A BAYESIAN DERIVATION OF GENERALIZED COHERENCE DETECTORS

Songsri Sirianunpiboon[⋆] *Stephen D. Howard*[⋆] *Douglas Cochran*[†]

[⋆]Defence Science and Technology Organisation
PO Box 1500, Edinburgh 5111, Australia.
[†]School of Mathematical and Statistical Sciences
Arizona State University, Tempe AZ 85287-5706 USA.

ABSTRACT

The generalized coherence (GC) estimate is a well studied statistic for detection of a common but unknown signal on several noisy channels. In this paper, it is shown that the GC detector arises naturally from a Bayesian perspective. Specifically, it is derived as a test of the hypothesis that the signals in the channels are independent Gaussian processes against the hypothesis that the processes have some arbitrary correlation. This is achieved by introducing suitable non-informative priors for the covariance matrices across the channels under the two hypotheses. Subsequently, reduced likelihoods are obtained by marginalizing the joint distribution of the data and the covariance matrix in each case. The likelihood ratio is then shown to be a monotonic function of the GC detection statistic. This derivation extends to the case of time-correlated signals, allowing comparison with the generalized likelihood ratio test (GLRT) recently proposed by Ramírez et al.

***Index Terms*—** Bayesian methods, generalized coherence estimate, multi-channel signal processing, non-informative priors.

1. INTRODUCTION

The problem of detecting a common but unknown signal on two or more noisy channels has a rich history and has recently received renewed interest [1]. Pioneering work in the context of sonar signal processing led to widespread use of the magnitude-squared coherence (MSC) estimate as a detection statistic in the two-channel setting [2]. The properties of the MSC estimate and the performance of detectors based upon it were studied extensively in the 1970s and 1980s [3, 4]. Extending the MSC concept to multiple-channel scenarios received considerable attention in the 1980s, leading to study of estimators of multiple coherence [5] and, somewhat later, generalized coherence (GC) as multi-channel detection statistics. The GC estimate was introduced in [6] in the context of multiple-channel detection, and the performance of the GC estimator as a detection statistic was documented in several works throughout the 1990s and into the early 2000s. Its symmetries and invariance properties were well studied during this period [7, 8], and the extent to which the GC estimate is canonical with respect to a desirable set of invariances and symmetries was examined in [9].

In this paper we take a fresh look at the GC detector from a Bayesian perspective. Specifically, we show that, from this point of view, it arises naturally as the statistic for a test of the null hypothesis that the signals in the channels are independent Gaussian processes against the alternative hypothesis that the processes are Gaussian with some arbitrary correlation. This is achieved by introducing suitable non-informative priors for the covariance matrices across the channels under the two hypotheses. Care must be taken in choosing the priors, as the two hypotheses correspond to parameter spaces of different dimensionality and one must choose consistent priors to avoid inadvertently favoring one of the hypotheses *a priori*. After choosing consistent "non-informative" priors, the marginalized likelihoods are obtained by marginalizing the joint distribution of the data and the covariance matrix in each case. The likelihood ratio is then shown to be a monotonic function of the GC detection statistic. This perspective directly connects the GC statistic to the Bayesian (as opposed to the Neyman-Pearson) detector for a precisely defined multi-channel problem.

The classical GC detector assumes that the time series in each channel are temporally white. In recent work, Ramírez et al. [1] have extended the GC detector to account for unknown temporal correlations in the multi-channel data. They assume that the data consists of L independent time blocks and derive a GLRT detector to test for spatial correlation among the channels in the presence of unknown temporal correlation. We show that this GLRT statistic also arises as a Bayesian marginalized likelihood ratio test.

Notation: In this paper we use bold lowercase symbols to represent complex vectors and bold uppercase symbols for complex matrices. $\dagger$ denotes Hermitian transpose, $|\cdot|$ determinant, and tr the trace operator. $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and $\mathbf{0}_n$ the $n \times n$ zero matrix. $\|\boldsymbol{x}\|^2$ is the squared (Euclidean) norm of $\boldsymbol{x}$, and $\boldsymbol{x} \sim \mathcal{CN}(\mu, \boldsymbol{R})$ means that $\boldsymbol{x}$ is a complex Gaussian random vector with mean μ and covariance $\boldsymbol{R}$.

2. PROBLEM FORMULATION

Consider M complex random processes representing the outputs of M spatially distributed sensors. For $m = 1, \ldots, M$, let $\boldsymbol{x}_m^\dagger = (x_{m,1}, \ldots, x_{m,N})$ be a vector of N complex measurements obtained from sensor m and $\boldsymbol{X}^\dagger = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$. The generalized coherence (GC) estimate [6, 9] is defined as

$$\gamma^2 = 1 - \frac{|\boldsymbol{X}\boldsymbol{X}^\dagger|}{\|\boldsymbol{x}_1\|^2 \ldots \|\boldsymbol{x}_M\|^2} \tag{1}$$

where $\boldsymbol{X}\boldsymbol{X}^\dagger$ is the $M \times M$ Gram matrix of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$.

The goal here is to derive the GC detector from the Bayesian perspective. The detector's value in determining whether a common signal is present on two or more noisy channels is predicated on distinguishing between hypotheses H_0, in which the vector $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ contains independent samples of independent zero-mean Gaussian processes with unknown covariance, and its complement H_1 in which these vectors contain unknown and correlated signals. Symbolically,

$$H_1 : \boldsymbol{X} \sim \mathcal{CN}(0, \boldsymbol{R}_1), \text{ for some covariance } \boldsymbol{R}_1$$
$$H_0 : \boldsymbol{X} \sim \mathcal{CN}(0, \boldsymbol{R}_0), \text{ for some diagonal covariance } \boldsymbol{R}_0.$$

Let $\boldsymbol{x}(n) = (x_{1,n}, x_{2,n}, \ldots, x_{M,n})^\dagger$ be a vector of samples at time n from M sensors. The likelihood function of H_0 is

$$\begin{aligned} p(\boldsymbol{X}|H_0, \boldsymbol{R}_0^{-1}) &= \prod_{n=1}^{N} p(\boldsymbol{x}(n)|H_0, \boldsymbol{R}_0) \\ &= \pi^{-MN} |\boldsymbol{R}_0^{-1}|^N e^{-\mathrm{tr}\{\boldsymbol{R}_0^{-1}\boldsymbol{W}\}} \end{aligned} \tag{2}$$

where $\boldsymbol{W} = \sum_{n=1}^{N} \boldsymbol{x}(n)\boldsymbol{x}(n)^\dagger = \boldsymbol{X}\boldsymbol{X}^\dagger$. Similarly, the likelihood function of H_1 is

$$p(\boldsymbol{X}|H_1, \boldsymbol{R}_1^{-1}) = \pi^{-MN} |\boldsymbol{R}_1^{-1}|^N e^{-\mathrm{tr}\{\boldsymbol{R}_1^{-1}\boldsymbol{W}\}}. \tag{3}$$

The parameters $\boldsymbol{R}_0$ and $\boldsymbol{R}_1$ are nuisance parameters in the likelihood which we want to eliminate. Since the likelihood is expressed in terms of $\boldsymbol{R}_0^{-1}$ and $\boldsymbol{R}_1^{-1}$, we regard these to be our nuisance parameters instead. The Bayesian treatment of such nuisance parameters is to marginalize the posterior distribution with respect to them. In this instance, we consider the integrated likelihoods

$$p(\boldsymbol{X}|H_0) = \int p(\boldsymbol{X}|H_0, \boldsymbol{R}_0^{-1}) p(\boldsymbol{R}_0^{-1}) d\boldsymbol{R}_0^{-1} \tag{4}$$

$$p(\boldsymbol{X}|H_1) = \int p(\boldsymbol{X}|H_1, \boldsymbol{R}_1^{-1}) p(\boldsymbol{R}_1^{-1}) d\boldsymbol{R}_1^{-1}. \tag{5}$$

Here

$$d\boldsymbol{R}_0^{-1} = \prod_{i=1}^{M} dr^{ii} \tag{6}$$

$$d\boldsymbol{R}_1^{-1} = \prod_{i=1}^{M} dr^{ii} \prod_{i<j} d\Re(r^{ij}) d\Im(r^{ij}) \tag{7}$$

where r^{ij} denotes the $(i,j)^{\text{th}}$ element of $\boldsymbol{R}^{-1}$ and $\Re$ and $\Im$ denote real and imaginary parts, respectively.

Through this procedure we have replaced the data models containing the unknown parameters $\boldsymbol{R}_0^{-1}$ and $\boldsymbol{R}_1^{-1}$, with the models $p(\boldsymbol{X}|H_0)$ and $p(\boldsymbol{X}|H_1)$ so that the detection problem compares

$$H_1 : \boldsymbol{X} \sim p(\,\cdot\,|H_1) \tag{8}$$
$$H_0 : \boldsymbol{X} \sim p(\,\cdot\,|H_0). \tag{9}$$

We proceed to show that the GC detection statistic is a monotonic function of the likelihood ratio, $\frac{p(\boldsymbol{X}|H_1)}{p(\boldsymbol{X}|H_0)}$.

3. PRIOR PROBABILITY DISTRIBUTION

In order to carry out the integrations in (4) and (5), we need prior distributions on $\boldsymbol{R}_0^{-1}$ and $\boldsymbol{R}_1^{-1}$. We would like these priors to be as non-informative as possible, while still leading to proper densities $p(\boldsymbol{X}|H_1)$ and $p(\boldsymbol{X}|H_0)$.

3.1. Jeffreys Priors

Consider the parameter manifold $\mathcal{M}_1$ of the positive definite matrix $\boldsymbol{R}_1$. Since every positive definite matrix has a unique inverse we can equally well parametrize this manifold in terms of inverse covariance matrices $\boldsymbol{R}_1^{-1}$. The manifold $\mathcal{M}_1$ is endowed with a Riemannian metric by the Fisher information associated with estimation of $\boldsymbol{R}_1^{-1}$ from data, based on the model H_1. That is,

$$F = \mathrm{E}_{\boldsymbol{X}} \left\{ d \log p(\boldsymbol{X}|\boldsymbol{R}_1^{-1}) \otimes d \log p(\boldsymbol{X}|\boldsymbol{R}_1^{-1}) \right\} \tag{10}$$

where d denotes the exterior derivative on $\mathcal{M}_1$ [10]. The Jeffreys prior on $\mathcal{M}_1$ is the volume form associated with the Fisher metric,

$$\mathrm{vol} = \sqrt{|F|} d\boldsymbol{R}_1^{-1}. \tag{11}$$

This is commonly used as a canonical non-informative prior, as it is invariant with respect to re-parametrization [11, 12]. For the complex data model, the Jeffreys prior for $\boldsymbol{R}_1^{-1}$ is given by

$$p_J(\boldsymbol{R}_1^{-1})\; d\boldsymbol{R}_1^{-1} = |\boldsymbol{R}_1^{-1}|^{-M} d\boldsymbol{R}_1^{-1}. \tag{12}$$

As a probability distribution, this is improper. Often this is not a problem as long as the posterior distribution can be normalized. In our case, the marginalized data model distribution is

$$\begin{aligned} p(\boldsymbol{X}|H_1) &\propto \frac{1}{\pi^{MN}} \int \det(\boldsymbol{R}_1^{-1})^{N-M} e^{-\mathrm{tr}\{\boldsymbol{R}_1^{-1}\boldsymbol{X}\boldsymbol{X}^\dagger\}} d\boldsymbol{R}_1^{-1} \\ &\propto \frac{\pi^{M(M-1)/2}}{\pi^{MN}} \prod_{i=1}^{M} \Gamma(N - M + i) |\boldsymbol{X}\boldsymbol{X}^\dagger|^{-N} \end{aligned} \tag{13}$$

which is still improper. Similarly, the Jeffreys prior for $\boldsymbol{R}_0^{-1}$ is

$$p_J(\boldsymbol{R}_0^{-1}) d\boldsymbol{R}_0^{-1} = |\boldsymbol{R}_0^{-1}|^{-1} d\boldsymbol{R}_0^{-1} \tag{14}$$

and the marginalized data model distribution is

$$p(\boldsymbol{X}|H_0) \propto \frac{1}{\pi^{MN}} \Gamma^M(N) \left(\prod_{i=1}^{M} \|\boldsymbol{x}_i\|^2 \right)^{-N}. \tag{15}$$

Again this is improper, so we need to be more careful in specifying a prior distribution. As choosing appropriate priors is crucial to the Bayesian approach, we proceed with a more judicious choice of prior distributions while retaining the Jeffreys priors as the basis of the choice.

3.2. Maximum Entropy Priors

In the previous section, we found that the use of the improper Jeffreys priors led to improper marginalized data model distributions. This suggests the need for more informative prior distributions. Our approach is to introduce enough prior information to give proper marginalized data model densities $p(\boldsymbol{X}|H_1)$ and $p(\boldsymbol{X}|H_0)$ and then relax this as much as possible in the final result in a manner we now describe.

The maximum entropy distribution, subject to certain constraints, represents the least informative distribution [13] satisfying those constraints. Our approach is to construct proper non-informative prior distributions by constructing maximum entropy distributions relative to the Jeffreys prior under each of the two hypotheses, subject to the conditions

$$\mathrm{E}(\mathrm{tr}(\boldsymbol{R}^{-1})) = \alpha_1 \tag{16}$$
$$\mathrm{E}(\log(|\boldsymbol{R}|)) = \alpha_2. \tag{17}$$

First consider the case that $\boldsymbol{R}_1 \in \mathcal{M}_1$. The maximum entropy distribution of $\boldsymbol{R}_1^{-1}$ relative to the Jeffreys prior, subject to (16) and (17) is given by

$$\begin{aligned} p(\boldsymbol{R}_1^{-1}) &= Z_1(\tau, q)^{-1} |\boldsymbol{R}_1^{-1}|^{-M} e^{-\tau \mathrm{tr}(\boldsymbol{R}_1^{-1}) - q \log(|\boldsymbol{R}_1|)} \\ &= Z_1(\tau, q)^{-1} |\boldsymbol{R}_1^{-1}|^{q-M} e^{-\mathrm{tr}(\tau \boldsymbol{R}_1^{-1})} \end{aligned} \tag{18}$$

where τ and q are chosen to satisfy the conditions (16) and (17) and $Z_1(\tau, q)$ is the normalization constant or partition function,

$$\begin{aligned} Z_1(\tau, q) &= \int_{\mathcal{M}_1} \det(\boldsymbol{R}_1^{-1})^{q-M} e^{-\mathrm{tr}(\tau \boldsymbol{R}_1^{-1})} d\boldsymbol{R}_1^{-1} \\ &= \pi^{M(M-1)/2} \tau^{-Mq} \prod_{i=1}^{M} \Gamma(q - M + i). \end{aligned} \tag{19}$$

The prior (18) is proper for $\tau > 0$ and $q > M - 1$. Notice that, in the limit $\tau, q \to 0$, the prior (18) becomes less informative and approaches the Jeffreys prior up to normalization.

The moments (16) and (17) can be computed using the partition function $Z_1(\tau, q)$. Specifically,

$$\mathrm{E}(\mathrm{tr}(\boldsymbol{R}_1^{-1})) = -\frac{\partial}{\partial \tau} \log Z_1(\tau, q) = Mq\tau^{-1} \tag{20}$$

and

$$\mathrm{E}(\log|\boldsymbol{R}_1|) = -\frac{\partial}{\partial q} \log Z_1(\tau, q) = M \log \tau - \sum_{i=1}^{M} \psi(q - M + i) \tag{21}$$

where ψ is the digamma function.

Now consider the prior under H_0. In this case $R_0 \in \mathcal{M}_0$, and the maximum entropy distribution relative to Jeffreys prior subject to the constraints (16) and (17) is

$$p(\boldsymbol{R}_0^{-1}) = Z_0(\tau, q)^{-1} |\boldsymbol{R}_0^{-1}|^{q-1} e^{-\mathrm{tr}(\tau \boldsymbol{R}_0^{-1})}. \tag{22}$$

In this case the partition function is

$$Z_0(\tau, q) = \int_{\mathcal{M}_0} |\boldsymbol{R}_0^{-1}|^{q-1} e^{-\mathrm{tr}(\tau \boldsymbol{R}_0^{-1})} d\boldsymbol{R}_0^{-1} = \tau^{-Mq} \Gamma^M(q) \tag{23}$$

implying that the prior (22) is proper for $\tau > 0$ and $q > 0$. We have the moments

$$\mathrm{E}(\mathrm{tr}(\boldsymbol{R}_0^{-1})) = -\frac{\partial}{\partial \tau} \log Z_0(\tau, q) = Mq\tau^{-1} \tag{24}$$

and

$$\mathrm{E}(\log|\boldsymbol{R}_0|) = -\frac{\partial}{\partial q} \log Z_0(\tau, q) = M \log \tau - M\psi(q). \tag{25}$$

We first note that (24) and (20) are consistent. Secondly, as $\tau \to 0$ we have,

$$\begin{aligned} \frac{\mathrm{E}(\log|\boldsymbol{R}_1|) - \mathrm{E}(\log|\boldsymbol{R}_0|)}{\mathrm{E}(\log|\boldsymbol{R}_0|)} &\sim \frac{M\psi(q) - \sum_{i=1}^{M} \psi(q - M + i)}{M \log \tau} \\ &= \frac{\sum_{i=1}^{M-1} \frac{i}{q-M+i}}{M \log \tau}, \end{aligned} \tag{26}$$

where the last equality uses the identity $\psi(x + 1) = \psi(x) + 1/x$. For $q = M$ this becomes

$$\frac{\mathrm{E}(\log|\boldsymbol{R}_1|) - \mathrm{E}(\log|\boldsymbol{R}_0|)}{\mathrm{E}(\log|\boldsymbol{R}_0|)} \sim \frac{M - 1}{M \log \tau}. \tag{27}$$

Thus, the prior probability distributions for both hypotheses are consistent with respect to the constraints (16) and (17) in the limit $\tau \to 0$.

4. GENERALIZED COHERENCE ESTIMATE AS A LIKELIHOOD RATIO

In this section we show that the GC detection statistic is obtained from the reduced likelihood ratio of the two hypotheses. We first calculate the posterior distributions for the priors on $\boldsymbol{R}_0^{-1}$ and $\boldsymbol{R}_1^{-1}$, discussed in the previous section. The reduced data models are obtained by marginalizing the likelihoods with respect to these parameters. Substituting (2) and (22) into (4), we obtain the marginalized likelihood for hypothesis H_0,

$$\begin{aligned} &p(\boldsymbol{X}|H_0) \\ &= \frac{\tau^{Mq} \pi^{-MN}}{\Gamma^M(q)} \int_{\mathcal{M}_0} |\boldsymbol{R}_0^{-1}|^{N+q-1} e^{-\mathrm{tr}(\boldsymbol{R}_0^{-1}(\boldsymbol{W} + \tau \boldsymbol{I}_M))} d\boldsymbol{R}_0^{-1} \\ &= \frac{\tau^{Mq} \pi^{-MN}}{\Gamma^M(q)} \prod_{i=1}^{M} \int_0^\infty \sigma_i^{-2(N+q)-1} e^{-\sigma_i^{-2}(\|\boldsymbol{x}_i\|^2 + \tau)} d\sigma_i \\ &= \frac{\tau^{Mq} \Gamma^M(N+q)}{\pi^{MN} \Gamma^M(q) \prod_{i=1}^{M} (\|\boldsymbol{x}_i\|^2 + \tau)^{N+q}}. \end{aligned} \tag{28}$$

For hypothesis H_1, substituting (3) and (18) into (5) yields

$$\begin{aligned} &p(\boldsymbol{X}|H_1) \\ &= \frac{\tau^{Mq} \int_{\mathcal{M}_1} |\boldsymbol{R}_1^{-1}|^{N+q-M} e^{-\mathrm{tr}\{\boldsymbol{R}_1^{-1}(\boldsymbol{W} + \tau \boldsymbol{I}_M)\}} d\boldsymbol{R}_1^{-1}}{\pi^{MN} \pi^{M(M-1)/2} \prod_{i=1}^{M} \Gamma(q - M + i)} \\ &= \frac{\tau^{Mq} \prod_{i=1}^{M} \Gamma(N + q - M + i)}{\pi^{MN} \prod_{i=1}^{M} \Gamma(q - M + i) |\boldsymbol{W} + \tau \boldsymbol{I}_M|^{N+q}}. \end{aligned}$$

The likelihood ratio statistic for testing H_1 against H_0 is thus given by

$$\frac{p(\boldsymbol{X}|H_1)}{p(\boldsymbol{X}|H_0)} = \frac{k(N + q, M)}{k(q, M)} \left(\frac{\prod_{i=1}^{M} (\|\boldsymbol{x}_i\|^2 + \tau)}{|\boldsymbol{W} + \tau \boldsymbol{I}_M|} \right)^{N+q} \tag{29}$$

where

$$k(x, M) = \prod_{i=1}^{M} \frac{\Gamma(x - M + i)}{\Gamma(x)}. \tag{30}$$

As $\tau \to 0$, the likelihood ratio becomes

$$\frac{p(\boldsymbol{X}|H_1)}{p(\boldsymbol{X}|H_0)} = \frac{k(N + q, M)}{k(q, M)} \left(\frac{\prod_{i=1}^{M} \|\boldsymbol{x}_i\|^2}{|\boldsymbol{X}\boldsymbol{X}^\dagger|} \right)^{N+q}. \tag{31}$$

The GC estimate γ^2 given in (1) [6, 9] is related to this likelihood ratio (31) by

$$\frac{p(\boldsymbol{X}|H_1)}{p(\boldsymbol{X}|H_0)} = \frac{k(N + q, M)}{k(q, M)} \frac{1}{(1 - \gamma^2)^{N+q}} \tag{32}$$

and is hence a monotonic function of the likelihood ratio. In the case $M = 2$, we have

$$\frac{p(\boldsymbol{X}|H_1)}{p(\boldsymbol{X}|H_0)} = \frac{q - 1}{N + q - 1} \left(1 - \frac{|\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle|^2}{\|\boldsymbol{x}_1\|^2 \|\boldsymbol{x}_2\|^2} \right)^{-(N+q)} \tag{33}$$

where $|\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle|^2$ is the absolute square of the inner product of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

5. DETECTION THRESHOLDS

The Neyman-Pearson detection thresholds for the generalized coherence detector have been explored in detail in [9]. Here we consider the Bayesian detection thresholds.

Let C_{ij} for $i, j = 0, 1$ denote the loss function for our decision problem; i.e., C_{ij} is the loss incurred by deciding in favor of H_i when H_j is true. We take $C_{00} = C_{11} = 0$. The expected losses incurred in deciding in favor of each hypothesis are

$$\rho(H_1) = C_{10} p(H_0|\boldsymbol{X}) \quad \text{and} \quad \rho(H_0) = C_{01} p(H_1|\boldsymbol{X}). \tag{34}$$

The Bayesian decision minimizes the expected loss; i.e., H_1 is favored if

$$\frac{\rho(H_1)}{\rho(H_0)} = \frac{C_{10} p(H_0|\boldsymbol{X})}{C_{01} p(H_1|\boldsymbol{X})} < 1, \tag{35}$$

which may be rewritten as

$$\frac{p(\boldsymbol{X}|H_1)}{p(\boldsymbol{X}|H_0)} > \frac{C_{10} p(H_0)}{C_{01} p(H_1)} \equiv \beta. \tag{36}$$

It follows from (31) and (32) that the Bayesian detection threshold for GC is given by

$$\gamma^2 > 1 - \left(\frac{k(N+q, M)}{k(q, M)} \frac{1}{\beta} \right)^{\frac{1}{N+q}}. \tag{37}$$

6. TIME-CORRELATED SIGNALS DETECTOR AS LIKELIHOOD RATIO

In this section we show that the GLRT for the case of time-correlated signals recently derived by Ramírez et al. [1] can be also obtained as a marginalized likelihood ratio from a Bayesian perspective. In this case, it is assumed that each of the L rows of the data matrix $\boldsymbol{X}$ is an N-vector of time samples. $\boldsymbol{X}$ is written as a LN-vector $\boldsymbol{z} = \text{vec}(\boldsymbol{X}^{\mathsf{T}})$. Consider the hypotheses

$$\begin{aligned} H_1 &: \boldsymbol{z} \sim \mathcal{CN}(0, \boldsymbol{R}_1) \\ H_0 &: \boldsymbol{z} \sim \mathcal{CN}(0, \boldsymbol{R}_0) \end{aligned} \tag{38}$$

where $\boldsymbol{R}_1$ is some unknown covariance matrix and $\boldsymbol{R}_0$ is some unknown block-diagonal covariance matrix; i.e., $\boldsymbol{R}_{ij} = \boldsymbol{0}_N, i \neq j$. Under H_0, the spatially uncorrelated vector-valued time series may be temporally correlated [1]. As in Section 2, through the procedure of marginalizing the unknown parameters $\boldsymbol{R}_0$ and $\boldsymbol{R}_1$, we can compare the two hypotheses as shown in (8) and (9).

Let $\boldsymbol{Z} = (\boldsymbol{z}(1) \ldots \boldsymbol{z}(L))$, the joint likelihood function of H_0 and H_1 are

$$p(\boldsymbol{Z}|H_0, \boldsymbol{R}_0) = \pi^{-MNL} |\boldsymbol{R}_0^{-1}|^L e^{-\text{tr}(\boldsymbol{R}_0^{-1}\boldsymbol{Q})} \tag{39}$$

$$p(\boldsymbol{Z}|H_1, \boldsymbol{R}_1) = \pi^{-MNL} |\boldsymbol{R}_1^{-1}|^L e^{-\text{tr}(\boldsymbol{R}_1^{-1}\boldsymbol{Q})} \tag{40}$$

where $\boldsymbol{Q} = \sum_{l=1}^{L} \boldsymbol{z}(l)\boldsymbol{z}(l)^\dagger$ with $\boldsymbol{Q}_{ij} = \sum_{l=1}^{L} \boldsymbol{x}_i(l)\boldsymbol{x}_j(l)^\dagger$ being its block element. Using the prior $p(\boldsymbol{R}_0^{-1})$ as given in (22) and marginalizing with respect to the nuisance parameters $\boldsymbol{R}_0^{-1}$, we obtain the reduced likelihood for hypothesis H_0; i.e.,

$$\begin{aligned} p(\boldsymbol{Z}|H_0) &= \frac{\pi^{-MNL}\tau^{Mq}}{\Gamma^M(q)} \int |\boldsymbol{R}_0^{-1}|^{L+q-1} e^{-\text{tr}\{\boldsymbol{R}_0^{-1}(\boldsymbol{Q}+\tau \boldsymbol{I}_{MN})\}} d\boldsymbol{R}_0^{-1} \\ &= \frac{\tau^{Mq}}{\pi^{MNL}\Gamma^M(q)} \prod_{i=1}^{M} \int |\boldsymbol{R}_{ii}^{-1}|^{L+q-1} e^{-\text{tr}\{\boldsymbol{R}_{ii}^{-1}(\boldsymbol{Q}_{ii}+\tau \boldsymbol{I}_N)\}} d\boldsymbol{R}_{ii}^{-1} \\ &= \frac{\tau^{Mq}\Gamma^M(L+q)}{\pi^{MNL}\Gamma^M(q)\prod_{i=1}^{M}|\boldsymbol{Q}_{ii}+\tau\boldsymbol{I}_N|^{L+q}}. \end{aligned} \tag{41}$$

Similarly, using the prior $p(\boldsymbol{R}_1^{-1})$ in (18) and marginalizing with respect to $\boldsymbol{R}_1^{-1}$, we obtain the marginalized likelihood for H_1 as

$$p(\boldsymbol{Z}|H_1) = \frac{\tau^{Mq}\prod_{i=1}^{M}\Gamma(L+q-M+i)}{\pi^{MNL}\prod_{i=1}^{M}(q-M+1)|\boldsymbol{Q}+\tau\boldsymbol{I}_{MN}|^{(L+q)}}. \tag{42}$$

As $\tau \to 0$, the likelihood ratio becomes

$$\frac{p(\boldsymbol{Z}|H_1)}{p(\boldsymbol{Z}|H_0)} \to \frac{k(L+q, M)}{k(q, M)} \gamma^{L+q}. \tag{43}$$

This is a monotonic function of the statistic

$$\gamma = \frac{\prod_{i=1}^{M}|\boldsymbol{Q}_{ii}|}{|\boldsymbol{Q}|}, \tag{44}$$

as is the GLRT detector given in [1]. Finally, the Bayesian detection thresholds can also be set as described in Section 5; i.e.,

$$\gamma > \left(\frac{\beta k(q, M)}{k(L+q, M)} \right)^{\frac{1}{L+q}}. \tag{45}$$

7. REFERENCES

[1] D. Ramírez, J. Vía, I. Santamaría, and L. L. Scharf, "Detection of spatially correlated Gaussian time series," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5006–5015, 2010.

[2] G. C. Carter, "Coherence and time delay estimation," *Proc. IEEE*, vol. 75, no. 2, pp. 236–255, 1987.

[3] A. H. Nuttall, "Invariance of distribution of coherence estimate to second-channel statistics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 2, pp. 120–122, 1981.

[4] H. Gish and D. Cochran, "Invariance of the magnitude-squared coherence estimate with respect to second-channel statistics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 12, pp. 1774–1776, 1987.

[5] R. D. Trueblood and D. L. Alspach, "Multiple coherence as a detection statistic," Naval Ocean Systems Center, Tech. Rep. NOSC 265, 1978.

[6] H. Gish and D. Cochran, "Generalized coherence," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing,*, Apr. 1988, pp. 2745–2748.

[7] A. Clausen and D. Cochran, "An invariance property of the generalized coherence estimate," *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 1065–1067, 1997.

[8] ——, "Asymptotic analysis of the generalized coherence estimate," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 45–53, 2001.

[9] D. Cochran, H. Gish, and D. Sinno, "A geometric approach to multiple-channel signal detection," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2049–2057, 1995.

[10] S. T. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, 2005.

[11] R. E. Kass and L. Wasserman, "Formal rules of selecting prior distributions: A review and annotated bibliography," *J. Amer. Statist. Assoc.*, vol. 91, pp. 1343–1370, 1996.

[12] H. Jeffreys, *Theory of Probability*. Oxford University Press, 1961.

[13] E. T. Jaynes, "Information theory and statistical mechanics," *The Physical Review*, vol. 106, no. 4, pp. 620–630, 1957.