Generalized 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
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: \( E[Y] = A\Phi \), where \( Y \) is a data matrix whose columns are independent random vectors with common covariance matrix \( \Sigma \), \( A \), and \( \Phi \) are known matrices, and \( X \) 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 \( X \) and \( \Sigma \) under the multivariate normal model for \( Y \). 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 \( y(t) \) 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, \( y(t) \) are independent sets of pixel observations, indexed by \( t \) that generally does not correspond to time.) Consider the following model for the received snapshots:

\[
y(t) = A(\theta)X\Phi(t, \eta) + e(t), \quad t = 1, \ldots, N, \tag{1}
\]

where the signal is described by

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( m \times r \) matrix \( A(\theta) \) (\( m \geq r \))

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.

\( \Delta \) an \( r \times d \) matrix \( X \) of unknown regression coefficients.
\( \hat{R}_{yy} \) is the sample cross-correlation matrix between \( y(t) \) and \( \phi(t, \eta) \)
\( \hat{S}_{y'y} \) is the sample correlation matrix of the received data projected onto the space orthogonal to the row space of \( \Phi(\eta) \).

Here, \( S_{y'y} \) and \( \hat{R}_{y'y} \) are functions of \( \eta \) and \( \tilde{T}_d \) is a function of \( \theta \) and \( \eta \). To simplify the notation, we omit these dependencies throughout this article. Assuming that

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

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

\[
\hat{X}(\theta, \eta) = \left[ A(\theta)^H S_{y'y}^{-1} A(\theta) \right]^{-1} A(\theta)^H \left( \hat{S}_{y'y}^{-1} \hat{R}_{y'y} - \hat{R}_{y'y} \right) + \left[ I_r - A(\theta)^{-1} A(\theta) \right] \Xi_1 + A(\theta)^H \Xi_2 \left[ I_d - \Phi(\eta)^T \Phi(\eta)^{-1} \right], \tag{5a}
\]

\[
\hat{\Sigma}(\theta) = \hat{S}_{y'y} + \left( I_m - \hat{T}_d S_{y'y}^{-1} \right) \hat{R}_{y'y} \left( I_m - \hat{T}_d S_{y'y}^{-1} \right)^H, \tag{5b}
\]

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

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

\[
\text{GLR}(\theta, \eta) = \frac{|\hat{R}_{yy}|}{|\hat{R}_{yy} - \hat{T}_d \hat{S}_{y'y}^{-1} \hat{R}_{y'y} \hat{R}_{y'y}|}, \tag{6}
\]

where \( | \cdot | \) denotes the determinant. (Concentrated likelihood function is also known as the profile likelihood; see [70, ch. 7.2.4.1] for its definition and properties) Here, the ML estimates of \( X \) and \( \Sigma \) follow by substituting \( \theta \) and \( \eta \) in \( (5a) \) and \( (5b) \) with \( \theta \) and \( \eta \). In [31], we also compute closed-form Cramér-Rao bound expressions for \( \theta \) and \( \eta \).

Detection

Expression \( (6) \) is written in the form of a generalized likelihood ratio (GLR) test statistic for testing \( H_0 : X = 0 \) versus \( H_1 : X \neq 0 \) (i.e., detecting the presence of signal) for the case of known \( \theta \) and \( \eta \). (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 \( (X \text{ and } \Sigma \text{ under } H_0 \text{ and } \Sigma \text{ under } H_1) \) replaced by their ML estimates; see also the Appendix. If \( \theta \) and \( \eta \) are unknown, the GLR test compares \( \max_{\theta, \eta} \text{GLR}(\theta, \eta) = \text{GLR}(\hat{\theta}, \hat{\eta}) \) with a threshold. Since \( (6) \) is concentrated with respect to the ML estimates of the nuisance parameters \( \Sigma \) in this case, it is also the maximized relative likelihood, as defined in [72]. Under \( H_0 \) and assuming known \( \theta \) and \( \eta \), \( 1/\text{GLR}(\theta, \eta) \) is distributed as complex Wilks’ lambda; see [53] and [73]. Since Wilks’ lambda distribution does not depend on the unknown parameters \( \Sigma \) 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 \( A(\theta) \)

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

GLR for Full-Rank \( A(\theta) \)
If \( A(\theta) \) has full rank \( r \), the second term in \( (5a) \) becomes zero, and \( (6) \) simplifies to

\[
\text{GLR}(\theta, \eta) = \frac{|A(\theta)^H S_{y'y}^{-1} A(\theta)|}{|A(\theta)^H \hat{S}_{y'y}^{-1} A(\theta)|}, \tag{8}
\]

see [31, App. A]. (In sensor array processing applications, \( (8) \) can be viewed as the ratio of the Capon spectral estimate in the direction \( \theta \) using the data \( Y \), and the Capon spectral estimate in the direction \( \theta \) using the projection of the data onto the space orthogonal to the row space of \( \Phi(\eta) \). In other words, it is the overall power arriving from the direction \( \theta \), normalized by the power of the noise only, arriving from the same direction \( \theta \) [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 $A(\theta)$ and $\Phi(\eta)$

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

$$\begin{align*}
\text{GLR}(\theta, \eta) &= \left| \frac{\hat{R}_{yy} - \hat{R}^{-1}_{yy} \mathcal{W}(\theta) \hat{R}_{yy}}{|S_{\theta}(\eta)|} \right|, \\
&= (9)
\end{align*}$$

where

$$\begin{align*}
\mathcal{W}(\theta) &= \hat{R}_{yy}^{-1} - \hat{R}_{yy}^{-1} A(\theta) \left[ A(\theta)^{H} \hat{R}_{yy}^{-1} A(\theta) \right]^{-1} A(\theta)^{H} \hat{R}_{yy}^{-1} \\
&= (10a)
\end{align*}$$

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

$$\begin{align*}
\text{GLR}(\theta, \eta) &= \\
&= \left| \frac{\hat{r}_{yy}^{H} \hat{R}_{yy}^{-1} A(\theta) \left[ A(\theta)^{H} \hat{R}_{yy}^{-1} A(\theta) \right]^{-1} A(\theta)^{H} \hat{R}_{yy}^{-1} \hat{r}_{yy}}{1 + \hat{r}_{yy}^{H} \hat{R}_{yy}^{-1} \hat{r}_{yy}} \right|. \\
&= (11)
\end{align*}$$

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 $A(\theta) = A$, 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:

$$\begin{align*}
\hat{R}_{yy}^{-1/2} \hat{R}_{yy}^{-1} \hat{R}_{yy}^{H} \hat{R}_{yy}^{-1/2} &= \hat{U} \hat{\Lambda}^{2} \hat{U}^{H}, \\
&= (12)
\end{align*}$$

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

$$\begin{align*}
\text{GLR}_{\text{low rank}}(\eta) &= \prod_{i=1}^{r} \frac{1}{1 - \hat{\lambda}^{2}(k)}, \\
&= (13)
\end{align*}$$

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, $\log[\text{GLR}_{\text{low rank}}(\eta)]$ is a measure of the (estimated) mutual information between $y(t)$ and $\Phi(t, \eta)$; see [77, sec. 9.2]. Using the results of [31, App. B.], the ML estimates of $H = AX$ and $\Sigma$ follow:

$$\begin{align*}
\hat{H}_{\text{low rank}}(\eta) &= \hat{A} \hat{X} = \hat{R}_{yy}^{1/2} \hat{U}(r)^{H} \hat{R}_{yy}^{-1/2} \hat{R}_{yy} \hat{R}_{yy}^{1}, \\
&+ \Xi \left[ I_{d} - \Phi(\eta) \Phi(\eta)^{-1} \right], \\
&= (14a)
\end{align*}$$

where $\Xi$ is an arbitrary matrix (of appropriate dimensions), $\hat{\Lambda}(r) = \text{diag}(\hat{\lambda}(1), \hat{\lambda}(2), \ldots, \hat{\lambda}(r))$, and $\hat{U}(r)$ is the matrix containing the first $r$ columns of $\hat{U}$. If $\Phi(\eta)$ 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 $B$ and $W$ that minimize the sample (estimated) geometric mean-square error of $B y(t) - W \Phi(t, \eta)$, or, equivalently, maximize its inverse:

$$\begin{align*}
l(\eta, B, W) &= \frac{1}{\left| (1/N) \cdot \varepsilon(\eta) \cdot \varepsilon(\eta)^{H} \right|}, \\
&= (15)
\end{align*}$$

subject to the normalizing constraint

$$\begin{align*}
\hat{B}_{yy}^{H} B_{yy} &= I_{r}, \\
&= (16)
\end{align*}$$

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

$$\begin{align*}
\hat{B}(\eta) &= \hat{U}(r)^{H} \hat{R}_{yy}^{-1/2}, \\
&= (17a)
\end{align*}$$

$$\begin{align*}
\hat{W}(\eta) &= \hat{B}(\eta) \hat{R}_{yy} \hat{R}_{yy}^{-1}, \\
&= (17b)
\end{align*}$$
\( l(\eta, \hat{B}(\eta), \hat{W}(\eta)) = \text{GLR}_{\text{low rank}}(\eta) \) (18)

is exactly the GLR expression for reduced-rank regression in (13). (Interestingly, a stronger result holds: if the optimal \( B \) and \( W \) in (17) simultaneously minimize all the eigenvalues of the sample mean-square error matrix \( \mathcal{E}(\eta): \mathcal{E}(\eta)^H / N \) subject to (16); see [42].) Note that the elements of \( \hat{B}(\eta)y(t) \) and \( \hat{W}(\eta)\Phi(t, \eta) \) are the estimated canonical variates of \( y(t) \) and \( \Phi(t, \eta) \) (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 \( A(\theta) \equiv I_{\omega} \), and hence the coefficient matrix becomes \( H = X \). Then, the measurement model (1) simplifies to

\[ y(t) = H\Phi(t, \eta) + e(t), \quad t = 1, \ldots, N. \] (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 \( H = X \) and \( \Sigma \) simplify to [using (5) or (14)]

\[ \text{GLR}(\eta) = \frac{|\hat{R}_y|}{|\hat{S}_{y|\eta}|} \] (20a)

\[ \hat{H}(\eta) = \hat{R}_{\rho y} \hat{R}_\rho^{-1} + \Xi \left[ I_d - \Phi(\eta)\Phi(\eta)^H \right], \] (20b)

\[ \hat{\Sigma}(\eta) = \hat{S}_{\nu|\eta}, \] (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 \( \Phi(\eta) \) 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 \( \eta \) and compute \( \hat{\Omega} = \hat{\Omega}(\eta) \) using

\[ \hat{\Omega}(\eta) = \left[ \hat{H}(\eta)^H \hat{R}_{\rho y}^{-1} \hat{H}(\eta) \right]^{-1} \hat{H}(\eta)^H \hat{R}_{\rho y}^{-1} \eta \] (21)

[where \( \hat{H}(\eta) = \hat{R}_{\rho y} \hat{R}_{\rho y}^{-1} \)].

**Step 2:** Fix \( \Omega \) and find \( \eta \) that minimizes

\[ \left[ \Omega - \Phi(\eta) \right] \left[ \Omega - \Phi(\eta) \right]^H \] (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 \( r \)th range gate into a vector \( y(t) \) (of size \( m = nP \)), we search for the presence of targets in one range gate at a time. Without loss of generality, let \( r = 1 \) be under test. Then, this radar array detection problem can be formulated within the GMANOVA framework in (1) with

\[ \Phi(\eta) = [1,0,0,\ldots,0] \text{ of size } 1 \times N, \] (23)

\[ \Delta X = x, \text{ an } r \times 1 \text{ vector of target amplitudes} \]

\[ \Delta A(\theta), \text{ an } m \times r \text{ spatiotemporal steering matrix of the targets} \]

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

We wish to test \( H_{\eta} : x = 0 \) (targets absent) versus \( H_{\nu} : x \neq 0 \) (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 \( y(2), y(3), \ldots, y(N) \), 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

\[ Z = [y(2) \ y(3) \ldots y(N)], \] (24)

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

\[ NS_{\nu|\theta} = ZZ^H, \] (25a)
After substituting (25) into (6), using the determinant formula \(|I + ab^H| = 1 + b^H a\) (see, e.g., [69, cor. 18.1.3 at p. 416]) and applying the monotonic transformation 

\[ 1 - 1 / \text{GLR}(\theta) \]

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: \(Y = [y(1) \ldots y(d)]\). This problem easily fits the GMANOVA framework in (1) with

\[
Y = [Y_T, Z],
\]

\[
Z = [y(d + 1) y(d + 2) \ldots y(N)],
\]

\[
\Phi(\eta) = [I_d, 0],
\]

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

\[
\text{GLR}(\theta) = \left| I_d - \left[ I_d + Y_T^H (ZZ^H)^{-1} Y_T \right]^{-1} \cdot Y_T^H (ZZ^H)^{-1} A(\theta) \right|.
\]

\[
\left[ A(\theta)^H (ZZ^H)^{-1} A(\theta) \right]^{-1} - A(\theta)^H (ZZ^H)^{-1} Y_T^{-1}
\]

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

**Range, Velocity, and Direction Estimation**

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

\[
A(\theta) = a(\theta),
\]

\[
\Phi(\eta) = \begin{bmatrix} I_d & 0 \end{bmatrix},
\]

\[
\begin{bmatrix} I_d + Y_T^H (ZZ^H)^{-1} Y_T \end{bmatrix}^{-1}
\]

\[
\left[ A(\theta)^H (ZZ^H)^{-1} A(\theta) \right]^{-1}
\]

\[
\left[ A(\theta)^H (ZZ^H)^{-1} Y_T^{-1} \right]
\]

\[
\text{GLR}_{\text{Kelly}}(\theta) = 1 - \frac{1}{\text{GLR}(\theta)} \frac{y(1)^H (ZZ^H)^{-1} A(\theta) \left[ A(\theta)^H (ZZ^H)^{-1} A(\theta) \right]^{-1} A(\theta)^H (ZZ^H)^{-1} y(1)}{1 + y(1)^H (ZZ^H)^{-1} y(1)}
\]

\[
\text{GLR}_{\text{Wang \\& Cai}}(\theta) = 1 - \frac{1}{\text{GLR}(\theta)} \frac{a(\theta)^H (ZZ^H)^{-1} Y_T^{-1} \left[ I_d + Y_T^H (ZZ^H)^{-1} Y_T \right]^{-1} Y_T^H (ZZ^H)^{-1} a(\theta)}{a(\theta)^H (ZZ^H)^{-1} a(\theta)}
\]
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 $\hat{r}_{t_0} = 1$, $a(\theta) = a(\omega)$, and

$$
\hat{r}_{t_0} = \hat{r}_{t_0}(\omega) = (1 / N) \cdot y_{\text{DFT}}(\omega),
$$

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 $X = x$ for given $\omega$:

$$
\hat{x}(\omega) = h_{\text{APES}}(\omega)^H \cdot \hat{r}_{t_0}(\omega),
$$

where

$$
h_{\text{APES}}(\omega) = \frac{\hat{S}_n^{-1} a(\omega)}{a(\omega)^H \hat{S}_n^{-1} a(\omega)}
$$

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 $\hat{S}_{t_0} = \hat{R}_{t_0} - \hat{r}_{t_0}(\omega) \hat{r}_{t_0}(\omega)^H$ 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), $\mathbf{a}(\theta) = [1, \exp(-\beta + j\omega), \ldots, \exp((-\beta + j\omega)(m - 1))]^T$, and $\phi(t, \eta) = \exp((-\beta + j\omega) t)$, $t = 1, \ldots, N$, where $\theta = \eta = [\omega, \beta]^T$ 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

$$
\mathbf{\phi}(\eta) = \phi(\eta)^T = [\exp(j\omega), \exp(j2\omega), \ldots, \exp(jN\omega)].
$$
\[ \theta = \eta = [\omega_1, \omega_2, \ldots, \omega_r]^T \]  
(39a)

\[ A(\theta) = [a(\omega_1) a(\omega_2) \ldots a(\omega_r)] \]  
(39b)

\[ \Phi(\eta) = \Phi(\theta) = [\varphi(\omega_1) \varphi(\omega_2) \ldots \varphi(\omega_r)]^T \]  
(39c)

where \( r \) is the number of sinusoids, and estimate the unknown frequencies \( \omega_1, \omega_2, \ldots, \omega_r \) by maximizing the GLR expressions in (6)-(9). Here

\[ \hat{R}_{yy} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y} \text{DTFT}(\omega_i) \mathbf{y}^H \text{DTFT}(\omega_i) \]  

where

**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, \( H \) is known as the channel response matrix. The MANOVA estimates of \( H \) 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 \( \eta \) 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 \( \eta \) can be estimated by maximizing the GLR(\( \eta \)) 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 \( t \)

\[ \mathbf{d} \] is a \( d \times 1 \) vector of symbols transmitted by an array of \( d \) antennas and received by the receiver array at time \( t \).

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

\[ \hat{\eta}_{\text{GLR}} = \arg \max_{\eta} \frac{\hat{\mathbf{R}}_{yy}^H}{\hat{\mathbf{S}}_{yy}^H} = \arg \max_{\eta} \left| \mathbf{Y} \left[ \mathbf{I}_N - \mathbf{II} \left( \Phi(\eta)^H \right) \right] \mathbf{Y}^H \right| \]  
(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

\[ \ln \left| \frac{\hat{\mathbf{R}}_{yy}^H}{\hat{\mathbf{S}}_{yy}^H} \right| = - \ln \left| \mathbf{I}_N - \mathbf{II} \left( \Phi(\eta)^H \right) \mathbf{II} \left( \mathbf{Y}^H \right) \right| = \text{tr} \left[ \mathbf{II} \left( \Phi(\eta)^H \right) \mathbf{II} \left( \mathbf{Y}^H \right) \right] \]  
(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 \( | \mathbf{I} + \mathbf{AB}| = | \mathbf{I} + \mathbf{BA}| \) (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:

\[ - \ln \left| \mathbf{I}_N - \mathbf{II} \left( \Phi(\eta)^H \right) \mathbf{II} \left( \mathbf{Y}^H \right) \right| = \text{tr} \left[ \mathbf{II} \left( \Phi(\eta)^H \right) \mathbf{II} \left( \mathbf{Y}^H \right) \right] + \frac{1}{2} \text{tr} \left[ \mathbf{II} \left( \Phi(\eta)^H \right) \mathbf{II} \left( \mathbf{Y}^H \right) \right]^2 + \ldots \]  
(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:

\[
\Phi(\eta) = \eta^T = [s(1), s(2), \ldots, s(N)].
\]  

Then

\[
\hat{R}_{\eta\eta} = \hat{r}_{\eta\eta} = (1 / N) \cdot \sum_{t=1}^{N} y(t)s(t)^*.
\]

(44a)

\[
\hat{R}_{\eta\eta} = \hat{r}_{\eta\eta} = (1 / N) \cdot \sum_{t=1}^{N} |s(t)|^2,
\]

(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 \( \eta \) and compute

\[
\omega(t) = \hat{\omega}(t, \eta) = \frac{\hat{r}_{\eta\eta}}{\hat{r}_{\eta\eta}^H \hat{R}_{\eta\eta}^{-1} \hat{r}_{\eta\eta}^H} y(t),
\]

(45)

\[ t = 1, 2, \ldots, N, \]

where \( \hat{r}_{\eta\eta} \) and \( \hat{R}_{\eta\eta} \) are defined in (44).

**Step 2:** Fix \( \omega(t), t = 1, 2, \ldots, N \), and update \( \eta \) as follows:

\[
\hat{\eta} = [\angle \omega(1), \angle \omega(2), \ldots, \angle \omega(N)]^T.
\]

(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 \( y(t) \) with frequency-shifted (by a constant \( \alpha \) and possibly conjugated replicas of \( y(t) \):

\[
\phi(t, \eta) = y(t) \exp(j2\pi\alpha) \quad \text{or} \quad \Phi(t, \eta) = y(t)^\ast \exp(j2\pi\alpha),
\]

(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

\[
B(t) = W \Phi(t, \eta) = \hat{b}^H y(t) - \hat{w}^H \Phi(t, \eta),
\]

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

\[
\hat{b} = \hat{R}_{\eta\eta}^{-1/2} \hat{U}(1)
\]

(52a)

\[
\hat{W} = \hat{w}^H = \hat{U}(1)^H \hat{R}_{\eta\eta}^{-1/2} \hat{R}_{\eta\eta} \hat{R}_{\eta\eta}^H
\]

(52b)

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

\[
\hat{\lambda}^2(1) \cdot \hat{R}_{\eta\eta} \hat{b} = \hat{R}_{\eta\eta} \hat{R}_{\eta\eta} \hat{b},
\]

(53a)

\[
\hat{\lambda}^2(1) \cdot \hat{R}_{\eta\eta} \hat{w} = \hat{R}_{\eta\eta} \hat{R}_{\eta\eta} \hat{w},
\]

(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:

\[ y(t) = h \cdot s(t) + e(t), \quad t = 1,2,\ldots,N, \quad (54) \]

where

- \( h \) is an unknown \( m \times 1 \) channel response vector
- \( s(t) \), \( t = 1,2,\ldots,N \) 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 \( \{s_1, s_2, \ldots, s_M\} \) with equal probability; the constant-modulus assumption implies that \( |s_n| = 1, n = 1,2,\ldots,M \). (These assumptions can be relaxed, resulting in more cumbersome computations.) As discussed before, to allow unique estimation of the channel \( h \) (i.e., to resolve the phase ambiguity), we also assume that a small number \( (N_T) \) of training symbols

\[ s_{\tau}(\tau), \quad \tau = 1,2,\ldots,N_T \quad (55) \]

is embedded in the transmission scheme. Denote the corresponding snapshots received by the array as \( y_{\tau}(\tau) \), \( \tau = 1,2,\ldots,N_T \). Then, the measurement model \((54)\) holds for the training symbols as well, with \( y(t) \) and \( s(t) \) replaced by \( y_{\tau}(\tau) \) and \( s_{\tau}(\tau) \), 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 \( h \) and spatial noise covariance \( \Sigma \). We now sketch the main ideas of this approach. We first computed the joint distribution of \( y(t) \), \( s(t) \) (for \( t = 1,2,\ldots,N \)), and \( y_{\tau}(\tau) \) (for \( \tau = 1,2,\ldots,N_T \)), which is also known as the complete-data likelihood function. Using this joint distribution, we then obtained complete-data sufficient statistics for estimating \( h \) and \( \Sigma \):

\[ \hat{h} = \frac{1}{N + N_T} \left[ \sum_{t=1}^{N} y(t) s(t)^* + \sum_{\tau=1}^{N_T} y_{\tau}(\tau) s_{\tau}(\tau)^* \right] \quad (57a) \]

\[ \hat{\Sigma} = \hat{R}_{yy} - \hat{h} \hat{h}^H \quad (57b) \]

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 \([\text{in (56)}]\) given the observed data \( y(t) \), \( t = 1,\ldots,N \) and \( y_{\tau}(\tau), s_{\tau}(\tau), \tau = 1,\ldots,N_T \).

▲ The maximization (M) step is reduced to finding the expressions for the complete-data ML estimates of the unknown parameters \( h \) 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 \( h \) and \( \Sigma \) follow as a special case \((\text{for } d = 1)\) of the MANOVA equations in \((20b)\) and \((20c)\):

\[ h_{\text{MLE}} = \frac{1}{N + N_T} \left[ \sum_{t=1}^{N} y(t) s(t)^* + \sum_{\tau=1}^{N_T} y_{\tau}(\tau) s_{\tau}(\tau)^* \right] \quad (56a) \]

\[ \Sigma_{\text{MLE}} = \hat{R}_{yy} - \hat{h} \hat{h}^H \quad (56b) \]

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

**Step 1:**

\[ h_{\text{EM}}^{(k+1)} = \frac{1}{N + N_T} \left[ \sum_{t=1}^{N} y(t) s(t)^* + \sum_{\tau=1}^{N_T} y_{\tau}(\tau) s_{\tau}(\tau)^* \right] \quad \text{(58a)} \]

**Step 2:**

\[ (\Sigma_{\text{MLE}}^{(k+1)})^{-1} = \hat{R}_{yy}^{-1} + \frac{\hat{R}_{yy}^{-1} h_{\text{EM}}^{(k+1)} (h_{\text{EM}}^{(k+1)})^H \hat{R}_{yy}^{-1}}{1 - (h_{\text{EM}}^{(k+1)})^H \hat{R}_{yy}^{-1} h_{\text{EM}}^{(k+1)}} \quad (58b) \]

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]):

$$\hat{s}(t) = \arg \max_{s(t) \in \mathbb{C}^{1 \times d}} \text{Re}\{ y(t)^H R_{y}^{-1} b^{\text{ML}} \cdot s(t) \},$$  

where $b^{\text{ML}}$ is the ML estimate of $b$ 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 $\mathbf{Y}$ 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:

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

Using the definitions in (27) and $\mathbf{A}(\mathbf{0}) = I_m$, we simplify (8) to the following GLR test statistic:

$$\text{GLR} = \frac{\left| y_1 y_1^H + ZZ^H \right|}{\left| ZZ^H \right|}$$

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

$$\text{GLR} = 1 + y_r^H (\mathbf{Z} \mathbf{Z}^H )^{-1} y_r.$$  \hspace{1cm} (62)

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

**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 $\mathbf{Y}_r$, 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 $\text{tr} (\mathbf{Y}_r \mathbf{Y}_r^H )$,  \hspace{1cm} (63)

which is simply the sum of squared magnitudes of all measurements within the window $\mathbf{Y}_r$; 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-
stantly, 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.

Alekandar Dogandzic 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, radiology, 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-Editor-in-chief 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 coreipient 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


Let $S$ be an $m \times m$ positive definite matrix. Then, for $a > 0, b > 0$,
\[
|\Sigma|^{-b} \exp \left[-a \text{tr}(\Sigma^{-1} S)\right] \leq |aS|^{b} \exp(-mb) \tag{A.1}
\]
for all $m \times m$ positive definite matrices $\Sigma$. Equality holds if and only if $\Sigma = aS / b$.

**Lemma 2**

Let $S_{m \times m}$ be a positive definite matrix, and $A : m \times r$ and $A_{1} : m \times s$ be two matrices such that $\text{rank}(A_{1}) = m - \text{rank}(A)$. Then
\[
S^{-1} - S^{-1} A (A^{H} S^{-1} A)^{-1} A^{H} S^{-1} = A_{1} (A_{1}^{H} S A_{1}^{-1}) A_{1}^{H} \tag{A.2}
\]
is a positive semidefinite matrix of rank $m - \text{rank}(A)$.

Under the measurement model in (1) and (2), the likelihood function is
\[
L(X, \theta, \eta, \Sigma) = \pi_{\Sigma}^{N} \cdot \exp \left(-\text{tr}\left\{ \Sigma^{-1} [Y - A(\theta)X\Phi(\eta)] [Y - A(\theta)X\Phi(\eta)]^{H} \right\} \right). \tag{A.3}
\]

Applying Lemma 1 to (A.3), with $b = N$ and $a = 1$, we obtain
\[
\begin{align*}
L(X, \theta, \eta, \Sigma) & \leq L(X, \theta, \eta,(1 / N) \cdot [Y - A(\theta)X\Phi(\eta)]
\[Y - A(\theta)X\Phi(\eta)]^{H} \\
& = |\pi \cdot (1 / N) \cdot [Y - A(\theta)X\Phi(\eta)]
[Y - A(\theta)X\Phi(\eta)]^{H}|^{-N} \cdot \exp(-mN), \tag{A.4}
\end{align*}
\]
where the equality holds if and only if
\[
\Sigma = (1 / N) \cdot [Y - A(\theta)X\Phi(\eta)] [Y - A(\theta)X\Phi(\eta)]^{H}. \tag{A.5}
\]
Clearly, the above expression is the ML estimate of the noise covariance $\Sigma$ for given $\theta, \eta$, and $X$, and (A.4) is the likelihood function, concentrated with respect to this estimate. Observe that, in the absence of signal (i.e., $X = 0$), the ML estimate of the noise covariance is simply $\hat{R}_{\eta}$ and (A.4) becomes
\[ L(\theta, \varphi, \Psi) \mid R \equiv \frac{\hat{R}_m}{N} \cdot \exp(-mN). \quad (A.6) \]

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:

\[ \text{GLR}(X, \varphi) \]

\[ = \left( \frac{1}{N} \right) \left| \begin{bmatrix} Y - A(\theta)X \Phi(\varphi) \end{bmatrix} \cdot \begin{bmatrix} Y - A(\theta)X \Phi(\varphi) \end{bmatrix} \right|^H \cdot \hat{R}_{\varphi} \quad (A.7) \]

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

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

\[ \hat{H}_{LS} = Y \Phi(\varphi)^H \left[ \Phi(\varphi) \Phi(\varphi)^H \right] = \hat{\Phi}_{\varphi}^H \]

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

\[ \hat{S}_{j0} = \left( \frac{1}{N} \right) \left[ Y - \hat{H}_{LS} \Phi(\varphi) \right] \left[ Y - \hat{H}_{LS} \Phi(\varphi) \right]^H \cdot \hat{R}_{\varphi} \quad (A.9) \]

Then, the decomposition

\[ \begin{bmatrix} Y - A(\theta)X \Phi(\varphi) \end{bmatrix} \cdot \begin{bmatrix} Y - A(\theta)X \Phi(\varphi) \end{bmatrix} \]

\[ = \begin{bmatrix} Y - \hat{H}_{LS} \Phi(\varphi) \end{bmatrix} \cdot \begin{bmatrix} Y - \hat{H}_{LS} \Phi(\varphi) \end{bmatrix}^H + \begin{bmatrix} \hat{H}_{LS} - A(\theta)X \end{bmatrix} \cdot \Phi(\varphi) \Phi(\varphi)^H \cdot \hat{H}_{LS}^H \quad (A.10) \]

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 \( X \), implying that \( \hat{S}_{j0} \) must also be positive definite (consider \( X = 0 \)). To ensure positive definiteness of \( \hat{S}_{j0} \) (with probability one), we impose condition (4), which follows using arguments similar to those in [78, th. 3.1.4]. Now we can write

\[ \left( \frac{1}{N} \right) \left[ Y - A(\theta)X \Phi(\varphi) \right] \cdot \begin{bmatrix} \hat{H}_{LS} - A(\theta)X \end{bmatrix}^H \cdot \hat{R}_{\varphi} \]

\[ = \hat{S}_{j0} \left( \begin{bmatrix} \hat{H}_{LS} - A(\theta)X \end{bmatrix} + \begin{bmatrix} \hat{R}_{\varphi} \end{bmatrix} \right)^H \cdot \hat{R}_{\varphi} \]

\[ = \hat{S}_{j0} \left( \begin{bmatrix} \hat{H}_{LS} - A(\theta)X \end{bmatrix} + \begin{bmatrix} \hat{R}_{\varphi} \end{bmatrix} \right)^H \cdot \hat{S}_{j0} \quad (A.11) \]

where we used the definitions in (3) and the determinant formula \(|I + AB| = |I + BA|\). Also

\[ \begin{bmatrix} \hat{H}_{LS} - A(\theta)X \end{bmatrix} \cdot \hat{S}_{j0}^{-1} \cdot \begin{bmatrix} \hat{H}_{LS} - A(\theta)X \end{bmatrix} = \Psi + \delta X^H \cdot A(\theta)^H \cdot \hat{S}_{j0}^{-1} \cdot A(\theta) \cdot \delta X \]

\[ (A.12) \]

where

\[ \Psi = \hat{H}_{LS}^H \cdot \left( \hat{S}_{j0}^{-1} - \hat{S}_{j0}^{-1} \cdot \hat{T}_A \cdot \hat{S}_{j0}^{-1} \right) \cdot \hat{H}_{LS} - X. \]

(A.13a)

\[ \delta X = \left[ A(\theta)^H \cdot \hat{S}_{j0}^{-1} \cdot A(\theta) \right] \cdot A(\theta)^H \cdot \hat{S}_{j0}^{-1} \cdot \hat{H}_{LS} - X. \]

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

\[ \left[ A(\theta)^H \cdot \hat{S}_{j0}^{-1} \cdot A(\theta) \right] \cdot A(\theta)^H = A(\theta)^H, \]

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

\[ \Gamma = I_N + (1/N) \cdot \Phi(\varphi)^H \cdot \Psi \cdot \Phi(\varphi) \]

(A.15)

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

\[ \left( \frac{1}{N} \right) \left| \begin{bmatrix} Y - A(\theta)X \Phi(\varphi) \end{bmatrix} \cdot \begin{bmatrix} Y - A(\theta)X \Phi(\varphi) \end{bmatrix} \right|^H \]

\[ = \hat{S}_{j0} \left( \begin{bmatrix} \hat{H}_{LS} - A(\theta)X \end{bmatrix} + \begin{bmatrix} \hat{R}_{\varphi} \end{bmatrix} \right)^H \cdot \hat{R}_{\varphi} \]

\[ = \hat{S}_{j0} \left( \begin{bmatrix} \hat{H}_{LS} - A(\theta)X \end{bmatrix} + \begin{bmatrix} \hat{R}_{\varphi} \end{bmatrix} \right)^H \cdot \hat{S}_{j0} \quad (A.16) \]

Clearly, (A.16) is minimized with respect to \( \delta X \) (and hence \( X \)) if and only if \( \Phi(\varphi)^H \cdot A(\theta)^H \cdot \hat{S}_{j0}^{-1} \cdot A(\theta) \cdot \delta X \Phi(\varphi) = 0 \), or, equivalently, \( A(\theta)^H \cdot \delta X \Phi(\varphi) = 0 \), or

\[ A(\theta)^H \cdot \Phi(\varphi) = 0. \]

(A.17)

Therefore,

\[ \text{GLR}(X, \theta, \varphi) \leq \text{GLR}(\theta, \varphi) = \left| \frac{\hat{R}_m}{N} \right| \cdot \left| \frac{\hat{S}_{j0}}{N} \right| \]

\[ = \frac{\left| \frac{R_{j0}}{N} \right|}{\left| \frac{S_{j0} + \hat{H}_{LS} \cdot \hat{R}_{\varphi} \cdot \hat{H}_{LS}^H - \hat{T}_A \cdot \hat{S}_{j0}^{-1} \cdot \hat{H}_{LS} \cdot \hat{R}_{\varphi} \cdot \hat{H}_{LS}^H}{N} \right|} \]

(A.18)

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