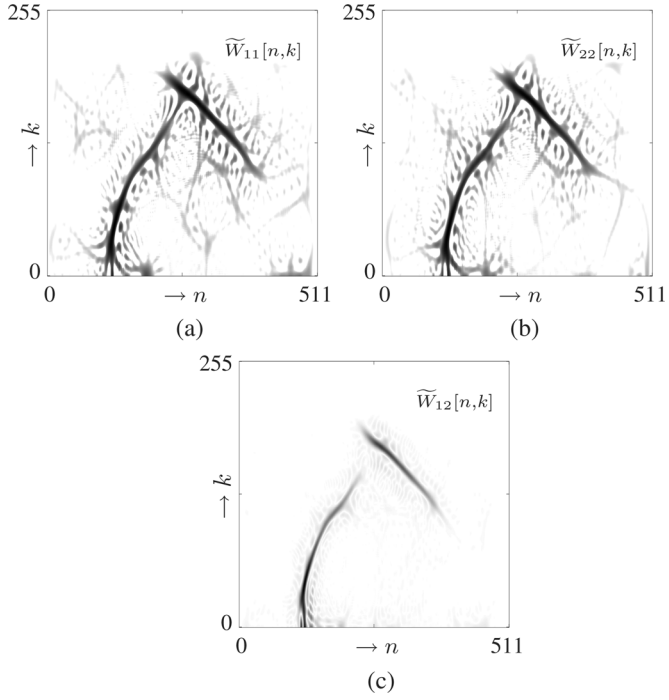

Fig. 6. Nonparametric nonstationary spectral analysis of the bivariate signal $\boldsymbol{x}[n]$ using the SPWD: (a) SPWD of $x_1[n]$, (b) SPWD of $x_2[n]$, (c) cross-SPWD of $x_1[n]$ and $x_2[n]$. Positive real parts are shown in (a) and (b), while the magnitude of the real part is shown in (c).

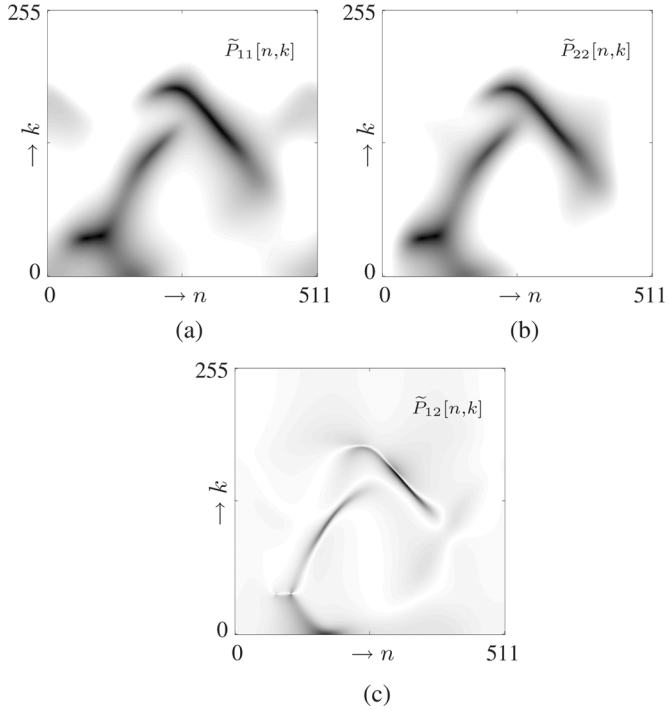

Fig. 7. Parametric nonstationary spectral analysis of the bivariate signal $\boldsymbol{x}[n]$ using estimated VTFAR(4, 2) spectra: (a) Spectrum of $x_1[n]$, (b) spectrum of $x_2[n]$, (c) cross-spectrum of $x_1[n]$ and $x_2[n]$. Positive real parts are shown in (a) and (b), while the magnitude of the real part is shown in (c).

analysis. More specifically, in either analysis, both source signals are visible in each of the two auto-spectra and in the cross-spectrum. While the SPWD spectra exhibit better TF concentration than the VTFAR spectra, the VTFAR spectra are free of oscillatory cross (interference) terms [32], [33]. The absence of cross terms is desirable in most applications and consistent with the implicit underspread assumption underlying VTFAR models with a small number of parameters. An underspread process does not feature any long-range correlation between TF disjoint process components (such correlation would give rise to oscillatory cross terms in the nonstationary spectrum) [14]. It should be noted at this point that all VTFAR spectra together are described by only 100 parameters, whereas the complete nonparametric description of the second-order process statistics via the two autocorrelation matrices and the cross-correlation matrix requires $2N(N+1)/2 + N^2 = 2N^2 + N \approx 5 \cdot 10^5$ numbers.

VII. Conclusion

We introduced the VTFAR model for nonstationary vector (multivariate) processes. The VTFAR model represents nonstationarity and spectral correlation in terms of frequency (Doppler) shifts. It is an extension of both the vector AR model of stationary vector processes and the TFAR model of nonstationary scalar processes, and it can be viewed as a time-varying vector AR model using an exponential (Fourier) basis expansion. It is parsimonious for the practically relevant class of underspread vector processes, which are nonstationary vector processes with rapidly decaying correlation in time and frequency. Even greater parsimony is achieved with a banded VTFAR model in which only the correlation with a certain number of neighboring signals is modeled.

For estimating the parameters of the VTFAR model, we developed a method based on a system of linear equations that we termed the multichannel time-frequency Yule-Walker (TFYW) equations. Under the underspread assumption, these equations can be approximated by equations with two-level block-Toeplitz structure. For the order-recursive solution of these "underspread multichannel TFYW equations," we developed a fast algorithm that extends the Wax–Kailath algorithm [7] to the case of vector processes. We also proposed information criteria for estimating the VTFAR model order as well as the matrix bandwidth of the banded VTFAR model. Finally, the performance of the proposed parameter and order estimators was assessed through numerical results, and the application of the VTFAR model and estimators to nonstationary multivariate spectral analysis was demonstrated.

Appendix
Multichannel Wax-Kailath Algorithm

We will develop an efficient order-recursive algorithm for solving the underspread multichannel TFYW equations in the form $\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{A}} = -\boldsymbol{\mathsf{F}}$ [see (39)]. This algorithm exploits the two-level block-Toeplitz (2LBT) structure of the matrix $\boldsymbol{\mathcal{F}}$; in fact, it can be used for any equation with Hermitian 2LBT structure. It is a multichannel extension of the Wax–Kailath algorithm [7] and will therefore be referred to as the *multichannel Wax-Kailath algorithm*. For a 2LBT matrix $\boldsymbol{\mathcal{F}}$ of size $MP \times MP$ and block size $P \times P$, its complexity is $\mathcal{O}(M^2P^3)$.

A. Order-Recursive Scheme

The multichannel Wax-Kailath algorithm is order-recursive with respect to the delay order $m \in \{1,\dots,M\}$ and operates on block-Toeplitz matrices of fixed size P (in our case, $P=(2L+1)D$, i.e., the product of Doppler order $2L+1$ and signal dimension D). At the mth recursion, the algorithm solves the underspread multichannel TFYW equations for delay order $m+1$, using the previously calculated solution to the underspread multichannel TFYW equations for delay order m. The underspread multichannel TFYW equations for delay order m are given by the $mP \times mP$ system with $mP \times D$ right-hand side[6]

$$\boldsymbol{\mathcal{F}}_{(m)}\boldsymbol{\mathcal{A}}_{(m)} = -\mathsf{F}_{(m)}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{F}}_{(m)} &\triangleq \operatorname{toep}\{\boldsymbol{\mathcal{F}}_{-m+1},\dots,\boldsymbol{\mathcal{F}}_{m-1}\}\\ &= \begin{bmatrix} \boldsymbol{\mathcal{F}}_0 & \cdots & \boldsymbol{\mathcal{F}}_{m-1}\\ \vdots & \ddots & \vdots\\ \boldsymbol{\mathcal{F}}_{-m+1} & \cdots & \boldsymbol{\mathcal{F}}_0 \end{bmatrix}_{mP}\\ \mathsf{F}_{(m)} &\triangleq \begin{bmatrix} \mathsf{F}_1\\ \vdots\\ \mathsf{F}_m \end{bmatrix}_{mP\times D}\end{aligned}$$

with $\boldsymbol{\mathcal{F}}_\nu$ and F_ν as defined in (37) and (38), respectively, and $\boldsymbol{\mathcal{A}}_{(m)}$ a matrix of size $mP\times D$. Note that $\boldsymbol{\mathcal{F}}_{(M)}=\boldsymbol{\mathcal{F}}, \mathsf{F}_{(M)}=\mathsf{F}$, and $\boldsymbol{\mathcal{A}}_{(M)}=\boldsymbol{\mathcal{A}}$. Furthermore, $\boldsymbol{\mathcal{F}}_{(m)}$ and $\mathsf{F}_{(m)}$ are parts of $\boldsymbol{\mathcal{F}}$ and F, respectively; however, $\boldsymbol{\mathcal{A}}_{(m)}$ is not a part of $\boldsymbol{\mathcal{A}}$. At the last recursion ($m=M-1$, where M is the delay order of the VTFAR model), the solution $\boldsymbol{\mathcal{A}}_{(M)}=\boldsymbol{\mathcal{A}}$ to the underspread multichannel TFYW equations for delay order $M, \boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{A}}=-\mathsf{F}$, is obtained.

B. Block Transpose and Block Persymmetry

We first introduce some background. The multichannel Wax–Kailath algorithm uses the matrix operation of *block-transposition* [34] with respect to blocks of size $D\times D$. Block transposition of a $KD\times KD$ block matrix $\boldsymbol{X}$ with $D\times D$ blocks $\boldsymbol{X}_{k,k'}$ is defined as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{1,1} & \boldsymbol{X}_{1,2} & \cdots & \boldsymbol{X}_{1,K}\\ \boldsymbol{X}_{2,1} & \boldsymbol{X}_{2,2} & \cdots & \boldsymbol{X}_{2,K}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{X}_{K,1} & \boldsymbol{X}_{K,2} & \cdots & \boldsymbol{X}_{K,K}\end{bmatrix}_{KD}$$

$$\longrightarrow \boldsymbol{X}^B \triangleq \begin{bmatrix} \boldsymbol{X}_{1,1} & \boldsymbol{X}_{2,1} & \cdots & \boldsymbol{X}_{K,1}\\ \boldsymbol{X}_{1,2} & \boldsymbol{X}_{2,2} & \cdots & \boldsymbol{X}_{K,2}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{X}_{1,K} & \boldsymbol{X}_{2,K} & \cdots & \boldsymbol{X}_{K,K}\end{bmatrix}_{KD}.$$

Note that the individual blocks $\boldsymbol{X}_{k,k'}$ remain unchanged, except for their position within $\boldsymbol{X}$.

In analogy to the persymmetry of Toeplitz/block-Toeplitz matrices [7], it can be shown that the 2LBT matrices $\boldsymbol{\mathcal{F}}_{(m)}$ feature the *block-persymmetry property*

$$\boldsymbol{E}_{(m)}\boldsymbol{\mathcal{F}}_{(m)}\boldsymbol{E}_{(m)} = \boldsymbol{\mathcal{F}}^B_{(m)} \tag{51}$$

[6]Subscripts in parentheses indicate the delay order at the respective recursion. Furthermore, for the sake of clarity, we will often indicate matrix dimensions as subscripts without parentheses, in the form of $a\times b$ for a rectangular matrix and a (short for $a\times a$) for a square matrix.

with the "block-reflection" matrices

$$\boldsymbol{E}_{(m)} \triangleq \begin{bmatrix} \boldsymbol{0} & & \boldsymbol{E}\\ & \iddots & \\ \boldsymbol{E} & & \boldsymbol{0}\end{bmatrix}_{mP} \text{ where } \boldsymbol{E} \triangleq \begin{bmatrix} \boldsymbol{0} & & \mathbf{I}_D\\ & \iddots & \\ \mathbf{I}_D & & \boldsymbol{0}\end{bmatrix}_{P}. \tag{52}$$

Note that $\boldsymbol{E}_{(m)}\boldsymbol{E}_{(m)}=\mathbf{I}_{mP}$ and thus $\boldsymbol{E}^{-1}_{(m)}=\boldsymbol{E}_{(m)}$. Block-transposition and matrix inversion do not commute. Taking the inverse of both sides of (51) yields the relation (cf. [34])

$$\left(\boldsymbol{\mathcal{F}}^B_{(m)}\right)^{-1} = \boldsymbol{E}_{(m)}\boldsymbol{\mathcal{F}}^{-1}_{(m)}\boldsymbol{E}_{(m)} \tag{53}$$

or equivalently

$$\boldsymbol{\mathcal{F}}^{-1}_{(m)} = \boldsymbol{E}_{(m)}\left(\boldsymbol{\mathcal{F}}^B_{(m)}\right)^{-1}\boldsymbol{E}_{(m)}. \tag{54}$$

C. Recursions

Our goal is to calculate $\boldsymbol{\mathcal{A}}_{(m+1)}$, the solution to the $(m+1)$th-order underspread multichannel TFYW equations $\boldsymbol{\mathcal{F}}_{(m+1)}\boldsymbol{\mathcal{A}}_{(m+1)} = -\mathsf{F}_{(m+1)}$, assuming that $\boldsymbol{\mathcal{A}}_{(m)}$ is known. Exploiting its 2LBT structure, we can express $\boldsymbol{\mathcal{F}}_{(m+1)}$ as

$$\boldsymbol{\mathcal{F}}_{(m+1)} = \begin{bmatrix} \boldsymbol{\mathcal{F}}_{(m)} & \boldsymbol{T}_{(m)}\\ \boldsymbol{S}_{(m)} & \boldsymbol{\mathcal{F}}_0\end{bmatrix}_{(m+1)P} \tag{55}$$

with the $P\times mP$ matrix $\boldsymbol{S}_{(m)} \triangleq [\boldsymbol{\mathcal{F}}_{-m}\cdots\boldsymbol{\mathcal{F}}_{-1}]$, the $mP\times P$ matrix $\boldsymbol{T}_{(m)} \triangleq [\boldsymbol{\mathcal{F}}^B_m\cdots\boldsymbol{\mathcal{F}}^B_1]^B$, and some $P\times P$ matrix $\boldsymbol{\mathcal{F}}_0$. Furthermore, we have

$$\mathsf{F}_{(m+1)} = \begin{bmatrix} \mathsf{F}_{(m)}\\ \mathsf{F}_{m+1}\end{bmatrix}_{(m+1)P\times D}. \tag{56}$$

Applying the partitioned matrix inversion theorem [35, Sec. 2.9] to (55), we obtain for the inverse of $\boldsymbol{\mathcal{F}}_{(m+1)}$

$$\boldsymbol{\mathcal{F}}^{-1}_{(m+1)} = \begin{bmatrix} \boldsymbol{\mathcal{F}}^{-1}_{(m)} + \boldsymbol{W}_{(m)}\boldsymbol{\Theta}^{-1}_m\boldsymbol{V}_{(m)} & \boldsymbol{W}_{(m)}\boldsymbol{\Theta}^{-1}_m\\ \boldsymbol{\Theta}^{-1}_m\boldsymbol{V}_{(m)} & \boldsymbol{\Theta}^{-1}_m\end{bmatrix}_{(m+1)P} \tag{57}$$

with the $mP\times P$ matrix $\boldsymbol{W}_{(m)} \triangleq -\boldsymbol{\mathcal{F}}^{-1}_{(m)}\boldsymbol{T}_{(m)}$, the $P\times mP$ matrix $\boldsymbol{V}_{(m)} \triangleq -\boldsymbol{S}_{(m)}\boldsymbol{\mathcal{F}}^{-1}_{(m)}$, and the $P\times P$ matrix $\boldsymbol{\Theta}_m \triangleq \boldsymbol{\mathcal{F}}_0 - \boldsymbol{S}_{(m)}\boldsymbol{\mathcal{F}}^{-1}_{(m)}\boldsymbol{T}_{(m)}$. From (57) and (56), it follows that the $(m+1)$th-order solution $\boldsymbol{\mathcal{A}}_{(m+1)} = -\boldsymbol{\mathcal{F}}^{-1}_{(m+1)}\mathsf{F}_{(m+1)}$ can be calculated recursively as

$$\boldsymbol{\mathcal{A}}_{(m+1)} = \begin{bmatrix}\boldsymbol{\mathcal{A}}_{(m)}\\ \boldsymbol{0}\end{bmatrix} - \begin{bmatrix}\boldsymbol{W}_{(m)}\\ \mathbf{I}_{LD}\end{bmatrix}\boldsymbol{\Theta}^{-1}_m\left(\boldsymbol{V}_{(m)}\mathsf{F}_{(m)} + \mathsf{F}_{m+1}\right). \tag{58}$$

We now need to compute the matrices $\boldsymbol{W}_{(m)}, \boldsymbol{V}_{(m)}$, and $\boldsymbol{\Theta}_m$ recursively. To this end, we first rewrite them using (54):

$$\boldsymbol{W}_{(m)} = -\boldsymbol{E}_{(m)}\left(\boldsymbol{\mathcal{F}}^B_{(m)}\right)^{-1}\boldsymbol{E}_{(m)}\boldsymbol{T}_{(m)} \tag{59a}$$

$$\boldsymbol{V}_{(m)} = -\boldsymbol{S}_{(m)}\boldsymbol{E}_{(m)}\left(\boldsymbol{\mathcal{F}}^B_{(m)}\right)^{-1}\boldsymbol{E}_{(m)} \tag{59b}$$

$$\boldsymbol{\Theta}_m = \boldsymbol{\mathcal{F}}_0 - \boldsymbol{S}_{(m)}\boldsymbol{E}_{(m)}\left(\boldsymbol{\mathcal{F}}^B_{(m)}\right)^{-1}\boldsymbol{E}_{(m)}\boldsymbol{T}_{(m)}. \tag{59c}$$

For a recursive calculation of the matrix $(\boldsymbol{\mathcal{F}}^B_{(m)})^{-1}$ occurring in (59), we apply the partitioned matrix inversion theorem to [see (55)]

$$\boldsymbol{\mathcal{F}}^B_{(m+1)} = \begin{bmatrix} \boldsymbol{\mathcal{F}}^B_{(m)} & \boldsymbol{S}^B_{(m)}\\ \boldsymbol{T}^B_{(m)} & \boldsymbol{\mathcal{F}}^B_0\end{bmatrix}_{(m+1)P}.$$

This yields

$$\left(\boldsymbol{\mathcal{F}}^B_{(m+1)}\right)^{-1} = \begin{bmatrix} \left(\boldsymbol{\mathcal{F}}^B_{(m)}\right)^{-1} + \boldsymbol{N}_{(m)}\boldsymbol{\Delta}_m^{-1}\boldsymbol{M}_{(m)} & \boldsymbol{N}_{(m)}\boldsymbol{\Delta}_m^{-1} \\ \boldsymbol{\Delta}_m^{-1}\boldsymbol{M}_{(m)} & \boldsymbol{\Delta}_m^{-1} \end{bmatrix}_{(m+1)P} \quad (60)$$

with the $mP \times P$ matrix $\boldsymbol{N}_{(m)} \triangleq -(\boldsymbol{\mathcal{F}}^B_{(m)})^{-1}\boldsymbol{S}^B_{(m)}$, the $P \times mP$ matrix $\boldsymbol{M}_{(m)} \triangleq -\boldsymbol{T}^B_{(m)}(\boldsymbol{\mathcal{F}}^B_{(m)})^{-1}$, and the $P \times P$ matrix $\boldsymbol{\Delta}_m \triangleq \boldsymbol{\mathcal{F}}^B_0 - \boldsymbol{T}^B_{(m)}(\boldsymbol{\mathcal{F}}^B_{(m)})^{-1}\boldsymbol{S}^B_{(m)}$. Using (53), these matrices are expressed as

$$\boldsymbol{N}_{(m)} = -\boldsymbol{E}_{(m)}\boldsymbol{\mathcal{F}}^{-1}_{(m)}\boldsymbol{E}_{(m)}\boldsymbol{S}^B_{(m)} \quad (61a)$$

$$\boldsymbol{M}_{(m)} = -\boldsymbol{T}^B_{(m)}\boldsymbol{E}_{(m)}\boldsymbol{\mathcal{F}}^{-1}_{(m)}\boldsymbol{E}_{(m)} \quad (61b)$$

$$\boldsymbol{\Delta}_m = \boldsymbol{\mathcal{F}}^B_0 - \boldsymbol{T}^B_{(m)}\boldsymbol{E}_{(m)}\boldsymbol{\mathcal{F}}^{-1}_{(m)}\boldsymbol{E}_{(m)}\boldsymbol{S}^B_{(m)}. \quad (61c)$$

Now inserting (60) into (59) (with m incremented by one), we obtain

$$\boldsymbol{W}_{(m+1)} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{W}_{(m)} \end{bmatrix} - \begin{bmatrix} \boldsymbol{E} \\ \tilde{\boldsymbol{N}}_{(m)} \end{bmatrix} \boldsymbol{\Delta}_m^{-1}\left(\boldsymbol{M}_{(m)}\tilde{\boldsymbol{T}}_{(m)} + \boldsymbol{E}\boldsymbol{\mathcal{F}}_{m+1}\right) \quad (62a)$$

$$\boldsymbol{V}_{(m+1)} = \left[\boldsymbol{0}\, \boldsymbol{V}_{(m)}\right] - \left(\tilde{\boldsymbol{S}}_{(m)}\boldsymbol{N}_{(m)} + \boldsymbol{\mathcal{F}}_{-m-1}\boldsymbol{E}\right)\boldsymbol{\Delta}_m^{-1}\left[\boldsymbol{E}\, \tilde{\boldsymbol{M}}_{(m)}\right] \quad (62b)$$

$$\boldsymbol{\Theta}_{m+1} = \boldsymbol{\Theta}_m - \left(\tilde{\boldsymbol{S}}_{(m)}\boldsymbol{N}_{(m)} + \boldsymbol{\mathcal{F}}_{-m-1}\boldsymbol{E}\right)\boldsymbol{\Delta}_m^{-1} \times \left(\boldsymbol{M}_{(m)}\tilde{\boldsymbol{T}}_{(m)} + \boldsymbol{E}\boldsymbol{\mathcal{F}}_{m+1}\right) \quad (62c)$$

with the shorthand notations $\tilde{\boldsymbol{S}}_{(m)} \triangleq \boldsymbol{S}_{(m)}\boldsymbol{E}_{(m)}, \tilde{\boldsymbol{T}}_{(m)} \triangleq \boldsymbol{E}_{(m)}\boldsymbol{T}_{(m)}, \tilde{\boldsymbol{N}}_{(m)} \triangleq \boldsymbol{E}_{(m)}\boldsymbol{N}_{(m)}, \tilde{\boldsymbol{M}}_{(m)} \triangleq \boldsymbol{M}_{(m)}\boldsymbol{E}_{(m)}$ and with $\boldsymbol{E}$ as defined in (52). We note that $\tilde{\boldsymbol{S}}_{(m)}, \tilde{\boldsymbol{T}}_{(m)}, \tilde{\boldsymbol{N}}_{(m)}$, and $\tilde{\boldsymbol{M}}_{(m)}$ can all be obtained recursively through simple concatenation operations, e.g., $\tilde{\boldsymbol{S}}_{(m+1)} = [\tilde{\boldsymbol{S}}_{(m)}\ \boldsymbol{\mathcal{F}}_{-m-1}\boldsymbol{E}]$. Similarly, plugging (57) into (61) yields

$$\boldsymbol{N}_{(m+1)} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{N}_{(m)} \end{bmatrix} - \begin{bmatrix} \boldsymbol{E} \\ \tilde{\boldsymbol{W}}_{(m)} \end{bmatrix} \boldsymbol{\Theta}_m^{-1}\left(\boldsymbol{V}_{(m)}\tilde{\boldsymbol{S}}^B_{(m)} + \boldsymbol{E}\boldsymbol{\mathcal{F}}^B_{-m-1}\right) \quad (63a)$$

$$\boldsymbol{M}_{(m+1)} = [\boldsymbol{0}\ \boldsymbol{M}_{(m)}] - \left(\tilde{\boldsymbol{T}}^B_{(m)}\boldsymbol{W}_{(m)} + \boldsymbol{\mathcal{F}}^B_{m+1}\boldsymbol{E}\right)\boldsymbol{\Theta}_m^{-1}\left[\boldsymbol{E}\, \tilde{\boldsymbol{V}}_{(m)}\right] \quad (63b)$$

$$\boldsymbol{\Delta}_{m+1} = \boldsymbol{\Delta}_m - \left(\tilde{\boldsymbol{T}}^B_{(m)}\boldsymbol{W}_{(m)} + \boldsymbol{\mathcal{F}}^B_{m+1}\boldsymbol{E}\right)\boldsymbol{\Theta}_m^{-1} \times \left(\boldsymbol{V}_{(m)}\tilde{\boldsymbol{S}}^B_{(m)} + \boldsymbol{E}\boldsymbol{\mathcal{F}}^B_{-m-1}\right) \quad (63c)$$

with $\tilde{\boldsymbol{W}}_{(m)} \triangleq \boldsymbol{E}_{(m)}\boldsymbol{W}_{(m)}$ and $\tilde{\boldsymbol{V}}_{(m)} \triangleq \boldsymbol{V}_{(m)}\boldsymbol{E}_{(m)}$.

D. Algorithm Summary

We now have developed all recursion steps of the multichannel Wax-Kailath algorithm. The algorithm can finally be summarized as follows.

- *Initialization*: set

$$\boldsymbol{W}_{(1)} = -\boldsymbol{\mathcal{F}}_0^{-1}\boldsymbol{\mathcal{F}}_1$$
$$\boldsymbol{V}_{(1)} = -\boldsymbol{\mathcal{F}}_{-1}\boldsymbol{\mathcal{F}}_0^{-1}$$
$$\boldsymbol{\Theta}_1 = \boldsymbol{\mathcal{F}}_0 - \boldsymbol{\mathcal{F}}_{-1}\boldsymbol{\mathcal{F}}_0^{-1}\boldsymbol{\mathcal{F}}_1$$
$$\boldsymbol{N}_{(1)} = -\left(\boldsymbol{\mathcal{F}}^B_0\right)^{-1}\boldsymbol{\mathcal{F}}^B_{-1}$$
$$\boldsymbol{M}_{(1)} = -\boldsymbol{\mathcal{F}}^B_1\left(\boldsymbol{\mathcal{F}}^B_0\right)^{-1}$$
$$\boldsymbol{\Delta}_1 = \boldsymbol{\mathcal{F}}^B_0 - \boldsymbol{\mathcal{F}}^B_1\left(\boldsymbol{\mathcal{F}}^B_0\right)^{-1}\boldsymbol{\mathcal{F}}^B_{-1}$$
$$\boldsymbol{\mathcal{A}}_{(1)} = -\boldsymbol{\mathcal{F}}_0^{-1}\mathsf{F}_1.$$

- *Recursion*: perform the update (58) for $m = 1, \ldots, M-1$ and the updates (62) and (63) for $m = 1, \ldots, M-2$. At the final recursion ($m = M-1$), (58) yields the solution $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}_{(M)}$ to the underspread multichannel TFYW equations $\boldsymbol{\mathcal{F}}\boldsymbol{\mathcal{A}} = -\mathsf{F}$. Note that (62) and (63) need not be calculated at the final recursion.

E. Special Cases

The multichannel Wax–Kailath algorithm simplifies for some special cases.

- For $D = 1$ (i.e., $\boldsymbol{x}[n]$ is a scalar TFAR process), the 2LBT matrix $\boldsymbol{\mathcal{F}}$ reduces to a Toeplitz/block-Toeplitz matrix and block-transposition reduces to ordinary transposition. Since ordinary transposition commutes with matrix inversion, a comparison of (57) and (60) shows that for $D = 1$, we have $\boldsymbol{\Theta}_{(m)} = \boldsymbol{\Delta}^T_{(m)}, \boldsymbol{V}_{(m)} = \boldsymbol{N}^T_{(m)}$, and $\boldsymbol{W}_{(m)} = \boldsymbol{M}^T_{(m)}$. With these simplifications, the Wax–Kailath algorithm [7] is reobtained.
- For $L = 0$ (i.e., $\boldsymbol{x}[n]$ is a stationary vector AR process), $\boldsymbol{\mathcal{F}}$ is a block-Toeplitz matrix. Here, the multichannel Levinson algorithm [2] is reobtained.
- Finally, for $L = 0$ and $D = 1$ (i.e., $\boldsymbol{x}[n]$ is a scalar and stationary AR process), $\boldsymbol{\mathcal{F}}$ is Toeplitz and our algorithm reduces to the Levinson–Durbin algorithm [1], [2].

ACKNOWLEDGMENT

The authors would like to thank Prof. M. Deistler, Prof. K. Gröchenig, and Prof. B. M. Pötscher for illuminating discussions and helpful comments.

REFERENCES

[1] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[2] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[3] M. Jachan and G. Matz, "Nonstationary vector AR modeling of wireless channels," in *Proc. IEEE SPAWC*, New York, Jun. 2005, pp. 648–652.
[4] W. Hesse, E. Möller, M. Arnold, and B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *J. Neurosci. Methods*, vol. 124, pp. 27–44, 2003.
[5] E. A. Robinson and S. Treitel, *Geophysical Signal Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
[6] M. Jachan, G. Matz, and F. Hlawatsch, "Time-frequency ARMA models and parameter estimators for underspread nonstationary random processes," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4366–4381, Sep. 2007.
[7] M. Wax and T. Kailath, "Efficient inversion of Toeplitz-block Toeplitz matrix," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, pp. 1218–1221, Oct. 1983.
[8] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. New York: Wiley, 1988.
[9] T. Subba Rao, "The fitting of non-stationary time-series models with time-dependent parameters," *J. Roy. Stat. Soc. Ser. B*, vol. 32, no. 2, pp. 312–322, 1970.
[10] L. A. Liporace, "Linear estimation of nonstationary signals," *J. Acoust. Soc. Amer.*, vol. 58, pp. 1288–1295, Dec. 1975.

[11] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, pp. 899–911, Aug. 1983.
[12] K. B. Eom, "Time-varying autoregressive modeling of high range resolution radar signatures for classification of noncooperative targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 974–988, Jul. 1999.
[13] M. Niedźwiecki, *Identification of Time-Varying Processes*. New York: Wiley, 2000.
[14] G. Matz and F. Hlawatsch, "Nonstationary spectral analysis based on time-frequency operator symbols and underspread approximations," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1067–1086, Mar. 2006.
[15] G. Matz and F. Hlawatsch, "Time-varying power spectra of nonstationary random processes," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed. Oxford, U.K.: Elsevier, 2003, ch. 9.4, pp. 400–409.
[16] W. Kozek, "On the underspread/overspread classification of nonstationary random processes," in *Proc. Int. Conf. Ind. Appl. Math.*, K. Kirchgässner, O. Mahrenholtz, and R. Mennicken, Eds., Mathematical Research, 1996, vol. 3, pp. 63–66.
[17] G. Matz and F. Hlawatsch, "Time-frequency transfer function calculus (symbolic calculus) of linear time-varying systems (linear operators) based on a generalized underspread theory," *J. Math. Phys. (Special Issue on Wavelet and Time-Frequency Analysis)*, vol. 39, pp. 4041–4071, Aug. 1998.
[18] K. Gröchenig and M. Leinert, "Wiener's lemma for twisted convolution and Gabor frames," *J. Amer. Math. Soc.*, vol. 17, no. 1, pp. 1–18, 2004.
[19] K. Gröchenig, private communication, 2009.
[20] L. A. Zadeh, "Frequency analysis of variable networks," *Proc. IRE*, vol. 76, pp. 291–299, Mar. 1950.
[21] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, pp. 360–393, 1963.
[22] G. Matz, "A time-frequency calculus for time-varying systems and nonstationary processes with applications," Ph.D. dissertation, Vienna Univ. of Technol., Vienna, Austria, Nov. 2000.
[23] M. Jachan, "Time-frequency autoregressive moving-average modeling," Ph.D. dissertation, Vienna Univ. of Technol., Vienna, Austria, Jul. 2006.
[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[25] G. Matz and F. Hlawatsch, "Wigner distributions (nearly) everywhere: Time-frequency analysis of signals, systems, random processes, signal spaces, and frames," *Signal Process.*, vol. 83, no. 7, pp. 1355–1378, 2003.
[26] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, pp. 181–208, Feb. 1993.
[27] M. Jachan, G. Matz, and F. Hlawatsch, "TFARMA models: Order estimation and stabilization," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Philadelphia, PA, Mar. 2005, vol. IV, pp. 301–304.
[28] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, pp. 716–723, Dec. 1974.
[29] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 6, pp. 465–471, 1978.
[30] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley, 2002.
[31] London Bat Group. [Online]. Available: http://www.londonbats.org.uk
[32] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. San Diego, CA: Academic, 1999.
[33] F. Hlawatsch and P. Flandrin, "The interference structure of the Wigner distribution and related time-frequency signal representations," in *The Wigner Distribution—Theory and Applications in Signal Processing*, W. Mecklenbräuker and F. Hlawatsch, Eds. Amsterdam, The Netherlands: Elsevier, 1997, pp. 59–133.
[34] H. Akaike, "Block Toeplitz matrix inversion," *SIAM J. Appl. Math.*, vol. 24, pp. 234–241, Mar. 1973.
[35] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

Michael Jachan received the M.Sc. degree in signal processing and the Ph.D. degree in telecommunications from the Vienna University of Technology in June 2001 and June 2006, respectively.

From 2001 to 2002 he was with the ftw. Telecommunications Research Center Vienna, where he was working on an xDSL system simulator. He was a Research and Teaching Assistant at the Institute of Telecommunications and Radio-Frequency Engineering, Vienna University of Technology, from 2002 to 2006. From 2006 to 2009, he was with the FDM, Freiburg Center for Data Analysis and Modeling, University of Freiburg, where he was active in a cooperation between the Physics Department and the Department of Neurology. Since April 2009, he has been with Brain Products GmbH working on algorithm implementation and graphics design. His research interests include statistical inference, time-frequency signal processing, and EEG processing.

Gerald Matz (S'95–M'01–SM'07) received the Dipl.-Ing. and Dr.Techn. degrees in electrical engineering in 1994 and 2000, respectively, and the Habilitation degree for Communication Systems in 2004, all from the Vienna University of Technology, Austria.

Since 1995 he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, where he currently holds a tenured position as Associate Professor. From March 2004 to February 2005 he was on leave as an Erwin Schrödinger Fellow with the Laboratoire des Signaux et Systèmes, Ecole Supérieure d'Electricité, France. During summer 2007 he was a guest researcher with the Communication Theory Lab at ETH Zurich, Switzerland. His research interests include wireless communications, statistical signal processing, and information theory.

Prof. Matz has directed or actively participated in several research projects funded by the Austrian Science Fund (FWF) and by the European Union. He has published about 120 papers in international journals, conference proceedings, and edited books. He serves as a member of the IEEE Signal Processing Society (SPS) Technical Committee on Signal Processing for Communications and Networking and as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and for *Signal Processing*. From 2004 to 2008 he was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He was Technical Program Co-Chair of EUSIPCO 2004 and has been a member of the Program Committee of numerous international conferences. In 2006, he received the Kardinal Initzer Most Promising Young Investigator Award.

Franz Hlawatsch (S'85–M'88–SM'00) received the Diplom-Ingenieur, Dr.Techn., and Univ.-Dozent (habilitation) degrees in electrical engineering/signal processing from the Vienna University of Technology, Vienna, Austria in 1983, 1988, and 1996, respectively.

Since 1983, he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, where he is currently an Associate Professor. During 1991–1992, as a recipient of an Erwin Schrödinger Fellowship, he spent a sabbatical year with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI. In 1999, 2000, and 2001, he held one-month Visiting Professor positions with INP/ENSEEIHT/TeSA, Toulouse, France and IRCCyN, Nantes, France. He (co)authored a book, a review paper that appeared in the *IEEE Signal Processing Magazine*, about 170 refereed scientific papers and book chapters, and three patents. He coedited two books. His research interests include signal processing for wireless communications, statistical signal processing, and compressive signal processing.

Prof. Hlawatsch was Technical Program Co-Chair of EUSIPCO 2004 and served on the technical committees of numerous IEEE conferences. From 2003 to 2007, he served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and since 2008, he has served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. Since 2004, he has been a member of the IEEE SPCOM Technical Committee. He is coauthor of a paper that won an IEEE Signal Processing Society Young Author Best Paper Award.