Two-Dimensional Multivariate Parametric Models for Radar Applications—Part I: Maximum-Entropy Extensions for Toeplitz-Block Matrices

Yuri I. Abramovich, *Senior Member, IEEE*, Ben A. Johnson, *Student Member, IEEE*, and Nicholas K. Spencer

***Abstract*—In a series of two papers, a new class of parametric models for two-dimensional multivariate (matrix-valued, space-time) adaptive processing is introduced. This class is based on the maximum-entropy extension and/or completion of partially specified matrix-valued Hermitian covariance matrices in both the space and time dimensions. This first paper considers the more restricted class of Toeplitz Hermitian covariance matrices that model stationary clutter. If the clutter is stationary only in time then we deal with a Toeplitz-block matrix, whereas clutter that is stationary in time and space is described by a Toeplitz-block-Toeplitz matrix. We first derive exact expressions for this new class of 2-D models that act as approximations for the unknown true covariance matrix. Second, we propose suboptimal (but computationally simpler) relaxed 2-D time-varying autoregressive models (“relaxations”) that directly use the non-Toeplitz Hermitian sample covariance matrix. The high efficiency of these parametric models is illustrated by simulation results using true ground-clutter covariance matrices provided by the DARPA KASSPER Dataset 1, which is a trusted phenomenological airborne radar model, and a complementary AFRL dataset.**

***Index Terms*—Adaptive processing, autoregressive, stationary interference, time-varying.**

I. Introduction and Background

IN modern radar applications, adaptive processing is often performed on two-dimensional (2-D) data streams. For example, efficient ground-clutter cancellation in airborne radar can only be achieved using space-time adaptive processing (STAP) on data that is collected simultaneously by M spatially distinct receive channels (antenna array sensors) over N temporally distinct waveform repetitions (“slow time”) that comprise a coherent processing interval (CPI), which is the fundamental set of data that a radar system processes. The main problem with the use of 2-D STAP algorithms is the need to adaptively estimate an MN-variate covariance matrix using, generally, a very limited number τ of training data samples. In airborne radar, there are very few range bins (the “fast-time” dimension) on the ground that produce sufficiently homogeneous backscattered signals due to both geometric considerations and terrain variations [1]–[3]. When no *a priori* assumptions regarding any further properties of the MN-variate Hermitian covariance matrix are made, the generic sample covariance matrix estimate is used in signal processing algorithms that are based on sample-matrix inversion (SMI). Unfortunately, the number of independent identically distributed (i.i.d.) training samples τ must be at least MN (otherwise the sample matrix is rank deficient), and moreover must be at least $2MN$ in order to ensure that the average signal-to-interference-plus-noise ratio (SINR) losses of the STAP filter are below 3 dB, relative to the optimum (clairvoyant) Wiener filter [4]. In most airborne radar scenarios, such a large number $\tau = 2MN$ of i.i.d. (i.e., sufficiently homogeneous) training samples is simply not available.

However, in most practical cases, some additional *a priori* information on the properties of the ground-clutter covariance matrix is available. In certain applications (and as modeled in DARPA’s high-fidelity airborne side-looking radar KASSPER Dataset 1 [5], [6]), the ground clutter is stationary in slow time. In this case, the (range-dependent) MN-variate true covariance matrices have a Toeplitz-block structure. Note that we carefully (but possibly unconventionally) use the term *Toeplitz-block matrix* for a blockwise-Toeplitz matrix composed of unstructured blocks, as distinct from a *block-Toeplitz matrix* which describes an overall unstructured matrix composed of individually Toeplitz blocks. (Of course, the term *Toeplitz-block-Toeplitz* means an overall Toeplitz matrix composed of Toeplitz blocks.)

Due to the embedded (block/matrix-valued) nature of the matrices we deal with, we also need to introduce our notation more carefully than usual. We use bold-face lower-case letters for vectors, block upper-case letters for block matrices, upper-case letters for simple matrices, and corresponding lower-case letters for the matrix elements. We use R and $\mathbb{R}$ to denote a general matrix, with T and $\mathbb{T}$ for a Toeplitz matrix. Z, $\boldsymbol{z}$, and z all refer to *a priori* unknown variables. As mentioned, a typical covariance (hence Hermitian) matrix $\mathbb{T}$ in this problem is also an MN-variate Toeplitz-block matrix, so that the macroscopic

Manuscript received October 30, 2007; revised July 10, 2008. First published August 29, 2008; current version published October 15,2008. The associate editor coordinating the review of this paper and approving it for publication was Dr. Sergiy A. Vorobyov. This work was funded in part under DSTO/RLM R&D Collaborative Agreement 290905. The work of N. K. Spencer was supported by DSTO through the Centre of Expertise in Phased Array and Microwave Radar Systems at The University of Adelaide. Some results of this paper were presented at the Fifteenth Annual Workshop on Adaptive Sensor Array Processing, MIT Lincoln Laboratory, Lexington, MA, 2007.

Y. I. Abramovich is with the Intelligence, Surveillance and Reconnaissance Division, Defence Science and Technology Organisation (DSTO), Adelaide 5111, Australia (e-mail: Yuri.Abramovich@dsto.defence.gov.au).

B. A. Johnson is with Lockheed Martin, Pty., Ltd., and the University of South Australia, Adelaide 5111, Australia (e-mail: ben.a.johnson@ieee.org).

N. K. Spencer is with Adelaide Research & Innovation, Pty., Ltd. (ARI), Adelaide 5000, Australia (e-mail: Nick.Spencer@adelaide.edu.au).

Digital Object Identifier 10.1109/TSP.2008.929868

structure is an $N \times N$ Toeplitz (matrix-valued) matrix, while the microscopic structure (each block) is an $M \times M$ (unstructured/general) matrix:

$$\begin{aligned}\mathbb{T} &\equiv \left\{r_{j-k}^{pq}\right\}_{j,k=1,\ldots,N}^{p,q=1,\ldots,M} \\ &\equiv \{R_{j-k}\}_{j,k=1,\ldots,N} \in \mathcal{T}_H^{MN\times MN} \\ R_{k-j}^H = R_{j-k} &\equiv \left\{r_{j-k}^{pq}\right\}^{p,q=1,\ldots,M} \in \mathcal{C}^{M\times M} \end{aligned} \tag{1}$$

where $\mathcal{C}^{m\times m}$ and $\mathcal{T}_H^{m\times m}$ are respectively the classes of complex and complex Hermitian Toeplitz m-variate matrices. In this notation, p and q are spatial indexes with M being the number of antenna sensors, while j and k are temporal ("slow-time") indexes with N being the number of waveform repetitions. The notation r_{j-k}^{pq} emphasizes that the elements (covariance lags) r are Toeplitz in the temporal dimension only. The default size of a matrix is $M \times M$. The default size of a block matrix is $MN \times MN$, unless we indicate otherwise by a subscript, e.g., $\mathbb{T}_{m(N+1)} \in \mathcal{T}^{m(N+1)\times m(N+1)}$; moreover, the first factor in the subscript indicates the "inner" dimension, with the second corresponding to the "outer," so that $\mathbb{T}_{m(N+1)}$ comprises $(N+1)\times(N+1)$ lots of m-variate blocks.

Another simplification occurs when we consider radars whose receive antenna is a uniform linear array (ULA), in which case the resulting MN-variate covariance matrix is Toeplitz-block-Toeplitz:

$$\begin{aligned}\mathbb{T} \equiv &\{T_{j-k}\}_{j,k=1,\ldots,N} \in \mathcal{T}_H^{MN\times MN} \\ T_{j-k} \equiv &\left\{r_{j-k}^{p-q}\right\}^{p,q=1,\ldots,M} \in \mathcal{T}^{M\times M}\end{aligned} \tag{2}$$

where $\mathcal{T}^{m\times m}$ is the class of m-variate complex Toeplitz matrices.

Restricting the class of admissible covariance matrices from Hermitian-block matrices to Toeplitz-block matrices to Toeplitz-block-Toeplitz matrices somewhat reduces the training-sample requirements from $\tau = 2MN$ [7]. Yet in most practical cases, the available number of suitable training samples τ is still too small. One remedy to the problem of sample-support shortage is a still further restriction of the admissible set of estimated covariance matrices to some parametric family, whose number of free parameters is sufficiently small to potentially decrease the required i.i.d. sample support.

Naturally, if the chosen parametric model is not accurately supported by the phenomenology of the underlying process, such as airborne radar ground clutter, it must be treated circumspectly. Even the "best" parametric approximation of a "true" covariance matrix will inevitably cause some STAP performance degradation with respect to the clairvoyant Wiener filter. In particular, the MN-variate "phenomenological" covariance matrices of the KASSPER dataset, apart from being Toeplitz-block matrices, do not exactly fit any known parametric model with a small number of free parameters.

Therefore, selecting an appropriate parametric model is equivalent to finding the best tradeoff between performance losses due to the mismatch between the true covariance matrix and its parametric approximation, and performance losses due to using a limited number of training samples ("finite sample support"). While the latter "stochastic losses" are expected to decrease as more parametric restrictions are imposed, the former "model-mismatch losses" should increase. A meaningful assessment and comparison of parametric model-based STAP could be performed only if a trustworthy phenomenological clairvoyant covariance matrix of ground clutter is provided. For this reason, our study relies upon the KASSPER Dataset 1, an elaborate high-fidelity phenomenological clutter model for a particular side-looking airborne radar. The KASSPER dataset uses the Splatter, Clutter and Target Signal model (SCATS) [8], which incorporates detailed modeling of range-specific clutter, interference and terrain-scattered interference, based on three-dimensional terrain datasets and rigorous terrain shadowing, scattering and diffraction modeling. An L-band radar front-end is modeled: $N = 32$ waveform repetitions per CPI and a nominally ULA with $M = 11$ sensors (subarrays), including three different types of antenna errors: physical positioning errors, differences in subarray beam-patterns, and calibration errors. These antenna errors mean that the MN-variate covariance matrices, provided for 1000 consecutive range bins, are Toeplitz-block matrices, not Toeplitz-block-Toeplitz ones.

Unfortunately, the KASSPER software does not have an option to "switch off" antenna errors [9], and so to validate our parametric models devised for a Toeplitz-block-Toeplitz covariance matrix, we used a complementary ground-clutter model from the U.S. Air Force Research Laboratory (AFRL) [10]. This model was derived for the same KASSPER radar scenario, but adopts slightly simplified electromagnetic simulation algorithms (e.g., no shadowing), and allows for an *ideal* ULA geometry.

Let the ith observed radar data "snapshot" be $\mathbb{x}_i \in \mathcal{C}^{MN\times 1}$, $(i = 1,\ldots,\tau)$, where $\mathbb{x}_i \sim \mathcal{CN}_{MN}(0,\mathbb{T})$, i.e., is an MN-variate random complex (circular) Gaussian training vector, with zero mean and covariance $\mathbb{T}$. Given the true covariance matrix $\mathbb{T}$ (for a particular range bin), the optimum Wiener STAP filter is $\mathbb{w}(\theta,\omega) = \mathbb{T}^{-1}\mathbb{s}(\theta,\omega)$, where $\mathbb{s}(\theta,\omega) \in \mathcal{C}^{MN\times 1}$ is the space-time "stacked steering vector" for a target located in the azimuthal direction θ and moving with the Doppler frequency ω. (We keep the elevation angle ϕ constant in this study; the steering vector $\mathbb{s}$ is calculated in the standard way [3].) For any parametric estimate (i.e., model approximation) of the covariance matrix $\hat{\mathbb{R}}(\Omega)$, where Ω is the set of parameters, the STAP filter is calculated as $\hat{\mathbb{w}}(\theta,\omega,\Omega) = \hat{\mathbb{R}}^{-1}(\Omega)\mathbb{s}(\theta,\omega)$. For this study, we have chosen the performance criterion for assessing and comparing this parametric model to be the SINR loss factor with respect to the optimum filter:

$$\eta(\theta,\omega) = \frac{\mathbb{s}^H\mathbb{T}^{-1}\mathbb{s}\mathbb{s}^H\hat{\mathbb{R}}^{-1}\mathbb{T}\hat{\mathbb{R}}^{-1}\mathbb{s}}{[\mathbb{s}^H\hat{\mathbb{R}}^{-1}\mathbb{s}]^2}. \tag{3}$$

If the parametric model $\mathbb{R}(\Omega)$ is calculated for the true (clairvoyant) covariance matrix $\mathbb{T}$, then the resulting loss factor η is the SINR performance degradation associated with the model mismatch only. More usually, when $\hat{\mathbb{R}}$ is an "adaptive" estimate, i.e., $\hat{\mathbb{R}}(\Omega) \equiv \hat{\mathbb{R}}(\Omega, \mathbb{x}_1, \ldots, \mathbb{x}_\tau)$, the loss factor η is a random number that incorporates both "model-mismatch" and "stochastic" losses.

Given the KASSPER or AFRL true covariance matrix $\mathbb{T}$, we are able to conduct Monte Carlo STAP simulations using a set of τ ideal (perfectly homogeneous) training samples, which are generated as

$$\mathbb{x}_i = \mathbb{T}^{\frac{1}{2}}\mathbb{C}_i, \quad \mathbb{C}_i \sim \mathcal{C}N_{MN}(0, I_{MN}) \tag{4}$$

where I_{MN} is the MN-variate identity matrix, as usual.

When considering STAP applications for a process that is stationary in slow time, an obvious parametric family is the multivariate autoregressive model $\mathrm{AR}_M(n)$ [11]–[13] with a relatively small order $(n \ll N)$. While this autoregressive (AR) clutter model does not have a straight-forward phenomenological (physical) basis, it has been traditionally used in univariate moving-target detection applications, and theoretically justifies the simplest k-stage moving-target indicator (MTI) filters [14]. $\mathrm{AR}_M(n)$ modeling leads to a two-fold reduction in the minimum sample-support requirement. First, the MN-variate Toeplitz-block matrix $\mathbb{T}$ is uniquely defined by its principal $M(n+1)$-variate block (the top-left corner, say)

$$\mathbb{T}_{M(n+1)} = \{R_{j-k}\}_{j,k=1,\ldots,n+1}. \tag{5}$$

Second, the ergodicity principle allows us to generate, *from a single MN-variate training sample*, $(N-n)$ different $M(n+1)$-variate identically distributed training samples to estimate the matrix $\mathbb{T}_{M(n+1)}$ by "sliding-window" averaging (temporal smoothing) over the N-long CPI. While these training samples are not statistically independent, their homogeneity makes them a valuable contribution to the $M(n+1)$-variate covariance matrix estimate $\hat{\mathbb{T}}_{M(n+1)}$. If the traditional sample matrix $\hat{\mathbb{R}}_{M(n+1)}$ is considered (as a sufficient statistic) for estimating the Toeplitz-block matrix $\mathbb{T}_{M(n+1)}$, then the minimum number of MN-variate training samples (range bins) $\mathbb{x}_i (i = 1, \ldots, \tau)$ is [15]

$$\tau_{\min}^{\mathrm{AR}_M(n)} = \left\lceil \frac{M(n+1)}{N-n} \right\rceil \tag{6}$$

where $\lceil\cdot\rceil$ denotes the smallest integer greater than the argument (the "ceiling" function).

For clutter that is stationary both in slow time (i.e., a strictly periodic radar waveform) and space (i.e., a perfectly ULA), whose MN-variate covariance matrix is Toeplitz-block-Toeplitz in structure, the natural choice for STAP applications is the 2-D autoregressive model $\mathrm{AR}(m,n)$ [16], [17]. For this model, the covariance matrix is uniquely specified by the $(m+1)(n+1)$-variate matrix [18]

$$\mathbb{T}_{(m+1)(n+1)} \equiv \left\{ r_{j-k}^{p-q} \right\}_{j,k=1,\ldots,n+1}^{p,q=1,\ldots,m+1} \tag{7}$$

and now $(M-m)(N-n)$ different (but dependent) $(m+1)(n+1)$-variate training samples can be generated from a single MN-variate one by sliding-window averaging over both the slow-time and space dimensions. Hence, the minimum sample support for the $\mathrm{AR}(m,n)$ model is

$$\tau_{\min}^{\mathrm{AR}(m,n)} = \left\lceil \frac{(m+1)(n+1)}{(M-m)(N-n)} \right\rceil \tag{8}$$

which means that *even a single training sample* $(\tau = 1)$ may be sufficient, if $m \ll M$ and $n \ll N$.

Despite promising results for $\mathrm{AR}_M(n)$ and $\mathrm{AR}(m,n)$ models [12], a number of important practical STAP application issues have not yet been addressed. Specifically, regarding the $\mathrm{AR}_M(n)$ model as follows.

(1.1) This model imposes parametric restrictions over only the temporal domain by limiting the AR order $(n \ll N)$. If similarly to the ideal ULA case, it is possible to impose parametric (order) restrictions over the spatial domain (for an arbitrary antenna array geometry), then a further sample-support reduction could be expected. We call this a "mixed AR model" because it is AR in its temporal dimension, with the spatial dimension having some sort of restriction.

(1.2) In most studies on parametric STAP [12], estimation of a stable (causal) $\mathrm{AR}_M(n)$ model is considered given a set of τ i.i.d. training samples. The various estimation procedures described in [16], for example, are not optimal in any particular sense (such as maximum likelihood), yet are computationally involved. Given a positive-definite (p.d.) Hermitian-block sample matrix $\hat{\mathbb{R}}_{M(n+1)}$, these methods reconstruct an $M(n+1)$-variate p.d. Toeplitz-block matrix that then uniquely specifies the $\mathrm{AR}_M(n)$ MN-variate covariance matrix estimate $\hat{\mathbb{T}}^{(n)}$. However, the possibility of reconstructing a MN-variate p.d. Hermitian-block matrix $\hat{\mathbb{R}}^{(n)}$ from the same sample matrix $\hat{\mathbb{R}}_{M(n+1)}$, that is efficient in terms of SINR performance, remains unexplored. If successful (as this paper will demonstrate), such an approach involves a deliberate inconsistency: stationarity in slow time (and therefore Toeplitz-block structure of the clutter covariance matrix) is exploited by sliding-window averaging over the CPI when estimating the $M(n+1)$-variate sample matrix $\hat{\mathbb{R}}_{M(n+1)}$; but this property is deliberately abandoned in our final estimate $\hat{\mathbb{R}}^{(n)}$, which then prompts us to use the term "relaxation" (for a relaxed model). Clearly, if the development of the "mixed" AR model in *(1.1)* is successful, a similar "relaxation" may also be available (and desirable) for this Toeplitz-block matrix. Finally, for any true MN-variate covariance matrix, as supplied in KASSPER or AFRL datasets, it remains unclear whether a rigorous (causal) $\mathrm{AR}_M(n)$ model with $n \ll N$ will always outperform its corresponding relaxation in the SINR sense.

Note that the "relaxation" idea is already known in the field of adaptive signal processing: the conventional (Hermitian) sample covariance matrix estimate (or its diagonally loaded version) is routinely used when applying an adaptive matched filter (AMF) or generalized likelihood-ratio (GLRT) detector [19], [20], despite often knowing properties of the underlying covariance matrix (e.g., Toeplitzness).

Regarding the causal 2-D $\mathrm{AR}(m,n)$ model, the as-yet unaddressed issues are as follows.

(2.1) Fairly recent results of Woerdeman *et al.* [18] have proven that an $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz matrix $\mathbb{T}_{(m+1)(n+1)}$ must satisfy special additional *structural requirements* in order to serve as the covariance matrix of a causal/stable $\mathrm{AR}(m,n)$ model. This is an important difference from the $\mathrm{AR}_M(n)$ model where an *arbitrary p.d. Toeplitz-block matrix* $\mathbb{T}_{M(n+1)}$ can be uniquely extended to be the MN-variate Toeplitz-block covariance matrix of a causal $\mathrm{AR}_M(n)$ model. In practice, this means that an $(m+1)(n+1)$-variate Hermitian-block sample matrix $\hat{\mathbb{R}}_{(m+1)(n+1)}$ must be somehow first converted into a p.d. Toeplitz-block-Toeplitz matrix $\hat{\mathbb{T}}_{(m+1)(n+1)}$, which then has to undergo further modifications to meet those structural requirements. While possible in principle, but not in the maximum-likelihood sense (e.g., see [18]), it is uncertain whether this is the only model that can achieve the dramatic sample-support reduction (8), and at the same time, produce an MN-variate covariance matrix estimate appropriate for STAP applications.

(2.2) For the "conventional" $\mathrm{AR}(m,n)$ model and other models derived for an arbitrary $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz matrix $\mathbb{T}_{(m+1)(n+1)}$, MN-variate Hermitian relaxations that can be directly calculated from the given $(m+1)(n+1)$-variate p.d. Hermitian-block sample matrix $\hat{\mathbb{R}}_{(m+1)(n+1)}$ should be considered for practical radar applications.

This paper presents the results of our study that addresses these issues. Specifically, in Section II for clutter that is stationary in slow time (hence a Toeplitz-block covariance matrix), we introduce a new class of 2-D mixed autoregressive models $\mathrm{AR}_M(n|m)$. Apart from being an $\mathrm{AR}_M(n)$-type model in the slow-time domain, we impose additional time-varying autoregressive (TVAR) order restrictions in the spatial domain that lead to requiring fewer training samples. Such "mixed models" are AR in one dimension and TVAR in the other. In Section III, we introduce a set of new 2-D models $\mathrm{TbT}_M(m,n)$, $\mathrm{AR}_M(n\|m)$ and $\mathrm{AR}_N(m\|n)$ that allows us to uniquely reconstruct an MN-variate Toeplitz-block matrix, *given an arbitrary* $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz matrix. The models introduced in Sections II and III serve as approximations of the true Toeplitz-block or Toeplitz-block-Toeplitz clutter covariance matrix, and therefore allow us to investigate "model-mismatch" SINR losses. Section IV introduces relaxations $\mathrm{TVAR}_M(n)$ and $\mathrm{TVAR}_M(n|m)$ corresponding to the stationary causal models $\mathrm{AR}_M(n)$ and $\mathrm{AR}_M(n|m)$ respectively. For ground clutter that is stationary in both dimensions, we also derive the relaxations $\mathrm{TVAR}_M(n\|m)$ and $\mathrm{TVAR}_N(m\|n)$. Model-mismatch losses for these models are analyzed in Section V for the phenomenological KASSPER and AFRL datasets. We also demonstrate high STAP SINR performance when using a small number of i.i.d. training samples, generated according to (4). (A more detailed analysis of parametric STAP performance for airborne radar applications is presented in a separate study [21].) Section VI summarizes this paper, while the Appendix reproduces the mathematical details of the Dym-Gohberg band-extension method that are necessary for understanding certain subtleties of our model derivations.

The second paper in this series (which follows this paper within this issue) investigates 2-D parametric models for Hermitian matrices.

II. 2-D Mixed Autoregressive Model $\mathrm{AR}_M(n|m)$

The 2-D mixed AR model $\mathrm{AR}_M(n|m)$ is designed for data that is stationary in the temporal dimension only (e.g., airborne radar ground-clutter data which is stationary in slow time). We find the covariance matrix for this model by solving a maximum-entropy (ME) matrix completion problem. Recall that in such a problem, a partially specified (incomplete) matrix is given, and the task is to find the value of all the unspecified elements such that its determinant is maximized [22], [23].

Let $\boldsymbol{z}$ be the L-variate vector of all (distinct) unspecified covariance lags z in the MN-variate complex matrix $\mathbb{T}$, then the covariance matrix $\mathbb{T}(\boldsymbol{z})$ for the $\mathrm{AR}_M(n|m)$ model is defined by

$$\text{Find} \qquad \max_{\boldsymbol{z}\in\mathcal{C}^{L\times 1}} \log\det \mathbb{T}(\boldsymbol{z}) \tag{9}$$

given a certain partially specified $\mathbb{T}$.

A typical *one-dimensional* problem might involve an incomplete M-variate m-banded matrix $R(\boldsymbol{z})$, i.e., only the elements R_{jk} inside the central band $|j-k|\le m$ (which has bandwidth $2m+1$) are specified; then the ME-completion problem would be to complete the band matrix so as to maximize $\det R(\boldsymbol{z})$. (Some authors call this a "band extension", others reserve the term *extension* for when the dimension of the matrix is increased.) In this paper, we deal with *two*-dimensional data and an MN-variate Toeplitz-*block* matrix $\mathbb{T}$, and while the same concept applies, the notation and structure of $\mathbb{T}$ are a little more complicated.

Specifically, the structure of our MN-variate Hermitian Toeplitz-block matrix $\mathbb{T}$ (comprising $N\times N$ unstructured blocks, each of $M\times M$ elements) is

$$\mathbb{T}(\boldsymbol{z}) \equiv \begin{cases} r_{j-k}^{pq} & \text{for } |j-k|\le n \cap |p-q|\le m \\ z_{j-k}^{pq} & \text{for } n<|j-k|<N \cup m<|p-q|<M \end{cases} \tag{10}$$

i.e., within its "2-D block-band," $\mathbb{T}(\boldsymbol{z})$ has the same (fixed/prescribed) elements as $\mathbb{R}$, while outside the "2-D block-band" it has unspecified elements z_{j-k}^{pq} [24]. For convenience, we let $\mathcal{BB}$ be the set of indexes that correspond to matrix elements within the 2-D block-band, with $\widetilde{\mathcal{BB}}$ being the complementary set:

$$\begin{aligned} \mathcal{BB} &\equiv \{j,k,p,q\colon |j-k|\le n\cap |p-q|\le m\}, \\ \widetilde{\mathcal{BB}} &\equiv \{j,k,p,q\colon n<|j-k|<N\cup m<|p-q|<M\}. \end{aligned} \tag{11}$$

As usual, the spatial ("inner") indexes $p,q=1,\ldots,M$ operate within each block, while the temporal ("outer") indexes $j,k=1,\ldots,N$ operate on blocks. This special matrix structure is merely a reflection of way the 2-D data is ordered, namely, the MN-variate data vector $\mathbb{x}$ is stacked as N lots of M data (i.e., all antenna sensor outputs in a group, repeated for each slow-

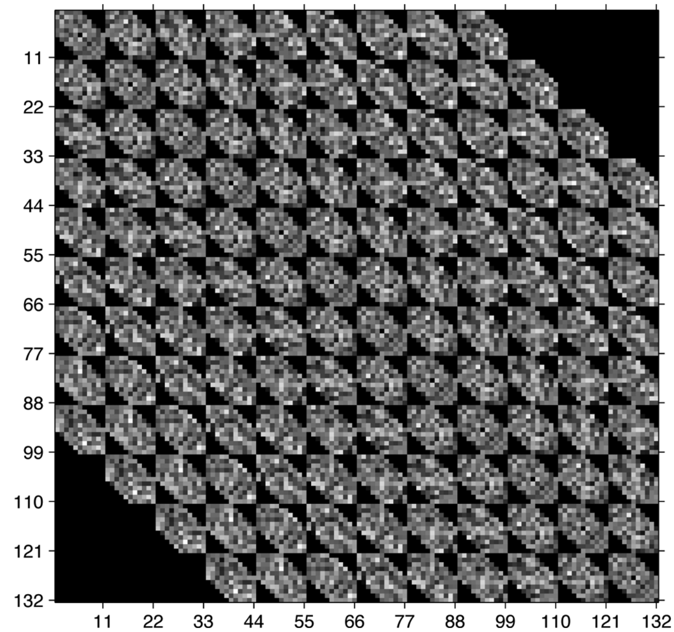

Fig. 1. Graphical representation of the structure of an example Hermitian Toeplitz-block matrix $\mathbb{T}$ with $M = 11$, $N = 12$ (rather than our usual $N = 32$), $m = 5$ and $n = 8$, where grayness represents element magnitude. The elements outside the "2-D block-band" (rendered black) are unspecified. The "2-D block-band" of specified elements ($\mathcal{BB}$) comprises $m + 1 = 6$ elements in the first row of each block, and $n + 1 = 9$ blocks in the first row of blocks.

time). This is the conventional radar ordering, though we shall see that the complementary ordering (i.e., all slow-time outputs in a group, repeated for each antenna sensor, producing a *block*-Toeplitz matrix) is useful.

Fig. 1 illustrates the structure of an example Hermitian Toeplitz-block matrix $\mathbb{T}$. We see that the ("inner", "microscopic," spatial) band within each M-variate block (where it exists) has width $2m + 1 = 11$, while the ("outer", "macroscopic," temporal) blockwise-band has bandwidth $2n + 1 = 17$ blocks. This 2-D block-band structure means that some of the M-variate blocks are partially specified, while others are completely unspecified.

In general, to define an entire MN-variate Hermitian Toeplitz-block matrix, it is (slightly more than) enough to specify the N distinct blocks (the top row, say). For almost every M-variate unstructured block (the top row, say, excluding the left block), it is necessary to specify all M^2 elements; for the exceptional Hermitian diagonal-block (the top-left block, say), there are only $M(M+1)/2$ distinct elements. In our particular 2-D block-banded ME-completion problem, exemplified by Fig. 1, let the $(n+1)$ distinct partially specified blocks be $R_0(\boldsymbol{z}^0), \ldots, R_n(\boldsymbol{z}^n)$, where

$$\boldsymbol{z}^{\kappa} \equiv \{z_{\kappa}^{pq}\}_{|\kappa| \leq n}^{m<|p-q|<M} \tag{12}$$

so that the unspecified vectors for this first group of blocks $\boldsymbol{z}^0, \ldots, \boldsymbol{z}^n$ each have length $L_1 \equiv (M-m-1)(M-m)$, except for the Hermitian block $\boldsymbol{z}^0$ that has length $L_0 \equiv L_1/2$. Let the remaining $(N-n-1)$ distinct fully unspecified blocks be $R_{n+1}(\boldsymbol{z}_{n+1}), \ldots, R_{N-1}(\boldsymbol{z}_{N-1})$, where

$$\boldsymbol{z}_{\kappa} \equiv \{z_{\kappa}^{pq}\}_{n<|\kappa|<N}^{p,q=1,\ldots,M} \tag{13}$$

so that this second group of unspecified vectors $\boldsymbol{z}_{n+1}, \ldots, \boldsymbol{z}_{N-1}$ each have length $L_2 \equiv M^2$. With this notation, the definition of the optimization function $\mathbb{T}(\boldsymbol{z})$ in the problem (9) can be recast as

$$\begin{aligned}\mathbb{T}(\boldsymbol{z}) = \mathrm{Toep}[&R_0(\boldsymbol{z}^0), \ldots, R_n(\boldsymbol{z}^n),\\ &R_{n+1}(\boldsymbol{z}_{n+1}), \ldots, R_{N-1}(\boldsymbol{z}_{N-1})]\end{aligned} \tag{14}$$

where $\mathrm{Toep}[\cdot]$ is the operator that forms a Hermitian Toeplitz matrix using its arguments.

First consider the subproblem that involves only the group of partially specified blocks

$$\begin{aligned}&\text{Find} \quad \max_{\boldsymbol{z} \in \mathcal{C}^{L \times 1}} \log\det \mathbb{T}_{M(n+1)}(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^n),\\ &\mathbb{T}_{M(n+1)}(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^n) \equiv \mathrm{Toep}\left[R_0(\boldsymbol{z}^0), \ldots, R_n(\boldsymbol{z}^n)\right].\end{aligned} \tag{15}$$

If we can somehow find the p.d. ME completion, $\mathbb{T}_{M(n+1)}^{(m)} \equiv \arg\max_{\boldsymbol{z}^0,\ldots,\boldsymbol{z}^n} \log\det \mathbb{T}_{M(n+1)}(\boldsymbol{z}^0, \ldots, \boldsymbol{z}^n) > 0$, then the overall completion problem (9), (14) (i.e., finding the vectors $\boldsymbol{z}_{n+1}, \ldots, \boldsymbol{z}_{N-1}$) is just the standard multivariate Burg problem. This suggests that we first deal with the partially specified blocks by ME-completing the $M(n+1)$-variate matrix $\mathbb{T}_{M(n+1)}$, and then deal with the fully unspecified blocks by ME-extending to the full-size MN-variate matrix $\mathbb{T}$.

A. Step 1: ME-Completion of $\mathbb{T}_{M(n+1)}$

Therefore the overall ME completion of $\mathbb{T}(\boldsymbol{z})$ (14) is feasible if the partially specified Toeplitz-block matrix $\mathbb{T}_{M(n+1)}$ is able to be ME-completed

$$\begin{aligned}\mathbb{T}_{M(n+1)}^{(m)} &\equiv \arg\max_{\boldsymbol{\zeta} \in \mathcal{C}^{L_1 \times 1}} \log\det \mathbb{T}_{M(n+1)}(\boldsymbol{\zeta}),\\ \mathbb{T}_{M(n+1)}(\boldsymbol{\zeta}) &= \mathrm{Toep}[R_0(\boldsymbol{z}^0), \ldots, R_n(\boldsymbol{z}^n)]\end{aligned} \tag{16}$$

$$\begin{aligned}R_{\kappa}(\boldsymbol{z}^{\kappa}) &\equiv \begin{cases} r_{\kappa}^{pq} & \text{for } |p-q| \leq m \\ z_{\kappa}^{pq} & \text{for } m < |p-q| < M \end{cases}\\ &|\kappa| \leq n, \qquad \boldsymbol{\zeta} \equiv \{\boldsymbol{z}^0, \ldots, \boldsymbol{z}^n\}.\end{aligned} \tag{17}$$

Instead of maximizing $\det \mathbb{T}_{M(n+1)}(\boldsymbol{\zeta})$, we may minimize the determinant of its inverse, $\mathbb{F}$ say, which immediately recasts this ME completion as a linear matrix inequality (LMI) [25], [26]. Due to the convexity of the LMI, existence and uniqueness of the solution is guaranteed. The following theorem may be viewed as a special case of in [22, Theorem 1.1], or as a multivariate generalization of in [27, Theorem 7].

Theorem 1: Let $\mathbb{T}$ be a partially specified p.d. Toeplitz-block matrix that admits p.d. completions. For any $\boldsymbol{\zeta}$ such that $\mathbb{T}(\boldsymbol{\zeta})$ is invertible, let

$$\mathbb{F}(\boldsymbol{\zeta}) \equiv \left\{f_{jk}^{pq}(\boldsymbol{\zeta})\right\}_{j,k=1,\ldots,n+1}^{p,q=1,\ldots,M} = \mathbb{T}^{-1}(\boldsymbol{\zeta}). \tag{18}$$

Then the determinant-maximizing p.d. Toeplitz-block matrix completion is unique and satisfies

$$\sum_{j=0}^{n-k} f_{j,j+k}^{pq} = 0 \quad \text{for} \quad 0 \le k \le n, \quad m < |p-q| < M. \tag{19}$$

In other words, the ME completion has the property that all elements in the inverse, in positions corresponding to the same completed element in the direct matrix, sum to zero.

One outcome of this theorem is that the "top-right block" of $\mathbb{F}$ is a band matrix

$$f_{0n}^{pq} = 0 \quad \text{for} \quad m < |p-q| < M \tag{20}$$

(consistent with the fact that the elements in $\mathbb{T}_{M(n+1)}$ in these positions are not replicated elsewhere).

Since this is a standard LMI, the solution ("analytic center") can be calculated by convex programming techniques, such as the interior-point methods, given a feasible starting point [26]. We choose to use the Newton method with Nesterov–Nemirovskii step-management [28]:

$$\boldsymbol{z} \longleftarrow \boldsymbol{z} - \alpha(z) H^{-1}(z) \boldsymbol{g}(\boldsymbol{z}) \tag{21}$$

where α is the "damping factor"

$$\alpha(\boldsymbol{z}) \equiv \begin{cases} 1, & \text{for } \delta(z) < \frac{1}{4} \\ \frac{1}{1+\delta(\boldsymbol{z})}, & \text{otherwise} \end{cases}$$
$$\delta(\boldsymbol{z}) \equiv \sqrt{\boldsymbol{g}^H(\boldsymbol{z}) H^{-1}(\boldsymbol{z}) \boldsymbol{g}(\boldsymbol{z})}. \tag{22}$$

We next specify the gradient $\boldsymbol{g}(\boldsymbol{z})$ and Hessian $H(\boldsymbol{z})$ of the objective function $\log\det \mathbb{T}_{M(n+1)}^{-1}(\boldsymbol{z})$.

Let $\boldsymbol{e}_j^p (j = 1,\ldots,n+1;\ p = 1,\ldots,M)$ be the $M(n+1)$-variate "element-selection vector", comprising a single nonzero element of unity at the position $M(j-1)+p$, so that the expression $\boldsymbol{e}_j^{pT} \mathbb{T} \boldsymbol{e}_k^q$ evaluates to (merely selects) the element of $\mathbb{T}$ in the block (p,q) at the local position (j,k). Then the gradient vector can be formally expressed as

$$\boldsymbol{g}(\boldsymbol{z}) \equiv \left[g_j^{pq}\right]_{1 \le j \le n+1; m<|p-q|<M}$$
$$g_j^{pq} = \sum_{i=0}^{n-j} \boldsymbol{e}_i^{p\,T} \mathbb{T}_{M(n+1)}^{-1}(\boldsymbol{z}) \boldsymbol{e}_{i+j}^q \tag{23}$$

and the number of (distinct) variables is $L = (2n+1)(M-m)(M-m-1)/2$. (The ordering of indexes is immaterial, as long as it is consistent.) The L-variate Hessian matrix is

$$\boldsymbol{H}\left(g_j^{pq}, g_k^{st}\right) \equiv \left\{ \sum_{i=0}^{n-j} \sum_{\ell=0}^{n-k} \left[\boldsymbol{e}_i^{pT} \mathbb{T}_{M(n+1)}^{-1}(\boldsymbol{z}) \boldsymbol{e}_\ell^s\right] \times \left[\boldsymbol{e}_{\ell+k}^{tT} \mathbb{T}_{M(n+1)}^{-1}(\boldsymbol{z}) \boldsymbol{e}_{i+j}^q\right] \right\}_{m<|p-q|,|s-t|<M}^{j,k=1,\ldots,n+1}. \tag{24}$$

Nesterov and Nemirovskii proved that this algorithm converges to the optimal solution $\mathbb{T}_{M(n+1)}^{(m)}$, where $\boldsymbol{g}(\boldsymbol{z}) = 0$ leads directly to (19), since

$$\boldsymbol{e}_i^{pT} \mathbb{T}_{M(n+1)}^{-1}(\boldsymbol{z}) \boldsymbol{e}_{i+j}^q = f_{i,i+j}^{pq}. \tag{25}$$

The Newton–Nesterov–Nemirovskii routine needs to begin at some feasible initial point. In this paper, we use this routine to find a parametric model for the given clairvoyant matrix, so naturally we use this as the initial point. Also, note that our attempts to avoid calculating the Hessian by using the computationally simpler gradient technique proved too demanding for the problem dimensions relating to KASSPER data (i.e., the convergence rate was impractically slow).

B. Step 2: ME-Extension of $\mathbb{T}_{M(n+1)}^{(m)}$ to $\mathbb{T}^{(n|m)}$

Now that we have a (fully specified) $M(n+1)$-variate matrix $\mathbb{T}_{M(n+1)}^{(m)}$, we wish to extend it to a full-size MN-variate matrix $\mathbb{T}^{(n|m)}$. As mentioned above, this is the standard multivariate Burg problem. The solution can be found using a special case of the Dym-Gohberg "band-extension" method (see the Appendix), and has the $\mathrm{AR}_M(n)$ property:

$$\begin{cases} \left\{\mathbb{T}^{(n|m)}\right\}_{j-k} = R_{j-k} & \text{for } |j-k| \le n \\ \left\{[\mathbb{T}^{(n|m)}]^{-1}\right\}_{jk} = 0 & \text{for } n < |j-k| < N. \end{cases} \tag{26}$$

The solution $\mathbb{T}^{(n|m)}$ uniquely specifies the mixed 2-D AR model $\mathrm{AR}_M(n|m)$.

The Dym-Gohberg band-extension theorem applies to any Hermitian-block matrix. For a more specifically Toeplitz-block matrix, we have

$$\begin{bmatrix} R_{ii} & \cdots & R_{i,i+n} \\ \vdots & & \vdots \\ R_{i+n,i} & \cdots & R_{i+n,i+n} \end{bmatrix} = \mathbb{T}_{M(n+1)} > 0, \quad \text{for } i = 1,\ldots,N-n \tag{27}$$

i.e., the "diagonally sliding block" is *constant* in the Toeplitz case, so the "full-height stacked matrices" $\mathbb{Y}_q$ in (68) are identical: $\mathbb{Y}_1 = \cdots = \mathbb{Y}_{N-n} \in \mathcal{C}^{M(n+1)\times M}$. Let 0_M be the M-variate zero matrix, then the "increasingly shorter stacked matrices" are

$$\mathbb{Y}_{N-n+j} = \begin{bmatrix} Y_{qq} \\ \vdots \\ Y_{Nq} \end{bmatrix} \begin{bmatrix} R_0 & \cdots & R_{j-n} \\ \vdots & \ddots & \vdots \\ R_{n-j} & \cdots & R_0 \end{bmatrix}^{-1} \begin{bmatrix} I_M \\ 0_M \\ \vdots \\ 0_M \end{bmatrix} \in \mathcal{C}^{M(n-j+1)\times M} \quad \text{for} \quad j = 1,\ldots,n. \tag{28}$$

Similarly, for a Toeplitz-block matrix, $\mathbb{x}_{n+1} = \mathbb{x}_{n+2} = \cdots = \mathbb{x}_N \in \mathcal{C}^{M(n+1)\times M}$, and

$$\mathbb{x}_q = \begin{bmatrix} R_0 & \cdots & R_{1-q} \\ \vdots & \ddots & \vdots \\ R_{1-q} & \cdots & R_0 \end{bmatrix}^{-1} \begin{bmatrix} 0_M \\ \vdots \\ 0_M \\ I_M \end{bmatrix} \in \mathcal{C}^{Mq\times M} \quad \text{for} \quad q = 1,\ldots,n. \tag{29}$$

Then the Dym-Gohberg factorization is

$$\mathbb{T}^{(n)} = (\mathbb{V}\mathbb{V}^H)^{-1} = (\mathbb{U}\mathbb{U}^H)^{-1} \tag{30}$$
$$V_{jk} = \begin{cases} Y_{jk} Y_{kk}^{-\frac{1}{2}}, & \text{for } k \le j \le \beta(k) \\ 0, & \text{otherwise,} \end{cases}$$

$$U_{jk} = \begin{cases} X_{jk} X_{kk}^{-\frac{1}{2}}, & \text{for } \gamma(k) \leq j \leq k \\ 0, & \text{otherwise.} \end{cases} \quad (31)$$

III. 2-D Mixed Autoregressive Models $\mathrm{TbT}_M(m,n)$, $\mathrm{AR}_M(n\|m)$ and $\mathrm{AR}_N(m\|n)$

As discussed in Section I, an $(m+1)(n+1)$-variate p.d. Hermitian Toeplitz-block-Toeplitz matrix is the covariance matrix of a causal (stable) 2-D autoregressive model $\mathrm{AR}(m,n)$, and has a unique MN-variate Toeplitz-block-Toeplitz extension, if and only if it satisfies the special structural conditions given by Woerdeman *et al.* [18, Theorem 2.1]. In this section, we consider models whose covariance matrix can be extended from *arbitrary* p.d. Toeplitz-block-Toeplitz to MN-variate.

A. $\mathrm{TbT}_M(m,n)$ *Model*

First, suppose that we require the completed MN-variate covariance matrix to remain p.d. and Toeplitz-block-Toeplitz. Then the ME completion problem is formally written as

$$\mathbb{T}^{(m,n)} \equiv \arg\max_{\boldsymbol{z}\in\mathcal{C}^{L\times 1}} \log\det \mathbb{T}(\boldsymbol{z}),$$
$$\mathbb{T}(\boldsymbol{z}) \equiv \begin{cases} r_{j-k}^{p-q} & \in \mathcal{BB} \\ z_{j-k}^{p-q} & \in \widetilde{\mathcal{BB}}, \end{cases} \quad \mathbb{T}_{(m+1)(n+1)} \equiv \left\{ r_{j-k}^{p-q} \right\}_{\in\mathcal{BB}}. \quad (32)$$

In other words, we seek the determinant-maximizing MN-variate matrix $\mathbb{T}^{(m,n)}$ that retains the (prescribed) elements of the given matrix $\mathbb{T}_{(m+1)(n+1)}$ within the 2-D block-band $\mathcal{BB}$ (see Fig. 1), whilst having to-be-determined elements z_{j-k}^{p-q} outside the 2-D block-band. As before, $\boldsymbol{z}$ is the L-variate vector of these (distinct) unknown complex numbers ordered in some arbitrary but consistent fashion.

The necessary and sufficient condition for the existence of a unique solution to this convex optimization problem (32) follows easily from the general result of Bakonyi and Woerdeman [22, Theorem 1.1]; then the solution must satisfy

$$\sum_{\substack{j,k=1\\|j-k|=\kappa}}^{N} \sum_{\substack{p,q=1\\|p-q|=\rho}}^{M} f_{jk}^{pq} = 0 \quad \text{for} \quad |j-k|>n, \quad |p-q|>m \quad (33)$$

where (as before) the matrix $\mathbb{F} \equiv \mathbb{T}^{-1}$ has elements f_{jk}^{pq} [cf. (19)]. As for the solution to every Toeplitz ME completion problem, this means that the elements in the inverse matrix *that correspond to the same completed covariance lag in the direct matrix* sum to zero. Hence, the MN-variate p.d. Toeplitz-block-Toeplitz ME completion of an arbitrary $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz matrix is generally *not* an autoregressive matrix, since groups of elements in the inverse sum to zero, but are not necessarily zero individually. For this reason, we call this model $\mathrm{TbT}_M(m,n)$, instead of using the AR nomenclature. While the completion is not an AR matrix, it does not preclude it from being used for STAP filter design.

To find the solution $\mathbb{T}^{(m,n)}$, (32) can be reformulated as the following LMI problem, similarly to the completion problem (16):

$$\text{Find} \quad \min \gamma(\boldsymbol{z}) \equiv \begin{cases} \log\det \mathbb{F}(\boldsymbol{z}), & \text{for } \mathbb{T}(\boldsymbol{z}) > 0 \\ \infty, & \text{otherwise.} \end{cases} \quad (34)$$

In order to calculate the gradient and Hessian, we again need a way to formally describe the matrix $\mathbb{T}$ as a function of its unknown variables $\boldsymbol{z}$. We introduce an "element-selection matrix", analogous to our definition of the element-selection vector $\boldsymbol{e}_j^p$ in the previous section.

Let D_κ^ρ be an MN-variate binary matrix whose only unity elements correspond to the positions of the group of elements relating to the same covariance lag (the same group mentioned above):

$$D_\kappa^\rho \equiv \left\{ d_{j-k}^{p-q}(\kappa,\rho) \right\}_{j,k=1,\ldots,N}^{p,q=1,\ldots,M}$$
$$d_{j-k}^{p-q}(\kappa,\rho) \equiv \delta(j-k-\kappa)\delta(p-q-\rho) \quad (35)$$

where $\delta(\cdot)$ is the Kronecker delta function. Let $\mathcal{P}$ be the sequence of indexes $\{1-M,\ldots,1-m, m+1,\ldots,M-1\}$. Then

$$\mathbb{T}(\boldsymbol{z}) = \mathbb{T}_0 + \sum_{\rho=m+1}^{M-1} \left[z_0^\rho D_0^\rho + \overline{z_0^\rho} \left(D_0^\rho\right)^T \right] + \sum_{\kappa=1}^{n} \sum_{\rho\in\mathcal{P}} \left[z_\kappa^\rho D_\kappa^\rho + \overline{z_\kappa^\rho} \left(D_\kappa^\rho\right)^T \right] \quad (36)$$
$$\boldsymbol{z} \equiv \left[z_{j-k=\kappa}^{p-q=\rho} \right]_{\kappa\leq n, m<|\rho|<M; n<\kappa<N, |m|<M}$$
$$\mathbb{T}_0 \equiv \begin{cases} r_{j-k}^{p-q} & \in \mathcal{BB} \\ 0 & \in \widetilde{\mathcal{BB}} \end{cases} \quad (37)$$

where the bar denotes complex conjugation. The number of variables in $\boldsymbol{z}$ is

$$L = (2M-1)(N-n-1) + (2n+1)(M-m-1). \quad (38)$$

With respect to the Nesterov–Nemirovskii algorithm, the gradient and Hessian are

$$\boldsymbol{g}(\boldsymbol{z}) \equiv \left[g_\kappa^\rho\right]_{\kappa,\rho\in\widetilde{\mathcal{BB}}},$$
$$g_\kappa^\rho = -\mathrm{tr}\mathbb{T}^{-1}(\boldsymbol{z}) \left(D_\kappa^\rho\right)^T \quad (39)$$
$$H\left(g_\kappa^\rho, g_s^t\right) = \left\{ \mathrm{tr}\mathbb{T}^{-1}(\boldsymbol{z}) D_s^t \mathbb{T}^{-1}(\boldsymbol{z}) \left(D_\kappa^\rho\right)^T \right\}_{\kappa,\rho,s,t\in\widetilde{\mathcal{BB}}}. \quad (40)$$

Again, the convergence condition $\boldsymbol{g}(\boldsymbol{z}) = 0$ coincides with the optimality condition (33).

In summary, our $\mathrm{TbT}_M(m,n)$ model enforces the Toeplitz-block-Toeplitz structure of the original covariance matrix onto the ME completion, but results in a non-AR model. Instead, we may wish to "approximate ME extensions with models from a suitably selected model set," as recommended in [23]. In this regard, let us now investigate what happens when we force the completion to be AR, but allow it to be a Toeplitz-block (not Toeplitz-block-Toeplitz) matrix.

B. $\mathrm{AR}_M(n\|m)$ Model

To find an $\mathrm{AR}_M(n)$-type ME completion, we first consider the completion of an $M(n+1)$-variate Toeplitz-block-Toeplitz matrix:

$$\mathbb{T}^{(m)}_{M(n+1)} \equiv \arg\max_{\boldsymbol{z}\in\mathcal{C}^{L\times 1}} \log\det \mathbb{T}_{M(n+1)}(\boldsymbol{z}),$$

$$\mathbb{T}_{M(n+1)}(\boldsymbol{z}) \equiv \begin{cases} r^{p-q}_{j-k} & \in \mathcal{BB} \\ z^{p-q}_{j-k} & \in \widetilde{\mathcal{BB}} \end{cases} \tag{41}$$

$$\boldsymbol{z} \equiv \left\{ z^{pq}_{jk} \right\}_{\in\widetilde{\mathcal{BB}}}. \tag{42}$$

The unique optimal solution to this LMI problem, $\mathbb{T}^{(m)}_{M(n+1)} \equiv \mathrm{Toep}[T^{(m)}_0,\ldots,T^{(m)}_n]$ say, can then be $\mathrm{AR}_M(n)$-type extended to an MN-variate p.d. Toeplitz-block matrix:

$$\mathbb{T}^{(n\|m)} \equiv \arg\max_{Z_{n+1},\ldots,Z_{N-1}} \log\det\mathrm{Toep} \times [T^{(m)}_0,\ldots,T^{(m)}_n,\, Z_{n+1},\ldots,Z_{N-1}] \quad Z_i\in\mathcal{C}^{M\times M}. \tag{43}$$

Indeed, the ME ("Burg") completion $\mathbb{T}^{(n\|m)}$ may again be calculated using the Dym-Gohberg band extension (26)–(31), i.e., we get an $\mathrm{AR}_M(n)$-type matrix with M-variate blocks, dependent on the spatial order m, generated from the $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz matrix. According to (26), this solution $\mathbb{T}^{(n\|m)}$ obeys the $\mathrm{AR}_M(n)$ property. This means that in the band within each block ($|j-k|\le n$), the solution has a Toeplitz-block-Toeplitz structure, as desired; outside this band, the completed matrices $Z_{n+1},\ldots,Z_{N-1}$ are arbitrary ones that meet our AR requirement.

We call this model $\mathrm{AR}_M(n\|m)$. The main difference between this model and $\mathrm{AR}_M(n|m)$ introduced in the previous section is that $\mathrm{AR}_M(n\|m)$ is for data that is stationary in *both* dimensions, and is represented by an $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz covariance matrix, despite the fact that the $\mathrm{AR}_M(n)$ property is retained in the slow-time dimension only. (Strictly speaking, the $\mathrm{AR}_M(n\|m)$ model should perhaps already be treated as a "relaxation", since stationarity in the space dimension is ignored.) On the other hand, the $\mathrm{AR}_M(n|m)$ model was introduced for data that is stationary only in time, and is represented by an incomplete $M(n+1)$-variate Toeplitz-block matrix.

C. $\mathrm{AR}_N(m\|n)$ Model

With respect to the $\mathrm{AR}_M(n\|m)$ model, the symmetry of space-time properties in a Toeplitz-block-Toeplitz matrix means that we can also define its dual model, $\mathrm{AR}_N(m\|n)$ say, which is represented by the covariance matrix $\mathbb{T}^{(m\|n)}_{NM}$. This dual matrix is still a Toeplitz-block matrix, but comprises $M\times M$ lots of $N\times N$ blocks (instead of *vice versa* for $\mathbb{T}^{(n\|m)}$).

Indeed, we can first complete the given matrix $\mathbb{T}_{(m+1)(n+1)}$ to an $N(m+1)$-variate Toeplitz-block-Toeplitz matrix by calculating the completion of the LMI problem

$$\mathbb{T}^{(n)}_{N(m+1)} \equiv \arg\max_{\boldsymbol{z}\in\mathcal{C}^{L\times 1}} \log\det \mathbb{T}_{N(m+1)}(\boldsymbol{z}),$$

$$\mathbb{T}_{N(m+1)}(\boldsymbol{z}) \equiv \begin{cases} r^{p-q}_{j-k} & \in \mathcal{BB} \\ z^{p-q}_{j-k} & \in \widetilde{\mathcal{BB}}, \end{cases} \quad \boldsymbol{z}\equiv\left\{z^{p-q}_{j-k}\right\}_{\in\widetilde{\mathcal{BB}}} \tag{44}$$

$$\mathbb{T}^{(n)}_{N(m+1)} \equiv \mathrm{Toep}\left[T^{(n)}_0,\ldots,T^{(n)}_m\right], \qquad T_i\in\mathcal{C}^{N\times N} \tag{45}$$

and then build the $\mathrm{AR}_N(m)$ model using the Dym-Gohberg technique

$$\mathbb{T}^{(m\|n)}_{NM} \equiv \arg\max_{Z_{m+1},\ldots,Z_{M-1}} \log\det\mathrm{Toep} \times \left[T^{(n)}_0,\ldots,T^{(n)}_m, Z_{m+1},\ldots,Z_{M-1}\right], \quad Z_i\in\mathcal{C}^{N\times N} \tag{46}$$

whose covariance matrix obeys the $\mathrm{AR}_N(m)$ property:

$$\begin{cases} \left\{\mathbb{T}^{(m\|n)}_{NM}\right\}_{p-q} = \left\{\mathbb{T}_{N(m+1)}\right\}_{p-q} & \text{for } |p-q|\le m \\ \left\{[\mathbb{T}^{(m\|n)}_{NM}]^{-1}\right\}_{pq} = 0 & \text{for } m<|p-q|<M. \end{cases} \tag{47}$$

D. Recap

The three 2-D mixed AR models that we have introduced, $\mathrm{TbT}_M(m,n)$, $\mathrm{AR}_M(n\|m)$ and $\mathrm{AR}_N(m\|n)$, are each uniquely defined by the corresponding completion of the given $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz matrix. Only if this matrix satisfies the special structural conditions given by Woerdeman *et al.* [18, Theorem 2.1] do all three of them collapse to the same causal 2-D model $\mathrm{AR}(m,n)$ whose MN-variate p.d. Toeplitz-block-Toeplitz covariance matrix $\mathbb{T}$ satisfies

$$\begin{cases} \{\mathbb{T}\}^{p-q}_{j-k} = r^{p-q}_{j-k} & \in\mathcal{BB} \\ \{\mathbb{T}^{-1}\}^{pq}_{jk} = 0 & \in\widetilde{\mathcal{BB}}. \end{cases} \tag{48}$$

Table I summarizes the characteristics of the stationary models introduced in this and the previous section.

IV. Time-Varying Autoregressive "Relaxations" $\mathrm{TVAR}_M(N)$, $\mathrm{TVAR}_M(n|m)$ and $\mathrm{TVAR}_M(n\|m)$

The 2-D TVAR relaxed models ("relaxations") $\mathrm{TVAR}_M(n)$ and $\mathrm{TVAR}_M(n|m)$ are designed for data that is stationary in the temporal dimension, while the model $\mathrm{TVAR}_M(n\|m)$ is for data that is stationary in both space and time; and are suboptimal parametric models associated with the above $\mathrm{AR}_M(n)$, $\mathrm{AR}_M(n|m)$ and $\mathrm{AR}_M(n\|m)$ models respectively.

As discussed earlier, the main reason for considering parametric models in STAP applications is that the sample support requirement is reduced. This is due to estimation of an $M(n+1)$- or $(m+1)(n+1)$-variate covariance matrix, instead of an MN-variate one, with ergodicity allowing training data to

TABLE I

SUMMARY OF THE 2-D MIXED STATIONARY MODELS. IN EACH CASE, THE GIVEN TEMPORAL COVARIANCE LAGS ARE $|j-k| \leq n$. THE FIRST TWO MODELS ARE FOR 1-D STATIONARY DATA; THE REMAINING THREE ARE FOR 2-D STATIONARY DATA

Model	Matrix	Given Spatial Lags	Problem Formulation
$\mathrm{AR}_M(n)$	$\mathbb{T}^{(n)}$	$p,q = 1, \ldots, M$	$\max\limits_{Z_{n+1},\ldots,Z_N} \log\det \mathrm{Toep}\left[R_0, \ldots, R_n, Z_{n+1}, \ldots, Z_N\right]$ $Z_{n+1}, \ldots, Z_N \in \mathcal{C}^{M\times M}, \quad R_\kappa \equiv \left\{r^{pq}_{j-k=\kappa}\right\}_{p,q=1,\ldots,M}$
$\mathrm{AR}_M(n\|m)$	$\mathbb{T}^{(n\|m)}$	$\|p-q\| \leq m$	$\max\limits_{Z_{n+1},\ldots,Z_N} \log\det \mathrm{Toep}\left[R_0(Z_0^{pq}), \ldots, R_n(Z_n^{pq}), Z_{n+1}, \ldots, Z_N\right]$ $Z^{pq}_{j-k=\kappa}, \quad \|\kappa\| \leq n, \quad m < \|p-q\| < M, \quad Z_{n+1}, \ldots, Z_N \in \mathcal{C}^{M\times M}$
$\mathrm{TbT}_M(m,n)$	$\mathbb{T}^{(m,n)}$	$\|p-q\| \leq m$	$\max\limits_{z^{p-q}_{j-k}} \log\det \mathbb{T} = \begin{cases} r^{p-q}_{j-k} & \in \mathcal{BB} \\ z^{p-q}_{j-k} & \in \widetilde{\mathcal{BB}}. \end{cases}$
$\mathrm{AR}_M(n\\|m)$	$\mathbb{T}^{(n\\|m)}$	$\|p-q\| \leq m$	$\max\limits_{Z_{n+1},\ldots,Z_N} \log\det \mathrm{Toep}\left[R_0(Z_0^{p-q}), \ldots, R_n(Z_n^{p-q}), Z_{n+1}, \ldots, Z_N\right]$ $Z^{p-q=\rho}_{j-k=\kappa}, \quad \|\kappa\| \leq n, \quad m < \|\rho\| < M, \quad Z_{n+1}, \ldots, Z_N \in \mathcal{C}^{M\times M}$
$\mathrm{AR}_N(m\\|n)$	$\mathbb{T}^{(m\\|n)}$	$\|p-q\| \leq m$	$\max\limits_{Z_{n+1},\ldots,Z_N} \log\det \mathrm{Toep}\left[\check{R}_0(Z^0_{j-k}), \ldots, \check{R}_n(Z^n_{j-k}), \check{Z}_{m+1}, \ldots, \check{Z}_M\right]$ $Z^{p-q=\kappa}_{j-k=\rho}, \quad \|\kappa\| \leq m, \quad n < \|\rho\| < N, \quad \check{Z}_{m+1}, \ldots, \check{Z}_M \in \mathcal{C}^{N\times N}$

be averaged over slow time (for 1-D stationary clutter data) or over both slow time and space (for 2-D stationary clutter).

A straight-forward idea that extends the framework of the stationary models developed in the previous two sections is to somehow convert the averaged sample (Hermitian-block) covariance matrix $\hat{\mathbb{R}}_{M(n+1)}$ or $\hat{\mathbb{R}}_{(m+1)(n+1)}$ into a p.d. Toeplitz-block or Toeplitz-block-Toeplitz matrix.

Unfortunately, even for the scalar problem, maximum-likelihood estimation of a p.d. Toeplitz covariance matrix is a difficult problem that does not have a known closed-form solution [29]. Brute-force approaches such as redundancy-averaging [30] usually give poor results, so that parametrization methods (e.g., reflection coefficients) are widely used. In [7], a p.d. Toeplitz matrix was reconstructed by estimating its maximum-entropy spectrum from a sample Hermitian matrix. Computationally, the problem was solved by factorizing a positive polynomial. In the multivariate case, the same approach can be taken for a positive matrix-valued polynomial [31], [32]. However, this method is complicated and seems to be of little practical value for real-time STAP applications.

Therefore, this section discusses an alternate approach, whereby we reconstruct an MN-variate Hermitian-block matrix from the given p.d. Hermitian-block sample matrix $\hat{\mathbb{R}}_{M(n+1)}$ or $\hat{\mathbb{R}}_{(m+1)(n+1)}$, with properties that are suitable for STAP applications.

A. $\mathrm{TVAR}_M(n)$ *Relaxation of the* $\mathrm{AR}_M(n)$ *Model*

Section II outlines the Dym-Gohberg band-extension technique for reconstructing an MN-variate Toeplitz-block $\mathrm{AR}_M(n)$-type matrix $\mathbb{T}^{(n)}$ given $\mathbb{T}_{M(n+1)}$. The reconstruction $\mathbb{T}^{(n)}$ has *exactly the same* multivariate ME spectrum as $\mathbb{T}_{M(n+1)}$, and in fact this equivalence uniquely defines the $\mathrm{AR}_M(n)$-type completion of an MN-variate Toeplitz-block matrix. In the scalar case, the ME spectrum is uniquely defined by the first column of the inverse matrix. In the 2-D case, we have

$$[\mathbb{T}^{(n)}]^{-1}\begin{bmatrix} I_M \\ 0_M \\ \vdots \\ 0_M \end{bmatrix} \equiv [\mathbb{T}^{(n)}]^{-1}\mathbb{I}_M = \begin{bmatrix} \mathbb{Y}_{M(n+1)} \\ 0_M \\ \vdots \\ 0_M \end{bmatrix},$$

$$\mathbb{Y}_{M(n+1)} \equiv \mathbb{T}^{-1}_{M(n+1)}\mathbb{I}_M \in \mathcal{C}^{M(n+1)\times M} \tag{49}$$

i.e., the first column-block of the inverse of the Toeplitz-block matrix defines the ME spectrum, and all nontrivial entries of this inverse are the same for the reconstructed $\mathbb{T}^{(n)}$ as for the original $\mathbb{T}_{M(n+1)}$.

Given the $M(n+1)$-variate Hermitian-block sample matrix $\hat{\mathbb{R}}_{M(n+1)}$, such that

$$\mathcal{E}\left\{\hat{\mathbb{R}}_{M(n+1)}\right\} = \mathbb{T}_{M(n+1)} \tag{50}$$

where $\mathcal{E}$ is the expectation operator, we propose constructing an MN-variate p.d. Hermitian-block matrix $\hat{\mathbb{R}}^{(n)}$ that meets two requirements. First, $\hat{\mathbb{R}}^{(n)}$ is $\mathrm{TVAR}_M(n)$-type covariance matrix:

$$\left\{[\hat{\mathbb{R}}^{(n)}]^{-1}\right\}_{jk} = 0 \quad \text{for} \quad n < |j-k| < N. \tag{51}$$

Second, the estimate of the multivariate ME (Burg) spectrum for $\hat{\mathbb{R}}^{(n)}$ is the same as for the given sample matrix:

$$[\hat{\mathbb{R}}^{(n)}]^{-1}\mathbb{I}_{MN} = \left[\mathbb{Y}^T_{M(n+1)}, 0_M, \ldots, 0_M\right]^T,$$

$$\mathbb{Y}_{M(n+1)} \equiv \hat{\mathbb{R}}^{-1}_{M(n+1)}\mathbb{I}_{M(n+1)}. \tag{52}$$

In fact, these two conditions *do not* uniquely specify the Hermitian extension of a given Hermitian-block matrix. With this extra freedom, it is important to note that we suggest the reconstruction process of the $\mathrm{TVAR}_M(n)$ model is performed the same as for Toeplitz-block matrices (27)–(31), *treating the Hermitian-block matrix as if it was Toeplitz-block*. Indeed, in those equations it is simply necessary to replace $\mathbb{T}_{M(n+1)}$ by $\hat{\mathbb{R}}_{M(n+1)}$.

The positive-definiteness of the sample matrix $\hat{\mathbb{R}}_{M(n+1)}$ means that $\hat{Y}_{qq} > 0$ for $q = 1, \ldots, N$ [cf. (28)], hence the block-lower-triangular matrix $\hat{\mathbb{V}}$ [cf. (30)] and, therefore, the resulting Hermitian-block matrix $\hat{\mathbb{R}}^{(n)} = [\hat{\mathbb{V}}\hat{\mathbb{V}}^H]^{-1}$ are p.d. Note that the above two properties are met by construction.

As before, we can also contemplate the "backward reconstruction" via $\hat{\mathbb{R}}^{(n)} = [\hat{\mathbb{U}}\hat{\mathbb{U}}^H]^{-1}$, where $\hat{\mathbb{U}}$ [cf. (31)] is a block-upper-triangular matrix. We may even consider computing the solution as

$$\hat{\mathbb{R}}^{(n)} = \frac{1}{2}[\hat{\mathbb{V}}\hat{\mathbb{V}}^H + \hat{\mathbb{U}}\hat{\mathbb{U}}^H]^{-1}. \tag{53}$$

Clearly, if the sample matrix $\hat{\mathbb{R}}_{M(n+1)}$ was a *Toeplitz*-block matrix, $\mathbb{T}_{M(n+1)}$ say, then all three options would yield the same result. For a merely *Hermitian*-block sample matrix, the results may differ, but these should not be statistically significant due to (50).

Recall that when the clairvoyant Toeplitz-block matrix $\mathbb{T}_{M(n+1)}$ is used instead of $\hat{\mathbb{R}}_{M(n+1)}$, we get the conventional $\mathrm{AR}_M(n)$ covariance matrix $\mathbb{T}^{(n)}$.

B. $\mathrm{TVAR}_M(n|m)$ *Relaxation of the* $\mathrm{AR}_M(n|m)$ *Model*

The mixed 2-D AR model $\mathrm{AR}_M(n|m)$ was introduced in Section II as the completion of the Toeplitz-block matrix $\mathbb{T}_{M(n+1)}(\boldsymbol{\zeta})$, followed by an $\mathrm{AR}_M(n)$-type extension to an MN-variate Toeplitz-block matrix $\mathbb{T}^{(m)}$. In the adaptive setting, the partially specified sample matrix is no longer Toeplitz-block:

$$\hat{\mathbb{R}}_{M(n+1)}(\boldsymbol{z}) \equiv \begin{cases} \hat{r}_{jk}^{pq} & \text{for } j,k = 1,\ldots,n+1; |p-q| \le m \\ z_{jk}^{pq} & \text{for } j,k = 1,\ldots,n+1; m < |p-q| < M \end{cases} \tag{54}$$

so that there is no need to retain the Toeplitz properties of the completed elements in $\hat{\mathbb{R}}_{M(n+1)}$, as in (17), hence the $\mathrm{TVAR}_M(n|m)$ model's covariance matrix is found by *unconstrained* ME completion:

$$\mathbb{R}^{(m)} \equiv \underset{\boldsymbol{z}\in\mathcal{C}^{L\times 1}}{\arg\max} \log\det \hat{\mathbb{R}}_{M(n+1)}(\boldsymbol{z}). \tag{55}$$

It is simple to show that the Dym-Gohberg band-extension method is directly applicable in this case.

We solve the problem by first simply reordering the sample matrix $\hat{\mathbb{R}}_{M(n+1)}$, with its $(n+1)\times(n+1)$ lots of $M\times M$ blocks, into a matrix $\check{\mathbb{R}}_{(n+1)M}$ that has $M\times M$ lots of $(n+1)\times(n+1)$ blocks:

$$\check{\mathbb{R}}_{(n+1)M} \equiv \left\{\check{R}_{pq}\right\}_{p,q=1,\ldots,M}. \tag{56}$$

This just means that we have swapped the spatial and temporal dimensions in the presentation of the radar data in the block matrix (the "inner" dimension is now temporal, while the "outer" is spatial). Formally, this transformation is accomplished by

$$\check{\mathbb{R}}_{(n+1)M} = \mathbb{J}\hat{\mathbb{R}}_{M(n+1)}\mathbb{J}^T \tag{57}$$

where $\mathbb{J}$ is an $M(n+1)$-variate unitary permutation matrix, whose columns are appropriately selected from $I_{M(n+1)}$. Our completion problem is therefore recast as

$$\begin{aligned} \check{\mathbb{R}}_{(n+1)M}^{(m)} &\equiv \underset{\boldsymbol{z}\in\mathcal{C}^{L\times 1}}{\arg\max} \log\det \check{\mathbb{R}}_{(n+1)M}(\boldsymbol{z}), \\ \check{\mathbb{R}}_{(n+1)M}(\boldsymbol{z}) &\equiv \begin{cases} \check{R}_{pq} & \text{for } |p-q| \le m \\ Z_{pq} & \text{for } m < |p-q| < M \end{cases} \end{aligned} \tag{58}$$

which coincides with the Dym-Gohberg band-extension formulation. The unique solution $\check{\mathbb{R}}_{(n+1)M}^{(m)}$ is found in a computationally efficient way, and has the $\mathrm{AR}_N(m)$ property:

$$\begin{cases} \left\{\check{\mathbb{R}}_{(n+1)M}^{(m)}\right\}_{jk} = \check{\mathbb{R}}_{pq} & \text{for } |p-q| \le m \\ \left\{\left[\check{\mathbb{R}}_{(n+1)M}^{(m)}\right]^{-1}\right\}_{jk} = 0 & \text{for } m < |p-q| < M. \end{cases} \tag{59}$$

Then we perform the inverse reordering to obtain the $M(n+1)$-variate Hermitian-block matrix (whose 2-D data structure is returned to usual):

$$\hat{\mathbb{R}}_{M(n+1)}^{(m)} = \mathbb{J}^T\check{\mathbb{R}}_{M(n+1)}^{(m)}\mathbb{J}. \tag{60}$$

Note that according to Theorem A.2, all $(m+1)(n+1)$-variate principal matrices $\check{\mathbb{R}}_{pq}(|p-q| \le m; j,k = 1,\ldots,M)$ must be p.d., which for a sample matrix $\hat{\mathbb{R}}_{M(n+1)}$ holds with probability one if $\tau > (m+1)(n+1)/(N-n)$; hence the sample matrix may be rank deficient.

Finally, given the completed p.d. Hermitian-block matrix $\hat{\mathbb{R}}_{M(n+1)}^{(m)}$, we compute the MN-variate $\mathrm{TVAR}_M(n)$-type covariance matrix $\hat{\mathbb{R}}^{(n|m)}$ using the restoration technique as for the $\mathrm{TVAR}_M(n)$ model. We therefore call this model $\mathrm{TVAR}_M(n|m)$.

The main difference between the $\mathrm{TVAR}_M(n)$ and $\mathrm{TVAR}_M(n|m)$ models is noticed when both models are applied to a Toeplitz-block matrix $\mathbb{T}_{M(n+1)}$. In this case, the $\mathrm{TVAR}_M(n)$ model coincides with the $\mathrm{AR}_M(n)$ one, and gives an MN-variate Toeplitz-block matrix $\mathbb{T}^{(n)}$; whereas the $\mathrm{TVAR}_M(n|m)$ model does not give a Toeplitz-block matrix when applied to the partially specified $\mathbb{T}_{M(n+1)}$. This is because the Dym-Gohberg band-extension method does not respect the Toeplitz nature of the N-variate blocks $\check{R}_{pq}$, and generates non-Toeplitz "missing" blocks that meet the requirement (59), hence the result $\hat{\mathbb{R}}_{M(n+1)}$ in (60) is no longer a Toeplitz-block matrix.

In the next section, we investigate how detrimental this "relaxation" is for STAP SINR performance.

C. $\mathrm{TVAR}_M(n\|m)$ *Relaxation of the* $\mathrm{AR}_M(n\|m)$ *Model*

In this case, we have data that is stationary in both dimensions (space and slow time, in the airborne radar application), and are given the $(m+1)(n+1)$-variate p.d. Hermitian-block sample matrix $\hat{\mathbb{R}}_{(m+1)(n+1)}$ that has been averaged in both dimensions, i.e., the prototype for this case is an M-sensor ULA collecting N periodic pulses.

We use the same approach as in the previous subsection: the given matrix is expanded in the spatial domain to the dimension $M(n+1)$, then convert this $M(n+1)$-variate into an

TABLE II
SUMMARY OF THE 2-D MIXED TIME-VARYING AUTOREGRESSIVE "RELAXATIONS." IN EACH CASE, THE GIVEN TEMPORAL COVARIANCE LAGS ARE $j, k = 1, \ldots, n+1$

Model	Prototype	Matrix	Given Spatial Cov. Lags	Properties
$\text{TVAR}_M(n)$	$\text{AR}_M(n)$	$\hat{\mathbb{R}}^{(n)}$	$p, q = 1, \ldots, M$	$\left\{[\hat{\mathbb{R}}^{(n)}]^{-1}\right\}_{jk} = 0$ for $n < \|j-k\| < N$, ME spectrum of $\hat{\mathbb{R}}^{(n)}$ same as $\hat{\mathbb{R}}_{M(n+1)}$
$\text{TVAR}_M(n\|m)$	$\text{AR}_M(n\|m)$	$\hat{\mathbb{R}}^{(n\|m)}$	$\|p-q\| \le m$	$\left\{[\hat{\mathbb{R}}^{(n\|m)}]^{-1}\right\}_{jk} = 0$ for $n < \|j-k\| < N$, ME spectrum of $\hat{\mathbb{R}}^{(n\|m)}$ same as $\hat{\mathbb{R}}^{(m)}_{M(n+1)}$
$\text{TVAR}_M(n\|\|m)$	$\text{AR}_M(n\|\|m)$	$\hat{\mathbb{R}}^{(n\|\|m)}$	$p, q = 1, \ldots, m+1$	$\left\{[\hat{\mathbb{R}}^{(n\|\|m)}]^{-1}\right\}_{jk} = 0$ for $n < \|j-k\| < N$, ME spectrum of $\hat{\mathbb{R}}^{(n\|\|m)}$ same as $\hat{\mathbb{R}}^{(m)}_{M(n+1)}$, ME spectrum of $\check{\mathbb{R}}^{(m)}_{(n+1)M}$ same as $\check{\mathbb{R}}_{(n+1)(m+1)}$
$\text{TVAR}_N(m\|\|n)$	$\text{AR}_N(m\|\|n)$	$\hat{\mathbb{R}}^{(m\|\|n)}$	$p, q = 1, \ldots, m+1$	$\left\{[\check{\mathbb{R}}^{(m\|\|n)}]^{-1}\right\}_{pq} = 0$ for $m < \|p-q\| < M$, ME spectrum of $\check{\mathbb{R}}^{(m\|\|n)}$ same as $\check{\mathbb{R}}^{(n)}_{(m+1)N}$, ME spectrum of $\hat{\mathbb{R}}^{(n)}_{N(m+1)}$ same as $\hat{\mathbb{R}}_{(m+1)(n+1)}$

MN-variate $\text{TVAR}_M(n)$-type covariance matrix. Briefly, the algorithm is to reorder the input matrix

$$\check{\mathbb{R}}_{(n+1)(m+1)} = \mathbb{J}\hat{\mathbb{R}}_{(m+1)(n+1)}\mathbb{J}^T \tag{61}$$

then restore

$$\check{\mathbb{R}}^{(m)}_{(n+1)M} = [\hat{\mathbb{V}}\hat{\mathbb{V}}^H]^{-1},$$
$$\hat{V}_{jk} \equiv \begin{cases} \hat{Y}_{jk}\hat{Y}_{kk}^{-\frac{1}{2}} & \text{for } k \le j \le \min\{N, k+m\} \\ 0 & \text{otherwise} \end{cases} \tag{62}$$

using the "full height stacked matrices"

$$[\hat{Y}_{ii}^T, \ldots, \hat{Y}_{i+m,i}^T]^T \equiv \hat{\mathbb{Y}}_{(m+1)(n+1)} \equiv \check{\mathbb{R}}^{-1}_{(n+1)(m+1)}\mathbb{I}_{n+1}, \quad \text{for } i = 1, \ldots, M-m \tag{63}$$

and the "increasingly shorter" [see (64) shown at the bottom of the page], then inverse reorder

$$\hat{\mathbb{R}}^{(m)}_{M(n+1)} = \mathbb{J}^T\check{\mathbb{R}}^{(m)}_{(n+1)M}\mathbb{J} \tag{65}$$

and finally restore the $\text{TVAR}_M(n)$-type MN-variate covariance matrix $\hat{\mathbb{R}}^{(n\|m)}$.

Naturally, we can also form the dual model $\text{TVAR}_N(m\|n)$ by similarly completing the given sample matrix as an $N(m+1)$-variate Hermitian-block matrix $\hat{\mathbb{R}}^{(n)}_{N(m+1)}$ (instead of $\hat{\mathbb{R}}^{(m)}_{M(n+1)}$), and then extending it to a $\text{TVAR}_N(m)$-type covariance matrix $\hat{\mathbb{R}}^{(m\|n)}$.

Again, it is important to note that, when applied to the clairvoyant $(m+1)(n+1)$-variate Toeplitz-block-Toeplitz matrix, the relaxations $\text{TVAR}_M(n\|m)$ and $\text{TVAR}_N(m\|n)$ are different to the models $\text{AR}_M(n\|m)$ and $\text{AR}_N(m\|n)$ respectively, mainly because the Dym-Gohberg method does not retain the Toeplitz-block-Toeplitz structure of the completed $\hat{\mathbb{R}}^{(m)}_{M(n+1)}$ or $\hat{\mathbb{R}}^{(n)}_{N(m+1)}$. Hence, the Toeplitz-block-Toeplitz or even Toeplitz-block structure is not necessarily retained in the relaxation covariance matrices $\hat{\mathbb{R}}^{(n\|m)}$ or $\hat{\mathbb{R}}^{(m\|n)}$, despite the 2-D stationarity being exploited by sliding-window averaging.

Table II summarizes the characteristics of the four relaxations introduced in this section.

V. SINR PERFORMANCE OF PARAMETRIC STAP: KASSPER AND AFRL DATA RESULTS

As discussed in the Introduction, we choose to evaluate different parametric models for STAP applications by analyzing the SINR loss factor with respect to the optimum (clairvoyant) Wiener filter (3) for phenomenological clutter covariance matrices, as provided by KASSPER and AFRL datasets. Recall that the KASSPER model has ULA antenna errors that prevent the true covariance matrix from being Toeplitz-block-Toeplitz, whereas the AFRL covariance matrix is based on the same KASSPER radar scenario, but with simplified electromagnetic calculations and an ideal ULA.

First, Fig. 2 presents the distribution of optimum STAP filter gains over the conventional matched filter for the (a) AFRL and (b) KASSPER model at the fixed elevation angle $\phi = -4.65°$ that is appropriate for range bin 200. (Recall that the "spatial aperture" M is 11 and the "temporal aperture" N is 32 in these datasets.) These gains are calculated on a discrete 2-D grid that cover the entire range of azimuthal angle θ and Doppler frequency ω; however, radar scientists sometimes prefer to work

$$\begin{bmatrix} \hat{Y}_{ii} \\ \vdots \\ \hat{Y}_{Mi} \end{bmatrix} = \begin{bmatrix} \check{R}_{i-M+m+1,i-M+m+1} & \cdots & \check{R}_{i-M+m+1,m+1} \\ \vdots & \ddots & \vdots \\ \check{R}_{m+1,i-M+m+1} & \cdots & \check{R}_{m+1,m+1} \end{bmatrix}^{-1} \mathbb{I}_{n+1} \quad \text{for } i = M-m+1, \ldots, M \tag{64}$$

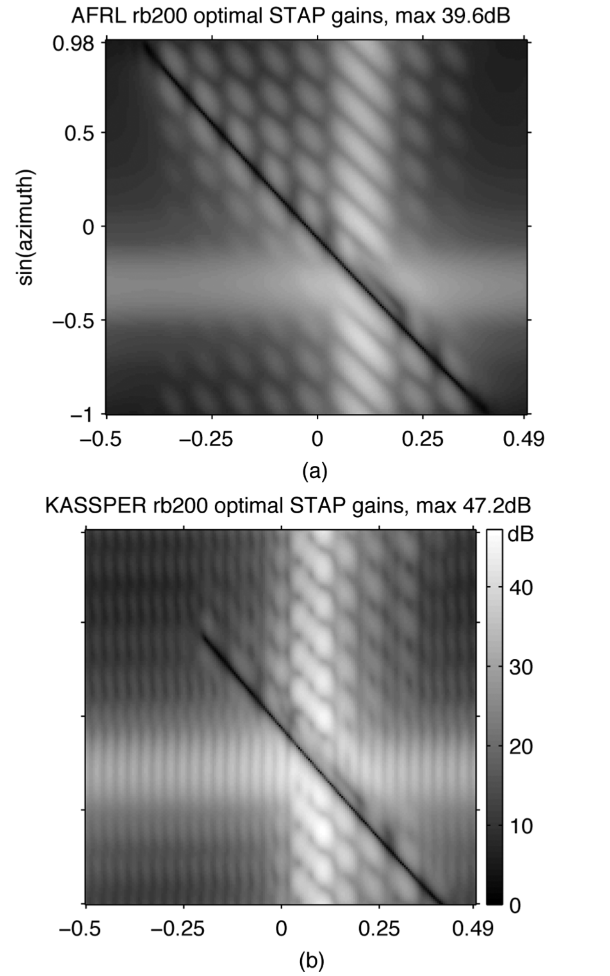

Fig. 2. Azimuth-Doppler distribution of optimal STAP gains with respect to the conventional matched filter for range bin 200 from both datasets.

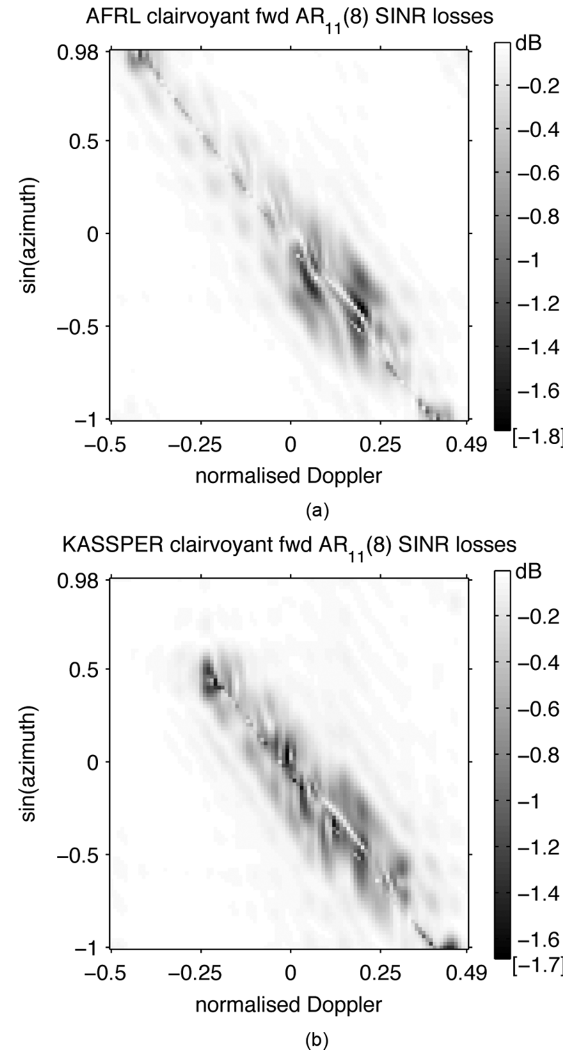

Fig. 3. Clairvoyant (forward-reconstructed) $\mathrm{AR}_{11}(8)$ SINR losses relative to optimal STAP filter processing for (a) AFRL and (b) KASSPER data.

in terms of $\sin\theta$ and normalized Doppler frequency $\omega/2\pi$. The optimum STAP gain q is calculated similarly to (3):

$$q(\theta,\omega) = \frac{\mathbb{s}^H \mathbb{T}^{-1} \mathbb{s}\mathbb{s}^H \mathbb{T}\mathbb{s}}{[\mathbb{s}^H\mathbb{s}]^2}. \tag{66}$$

Fig. 2 shows that the antenna errors present in the KASSPER data do not significantly affect the optimal STAP gain in the AFRL data. As expected for a target with zero radial speed, along the so-called clutter ridge (the long thin dark diagonal feature) there is no gain, but even in close proximity to the clutter ridge the gain can exceed 40 dB for KASSPER data. The slightly different additive-noise power in the AFRL data reduces the global maximum gain to about 40 dB, instead of about 47 dB for KASSPER data, but otherwise the distributions in Fig. 2 are very similar.

Fig. 3 illustrates the SINR losses (3) for the traditional [11], [12], [13], [15], $\mathrm{AR}_M(n)$ model of (temporal) order $n = 8$, calculated for the clairvoyant matrices for range bin 200 in each dataset, with almost the same distribution of losses. The maximum SINR loss factor for the AFRL Toeplitz-block-Toeplitz matrix is -1.8 dB, while for the KASSPER Toeplitz-block matrix it reaches -1.9 dB. As expected for a low-order $\mathrm{AR}_M(n)$ model, maximum losses occur close to the clutter ridge, whereas away from the ridge the losses drop to only a fraction of a decibel.

These results confirm the very high potential efficiency of the $\mathrm{AR}_M(n)$ models with $n \ll N$ for STAP applications, as demonstrated in [12], [13], but also show that this model, which does not rely on an ideal ULA, responds equally well to the ideal and perturbed ULA geometries. Unfortunately, we will see that this equal-responsiveness breaks down as soon as we apply "spatial averaging" over the clairvoyant covariance matrix.

Indeed, Fig. 4 presents the AFRL/KASSPER comparison in SINR loss factor for the $\mathrm{TVAR}_M(n\|m)$ model with orders $m = 5$ and $n = 8$. In the AFRL case (ideal ULA, Toeplitz-block-Toeplitz covariance matrix), this 2-D parametric model results in -2.9-dB maximum losses, which is only -1.1 dB worse than for the traditional 1-D parametric $\mathrm{AR}_M(n)$ model of Fig. 3(a). For the KASSPER data (imperfect ULA, Toeplitz-block covariance matrix, other antenna errors), the $\mathrm{TVAR}_M(n\|m)$ model (that incorporates spatial averaging) leads to disastrous SINR losses as great as -27 dB. Clearly, this particular parametric model that relies on an ideal ULA is completely unsuitable for an antenna array with such errors,

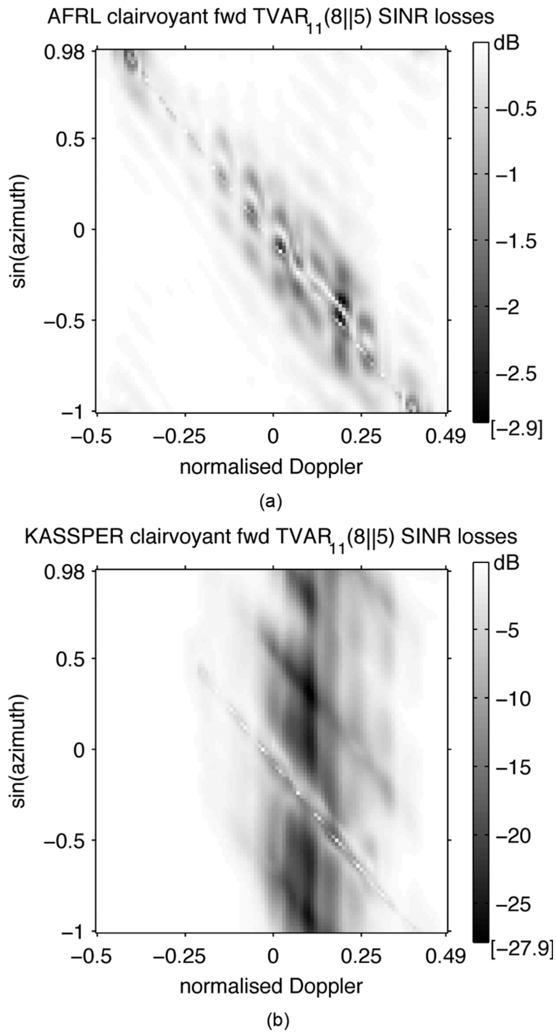

Fig. 4. Clairvoyant (forward-reconstructed) $\mathrm{TVAR}_{11}(8\|5)$ SINR losses for (a) AFRL and (b) KASSPER data.

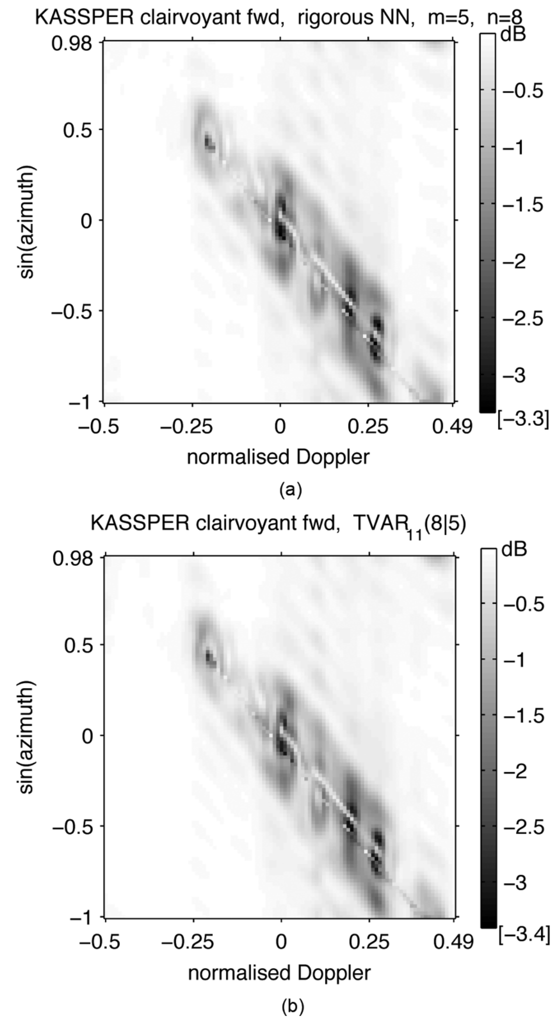

Fig. 5. Clairvoyant SINR losses for (a) the "rigorous" Nesterov–Nemirovskii and (b) $\mathrm{TVAR}_{11}(8|5)$ models.

while $\mathrm{AR}_M(n|m)$ could still be applied, and so we hope that some sample-support reductions may be achieved for the KASSPER antenna array by considering our new $\mathrm{AR}_M(n|m)$ model and/or its $\mathrm{TVAR}_M(n|m)$ relaxation instead of the unsuitable $\mathrm{AR}_M(n\|m)$.

Fig. 5 shows the SINR losses for the latter two models when applied to the clairvoyant Toeplitz-block matrix $\mathbb{T}_{M(n+1)}$, whereby the solutions $\mathbb{T}^{(n|m)}$ and $\mathbb{R}^{(n|m)}$ differ, as discussed above. Yet, in terms of SINR losses, this difference is negligible. Fig. 5(a) is for the model $\mathrm{AR}_{11}(8|5)$ that is rigorously calculated using the Nesterov–Nemirovskii routine for ME completion; Fig. 5(b) is calculated for the computationally simpler $\mathrm{TVAR}_M(n|m)$ relaxation. The results are almost indistinguishable, and in both cases the loss is bounded by -3.4 dB. Since the number of parameters in the $\mathrm{AR}_M(n|m)$ and $\mathrm{TVAR}_M(n|m)$ models is less than in the $\mathrm{AR}_M(n)$ and $\mathrm{TVAR}_M(n)$ models respectively, we can expect that the greater model-mismatch losses in the $\mathrm{AR}_M(n|m)$ or $\mathrm{TVAR}_M(n|m)$ model, relative to the $\mathrm{AR}_M(n)$ or $\mathrm{TVAR}_M(n)$ one, are offset by smaller finite sample-support losses.

A detailed analysis of SINR losses in parametric STAP with limited samples runs beyond the scope of this study, and so is to be presented in a separate paper [21]. However, at Fig. 6 we illustrate a sample stochastic (a particular realization of) SINR loss-factor distribution for $\tau = 5$ i.i.d. training samples, generated using (4). Here we compare $\mathrm{TVAR}_{11}(8|5)$-based STAP with one based on the $\mathrm{TVAR}_{11}(8)$ parametric model. As expected, the maximum SINR loss for the more restrictive $\mathrm{TVAR}_{11}(8|5)$ model (-8.8 dB in this instance) are not as severe as in the $\mathrm{TVAR}_{11}(8)$ model (-13.3 dB) that does not impose any spatial restrictions. Here we average the standard $M(n+1)$-variate sample matrix $\hat{\mathbb{R}}_{M(n+1)}$ over $N-n=24$ "temporal shifts" and $\tau=5$ i.i.d. training samples, so that the total number of $M(n+1)$-variate training samples is $\tau(N-n)=120$, which marginally exceeds the matrix dimension $M(n+1)=99$.

Comparing the accurate stationary (causal) $\mathrm{AR}_M(n|m)$ model performance in Fig. 5(a), and that of its $\mathrm{TVAR}_M(n|m)$ relaxation for the clairvoyant and sample KASSPER cases in Fig. 6(a), reveals that relaxation-based STAP is quite efficient for a "real-world" environment, but also that an exact restoration of the Toeplitz properties in the reconstructed MN-variate covariance matrix model *is unnecessary* for STAP design. The same conclusion follows from comparing Fig. 3(a) and Fig. 4(a): while the rigorous $\mathrm{AR}_M(n)$ model (which is

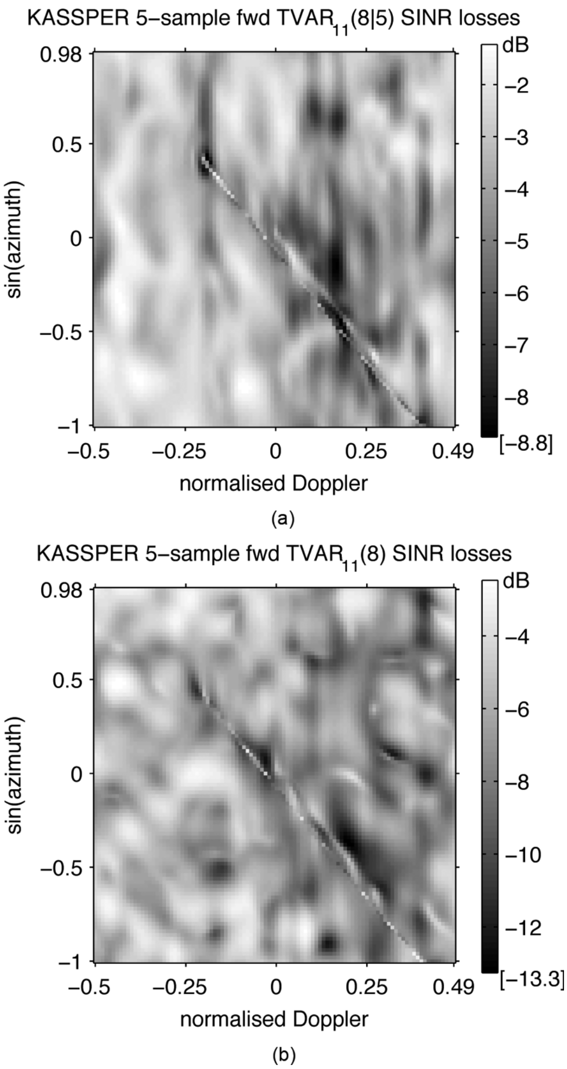

Fig. 6. Sample KASSPER SINR losses for (a) $\mathrm{TVAR}_{11}(8|5)$, and (b) $\mathrm{TVAR}_{11}(8)$, for $\tau = 5$ snapshots.

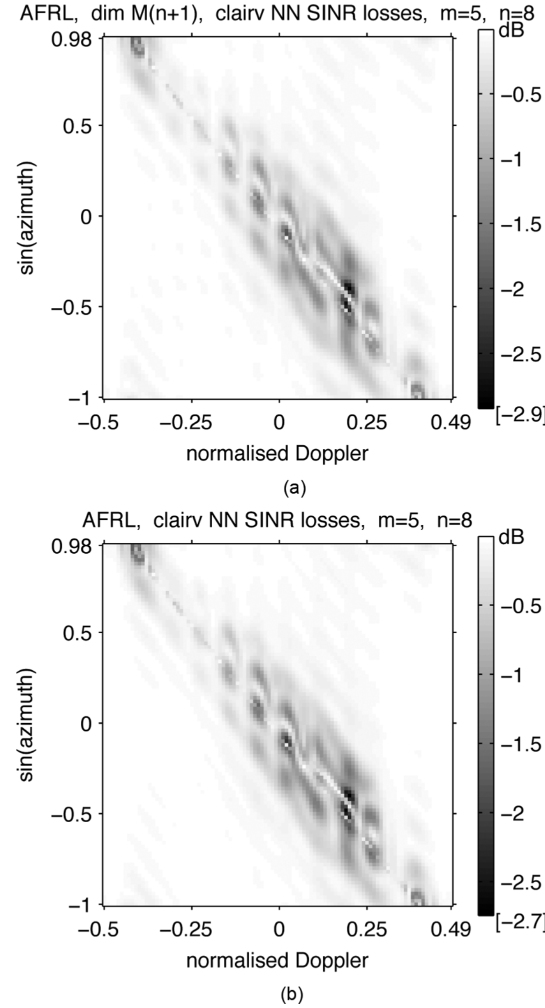

Fig. 7. Comparison of clairvoyant SINR losses for the rigorous Nesterov-Nemirovskii $\mathrm{TbT}_{11}(5,8)$ model for the (a) partial $M(n+1)$-variate and (b) full MN-variate AFRL data.

equivalent to the $\mathrm{AR}_M(n|m)$ and $\mathrm{AR}_M(n\|m)$ models when $m = M-1$) leads to -1.8-dB maximum SINR losses, the $\mathrm{TVAR}_{11}(8\|5)$ model is only -1.1 dB worse. This means that the large spatial-order reduction from $m_{\max} = M-1 = 10$ to $m = 5$, and all possible additional SINR degradations with respect to the rigorous stationary models $\mathrm{TbT}_M(m,n)$, $\mathrm{AR}_M(n\|m)$ or $\mathrm{AR}_N(m\,\|\,n)$, *all together* cost only -1.1 dB in (maximal) SINR degradation!

This fact, along with Fig. 5, removes any motivation to analyze SINR losses for the rigorous stationary models for this particular airborne radar scenario. Yet, at Fig. 7 we show the loss distributions for the $\mathrm{TbT}_{11}(5,8)$ and $\mathrm{AR}_{11}(8\|5)$ models calculated for the clairvoyant matrix $\mathbb{T}_{(m+1)(n+1)}$ that, as expected, is hardly distinguishable from Fig. 4(a).

Finally, at Fig. 8 we present an example stochastic realization of SINR losses for $\tau = 5,\,4,\,2,\,1$ i.i.d. training samples calculated for the $\mathrm{TVAR}_{11}(8\|5)$ relaxation model. Of course, for a single training snapshot we observe significant SINR losses, up to -11.6 dB, but already for five samples the SINR losses do not exceed -3.9 dB.

Clearly, spatial stationarity and ergodicity permit a noticeable STAP performance improvement, though for the KASSPER/AFRL radar model with $M \ll N$, this improvement is less profound than the one due to temporal-order reduction ($n \ll N$). We may expect different results for applications other than the KASSPER/AFRL scenario, however, the proposed set of 2-D parametric stationary models and their time-varying autoregressive “relaxations” provide a good “tool-box” for parametric STAP design.

VI. Summary and Conclusion

For a multivariate process that is stationary in its temporal dimension (“slow time” in the radar application), we have introduced various 2-D parametric models that rely on autoregressive and maximum-entropy extension principles, motivated by their application in STAP radar space–time adaptive processing. These parametric models have fewer free parameters than the model for an arbitrary MN-variate Hermitian covariance matrix, which results in our goal of a reduced requirement for training-sample support, but unfortunately also leads to an extra STAP performance degradation caused by the mismatch

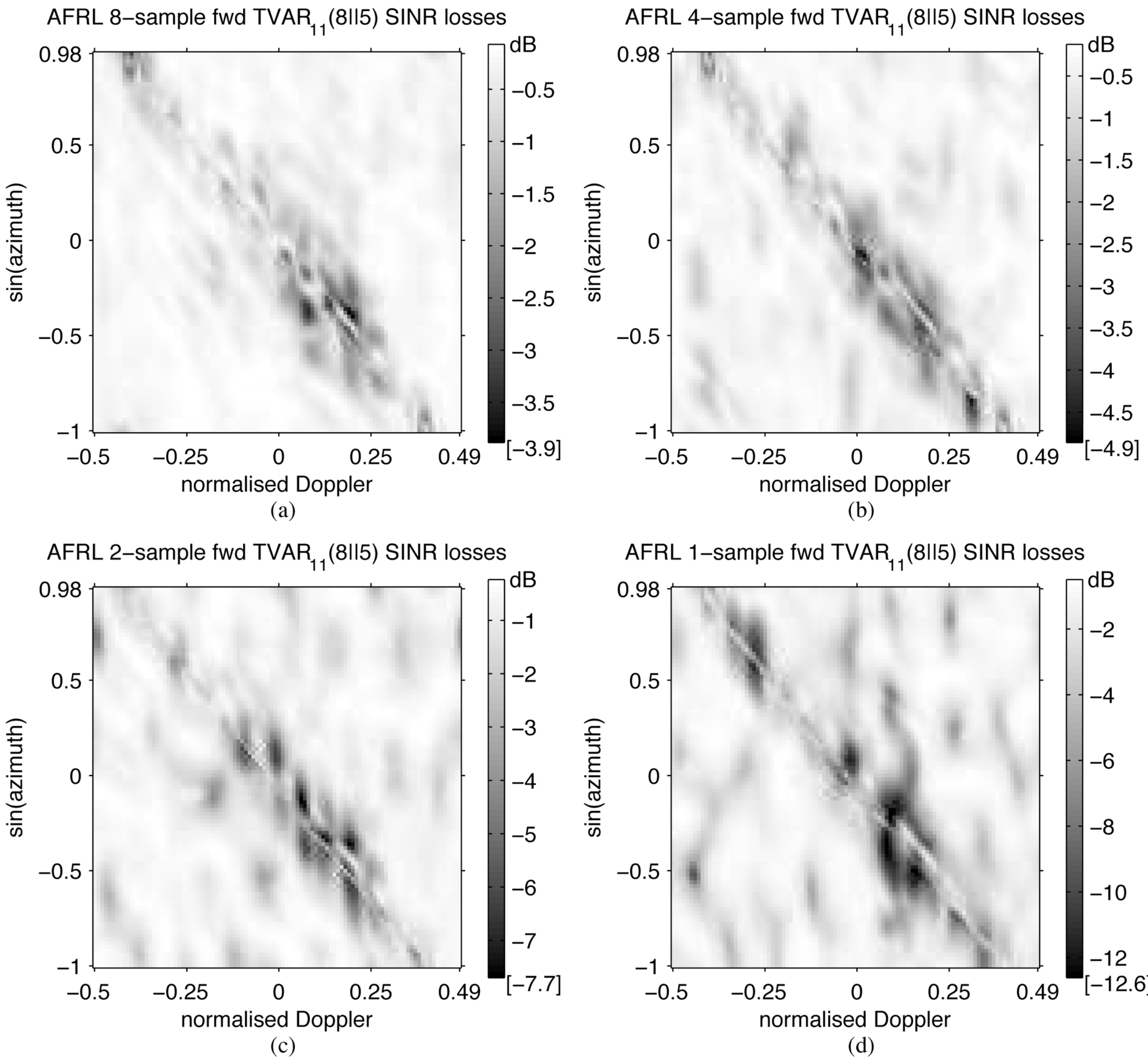

Fig. 8. Sample AFRL $\mathrm{TVAR}_{11}(8\|5)$ SINR losses for (a) $\tau = 8$, (b) $\tau = 4$, (c) $\tau = 2$, and (d) $\tau = 1$ snapshots.

between any true clutter covariance matrix and its parametric model.

To quantify the performance of the introduced models, we therefore focused on the STAP filter performance degradation (SINR loss) due to replacing the true covariance matrix in the optimum Wiener filter by the estimate from the parametric model. This criterion enables any invertible covariance matrix model to be considered for STAP design. On the other hand, meaningful recommendations regarding these parametric models can only be justified for a particular application and having a trustworthy high-fidelity phenomenological clutter covariance matrix provided. For this reason, our performance analysis has been conducted for the radar clutter covariance matrix model (DARPA's "KASSPER Dataset 1") of an airborne side-looking radar, with a nominally uniformly spaced linear antenna array ($M = 11$ sensors) and a periodic waveform ($N = 32$ repetitions). This clutter is stationary in slow time over the coherent processing interval, and is described by a range-dependent MN-variate Toeplitz-block matrix, whose structure is $N \times N$ lots of $M \times M$ ("inner") spatial blocks. Due to a number of different antenna errors incorporated into the radar front-end model, these spatial blocks are not Toeplitz matrices, as would be the case for an ideal ULA.

In order to explore STAP enhancement for a perfect ULA (or at least sufficiently accurate) geometry, we therefore considered a complementary model, devised by the AFRL. For the same KASSPER scenario, this model adopts slightly less sophisticated electromagnetic phenomena, but has flexible antenna errors, including none. Stationarity (and therefore ergodicity) in one (slow-time) or both (slow-time and ULA-aperture) dimensions, allows us use sliding-window averaging over these dimensions when estimating covariance matrices. This makes the traditional choice of the multivariate autoregressive model $\mathrm{AR}_M(n)(n \ll N)$ natural for slow-time stationarity only, and the 2-D AR model $\mathrm{AR}(m,n)$ for an ideal ULA and slow-time stationarity.

The family of existing parametric STAP models has been expanded in this paper. For an arbitrary antenna array geometry, we have introduced a "mixed" AR model $\mathrm{AR}_M(n|m)$. In addition to the well-known properties of the M-variate model $\mathrm{AR}_M(n)$, our $\mathrm{AR}_M(n|m)$ model considers spatial lags observed within the limited band $|p-q| \leq m$, and then the missing covariance matrix elements are completed in a maximum-entropy fashion. Relative to the conventional $\mathrm{AR}_M(n)$ model, our $\mathrm{AR}_M(n|m)$ model has a reduction in (minimal) training-sample support by a factor of $M(M-m)$. The unique optimum solution for the $\mathrm{AR}_M(n|m)$ model (given set of true covariance lags) has been derived as a convex-programming solution ("analytic center") of a linear matrix inequality problem.

Whereas the $\mathrm{AR}_M(n)$ and $\mathrm{AR}_M(n|m)$ models can be calculated for an *arbitrary* $M(n+1)$-variate p.d. Toeplitz-block matrix, the 2-D causal $\mathrm{AR}(m,n)$ model may not exist for an arbitrary $(m+1)(n+1)$-variate Toeplitz-block-Toeplitz clutter covariance matrix, where the clutter is stationary in two dimensions. This result was recently proven by Geronimo and Woerdeman, who derived structural conditions that must be met by a Toeplitz-block-Toeplitz matrix to serve as the covariance matrix of a causal $\mathrm{AR}(m,n)$ model.

As an alternative to the 2-D $\mathrm{AR}(m,n)$ model, therefore, we suggested a new family of parametric models that are uniquely specified by an *arbitrary* $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz matrix, and introduced them as ME completions. In particular, we demonstrated that an ME completion that is also constrained to be Toeplitz-block-Toeplitz is not necessarily an AR-type matrix (but this does not preclude it for STAP filter design). In our two other completions $\mathrm{AR}_M(n\|m)$ and $\mathrm{AR}_N(m\|n)$, we enforced the AR properties in the completion, but only demanded Toeplitz-block (rather than Toeplitz-block-Toeplitz) properties, so that some completed blocks are non-Toeplitz. All the introduced ME completions of an arbitrary $(m+1)(n+1)$-variate p.d. Toeplitz-block-Toeplitz matrix collapse to the same $\mathrm{AR}(m,n)$ model if the necessary and sufficient conditions given by Geronimo and Woerdeman are satisfied.

Strictly speaking, the two models $\mathrm{AR}_M(n\|m)$ and $\mathrm{AR}_N(m\|n)$ must be treated as "relaxations" (relaxed models), since the fundamental Toeplitz-block-Toeplitz property relied on for sliding-window averaging over both dimensions is no longer retained in the final covariance matrix model. Yet, we demonstrated that for AFRL data this deliberate "inconsistency" does not lead to a noticeable STAP performance degradation.

Our sole parametric model performance criterion has been radar-relevant STAP SINR, and this encouraged us to expand the (pre-existing but vaguely defined) relaxation idea much further. Most studies on parametric STAP consider estimation of a causal AR model that is consistent with the adopted ergodicity properties. In adaptive settings, this means that when constructing an accurate Toeplitz-block or Toeplitz-block-Toeplitz matrix model, the sample Hermitian $M(n+1)$- or $(m+1)(n+1)$-variate matrix is first transformed into a p.d. Toeplitz-block or Toeplitz-block-Toeplitz matrix respectively, and then completed using the convex programming technique, for example. While possible, this "rigorous" approach is impractical due to significant computational requirements.

Rather than Toeplitz-block or Toeplitz-block-Toeplitz solutions, we therefore proposed a new class of Hermitian TVAR relaxations, that can be directly calculated from the $M(n+1)$- or $(m+1)(n+1)$-variate Hermitian sample matrices. When constructing our TVAR relaxations, we followed the same fundamental principle as for the "rigorous" causal $\mathrm{AR}_M(n)$ model. Indeed, the $\mathrm{AR}_M(n)$-type MN-variate Toeplitz-block matrix extension of a given $M(n+1)$-variate p.d. Toeplitz-block matrix may be treated as the unique extension that retains the multivariate ME (Burg) spectrum of the given $M(n+1)$-variate matrix. We have derived a reconstruction of the $\mathrm{TVAR}_M(n)MN$-variate Hermitian relaxation using a Cholesky factorization of its inverse, which mimics the same factorization of the MN-variate Toeplitz-block matrix, and therefore retains the same multivariate ME spectrum estimate in the MN-variate extended matrix as in the given $M(n+1)$-variate sample matrix.

These relaxations were derived for adaptive applications with $M(n+1)$- and $(m+1)(n+1)$-variate sample Hermitian-block matrices, and when applied to the true $M(n+1)$-variate Toeplitz block or $(m+1)(n+1)$-variate Toeplitz-block-Toeplitz matrices, respectively, they generally lead to models different from $\mathrm{AR}_M(n|m)$, $\mathrm{AR}_M(n\|m)$ or $\mathrm{AR}_N(m\|n)$. Yet, for the KASSPER covariance matrix for range bin 200, the rigorous mixed model $\mathrm{AR}_{11}(8|5)$ results in -2.88-dB maximum SINR losses, while its corresponding $\mathrm{TVAR}_{11}(8|5)$ relaxation has almost identical -2.86-dB losses. Similarly, the performance of the rigorous $\mathrm{AR}_M(n\|m)$ model calculated for the 200th AFRL covariance matrix was practically the same as that of its relaxation $\mathrm{TVAR}_{11}(8\|5)$.

This analysis demonstrated that significant spatial-order reduction from $m_{\max} = M-1 = 10$ to $m = 5$ in the mixed model $\mathrm{AR}_{11}(8|5)$ increases the model-mismatch losses from -1.8 dB in the conventional $\mathrm{AR}_{11}(8)$ model to only -2.9 dB, while almost doubling the effective number of training samples $(M/(M-m) = 1.83)$. Spatial averaging for the ideal ULA geometry (AFRL data) leads to a much more significant sample-support increase, while the model-mismatch losses in the $\mathrm{TVAR}_{11}(8\|5)$ model are practically the same $(-2.9$ dB) as in the $\mathrm{AR}_{11}(8\|5)$ and $\mathrm{TVAR}_{11}(8|5)$ models for the perturbed-ULA (KASSPER) data. Unfortunately, the KASSPER antenna errors lead to a catastrophic SINR performance degradation when spatial averaging is used to calculate the $\mathrm{TVAR}_M(n\|m)$ model.

We decided that a detailed statistical analysis of parametric STAP performance was beyond the scope of this paper, mainly due to various possible modifications (regularizations) of the $M(n+1)$- or $(m+1)(n+1)$-variate sample covariance matrix estimate that could be considered. This analysis, as well as the antenna array accuracy requirements that permit spatial averaging, is to be presented separately. However, we provided a few examples that demonstrate an extremely small training sample support (five samples, say) gives good performance for a sufficiently accurate ULA geometry.

Overall, for a high-fidelity trusted phenomenological clutter covariance matrix model, there is a modest performance degradation associated with our 2-D parametric models which offer significant reductions in training-sample support requirements and calculations for STAP filter design.

APPENDIX
DYM-GOHBERG BAND-EXTENSION METHOD

This material is reproduced from in [33, Chapter III]; see also from [34, Theorem 1.1], and [35].

Theorem A.1: For $i,j=1,\dots,n$ and $|i-j|\le p$ let $A_{ij}=A_{ji}^H$ be a given operator acting from one Hilbert space $\mathcal{H}_j$ onto another $\mathcal{H}_i$, and suppose that

$$\begin{bmatrix} A_{ii} & \cdots & A_{i,i+p} \\ \vdots & & \vdots \\ A_{i+p,i} & \cdots & A_{i+p,i+p} \end{bmatrix} > 0 \quad \text{for } i=1,\dots,n-p. \quad (67)$$

For $q=1,\dots,n$ let

$$\mathbb{Y}_q \equiv \begin{bmatrix} Y_{qq} \\ \vdots \\ Y_{\beta(q),q} \end{bmatrix} \begin{bmatrix} A_{qq} & \cdots & A_{q,\beta(q)} \\ \vdots & & \vdots \\ A_{\beta(q),q} & \cdots & A_{\beta(q),\beta(q)} \end{bmatrix}^{-1} \begin{bmatrix} 0_q \\ \vdots \\ 0_q \\ I_q \end{bmatrix},$$

$$\beta(q) \equiv \min\{n, q+p\} \quad (68)$$

$$\mathbb{X}_q \equiv \begin{bmatrix} X_{\gamma(q),q} \\ \vdots \\ X_{qq} \end{bmatrix} \begin{bmatrix} A_{\gamma(q),\gamma(q)} & \cdots & A_{\gamma(q),q} \\ \vdots & & \vdots \\ A_{q,\gamma(q)} & \cdots & A_{qq} \end{bmatrix}^{-1} \begin{bmatrix} 0_q \\ \vdots \\ 0_q \\ I_q \end{bmatrix},$$

$$\gamma(q) \equiv \max\{1, q-p\}. \quad (69)$$

Let the $n\times n$ triangular operator matrices U and V be defined by

$$V_{ij} = \begin{cases} Y_{ij}Y_{jj}^{-\frac{1}{2}}, & \text{for } j\le i\le \beta(j) \\ 0, & \text{otherwise} \end{cases}$$
$$U_{ij} = \begin{cases} X_{ij}X_{jj}^{-\frac{1}{2}}, & \text{for } \gamma(j)\le i\le j \\ 0, & \text{otherwise.} \end{cases} \quad (70)$$

Then the $n\times n$ operator $F\equiv (UU^H)^{-1}=(VV^H)^{-1}$ is the unique p.d. operator matrix with

$$F_{ij}=A_{ij}, \quad |i-j|\le p$$

and

$$\{F^{-1}\}_{ij}=0, \quad |i-j|>p. \quad (71)$$

Remark: Note that the theorem applies for different dimensions of the Hilbert spaces $\mathcal{H}_1,\dots,\mathcal{H}_n$. For STAP applications, this means the MN-variate data vector can be divided into arbitrary blocks.

Theorem A.2: For $i,j=1,\dots,n$ and $|i-j|\le p$ let $A_{ij}=A_{ji}^H$ be a given operator acting from one Hilbert space $\mathcal{H}_j$ onto another $\mathcal{H}_i$. In order that there exists a positive extension of the band $\{A_{ij}\}_{|i-j|\le p}$ it is necessary and sufficient that (67) holds.

Remark: This can be used to specify the minimum sample volume required for a p.d. band extension to exist for a sample matrix.

Corollary A.1: For $i,j=1,\dots,n$ and $|i-j|\le p$ let $A_{ij}=A_{ji}^H$ be a given operator acting from one finite-dimensional Hilbert space $\mathcal{H}_j$ onto another $\mathcal{H}_i$, and suppose that (67) holds. Then for any positive extension A of the given band, $\det A\le \prod_{q=1}^{n}\det M_q^{-1}$, where

$$M_q \equiv [0_q,\dots,0_q,\, I_q] \begin{bmatrix} A_{\gamma(q),\gamma(q)} & \cdots & A_{\gamma(q),q} \\ \vdots & & \vdots \\ A_{q,\gamma(q)} & \cdots & A_{qq} \end{bmatrix}^{-1} \times \begin{bmatrix} 0_q \\ \vdots \\ 0_q \\ I_q \end{bmatrix} \quad \text{for } q=1,\dots,n \quad (72)$$

with equality holding iff A is the unique band extension of the given band.

Remark: This proves the ME property of the unique band extension (for a Gaussian ensemble).

ACKNOWLEDGMENT

The authors would like to thank E. Baranoski (DARPA) and T. Cushman (AFRL) for supplying DARPA KASSPER Dataset 1, and P.M. Corbell and M. Rangaswamy (AFRL) for supplying the complementary AFRL dataset, and a number of useful discussions.

REFERENCES

[1] R. Nitzberg, "An effect of range-heterogeneous clutter on adaptive Doppler filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 3, pp. 475–480, May 1990.
[2] W. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 621–633, Apr. 2000.
[3] R. Klemm, *Space-Time Adaptive Processing: Principles and Applications*. London, U.K.: IEE, 1998.
[4] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, no. 6, pp. 853–863, Nov. 1974.
[5] J. Bergin and P. Techau, "High-fidelity site-specific radar data set," in *Proc. 2nd DARPA KASSPER Workshop*, Apr. 2002, pp. 1–8.
[6] J. Guerci and E. Baranoski, "Knowledge-aided adaptive radar at DARPA: An overview," *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 41–50, Jan. 2006.
[7] Y. Abramovich, D. Arov, and V. Kachur, "Adaptive cancellation filters for stationary interference with a Toeplitz correlation matrix," *Sov. J. Commun. Technol. Electron.*, vol. 33, no. 4, pp. 54–61, 1988, English translation of *Radioteknika i Elektronika*.
[8] P. Techau, D. Barrick, and A. Schnittman, "The two-scale bistatic rough surface scattering model," presented at the 2nd ARPA Advanced Signal Processing Hot Clutter Technical Interchange Meeting, Rome Laboratory, NY, Sep. 1994.
[9] J. Bergin, private communication, Nov. 2007.
[10] P. Corbell, private communication, Nov. 2007.
[11] J. Guerci and E. Feria, "Application of a least squares predictive-transform modeling methodology to space-time adaptive array processing," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1825–1833, Jul. 1996.

[12] J. Michels, M. Rangaswamy, and B. Himed, “Performance of parametric and covariance based STAP tests in compound-Gaussian clutter,” *Digit. Signal Process.*, vol. 12, pp. 307–388, 2002.

[13] A. De Maio and A. Farina, “A maximum entropy framework for spacetime adaptive processing,” *Signal Process.*, vol. 84, no. 9, pp. 1637–1652, Sep. 2004.

[14] D. Barton, “Low-angle radar tracking,” *Proc. IEEE*, vol. 62, no. 6, pp. 687–704, Jun. 1974.

[15] P. Parker and A. Swindlehurst, “Space-time autoregressive filtering for matched subspace STAP,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 2, pp. 510–520, Apr. 2003.

[16] S. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[17] S. Marple, P. Corbell, and M. Rangaswamy, “Multi-channel parametric estimator for fast block matrix inverses,” in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Honolulu, HI, 2007, vol. 2, pp. 1137–1140.

[18] H. Woerdeman, J. Geronimo, and G. Castro, “A numerical algorithm for stable 2-D autoregressive filter design,” *Signal Process.*, vol. 83, no. 6, pp. 1299–1308, Jun. 2003.

[19] E. Kelly, “An adaptive detection algorithm,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 115–127, Mar. 1986.

[20] F. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, “A CFAR adaptive matched filter detector,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[21] Y. Abramovich, M. Rangaswamy, B. Johnson, P. Corbell, and N. Spencer, “Performance analysis of two-dimensional parametric STAP for airborne radar using KASSPER data,” *IEEE Trans. Aerosp. Electron. Syst.*, submitted for publication.

[22] M. Bakonyi and H. Woerdeman, “Maximum entropy elements in the intersection of an affine space and cone of positive definite matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 18, no. 2, pp. 369–376, Apr. 1995.

[23] H. Lev-Ari, S. Parker, and T. Kailath, “Multidimensional maximum-entropy extension,” *IEEE Trans. Inf. Theory*, vol. 35, pp. 497–508, May 1989.

[24] H. Nelis, E. Deprettere, and P. Dewilde, “Approximate inversion of partially specified positive definite matrices,” in *Numerical Linear Algebra, Digital Signal Processing, and Parallel Algorithms*, G. Golub and P. Dooren, Eds. New York: Springer, 1991, vol. F70, pp. 559–567, NATO Adv. Sci. Inst..

[25] G. Sonnevend, “Applications of analytic centers,” in *Numerical Linear Algebra, Digital Signal Processing, and Parallel Algorithms*, G. Golub and P. Dooren, Eds., 1991, vol. F70, pp. 617–632, Springer NATO Adv. Sci. Inst..

[26] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, ser. SIAM Studies in Applied Mathematics. Philadelphia, PA: SIAM, 1994, vol. 15.

[27] M. Lundquist and C. Johnson, “Linearly constrained positive definite completions,” *Linear Algebra Appl.*, vol. 150, pp. 195–208, May 1991.

[28] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, ser. SIAM Studies in Applied Mathematics. Philadelphia, PA: SIAM, 1994, vol. 13.

[29] E. Polak, “Algorithms for maximum likelihood constrained covariance estimation,” CSSIP Short Courses Program. Adelaide, Australia, Sep. 2001.

[30] K. Indukumar and V. Reddy, “A note on redundancy averaging,” *IEEE Trans. Signal Process.*, vol. 40, no. 2, pp. 466–469, Feb. 1992.

[31] B. Fritzsche and B. Kirstein, “Inverse problems for positive Hermitian block Toeplitz matrices and nondegenerate Schur sequences,” *Linear Algebra Appl.*, vol. 179, pp. 237–270, Jan. 1993.

[32] Y. Abramovich, B. Johnson, and N. Spencer, “Multivariate spectral reconstruction of STAP covariance matrices: Toeplitz-block solution,” presented at the SAM Conference, Darmstadt, Germany, 2008.

[33] H. Woerdeman, “Matrix and operator extensions,” Ph.D. dissertation, Vrije Universiteit, Amsterdam, The Netherlands, 1989.

[34] H. Dym and I. Gohberg, “Extensions of band matrices with band inverses,” *Linear Algebra Appl.*, vol. 36, pp. 1–24, Mar. 1981.

[35] H. Lev-Ari and T. Kailath, “Autoregressive models for non-stationary discrete time processes,” presented at the 1981 IEEE Int. Symp. Inform. Theory, Santa Monica, CA, 1982.

Yuri I. Abramovich (M’96–SM’06) received the Dipl. Eng. (Hons.) degree in radio electronics and the Cand.Sci. degree (Ph.D. equivalent) in theoretical radio techniques, both from the Odessa Polytechnic University, Odessa, Ukraine, Russia, in 1967 and 1981, respectively, and the D.Sc. degree in radar and navigation from the Leningrad Institute for Avionics, Leningrad, Russia, in 1971.

From 1968 to 1994, he was with the Odessa State Polytechnic University, Odessa, Ukraine, as a Research Fellow, Professor, and ultimately as Vice-Chancellor of Science and Research. From 1994 to 2006, he was at the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), Adelaide, Australia. Since 2000, he has been with the Australian Defence Science and Technology Organisation (DSTO), Adelaide, as Principal Research Scientist, seconded to CSSIP until its closure. His research interests are in signal processing (particularly spatio–temporal adaptive processing, beamforming, signal detection and estimation), its application to radar (particularly over-the-horizon radar), electronic warfare, and communication.

Dr. Abramovich served as Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005. Since 2007, he has served as Associate Editor of IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS.

Ben A. Johnson (S’04) received the B.S. (*cum laude*) degree in physics from Washington State University in 1984 and the M.S. degree in digital signal processing from the University of Southern California, Los Angeles, in 1988. He is currently working towards the Ph.D. degree at the Institute of Telecommunications Research, University of South Australia, focusing on applications of spatio–temporal adaptive processing in HF radar.

From 1984 to 1989, he was a Systems Engineer in airborne radar at Hughes Aircraft Company (now Raytheon). From 1989 to 1998, he was a Senior Radar Engineer in ground-based surveillance systems with Sensis Corporation. Since 1998, he has been with Lockheed Martin (seconded to RLM Management) on the Jindalee over-the-horizon Operational Radar Network (JORN), first as a Senior Test Engineer and then as Technical Director.

Nicholas K. Spencer received the B.Sc. (Hons.) degree in applied mathematics and the M.Sc. degree in computational mathematics both from the Australian National University, Canberra, in 1985 and 1992, respectively.

He has been with the Australian Department of Defence, Canberra; the Flinders University of South Australia, Adelaide; the University of Adelaide; the Australian Centre for Remote Sensing, Canberra; and the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), Adelaide, in the areas of computational and mathematical sciences. He is currently a Senior Researcher at Adelaide Research & Innovation Pty. Ltd. (ARI), Australia. His research interests include array signal processing, parallel and supercomputing, software best-practice, human–machine interfaces, multilevel numerical methods, modeling and simulation of physical systems, theoretical astrophysics, and cellular automata.