Geodesic Paths for Time-Dependent Covariance Matrices in a Riemannian Manifold

Avishai Ben-David and Justin Marks

***Abstract*—Time-dependent covariance matrices are important in remote sensing and hyperspectral detection theory. The difficulty is that $\mathbf{C}(t)$ is usually available only at two endpoints $\mathbf{C}(t_0) = A$ and $\mathbf{C}(t_1) = B$ where $\mathbf{C}(t_0 < t < t_1)$ is needed. We present the Riemannian manifold of positive definite symmetric matrices as a framework for predicting a geodesic time-dependent covariance matrix. The geodesic path $\mathbf{A} \rightarrow \mathbf{B}$ is the shortest and most efficient path (minimum energy). Although there is no guarantee that data will necessarily follow a geodesic path, the predicted geodesic $\mathbf{C}(t)$ is of value as a concept. The path for the inverse covariance is also geodesic and is easily computed. We present an interpretation of $\mathbf{C}(t)$ with coloring and whitening operators to be a sum of scaled, stretched, contracted, and rotated ellipses.**

***Index Terms*—Background characterization, detection algorithms, geodesic path, hyperspectral remote sensing, matched filters, Riemannian manifold, signal processing algorithms, statistical modeling, time-dependent covariance matrices.**

I. Introduction

THE covariance matrix is a key concept in hyperspectral discrimination algorithms [1]–[4]. Let $\boldsymbol{x}$ be a random variable with a time-dependent covariance matrix $\mathbf{C}(t)$; hence, $\boldsymbol{x} = \boldsymbol{x}(t)$. Given a test vector (data) $\boldsymbol{x}(t)$, we pose the question "should $\boldsymbol{x}(t)$ be labeled as a signal of interest? Does it belong to H_0 or to H_1?" where with the binary Neyman–Pearson hypothesis testing [4] regarding the data, H_0 is the null hypothesis that the data do not contain the signal of interest (e.g., a target), and H_1 is the alternative hypothesis that the data do contain the signal of interest in addition to noise and interferences. The noise and interferences (the "background") during the H_1 scenario are characterized by the covariance matrix $\mathbf{C}|H_0$. To answer this question, we need to have $\mathbf{C}(t)|H_0$ at all times—but, when the test vector $\boldsymbol{x}(t)$ is measured the condition may be H_1 (target plus background) and cannot be used for $\boldsymbol{C}(t)|H_0$. At best, we have access only to $\mathbf{C}(t)|H_0$ prior and post the H_1 scenario. Given the boundary conditions $\mathbf{C}(t_0) = \mathbf{A}$ and $\mathbf{C}(t_1) = \mathbf{B}$, the objective is to find the most "reasonable" time-dependent covariance $\mathbf{C}(t_0 \leq t \leq t_1)$, where for convenience we set $t_0 = 0$ and $t_1 = 1$. This scenario occurs in many field tests where a target (chemical cloud) is disseminated at $t_0 < t < t_1$ for which $\mathbf{C}(t)|H_0$ is inaccessible. Performing the interpolation between two matrices ($\mathbf{A}$ and $\mathbf{B}$) is a basic task in temporal monitoring applications (e.g., anomalous change detection). The key question is how to define "reasonable" when we compute (predict) $\mathbf{C}(t)$ by interpolation between $\mathbf{A}$ and $\mathbf{B}$. Interpolation implies missing data points; hence, the type of interpolation used to fill in the missing data can only be judged by our prejudice (or prior knowledge) of what is "reasonable." Our approach is to perform the interpolation between the endpoints $\mathbf{A}$ and $\mathbf{B}$ in the Riemannian manifold of covariance matrices (symmetric matrices with positive eigenvalues) [5]. It can be shown (see proposition 5.3.2 in [6]) that this path for $\mathbf{C}(t)$ from $\mathbf{A} \rightarrow \mathbf{B}$, where $0 \leq t \leq 1$ is the shortest (this path is known as a geodesic path) and has the least energy of all possible paths.

The premise of this work is that nature is an efficient machine and thus the geodesic path of $\mathbf{C}(\mathrm{t})$ that is the most efficient (shortest and with minimal energy) path is chosen. The argument of "efficiency" is used in the same way that the chosen path of light particles (photons) from $\mathbf{A} \rightarrow \mathbf{B}$ is such that it minimizes the time (Fermat's principle). In this letter, we do not address how a real atmosphere behaves and what atmospheric conditions, topography, etc., may impose on $\mathbf{C}(\mathrm{t})$, where $\mathbf{C}(\mathrm{t})$ may vary due to particulates and gases in the atmosphere. We view the geodesic path in Riemannian space with an analogy to airplane flight path between two cities. The "best" flight path is the shortest path achieved in a path along great circles, where the metric for best is shortest path length, hence shortest time, efficient, and with least energy (fuel) consumption. That said, it does not mean that "best" flight paths always follow great circles (e.g., a change of a path due to local atmospheric weather disturbances). The Riemannian manifold is a space with a structure that provides the "best" (shortest with least energy) path for the trajectory of time-dependent covariance matrix $\mathbf{C}(\mathrm{t})$ from $\mathbf{A} \rightarrow \mathbf{B}$. The Riemannian manifold is a good starting point (a reasonable framework) for thinking about time-dependent covariance matrices in remote sensing.

The discussion of geodesic paths in Riemannian manifolds has been so far mainly limited to geometric analysis within the mathematical community. In this letter, we want to introduce the relevance of the geodesic path to the hyperspectral remote sensing community. Although we address the temporal change in $\mathbf{C}(t)$ for a single-pixel hyperspectral sensor, the Riemannian framework for $\mathbf{C}(t)$ is valid for imaging sensors where $\mathbf{A}$ and $\mathbf{B}$ are estimated from 2-D pixels at t_0 and t_1, and $\mathbf{C}(t)$ is the geodesic interpolation between the two matrices. We follow the seminal work by Pennec *et al.* [5] and illuminate the mechanics of the geodesic paths $\mathbf{C}(t)$ and $\mathbf{C}^{-1}(t)$ in the Riemannian manifold and contrast it with the "naive" linear

Manuscript received October 1, 2013; revised November 13, 2013; accepted December 23, 2013. Date of publication January 28, 2014; date of current version March 14, 2014.

A. Ben-David is with RDECOM, Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD 21010, USA (e-mail: avishai.bendavid@us.army.mil).

J. Marks was with the Air Force Institute of Technology, Wright-Patterson AFB, OH 45433 USA. He is currently with Bowdoin College, Brunswick, ME 04011 USA (e-mail: jmarks@bowdoin.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2013.2296833

path $\mathbf{A} \to \mathbf{B}$ in the Euclidean space, given by $\mathbf{C}_{\rm lin}(t) = \mathbf{A} + t(\mathbf{B} - \mathbf{A})$. "Naive" does not necessarily imply "simpler." We show that the geodesic inverse path $\mathbf{C}_{\rm geo}^{-1}(t)$ is actually easier to compute than $\mathbf{C}_{\rm lin}^{-1}(t)$. Our letter gives a reasonable way to interpolate the missing $\mathbf{C}(t)$ data along the path $\mathbf{A} \to \mathbf{B}$. We give an interpretation for the geodesic $\mathbf{C}(t)$ with whitening and coloring operations [7].

II. Time-Dependent Covariance Matrix

In the Riemannian manifold, the gradient (tangent) matrix $\mathbf{G}$ and the time-dependent geodesic covariance matrix [5] are computed for $0 \le t \le 1$ with exponential logarithm mapping given by (1), shown at the bottom of the page, where $\mathbf{G} = \mathbf{A}^{0.5}\log(\mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5})\mathbf{A}^{0.5}$; $\mathbf{M} = \log(\mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5})$, and $\exp(\cdot)$ and $\log(\cdot)$ are the matrix exponential and logarithm functions (given in MATLAB by expm and logm functions). Note that $\mathbf{C}_{\rm geo}^{-1}(t)$ is given without an explicit inverse operation. The local geodesic tangent $\mathbf{G}(t)$ along the geodesic path $\mathbf{A} \to \mathbf{B}$ is given by $\mathbf{G}_{\rm geo}(t) = (1-t)^{-1}\mathbf{C}_{\rm geo}^{-0.5}\log(\mathbf{C}_{\rm geo}^{-0.5}\mathbf{B}\mathbf{C}_{\rm geo}^{-0.5})\mathbf{C}_{\rm geo}^{-0.5}$ where the local gradient is a function of location and is scaled by $(1-t)^{-1}$ for the remainder of the distance $\mathbf{C}_{\rm geo}(t) \to \mathbf{B}$. One can view the tangent as "velocity" (i.e., speed and direction) toward the endpoint $(\mathbf{B})$ of the path. The tangent plays an important role in charting a geodesic (shortest) path because a length of a parametric curve can be given via integration of the time-dependent (local) tangent—and finding the geodesic path involves constraining the tangents along the path. In numerical implementation, we suggest to force symmetry (to counteract lack of symmetry due to rounding errors) for all the matrix terms by computing $\mathbf{X} \to (\mathbf{X} + \mathbf{X}^{\rm T})/2$ where superscript T is the matrix transpose.

Eigenvalues and the Eigenvectors: The exponential matrix $\exp(t\mathbf{Q}) = \mathbf{U}\mathbf{\Lambda}(t)\mathbf{U}^{\rm T}$ using singular value decomposition (SVD) where $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\rm T}$ is given with only time-dependent eigenvalues $\mathbf{\Lambda}(t) = \exp(t \cdot {\rm eig}(\mathbf{Q}))$. Nevertheless, the eigenvalues and eigenvectors of $\mathbf{C}_{\rm geo}(t) = \mathbf{A}^{0.5}\exp(t\mathbf{M})\mathbf{A}^{0.5}$ are time dependent, and thus $\mathbf{C}(t) = \mathbf{U}(t)\mathbf{\Lambda}(t)\mathbf{U}^{\rm T}(t)$.

A few properties of geodesic $\mathbf{C}(t)$ are of particular interest.

Reversibility: The geodesic path $\mathbf{B} \to \mathbf{A}$ is the same as the geodesic path $\mathbf{A} \to \mathbf{B}$; $\mathbf{C}_{\rm geo}(1-t, \mathbf{B} \to \mathbf{A}) = \mathbf{C}_{\rm geo}(t, \mathbf{A} \to \mathbf{B})$. The path for $\mathbf{A}^{-1} \to \mathbf{B}^{-1}$ is also a geodesic path and can be derived by noting that $[\exp(t\mathbf{X})]^{-1} = \exp(-t\mathbf{X})$. The geodesic inverse path is also reversible: $\mathbf{C}_{\rm geo}^{-1}(t, \mathbf{A} \to \mathbf{B}) = \mathbf{C}_{\rm geo}(1-t, \mathbf{B}^{-1} \to \mathbf{A}^{-1})$. The gradient matrix $\mathbf{G}$ is guaranteed to be real and symmetric, but not necessarily positive definite (i.e., the eigenvalues are real but can be negative).

Distance: The distance [5], [8] between two matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ is given by $d(\mathbf{C}_1, \mathbf{C}_2) = d(\mathbf{I}, \mathbf{C}_1^{-0.5}\mathbf{C}_2\mathbf{C}_1^{-0.5}) = [\sum \log(s_{\rm i})^2]^{0.5}$ where $s_{\rm i}$ is the ith eigenvalue of $\mathbf{C}_1^{-0.5}\mathbf{C}_2\mathbf{C}_1^{-0.5}$; $d(\mathbf{C}_1, \mathbf{C}_2) = d(\mathbf{C}_2, \mathbf{C}_1)$, and $d(\mathbf{C}_1, \mathbf{C}_2) \le d(\mathbf{I}, \mathbf{C}_1) + d(\mathbf{I}, \mathbf{C}_2)$.

Using series expansion for matrix exponential and its inverse, we can rewrite (1) by

$$\left\{\begin{array}{l}\mathbf{C}_{\rm geo}(t, \mathbf{A} \to \mathbf{B}) = \mathbf{A} + \sum\limits_{n=1}^{\infty} \frac{t^n}{n!}\mathbf{A}^{0.5}\mathbf{M}^n\mathbf{A}^{0.5} \\ \mathbf{C}_{\rm geo}^{-1}(t, \mathbf{A} \to \mathbf{B}) = \mathbf{A}^{-1} + \sum\limits_{n=1}^{\infty} (-1)^n \frac{t^n}{n!}\mathbf{A}^{-0.5}\mathbf{M}^n\mathbf{A}^{-0.5}\end{array}\right\}. \tag{2}$$

Due to the fact that $0 \le t \le 1$ and the reciprocity $\mathbf{C}_{\rm geo}(t, \mathbf{A} \to \mathbf{B}) = \mathbf{C}_{\rm geo}(1-t, \mathbf{B} \to \mathbf{A})$, the series needs to be computed only for $0 \le t \le 0.5$, and usually only a few terms.

We use the series expansion representation to interpret the geodesic time-dependent $\mathbf{C}(t)$ and its inverse along the path. Our interpretation relies on the concepts of rotations and scaling which we explain with whitening and coloring operations on normally distributed multivariate vectors. Let $\boldsymbol{x} \sim N(0, \mathbf{C})$ be a p-by-1 zero mean vector with normal probability density function (pdf) where $\mathbf{C} = {\rm E}(\boldsymbol{x}\boldsymbol{x}^{\rm T})$ is the p-by-p population covariance matrix, ${\rm E}[\cdot]$ denotes expectation process. The sampled covariance matrix $\hat{\mathbf{C}} \sim W(m^{-1}\mathbf{C}, m)$ is distributed as a central Wishart distribution [9], [10] with m degrees of freedom, where $m > p$ and $m+1$ is the number of vectors $\boldsymbol{x}$ that are used in the expectation. The variances of the sampled covariance $\hat{\mathbf{C}}$ and the inverse covariance $\hat{\mathbf{C}}^{-1}$[11], [12] are of importance when field data are to be compared with theoretical predictions of population covariance $\mathbf{C}$. The sampled variance of $\hat{\mathbf{C}}$ is given by ${\rm Var}(\hat{\mathbf{C}}) = m^{-1}(\mathbf{C}\ {\rm tr}(\mathbf{C}) + \mathbf{C}^2)$, where ${\rm tr}(\cdot)$ is matrix trace. The ratio m/p is important in the performance of detection algorithms [13]. We compute $d(\mathbf{C}_{\rm geo}, \hat{\mathbf{C}} \pm [{\rm Var}(\hat{\mathbf{C}})]^{0.5})$ in Section V for field data $\hat{\mathbf{C}}$.

III. Interpretation of $\mathbf{C}(t)$ as a Sum of Rotated and Scaled Ellipses

A covariance matrix (a quadratic form) represents a p-dimensional ellipsoid whose principal directions and axis lengths are the eigenvectors and eigenvalues, respectively. The population covariance is an ellipsoid with a specific orientation and aspect ratio. Let $\mathbf{F}$ be a p-by-p positive definite matrix with SVD decomposition $\mathbf{F} = \mathbf{U}\boldsymbol{S}\mathbf{U}^{-1}$. The operators $\mathbf{F}^{\pm 0.5}$ are rotation (by eigenvectors $\mathbf{U}$) and scaling operators (by eigenvalues $\boldsymbol{S}^{\pm 0.5}$) where $\mathbf{F}^{\pm 0.5} = \mathbf{U}\boldsymbol{S}^{\pm 0.5}\mathbf{U}^{-1}$, and $\mathbf{F}^{\pm 0.5}\boldsymbol{x}$ reshapes the elliptical covariance of $\boldsymbol{x}$ ($\boldsymbol{x}$ does not have to be normally distributed). When $\boldsymbol{x}$ is normal, the density of the contours of $\boldsymbol{x}$ is elliptical, and it is convenient to think of the operation $\mathbf{F}^{\pm 0.5}\boldsymbol{x}$ as whitening and coloring. The *pdf* of $\boldsymbol{x} \sim N(0, \mathbf{C})$ has elliptical density (hyper ellipsoid in p-D). Each of the p axes of the ellipsoid can be thought of as a color and the length of the axes (the eigenvalue of $\mathbf{C}$) is the intensity of the color. The color of x is captured in the structure of its covariance. White light is composed of all colors with equal intensity. The transformation $\boldsymbol{x} \to \mathbf{F}^{-0.5}\boldsymbol{x}$ produces

$$\left\{\begin{array}{l}\mathbf{C}_{\rm geo}(t, \mathbf{A} \to \mathbf{B}) = \mathbf{A}^{0.5}\exp(t\mathbf{A}^{-0.5}\mathbf{G}\mathbf{A}^{-0.5})\mathbf{A}^{0.5} = \mathbf{A}^{0.5}\exp(t\mathbf{M})\mathbf{A}^{0.5} \\ \mathbf{C}_{\rm geo}^{-1}(t) = [\mathbf{C}_{\rm geo}(t, \mathbf{A} \to \mathbf{B})]^{-1} = \mathbf{A}^{-0.5}\exp(-t\mathbf{M})\mathbf{A}^{-0.5} \\ \quad = \mathbf{C}_{\rm geo}(t, \mathbf{A}^{-1} \to \mathbf{B}^{-1}) = \mathbf{A}^{-0.5}\exp(-t\mathbf{A}^{-0.5}\mathbf{G}\mathbf{A}^{-0.5})\mathbf{A}^{-0.5}\end{array}\right\} \tag{1}$$

$\mathbf{F}^{-0.5}\boldsymbol{x} \sim N(0, \mathbf{F}^{-0.5}\mathbf{C}\mathbf{F}^{-0.5})$. We think of $\mathbf{F}^{-0.5}\boldsymbol{x}$ as removing color $\mathbf{F}$ from $\mathbf{C}$. When $\mathbf{F} = \mathbf{C}$, $\mathbf{F}^{-0.5}\boldsymbol{x} \sim N(0, \mathbf{I})$ is a whitening operation that produces a whitened vector with a diagonal covariance matrix of equal values (hence, the analogy to a white light) where all the "color" of $\boldsymbol{x}$ was removed. The whitened vector is also termed a "sphered" vector (i.e., unit variance in all directions). A coloring operation is an operation that imposes (adds) color on $\boldsymbol{x}$. When $\boldsymbol{x} \sim N(0, \mathbf{C})$, the operation $\mathbf{F}^{0.5}\boldsymbol{x} \sim N(0, \mathbf{F}^{0.5}\mathbf{C}\mathbf{F}^{0.5})$ is an attempt to put color $\mathbf{F}$ on the original color $\mathbf{C}$. When $\mathbf{C} = \mathbf{I}$, $\mathbf{F}^{0.5}\boldsymbol{x}$ results with only color $\mathbf{F}$.

Equation (2) presents the geodesic covariance $\mathbf{C}(t)$ as a sum of rotated and scaled ellipses. It is convenient to show this by tracking a normal p-by-1 vector, $\boldsymbol{x}(t)$, where at time $t = 0$ $\boldsymbol{x}(0) \sim N(0, \mathbf{A})$, at $t = 1$ $\boldsymbol{x}(1) \sim N(0, \mathbf{B})$, and at $0 < t < 1$ $\boldsymbol{x}(t) \sim N(0, \mathbf{C}(t))$ where $\mathbf{C}(t)$ is given by (2). $\boldsymbol{x} \sim N(0, \mathbf{M})$ is a hyper-ellipsoid whose axes are the eigenvectors, and the magnitude of the axes (the eigenvalues) determines the eccentricity of an ellipsoid *must* be positive. The matrix $\mathbf{M}$ in (1), (2) is positive definite if the eigenvalues of $\mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5}$ are greater than one. For the computation of the geodesic covariance, it is unimportant whether $\mathbf{M}$ is positive definite. However, for the geometric interpretation that we provide using (2), hyper-ellipses only make sense when $\mathbf{M}$ is positive definite. Therefore, to aid the geometric interpretation, we define scaling constants k_a and k_b, $\mathbf{A} \to k_a\mathbf{A}$ and $\mathbf{B} \to k_b\mathbf{B}$, to force $\mathbf{M}$ to be positive definite. The eigenvalues of $\mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5}$ with this scaling are >1 when $k_b k_a^{-1} > 1/\min(\mathrm{eig}(\boldsymbol{A}^{-0.5}\boldsymbol{B}\boldsymbol{A}^{-0.5}))$. The matrix $\mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5}$ is the attempt to remove color from $\boldsymbol{x}(t = 1)$ with the covariance of $\boldsymbol{x}(t = 0)$ where $\mathbf{A}^{-0.5}\boldsymbol{x}(t = 1) \sim N(0, \mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5})$. The matrices $\mathbf{M}$ and the matrix $\mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5}$ share the same eigenvectors, but with different eigenvalues; $\mathrm{eig}(\mathrm{M}) = \log(\mathrm{eig}(\mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5}))$. In 2-D, $\boldsymbol{x} \sim N(0, \mathbf{M})$ is a shrunken ellipse (with the same orientation) as the ellipse represented by $N(0, \mathbf{A}^{-0.5}\mathbf{B}\mathbf{A}^{-0.5})$. The vector $\boldsymbol{x} \sim N(0, \mathbf{M}^{\mathrm{n}})$ is also an ellipse at the same orientation as of $\boldsymbol{x} \sim N(0, \mathbf{M})$, but with eigenvalues $\mathrm{eig}(\mathbf{M}^{\mathrm{n}}) = \mathrm{eig}(\mathbf{M})^n$; hence, the eccentricity of the ellipse is now accentuated in a nonlinear manner. In (2), $\mathbf{A}^{0.5}\mathbf{M}^n\mathbf{A}^{0.5}$ is coloring of $x \sim N(0, \mathbf{M}^n)$ with the covariance at $t = 0$, where $\mathbf{A}^{0.5}N(0, \mathbf{M}^{\mathrm{n}}) \sim \mathrm{N}(0, \mathbf{A}^{0.5}\mathbf{M}^n\mathbf{A}^{0.5})$ is rotated due to the coloring operation. The scalar $(n!)^{-1}t^n$ in (2) is a simple scaling that expands (or shrinks) the ellipse in equal amounts in all directions. Thus, (2) is a sum of rotated ellipses (rotated and colored by $\mathbf{A}^{0.5}$ at $t = 0$ where ellipses change their eccentricities (by $\mathbf{M}^n$), are scaled (by $(n!)^{-1}t^n$), and the matrix $\mathbf{M}$ is related to (attempted) whitening of the boundary condition $(\mathbf{B})$ at $t = 1$ by the boundary condition $(\mathbf{A})$ at $t = 0$. The interpretation of the inverse covariance $\mathbf{C}^{-1}(t, \mathbf{A} \to \mathbf{B})$ is similar, but $\mathbf{M}^n$ in (2) is whitened (sphered) by the boundary condition $\mathbf{A}^{-0.5}$ at $t = 0$ and we sum and subtract a series of whitened and scaled ellipses due to altering signs of $(-1)^n$.

Distance: The distance $d(\mathbf{C}_1, \mathbf{C}_2)$ between two matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ can be interpreted as a measure of deviation (distance) from perfect whitening operation. We measure the quality of whitening of a vector $\boldsymbol{x} \sim N(0, \mathbf{C}_2)$ by $\mathbf{C}_1^{-0.5}$, $\mathbf{C}_1^{-0.5}\boldsymbol{x} \sim N(0, \mathbf{C}_1^{-0.5}\mathbf{C}_2\mathbf{C}_1^{-0.5})$ with respect to a perfect whitening given by $\boldsymbol{x} \sim N(0, \mathbf{I})$. The mapping between k-variance units to a distance is given by $d(\mathbf{I}, (1 + k)\mathbf{I}) = p^{0.5}\log(1 + k)$ for $\mathbf{C}_1 = \mathbf{I}$ and $\mathbf{C}_2 = (1 + k)\mathbf{I}$. Let us assume that an identity covariance matrix $\mathbf{C} = \mathbf{I}$ is sampled and is distributed as a central Wishart $\hat{\mathbf{I}} \sim W(m^{-1}\mathbf{I}, m)$ where the sampled variance is $\mathrm{Var}(\hat{\mathbf{I}}) = m^{-1}(1 + \mathrm{p})\mathbf{I}$. The effect of sampling on the distance $d(\mathbf{I}, \hat{\mathbf{I}})$ due to the number (m) of sampled $\hat{\boldsymbol{x}}$ vectors used in the estimate $\hat{\mathbf{I}} = E(\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T)$ may be approximated by computing $d(\mathbf{I}, \hat{\mathbf{I}}) = d(\mathbf{I}, \mathbf{I} \pm \mathrm{Var}(\hat{\mathbf{I}})) = p^{0.5}|\log(1 \pm m^{-1}(1 + p))|$, where $d(\mathbf{I}, \hat{\mathbf{I}}) \to 0$ and $\mathrm{Var}(\hat{\mathbf{I}}) \to 0$ as $m \to \infty$.

IV. Implications for Matched Filter Detection

The impact of using a linearly time-dependent covariance $\mathbf{C}_{\mathrm{lin}}(t) = \mathbf{A} + (\mathbf{B} - \mathbf{A})t$ instead of using geodesic time-dependent covariance $\mathbf{C}_{\mathrm{geo}}(t)$ in (1) on the standard matched filter detection algorithm [3], [4] is easy to see when we assume that the true time-dependent covariance is the geodesic covariance. The time-dependent detection score (S) for a matched filter (MF) detector is given by $S(t) = \boldsymbol{x}^{\mathrm{T}}(t)\mathbf{C}^{-1}(t)\boldsymbol{\delta}/[\boldsymbol{\delta}^{\mathrm{T}}\mathbf{C}^{-1}(t)\boldsymbol{\delta}]$ where $\boldsymbol{\delta}$ is the target vector, and the zero mean test vector $\boldsymbol{x}$ is distributed as $\boldsymbol{x}(t)|H_0 \sim N(0, \mathbf{C}_{\mathrm{geo}}(t))$ under the H_0 (target absent) scenario and as $\boldsymbol{x}(t)|H_1 \sim N(\boldsymbol{\delta}, \mathbf{C}_{\mathrm{geo}}(t))$ under the H_1 (target present scenario). We have two choices for $\mathbf{C}^{-1}(t)$ in computing detection scores: with $\mathbf{C}^{-1}(t) \to \mathbf{C}_{\mathrm{geo}}^{-1}(t)$, or with $\mathbf{C}^{-1}(t) \to \mathbf{C}_{\mathrm{lin}}^{-1}(t)$. When the whitening $\mathbf{C}^{-1}(t)$ is $\mathbf{C}_{\mathrm{geo}}^{-1}(t)$, the test vector $\mathbf{C}^{-0.5}\boldsymbol{x}$ is perfectly whitened and the decision boundary between H_0 and H_1 is optimal, $\mathbf{C}^{-0.5}\boldsymbol{x}|H_0 \sim \mathrm{N}(0, \mathbf{I})$ and $\mathbf{C}^{-0.5}\boldsymbol{x}|H_1 \sim \mathrm{N}(\mathbf{C}^{-0.5}\boldsymbol{\delta}, \mathbf{I})$; detection scores are distributed as $\mathrm{S}_{\mathrm{geo}}(t)|H_0 \sim N(0, \mathrm{Var}_{\mathrm{geo}})$ and $\mathrm{S}_{\mathrm{geo}}(t)|H_1 \sim N(1, \mathrm{Var}_{\mathrm{geo}})$ where the geodesic variance is $\mathrm{Var}_{\mathrm{geo}} = [\boldsymbol{\delta}^{\mathrm{T}}\mathbf{C}_{\mathrm{geo}}^{-1}\boldsymbol{\delta}]^{-1}$. However, if the whitening $\mathbf{C}^{-1}(t)$ is taken to be $\mathbf{C}_{\mathrm{lin}}^{-1}(t)$ the whitening is imperfect, $\mathbf{C}^{-0.5}x|H_0 \sim \mathrm{N}(0, \mathbf{C}_{\mathrm{lin}}^{-0.5}\mathbf{C}_{\mathrm{geo}}\mathbf{C}_{\mathrm{lin}}^{-0.5})$, and $\mathbf{C}^{-0.5}\boldsymbol{x}|H_1 \sim \mathrm{N}(\mathbf{C}_{\mathrm{lin}}^{-0.5}\boldsymbol{\delta}, \mathbf{C}_{\mathrm{lin}}^{-0.5}\mathbf{C}_{\mathrm{geo}}\mathbf{C}_{\mathrm{lin}}^{-0.5})$; detection scores are distributed as $\mathrm{S}_{\mathrm{lin}}(t)|H_0 \sim \mathrm{N}(0, \mathrm{Var}_{\mathrm{lin}})$ and $\mathrm{S}_{\mathrm{lin}}(t)|H_1 \sim N(1, \mathrm{Var}_{\mathrm{lin}})$ with $\mathrm{Var}_{\mathrm{lin}} = [\boldsymbol{\delta}^{\mathrm{T}}\mathbf{C}_{\mathrm{lin}}^{-1}\mathbf{C}_{\mathrm{geo}}\mathbf{C}_{\mathrm{lin}}^{-1}\boldsymbol{\delta}][\boldsymbol{\delta}^{\mathrm{T}}\mathbf{C}_{\mathrm{lin}}^{-1}\boldsymbol{\delta}]^{-2}$. Due to the imperfect whitening by $\mathbf{C}_{\mathrm{lin}}$, the score variances $\mathrm{Var}_{\mathrm{lin}} \geq \mathrm{Var}_{\mathrm{geo}}$ and the detection and false alarm probabilities (given by the cumulative probability density of the H_1 and H_0 scores, respectively) are worse when the naive linear time-dependent covariance is used: $\mathrm{P}_{\mathrm{detect}}(\mathrm{linear}) \leq \mathrm{P}_{\mathrm{detect}}(\mathrm{geodesic})$ for a given value of false alarm probability, and the false alarm probability $\mathrm{P}_{\mathrm{false}}(\mathrm{linear}) \geq \mathrm{P}_{\mathrm{false}}(\mathrm{geodesic})$ for a given value of probability of detection.

V. Results

We explored (with simulations) the difference between $\mathbf{C}_{\mathrm{geo}}(t)$ and $\mathbf{C}_{\mathrm{lin}}(t)$. We also experimented with linear interpolation in the inverse space $\mathbf{A}^{-1} \to \mathbf{B}^{-1}$ and compared it to $\mathbf{C}^{-1}{}_{\mathrm{geo}}(t)$ in (1). We found that the volume of $\mathbf{C}_{\mathrm{lin}}(t)$ (volume equals the product of eigenvalues) is always larger than that of $\mathbf{C}_{\mathrm{geo}}(t)$, $V(\mathbf{C}_{\mathrm{lin}}(t)) > V(\mathbf{C}_{\mathrm{geo}}(t))$, and that the instantaneous distance $d(t) = d(\mathbf{C}_{\mathrm{geo}}(t), \mathbf{C}_{\mathrm{lin}}(t))$ can be up to tens of percents of the distance $d(\mathbf{A}, \mathbf{B})$. The objective of detection is to separate two populations H_0 and H_1, where the orientation of H_1 depends on the target direction, and detection is implemented with a specific algorithm (e.g., MF, spectral angle mapper, adaptive cosine estimator). The volume

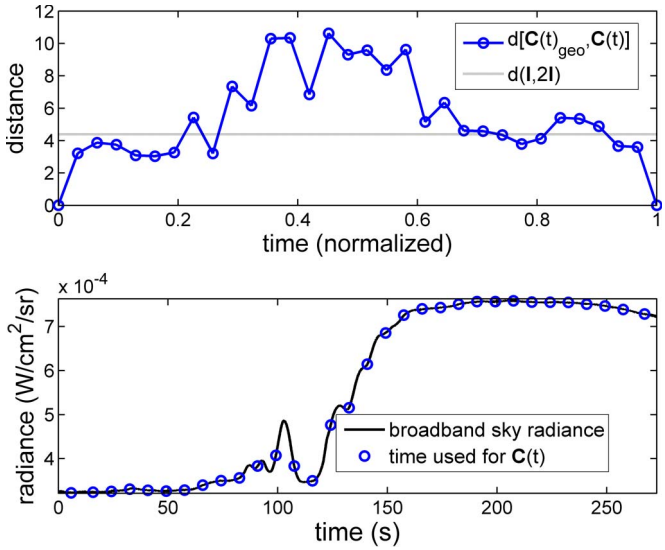

Fig. 1. Distance between predicted geodesic covariance $\mathbf{C}_{\text{geo}}(t)$ and field data $\mathbf{C}(t)$ as a function of time (upper panel). Broadband radiance integrated over all spectral bands (lower panel).

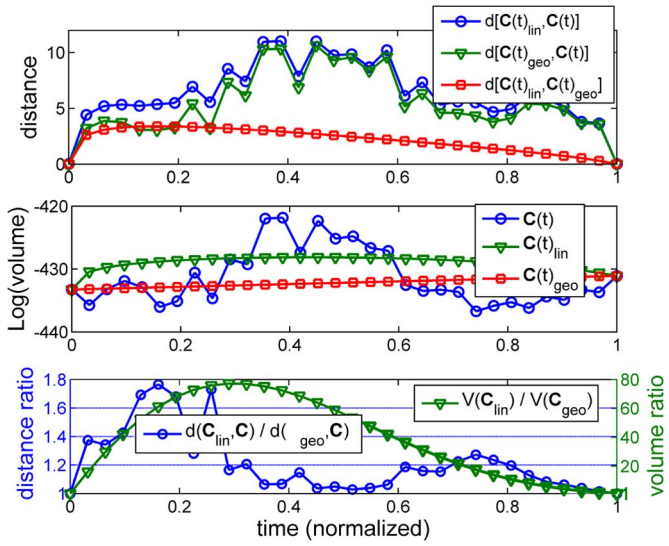

Fig. 2. Distances between covariance matrices (top panel). Volumes of covariance matrices (middle panel). Ratio of distances and ratio of volumes (bottom panel). $\mathbf{C}(t)$ is field data covariance.

of $\mathbf{C}(t)$ is related to separation between H_0 and H_1 data. As the separation increases, the detection performance improves. As the H_0 space increases in the direction of the target, the H_1 target space is more likely to be hidden inside the H_0 space and separation between $\boldsymbol{x}|H_0$ and $\boldsymbol{x}|H_1$ is more difficult. As the H_0 space (i.e., the volume of $\mathbf{C}(t)$) increases, the performance of a detector deteriorates because one must set a higher threshold around the H_0 data to minimize false alarm at the price of reduced probability of detection. When the change of volume is due to shape (structure) of $\mathbf{C}(t)$, i.e., rotation of eigenvectors and change in relative magnitude of eigenvalues, the detection scores (which depends on target direction) will be affected.

We present field data $\mathbf{C}(t)$ and compare it to geodesic $\mathbf{C}_{\text{geo}}(t)$ that is estimated from the boundary conditions $\mathbf{C}(t_0)$ and $\mathbf{C}(t_1)$ of the data—even though there is no guarantee that field data will be geodesic (as previously discussed). The measurements were collected on February 20, 2013 at MESH Inc., Oxford, Pennsylvania with MCAD sensor operating at 30 spectra/s acquisition rate; 8 cm^{-1} resolution; spectral range 770 cm^{-1} to 1650 cm^{-1}. The measurements were conducted at *very* variable conditions in order to maximize the change in the field data $\mathbf{C}(t)$. During the time of the measurements (12:32 to 12:36, local time) clouds drifted in and out of the field of view of the sensor that was pointing at 20° above the horizon; temperature changed from 271.6 K to 271.2 K (below freezing); relative humidity 57%; dew point −8.9 °C to −9.44 °C; high wind speed 4.9 m/s to 5.8 m/s with gusts; wind direction 302° to 322°. We selected 40 bands (p) uniformly spaced between 850 and 1150 cm^{-1} and computed $\mathbf{C}(t)$ with 250 spectra (m), and 350 spectra for the boundary conditions in order to enhance the quality of the boundary estimates $\mathbf{A}$ and $\mathbf{B}$. Hence, the ratio m/p in Wishart statistics was low (6.25) and therefore we used the algorithm of Ben–David and Davidson [14] to estimate $\mathbf{C}(t)$.

The effect of variable conditions during the measurements is shown in the broadband radiance (integrated over all spectral bands) in the bottom panel of Fig. 1. In Fig. 1 (top panel), we show the distance between the geodesic $\mathbf{C}_{\text{geo}}(t)$ and the field data $\mathbf{C}(t)$. In the upper plot, we also show the distance $d(\mathbf{I}, 2\mathbf{I}) = 4.4$ between two whitened distributions $N(0, \mathbf{I})$ and $N(0, 2\mathbf{I})$ that have a difference of 1-unit of variance (for $d(\mathbf{I}, 3\mathbf{I})$ the distance is 6.9). The distance due to sampling error with $m = 350$ and $p = 40$ is $d(\mathbf{I}, \hat{\mathbf{I}}) = d(\mathbf{I}, \mathbf{I} + \text{Var}(\hat{\mathbf{I}})) = 0.7$ and $d(\mathbf{I}, \mathbf{I} - \text{Var}(\hat{\mathbf{I}})) = 0.79$ which is small compared to the distances in Fig. 1. The figure shows that much of the distance between $\boldsymbol{C}_{\text{geo}}(t)$ and $\mathbf{C}(t)$ falls below the line $d(\mathbf{I}, 2\mathbf{I})$ and thus suggests that the predicted $\mathbf{C}_{\text{geo}}(t)$ is "close" to the field data. Furthermore, the effect of uncertainty $\text{Var}(\mathbf{C}(t))$ of the sampled field data $\mathbf{C}(t)$ due to the small number of spectra (vectors) on computing the distance between $\mathbf{C}_{\text{geo}}(t)$ and $\mathbf{C}(t) \pm [\text{Var}(\mathbf{C}(t))]^{0.5}$ is about 8 distance-units for the first and last third portions of the data $t < 1/3$ and $t > 2/3$, and 16 distance-units between $1/3 < t < 2/3$; hence, $\mathbf{C}_{\text{geo}}(t)$ is within the variation of $\mathbf{C}(t)$.

In Fig. 2 (top panel), we show the three distances: $\mathbf{C}_{\text{lin}}(t)$ to the field data $\mathbf{C}(t)$, $\mathbf{C}_{\text{geo}}(t)$ to $\mathbf{C}(t)$, and $\mathbf{C}_{\text{lin}}(t)$ to $\mathbf{C}_{\text{geo}}(t)$. In comparing these distances, we regard the field data $\mathbf{C}(t)$ as "error free" in order to explore the difference between $\mathbf{C}_{\text{geo}}(t)$ predictions and the "naive" $\mathbf{C}_{\text{lin}}(t)$ predictions. The figure (top panel) shows that the distance between $\mathbf{C}_{\text{geo}}(t)$ and $\mathbf{C}_{\text{lin}}(t)$ is small, but nevertheless, the distance $\mathbf{C}_{\text{lin}}(t)$ to $\mathbf{C}(t)$ is larger than the distance $\mathbf{C}_{\text{geo}}(t)$ to $\mathbf{C}(t)$; the ratio between the two is shown in the bottom panel (y-axis on the left) to be up to two-folds; hence, the geodesic prediction is much closer to the field data than the linear predictions. In the middle panel, we show the logarithm of the p-D volume, $\log(V) = \sum \log(eigenvalues)$, for the hyper-ellipsoids, $\mathbf{C}(t)$, $\mathbf{C}_{\text{lin}}(t)$, and $\mathbf{C}_{\text{geo}}(t)$. The figure shows that the volume of $\mathbf{C}_{\text{lin}}(t)$ is much larger than of $\mathbf{C}_{\text{geo}}(t)$; the ratio $V_{\text{lin}}(t)/V_{\text{geo}}(t) = \exp[\log(V_{\text{lin}}(t)) - \log(V_{\text{geo}}(t))]$ is shown in the bottom panel (y-axis on the right) to be up to 80-fold.

In Fig. 3, we show the signal-to-noise ratio (SNR) for a matched filter detector for the following scenarios: $\boldsymbol{x}(t)|H_0 \sim N(0, \mathbf{C}(t))$, $\boldsymbol{x}(t)|H_1 \sim N(\boldsymbol{\delta}, \mathbf{C}(t))$, $\mathbf{C}(t)$ is the field data. The SNR[4], $SNR = [\text{E}(S|H_1) - \text{E}(S|H_0)]^2/\text{Var}(S|H_0)$, when

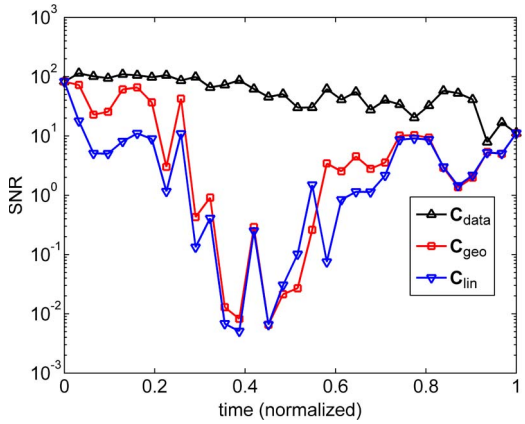

Fig. 3. Signal-to-noise ratio for matched filter scores.

the whitening in the MF is done with the data $\mathbf{C}^{-1}(t)$, with $\mathbf{C}_{\text{geo}}^{-1}(t)$, or with $\mathbf{C}_{\text{lin}}^{-1}(t)$ is

$$\left.\begin{aligned} SNR_{\text{geo}}(t) &= \left[\boldsymbol{\delta}^T \mathbf{C}_{\text{geo}}^{-1}(t)\boldsymbol{\delta}\right]^2 \left[\boldsymbol{\delta}^T \mathbf{C}_{\text{geo}}^{-1}(t)\mathbf{C}(t)\mathbf{C}_{\text{geo}}^{-1}(t)\boldsymbol{\delta}\right]^{-1} \\ SNR_{\text{lin}}(t) &= \left[\boldsymbol{\delta}^T \mathbf{C}_{\text{lin}}^{-1}(t)\boldsymbol{\delta}\right]^2 \left[\boldsymbol{\delta}^T \mathbf{C}_{\text{lin}}^{-1}(t)\mathbf{C}(t)\mathbf{C}_{\text{lin}}^{-1}(t)\boldsymbol{\delta}\right]^{-1} \\ SNR_{\text{data}}(t) &= \boldsymbol{\delta}^T \mathbf{C}^{-1}(t)\boldsymbol{\delta} \end{aligned}\right\}. \tag{3}$$

SNR solely determines the performance of MF detectors [Ch. 4.3.2 in 4]. The optimal whitening is done with the data $\mathbf{C}^{-1}(t)$ for which SNR_{data} is optimal. However, in most scenarios, $\mathbf{C}(t)$ is not available and we resort to prediction for $\mathbf{C}^{-1}(t)$. The target vector $\boldsymbol{\delta}$ was chosen to be the $\mathbf{1}$ vector (all p elements equal to 1) in order to give equal weight to all the p eigenvectors (directions) of $\mathbf{C}(t)$ and was scaled to have a norm of 0.1. SNR is highly dependent on the spectral direction of the target with respect to the eigenvectors of $\mathbf{C}(t)$. Fig. 3 shows that SNR using the geodesic predictions for whitening, for this target, is higher (by $\sim$3.6 on average) than if linear predictions are used for whitening. However, the drop of SNR from the optimal SNR (top curve) shows that the geodesic SNR predictions are not very good near the middle of the time interval $(t \sim 0.4)$ where large changes in radiance occur (see Fig. 1, bottom panel, for t around 100 to 150 s).

VI. Conclusion

Time-dependent covariance matrix $\mathbf{C}(t)$ is important in remote sensing and hyperspectral detection theory. The difficulty is that $\mathbf{C}(t)$ is usually available only at $\mathbf{C}(t_0) = \mathbf{A}$ and $\mathbf{C}(t_1) = \mathbf{B}$ whereas $\mathbf{C}(t_0 < t < t_1)$ is needed for detection algorithms. We presented the Riemannian manifold of positive definite symmetric matrices as a framework for predicting geodesic time-dependent covariance matrix $\mathbf{C}(t_0 < t < t_1)$. The geodesic path for $\mathbf{C}(t)$ is the shortest and most efficient path (minimum energy) between two boundary conditions. In this letter, we wanted to introduce the relevance of the geodesic concept to the hyperspectral remote sensing community. The path for inverse covariance $\mathbf{C}^{-1}(t)$ from $\mathbf{A} \to \mathbf{B}$ is also geodesic (inverse covariance matrices play a key role in detection theory). We compared the geodesic covariance and the naive linearly predicted covariance $\mathbf{C}_{\text{lin}}(t)$ that linearly traverses the path $\mathbf{A} \to \mathbf{B}$ in simulations and in field data and showed the difference in performance of matched-filter detection algorithm. We showed that the volume of the geodesic covariance is smaller than that of $\mathbf{C}_{\text{lin}}(t)$ and thus using $\mathbf{C}_{\text{geo}}(t)$ will increase the separation between the H_0 and H_1 detection scores and the detection performance will improve (false alarm and detection probabilities depend on the detection algorithm and location of targets). The distance $d(\mathbf{C}_1, \mathbf{C}_2)$ is interpreted as deviation from sphericity in attempting to whiten one matrix by the other. We presented the covariance $\mathbf{C}_{\text{geo}}(t)$ as the sum of scaled, stretched, contracted, and rotated elliptical covariance matrices. For hyperspectral vector $\boldsymbol{x}(t) \sim N(0, \mathbf{C}(t))$, we interpret the geodesic time-dependent covariance a sum of rotated ellipses (rotated and colored by the boundary condition $\mathbf{A}^{0.5}$ at $t = 0$) where ellipses change their eccentricities (by $\mathbf{M}^n$), are scaled (by $(n!)^{-1}t^n$), and the matrix $\mathbf{M}$ is related to (attempted) whitening of the boundary condition at $\mathrm{t} = 1$ (covariance $\mathbf{B}$) by the boundary condition at $t = 0$ (covariance $\mathbf{A}$). The eigenvalues and the eigenvectors of $\mathbf{C}_{\text{geo}}(t)$ are time dependent; thus, in simulations for $\mathbf{C}(t)$, we suggest to construct $\mathbf{C}(t) = \mathbf{U}(t)\boldsymbol{\Lambda}(t)\mathbf{U}^T(t)$. Although there is no guarantee that covariance data will necessarily follow a geodesic path (in the same way that airplanes do not always fly in the shortest path between two cities along great circles), the geodesic time-dependent covariance $\mathbf{C}(t)$ is of value as a concept that gives insight to hyperspectral remote sensing algorithms where $\mathbf{C}(t)$ plays a key role.

Acknowledgment

The authors extend heart-felt thanks to C. Davidson (STC Corp.) for his insight, discussions, and help in writing software. The authors thank T. Gruber and Mesh Inc. for providing field data.

References

[1] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?" in *Proc. SPIE Algorithms Technol. Multispectr., Hyperspectr., Ultraspectr. Imagery XV*, 2009, vol. 7334, no. 1, pp. 733402-1–733402-16.

[2] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[3] L. L. Scharf, *Statistical Signal Processing, Detection, Estimation, Time Series Analysis*. Reading, MA, USA: Addison-Wesley, 1991.

[4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. II. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[5] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, Jan. 2006.

[6] R. L. Bishop and S. Goldberg, *Tensor Analysis on Manifolds*. New York, NY, USA: Dover, 1980.

[7] A. Ben-David, C. E. Davidson, and R. G. Vanderbeek, "Lidar detection algorithm for time and range anomalies," *Appl. Opt.*, vol. 46, pp. 7275–7288, Oct. 2007.

[8] O. Tuzel, F. Porikli, and P. Meer, "Human detection via classification on Riemannian manifolds," in *Proc. IEEE CVPR*, Jun. 2007, pp. 1–8.

[9] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20A, no. 1/2, pp. 32–52, Jul. 1928.

[10] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Hoboken, NJ, USA: Wiley, 2003.

[11] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall, 2000.

[12] G. Letac and H. Massam, "All invariant moments of the Wishart distribution," *Scand. J. Stat.*, vol. 31, no. 2, pp. 295–318, Jun. 2004.

[13] C. E. Davidson and A. Ben-David, "Performance loss of multivariate detection algorithms due to covariance estimation," in *Proc. SPIE Image Signal Process. Remote Sens. XV*, Berlin, Germany, Aug. 2009, pp. 74770J-1–74770J-12.

[14] A. Ben-David and C. E. Davidson, "Eigenvalues estimation of hyperspectral Wishart covariance matrices from limited number of samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4384–4396, Nov. 2012.