ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3315-6701-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICASSP55912.2026.11463389

FIRST-ORDER AND SECOND-ORDER DETECTORS FOR MATCHED SUBSPACE DETECTION ON GRAPHS

David Ramírez[1,2]*, Chengen Liu*[3]*, Victor M. Tenorio*[4]*, Elvin Isufi*[3]*, and Antonio G. Marques*[4]

[1] Dept. of Signal Theory and Communications, Universidad Carlos III de Madrid, Spain
[2] Gregorio Marañón Health Research Institute, Madrid, Spain
[3] Dept. of Intelligent Systems, Delft University of Technology, The Netherlands
[4] Dept. of Signal Theory and Communications, Universidad Rey Juan Carlos, Spain

ABSTRACT

Matched subspace detection (MSD) is a powerful tool recently generalized from Euclidean data to graph signal processing. However, existing graph-based MSD methods are often limited by assumptions of known noise variance and by overlooking the statistical properties of the graph Fourier transform (GFT) coefficients thereby limiting practical applicability. To address these gaps, this paper introduces two novel generalized likelihood ratio (GLR) tests for graph-based MSD. The first-order GLR test operates without knowledge of the noise variance and the GFT coefficients by estimating them via maximum likelihood. The second-order GLR test further incorporates a Gaussian prior on the GFT coefficients, yielding a more powerful and comprehensive statistical model. Experimental results demonstrate that our proposed detectors are robust and effective, particularly in challenging noisy scenarios, highlighting their importance for detection tasks in graph signal processing.

***Index Terms*—** Generalized likelihood ratio test, graph signal processing, matched subspace detection, maximum likelihood estimation

1. INTRODUCTION

Matched subspace detection (MSD) is a fundamental signal processing tool developed to determine whether a signal resides within a specific subspace [1]. Its effectiveness has led to widespread adoption in diverse fields, such as medical imaging [2], hyperspectral imaging [3], and communications [4].

Traditionally, MSD has been applied to data with an underlying Euclidean structure, such as images or space-time series. However, the irregular structure inherent in much real-world data is better captured using graphs [5], for instance for brain activity patterns [6], social networks [7], or sensor measurements [8]. Consequently, there is an increasing interest in adapting MSD for graph-based tasks. For instance, [9] formulated an MSD problem to identify the underlying graph topology for smooth or bandlimited signals, while [10] developed detectors to select which of two candidate graph topologies best fits an observed signal. Other notable work includes a blind detector for graph topology changes [11] and an extension of MSD to higher-order topological structures [12].

Despite these advances, existing methods suffer from key limitations that hinder their practical deployment. First, they typically assume that the noise variance is known, a condition rarely met in real-world applications. Second, they often neglect potential prior statistical information about the graph Fourier transform (GFT) coefficients, missing an opportunity to enhance detection performance.

To address these limitations, we propose two novel generalized likelihood ratio (GLR) tests for the graph topology MSD task. The first-order GLR test tackles the issue of unknown noise by estimating its variance via maximum likelihood estimation (MLE). Contrary to the first-order GLR that assumes deterministic, but unknown, GFT coefficients, the second-order GLR test incorporates a Gaussian prior on the GFT coefficients, creating a more powerful statistical model that accounts for both unknown noise and signal characteristics. Our experimental results verify that these methods are effective and robust, confirming their value in practical applications.

2. PROBLEM FORMULATION AND PREVIOUS APPROACHES

This work addresses the problem of selecting the true underlying graph structure for a signal from a pair of candidate graphs, where the assumption is that the observed signal is a low-pass graph signal corrupted by additive white Gaussian noise.

To properly define low-pass graph signals, we shall now review basic GSP concepts and introduce the required notation. For more details, we refer the reader to [13, 14]. We consider a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of N nodes and $\mathcal{E}$ is the set of links, such that the pair (i, j) belongs to $\mathcal{E}$ if there exists a link from node i to node j. The structure of the graph is captured by the adjacency matrix, $\mathbf{A} \in \mathbb{R}^{N\times N}$, which is a sparse matrix with the non-zero entries associated to the elements in $\mathcal{E}$. Moreover, for a given $\mathcal{G}$, we can define a graph signal as the vector $\mathbf{x} = [x_1, \ldots, x_N]^T \in \mathbb{R}^N$, with x_i representing the signal value at node i.

The graph Laplacian is [13]

$$\mathbf{L} = \operatorname{diag}(\mathbf{A1}) - \mathbf{A} = \mathbf{U} \operatorname{diag}(\lambda_1, \ldots, \lambda_N)\mathbf{U}^T,$$

The work of D. Ramírez was partially supported by MICIU/AEI/10.13039/501100011033/FEDER, UE, under grant PID2021-123182OB-I00 (EPiCENTER) and PID2024-157856NB-I00 (CARTESIAN), by the Office of Naval Research (ONR) Global under contract N62909-23-1-2002, and by the Comunidad de Madrid under grant TEC-2024/COM-89 (IDEA-CM). The work of E. Isufi was partially supported by the TU Delft AI Labs programme, the NWO OTP GraSPA proposal #19497, and the NWO VENI proposal 222.032. The work of V. Tenorio and A. Marques was partially supported by the Spanish AEI (AEI/10.13039/501100011033) under grants PID2022-136887NB-I00 and PID2023-149457OB-I00, and FPU20/05554, by the Community of Madrid via IDEA-CM (TEC-2024/COM-89) and the Ellis Madrid Unit, and by the EU H2020 Grant Tailor (No 952215, agreements 76 and 82). C. Liu receives funding from the China Scholarship Council.

where $\lambda_i, i = 1, \ldots, N$, with $\lambda_i \leq \lambda_{i+1}$, are the graph frequencies. The eigenvectors of the graph Laplacian can be used to define the GFT and inverse GFT, which are given by

$$\tilde{\mathbf{x}} = \mathbf{U}^T\mathbf{x}, \qquad \mathbf{x} = \mathbf{U}\tilde{\mathbf{x}},$$

respectively. Hence, a low-pass graph signal of bandwidth F is defined as $\mathbf{x} = \mathbf{U}_{\mathcal{F}}\tilde{\mathbf{x}}_{\mathcal{F}}$, where $\mathbf{U} = \left[\mathbf{U}_{\mathcal{F}}\ \mathbf{U}_{\bar{\mathcal{F}}}\right]$, with $\mathbf{U}_{\mathcal{F}} \in \mathbb{R}^{N\times F}$, $\mathbf{U}_{\bar{\mathcal{F}}} \in \mathbb{R}^{N\times(N-F)}$, and $\tilde{\mathbf{x}}_{\mathcal{F}}$ are the GFT coefficients. Note that $\mathbf{U}_{\mathcal{F}}$ and $\mathbf{U}_{\bar{\mathcal{F}}}$ are orthonormal bases for orthogonal subspaces, i.e., $\mathbf{U}_{\mathcal{F}}^T\mathbf{U}_{\mathcal{F}} = \mathbf{I}_F$, $\mathbf{U}_{\bar{\mathcal{F}}}^T\mathbf{U}_{\bar{\mathcal{F}}} = \mathbf{I}_{N-F}$, and $\mathbf{U}_{\mathcal{F}}^T\mathbf{U}_{\bar{\mathcal{F}}} = \mathbf{0}$.

Based on this signal model, the question of whether an observed signal has been generated from graph $\mathcal{G}_0$ or $\mathcal{G}_1$ can be formulated as

$$\begin{aligned}\mathcal{H}_1 &: \mathbf{y} = \mathbf{U}_{\mathcal{F},1}\tilde{\mathbf{x}}_{\mathcal{F},1} + \mathbf{n},\\ \mathcal{H}_0 &: \mathbf{y} = \mathbf{U}_{\mathcal{F},0}\tilde{\mathbf{x}}_{\mathcal{F},0} + \mathbf{n},\end{aligned} \tag{1}$$

where $\mathbf{n} \sim \mathcal{N}_N(\mathbf{0}, \sigma^2\mathbf{I}_N)$, with unknown σ^2, is additive white Gaussian noise, $\mathbf{U}_{\mathcal{F},i}$ contains the eigenvectors associated with the F smallest eigenvalues of the known graph Laplacian $\mathbf{L}_i$, corresponding to graph $\mathcal{G}_i$, and $\tilde{\mathbf{x}}_{\mathcal{F},i}$ denotes the unknown GFT in $\mathcal{G}_i$ of the signal.

The detection problem in (1) has been previously addressed in [9]. Concretely, without prior information, the GLR test derived in that work is [9, Equation (17)]

$$T_{\text{SMSD}} = \|\tilde{\mathbf{y}}_{\mathcal{F},0}\|^2 - \|\tilde{\mathbf{y}}_{\mathcal{F},1}\|^2 = \|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},1}\|^2 - \|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},0}\|^2 \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma,$$

where $\tilde{\mathbf{y}}_{\mathcal{F},i} = \mathbf{U}_{\mathcal{F},i}^T\mathbf{y}$ and $\tilde{\mathbf{y}}_{\bar{\mathcal{F}},i} = \mathbf{U}_{\bar{\mathcal{F}},i}^T\mathbf{y}$. Nevertheless, as we will show in the next section, T_{SMSD} is not the GLR for unknown noise variance, but for known variance . This is easy to verify since, for unknown σ^2 and $\tilde{\mathbf{x}}_{\mathcal{F},i}$, the detection problem in (1) is invariant to scalings, and T_{SMSD} is not.

3. DERIVATION OF THE GLRS

This section derives the GLR tests for the hypothesis testing problem formulated in (1). Although GLR tests are not generally Uniformly Most Powerful (UMP), they are widely adopted for their mathematical tractability and robust performance [15]. We consider two distinct modeling assumptions for the coefficient vectors $\{\tilde{\mathbf{x}}_{\mathcal{F},i}\}_{i\in\{0,1\}}$: first, that they are deterministic unknown parameters; and second, that they are random vectors governed by Gaussian prior distributions. Following the nomenclature in [16], we designate these frameworks as first-order and second-order GLRs, respectively.

3.1. First-order GLR

For deterministic GFT coefficients and Gaussian noise, (1) can be more formally formulated as

$$\begin{aligned}\mathcal{H}_1 &: \mathbf{y} \sim \mathcal{N}_N(\mathbf{U}_{\mathcal{F},1}\tilde{\mathbf{x}}_{\mathcal{F},1}, \sigma^2\mathbf{I}_N),\\ \mathcal{H}_0 &: \mathbf{y} \sim \mathcal{N}_N(\mathbf{U}_{\mathcal{F},0}\tilde{\mathbf{x}}_{\mathcal{F},0}, \sigma^2\mathbf{I}_N),\end{aligned} \tag{2}$$

which is similar to the MSD problem in [1]; the only difference being that in [1] the null hypothesis contains only noise. The GLR test for (2) is

$$\Lambda_1 = \frac{\max\limits_{\tilde{\mathbf{x}}_{\mathcal{F},1},\sigma^2} \ell(\tilde{\mathbf{x}}_{\mathcal{F},1}, \sigma^2; \mathbf{y})}{\max\limits_{\tilde{\mathbf{x}}_{\mathcal{F},0},\sigma^2} \ell(\tilde{\mathbf{x}}_{\mathcal{F},0}, \sigma^2; \mathbf{y})} = \frac{\ell(\hat{\tilde{\mathbf{x}}}_{\mathcal{F},1}, \hat{\sigma}_1^2; \mathbf{y})}{\ell(\hat{\tilde{\mathbf{x}}}_{\mathcal{F},0}, \hat{\sigma}_0^2; \mathbf{y})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{3}$$

where $\hat{\tilde{\mathbf{x}}}_{\mathcal{F},i}$ and $\hat{\sigma}_i^2$ are the maximum likelihood (ML) estimates of the parameters under hypothesis i and $\ell(\tilde{\mathbf{x}}_{\mathcal{F},i}, \sigma^2; \mathbf{y})$ is the corresponding likelihood given by

$$\ell(\tilde{\mathbf{x}}_{\mathcal{F},i}, \sigma^2; \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{U}_{\mathcal{F},i}\tilde{\mathbf{x}}_{\mathcal{F},i}\|^2\right\}.$$

Let us first obtain the ML estimate of $\tilde{\mathbf{x}}_{\mathcal{F},i}$ by setting to zero the derivative of $\log \ell(\tilde{\mathbf{x}}_{\mathcal{F},i}, \sigma^2; \mathbf{y})$ with respect to $\tilde{\mathbf{x}}_{\mathcal{F},i}$, which yields $\hat{\tilde{\mathbf{x}}}_{\mathcal{F},i} = \mathbf{U}_{\mathcal{F},i}^T\mathbf{y}$. Plugging now $\hat{\tilde{\mathbf{x}}}_{\mathcal{F},i}$ back into the log-likelihood, and maximizing this compressed log-likelihood, the ML estimate of σ^2 becomes $\hat{\sigma}_i^2 = \frac{1}{N}\mathbf{y}^T\mathbf{P}_{\mathcal{F},i}^{\perp}\mathbf{y}$, where $\mathbf{P}_{\mathcal{F},i} = \mathbf{U}_{\mathcal{F},i}\mathbf{U}_{\mathcal{F},i}^T$ is the orthogonal projector onto the subspace spanned by the columns of $\mathbf{U}_{\mathcal{F},i}$ and $\mathbf{P}_{\mathcal{F},i}^{\perp} = \mathbf{I}_N - \mathbf{P}_{\mathcal{F},i}$ is the complementary projector.

Taking into account these ML estimates, the GLR test in (3) becomes

$$\Lambda_1^{1/N} = \frac{\mathbf{y}^T\mathbf{P}_{\bar{\mathcal{F}},0}^{\perp}\mathbf{y}}{\mathbf{y}^T\mathbf{P}_{\bar{\mathcal{F}},1}^{\perp}\mathbf{y}} = \frac{\|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},0}\|^2}{\|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},1}\|^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma,$$

where we have used $\mathbf{P}_{\bar{\mathcal{F}},i}^{\perp} = \mathbf{U}_{\bar{\mathcal{F}},i}\mathbf{U}_{\bar{\mathcal{F}},i}^T$. This shows that Λ_1 and T_{SMSD} are both functions of $\|\hat{\mathbf{y}}_{\bar{\mathcal{F}},0}\|^2$ and $\|\hat{\mathbf{y}}_{\bar{\mathcal{F}},1}\|^2$ but the functions that fuse these energies are different. In fact, it could be shown that if in the previous steps we did not estimate σ^2 and consider it known, we obtain T_{SMSD}. Nevertheless, although T_{SMSD} does not depend on σ^2, its distribution does. Further comments will be made about this in the following section.

3.2. Second-order GLR

This section considers that $\tilde{\mathbf{x}}_{\mathcal{F},i}$ is an unknown stochastic signal distributed as $\tilde{\mathbf{x}}_{\mathcal{F},i} \sim \mathcal{N}_F(\mathbf{0}, \mathbf{D}_i)$, with $\mathbf{D}_i = \text{diag}(d_{i,1}, \ldots, d_{i,F})$ being its covariance matrix. Hence, the detection problem in (1) can be formulated as

$$\begin{aligned}\mathcal{H}_1 &: \mathbf{y} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{R}_1),\\ \mathcal{H}_0 &: \mathbf{y} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{R}_0),\end{aligned} \tag{4}$$

where $\mathbf{R}_i = \mathbf{U}_{\mathcal{F},i}\mathbf{D}_i\mathbf{U}_{\mathcal{F},i}^T + \sigma^2\mathbf{I}_N$ is the covariance matrix under hypothesis i, with $\mathbf{D}_i$ and σ^2 unknown.

The GLR test for the second-order detection problem in (4) is

$$\Lambda_2 = \frac{\max\limits_{\mathbf{D}_1,\sigma^2} \ell(\mathbf{D}_1, \sigma^2; \mathbf{y})}{\max\limits_{\mathbf{D}_0,\sigma^2} \ell(\mathbf{D}_0, \sigma^2; \mathbf{y})} = \frac{\ell(\hat{\mathbf{D}}_1, \hat{\sigma}^2; \mathbf{y})}{\ell(\hat{\mathbf{D}}_0, \hat{\sigma}^2; \mathbf{y})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma,$$

where $\ell(\mathbf{D}_i, \sigma^2; \mathbf{y})$ is the likelihood under the ith hypothesis given by

$$\ell(\mathbf{D}_i, \sigma^2; \mathbf{y}) = \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{R}_i)}} \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{R}_i^{-1}\mathbf{y}\right).$$

Computing the inverse and determinant of the covariance matrix, the log-likelihood becomes

$$\begin{aligned}\log \ell(\mathbf{D}_i, \sigma^2; \mathbf{y}) = &-\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{n=1}^{F}\log(\bar{d}_{i,n} + 1)\\ &- \frac{1}{2\sigma^2}\left(\sum_{n=1}^{F}\frac{\left|\tilde{y}_{\mathcal{F},i}^{(n)}\right|^2}{\bar{d}_{i,n} + 1} + \|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},i}\|^2\right),\end{aligned} \tag{5}$$

where $\tilde{y}_{\mathcal{F},i}^{(n)}$ is the nth component of $\tilde{\mathbf{y}}_{\mathcal{F},i}$ and $\bar{d}_{i,n} = d_{i,n}/\sigma^2$. Then, we will obtain the compressed likelihoods by estimating $\bar{d}_{i,n}$ and σ^2, instead of $d_{i,n}$ and σ^2.

Now, we will follow an approach similar to [17] to obtain the ML estimates of $\bar{d}_{i,n}$ and σ^2. First, setting the derivative of (5) with respect to σ^2 to zero yields

$$\hat{\sigma}^2 = \frac{1}{N}\left(\sum_{n=1}^{F}\frac{\left|\tilde{y}_{\mathcal{F},i}^{(n)}\right|^2}{\bar{d}_{i,n}+1} + \|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},i}\|^2\right). \tag{6}$$

Typically, we would plug $\hat{\sigma}^2$ into (5) to obtain the compressed log-likelihood and maximize it with respect to the remaining parameters $\{\bar{d}_{i,n}\}_{n=1}^{F}$. This approach, however, yields a complicated nonlinear function that cannot be optimized in closed form. To avoid this issue, we will optimize (5) with respect to $\bar{d}_{i,n}$ to obtain

$$\hat{\bar{d}}_{i,n} = \max\left(\frac{\left|\tilde{y}_{\mathcal{F},i}^{(n)}\right|^2}{\sigma^2} - 1, 0\right). \tag{7}$$

The ML estimates in (6) and (7) are coupled: (6) depends on $\bar{d}_{i,n}$ and (7) depends on σ^2. Then, substitute (7) into (6) to get

$$N\hat{\sigma}^2 = \sum_{n=1}^{F}\min\left(\left|\tilde{y}_{\mathcal{F},i}^{(n)}\right|^2, \hat{\sigma}^2\right) + \|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},i}\|^2, \tag{8}$$

which depends now only on $\hat{\sigma}^2$, but in a non-linear fashion with no closed-form solution. However, similar to [17], we can propose a simple algorithm to solve (8). Concretely, [17] proposes to rewrite (8) as $f_1(\hat{\sigma}^2) = f_2(\hat{\sigma}^2)$, where

$$f_1(\hat{\sigma}^2) = N\hat{\sigma}^2 - \|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},i}\|^2,$$
$$f_2(\hat{\sigma}^2) = \sum_{n=1}^{F}\min\left(\left|\tilde{y}_{\mathcal{F},i}^{(n)}\right|^2, \hat{\sigma}^2\right).$$

Thus, $\hat{\sigma}^2$ is obtained as the intersection between an affine function, $f_1(\hat{\sigma}^2)$, and a piecewise-linear one, $f_2(\hat{\sigma}^2)$, and this intersection was shown to be unique [17]. Before proceeding, let us denote by q_i the unique integer that satisfies

$$w_{i,q_i+1} \le \hat{\sigma}^2 < w_{i,q_i}, \tag{9}$$

where $w_{i,n}$ is the nth largest value among $\left|\tilde{y}_{\mathcal{F},i}^{(1)}\right|^2, \ldots, \left|\tilde{y}_{\mathcal{F},i}^{(F)}\right|^2$, and define $w_{i,0} = \infty$ and $w_{i,F+1} = 0$. Moreover, we define $\mathscr{F}_i$ as the set of frequencies that correspond to the $F - q_i$ smallest values of $\left|\hat{y}_{\mathcal{F},i}^{(n)}\right|^2$ and $\bar{\mathscr{F}}_i$ is the set containing the remaining q_i frequencies. Taking these definitions into account, the ML estimate of the noise variance is

$$\hat{\sigma}^2 = \frac{1}{\bar{N}_i}\left(\sum_{n\in\mathscr{F}_i}\left|\tilde{y}_{\mathcal{F},i}^{(n)}\right|^2 + \|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},i}\|^2\right). \tag{10}$$

where $\bar{N}_i = N - q_i$. Note that, contrary to the first-order model ML estimate that only uses the out-of-band energy, (10) could also take into account weak in-band components. To obtain q_i, we must sweep it from F to 0, compute (10), and keep the value fulfilling (9). Plugging back (10) into (7), we get the ML estimator of $\bar{d}_{i,n}$, which allows us to obtain the compressed likelihood under $\mathcal{H}_i$.

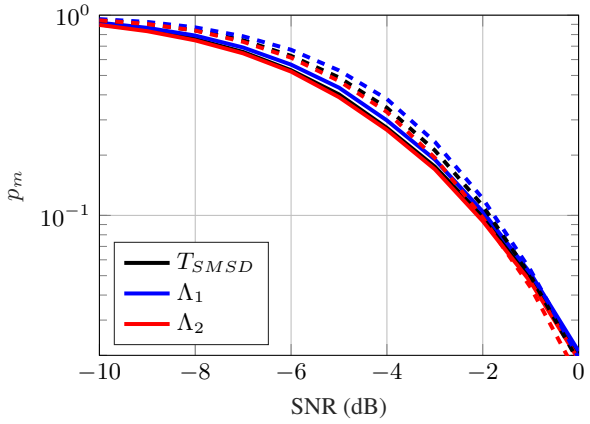

Fig. 1. Probability of missed detection vs. SNR for $p_{fa} = 10^{-3}$ in a scenario with $\eta = 0.1$, and two different cases for F. 1) Dashed line: $F = 30$; 2) Solid line: $F = 20$

Equipped with both compressed likelihoods, the GLR test becomes

$$\Lambda_2 = \frac{\left[\frac{1}{\bar{N}_0}\left(\|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},0}\|^2 + \sum_{n\in\bar{\mathscr{F}}_0}\left|\tilde{y}_{\mathcal{F},0}^{(n)}\right|^2\right)\right]^{\bar{N}_0}\prod_{n\in\bar{\mathscr{F}}_0}\left|\tilde{y}_{\mathcal{F},0}^{(n)}\right|^2}{\left[\frac{1}{\bar{N}_1}\left(\|\tilde{\mathbf{y}}_{\bar{\mathcal{F}},1}\|^2 + \sum_{n\in\bar{\mathscr{F}}_1}\left|\tilde{y}_{\mathcal{F},1}^{(n)}\right|^2\right)\right]^{\bar{N}_1}\prod_{n\in\bar{\mathscr{F}}_1}\left|\tilde{y}_{\mathcal{F},1}^{(n)}\right|^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma.$$

4. NUMERICAL RESULTS

4.1. Synthetic data

In the following experiments, we have considered Watts-Strogatz networks with $N = 100$ nodes and $\tilde{\mathbf{x}}_{\mathcal{F},i} \sim \mathcal{N}_F(\mathbf{0}, \mathbf{D}_i)$. Under $\mathcal{H}_1$, the mean node degree is $2K_1 = 40$ and the rewiring probability is $\beta_1 = 0.2$, whereas under $\mathcal{H}_0$, these parameters are $2K_0 = 12$ and $\beta_0 = 0.1$. Similar to [18], we have considered that $d_{i,n} = a^{n-1}, n = 1, \ldots, F$. Of course, this exponential profile is a simplification, but it allows for an insightful comparison and an easy way to control the spectral flatness, defined as

$$\eta = \frac{\left(\prod_{n=1}^{F} d_{i,n}\right)^{1/F}}{\frac{1}{F}\sum_{n=1}^{F} d_{i,n}}, \quad i = 0, 1,$$

which will be a critical parameter for determining the performance of the detectors. The case of $\eta = 1$ corresponds to a white signal (that is, flat spectrum), whereas small values of η correspond to colored signals with a highly asymmetric spectrum. Moreover, the signal-to-noise ratio is defined as

$$\text{SNR (dB)} = 10\log_{10}\left(\frac{\sum_{n=1}^{F} d_{i,n}}{\sigma^2 N}\right).$$

The first experiment considers a scenario with $\eta = 0.1$, and two different bandwidths: $F = 30$ and $F = 20$. Fig. 1 depicts the probability of missed detection, p_m, vs. SNR for a probability of false alarm $p_{fa} = 10^{-3}$. As this figure shows, for this example with highly asymmetric power profile, the second-order GLR, Λ_2, performs better than the first-order GLR, Λ_1, and T_{SMSD}, although the

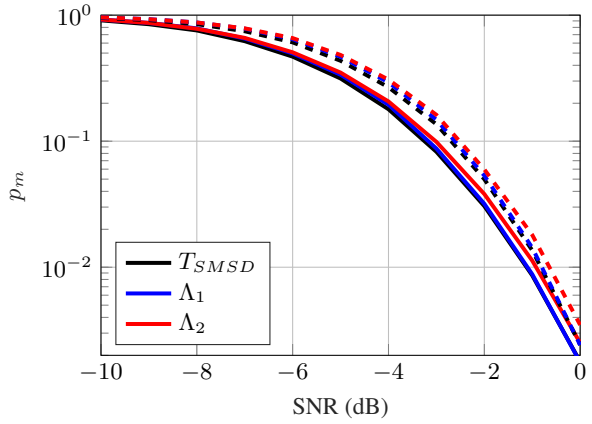

Fig. 2. Probability of missed detection vs. SNR for $p_{fa} = 10^{-3}$ in a scenario with $\eta = 0.5$, and two different cases for F. 1) Dashed line: $F = 30$; 2) Solid line: $F = 20$.

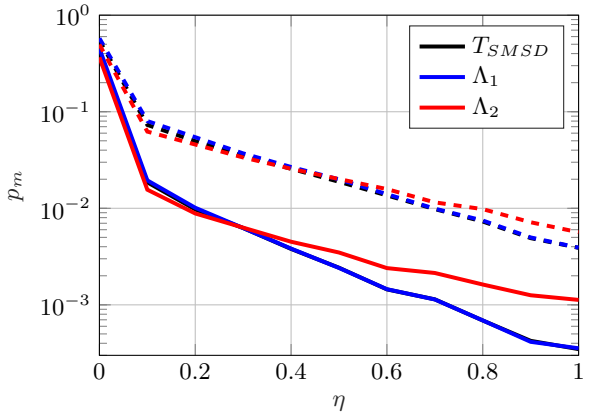

Fig. 3. Probability of missed detection vs. η for $p_{fa} = 10^{-3}$ in a scenario with SNR $= 0$, $F = 30$, and two different distributions. 1) Dashed line: Laplacian; 2) Solid line: Gaussian

performance differences are small. Moreover, the effect of increasing the bandwidth from $F = 20$ to $F = 30$ is not too severe.

The second experiment considers a similar scenario where the only difference is the spectral flatness, which is now $\eta = 0.5$. In this case, we can observe in Fig. 2 that the order of the different detectors has reversed, and now Λ_1 and T_{SMSD} outperform Λ_2. Nevertheless, the performance differences are also negligible. Similar conclusions can also be drawn regarding the effect of increasing the bandwidth and the order among Λ_1 and T_{SMSD}.

In the two previous experiments, as we already pointed out, we have seen that η is a critical parameter that determines the performance of the detectors. In this experiment, we corroborate this behavior by evaluating the probability of missed detection against η for $p_{fa} = 10^{-3}$ in a scenario with SNR $= 0$ dBs and $F = 30$. The results, which are depicted in Fig. 3, show that for small values of η the second-order GLR presents the best performance, but it is outperformed as η increases. Additionally, this figure also shows the results when $\tilde{\mathbf{x}}_{\mathcal{F},i}$ is drawn from a Laplacian distribution instead of a Gaussian one. The motivation for this new experiment is to show that the previous insights hold for other distributions different from the Gaussian, which was the one used in the derivation of the second-order GLR.

One final comment is in order. Based on these experiments, one could conclude that T_{SMSD} should be the detector of choice. It presents a slightly better performance than Λ_1 for large η and only slightly worse than Λ_2 for small η. However, the distribution under $\mathcal{H}_0$ of T_{SMSD} depends on σ^2 and $\tilde{\mathbf{x}}_{\mathcal{F},0}$, which are unknown; a problem shared by Λ_2 that requires the prior of $\tilde{\mathbf{x}}_{\mathcal{F},0}$. On the contrary, the null distribution of Λ_1 does not depend on neither σ^2 nor $\tilde{\mathbf{x}}_{\mathcal{F},0}$. That is, Λ_1 is a constant false alarm rate (CFAR) detector.

4.2. Real-world data

In this experiment, we evaluate our method on the Molene dataset, which consists of hourly temperature records from 37 weather stations around Brest, France, spanning 744 hours [19]. Under $\mathcal{H}_0$, the graph is constructed by connecting pairs of stations based on a distance threshold, while the signal corresponds to the original temperature data. Under $\mathcal{H}_1$, the graph is perturbed via random rewiring, and the signal is generated according to the perturbed topology using the same GFT coefficients as under $\mathcal{H}_0$.

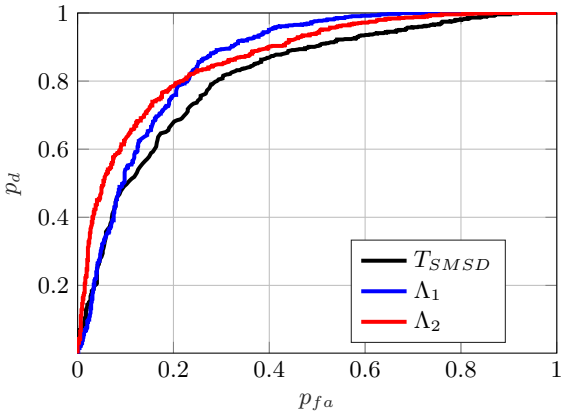

Fig. 4. ROC of three different detectors for the Molene dataset

The receiver operating characteristic (ROC) curves are depicted in Fig. 4, which shows that all three detectors are capable of identifying the underlying graph, as the temperature data exhibits smoothness. Both, the first- and second-order GLR detectors outperform T_{SMSD}, probably due to the absence of explicit knowledge about the noise variance in the signal.

5. CONCLUSIONS

In this paper, we have developed two MSD for graph-based data: first-order and second-order GLR tests. By addressing the limitations of prior methods, such as reliance on known noise variance and overlooking the statistical properties of graph Fourier transform coefficients, our proposed solutions offer improved generalization and robustness. The first-order GLR estimates the graph Fourier transform (GFT) coefficients and the noise variance via maximum likelihood, while the second-order GLR leverages Gaussian-distributed GFT coefficients for richer statistical modeling. Experimental results demonstrate the effectiveness of both methods.

6. REFERENCES

[1] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[2] Z. Li, Q. Li, X. Yu, P. S. Conti, and R. M. Leahy, "Lesion detection in dynamic FDG-PET using matched subspace detection," *IEEE Trans. on Medical Imaging*, vol. 28, no. 2, pp. 230–240, 2008.

[3] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, 2002.

[4] M. L McCloud and L. L. Scharf, "Interference estimation with applications to blind multiple-access communication over fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 947–961, 2002.

[5] A. Ortega, P. Frossard, J. Kovavcević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[6] W. H. Kim et al., "Multi-resolutional brain network filtering and analysis via wavelets on non-Euclidean space," in *Int. Conf. on Medical Image Comp. and Computer-Assisted Intervention*, 2013, pp. 643–651.

[7] F. B. Viégas and J. Donath, "Social network visualization: Can we go beyond the graph?," in *Work. on Social Networks, CSCW*, 2004, vol. 4, pp. 6–10.

[8] X. Zhu and M. Rabbat, "Graph spectral compressed sensing for sensor networks," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, 2012, pp. 2865–2868.

[9] C. Hu, J. Sepulcre, K. A. Johnson, G. E. Fakhri, Y. M. Lu, and Q. Li, "Matched signal detection on graphs: Theory and application to brain imaging data classification," *NeuroImage*, vol. 125, pp. 587–600, 2016.

[10] S. P. Chepuri and G. Leus, "Subgraph detection using graph signals," in *Asilomar Conf. Signals, Systems, and Computers*, 2016, pp. 532–534.

[11] E. Isufi, A. S. U. Mahabir, and G. Leus, "Blind graph topology change detection," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 655–659, 2018.

[12] C. Liu, V. M. Tenorio, A. G. Marques, and E. Isufi, "Matched topological subspace detector," *arXiv:2504.05892*, 2025.

[13] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.

[14] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, 2014.

[15] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, New York: Academic, 1979.

[16] D. Ramírez, I. Santamaría, and L. Scharf, *Coherence: In Signal Processing and Machine Learning*, Springer Nature, 2023.

[17] Y. Bresler, "Maximum likelihood estimation of a linearly structured covariance with application to antenna array processing," in *Annual ASSP Work. Spectrum Estimation and Modeling*, 1988, pp. 172–175.

[18] I. Santamaría, L. L. Scharf, and D. Ramírez, "Scale-invariant subspace detectors based on first- and second-order statistical models," *IEEE Trans. Signal Process.*, vol. 68, pp. 6432–6443, 2020.

[19] B. Girault, S. S. Narayanan, and A. Ortega, "Local stationarity of graph signals: Insights and experiments," in *Wavelets and Sparsity XVII*, 2017, vol. 10394, pp. 340–356.