

# Generalized Multivariate Analysis of Variance

A unified framework for signal processing in correlated noise

*Aleksandar Dogandžić and Arye Nehorai*

**G**eneralized multivariate analysis of variance (GMANOVA) [1]-[9] and related reduced-rank regression [10]-[15] are general statistical models that comprise versions of regression, canonical correlation, and profile analyses as well as analysis of variance (ANOVA) and covariance in univariate and multivariate settings. It is a powerful and, yet, not very well-known tool. In this article, we develop a unified framework for explaining, analyzing, and extending signal processing methods based on GMANOVA. We show the applicability of this framework to a number of detection and estimation problems in signal processing and communications and provide new and simple ways to derive numerous existing algorithms for

▲ synchronization and space-time channel and noise estimation in [16]-[33]

▲ space-time symbol detection in [23] and [30]-[38]

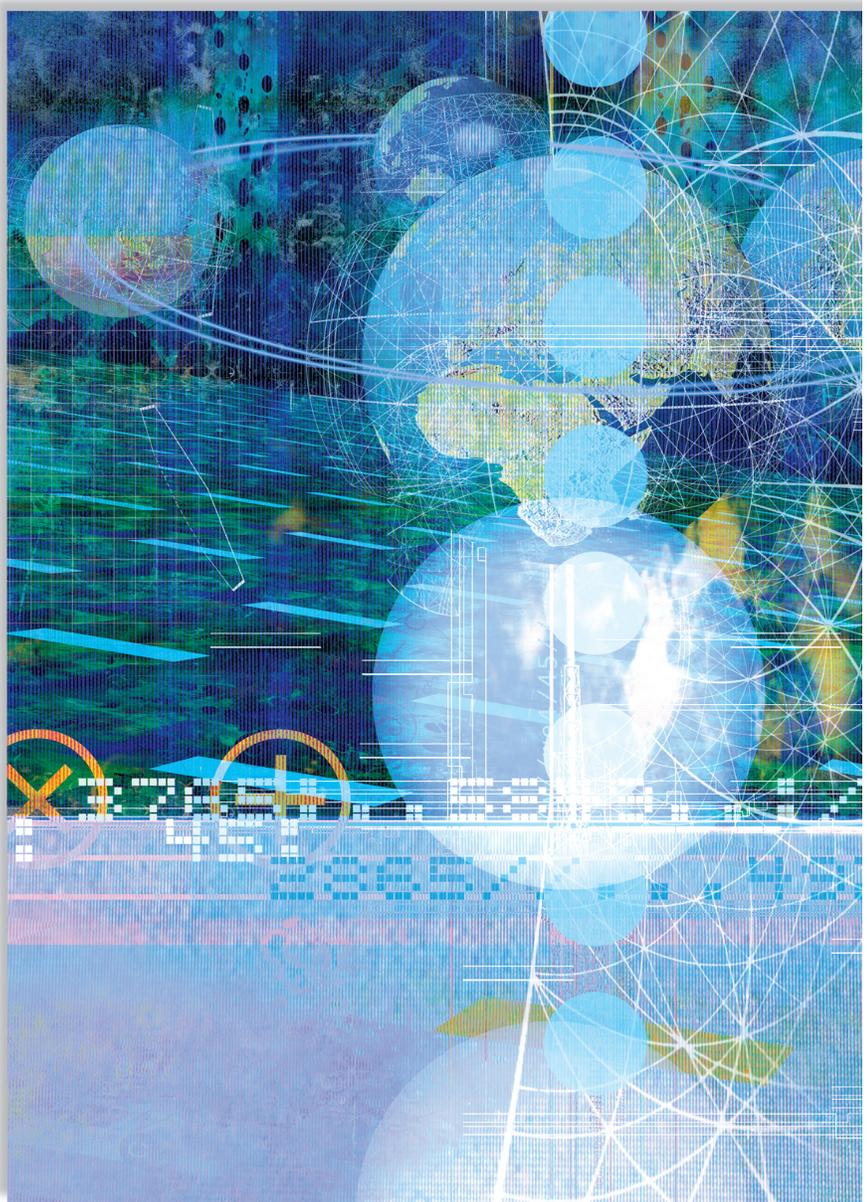
▲ blind and semiblind channel equalization, estimation, and signal separation in [39]-[44]

▲ source location using parametric signal models in [17], [23], [28], [31], and [45]-[51]

▲ radar target estimation and detection in [45]-[50] and [52]-[55]

▲ spectral analysis [56], [57] and nuclear magnetic resonance (NMR) spectroscopy [58].

Many of the above methods were originally derived “from scratch,” without knowledge of their close relationship with the GMANOVA model. We explicitly show this relationship and present new insights and guidelines for generalizing these methods. We also acknowledge the pioneering works of Brillinger (on frequency wavenumber analysis; see [59]) and Kelly and Forsythe (on radar detection; see [53]) who



©DIGITAL VISION, LTD.

first applied GMANOVA to signal processing problems. Note that special cases of GMANOVA have also been applied to time-delay estimation for proximity acoustic sensors [60], synthetic aperture radar (SAR) [61], [62], inverse SAR (ISAR) of maneuvering targets [63], and hyperspectral image data analysis [64]-[66]; for applications of related reduced-rank regression methods to system identification, see [67] and references therein. Our results could inspire applications of the general framework of GMANOVA to new problems in signal processing. We will present such an application to flaw detection in nondestructive evaluation (NDE) of materials. A promising area for future growth is image processing, as is shown in [61]-[66].

## Problem Formulation and Main Results

### Historical Background

The GMANOVA model was first formulated by Potthoff and Roy [1], who were interested in fitting the following patterned-mean problem:  $E[\mathcal{Y}] = \mathbf{A}\mathbf{X}\mathbf{\Phi}$ , where  $\mathcal{Y}$  is a data matrix whose columns are independent random vectors with common covariance matrix  $\Sigma$ ,  $\mathbf{A}$ , and  $\mathbf{\Phi}$  are known matrices, and  $\mathbf{X}$  is a matrix of unknown regression coefficients. In [1], this model was applied to fitting growth patterns of groups of individuals, hence also the name *growth-curve model* [1]-[8]. (Other common statistical applications are: clinical trials of pharmaceutical drugs, agronomical investigations, and business surveys; see [6]-[9] for illustrative examples.) In [2], Khatri computed maximum likelihood (ML) estimates of  $\mathbf{X}$  and  $\Sigma$  under the multivariate normal model for  $\mathcal{Y}$ . Khatri's results are closely related to the concomitant-variable method, independently developed by Rao [3], [4]. In [68], it was shown that the estimates of the regression coefficients and corresponding generalized likelihood ratio tests developed in [2] are robust when the errors are not normal.

In the following, we describe the measurement model and state the main results and important special cases.

### Measurement Model

We now present the general measurement model that will be examined in this article. Let  $\mathbf{y}(t)$  be an  $m \times 1$  complex data vector (snapshot) received at time  $t$  and assume that we have collected  $N$  snapshots. (Note that in image processing applications,  $\mathbf{y}(t)$  are independent sets of pixel observations, indexed by  $t$  that generally does not correspond to time.) Consider the following model for the received snapshots:

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\phi}(t, \boldsymbol{\eta}) + \mathbf{e}(t), \quad t = 1, \dots, N, \quad (1)$$

where the signal is described by

- ▲ an  $m \times r$  matrix  $\mathbf{A}(\boldsymbol{\theta})$  ( $m \geq r$ )
- ▲  $d \times 1$  vectors  $\boldsymbol{\phi}(t, \boldsymbol{\eta})$ ,  $t = 1, \dots, N$

- ▲ an  $r \times d$  matrix  $\mathbf{X}$  of unknown regression coefficients. Here,  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  are parameter vectors (unknown in general) and  $\mathbf{e}(t)$  is temporally white and circularly symmetric zero-mean complex Gaussian noise vector with unknown positive definite spatial covariance  $\Sigma$ , i.e.,

$$E[\mathbf{e}(t)\mathbf{e}(\tau)^H] = \Sigma \delta_{t-\tau}, \quad t, \tau \in \{1, 2, \dots, N\}. \quad (2)$$

In sensor array processing applications,  $\boldsymbol{\theta}$  is usually a vector of *spatial* parameters describing source locations, and  $\mathbf{A}(\boldsymbol{\theta})$  is the array response (steering) matrix. The vector  $\boldsymbol{\eta}$  usually consists of *temporal* parameters. More details on the choices of  $\boldsymbol{\eta}$  in various applications will be given later. Also,  $\delta_t$  denotes the Kronecker delta symbol and " $^H$ " the Hermitian (conjugate) transpose. Our goal is to present methods for

- ▲ estimating the unknown signal and noise parameters  $\mathbf{X}$ ,  $\boldsymbol{\theta}$ ,  $\boldsymbol{\eta}$ , and  $\Sigma$
- ▲ detecting the presence of signal (e.g., in radar)
- ▲ demodulating the received signal (communications).

### Main Results

We present the basic results of this article. Details of their derivation are relegated to the Appendix. First, define the following matrices:

$$\mathcal{Y} = [\mathbf{y}(1) \cdots \mathbf{y}(N)], \quad (3a)$$

$$\Phi(\boldsymbol{\eta}) = [\boldsymbol{\phi}(1, \boldsymbol{\eta}) \cdots \boldsymbol{\phi}(N, \boldsymbol{\eta})], \quad (3b)$$

$$\hat{\mathbf{R}}_{yy} = (1/N) \cdot \mathcal{Y}\mathcal{Y}^H, \quad (3c)$$

$$\hat{\mathbf{R}}_{\phi\phi} = (1/N) \cdot \Phi(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})^H, \quad (3d)$$

$$\hat{\mathbf{R}}_{y\phi} = \hat{\mathbf{R}}_{\phi y}^H = (1/N) \cdot \mathcal{Y}\Phi(\boldsymbol{\eta})^H, \quad (3e)$$

$$\begin{aligned} \hat{\mathbf{S}}_{y|\phi} &= \hat{\mathbf{R}}_{yy} - \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \hat{\mathbf{R}}_{\phi y}^H \\ &= (1/N) \cdot \mathcal{Y} \left[ \mathbf{I}_N - \Pi(\Phi(\boldsymbol{\eta})^H) \right] \mathcal{Y}^H \end{aligned} \quad (3f)$$

$$\hat{\mathbf{T}}_A = \mathbf{A}(\boldsymbol{\theta}) \left[ \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta}) \right]^{-1} \mathbf{A}(\boldsymbol{\theta})^H, \quad (3g)$$

where  $\mathbf{I}_m$  denotes the identity matrix of size  $m$ ,  $\Pi(\mathbf{B}) = \mathbf{B}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H$  the projection matrix onto the column space of a matrix  $\mathbf{B}$ , and " $^{-1}$ " a generalized inverse of a matrix, respectively. (A generalized inverse of a matrix  $\mathbf{A}$  is defined as any matrix  $\mathbf{A}^-$  such that  $\mathbf{A}\mathbf{A}^- \mathbf{A} = \mathbf{A}$ ; see, e.g., [5, ch. 1.6] and [69, ch. 9].) Note that

- ▲  $\hat{\mathbf{R}}_{yy}$  is the sample correlation matrix of the received data  $\mathbf{y}(t)$
- ▲  $\hat{\mathbf{R}}_{\phi\phi}$  is the sample correlation matrix of  $\boldsymbol{\phi}(t, \boldsymbol{\eta})$

▲  $\hat{\mathbf{R}}_{y\phi}$  is the sample cross-correlation matrix between  $\mathbf{y}(t)$  and  $\phi(t, \boldsymbol{\eta})$

▲  $\hat{\mathbf{S}}_{y|\phi}$  is the sample correlation matrix of the received data projected onto the space orthogonal to the row space of  $\Phi(\boldsymbol{\eta})$ .

Here,  $\hat{\mathbf{S}}_{y|\phi}$ ,  $\hat{\mathbf{R}}_{\phi\phi}$ , and  $\hat{\mathbf{R}}_{y\phi}$  are functions of  $\boldsymbol{\eta}$ , and  $\hat{\mathbf{T}}_A$  is a function of  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ . To simplify the notation, we omit these dependencies throughout this article. Assuming that

$$N \geq \text{rank}[\Phi(\boldsymbol{\eta})] + m, \quad (4)$$

$\hat{\mathbf{S}}_{y|\phi}$  will be a positive definite matrix (with probability one), and the maximum likelihood (ML) estimates of  $\boldsymbol{X}$  and  $\boldsymbol{\Sigma}$  for known  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  are (see the Appendix)

$$\begin{aligned} \hat{\mathbf{X}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = & \left[ \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta}) \right]^{-1} \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \\ & + \left[ \mathbf{I}_r - \mathbf{A}(\boldsymbol{\theta})^{-1} \mathbf{A}(\boldsymbol{\theta}) \right] \Xi_1 \\ & + \mathbf{A}(\boldsymbol{\theta})^H \Xi_2 \left[ \mathbf{I}_d - \Phi(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})^{-1} \right], \end{aligned} \quad (5a)$$

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = & \hat{\mathbf{S}}_{y|\phi} + \left( \mathbf{I}_m - \hat{\mathbf{T}}_A \hat{\mathbf{S}}_{y|\phi}^{-1} \right) \\ & \cdot \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \hat{\mathbf{R}}_{y\phi}^H \left( \mathbf{I}_m - \hat{\mathbf{T}}_A \hat{\mathbf{S}}_{y|\phi}^{-1} \right)^H, \end{aligned} \quad (5b)$$

where  $\Xi_1$  and  $\Xi_2$  are arbitrary matrices (of appropriate dimensions). Here,  $\mathbf{I}_r - \mathbf{A}(\boldsymbol{\theta})^{-1} \mathbf{A}(\boldsymbol{\theta})$  is a matrix whose columns span the space orthogonal to the column space of  $\mathbf{A}(\boldsymbol{\theta})^H$  and  $[\mathbf{I}_d - \Phi(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})^{-1}]^H$  is a matrix whose columns span the space orthogonal to the column space of  $\Phi(\boldsymbol{\eta})$ . Therefore, premultiplying (5a) by  $\mathbf{A}(\boldsymbol{\theta})$  and postmultiplying by  $\Phi(\boldsymbol{\eta})$  reduces the second and third terms in (5a) to zero, implying that the estimate of the mean  $\mathbf{A}(\boldsymbol{\theta})\hat{\mathbf{X}}(\boldsymbol{\theta}, \boldsymbol{\eta})\Phi(\boldsymbol{\eta})$  is unique (and is equal to (A.17) in the Appendix).

For unknown  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ , their ML estimates  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\eta}}$  can be obtained by maximizing the concentrated likelihood function (see the Appendix):

$$\text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\hat{\mathbf{R}}_{yy}|}{|\hat{\mathbf{R}}_{yy} - \hat{\mathbf{T}}_A \hat{\mathbf{S}}_{y|\phi}^{-1} \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \hat{\mathbf{R}}_{y\phi}^H|}, \quad (6)$$

where  $|\cdot|$  denotes the determinant. (Concentrated likelihood function is also known as the profile likelihood; see [70, ch. 7.2.4.] for its definition and properties) Here, the ML estimates of  $\boldsymbol{X}$  and  $\boldsymbol{\Sigma}$  follow by substituting  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  in (5a) and (5b) with  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\eta}}$ . In [31], we also compute closed-form Cramér-Rao bound expressions for  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ .

$$\text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\hat{\mathbf{R}}_{yy}|}{|\hat{\mathbf{S}}_{y|\phi} + \hat{\mathbf{S}}_{y|\phi} \mathbf{A}_\perp(\boldsymbol{\theta}) \left[ \mathbf{A}_\perp(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}_\perp(\boldsymbol{\theta}) \right]^{-1} \mathbf{A}_\perp(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \hat{\mathbf{R}}_{y\phi}^H|} \quad (7)$$

## Detection

Expression (6) is written in the form of a generalized likelihood ratio (GLR) test statistic for testing  $H_0: \boldsymbol{X} = \mathbf{0}$  versus  $H_1: \boldsymbol{X} \neq \mathbf{0}$  (i.e., detecting the presence of signal) for the case of known  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ . (See, e.g., [4, p. 418], [71], and [74, ch. 6.4.2.] for the definition of the generalized likelihood ratio test.) The GLR test computes the ratio of likelihood functions under the two hypotheses, with unknown parameters ( $\boldsymbol{X}$  and  $\boldsymbol{\Sigma}$  under  $H_1$  and  $\boldsymbol{\Sigma}$  under  $H_0$ ) replaced by their ML estimates; see also the Appendix. If  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  are unknown, the GLR test compares  $\max_{\boldsymbol{\theta}, \boldsymbol{\eta}} \text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \text{GLR}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$  with a threshold. Since (6) is concentrated with respect to the ML estimates of the nuisance parameters ( $\boldsymbol{\Sigma}$  in this case), it is also the maximized relative likelihood, as defined in [72]. Under  $H_0$  and assuming known  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ ,  $1 / \text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta})$  is distributed as complex Wilks' lambda; see [53] and [73]. Since Wilks' lambda distribution does not depend on the unknown parameters ( $\boldsymbol{\Sigma}$  in this case), we can compute a threshold (with which the above test statistic should be compared) that maintains a constant probability of false alarm. Such a detector is referred to as a constant false alarm rate (CFAR) detector; see, e.g., [74].

## GLR as a Function of $\mathbf{A}_\perp(\boldsymbol{\theta})$

In some applications, it may be convenient to express the above GLR test statistic in terms of a matrix  $\mathbf{A}_\perp(\boldsymbol{\theta})$  whose columns span the space orthogonal to the column space of  $\mathbf{A}(\boldsymbol{\theta})$  [see (7) at the bottom of the page], which follows by applying Lemma 2 from the Appendix, with  $\mathbf{S} = \hat{\mathbf{S}}_{y|\phi}$  and  $\mathbf{A}_\perp = \mathbf{A}_\perp(\boldsymbol{\theta})$ , to (6). For example, if  $\mathbf{A}(\boldsymbol{\theta})$  is a Vandermonde matrix, we can easily construct a corresponding  $\mathbf{A}_\perp(\boldsymbol{\theta})$  and apply polynomial-rooting based ideas to estimate  $\boldsymbol{\theta}$ ; see e.g., [75] and [76].

## GLR for Full-Rank $\mathbf{A}(\boldsymbol{\theta})$

If  $\mathbf{A}(\boldsymbol{\theta})$  has full rank  $r$ , the second term in (5a) becomes zero, and (6) simplifies to

$$\text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta})|}{|\mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{yy}^{-1} \mathbf{A}(\boldsymbol{\theta})|}, \quad (8)$$

see [31, App. A]. (In sensor array processing applications, (8) can be viewed as the ratio of the Capon spectral estimate in the direction  $\boldsymbol{\theta}$  using the data  $\mathcal{Y}$ , and the Capon spectral estimate in the direction  $\boldsymbol{\theta}$  using the projection of the data onto the space orthogonal to the row space of  $\Phi(\boldsymbol{\eta})$ . In other words, it is the overall power arriving from the direction  $\boldsymbol{\theta}$ , normalized by the power of the noise only, arriving from the same direction  $\boldsymbol{\theta}$  [59].) We will use (8) shortly to derive the reduced-rank regression equations in (14) and the corresponding GLR expression in (13).

### GLR for Full-rank $\mathbf{A}(\boldsymbol{\theta})$ and $\Phi(\boldsymbol{\eta})$

If, in addition to  $\mathbf{A}(\boldsymbol{\theta})$ ,  $\Phi(\boldsymbol{\eta})$  has full rank (equal to  $d$ ), then both the second and third terms in (5a) are zero,  $\hat{\mathbf{X}}(\boldsymbol{\theta}, \boldsymbol{\eta})$  is unique, and another interesting expression for  $\text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta})$  follows (see [31]):

$$\text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\hat{\mathbf{R}}_{\phi\phi} - \hat{\mathbf{R}}_{y\phi}^H \mathcal{W}(\boldsymbol{\theta}) \hat{\mathbf{R}}_{y\phi}|}{|\hat{\mathbf{S}}_{\phi|y}(\boldsymbol{\eta})|}, \quad (9)$$

where

$$\mathcal{W}(\boldsymbol{\theta}) = \hat{\mathbf{R}}_{yy}^{-1} - \hat{\mathbf{R}}_{yy}^{-1} \mathbf{A}(\boldsymbol{\theta}) \left[ \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{yy}^{-1} \mathbf{A}(\boldsymbol{\theta}) \right]^{-1} \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{yy}^{-1} \quad (10a)$$

$$\hat{\mathbf{S}}_{\phi|y}(\boldsymbol{\eta}) = \hat{\mathbf{R}}_{\phi\phi} - \hat{\mathbf{R}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{R}}_{y\phi}. \quad (10b)$$

Assuming  $d=1$ ,  $\Phi(\boldsymbol{\eta})$  becomes a  $1 \times N$  vector,  $\hat{\mathbf{R}}_{y\phi} = \hat{r}_{y\phi}$  reduces to an  $m \times 1$  vector, and  $\hat{\mathbf{R}}_{\phi\phi} = \hat{r}_{\phi\phi}$  to a scalar. Then, (9) simplifies to

$$\text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{\hat{r}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \mathbf{A}(\boldsymbol{\theta}) \left[ \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{yy}^{-1} \mathbf{A}(\boldsymbol{\theta}) \right]^{-1} \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{yy}^{-1} \hat{r}_{y\phi}}{\hat{r}_{\phi\phi} - \hat{r}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{r}_{y\phi}}. \quad (11)$$

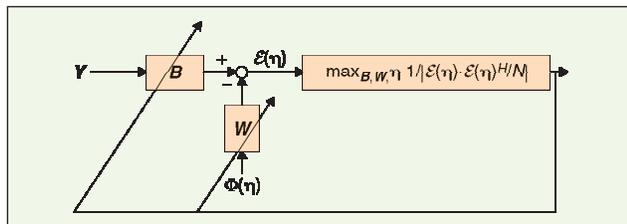
Special cases of the above expression have been used for target parameter estimation with radar arrays [48]-[50] and target detection in hyperspectral images [64], [65], which will be discussed in the Applications Section (Radar Array Processing).

### Reduced-Rank Regression

Consider a nonparametric model for the matrix  $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}$ , i.e., assume that it is completely unknown having full rank  $r \leq \min(d, m)$ . To solve this problem, it is useful to perform eigenvalue decomposition of the following matrix:

$$\hat{\mathbf{R}}_{yy}^{-1/2} \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \hat{\mathbf{R}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1/2} = \hat{\mathbf{U}} \hat{\boldsymbol{\Lambda}}^2 \hat{\mathbf{U}}^H, \quad (12)$$

where  $\hat{\boldsymbol{\Lambda}}^2 = \text{diag}\{\hat{\lambda}^2(1), \hat{\lambda}^2(2), \dots, \hat{\lambda}^2(m)\}$  and  $\hat{\lambda}(1) \geq \hat{\lambda}(2) \geq \dots \geq \hat{\lambda}(m) \geq 0$ . [Here,  $\mathbf{R}^{1/2}$  denotes a Hermitian square root of a Hermitian matrix  $\mathbf{R}$ , and  $\mathbf{R}^{-1/2} = (\mathbf{R}^{1/2})^{-1}$ .] Again, for notational simplicity we omit the dependence of the above quantities on  $\boldsymbol{\eta}$ . Note that  $\hat{\lambda}(k)$  are the sample canonical correlations between  $\mathbf{y}(t)$  and  $\boldsymbol{\phi}(t, \boldsymbol{\eta})$ ; see, e.g., [11]. Note that (8) with unstructured  $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}$  can be interpreted as a multivariate



▲ 1. Canonical correlation analysis based on GMANOVA.

Rayleigh quotient and is easily maximized with respect to  $\mathbf{A}$ , yielding

$$\text{GLR}_{\text{low rank}}(\boldsymbol{\eta}) = \prod_{k=1}^r \frac{1}{1 - \hat{\lambda}^2(k)}. \quad (13)$$

For details of the proof, see the derivation in [31, App. B]. The above result is also closely related to the Poincaré separation theorem [4, pp. 64-65], which addresses the problem of maximizing the multivariate Rayleigh quotient. Interestingly,  $\log[\text{GLR}_{\text{low rank}}(\boldsymbol{\eta})]$  is a measure of the (estimated) mutual information between  $\mathbf{y}(t)$  and  $\boldsymbol{\phi}(t, \boldsymbol{\eta})$ ; see [77, sec. 9.2]. Using the results of [31, App. B.], the ML estimates of  $\mathbf{H} = \mathbf{A}\mathbf{X}$  and  $\Sigma$  follow:

$$\hat{\mathbf{H}}_{\text{low rank}}(\boldsymbol{\eta}) = \hat{\mathbf{A}} \hat{\mathbf{X}} = \hat{\mathbf{R}}_{yy}^{1/2} \hat{\mathbf{U}}(r) \hat{\mathbf{U}}(r)^H \hat{\mathbf{R}}_{yy}^{-1/2} \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} + \Xi \cdot [\mathbf{I}_d - \Phi(\boldsymbol{\eta})\Phi(\boldsymbol{\eta})^{-1}], \quad (14a)$$

$$\hat{\Sigma}(\boldsymbol{\eta}) = \hat{\mathbf{R}}_{yy} - \hat{\mathbf{R}}_{yy}^{1/2} \hat{\mathbf{U}}(r) \hat{\boldsymbol{\Lambda}}^2(r) \hat{\mathbf{U}}(r)^H \hat{\mathbf{R}}_{yy}^{1/2}, \quad (14b)$$

where  $\Xi$  is an arbitrary matrix (of appropriate dimensions),  $\hat{\boldsymbol{\Lambda}}^2(r) = \text{diag}\{\hat{\lambda}^2(1), \hat{\lambda}^2(2), \dots, \hat{\lambda}^2(r)\}$ , and  $\hat{\mathbf{U}}(r)$  is the matrix containing the first  $r$  columns of  $\hat{\mathbf{U}}$ . If  $\Phi(\boldsymbol{\eta})$  has full rank, the second term in (14a) disappears, and (14a) and (14b) reduce to the complex versions of the reduced-rank regression and noise covariance estimates in [10]-[14].

### Canonical Correlation Analysis and Reduced-Rank Regression

Consider the problem depicted in Figure 1: we wish to find the  $r \times m$  and  $r \times d$  matrices  $\mathbf{B}$  and  $\mathbf{W}$  that minimize the sample (estimated) geometric mean-square error of  $\mathbf{B}\mathbf{y}(t) - \mathbf{W}\boldsymbol{\phi}(t, \boldsymbol{\eta})$ , or, equivalently, maximize its inverse:

$$l(\boldsymbol{\eta}, \mathbf{B}, \mathbf{W}) = \frac{1}{|(1/N) \cdot \mathcal{E}(\boldsymbol{\eta}) \cdot \mathcal{E}(\boldsymbol{\eta})^H|}, \quad (15)$$

where  $\mathcal{E}(\boldsymbol{\eta}) = \mathbf{B}\boldsymbol{\Upsilon} - \mathbf{W}\Phi(\boldsymbol{\eta})$ ,

subject to the normalizing constraint

$$\mathbf{B}\hat{\mathbf{R}}_{yy}\mathbf{B}^H = \mathbf{I}_r, \quad (16)$$

which prevents the trivial solution (in which  $\mathbf{B}$  and  $\mathbf{W}$  equal zero), and decorrelates the rows of the filtered data matrix  $\mathbf{B}\boldsymbol{\Upsilon}$ . The optimal  $\mathbf{B}$  and  $\mathbf{W}$  for the above problem are (see [42]):

$$\hat{\mathbf{B}}(\boldsymbol{\eta}) = \hat{\mathbf{U}}(r)^H \hat{\mathbf{R}}_{yy}^{-1/2}, \quad (17a)$$

$$\hat{\mathbf{W}}(\boldsymbol{\eta}) = \hat{\mathbf{B}}(\boldsymbol{\eta}) \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \quad (17b)$$

and

$$l(\boldsymbol{\eta}, \hat{\mathbf{B}}(\boldsymbol{\eta}), \hat{\mathbf{W}}(\boldsymbol{\eta})) = \text{GLR}_{\text{low rank}}(\boldsymbol{\eta}) \quad (18)$$

is exactly the GLR expression for reduced-rank regression in (13). (Interestingly, a stronger result holds: the optimal  $\mathbf{B}$  and  $\mathbf{W}$  in (17) *simultaneously* minimize all the eigenvalues of the sample mean-square error matrix  $\mathcal{E}(\boldsymbol{\eta}) \cdot \mathcal{E}(\boldsymbol{\eta})^H \ll N$  subject to (16); see [42].) Note that the elements of  $\hat{\mathbf{B}}(\boldsymbol{\eta})\mathbf{y}(t)$  and  $\hat{\mathbf{W}}(\boldsymbol{\eta})\boldsymbol{\phi}(t, \boldsymbol{\eta})$  are the estimated canonical variates of  $\mathbf{y}(t)$  and  $\boldsymbol{\phi}(t, \boldsymbol{\eta})$  (see, e.g., [11] and [42]). The above results can be used to derive blind adaptive signal extraction algorithms in [40]; see the Applications section (Wireless Communications).

### MANOVA

Multivariate analysis of variance (MANOVA) is an important special case of GMANOVA where  $\mathbf{A}(\boldsymbol{\theta}) \equiv \mathbf{I}_m$ , and hence the coefficient matrix becomes  $\mathbf{H} = \mathbf{X}$ . Then, the measurement model (1) simplifies to

$$\mathbf{y}(t) = \mathbf{H}\boldsymbol{\phi}(t, \boldsymbol{\eta}) + \mathbf{e}(t), \quad t = 1, \dots, N. \quad (19)$$

The MANOVA model dates back to the first half of the 20th century and is a standard part of modern textbooks on multivariate statistical analysis; see, e.g., [4]-[7], [9], [11], [78]. The GLR in (8) and ML estimates of  $\mathbf{H} = \mathbf{X}$  and  $\Sigma$  simplify to [using (5) or (14)]

$$\text{GLR}(\boldsymbol{\eta}) = \frac{|\hat{\mathbf{R}}_{yy}|}{|\hat{\mathbf{S}}_{y|\phi}|}, \quad (20a)$$

$$\hat{\mathbf{H}}(\boldsymbol{\eta}) = \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} + \Xi \cdot [\mathbf{I}_d - \boldsymbol{\Phi}(\boldsymbol{\eta})\boldsymbol{\Phi}(\boldsymbol{\eta})^{-1}], \quad (20b)$$

$$\hat{\Sigma}(\boldsymbol{\eta}) = \hat{\mathbf{S}}_{y|\phi}, \quad (20c)$$

where, as before,  $\Xi$  is an arbitrary matrix of appropriate dimensions. The above GLR can be used for noncoherent detection of space-time codes, as will be discussed in the Applications section (Wireless Communications). Its recursive implementation was derived in [31]. Interestingly, if  $\boldsymbol{\Phi}(\boldsymbol{\eta})$  has full rank (equal to  $d$ ) and  $d < m$ , it can be shown that the concentrated likelihood function in (20a) increases by iterating between the following two steps.

*Step 1:* Fix  $\boldsymbol{\eta}$  and compute  $\boldsymbol{\Omega} = \hat{\boldsymbol{\Omega}}(\boldsymbol{\eta})$  using

$$\hat{\boldsymbol{\Omega}}(\boldsymbol{\eta}) = [\hat{\mathbf{H}}(\boldsymbol{\eta})^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{H}}(\boldsymbol{\eta})]^{-1} \cdot \hat{\mathbf{H}}(\boldsymbol{\eta})^H \hat{\mathbf{R}}_{yy}^{-1} \boldsymbol{\gamma} \quad (21)$$

[where  $\hat{\mathbf{H}}(\boldsymbol{\eta}) = \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1}$ ].

*Step 2:* Fix  $\boldsymbol{\Omega}$  and find  $\boldsymbol{\eta}$  that minimizes

$$|[\boldsymbol{\Omega} - \boldsymbol{\Phi}(\boldsymbol{\eta})] \cdot [\boldsymbol{\Omega} - \boldsymbol{\Phi}(\boldsymbol{\eta})]^H|. \quad (22)$$

The derivation of this result is based on identity (18); see [42]. The above iteration will be used in the following discussion to develop blind equalization (DW-ILSP and LSCMA) algorithms; see the Applications section (Wireless Communications). An alternative way to maximize (20a) is by using the *cyclic ML* approach in [37, sec. III-D]; see also [42, sec. V].

In the following, we review several important signal processing applications of GMANOVA and MANOVA models.

## Applications

We discuss the applications of GMANOVA to radar array processing, spectral analysis, and wireless communications. We also derive a multivariate energy detector and outline how it can be applied to NDE flaw detection in correlated interference.

### Radar Array Processing

#### Kelly's Detector and Extensions

Assume that an  $n$ -element radar array receives  $P$  pulse returns, where each pulse provides  $N$  range-gate samples. After collecting spatio-temporal data from the  $t$ th range gate into a vector  $\mathbf{y}(t)$  (of size  $m = nP$ ), we search for the presence of targets in one range gate at a time. Without loss of generality, let  $t = 1$  be under test. Then, this radar array detection problem can be formulated within the GMANOVA framework in (1) with

$$\boldsymbol{\Phi}(\boldsymbol{\eta}) = [1, 0, 0, \dots, 0] \text{ of size } 1 \times N, \quad (23)$$

▲  $\mathbf{X} = \mathbf{x}$ , an  $r \times 1$  vector of target amplitudes

▲  $\mathbf{A}(\boldsymbol{\theta})$ , an  $m \times r$  spatiotemporal steering matrix of the targets

▲  $\boldsymbol{\theta}$ , a vector of target parameters, e.g., directions of arrival (DOAs) and Doppler shifts; see [52] and [79].

We wish to test  $H_0: \mathbf{x} = \mathbf{0}$  (targets absent) versus  $H_1: \mathbf{x} \neq \mathbf{0}$  (targets present). The unknown noise covariance  $\Sigma$  accounts for broadband noise, clutter, and jamming. To be able to estimate  $\Sigma$ , we need noise-only snapshots  $\mathbf{y}(2), \mathbf{y}(3), \dots, \mathbf{y}(N)$ , where  $N \geq m + 1$ ; see (4). In [52], Kelly derived the GLR test for the above problem assuming one target ( $r = 1$ ). It was originally derived from scratch, but Kelly and Forsythe recognized its close relationship with GMANOVA in [53].

We now show how celebrated Kelly's detector and its extensions follow from the GMANOVA framework. Collecting all noise-only snapshots into one matrix

$$\mathbf{Z} = [\mathbf{y}(2) \ \mathbf{y}(3) \ \dots \ \mathbf{y}(N)], \quad (24)$$

and substituting (23)-(24) into (3), we obtain

$$N\hat{\mathbf{S}}_{y|\phi} = \mathbf{Z}\mathbf{Z}^H, \quad (25a)$$

$$N\hat{\mathbf{R}}_{y\phi} = \mathbf{y}(1), \quad (25b)$$

$$N\hat{\mathbf{R}}_{\phi\phi} = \mathbf{1}. \quad (25c)$$

After substituting (25) into (6), using the determinant formula  $|\mathbf{I} + \mathbf{a}\mathbf{b}^H| = 1 + \mathbf{b}^H \mathbf{a}$  (see, e.g., [69, cor. 18.1.3 at p. 416]) and applying the monotonic transformation  $1 - 1/\text{GLR}(\boldsymbol{\theta})$ , we have (26), found at the bottom of the page, which is a multivariate extension (for  $r > 1$ ) of the Kelly's detector.

The above detector can be further generalized to simultaneously testing multiple ( $d$ ) snapshots. Without loss of generality, choose the first  $d$  snapshots to be under test:  $\mathbf{Y}_T = [\mathbf{y}(1) \cdots \mathbf{y}(d)]$ . This problem easily fits the GMANOVA framework in (1) with

$$\mathbf{Y} = [\mathbf{Y}_T, \mathbf{Z}], \quad (27a)$$

$$\mathbf{Z} = [\mathbf{y}(d+1) \ \mathbf{y}(d+2) \cdots \mathbf{y}(N)], \quad (27b)$$

$$\boldsymbol{\Phi}(\boldsymbol{\eta}) = [\mathbf{I}_d, \mathbf{0}], \quad (27c)$$

where  $N \geq m + d$ ; see (4). Substituting (27) into (6) yields

$$\begin{aligned} \text{GLR}(\boldsymbol{\theta}) = & \left| \mathbf{I}_d - \left[ \mathbf{I}_d + \mathbf{Y}_T^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{Y}_T \right]^{-1} \right. \\ & \cdot \mathbf{Y}_T^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{A}(\boldsymbol{\theta}) \\ & \cdot \left[ \mathbf{A}(\boldsymbol{\theta})^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{A}(\boldsymbol{\theta}) \right]^{-1} \\ & \left. \cdot \mathbf{A}(\boldsymbol{\theta})^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{Y}_T \right|^{-1} \end{aligned}$$

which is a multivariate extension (for  $r > 1$ ) of Wang and Cai's detector in [54]. Indeed, for one target ( $r = 1$ ) we have  $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})$ , and the above expression simplifies to (28) at the bottom of the page, which is exactly the detector in [54], originally derived "from scratch."

$$\text{GLR}_{\text{Kelly}}(\boldsymbol{\theta}) = 1 - \frac{1}{\text{GLR}(\boldsymbol{\theta})} = \frac{\mathbf{y}(1)^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{A}(\boldsymbol{\theta}) \left[ \mathbf{A}(\boldsymbol{\theta})^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{A}(\boldsymbol{\theta}) \right]^{-1} \mathbf{A}(\boldsymbol{\theta})^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{y}(1)}{1 + \mathbf{y}(1)^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{y}(1)} \quad (26)$$

$$\text{GLR}_{\text{Wang \& Cai}}(\boldsymbol{\theta}) = 1 - \frac{1}{\text{GLR}(\boldsymbol{\theta})} = \frac{\mathbf{a}(\boldsymbol{\theta})^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{Y}_T \left[ \mathbf{I}_d + \mathbf{Y}_T^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{Y}_T \right]^{-1} \mathbf{Y}_T^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{a}(\boldsymbol{\theta})}{\mathbf{a}(\boldsymbol{\theta})^H (\mathbf{Z}\mathbf{Z}^H)^{-1} \mathbf{a}(\boldsymbol{\theta})} \quad (28)$$

### Range, Velocity, and Direction Estimation

Radar array estimation algorithms in [47]-[49] can also be cast into the GMANOVA framework. Equation (1) with

$$\mathbf{A}(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta}), \quad (29a)$$

$$\mathbf{X} = x, \quad (29b)$$

$$\boldsymbol{\Phi}(t, \boldsymbol{\eta}) = s(t - \tau) \cdot \exp[j\omega_D(t - \tau)], \quad t = 1, \dots, N \quad (29c)$$

can be used to model the signal reflected from a single point target and received by an  $m$ -element radar array. Here,  $\mathbf{y}(t)$  contains the array measurements at time  $t$ ,  $\mathbf{a}(\boldsymbol{\theta})$  is the array response to a planewave reflected from the target,  $\boldsymbol{\theta}$  is the vector of DOA parameters (e.g., azimuth and elevation),  $x$  is the (scalar) complex amplitude of the received target signal,  $s(t)$  is the transmitted waveform,  $\tau$  is the time delay (proportional to the target's range), and  $\omega_D$  is the Doppler shift (proportional to the target's radial velocity). Then  $\boldsymbol{\eta} = [\tau, \omega_D]^T$  and

$$\hat{\mathbf{R}}_{y\phi} = \hat{r}_{y\phi} = \frac{1}{N} \sum_{t=1}^N \mathbf{y}(t) \left\{ s(t - \tau) \cdot \exp[j\omega_D(t - \tau)] \right\}^*, \quad (30a)$$

$$\hat{\mathbf{R}}_{\phi\phi} = \hat{r}_{\phi\phi} = \frac{1}{N} \sum_{t=1}^N |s(t - \tau)|^2, \quad (30b)$$

where “ $*$ ” denotes complex conjugation. Assuming that the entire signal  $s(t - \tau)$  is included in the observation interval  $t = 1, 2, \dots, N$  and for integer delay  $\tau$ ,  $\hat{r}_{\phi\phi}$  in (30b) simplifies to the signal energy (which is independent of  $\boldsymbol{\eta}$ ) and  $\hat{r}_{y\phi}$  in (30a) can be computed using the Parseval's identity:

$$\hat{r}_{y\phi} = \frac{1}{2\pi N} \int_{-\pi}^{\pi} \mathcal{Y}_{\text{DFTF}}(\omega) \Phi_{\text{DFTF}}(\omega, \boldsymbol{\eta})^* d\omega, \quad (31)$$

where

$$\mathcal{Y}_{\text{DFTF}}(\omega) = \sum_{t=1}^N \mathbf{y}(t) \exp(-j\omega t), \quad (32a)$$

$$s_{\text{DTFT}}(\omega) = \sum_{t=1}^N s(t) \exp(-j\omega t), \quad (32b)$$

$$\phi_{\text{DTFT}}(\omega, \boldsymbol{\eta}) = s_{\text{DTFT}}(\omega - \omega_D) \cdot \exp(-j\omega \tau). \quad (32c)$$

Note that  $y_{\text{DTFT}}(\omega)$  and  $\phi_{\text{DTFT}}(\omega, \boldsymbol{\eta})$  are the discrete-time Fourier transforms (DTFTs) of  $y(t)$  and  $\phi(t, \boldsymbol{\eta})$ . Substituting (30) into (11), we obtain the concentrated likelihood function

$$\text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = 1 + \frac{|\hat{\mathbf{r}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \mathbf{a}(\boldsymbol{\theta})|^2}{\mathbf{a}(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{yy}^{-1} \mathbf{a}(\boldsymbol{\theta}) \cdot (\hat{r}_{\phi\phi} - \hat{\mathbf{r}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{r}}_{y\phi})}, \quad (33)$$

which needs to be maximized to obtain the ML estimates of  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ ; see also [49]. A similar expression was used for target detection in hyperspectral images; see [65, eq. (3-6)].

If we are interested in estimating  $\boldsymbol{\eta}$  (range and velocity) only, we can apply the MANOVA model. Then, substituting  $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}_m$  and (29c) into (11) yields  $\text{GLR}(\boldsymbol{\eta}) = \hat{r}_{\phi\phi} / (\hat{r}_{\phi\phi} - \hat{\mathbf{r}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{r}}_{y\phi})$ . After the monotonic transformation  $1 - 1 / \text{GLR}(\boldsymbol{\eta})$ , we obtain a simpler form of the GLR:

$$\text{GLR}'(\boldsymbol{\eta}) = \frac{\hat{\mathbf{r}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{r}}_{y\phi}}{\hat{r}_{\phi\phi}}, \quad (34)$$

see also [49] and [64], where it was applied to target detection in hyperspectral images. To maximize (33) and (34) with respect to noninteger delays  $\tau$ , we can use (31) to compute  $\hat{r}_{y\phi}$ ; see also [74, sec. 7.5].

If  $s(t) \equiv 1$  and  $\tau \equiv 0$ , then  $\hat{r}_{\phi\phi} = 1$  and  $\hat{r}_{y\phi}$  in (30a) becomes proportional to the DTFT of  $y(t)$ , evaluated at  $\omega_D$ . This scenario, analyzed extensively in [48], implies that matched filtering has been performed beforehand [i.e., the snapshot  $y(t)$  corresponds to the matched-filtered return from the  $t$ th pulse] and the DOA and Doppler shift are estimated using the filtered data. Under this model, (33)-(34) reduce to [48, eqs. (16) and (32)], and (11) to the concentrated likelihood function for low-angle target estimation in [50, eq. (32)].

### Spectral Analysis

Consider estimating the parameters of a single complex sinusoid from a noisy discrete-time complex data sequence  $y(t)$ . Using common spectral estimation methodology, construct  $N$  data snapshots as follows:

$$y(t) = [y(t), y(t+1), \dots, y(t+m-1)]^T, \quad t = 1, 2, \dots, N, \quad (35)$$

and choose

- ▲  $\boldsymbol{\theta} = \boldsymbol{\eta} = \omega$  (angular frequency of the sinusoid)
- ▲  $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{a}(\omega) = [1, \exp(j\omega), \dots, \exp(j(m-1)\omega)]^T$
- ▲  $\boldsymbol{\Phi}(\boldsymbol{\eta}) = \boldsymbol{\phi}(\omega)^T = [\exp(j\omega), \exp(j2\omega), \dots, \exp(jN\omega)]$ .

## We develop a unified framework for explaining, analyzing, and extending signal processing methods based on the generalized multivariate analysis of variance.

Then the concentrated likelihood for estimating  $\omega$  simplifies to (33) with  $\hat{r}_{\phi\phi} = 1$ ,  $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{a}(\omega)$ , and

$$\hat{r}_{y\phi} = \hat{r}_{y\phi}(\omega) = (1/N) \cdot y_{\text{DTFT}}(\omega), \quad (36)$$

see also (32a).

### Amplitude and Phase Estimation of a Sinusoid (APES)

Substituting the above model into (5a) yields the GMANOVA estimate of the complex amplitude  $\mathbf{X} = x$  for given  $\omega$ :

$$\hat{x}(\omega) = \mathbf{h}_{\text{APES}}(\omega)^H \cdot \hat{r}_{y\phi}(\omega), \quad (37)$$

where

$$\mathbf{h}_{\text{APES}}(\omega) = \frac{\hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{a}(\omega)}{\mathbf{a}(\omega)^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{a}(\omega)} \quad (38)$$

is exactly the (forward) amplitude and phase estimation of a sinusoid (APES) filter proposed in [56] and [57]. (An extension to forward-backward APES is straightforward; see [56].) It was derived “from scratch” in [56] using the ML-based approach. An alternative, nonparametric derivation of APES was presented in [57]. Note that  $\hat{\mathbf{S}}_{y|\phi} = \hat{\mathbf{R}}_{yy} - \hat{r}_{y\phi}(\omega) \hat{r}_{y\phi}(\omega)^H$  and its inverse can be efficiently computed using the matrix inversion lemma; see [56, eq. (23)].

### APES for Damped Sinusoids

A 2-D APES filter for damped sinusoids in [58] follows from the GMANOVA framework by choosing (35),  $\mathbf{a}(\boldsymbol{\theta}) = [1, \exp(-\beta + j\omega), \dots, \exp\{(-\beta + j\omega)(m-1)\}]^T$ , and  $\boldsymbol{\phi}(t, \boldsymbol{\eta}) = \exp\{(-\beta + j\omega)t\}$ ,  $t = 1, \dots, N$ , where  $\boldsymbol{\theta} = \boldsymbol{\eta} = [\omega, \beta]^T$  and  $\beta > 0$  is the damping factor. In [58], this filter was applied to NMR spectroscopy.

An extension of APES for chirp signals was derived in [63] and applied to ISAR imaging of maneuvering targets.

### Estimating Frequencies of Multiple Sinusoids

Simultaneous estimation of the frequencies of multiple sinusoids can be easily cast into the GMANOVA framework: choose

$$\boldsymbol{\theta} = \boldsymbol{\eta} = [\omega_1, \omega_2, \dots, \omega_r]^T \quad (39a)$$

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\omega_1) \ \mathbf{a}(\omega_2) \ \dots \ \mathbf{a}(\omega_r)], \quad (39b)$$

$$\boldsymbol{\Phi}(\boldsymbol{\eta}) = \boldsymbol{\Phi}(\boldsymbol{\theta}) = [\varphi(\omega_1) \ \varphi(\omega_2) \ \dots \ \varphi(\omega_r)]^T, \quad (39c)$$

where  $r$  is the number of sinusoids, and estimate the unknown frequencies  $\omega_1, \omega_2, \dots, \omega_r$  by maximizing the GLR expressions in (6)-(9). Here  $\hat{\mathbf{R}}_{y\phi} = (1/N) \cdot [\mathbf{y}_{\text{DFT}}(\omega_1), \mathbf{y}_{\text{DFT}}(\omega_2), \dots, \mathbf{y}_{\text{DFT}}(\omega_r)]$ .

### Wireless Communications

In wireless communications, the simple MANOVA measurement model (19) is by far the most predominantly used, although it is not referred to as such. Here,  $\mathbf{H}$  is known as the *channel response matrix*. The MANOVA estimates of  $\mathbf{H}$  and the noise covariance matrix  $\Sigma$  [given in (20b)-(20c)] and the GLR in (20a) have been utilized in numerous recent algorithms for channel and noise estimation [17], [27], [32], [33], synchronization [18], [19], [20], [21], [25], [26], and symbol detection [23], [31]-[38]. (In most of the above references, the MANOVA equations and corresponding GLR were derived from scratch, see [17]-[21], [25], [26], [33], [35]-[37].) Special cases of the more complex GMANOVA and reduced-rank models have also been applied to channel estimation and synchronization; see [16], [22]-[24], [28]-[31]. For example, the reduced-rank regression results in (14) have been applied to low-rank channel estimation in [22], [23], and [29]-[31]. The temporal parameter vector  $\boldsymbol{\eta}$  typically contains

▲ i) unknown time delays or Doppler shifts, or both (in channel estimation for wireless communications and radar target estimation)

▲ ii) unknown frequencies (spectral analysis)

▲ iii) unknown symbols (blind equalization and noncoherent detection)

▲ iv) unknown phases of the received signal (constant-modulus blind equalization).

The vector  $\boldsymbol{\eta}$  can be estimated by maximizing the GLR( $\boldsymbol{\eta}$ ) functions in (13) and (20a) under the reduced-rank and MANOVA models, respectively. Below, we discuss applications of these models to noncoherent space-time detection and blind and semiblind channel equalization and estimation.

### Noncoherent Space-Time Detection

We derive methods for noncoherent space-time detection in spatially correlated noise with unknown covariance.

We use the MANOVA model in (19) to describe a flat-fading multi-input, multi-output (MIMO) wireless channel with antenna arrays employed at both ends of the wireless link. Here

▲  $\mathbf{y}(t)$  is an  $m \times 1$  measurement vector received by the receiver array at time  $t$

▲  $\boldsymbol{\phi}(t, \boldsymbol{\eta})$  is the  $d \times 1$  vector of symbols transmitted by an array of  $d$  antennas and received by the receiver array at time  $t$ .

The matrix  $\boldsymbol{\Phi}(\boldsymbol{\eta})$  contains one or more space-time codewords to be detected. Assume that the transmitted space-time codewords are uniquely described by  $\boldsymbol{\eta}$ . Then, the MANOVA-based GLR demodulation scheme consists of finding  $\boldsymbol{\eta}$  that maximizes GLR( $\boldsymbol{\eta}$ ) in (20a):

$$\begin{aligned} \hat{\boldsymbol{\eta}}_{\text{GLR}} &= \arg \max_{\boldsymbol{\eta}} \frac{|\hat{\mathbf{R}}_{yy}|}{|\hat{\mathbf{S}}_{y\phi}|} \\ &= \arg \min_{\boldsymbol{\eta}} \left| \boldsymbol{\Upsilon} [\mathbf{I}_N - \boldsymbol{\Pi}(\boldsymbol{\Phi}(\boldsymbol{\eta})^H)] \boldsymbol{\Upsilon}^H \right| \end{aligned} \quad (40)$$

which is exactly the GLR detector proposed and analyzed in [23], [31], and [34]-[37]. The above detector can be viewed as a multivariate extension (accounting for multiple receive antennas and spatially correlated noise) of the multiuser detector in [80]. For a single-input, single-output (SISO) scenario with  $m = d = 1$ , it further reduces to the standard noncoherent detector in, e.g., [81, sec. 5.4]. The logarithm of the GLR expression in (20a) and (40) can be approximated as

$$\begin{aligned} \ln \left[ \frac{|\hat{\mathbf{R}}_{yy}|}{|\hat{\mathbf{S}}_{y\phi}|} \right] &= -\ln \left| \mathbf{I}_N - \boldsymbol{\Pi}(\boldsymbol{\Phi}(\boldsymbol{\eta})^H) \boldsymbol{\Pi}(\boldsymbol{\Upsilon}^H) \right| \\ &\approx \text{tr} \left[ \boldsymbol{\Pi}(\boldsymbol{\Phi}(\boldsymbol{\eta})^H) \boldsymbol{\Pi}(\boldsymbol{\Upsilon}^H) \right], \end{aligned} \quad (41)$$

which is the subspace-invariant detector in [38] and can be viewed as a multivariate extension of (34). Here, the equality follows by using the determinant formula  $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$  (see, e.g., [69, cor. 18.1.2, p. 416]) and the approximate expression is obtained by keeping only the first term in the Taylor-series expansion:

$$\begin{aligned} -\ln \left| \mathbf{I}_N - \boldsymbol{\Pi}(\boldsymbol{\Phi}(\boldsymbol{\eta})^H) \boldsymbol{\Pi}(\boldsymbol{\Upsilon}^H) \right| \\ &= \text{tr} \left[ \boldsymbol{\Pi}(\boldsymbol{\Phi}(\boldsymbol{\eta})^H) \boldsymbol{\Pi}(\boldsymbol{\Upsilon}^H) \right] \\ &\quad + \frac{1}{2} \text{tr} \left\{ \left[ \boldsymbol{\Pi}(\boldsymbol{\Phi}(\boldsymbol{\eta})^H) \boldsymbol{\Pi}(\boldsymbol{\Upsilon}^H) \right]^2 \right\} + \dots \end{aligned} \quad (42)$$

### Blind Equalization

We utilize the proposed GMANOVA framework to derive algorithms for blind channel equalization and signal separation in [39]-[41].

### Iterative Least Squares with Projection (ILSP) and Least-Squares Constant Modulus Algorithm (LSCMA)

We derive ILSP [41] and LSCMA [39] algorithms using the iteration (21)-(22). First, we specialize the MANOVA model in (19) to the single-input, multi-out-

put (SIMO) flat-fading scenario, i.e., assuming that  $d=1$ . Define the vector of (unknown) received symbols:

$$\Phi(\boldsymbol{\eta}) = \boldsymbol{\eta}^T = [s(1), s(2), \dots, s(N)]. \quad (43)$$

Then

$$\hat{\mathbf{R}}_{y\phi} = \hat{\mathbf{r}}_{y\phi} = (1/N) \cdot \sum_{t=1}^N \mathbf{y}(t) s(t)^* \quad (44a)$$

$$\hat{\mathbf{R}}_{\phi\phi} = \hat{\mathbf{r}}_{\phi\phi} = (1/N) \cdot \sum_{t=1}^N |s(t)|^2, \quad (44b)$$

and the concentrated likelihood is given in (34). Hence, we need to find the most likely symbol sequence (43) that maximizes (34). To accomplish this task, apply the iteration (21)-(22) with (43)-(44):

*Step 1:* Fix  $\boldsymbol{\eta}$  and compute

$$\omega(t) = \hat{\omega}(t, \boldsymbol{\eta}) = \frac{\hat{\mathbf{r}}_{\phi\phi}}{\hat{\mathbf{r}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{r}}_{y\phi}} \cdot \hat{\mathbf{r}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \mathbf{y}(t), \quad t=1, 2, \dots, N, \quad (45)$$

[where  $\hat{\mathbf{r}}_{y\phi}$  and  $\hat{\mathbf{r}}_{\phi\phi}$  are defined in (44)].

*Step 2:* Fix  $\omega(t), t=1, 2, \dots, N$ , and minimize

$$\sum_{t=1}^N |\omega(t) - s(t)|^2 \quad (46)$$

with respect to  $\boldsymbol{\eta}$ .

Based on the finite-alphabet property of the received symbols, Step 2 reduces to projecting each  $\omega(t), t=1, 2, \dots, N$  onto finite alphabet. In this case, the above iteration is identical to the decoupled weighted iterative least squares with projection algorithm in (DW-ILSP) [41]. To resolve phase ambiguity, a small number of *known* (training) symbols is typically embedded in the transmission scheme (see [41]); the above algorithm can be easily modified to utilize the training data.

If the transmitted symbols belong to a constant-modulus constellation, we can model the received signal as follows:

$$\Phi(\boldsymbol{\eta}) = [\exp(j\vartheta(1)), \exp(j\vartheta(2)), \dots, \exp(j\vartheta(N))], \quad (47a)$$

$$\boldsymbol{\eta} = [\vartheta(1), \vartheta(2), \dots, \vartheta(N)]^T, \quad (47b)$$

$$\hat{\mathbf{r}}_{y\phi} = (1/N) \cdot \sum_{t=1}^N \mathbf{y}(t) \exp(-j\vartheta(t)), \quad (47c)$$

$$\hat{\mathbf{r}}_{\phi\phi} = \mathbf{1}, \quad (47d)$$

where  $\vartheta(t), t=1, 2, \dots, N$  are the unknown phases. In this case, iteration (21)-(22) simplifies to the following.

*Step 1:* Fix  $\boldsymbol{\eta} = [\vartheta(1), \vartheta(2), \dots, \vartheta(N)]^T$  and compute

$$\omega(t) = \hat{\omega}(t, \boldsymbol{\eta}) = \frac{1}{\hat{\mathbf{r}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{r}}_{y\phi}} \cdot \hat{\mathbf{r}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \mathbf{y}(t), \quad t=1, 2, \dots, N, \quad (48)$$

[where  $\hat{\mathbf{r}}_{y\phi}$  is defined in (47c)].

*Step 2:* Fix  $\omega(t), t=1, 2, \dots, N$ , and update  $\boldsymbol{\eta}$  as follows:

$$\hat{\boldsymbol{\eta}} = [\angle\omega(1), \angle\omega(2), \dots, \angle\omega(N)]^T. \quad (49)$$

The above algorithm is identical to the least-squares constant modulus algorithm (LSCMA) in [39].

The DW-ILSP and LSCMA algorithms were originally derived using approaches very different from the ML-based methodology presented here; see [41] and [39]. Note that our approach provides a framework for extending these algorithms to the MIMO scenario, based on the iteration (21)-(22).

### Spectral Self-Coherence Restoral (SCORE) Algorithms

Consider the problem of “matching” the receiver array measurements  $\mathbf{y}(t)$  with frequency-shifted (by a constant  $\alpha$ ) and possibly conjugated replicas of  $\mathbf{y}(t)$ :

$$\Phi(t, \boldsymbol{\eta}) = \mathbf{y}(t) \exp(j2\pi\alpha t) \text{ or } \Phi(t, \boldsymbol{\eta}) = \mathbf{y}(t)^* \exp(j2\pi\alpha t), \quad (50)$$

see [40, eq. (31)]. We now adopt the reduced-rank canonical correlation model with  $r=1$ , i.e., we wish to minimize the sample mean-square error of

$$\mathbf{B}\mathbf{y}(t) - \mathbf{W}\Phi(t, \boldsymbol{\eta}) = \mathbf{b}^H \mathbf{y}(t) - \mathbf{w}^H \Phi(t, \boldsymbol{\eta}), \quad (51)$$

subject to  $\mathbf{b}^H \hat{\mathbf{R}}_{yy} \mathbf{b}$ ; see (16). Then, the optimal  $\hat{\mathbf{b}}$  and  $\hat{\mathbf{w}}$  are

$$\hat{\mathbf{b}} = \hat{\mathbf{R}}_{yy}^{-1/2} \hat{\mathbf{U}}(1) \quad (52a)$$

$$\hat{\mathbf{W}} = \hat{\mathbf{w}}^H = \hat{\mathbf{U}}(1)^H \hat{\mathbf{R}}_{yy}^{-1/2} \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1}, \quad (52b)$$

see (17). [Note that (50) together with (4) implies that both  $\hat{\mathbf{R}}_{yy}$  and  $\hat{\mathbf{R}}_{\phi\phi}$  are positive definite with probability one. However, due to generality, we present expressions (52) that allow for singular  $\hat{\mathbf{R}}_{\phi\phi}$ , which may be useful in other applications.] The above solutions satisfy

$$\hat{\lambda}^2(1) \cdot \hat{\mathbf{R}}_{yy} \hat{\mathbf{b}} = \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \hat{\mathbf{R}}_{y\phi}^H \hat{\mathbf{b}}, \quad (53a)$$

$$\hat{\lambda}^2(1) \cdot \hat{\mathbf{R}}_{\phi\phi} \hat{\mathbf{w}} = \hat{\mathbf{R}}_{y\phi}^H \hat{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{w}}, \quad (53b)$$

which are exactly the cross-SCORE eigenequations in [40, eqs. (35) and (36)].

## The MANOVA model dates back to the first half of the 20th century and is a standard part of modern textbooks on multivariate statistical analysis.

### Semiblind Channel and Noise Estimation Using the EM Algorithm

Consider a SIMO flat-fading channel described by the following equation:

$$\mathbf{y}(t) = \mathbf{h} \cdot s(t) + \mathbf{e}(t), \quad t = 1, 2, \dots, N, \quad (54)$$

where

- ▲  $\mathbf{h}$  is an unknown  $m \times 1$  channel response vector
- ▲  $s(t)$ ,  $t = 1, 2, \dots, N$  are the unknown symbols received by the array.

The above equation is a special case of the MANOVA model (19) with  $d = 1$ ; this model has also been used to derive the DW-ILSP algorithm. However, unlike the DW-ILSP approach which treats the unknown symbols  $s(t)$  as deterministic parameters, we model them as independent, identically distributed (i.i.d.) random variables that take values from an  $M$ -ary constant-modulus constellation  $\{s_1, s_2, \dots, s_M\}$  with equal probability; the constant-modulus assumption implies that  $|s_n| = 1$ ,  $n = 1, 2, \dots, M$ . (These assumptions can be relaxed, resulting in more cumbersome computations.) As discussed before, to allow unique estimation of the channel  $\mathbf{h}$  (i.e., to resolve the phase ambiguity), we also assume that a small number ( $N_T$ ) of training symbols

$$s_T(\tau), \quad \tau = 1, 2, \dots, N_T \quad (55)$$

is embedded in the transmission scheme. Denote the corresponding snapshots received by the array as  $\mathbf{y}_T(\tau)$ ,  $\tau = 1, 2, \dots, N_T$ . Then, the measurement model (54) holds for the training symbols as well, with  $\mathbf{y}(t)$  and  $s(t)$  replaced by  $\mathbf{y}_T(\tau)$  and  $s_T(\tau)$ , respectively.

In [43] and [44], we treat the unknown symbols as the *unobserved* (or missing) data and combine the MANOVA model with the expectation-maximization (EM) algorithm to estimate the channel  $\mathbf{h}$  and spatial noise covariance  $\Sigma$ . We now sketch the main ideas of this approach. We first computed the joint distribution of  $\mathbf{y}(t)$ ,  $s(t)$  (for  $t = 1, 2, \dots, N$ ), and  $\mathbf{y}_T(\tau)$  (for  $\tau = 1, 2, \dots, N_T$ ), which is also known as the complete-data likelihood function. Using this joint distribution, we then obtained complete-data sufficient statistics for estimating  $\mathbf{h}$  and  $\Sigma$ :

$$\hat{\mathbf{r}}_{y_0} = \frac{1}{N + N_T} \left[ \sum_{t=1}^N \mathbf{y}(t) s(t)^* + \sum_{\tau=1}^{N_T} \mathbf{y}_T(\tau) s_T(\tau)^* \right], \quad (56a)$$

$$\hat{\mathbf{R}}_{yy} = \frac{1}{N + N_T} \left[ \sum_{t=1}^N \mathbf{y}(t) \mathbf{y}(t)^H + \sum_{\tau=1}^{N_T} \mathbf{y}_T(\tau) \mathbf{y}_T(\tau)^H \right] \quad (56b)$$

and observed that the complete-data likelihood belongs to the *exponential family* of distributions, i.e., its logarithm is a linear function of the above *natural sufficient statistics* (see e.g., [82] for the definition and properties of the exponential family). If the complete-data likelihood belongs to the exponential family and if  $N + N_T \geq m + 1$  [see (4)], the EM algorithm is easily derived as follows.

▲ The expectation (E) step is reduced to computing conditional expectations of the complete-data sufficient statistics [in (56)] given the observed data  $\mathbf{y}(t)$ ,  $t = 1, \dots, N$  and  $\mathbf{y}_T(\tau)$ ,  $s_T(\tau)$ ,  $\tau = 1, \dots, N_T$ .

▲ The maximization (M) step is reduced to finding the expressions for the complete-data ML estimates of the unknown parameters  $\mathbf{h}$  and  $\Sigma$  and replacing the complete-data natural sufficient statistics (56) that occur in these expressions with their conditional expectations computed in the E step.

In our problem, the complete-data ML estimates of  $\mathbf{h}$  and  $\Sigma$  follow as a special case (for  $d = 1$ ) of the MANOVA equations in (20b) and (20c):

$$\hat{\mathbf{h}} = \frac{1}{N + N_T} \left[ \sum_{t=1}^N \mathbf{y}(t) s(t)^* + \sum_{\tau=1}^{N_T} \mathbf{y}_T(\tau) s_T(\tau)^* \right], \quad (57a)$$

$$\hat{\Sigma} = \hat{\mathbf{R}}_{yy} - \hat{\mathbf{h}} \hat{\mathbf{h}}^H, \quad (57b)$$

where we used the constant-modulus property of the transmitted symbols. Following the above procedure, we derive the EM algorithm for estimating  $\mathbf{h}$  and  $\Sigma$ :

*Step 1:*

$$\mathbf{h}^{(k+1)} = \frac{1}{N + N_T} \left[ \sum_{t=1}^N \mathbf{y}(t) \frac{\sum_{n=1}^M s_n^* \exp[2\text{Re}\{\mathbf{y}(t)^H (\Sigma^{(k)})^{-1} \mathbf{h}^{(k)} s_n\}]}{\sum_{l=1}^M \exp[2\text{Re}\{\mathbf{y}(t)^H (\Sigma^{(k)})^{-1} \mathbf{h}^{(k)} s_l\}]} + \sum_{\tau=1}^{N_T} \mathbf{y}_T(\tau) s_T(\tau)^* \right], \quad (58a)$$

*Step 2:*

$$(\Sigma^{(k+1)})^{-1} = \mathbf{R}_{yy}^{-1} + \frac{\mathbf{R}_{yy}^{-1} \mathbf{h}^{(k+1)} (\mathbf{h}^{(k+1)})^H \mathbf{R}_{yy}^{-1}}{1 - (\mathbf{h}^{(k+1)})^H \mathbf{R}_{yy}^{-1} \mathbf{h}^{(k+1)}}. \quad (58b)$$

Note that (58a) and (58b) each incorporate both E and M steps. To avoid matrix inversion, we applied the matrix inversion lemma (see, e.g., [69, cor. 18.2.10, p. 424]) to

directly compute the estimates of  $\Sigma^{-1}$ ; see (58b). We now utilize the above channel estimates to detect the unknown transmitted symbols  $s(t)$  (see [43] and [44]):

$$\hat{s}(t) = \arg \max_{s(t) \in \{s_1, s_2, \dots, s_M\}} \text{Re}\{y(t)^H \mathbf{R}_y^{-1} \mathbf{b}^{(\infty)} \cdot s(t)\}, \quad (59)$$

where  $\mathbf{b}^{(\infty)}$  is the ML estimate of  $\mathbf{b}$  obtained from the EM iteration (58a)-(58b).

In Figure 2, we compare symbol error rates of the detector (59) and the DW-ILSP detector in (45)-(46). We consider an array of  $m=5$  receiver antennas. The transmitted symbols were generated from an uncoded QPSK modulated constellation (i.e.,  $M=4$ ) with normalized energy. We added a three-symbol training sequence ( $N_T=3$ ), which was utilized to obtain initial estimates of the channel coefficients. (For further details of the simulation scenario, see [44].) The symbol error rates averaged over random channel realizations are shown as functions of the bit signal-to-noise ratio (SNR) per receiver antenna for block lengths  $N=50, 100$ , and  $150$ . An intuitive explanation for the better performance of the EM-based detector is that the EM algorithm exploits additional information provided by the distribution of the unknown symbols. Note also that the number of real parameters in the random-symbol measurement model equals  $m^2 + 2m$ , and, therefore, is independent of  $N$ . This is in contrast with DW-ILSP and other deterministic ML methods (e.g., [37], [42], and [83]) where the number of parameters grows with  $N$ .

### Other Applications

#### Multivariate Weighted Energy Detector

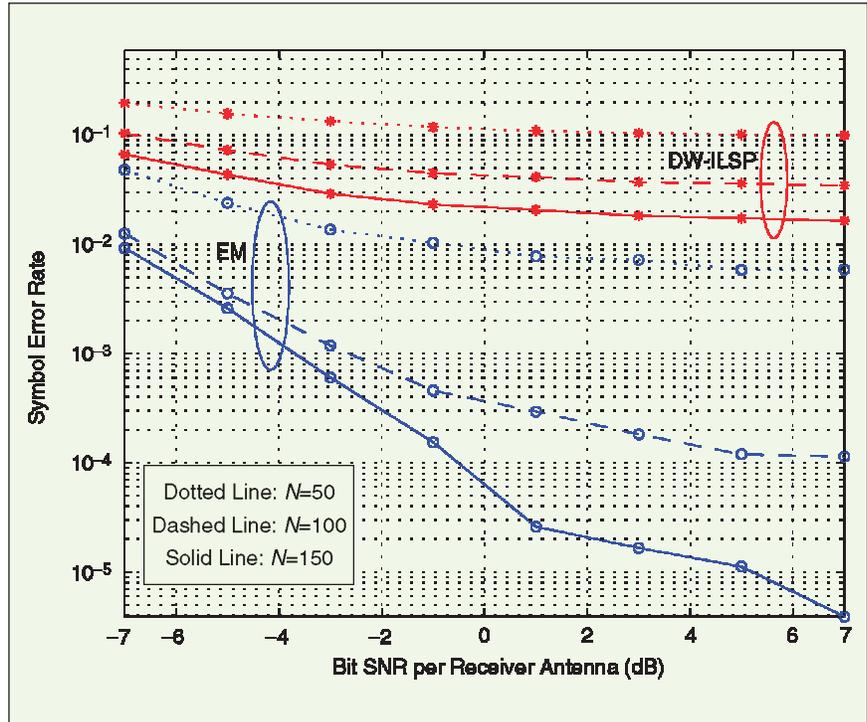
Consider the problem of detecting the presence of a signal in a data matrix under test  $\mathbf{Y}_T$  of size  $m \times d$ , where noise-only data matrix  $\mathbf{Z}$  of size  $m \times (N-d)$  is available, and  $N \geq m+d$ . If we do not have any additional information about the nature of the signal to be detected, we can choose a *nonparametric* model for the signal mean:

$$E[\mathbf{Y}_T] = \mathbf{X}. \quad (60)$$

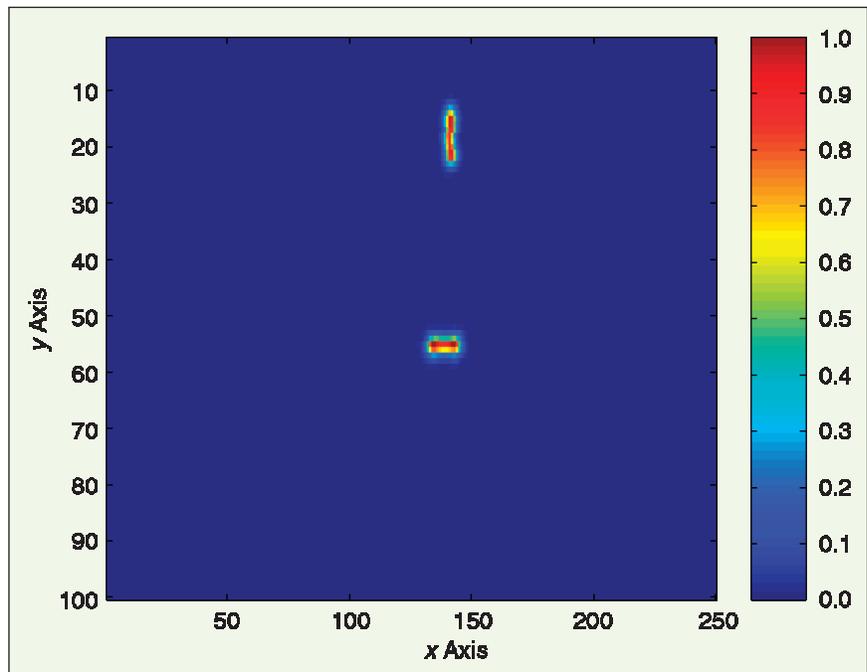
Using the definitions in (27) and  $\mathbf{A}(\Theta) = \mathbf{I}_m$ , we simplify (8) to the following GLR test statistic:

$$\text{GLR} = \frac{|\mathbf{Y}_T \mathbf{Y}_T^H + \mathbf{Z} \mathbf{Z}^H|}{|\mathbf{Z} \mathbf{Z}^H|} \quad (61)$$

for testing  $H_0: \mathbf{X} = \mathbf{0}$  versus  $H_1: \mathbf{X} \neq \mathbf{0}$ . The above statistic can be viewed as a multivariate extension of the classical energy detector; indeed, for  $m=1$ , it simplifies to the en-



▲ 2. Symbol error rates of the EM-based and DW-ILSP detectors as functions of the SNR per receiver antenna.

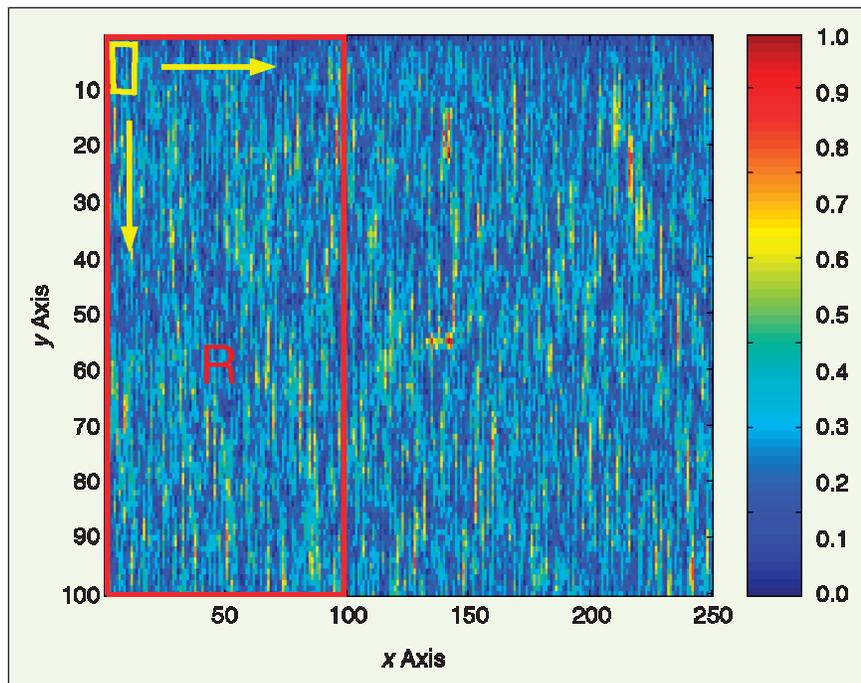


▲ 3. Magnitude plot of low-noise NDE measurements with peak value normalized to one.

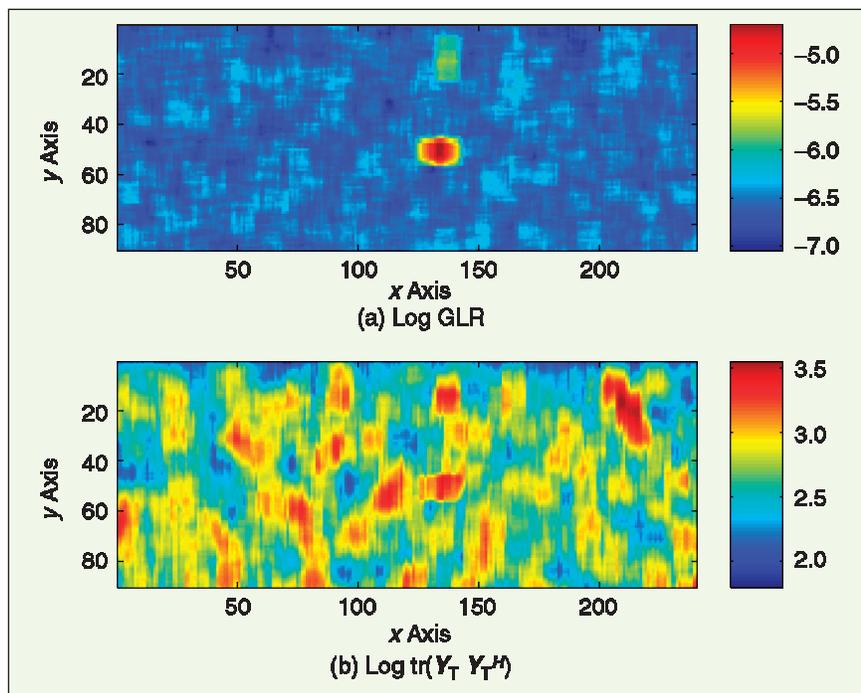
ergy detector in, e.g., [74, ch. 7.3.]. Expression (61) simplifies also when the presence of a signal is tested in one snapshot at a time (i.e.,  $d=1$  and hence  $\Upsilon_T = y_T$ ):

$$\text{GLR} = 1 + y_T^H (ZZ^H)^{-1} y_T. \quad (62)$$

which is the weighted energy detector in [62, eq. (37)] and [66, eq. (20)].



▲ 4. Magnitude plot of noisy NDE measurements.



▲ 5. Logarithms of (a) the proposed GLR test statistic in (61) and (b) classical energy detector for uncorrelated noise in (63).

#### Flaw Detection for Nondestructive Evaluation of Materials:

We now apply the above test to NDE flaw detection in correlated noise; see also [85]. In NDE, correlated noise is typically caused by

- ▲ backscattered “clutter” in ultrasonic NDE array systems (similar to the clutter in radar) [84]
- ▲ random liftoff variations between measurement locations in eddy-current systems [86]. (Liftoff is the distance between the probe and the testpiece surface.)

A key aim of eddy-current NDE is to quantify flaws in conductors using changes of the probe impedance due to defects; see [86]. Figure 3 shows a magnitude plot of low-noise experimental eddy-current impedance measurements in a sample containing two realistic flaws, where each pixel corresponds to a measurement location. The data was collected by scanning the testpiece surface columnwise (parallel to the  $y$  axis). To model liftoff variations, we added complex Gaussian noise, correlated along  $y$  direction (i.e., between rows) and uncorrelated along  $x$  direction (i.e., independent columns). Figure 4 shows a magnitude plot of noisy measurements. We used a region  $\mathbf{R}$  of the image to generate the noise-only data matrix  $\mathbf{Z}$ . A window  $\Upsilon_T$  of size  $m \times d = 10 \times 10$  was swept across the noisy image, as depicted in Figure 4. For each location of the window, we computed the (logarithms of)

- ▲ the proposed GLR test statistic in (61)
- ▲ the classical energy detector for white noise

$$\text{tr}(\Upsilon_T \Upsilon_T^H), \quad (63)$$

which is simply the sum of squared magnitudes of all measurements within the window  $\Upsilon_T$ ; see Figure 5. Clearly, the proposed detector which accounts for noise correlation outperforms the classical detector, which breaks down in this scenario.

### Concluding Remarks

We reviewed GMANOVA and its applications to numerous problems in signal processing and communications. We presented a unified framework for developing GMANOVA-based methods and showed that many existing algorithms readily follow as its special cases. More impor-

tantly, insights gained from this framework allow generalizations of many of these methods. A novel application to flaw detection for nondestructive evaluation of materials was proposed. We hope that our results would lead to successful applications of this powerful tool to new and exciting signal processing problems.

## Acknowledgment

This work was supported by the NSF Industry-University Cooperative Research Program, Center for Nondestructive Evaluation, Iowa State University, the Air Force Office of Scientific Research under Grants F49620-00-1-0083 and F49620-02-1-0339, the National Science Foundation under Grant CCR-0105334, and the Office of Naval Research under Grant N00014-01-1-0681.

*Aleksandar Dogandžić* received the Dipl. Ing. degree (summa cum laude) in electrical engineering from the University of Belgrade, Yugoslavia, in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of Illinois at Chicago (UIC) in 1997 and 2001, respectively. In August 2001, he joined the Department of Electrical and Computer Engineering, Iowa State University, Ames, as an Assistant Professor. His research interests are in statistical signal processing and its applications to wireless communications, nondestructive evaluation of materials, radar, and biomedicine. He received UIC's 2001 Outstanding Thesis Award in the Division of Engineering, Mathematics, and Physical Sciences.

*Arye Nehorai* received the Ph.D. degree from Stanford University. From 1985 to 1995 he was a faculty member at Yale University. In 1995 he joined the University of Illinois at Chicago as a Full Professor where he was later named University Scholar. He is vice president-Publications of the IEEE Signal Processing Society and was editor-in-chief of *IEEE Transactions on Signal Processing*. He served as chair of the IEEE Signal Processing Society Technical Committee on Sensor Array and Multichannel Processing and was corecipient of the 1989 IEEE SPS Senior Award for Best Paper. He was elected Distinguished Lecturer of the IEEE Signal Processing Society from 2004 to 2005. He has been Fellow of the IEEE since 1994 and of the Royal Statistical Society since 1996.

## References

- [1] R.F. Potthoff and S.N. Roy, "A generalized multivariate analysis of variance model useful especially for growth curve problems," *Biometrika*, vol. 51, no. 3/4, pp. 313-326, 1964.
- [2] C.G. Khatri, "A note on a MANOVA model applied to problems in growth curve," *Ann. Inst. Statist. Math.*, vol. 18, pp. 75-86, 1966.
- [3] C.R. Rao, "The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves," *Biometrika*, vol. 52, no. 3/4, pp. 447-458, 1965.
- [4] C.R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.
- [5] M.S. Srivastava and C.G. Khatri, *An Introduction to Multivariate Statistics*. New York: North-Holland, 1979.
- [6] M.S. Srivastava and E.M. Carter, *An Introduction to Applied Multivariate Statistics*. New York, North-Holland, 1983.
- [7] G.A.F. Seber, *Multivariate Observations*. New York: Wiley, 1984.
- [8] A.M. Kshirsagar and W.B. Smith, *Growth Curves*. New York: Marcel Dekker, 1995.
- [9] E.F. Vonesh and V.M. Chinchilli, *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker, 1997.
- [10] T.W. Anderson, "Estimating linear restrictions on regression coefficients for multivariate normal distributions," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 327-351, Sept. 1951.
- [11] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.
- [12] P.M. Robinson, "Identification, estimation and large-sample theory for regressions containing unobservable variables," *Int. Econ. Rev.*, vol. 15, no. 3, pp. 680-692, Oct. 1974.
- [13] M.K.-S. Tso, "Reduced-rank regression and canonical analysis," *J. R. Stat. Soc., Ser. B*, vol. 43, no. 2, pp. 183-189, 1981.
- [14] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Processing*, vol. 44, pp. 3069-3078, Dec. 1996.
- [15] G. Reinsel and R. Velu, "Reduced-rank growth curve models," *J. Stat. Planning and Inference*, vol. 114, no. 1/2, pp. 107-129, June 1, 2003.
- [16] D.M. Dlugos and R.A. Scholtz, "Acquisition of spread spectrum signals by an adaptive array," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1253-1270, Aug. 1989.
- [17] M. Cedervall and A. Paulraj, "Joint channel and space-time parameter estimation," in *Proc. 30th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1996, pp. 375-379.
- [18] D. Zheng, J. Li, S.L. Miller, and E.G. Ström, "An efficient code-timing estimator for DS-CDMA signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 82-89, Jan. 1997.
- [19] Z.-S. Liu, J. Li, and S. Miller, "An efficient code-timing estimator for receiver diversity DS-CDMA systems," *IEEE Trans. Commun.*, vol. 46, pp. 826-835, June 1998.
- [20] A. Jakobsson, A.L. Swindlehurst, D. Astély, and C. Tidedav, "A blind frequency domain method for DS-CDMA synchronization using antenna arrays," in *Proc. 32nd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1998, pp. 1848-1852.
- [21] D. Astély, A. Jakobsson, and A.L. Swindlehurst, "Burst synchronization on unknown frequency selective channels with co-channel interference using an antenna array," in *Proc. 49th Veh. Technol. Conf.*, Houston, TX, May 1999, pp. 2363-2367.
- [22] E. Lindskog and C. Tidedav, "Reduced rank channel estimation," in *Proc. 49th Veh. Technol. Conf.*, Houston, TX, May 1999, pp. 1126-1130.
- [23] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation in unknown spatially correlated noise," in *Proc. 37th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sept. 1999, pp. 948-957.
- [24] E.G. Ström and F. Malmsten, "A maximum likelihood approach for estimating DS-CDMA multipath fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 132-140, Jan. 2000.
- [25] G. Seco, A.L. Swindlehurst, and D. Astély, "Exploiting antenna arrays for synchronization," in *Signal Processing Advances in Communications: Trends in Single- and Multi-User Systems*, vol. 2, G.B. Giannakis Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2001, ch. 10, pp. 403-430.

- [26] G. Seco, J.A. Fernández-Rubio, and A.L. Swindlehurst, "Code-timing synchronization in DS-CDMA systems using space-time diversity," *Signal Process.*, vol. 81, no. 8, pp. 1581-1602, Aug. 2001.
- [27] C. Sengupta, J.R. Cavallaro, and B. Aazhang, "On multipath channel estimation for CDMA systems using multiple sensors," *IEEE Trans. Commun.*, vol. 49, pp. 543-553, Mar. 2001.
- [28] K. Wang and H. Ge, "Joint space-time channel parameter estimation for DS-CDMA system in multipath Rayleigh fading channels," *Electron. Lett.*, vol. 37, no. 7, pp. 458-460, Mar. 2001.
- [29] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of fast-varying space-time communication channels," *IEEE Trans. Signal Processing*, vol. 51, pp. 1184-1195, May 2003.
- [30] U. Spagnolini, "Adaptive rank-one receiver for GSM/DCS systems," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1264-1271, Sept. 2002.
- [31] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation and symbol detection in unknown spatially correlated noise," *IEEE Trans. Signal Processing*, vol. 50, pp. 457-474, Mar. 2002.
- [32] L. Li, H. Li, and Y.-D. Yao, "Channel estimation and interference suppression in frequency-selective fading channels," *Electron. Lett.*, vol. 38, no. 8, pp. 383-385, Apr. 2002.
- [33] L. Li, H. Li, and Y.-D. Yao, "Channel estimation and interference suppression for space-time coded systems in frequency-selective fading channels," *Wireless Commun. Mob. Comput.*, vol. 2, no. 7, pp. 751-761, Nov. 2002.
- [34] K.W. Forsythe, D.W. Bliss, and C.M. Keller, "Multichannel adaptive beamforming and interference mitigation in multiuser CDMA systems," in *Proc. 33rd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 1999, pp. 506-510.
- [35] E.G. Larsson, J. Liu, and J. Li, "Demodulation of space-time codes in the presence of interference," *Electron. Lett.*, vol. 37, pp. 697-698, May 2001.
- [36] E.G. Larsson, P. Stoica, and J. Li, "On maximum-likelihood detection and decoding for space-time coding systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 937-944, Apr. 2002.
- [37] E.G. Larsson, P. Stoica, and J. Li, "Orthogonal space-time block codes: Maximum likelihood detection for unknown channels and unstructured interferences," *IEEE Trans. Signal Processing*, vol. 51, pp. 362-372, Feb. 2003.
- [38] K.W. Forsythe, "Performance of space-time codes over a flat-fading channel using a subspace-invariant detector," in *Proc. 36th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2002, pp. 750-755.
- [39] B.G. Agee, "The least-squares CMA: A new technique for rapid correction of constant modulus signals," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, Apr. 1986, pp. 953-956.
- [40] B.G. Agee, S.V. Schell, and W.A. Gardner, "Spectral self-coherence restoral: A new approach to blind adaptive signal extraction using antenna arrays" *Proc. IEEE*, vol. 78, pp. 753-767, Apr. 1990.
- [41] A. Ranheim, "A decoupled approach to adaptive signal separation using an antenna array," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 676-682, May 1999.
- [42] A. Dogandžić and A. Nehorai, "Finite-length MIMO adaptive equalization using canonical correlation analysis," *IEEE Trans. Signal Processing*, vol. 50, pp. 984-989, Apr. 2002.
- [43] A. Dogandžić, W. Mo, and Z. Wang, "Maximum likelihood semi-blind channel and noise estimation using the EM algorithm," in *Proc. 37th Annu. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 2003, pp. WA3.1-WA3.6.
- [44] A. Dogandžić, W. Mo, and Z. Wang, "Semi-blind channel and noise estimation using the EM algorithm," *IEEE Trans. Signal Processing*, to be published.
- [45] M. Viberg, B. Ottersten, and O. Erikmats, "A comparison of model-based detection and adaptive sidelobe cancelling for radar array processing," in *Proc. Nordic Antenna Symp.*, Eskilstuna, Sweden, May 1994.
- [46] J. Li, B. Halder, P. Stoica, and M. Viberg, "Computationally efficient angle estimation for signals with known waveforms," *IEEE Trans. Signal Processing*, vol. 43, pp. 2154-2163, Sept. 1995.
- [47] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 996-1004, Apr. 1997.
- [48] A.L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proc. IEEE*, vol. 86, pp. 421-441, Feb. 1998.
- [49] A. Dogandžić and A. Nehorai, "Estimating range, velocity, and direction with a radar array," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, AZ, Mar. 1999, pp. 2773-2776.
- [50] K. Boman and P. Stoica, "Low angle estimation: Models, methods, and bounds," *Dig. Signal Process.*, vol. 11, no. 1, pp. 35-79, Jan. 2001.
- [51] A. Dogandžić and A. Nehorai, "Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays," *IEEE Trans. Signal Processing*, vol. 48, pp. 13-25, Jan. 2000.
- [52] E.J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 115-127, Mar. 1986.
- [53] E.J. Kelly and K.M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Lincoln Lab., Mass. Inst. Technol., Lexington, MA, Tech. Rep. 848, Apr. 1989.
- [54] H. Wang and L. Cai, "On adaptive multiband signal detection with GLR algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, pp. 225-233, Mar. 1991.
- [55] K.A. Burgess and B.D. Van Veen, "Subspace-based adaptive generalized likelihood ratio detection," *IEEE Trans. Signal Processing*, vol. 44, pp. 912-927, Apr. 1996.
- [56] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Processing*, vol. 44, pp. 1469-1484, June 1996.
- [57] P. Stoica, H. Li, and J. Li, "A new derivation of the APES filter," *IEEE Signal Processing Lett.*, vol. 6, pp. 205-206, Aug. 1999.
- [58] P. Stoica and T. Sundin, "Nonparametric NMR spectroscopy," *J. Magn. Reson.*, vol. 152, no. 1, pp. 57-69, Sept. 2001.
- [59] D.R. Brillinger, "A maximum likelihood approach to frequency-wave number analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 1076-1085, Oct. 1985.
- [60] X. Li, E.G. Larsson, M. Sheplak, and J. Li, "Phase-shift-based time-delay estimators for proximity acoustic sensors," *IEEE J. Oceanic Eng.*, vol. 27, pp. 47-56, Jan. 2002.
- [61] H.S. Kim and A.O. Hero, "Comparison of GLR and invariant detectors under structured clutter covariance," *IEEE Trans. Image Processing*, vol. 10, pp. 1509-1520, Oct. 2001.
- [62] L.M. Novak and M.C. Burl, "Optimal speckle reduction in polarimetric SAR imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, pp. 293-305, Mar. 1990.
- [63] G. Wang, X.-G. Xia, and V.C. Chen, "Adaptive filtering approach to chirp estimation and inverse synthetic aperture radar imaging of maneuvering targets," *Opt. Eng.*, vol. 42, no. 1, pp. 190-199, Jan. 2003.
- [64] I.S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1760-1770, Oct. 1990.
- [65] X. Yu, I.S. Reed, and A.D. Stocker, "Comparative analysis of adaptive multispectral detectors," *IEEE Trans. Signal Processing*, vol. 41, pp. 2639-2656, Aug. 1993.

- [66] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Mag.*, vol. 19, pp. 29-43, Jan. 2002.
- [67] B.C. Juricek, D.E. Seborg, and W.E. Larimore, "Identification of multivariable, linear, dynamic models: Comparing regression and subspace techniques," *Ind. Eng. Chem. Res.*, vol. 41, no. 9, pp. 2185-2203, 2002.
- [68] C.G. Khatri, "Robustness study for a linear growth model," *J. Multivariate Analysis*, vol. 24, no. 1, pp. 66-87, 1989.
- [69] D.A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York: Springer-Verlag, 1997.
- [70] P. McCullagh and J.A. Nelder, *Generalized Linear Models*, 2nd ed. London, U.K.: Chapman & Hall, 1989.
- [71] H.L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968, pt. I.
- [72] J.D. Kalbfleisch and D.A. Sprott, "Application of likelihood methods to models involving large numbers of parameters," *J. R. Stat. Soc., Ser. B*, vol. 32, no. 2, pp. 175-208, 1970.
- [73] C.G. Khatri, "Classical statistical analysis based on a certain multivariate complex Gaussian distribution," *Ann. Math. Stat.*, vol. 36, no. 1, pp. 98-114, Feb. 1965.
- [74] S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice Hall, 1998, pt. II.
- [75] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1081-1089, Oct. 1986.
- [76] H. Krim and M. Viberg, "Two decades of array signal processing research—The parametric approach," *IEEE Signal Processing Mag.*, vol. 13, pp. 67-94, July 1996.
- [77] M.S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden-Day, 1964.
- [78] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
- [79] J. Ward, E.J. Baranoski, and R.A. Gabel, "Adaptive processing for airborne surveillance radar," in *Proc. 30th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1996, pp. 566-571.
- [80] E. Visotsky and U. Madhoo, "Noncoherent multiuser detection for CDMA systems with nonlinear modulation: A non-Bayesian approach," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1352-1367, May 2001.
- [81] J.G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.
- [82] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2000.
- [83] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, pp. 1000-1011, Feb.-Apr. 1994.
- [84] R.B. Thompson and T.A. Gray, "Use of ultrasonic models in the design and validation of new NDE techniques," *Phil. Trans. R. Soc. Lond., A*, vol. 320, no. 1554, pp. 329-340, 1986.
- [85] A. Dogandžić and N. Eua-Anant, "Flaw detection in correlated noise," in *Proc. Annu. Rev. Progress Quantitative Nondestructive Evaluation*, Green Bay, WI, July 2003.
- [86] B.A. Auld and J.C. Moulder, "Review of advances in quantitative eddy current nondestructive evaluation," *J. Nondestructive Eval.*, vol. 18, no. 1, pp. 3-36, Mar. 1999.

## Appendix

We derive the ML estimates of  $\mathbf{X}$  and  $\Sigma$  in (5) and the GLR expression in (6). To derive (5), we follow an ap-

proach similar to that of Srivastava and Khatri [5]. (For an alternative, conditional approach to solving this problem, see e.g., [2], [8], and [9].) Then, we compute the GLR in (6) by substituting the estimates of  $\Sigma$  (under  $H_0: \mathbf{X} = \mathbf{0}$ ) and  $\mathbf{X}$  and  $\Sigma$  [under  $H_1: \mathbf{X} \neq \mathbf{0}$ ; see (5)] into the likelihood ratio. Note that our concentrated likelihood function in (6) is simpler than the one that follows from [5, th. 1.10.3].

For completeness, we first state the following two lemmas from [5], which will be used in the derivation. They are also of general interest to the signal processing audience. (Special cases of both lemmas have been widely used in signal processing literature.)

### Lemma 1

Let  $\mathbf{S}$  be an  $m \times m$  positive definite matrix. Then, for  $a > 0, b > 0$ ,

$$|\Sigma|^{-b} \exp[-a \operatorname{tr}(\Sigma^{-1} \mathbf{S})] \leq |a\mathbf{S} / b|^{-b} \exp(-mb) \quad (\text{A.1})$$

for all  $m \times m$  positive definite matrices  $\Sigma$ . Equality holds if and only if  $\Sigma = a\mathbf{S} / b$ .

### Lemma 2

Let  $\mathbf{S}: m \times m$  be a positive definite matrix, and  $\mathbf{A}: m \times r$  and  $\mathbf{A}_\perp: m \times s$  be two matrices such that  $\operatorname{rank}(\mathbf{A}_\perp) = m - \operatorname{rank}(\mathbf{A})$  and  $\mathbf{A}^H \mathbf{A}_\perp = \mathbf{0}$ . Then

$$\mathbf{S}^{-1} - \mathbf{S}^{-1} \mathbf{A} (\mathbf{A}^H \mathbf{S}^{-1} \mathbf{A})^{-1} \mathbf{A}^H \mathbf{S}^{-1} = \mathbf{A}_\perp (\mathbf{A}_\perp^H \mathbf{S} \mathbf{A}_\perp)^{-1} \mathbf{A}_\perp^H \quad (\text{A.2})$$

is a positive semidefinite matrix of rank  $m - \operatorname{rank}(\mathbf{A})$ .

Under the measurement model in (1) and (2), the likelihood function is

$$L(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\eta}, \Sigma) = |\pi \Sigma|^{-N} \cdot \exp\left(-\operatorname{tr}\left\{\Sigma^{-1} [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta}) \mathbf{X} \Phi(\boldsymbol{\eta})] \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta}) \mathbf{X} \Phi(\boldsymbol{\eta})]^H\right\}\right). \quad (\text{A.3})$$

Applying Lemma 1 to (A.3), with  $b = N$  and  $a = 1$ , we obtain

$$\begin{aligned} L(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\eta}, \Sigma) &\leq L(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\eta}, (1/N) \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta}) \mathbf{X} \Phi(\boldsymbol{\eta})] \\ &\quad \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta}) \mathbf{X} \Phi(\boldsymbol{\eta})]^H) \\ &= |\pi \cdot (1/N) \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta}) \mathbf{X} \Phi(\boldsymbol{\eta})] \\ &\quad \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta}) \mathbf{X} \Phi(\boldsymbol{\eta})]^H|^{-N} \cdot \exp(-mN), \quad (\text{A.4}) \end{aligned}$$

where the equality holds if and only if

$$\Sigma = (1/N) \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta}) \mathbf{X} \Phi(\boldsymbol{\eta})] \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta}) \mathbf{X} \Phi(\boldsymbol{\eta})]^H. \quad (\text{A.5})$$

Clearly, the above expression is the ML estimate of the noise covariance  $\Sigma$  for given  $\boldsymbol{\theta}, \boldsymbol{\eta}$ , and  $\mathbf{X}$ , and (A.4) is the likelihood function, concentrated with respect to this estimate. Observe that, in the absence of signal (i.e.,  $\mathbf{X} = \mathbf{0}$ ), the ML estimate of the noise covariance is simply  $\hat{\mathbf{R}}_{yy}$  and (A.4) becomes

$$L(\mathbf{0}, \boldsymbol{\theta}, \boldsymbol{\eta}, \hat{\mathbf{R}}_{yy}) = \left| \pi \hat{\mathbf{R}}_{yy} \right|^{-N} \cdot \exp(-mN). \quad (\text{A.6})$$

Computing the ratio between the concentrated likelihood functions (A.4) and (A.6) and then raising it to the power  $1/N$  yields the following GLR test statistic:

$$\text{GLR}(X, \boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\hat{\mathbf{R}}_{yy}|}{|(1/N) \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})] \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})]^H|}, \quad (\text{A.7})$$

for testing  $H_0: \mathbf{X} = \mathbf{0}$  versus  $H_1: \mathbf{X} = \mathbf{X}$ . To be able to compute the above expression, we require that  $[\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})] \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})]^H$  is positive definite for every  $\mathbf{X}$ ,  $\boldsymbol{\theta}$ , and  $\boldsymbol{\eta}$ .

We now maximize (A.7) with respect to the regression coefficient matrix  $\mathbf{X}$ . Let

$$\hat{\mathbf{H}}_{\text{LS}} = \boldsymbol{\Upsilon}\boldsymbol{\Phi}(\boldsymbol{\eta})^H [\boldsymbol{\Phi}(\boldsymbol{\eta})\boldsymbol{\Phi}(\boldsymbol{\eta})^H]^{-1} = \hat{\mathbf{R}}_{y\phi} \hat{\mathbf{R}}_{\phi\phi}^{-1} \quad (\text{A.8})$$

denote a least-squares (LS) estimate of the coefficient matrix  $\mathbf{H} [\equiv \mathbf{A}(\boldsymbol{\theta})\mathbf{X}]$  in the MANOVA model (19). To simplify the notation, we omit the dependence of  $\hat{\mathbf{H}}_{\text{LS}}$  on  $\boldsymbol{\eta}$ . Expression (3f) can be written in terms of  $\hat{\mathbf{H}}_{\text{LS}}$  as

$$\hat{\mathbf{S}}_{y|\phi} = (1/N) \cdot [\boldsymbol{\Upsilon} - \hat{\mathbf{H}}_{\text{LS}}\boldsymbol{\Phi}(\boldsymbol{\eta})] \cdot [\boldsymbol{\Upsilon} - \hat{\mathbf{H}}_{\text{LS}}\boldsymbol{\Phi}(\boldsymbol{\eta})]^H. \quad (\text{A.9})$$

Then, the decomposition

$$\begin{aligned} & [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})] \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})]^H = [\boldsymbol{\Upsilon} - \hat{\mathbf{H}}_{\text{LS}}\boldsymbol{\Phi}(\boldsymbol{\eta})] \\ & \cdot [\boldsymbol{\Upsilon} - \hat{\mathbf{H}}_{\text{LS}}\boldsymbol{\Phi}(\boldsymbol{\eta})]^H + [\hat{\mathbf{H}}_{\text{LS}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}] \cdot \boldsymbol{\Phi}(\boldsymbol{\eta})\boldsymbol{\Phi}(\boldsymbol{\eta})^H \\ & \cdot [\hat{\mathbf{H}}_{\text{LS}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}]^H \end{aligned} \quad (\text{A.10})$$

is obtained by completing the squares and using basic properties of generalized inverses; see [5, th. 1.10.3]. As discussed before, we require that the left-hand side of the above expression is positive definite for every  $\mathbf{X}$ , implying that  $\hat{\mathbf{S}}_{y|\phi}$  must also be positive definite (consider  $\mathbf{X} = \mathbf{0}$ ). To ensure positive definiteness of  $\hat{\mathbf{S}}_{y|\phi}$  (with probability one), we impose condition (4), which follows using arguments similar to those in [78, th. 3.1.4]. Now we can write

$$\begin{aligned} & |(1/N) \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})] \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})]^H| \\ & = \left| \hat{\mathbf{S}}_{y|\phi} \cdot \left[ \mathbf{I}_m + \hat{\mathbf{S}}_{y|\phi}^{-1} \cdot [\hat{\mathbf{H}}_{\text{LS}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}] \cdot \hat{\mathbf{R}}_{\phi\phi} \right. \right. \\ & \quad \left. \left. \cdot [\hat{\mathbf{H}}_{\text{LS}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}]^H \right| \right| \\ & = \left| \hat{\mathbf{S}}_{y|\phi} \cdot \left[ \mathbf{I}_d + \hat{\mathbf{R}}_{\phi\phi} \cdot [\hat{\mathbf{H}}_{\text{LS}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}]^H \cdot \hat{\mathbf{S}}_{y|\phi}^{-1} \right. \right. \\ & \quad \left. \left. \cdot [\hat{\mathbf{H}}_{\text{LS}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}] \right] \right|, \end{aligned} \quad (\text{A.11})$$

where we used the definitions in (3) and the determinant formula  $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$ . Also

$$\begin{aligned} & [\hat{\mathbf{H}}_{\text{LS}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}]^H \cdot \hat{\mathbf{S}}_{y|\phi}^{-1} \cdot [\hat{\mathbf{H}}_{\text{LS}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}] = \boldsymbol{\Psi} + \delta\mathbf{X}^H \\ & \cdot \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta}) \cdot \delta\mathbf{X}, \end{aligned} \quad (\text{A.12})$$

where

$$\boldsymbol{\Psi} = \hat{\mathbf{H}}_{\text{LS}}^H \cdot \left[ \hat{\mathbf{S}}_{y|\phi}^{-1} - \hat{\mathbf{S}}_{y|\phi}^{-1} \hat{\mathbf{T}}_{\text{A}} \hat{\mathbf{S}}_{y|\phi}^{-1} \right] \cdot \hat{\mathbf{H}}_{\text{LS}}, \quad (\text{A.13a})$$

$$\delta\mathbf{X} = \left[ \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta}) \right]^{-1} \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \hat{\mathbf{H}}_{\text{LS}} - \mathbf{X}. \quad (\text{A.13b})$$

To derive (A.12), we have used (3g) and the identity

$$\begin{aligned} & \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \hat{\mathbf{T}}_{\text{A}} = \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta}) \left[ \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta}) \right]^{-1} \\ & \mathbf{A}(\boldsymbol{\theta})^H = \mathbf{A}(\boldsymbol{\theta})^H, \end{aligned} \quad (\text{A.14})$$

see [69, th. 14.12.11(5)]. By Lemma 2,  $\boldsymbol{\Psi}$  is positive semidefinite, and hence

$$\boldsymbol{\Gamma} = \mathbf{I}_N + (1/N) \cdot \boldsymbol{\Phi}(\boldsymbol{\eta})^H \boldsymbol{\Psi} \boldsymbol{\Phi}(\boldsymbol{\eta}) \quad (\text{A.15})$$

is positive definite. Thus, substituting (A.12) into (A.11) yields

$$\begin{aligned} & |(1/N) \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})] \cdot [\boldsymbol{\Upsilon} - \mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta})]^H| \\ & = |\hat{\mathbf{S}}_{y|\phi}| |\boldsymbol{\Gamma}| |\mathbf{I}_N + \boldsymbol{\Gamma}^{-1/2} \\ & \cdot (1/N) \boldsymbol{\Phi}(\boldsymbol{\eta})^H \delta\mathbf{X}^H \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta}) \delta\mathbf{X} \boldsymbol{\Phi}(\boldsymbol{\eta}) \cdot \boldsymbol{\Gamma}^{-1/2}|. \end{aligned} \quad (\text{A.16})$$

Clearly, (A.16) is minimized with respect to  $\delta\mathbf{X}$  (and hence  $\mathbf{X}$ ) if and only if  $\boldsymbol{\Phi}(\boldsymbol{\eta})^H \delta\mathbf{X}^H \mathbf{A}(\boldsymbol{\theta})^H \hat{\mathbf{S}}_{y|\phi}^{-1} \mathbf{A}(\boldsymbol{\theta}) \delta\mathbf{X} \boldsymbol{\Phi}(\boldsymbol{\eta}) = \mathbf{0}$ , or, equivalently,  $\mathbf{A}(\boldsymbol{\theta}) \delta\mathbf{X} \boldsymbol{\Phi}(\boldsymbol{\eta}) = \mathbf{0}$ , or

$$\mathbf{A}(\boldsymbol{\theta})\mathbf{X}\boldsymbol{\Phi}(\boldsymbol{\eta}) = \hat{\mathbf{T}}_{\text{A}} \hat{\mathbf{S}}_{y|\phi}^{-1} \boldsymbol{\Upsilon} \boldsymbol{\Pi}(\boldsymbol{\Phi}(\boldsymbol{\eta})^H). \quad (\text{A.17})$$

Therefore,

$$\begin{aligned} \text{GLR}(X, \boldsymbol{\theta}, \boldsymbol{\eta}) & \leq \text{GLR}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\hat{\mathbf{R}}_{yy}|}{|\hat{\mathbf{S}}_{y|\phi}| |\boldsymbol{\Gamma}|} \\ & = \frac{|\hat{\mathbf{R}}_{yy}|}{|\hat{\mathbf{S}}_{y|\phi} + \hat{\mathbf{H}}_{\text{LS}} \hat{\mathbf{R}}_{\phi\phi} \hat{\mathbf{H}}_{\text{LS}}^H - \hat{\mathbf{T}}_{\text{A}} \hat{\mathbf{S}}_{y|\phi}^{-1} \hat{\mathbf{H}}_{\text{LS}} \hat{\mathbf{R}}_{\phi\phi} \hat{\mathbf{H}}_{\text{LS}}^H|}, \end{aligned} \quad (\text{A.18})$$

which is equal to (6), the ML estimates of  $\mathbf{X}$  in (5a) follow from (A.17), and the ML estimate of  $\boldsymbol{\Sigma}$  in (5b) follows by substituting (A.17) into (A.5).