MAXIMUM-ENTROPY SURROGATION IN NETWORK SIGNAL DETECTION

D. Cochran[1], *S. D. Howard*[2], *B. Moran*[3], *H. A. Schmitt*[4]

[1]Arizona State University, Tempe AZ, USA
[2]Defence Science & Technology Organisation, Edinburgh SA, Australia
[3]Defence Science Institute, Parkville VIC, Australia
[4]Physical Layer Systems LLC, Tucson AZ, USA

ABSTRACT

Multiple-channel detection is considered in the context of a sensor network where raw data are shared only by nodes that have a common edge in the network graph. Established multiple-channel detectors, such as those based on generalized coherence or multiple coherence, use pairwise measurements from every pair of sensors in the network and are thus directly applicable only to networks whose graphs are completely connected. An approach is introduced that uses a maximum-entropy technique to formulate surrogate values for missing measurements corresponding to pairs of nodes that do not share an edge in the network graph. The broader potential merit of maximum-entropy baselines in quantifying the value of information in sensor network applications is also noted.

Index Terms— Sensor networks, Multiple-channel detection, Generalized coherence, Maximum entropy, Value of information

1. INTRODUCTION

Established methods in coherent multiple-channel signal detection typically assume that data from all sensors are collected at a single location for processing. In particular, this view is implicit in the generalized coherence (GC) approach introduced in [1] and elaborated and extended in numerous other works (*e.g.*, [2, 3, 4, 5, 6]). In signal processing for sensor networks, it is frequently desirable to process data locally at (or in local neighborhoods of) the nodes and reduce, or even eliminate, the need for aggregation of data at a "fusion center."

In GC-based detection and related methods, such as those using multiple coherence [7], processing entails computing inner products (correlations) between segments of time series data collected at each pair of nodes in the network. When all data are collected at a fusion center, this is not an issue. This

paper proposes an approach for implementing a GC detector that incorporates a high degree of local data reduction by using only inner products of time series segments collected at pairs of nodes that are adjacent in the topology of the network. This is equivalent to traditional GC detection only when the network is fully connected. Otherwise, the detector must operate without using the inner products associated with pairs of nodes that do not share an edge in the network graph.

The approach introduced here replaces each missing inner product datum by a surrogate value obtained via a maximum entropy method, and then proceeds to apply a standard GC detector as though all data were available. The paper begins in Section 2 by summarizing the basics of GC detection, and proceeds in Section 3 to describe the maximum entropy procedure that allows surrogation for missing data in networks that are not fully connected. Section 4 shows simulation results for small networks that illustrate the performance of the approach in such settings. The paper concludes in Section 5 with a discussion of further work needed to make this technique viable for larger sensor networks and also how this approach suggests a point of view about quantifying the value of information in network signal processing.

2. THE GENERALIZED COHERENCE DETECTOR

The GC estimate is an established statistic for detection of a common signal on several noisy channels [2]. Its properties and applications have been well documented, and it continues to be extended and studied in various contexts [5, 6]. A crucial drawback of GC-based methods in the sensor network setting is that standard implementations require all the raw sensor data to be collected in one place (*i.e.*, a "fusion center") to perform the processing.

Given M measurements $X_1, \ldots, X_M$ at the nodes of a network with each $X_m \in \mathbb{C}^N$, define normalized measurement vectors $U_m = X_m / \|X_m\|$ for $m = 1, \ldots, M$. The GC estimate obtained from these measurements is $\hat{\gamma}^2(X_1, \ldots, X_M) = 1 - \det G(U_1, \ldots, U_M)$, where the

matrix $G(U_1, \ldots, U_M)$ is the $M \times M$ Gram matrix

$$\begin{bmatrix} 1 & \langle U_1, U_2 \rangle & \cdots & \langle U_1, U_M \rangle \\ \langle U_2, U_1 \rangle & 1 & \cdots & \langle U_2, U_M \rangle \\ \vdots & & \ddots & \vdots \\ \langle U_M, U_1 \rangle & \cdots & \langle U_M, U_{M-1} \rangle & 1 \end{bmatrix} \quad (1)$$

formed from the normalized data vectors. In typical multiple-channel detection applications, the value of $\hat{\gamma}^2$ is compared to a threshold to decide between signal-present (H_1) and signal-absent (H_0) hypotheses.

In a completely connected network, such as those depicted in Figures 1(a) and 2(a), all of the inner products comprising the Gram matrix may be computed locally; *i.e.*, by exchange of data between nodes that share an edge in the network graph. This enables dramatic reduction in the amount of data necessary to communicate to the fusion center in order to implement a GC detector for the network. When the network graph is not complete, however, inner products of data vectors corresponding to nodes that do not share an edge cannot be computed locally, and transmission of these scalar values corresponding to each edge in the network graph is not sufficient to enable the fusion center to compute the GC estimate.
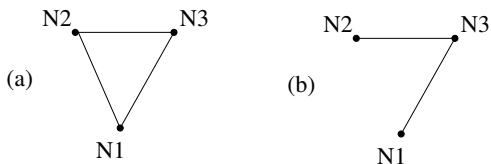
Fig. 1. (a) A completely connected three-node network. (b) A three-node network with no edge between nodes one and two. Every three-node network with two edges has the same topology (*i.e.*, linear).

Considering the three-node case depicted in Figure 1(b), note that the value of $\langle U_1, U_2 \rangle$ is generally not determined by the values of $\langle U_1, U_3 \rangle$ and $\langle U_2, U_3 \rangle$. But it is not independent of these values either; *e.g.*, G must be non-negative definite. In the approach described in Sec. 3, the values of the "missing" inner products (*i.e.*, those corresponding to pairs of nodes that do not share an edge) will be replaced by surrogate values obtained via a maximum entropy technique, thereby enabling a GC test to be performed despite the missing data. Sec. 4 considers the detection performance of GC detectors implemented in this way.

3. MAXIMUM-ENTROPY SURROGATION

Assume that there is a complex random variable $\mathbf{x}_m$ associated with each network node m, modeling data samples collected at that node. Collected samples at node m are hence

realizations of $\mathbf{x}_m$ and can be used to estimated the mean and variance of $\mathbf{x}_m$ in standard ways [8]. The ability to communicate between nodes m and j linked by an edge permits estimation of the covariance $\text{cov}(\mathbf{x}_m, \mathbf{x}_j)$ by similar methods. For a complete graph (*i.e.*, one in which each pair of nodes shares an edge), it is thus possible to estimate the full $M \times M$ covariance matrix C of the variables $\mathbf{x}_1, \ldots, \mathbf{x}_M$. Indeed, assuming the X_m have mean zero, if N independent samples of $\mathbf{x}_m$ are collected at node m for each m and are assimilated into complex N-vectors $\mathbf{X}_1, \ldots, \mathbf{X}_M$ of sample values, then the standard estimator $\widehat{C}$ of C is proportional to the Gram matrix $G(\mathbf{X}_1, \ldots, \mathbf{X}_M)$. The GC detector can thus be viewed as a test on the estimated covariance matrix of the variates $\mathbf{x}_1, \ldots, \mathbf{x}_M$. If each $\mathbf{x}_m$ is further assumed to be normalized to unit variance, using the true variance in place of an estimate on the main diagonal gives a test matrix of the form (1).

The maximum-entropy method [9] holds that missing values in C should be surrogated in such a way as to introduce no new assumptions about the nature of the random variables or of the network. The joint distribution of the random variables $\mathbf{x}_1, \ldots, \mathbf{x}_M$ that best describes current knowledge (*i.e.*, the covariance estimates for all pairs of directly connected nodes) with no further assumptions is the maximum entropy distribution constrained by the available data.

The problem of finding the maximum-entropy completion of a covariance matrix has been studied in prior literature (see, *e.g.*, [10] and references cited therein). The maximum-entropy probability density $p(x_1, \ldots, x_M)$ consistent with the estimated covariances must be of the form

$$\frac{1}{Z} \exp \left\{ -\left(\sum_{m=1}^{M} \lambda_m x_m^* + \frac{1}{2} \sum_{(m,j) \in \mathcal{E}} \mu_{m,j} x_m x_j \right) \right\} \quad (2)$$

In this expression λ_m for $m = 1, \ldots, M$ and $\mu_{m,j} = \mu_{j,m}$ for values of (m, j) corresponding to the edge set $\mathcal{E}$ of the network graph are Lagrange parameters in the constrained optimization problem that arises in maximization of entropy subject to the constraints imposed by knowledge of the covariance values estimated from available measurements. For complex random variables, this will be a complex normal density and hence completely specified by its mean and covariance matrix.

It follows from (2), and is also noted in past literature [10, Sec. 2.2], that the covariance matrix of this maximum-entropy distribution will have the property that its inverse will has zeros in positions corresponding to the missing covariance values. This observation gives a direct means for calculating the needed surrogate values in small networks.

Example: Consider the three node network depicted in Fig. 1(b). Writing the estimated covariance matrix as

$$\hat{C} = \begin{bmatrix} 1 & s & \hat{a} \\ s^* & 1 & \hat{b} \\ \hat{a}^* & \hat{b}^* & 1 \end{bmatrix}$$

the value of the covariance estimate $\hat{a}$ is assumed to be obtained by exchange of data between nodes 1 and 3 and the value $\hat{b}$ is obtained by exchange of data between nodes 2 and 3. A maximum-entropy surrogate value for s is obtained by noting

$$\hat{C}^{-1} = \frac{1}{D} \left[\begin{array}{ccc} 1 - |\hat{b}|^2 & \hat{a}\hat{b}^* - s & s\hat{b} - \hat{a} \\ \hat{a}^*\hat{b} - s^* & 1 - |\hat{a}|^2 & s^*\hat{a} - \hat{b} \\ s^*\hat{b}^* - \hat{a}^* & s\hat{a}^* - \hat{b}^* & 1 - |s|^2 \end{array} \right]$$

where $D = \det \hat{C}$. The 1-2 entry of $\hat{C}^{-1}$ will be zero when s assumes the desired value. Hence, $s = \hat{a}\hat{b}^*$ and the maximum-entropy completion of $\hat{C}$ is

$$\hat{C} = \left[\begin{array}{ccc} 1 & \hat{a}\hat{b}^* & \hat{a} \\ \hat{a}^*\hat{b} & 1 & \hat{b} \\ \hat{a}^* & \hat{b}^* & 1 \end{array} \right]$$

This direct calculation method was used to find surrogate values in the small network examples discussed above and evaluated in Sec. 4. In general, a necessary and sufficient condition for the existence of the maximum entropy distribution as a (normalizable) probability distribution is $\hat{C}$ is invertible as a symbolic matrix. Finding k surrogate values in this way requires solving k equations in k unknowns, which becomes prohibitively cumbersome for even small networks (*e.g.*, $M > 5$). Fortunately, maximum-entropy covariance matrix completion problems [11] fall into a class of determinant maximization problems (entropy in this setting is $\log \det \hat{C}$) that can be efficiently solved by convex programming techniques [12, 10].

The GC statistic is invariant to re-indexing of the M channels [2]. This is in contrast to the multiple coherence statistic [7], which distinguishes a reference channel. Consequently, networks that are topologically isomorphic will have identical detection performance characteristics under the approach introduced here. For example, for a network having four nodes and five edges, it suffices to analyze the case depicted in Fig. 2(b) since all other cases are topologically equivalent to this one (and similarly for all M-node networks whose graphs are one edge short of being complete). Fig. 3 shows that there are topologically distinct cases of four-node networks with four edges. In general, it will be necessary to analyze the performance of each such equivalence class separately, and it is anticipated that different topologies will provide distinct detection performance characteristics. While further study of this aspect of the approach is beyond the scope of this paper, it is expected that comparison of network topologies based on the detection performance they support when completed via a maximum-entropy method will provide insight into sensor network design and optimization.

4. RESULTS FOR SMALL NETWORKS OF SENSORS

Figure 4 shows receiver operating characteristic (ROC) curves for detection of a white complex Gaussian signal
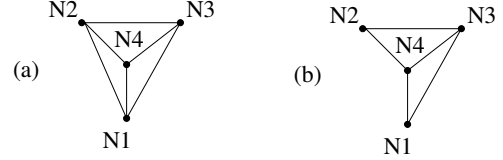
Fig. 2. (a) A completely connected four-node network. (b) A four-node network with no edge between nodes one and two. Every four-node network with five edges has the same topology.
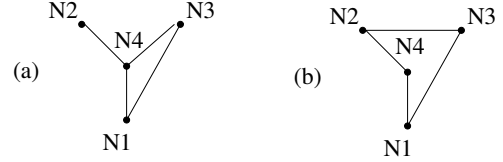

Fig. 3. Two four-node networks with no edge between nodes one and two. (a) This network also lacks a 2-3 edge. (b) This one lacks edge 3-4 as well as 1-2. Every connected four-node network with four edges has one of these two topologies.

vector in white complex Gaussian noise in a 4-node sensor network. The vectors are of length $N = 64$. The signal-to-noise ratio is identical at each sensor, with the top set of curves at -3 dB, the center set at -4.5 dB, and the bottom set at -6 dB. Within each set, the top curve is for the complete network (Figure 2(a)), the middle curve for the network with no link between nodes 1 and 2 (Figure 2(b)), and the bottom curve for the network with no links between nodes 1 and 2 or nodes 2 and 3 (Figure 3(a)).

These curves indicate that the equal-channel SNR detection performance lost when network connectivity is reduced by the removal of one or two links is modest – much less significant than the detection performance would be diminished by 1 dB of SNR at each node. While this experimental finding with such a small network is only relevant to a small application regime, it indicates that further study with larger networks is warranted.

5. DISCUSSION AND CONCLUSIONS

A maximum entropy method to enable the implementation of generalized coherence detectors on incompletely connected sensor networks without aggregation of raw data at a fusion center has been described and demonstrated with small networks. Performance degradation in the cases studied was modest, though more compete evaluation of the method with larger and sparser networks, unequal signal-to-noise ratios at the sensor nodes, and signals of rank greater than one is essential to establish a comprehensive understanding of this approach in detection of signals spread across a sensor network.
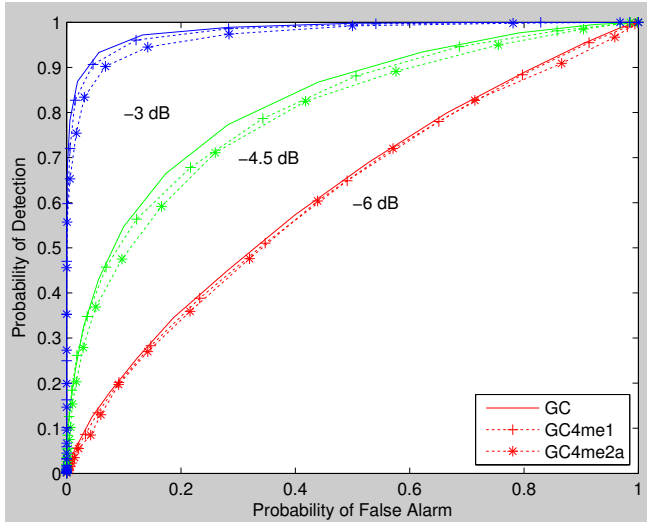
Fig. 4. ROC curves for detection of a 64-component white complex Gaussian signal vector in white complex Gaussian noise in a 4-node sensor network with identical SNR at each sensor. The top set of curves is at -3 dB, the center set at -4.5 dB, and the bottom set at -6 dB. Within each set, the top curve is for the complete network (Figure 2(a)), the middle curve for the network with no link between nodes 1 and 2 (Figure 2(b)), and the bottom curve for the network with no links between nodes 1 and 2 or nodes 2 and 3 (Figure 3(a)).

In ongoing work, known numerical methods for covariance matrix completion (*i.e.,* via optimization algorithms) are being used to enable application of this approach to sensor networks of moderate size (*e.g.*, 10–20 nodes). This appears necessary because direct calculation of maximum-entropy surrogate values becomes formidable even for smaller numbers of nodes.

It is noteworthy that the use of maximum entropy baselines in this application provides a mechanism for quantifying the value of information sharing within the sensor network; *i.e.*, in this setting, a link is precisely as valuable as the performance gain it enables over the use of a maximum entropy surrogate in place of its datum.

6. REFERENCES

[1] Herbert Gish and Douglas Cochran, "Generalized coherence," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1987.

[2] Douglas Cochran, Herbert Gish, and Dana Sinno, "A geometric approach to multiple-channel signal detection," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2049–2057, September 1995.

[3] Axel Clausen and Douglas Cochran, "An invariance property of the generalized coherence estimate," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 1065–1067, April 1997.

[4] Axel Clausen and Douglas Cochran, "Asymptotic analysis of the generalized coherence estimate," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 45–53, January 2001.

[5] David Ramírez, Javier Vía, Ignacio Santamaría, and Louis L. Scharf, "Detection of spatially correlated Gaussian time series," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5006–5015, October 2010.

[6] Songsri Sirianunpiboon, Stephen D. Howard, and Douglas Cochran, "A Bayesian derivation of the generalized coherence detector," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2012.

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

[8] Athanasios Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 4th edition, 2002.

[9] Edwin T. Jaynes, "Where do we stand on maximum entropy?," in *The Maximum Entropy Formalism*, R. D. Levine and M. Tribus, Eds., pp. 15–104. MIT Press, 1979.

[10] Lieven Vandenberghe, Stephen Boyd, and Shao-Po Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 2, pp. 499–533, 1998.

[11] Israel Gohberg, Marinus A. Kaashoek, and Frederik van Schagen, *Partially Specified Matrices and Operators: Classification, Completion, Applications*, Birkhäuser, 1995.

[12] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.