Blind Graph Topology Change Detection:

A Graph Signal Processing approach

by

A.S.U. Mahabir

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on xxxxx.

Student number:	4043278
Faculty:	EEMCS
Master Program:	Electrical Engineering
Track:	Signals and Systems
Thesis committee:	

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Circuits and Systems Group

Abstract

Graphs are used to model irregular data structures and serve as models to represent/capture the interrelationships between data. The data in graphs are also referred as graph signals. Graph signal processing (GSP) can then be applied which basically extends classical signal processing to solve problems. Anomaly detection is an example of such a problem. Two hypothetical situations are given, and a detector has to be designed to distinguish between these. Under the null hypothesis, graph structures are considered to be untouched. Under the alternative hypothesis, (unknown) topological changes might have occurred. Now by incorporating a priori knowledge about the graphs, the decision making process should improve.

In most works, a priori knowledge of the graphs under the null and alternative hypothesis was incorporated. This means that detectors were designed which were able to anticipate on possible topological changes. In this thesis, the problem is considered where only a priori knowledge of the graph under the null hypothesis is exploited. This means that detectors are not able to anticipate on potential changes and this where blind detection comes into play. Blind detection is important because it considers a more realistic scenario. In this work, the blind topology change detector (BTCD) and the constrained blind topology change detector (CTCD) are derived which exploit different properties of the data related to the known graph structure. For the BTCD, the bandlimitedness of graph signals was exploited and for the CTCD, the graph signal smoothness. The main question in this work, was to find out what the potentials are with the blind detection principle for graph change detection.

Different test scenarios are used to evaluate the detectors on both synthetic and real data. For the BTCD, the obtained results compare well when information about the alternative graph is available. For this detector, the potential of blind detection was highly visible. For bandlimited graph signals, the BTCD as good as detectors using full information. For the CTCD, comparable results (with detectors using full information) are attained for just a few test scenarios. For small changes, the graph signal smoothness seems to be less powerful as to the graph signal bandlimitedness.

This study showed that graph change detection is still possible without having full information. Some graph signal properties are more powerful w.r.t. others.

Preface

My interest in signal processing began in the first semester of my second year of my Bachelor studies here at TU Delft. To be able to understand concepts on how information bearing signals are analyzed, synthesized and modified, was just wonderful. The possibilities with signal processing are endless and to be part of a fairly new field within signal processing, gives me a sense of pride. Graph signal processing basically extends classical signal processing and all its applications, such as detection theory, to a new direction.

Towards the end of my master, I was presented with a rather interesting research topic, if graph signal processing could be used to detect graph anomalies. What differs this problem with existing ones, is that detection, in this work, comes into force using less information. The lack of information is due to the fact that anomalies are often not known in advance. This is where the blind detection comes into play. The analogy between classical signal processing made this project appealing to me, since it allowed me to use knowledge gained from previous years.

I hope that this report conveys to the reader my thoughts and ideas. It pleases me when I look back on the progression of the project. From reading existing papers on graph signal processing and graph anomaly detection for an appropriate research question, towards the final phase in obtaining my degree as a Master of Science in Electrical Engineering.

A.S.U. Mahabir
Delft, September 2017

Acknowledgements

There have been quite a few people who were of paramount importance throughout the writing of my thesis. Firstly, I would like to thank my family, friends and classmates for their support. Without a doubt I would like to thank my parents for their unconditional support and for granting me the opportunity to chase my dreams of becoming an electrical engineer. I am truly grateful for the amount of motivation given to me by my friends who did not hesitate to encourage me whenever I needed it and who always offered a listening ear. Special thanks goes to my Professor Geert Leus and also his Ph.D. student Elvin Isufi and Ana. They encouraged me to push further past my boundaries when I felt I had reached my limit. From the moment I embarked on this journey I truly felt that I could count on them.

A.S.U. Mahabir
Delft, September 2017

Contents

1 Introduction

In a social media network nowadays, users interact with each other through online relationships. Each user of such a network can be seen as a node in a complex network which can also be referred to a graph [1]. The interdependencies between nodes can be modeled using links which, for example, could represent connections that nodes have in common, such as, friendships. An example of such a graph is given in Figure 1.1. A graph also tends to be useful in generating/storing data [2]. For example, a sensor network for environment monitoring systems. The data that is collected is then indexed by the nodes of the graph representing the sensor network [3]. Such data are also referred as graph signals. What differs a graph data storage structure with respect to a classical one (e.g. matrix like storage structure), is that there are no well ordered samples anymore. This is because a node can be linked to any other node in the graph.

Figure 1.1: Example of a social network where links represent connections that people have in common, e.g., friendships.

In practice, there are many signals, derived from a large variety of fields such as social, economic, epidemiology, biological and transportation networks. All these signals can be modeled as signals on graphs [4]. The interdependencies of these signals can then be captured by the structure of a graph [2] [5]. Another application is modeling images with graphs. For example, video signals own the property that they are ordered according to their spatial position/time instances defined in the sampling process [6]. Nodes represent pixels and links reveal how pixels are related to their neighborhood. Together they form the structure and size of the image plus the mutual dependencies between the pixels. The graph signal in this example are then RGB values which are stored on each node/pixel (see Figure 1.2).

Graph signals show key dependencies which arise from the irregular data domain. This means that classical signal processing tools can no longer be applied and give reason to extend the analysis to a new paradigm, called graph signal processing (GSP) [7]. A common problem in the context of signal processing is the task of detecting events of interest by analyzing signals. This is also known as detection theory [8]. Detection theory can then be used to decide between two or more hypothetical situations. An example of such a situation, is to decide whether a signal (embedded in noise) is present or not in a sensor measurement.

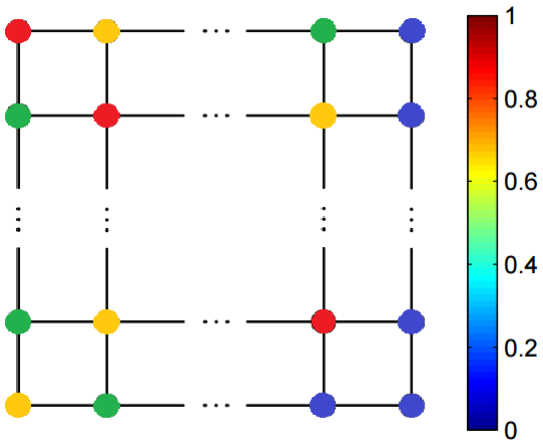


Figure 1.2: Example of an image represented as a (ordered) graph. Each node represents a pixel and the links represent how each pixel is related to its neighborhood. The graph signal in this case is a specific RGB value indexed by a pixel [5] [6].

A problem that is common for graphs, consists of detecting topology anomalies which are described by graph topology changes. Graph anomaly detection is applied in different areas such as security, finance, health care, and law enforcement [9]. A detector must then be designed to distinguish between the nominal hypothesis $\mathcal{H}_0$ where the topology of the graph remains untouched and the alternative hypothesis $\mathcal{H}_1$ where graph topology changes have occurred:

$$\begin{aligned} \mathcal{H}_0 &: \quad \text{Nominal graph} \\ \mathcal{H}_1 &: \quad \text{Alternative graph} \end{aligned}$$

In Figure 1.3 an example is given where two graphs are shown with the same nodes but with different links. This indicates a change in the graph topology. Graph anomaly detection is accomplished by collecting graph signal observations and applying GSP techniques since the statistics of these signals reveal graph topological structures [10]. Later in this work, where the detectors are designed, the correlation between graph signal statistics and graph topology will be established and explained.

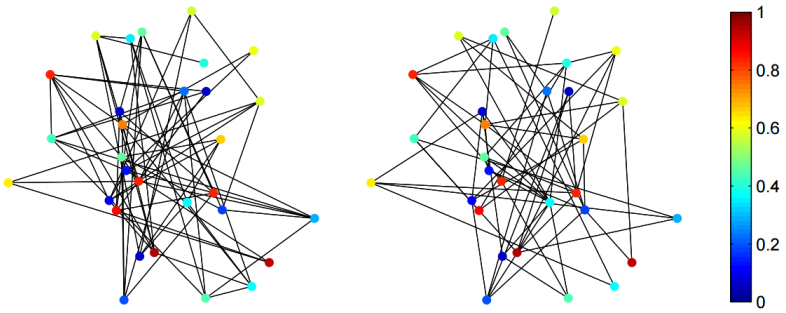


Figure 1.3: Example of two random graphs carrying different graph signals. Both graphs have the same nodes but the graph on the left has different links w.r.t. the graph on the right. This implies the occurrence of a topology change [11].

In previous works, cases were considered where the nominal and the alternative graph were known. An example of such anomaly detection problem is described in [12] [13] where the authors attempted to detect known target graphs embedded in a noisy background graph. More formally, this problem is written as:

$$\begin{aligned} \mathcal{H}_0: &\quad \text{noisy background graph} \\ \mathcal{H}_1: &\quad \text{noisy background graph + target graph} \end{aligned}$$

Meanwhile, in [11], GSP is applied to fMRI data data to detect early stages of Alzheimer. A graph signal was fit to known candidate graphs where one graph represents a graph under normal circumstances while the alternative graphs represent anomalous graphs. What is meant by a graph under normal circumstances, is a healthy brain while the anomalous graphs represent brain structures dealing with diseases, such as, schizophrenia, autism, strokes and Alzheimer [10]. Such a problem can then be formulated as follows:

$$\begin{aligned} \mathcal{H}_0: &\quad \text{candidate graph: healthy brain} \\ \mathcal{H}_1: &\quad \text{candidate graphs: brain with a set of possible diseases} \end{aligned}$$

Knowing graph topologies a priori plays an important role for the detection performance. A detector is then capable to anticipate anomalies since alternative graphs (with topological changes included) are known in advance. In this work, a detection problem is considered where only the nominal graph is known. This affects the performance significantly, but it considers a more realistic case since topology changes are usually not known a priori. Throughout this thesis, the term blind detection will be used to point to anomaly detection without knowing the alternative graph, i.e., being able to anticipate on possible structural changes. For instance, when one only knows that the brain is not healthy, the problem is formulated as follows:

$$\begin{aligned} \mathcal{H}_0: &\quad \text{candidate graph: healthy brain} \\ \mathcal{H}_1: &\quad \text{candidate graphs: } \textbf{not} \text{ healthy brain with an unknown set of possible diseases} \end{aligned}$$

The scope of this work is to study the possibilities with blind detection and to evaluate detectors that are designed according to this principle. For the different detectors, signal properties such as graph signal bandlimitedness [11], graph signal smoothness [11] and graph signal stationarity [14] [15] [16] are exploited. A sub-question is therefore: how do topology changes reflect in the statistics of an observed graph signal when, the blind detection principle is used. The work of [11] is used as reference and to consider a more realistic case, small topology changes are also considered. Besides using synthetic graphs, real graph are also considered as well to see how discriminative the mentioned graph signal properties are.

A possible example of a more realistic (blind detection) problem is to detect whether the topology in a brain has changed over time without using candidate graphs (see Figure 1.4). The candidate graphs could capture a set of (unknown) frequently occurring topological changes. Using these candidate graphs would lead to a detector that is capable to forecast possible diseases.

The proposed detection technique can then be used as a diagnostic tool to detect such diseases in early stages. The results with the blind detection principle can be used for a deeper analysis on targeted patients.

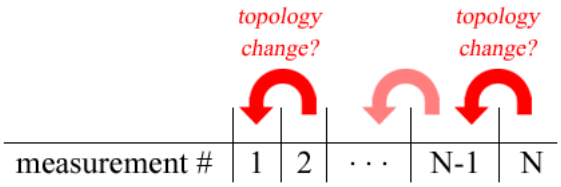


Figure 1.4: Example of blind detection with a followup scheme of brain measurements. For every consecutive measurement, one can wonder whether the new observation has changed w.r.t. the previous one.

The remaining part of this work is organized as follows. First the preliminaries are presented in chapter 2. In this chapter some detailed information regarding GSP and detection theory are given. Also the signal properties used for hypothesis discrimination will be discussed in detail.

In chapter 3, a detector is designed which discriminates between hypotheses on the basis of graph signal bandlimitedness. The energy of the signal is assumed to be bounded for the nominal graph and this forms the basis for the detector derived in this chapter.

In chapter 4 a detector is designed exploiting graph signal smoothness. Besides for the energy being bounded for the nominal graph, its energy decay is also expected to be bounded. The distribution of the signal energy over its spectrum could then reveal potential graph changes.

Finally in section 5, a glance is given on possible future work. Graph signal stationarity, compressive sampling and compressive sensing will be discussed on possible research directions. Stationarity for graph signals is graph topology dependent. This property could get (partially) lost as the graph structure changes and could form a basis of a detector design. Compressive sampling/sensing are techniques which can be used to speed up the signal processing procedure.

In the very last chapter of this work, an overall conclusion is given on the attainable detection performance of blind detection.

2

Preliminaries and Problem definition

Before going into detail of the considered problem in this chapter, some background information on GSP and detection theory is given in Sections 2.1 and 2.2 respectively. Different signal properties (such as graph signal bandlimitedness, smoothness and stationarity) and their potential discriminative power will be discussed as well. Subsequently, the problem of interest is given in terms of GSP and detection theory in Section 2.3.

2.1. Graph signal processing

A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of an edge set $\mathcal{E}$ and a node set $\mathcal{V}$. The edge set $\mathcal{E}$ contains all the links which connect the nodes. In this work, undirected graphs are considered. An undirected graph is a model where links have no orientation, which means that a link from node A to node B is the same as the link from node B to node A. Also, it is assumed that the graph is unweighted which means that the links in the graph carry no weight, i.e., either a link carries an one if there is a link or a zero otherwise. The number of nodes is N, i.e., $|\mathcal{V}| = N$. A graph signal $\mathbf{x} \in \mathbb{R}^N$ is then defined on the N vertices of a graph and is a mapping from the vertex domain to the set of real numbers, i.e., $\mathbf{x} : \mathcal{V} \to \mathbb{R}^N$. Graph structures are described by their so called adjacency matrices $\mathbf{A} \in \mathbb{R}^{N \times N}$. Elements of the adjacency matrix $a_{i,j}$, $i, j \in \{1, 2, \cdots, N\}$, show whether vertex pairs are adjacent or not in the graph [7]. Since unweighted graphs are considered, a 1 is placed for nodes which are adjacent, as illustrated in Figure 2.1:

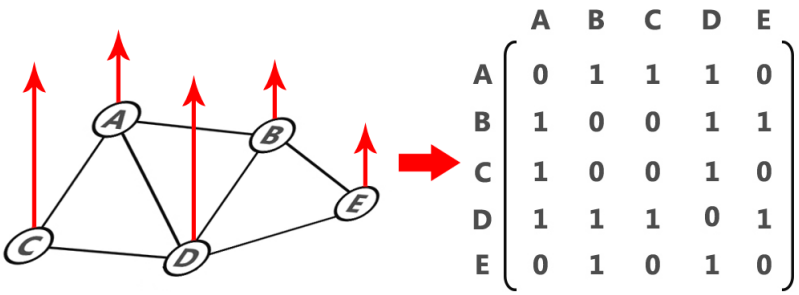


Figure 2.1: Example of a undirected and unweighted graph (left) carrying a graph signal on each node. Also its corresponding adjacency matrix (right) is shown.

In the field of GSP, the degree matrix $\mathbf{D} \in \mathbb{R}^{N \times N}$ is a diagonal matrix which contains information about the degree of each node. The degree matrix $\mathbf{D}$ gives the number of links that are coupled to each node in the unweighted graph. Suppose $d_{i,j}$ with $i, j \in \{1, 2, \cdots, N\}$ is an element of $\mathbf{D}$. Also assume that v_i is a node from set $\mathcal{V}$. The degree matrix is then defined as:

$$d_{i,j} = \begin{cases} \deg(v_i) & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

$\deg(v_i)$ the number of links terminating node v_i. Another common graph structure descriptor is the graph Laplacian $\mathbf{L} \in \mathbb{R}^{N \times N}$. Matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is a real symmetric matrix for undirected graphs. Since $\mathbf{L}$ is real and symmetric, an eigenvalue decomposition exists:

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \tag{2.2}$$

where $\mathbf{U}$ is the unitary eigenvector matrix and $\mathbf{\Lambda}$ is the diagonal eigenvalue matrix. The graph Laplacian forms the basis for spectral graph theory and is used to define the graph Fourier transform (GFT) [17]. The GFT $\hat{\mathbf{x}}$ of $\mathbf{x}$ is defined as the projection of $\mathbf{x}$ onto the eigenvectors of the graph Laplacian, i.e.,:

$$\hat{\mathbf{x}} = \mathbf{U}^H\mathbf{x}. \tag{2.3}$$

Similarly, the inverse GFT is defined as $\mathbf{x} = \mathbf{U}\hat{\mathbf{x}}$. The eigenvalues of the graph Laplacian $\mathbf{L}$, i.e., $\mathbf{\Lambda} = \text{diag}([\lambda_1\ \lambda_2\ \cdots\ \lambda_N])$, are seen as graph frequencies and show a similar behavior as frequencies in the classical setting [7] [17]. In the classical setting, the Fourier transform is defined as an expansion of a signal in terms of complex exponentials. These complex exponentials are known to be eigenfunctions of the one-dimensional Laplace operator [2]. This is the reason why graph frequencies are considered as an extension of the classical definition of temporal frequencies. The GFT expands a graph signal in terms of the the eigenvectors of the graph Laplacian $\mathbf{L}$ which is seen in (2.3). For an even deeper understanding of the GFT, the reader can consult [7] [17].

2.1.1. Graph signal bandlimitedness

Since the analogy between classical frequencies and graph frequencies has been established, graph spectrum analysis can be carried out now in a similar way. Just like bandlimited signals exist for classical signals, bandlimited graph signals also exist. Bandlimited signals are used for different techniques. They are often made bandlimited with windowing/sampling techniques on the Fourier transform of a signal to prevent any spurious emissions, i.e., unwanted frequencies. For certain eigenvectors, $\hat{\mathbf{x}}$ has no content. Bandlimited graph signals are defined as follows:

$$\hat{x}(\lambda_i) = 0, \quad \forall \quad \lambda_i > \lambda_K, \tag{2.4}$$

which implies that the graph frequency content of a graph signal is zero only on a subset of graph frequencies (see Figure 2.3). The property of a signal being bandlimited can be exploited using subspace sampling for example, to suppress unwanted frequencies [18] [19]. Also filtering is possible in the context of graphs where certain frequencies are either suppressed or passed in a certain band [2] [20]. This is accomplished with the following: instead of using all eigenvectors of $\mathbf{L}$, one can only use the eigenvectors corresponding to frequencies of interest, e.g., the frequencies corresponding to the out of band content:

$$\mathbf{U}_K^c = [\mathbf{u}_{K+1} \cdots \mathbf{u}_N]. \tag{2.5}$$

Applying this transform matrix $\mathbf{U}_K^c$ should be seen as a form of high pass filtering since it can suppress graph frequencies. If the graph frequency content is band limited, its energy is zero out of band. This could be helpful to distinguish between graphs since GFT coefficients generally depend on graph topology characteristics. So, the property of $\hat{\mathbf{x}}$ being bandlimited on a graph $\mathcal{G}$ can be exploited since another frequency graph signal $\hat{\mathbf{x}}' \neq \hat{\mathbf{x}}$ might not have the same spectral content.

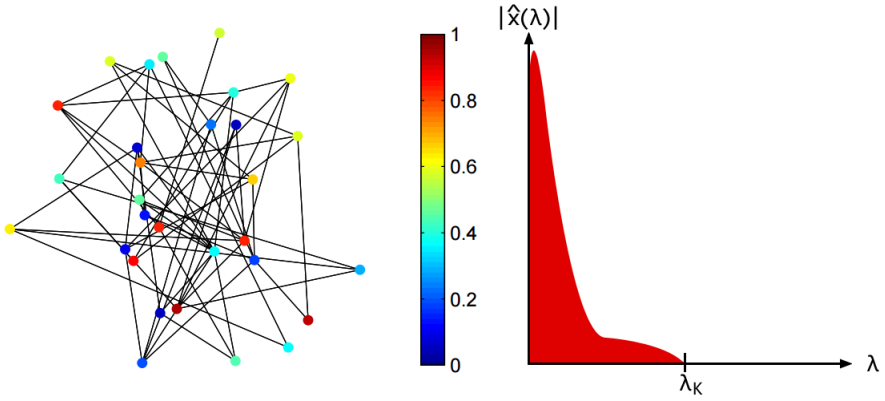


Figure 2.2: Example of a graph with a bandlimited graph signal spectrum.

In Chapter 3, the bandlimitedness of $\hat{\mathbf{x}}$ is exploited where K is assumed to be known under $\mathcal{H}_0$.

2.1.2. Graph signal smoothness

Another property that can be exploited in order to distinguish between graphs, is the graph signal smoothness. The property of $\hat{\mathbf{x}}$ being smooth on a graph $\mathcal{G}$ can be exploited since another frequency graph signal $\hat{\mathbf{x}}' \neq \hat{\mathbf{x}}$ might not have the same energy decay in its spectrum. Smoothness in the context of GSP has a similar meaning as for the classical setting, where Fourier coefficients tend to decay rapidly in its spectral domain. The same thing can be said for smooth graph signals [2] [17] (see Figure 2.3). Graph signal smoothness can be quantified with the following measure [11]:

$$C(\hat{\mathbf{x}}) = \frac{\hat{\mathbf{x}}^H \mathbf{\Lambda} \hat{\mathbf{x}}}{\hat{\mathbf{x}}^H \hat{\mathbf{x}}} = \frac{\mathbf{x}^H \mathbf{L} \mathbf{x}}{\mathbf{x}^H \mathbf{x}}. \tag{2.6}$$

This quantity is small when $\mathbf{x}$ tends to be constant in a certain neighborhood of vertices [2]. A signal being smooth in practice means that $\hat{\mathbf{x}}$ consist mostly out of low pass frequency content, i.e., that a smooth $\hat{\mathbf{x}}$ is slow-varying on $\mathcal{G}$. Now if the structure of the graph changes, the smoothness of $\hat{\mathbf{x}}$ on the changed graph is expected to change as well. If this change is significant w.r.t. the smoothness measured under the nominal graph, one should be able to distinguish between graphs. This is done by observing directly the graph signal and its relation w.r.t. these graphs.

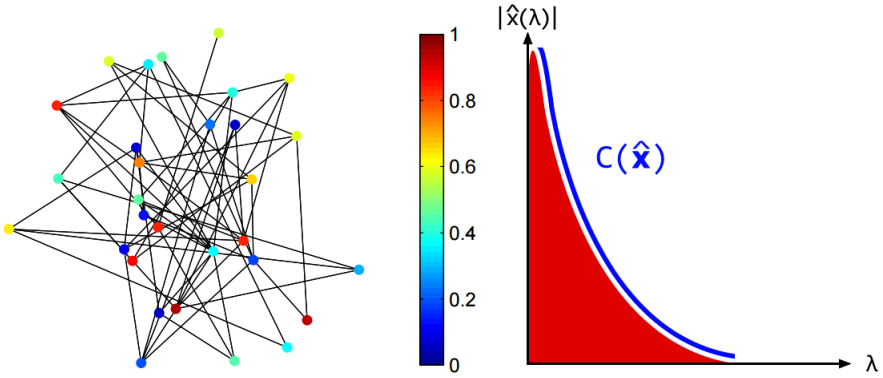


Figure 2.3: Example of a graph with a smooth graph signal spectrum which is bounded.

In Chapter 4, the smoothness of $\hat{\mathbf{x}}$ is exploited. The smoothness measure $C(\hat{\mathbf{x}})$ is assumed to be bounded with a known upper bound under $\mathcal{H}_0$.

2.1.3. Graph signal stationarity

In the last problem of this work, detection based on (weakly) stationary graph signals is examined. The property of $\hat{\mathbf{x}}$ being stationary on a graph $\mathcal{G}$ can be exploited since another frequency graph signal $\hat{\mathbf{x}}' \neq \hat{\mathbf{x}}$ might have lost this property due to unknown graph changes. Such stationary signals have a stochastic nature and can be modeled with parametric models, i.e., $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_x)$ where $\mathbf{C}_x \in \mathbb{R}^{N \times N}$ is the covariance matrix of $\mathbf{x}$. For a graph $\mathcal{G}$ with laplacian matrix $\mathbf{L}$ which has an eigenvalue decomposition as in (2.2). A graph signal $\mathbf{x}$ is then said to be stationary w.r.t. $\mathcal{G}$ if the covariance matrix $\mathbf{C}_x$ and the Laplacian matrix $\mathbf{L}$ are simultaneously diagonalizable [14] [16], i.e.,:

$$\begin{aligned} \mathbf{\Lambda} &= \mathbf{U}^H \mathbf{L} \mathbf{U} \\ \text{diag}(\mathbf{p}_x) &= \mathbf{U}^H \mathbf{C}_x \mathbf{U} \end{aligned} \tag{2.7}$$

where the vector $\mathbf{p}_x$ is known as the power spectral density (PSD) of $\mathbf{x}$. Now to quantify stationarity, a measure was adopted from [15]. The level of stationarity of a signal $\mathbf{x}$, is based on how diagonalizable its covariance matrix $\mathbf{C}_x$ is, using the eigenvector matrix $\mathbf{U}$ of $\mathcal{G}$. With this in place, the stationarity level measure $s(\mathbf{C}_x, \mathbf{U})$ is defined as:

$$s(\mathbf{C}_x, \mathbf{U}) = \frac{||\text{diag}(\mathbf{U}^H \mathbf{C}_x \mathbf{U})||_2}{||\mathbf{U}^H \mathbf{C}_x \mathbf{U}||_F}, \tag{2.8}$$

where the operators $||\bullet||_2$ and $||\bullet||_F$ stand for the 2- and Frobenius norm respectively.

When $s(\mathbf{C}_x, \mathbf{U})$ is close to 1, the more diagonal the matrix $\mathbf{U}^H \mathbf{C}_x \mathbf{U}$ is and the more stationary the signal $\mathbf{x}$ is w.r.t. the associated graph. This quantification of the stationarity is then expected to be powerful enough to distinguish between the hypothetical graphs. Under $\mathcal{H}_0$, this quantity is expected to be upper bounded with a known bound. So, if unknown changes occur, the stationarity level of a graph signal is expected to be less on the alternative graph when compared to the nominal graph.

2.2. Detection theory

A common goal within the field of signal processing, is the goal of being able to decide whether an even of interest occurs or not. One of the simplest detection problem is to find out whether a signal $\mathbf{x}$ (embedded in noise $\mathbf{n}$) is present in a measurement $\mathbf{y}$ or not. For the latter, this means that solely noise is present. The hypothesis, where a signal $\mathbf{x}$ is present, is also known as the alternative hypothesis $\mathcal{H}_1$. When only the noise is present, this is known as the null hypothesis $\mathcal{H}_0$:

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathbf{y} = \mathbf{n} \\ \mathcal{H}_1 : &\quad \mathbf{y} = \mathbf{x} + \mathbf{n}. \end{aligned} \tag{2.9}$$

The problem then exists of deciding between two possible hypotheses and is also referred as the binary hypothesis testing problem. The goal is to exploit the observed data $\mathbf{y}$ as efficient as possible, such that the correct hypothesis is decided most of the time in the decision making. The measurement data are random in nature. Statistical models (e.g. Gaussian models) are then used to model the distributions behind these potential hypotheses. For instance, if $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I})$ and $\mathbf{x}$ is deterministic, the following is true for the distribution of $\mathbf{y}$:

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I}) \\ \mathcal{H}_1 : &\quad \mathbf{y} \sim \mathcal{N}(\mathbf{x}, \sigma_n^2\mathbf{I}). \end{aligned} \tag{2.10}$$

The distribution under $\mathcal{H}_0$ can be denoted by $p(\mathbf{y}; \mathcal{H}_0)$ and under $\mathcal{H}_1$, by $p(\mathbf{y}; \mathcal{H}_1)$. There are several decisions that the detector can take which are reflected in different probabilities. The probability of false alarm P_{FA} is defined as $P(\mathcal{H}_1; \mathcal{H}_0)$. This is the probability of deciding $\mathcal{H}_1$ while in fact $\mathcal{H}_0$ was true. The probability of detection P_D is defined as $P(\mathcal{H}_1; \mathcal{H}_1)$ which means that $\mathcal{H}_1$ is decided while this same hypothesis is true. This problem setup is referred as the Neyman-Pearson (NP) approach for hypothesis testing/signal detection [8]. With the NP approach, one designs a detector that tries to maximize the probability of detection P_D, while the probability of false alarm P_{FA} is kept fixed, e.g., $P_{FA} = \alpha$. This is accomplished by applying the so called likelihood ratio test $L(\mathbf{y})$ on the observed data. The $L(\mathbf{y})$ is then compared against a detection threshold γ to decide which hypothesis is true:

$$L(\mathbf{y}) = \frac{p(\mathbf{y}; \mathcal{H}_1)}{p(\mathbf{y}; \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{2.11}$$

where γ is found from:

$$P_{FA} = \int p(\mathbf{y}; \mathcal{H}_0) d\mathbf{y} = \alpha. \tag{2.12}$$

By simplifying and by mathematical reduction of $L(\mathbf{y})$, a detector $T(\mathbf{y})$ can be derived. In Figure 2.4, a block scheme is given for a detector.

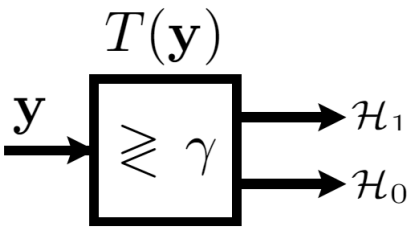


Figure 2.4: Classical detector scheme: An observed signal is fed into the detector and a decision is made.

A more complicated detection problem arises when the parametric models, describing the distributions under both hypothesis, contain unknown parameters. If this is the case, they have to be estimated. The data that is collected is then used to perform a maximum likelihood estimation (MLE) in order to find these parameters. Suppose that $\boldsymbol{\theta}$ is a vector which contain unknown parameters. Under $\mathcal{H}_0$, this unknown vector is denoted by $\boldsymbol{\theta}_0$ and under $\mathcal{H}_0$, by $\boldsymbol{\theta}_1$. In this specific scenario, one has the find the MLES $\boldsymbol{\theta}_0^*$ and $\boldsymbol{\theta}_1^*$. These MLEs are then used in the so called generalized likelihood ratio test $L_G(\mathbf{y})$ which is also a function of the data $\mathbf{y}$. This test is again used to derive a detector $T(\mathbf{y})$ which is again compared against a threshold γ for the decision making:

$$L(\mathbf{y}) = \frac{p(\mathbf{y}; \boldsymbol{\theta}_1^*, \mathcal{H}_1)}{p(\mathbf{y}; \boldsymbol{\theta}_0^*, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{2.13}$$

where $\boldsymbol{\theta}_0^*$ and $\boldsymbol{\theta}_1^*$ are found from:

$$\boldsymbol{\theta}_0^* = \arg\max_{\boldsymbol{\theta}} \quad p(\mathbf{y}; \boldsymbol{\theta}_0, \mathcal{H}_0) \tag{2.14}$$

$$\boldsymbol{\theta}_1^* = \arg\max_{\boldsymbol{\theta}} \quad p(\mathbf{y}; \boldsymbol{\theta}_1, \mathcal{H}_1) \tag{2.15}$$

The performance of the detector will be evaluated with receiver operating characteristic (ROC) curves which plots the probability of detection P_D against the probability of false alarm P_{FA} [8]. Each point in a ROC curve represents to a (P_{FA}, P_D) pair for a given detection threshold γ.

Observed graph signals in this work are of a stochastic nature as nodal data is corrupted by noise. Parametric models are then also used to model these distributions in order to analyze graph topological changes as the parameters of these models tend to vary as topological changes occur. The null hypothesis $\mathcal{H}_0$ is the hypothesis where the graph remains untouched. The alternative hypothesis $\mathcal{H}_1$, is the hypothesis where the graph topology is facing (unknown) changes. This means that each hypothesis is characterized by a distribution with unique parameters, especially under $\mathcal{H}_1$. Based on an observation graph signal $\mathbf{y} \in \mathbb{R}^N$, a detector $T(\mathbf{y}) \in \mathbb{R}$ is designed. The detector is then compared to a threshold γ in order to detect graph topological changes.

2.3. Hypothesis testing on graphs

Given a fixed set of vertices $\mathcal{V}$ with N vertices, let graph $\mathcal{G}_0 = (\mathcal{V}, \mathcal{E}_0)$ denote the nominal graph with edge set $\mathcal{E}_0$. The anomalous graph is given by $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ with $\mathcal{E}_1 \neq \mathcal{E}_0$. An example of such a case is given by [11] where the nominal graph represents the brain structure of a healthy person and where the alternative graph represents the structure of the brain of a patient suffering from Alzheimer. Figure 1.3 illustrates two graphs which have the same set of nodes but different edges. Functional magnetic resonance imaging (fMRI) measurements then represent graph signals and based on these, the task is to detect brain structure changes as it might indicate whether a person is in the early stages of Alzheimer. The graph Laplacian $\mathbf{L}$ associated with graph $\mathcal{G}_i$ is denoted by $\mathbf{L}_i = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{U}_i^H$ for $i \in \{0, 1\}$. The graph signal $\mathbf{y} \in \mathbb{R}^N$ is considered as the observation signal and is defined on the N vertices of the graph. The problem described by [11] is as follows: Given an observation graph signal $\mathbf{y}$, the goal is to design hypothesis tests to decide which graph ($\mathcal{G}_0$ or $\mathcal{G}_1$) fits the signal more accurately. The observed graph signal $\mathbf{y}$, is modeled as contaminated version of the graph signals $\mathbf{U}_0\hat{\mathbf{x}} \in \mathbb{R}^N$ or $\mathbf{U}_1\hat{\mathbf{x}} \in \mathbb{R}^N$ where $\hat{\mathbf{x}} \in \mathbb{R}^N$ is the clean signal defined in the frequency domain. Furthermore, the clean frequency signal $\hat{\mathbf{x}}$ is considered to be unknown but contains properties in its related graph (e.g. graph signal bandlimitedness, smoothness and stationary) which are exploited. Graph signals $\mathbf{U}_0\hat{\mathbf{x}}$ or $\mathbf{U}_1\hat{\mathbf{x}}$ are contaminated with zero mean white Gaussian noise $\mathbf{n} \in \mathbb{R}^N$, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. In the work of [11], it is assumed that $\mathcal{G}_0$ and $\mathcal{G}_1$ are known. Their hypothesis testing problem is formulated as:

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathbf{y} = \mathbf{U}_0\hat{\mathbf{x}} + \mathbf{n} \\ \mathcal{H}_1 : &\quad \mathbf{y} = \mathbf{U}_1\hat{\mathbf{x}} + \mathbf{n}. \end{aligned} \tag{2.16}$$

Regarding the problem in this work, it is assumed that solely $\mathcal{G}_0$ is known as specified in Chapter 1. Now based on the noisy observation signal $\mathbf{y}$, the objective is to find out if $\mathbf{y}$ fits to $\mathcal{G}_0$ or to a yet unknown graph. This gives rise to the notion of blind topology change detection and changes problem (2.16) into:

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathbf{y} = \mathbf{U}_0\hat{\mathbf{x}} + \mathbf{n} \\ \mathcal{H}_1 : &\quad \mathbf{y} \neq \mathbf{U}_0\hat{\mathbf{x}} + \mathbf{n}. \end{aligned} \tag{2.17}$$

Now, by exploiting different properties of $\hat{\mathbf{x}}$ (or $\mathbf{x}$), such as band limitedness, smoothness and stationarity, the task is to detect unknown graph topology changes in a blind fashion. In Figure 2.5, a block scheme representation is given for both problems.

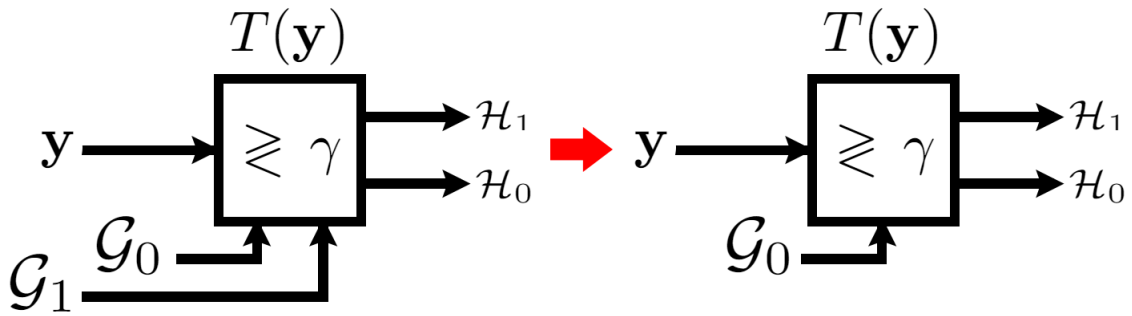


Figure 2.5: The scheme on the left represents the detector where information about both graphs is utilized. The scheme on the right represents the blind detection case where only information of $\mathcal{G}_0$ is used. In both cases, properties about $\hat{\mathbf{x}}$ are exploited.

3

Exploiting bandlimited signals

This chapter is dedicated to the study of the blind graph topology change detection by exploiting the bandlimitedness of graph signals. The chapter is organized as follows: Section 3.1 firstly gives a short introduction on how bandlimitedness can be exploited using a form of graph filtering. Subsequently, the first detector is derived. In Section 3.2, the detector derived by [11] is given. Their detector is used as reference for the performance analysis which is carried out in Section 3.3.

3.1. Blind topology change detector

Before the derivations of the blind topology change detector (BTCD) are given, the band limitedness of $\hat{\mathbf{x}}$ is exploited via the principle given in Section 2.1.1. The following transformation matrix is applied to achieve this:

$$\begin{aligned}\hat{\mathbf{y}}_K^c = \mathbf{U}_{0,K}^{c\,H}\mathbf{y} &= \mathbf{U}_{0,K}^{c\,H}\mathbf{U}_0\hat{\mathbf{x}} + \mathbf{U}_{0,K}^{c\,H}\mathbf{n} \\ &\triangleq \boldsymbol{\theta} + \hat{\mathbf{n}}_K^c\end{aligned} \tag{3.1}$$

where the transformation matrix $\mathbf{U}_{0,K}^c \in \mathbb{R}^{N\times(N\text{-}K)}$ contains the eigenvectors corresponding to the out of band frequencies ($\lambda_i > \lambda_K$). This matrix $\mathbf{U}_{0,K}^c$ consist of high pass filtering on the graph.

From (3.1), vectors $\hat{\mathbf{y}}_K^c \in \mathbb{R}^{N\text{-}K}$ and $\hat{\mathbf{n}}_K^c \in \mathbb{R}^{N\text{-}K}$ represents the out of band frequency content of the observed signal $\mathbf{y}$ and noise $\mathbf{n}$, respectively. Vector $\boldsymbol{\theta} = \mathbf{U}_{0,K}^{c\,H}\mathbf{U}_0\hat{\mathbf{x}} \in \mathbb{R}^{N\text{-}K}$ represents the out of band frequency content of $\mathbf{x}$. This parameter vector $\boldsymbol{\theta}$ is considered to be unknown due to $\hat{\mathbf{x}}$ and/or unknown topological changes which might occur under $\mathcal{H}_1$. The noise vector is now distributed as $\hat{\mathbf{n}}_K^c \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{U}_{0,K}^{c\,H}\mathbf{U}_{0,K}^c)$ which yields that $\hat{\mathbf{y}}_K^c$ is distributed as:

$$\hat{\mathbf{y}}_K^c \sim \mathcal{N}\big(\boldsymbol{\theta}, \boldsymbol{\Sigma}\big) \quad \text{with} \quad \boldsymbol{\Sigma} = \sigma^2\mathbf{U}_{0,K}^{c\,H}\mathbf{U}_{0,K}^c. \tag{3.2}$$

This distribution will be denoted by $p(\hat{\mathbf{y}}_K^c; \boldsymbol{\theta}, \mathcal{H}_0)$ as the bandlimited graph signal is observed under the nominal graph hypothesis. Under $\mathcal{H}_1$, the same model is used which is denoted by $p(\hat{\mathbf{y}}_K^c; \boldsymbol{\theta}, \mathcal{H}_1)$ where $\boldsymbol{\theta}$ is once again the unknown parameter vector. With this in place, the detection problem (2.16) is again reformulated into the following equivalent composite binary parameter test in terms of signal energy in $\boldsymbol{\theta}$ [8]. Under $\mathcal{H}_0$, no signal energy is expected out of band. Under $\mathcal{H}_1$, energy might be present due to topological changes:

$$\begin{aligned}\mathcal{H}_0 :&\quad \boldsymbol{\theta}^H\boldsymbol{\theta} = 0 \\ \mathcal{H}_1 :&\quad \boldsymbol{\theta}^H\boldsymbol{\theta} > 0.\end{aligned} \tag{3.3}$$

To exploit the bandlimitedness of graph signal $\hat{\mathbf{x}}$, the (known) parameter K has to be found. First, an energy percentage $E_{\%}$ was defined. The energy percentage is bounded with $0 < E_{\%} < 1$ and indicates how much energy one wants to capture in the band. Define $\hat{\mathbf{y}}_0 = \mathbf{U}_0^H\mathbf{y}$ as the projection of the observation signal on the eigenvectors of $\mathcal{G}_0$. Subsequently, define $\hat{\mathbf{y}}_{0,i} \in \mathbb{R}^i$ with $i \in \{1, \cdots, N\}$. This vector is a truncated version of $\hat{\mathbf{y}}_0$ where i decides the length of this vector, i.e., $\hat{\mathbf{y}}_{0,i} \in \hat{\mathbf{y}}_0$. Now

by using the definition of $E_{\%}$ and $\hat{\mathbf{y}}_{0,i}$, the following rule was used to estimate K from the observation signal $\mathbf{y}$:

$$K = \max \left\{ 1 < i < N \quad \middle| \quad \frac{\hat{\mathbf{y}}_{0,i}^H \hat{\mathbf{y}}_{0,i}}{\hat{\mathbf{y}}_0^H \hat{\mathbf{y}}_0} \leq E_{\%} \right\}, \tag{3.4}$$

which means that K is found, whereas the fraction of the in band energy ($\hat{\mathbf{y}}_{0,i}^H \hat{\mathbf{y}}_{0,i}$) and total energy ($\hat{\mathbf{y}}_0^H \hat{\mathbf{y}}_0$) is maximized and stays bounded by $E_{\%}$.

The distributions of $\hat{\mathbf{y}}_K^c$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ are the same except that the value of the unknown parameter vector $\boldsymbol{\theta}$ is different. With this in mind, the generalized likelihood ratio test (GLRT) $L_G(\hat{\mathbf{y}}_K^c)$ is used to derive the detector [8]. Vector $\boldsymbol{\theta}$ under $\mathcal{H}_0$, denoted by $\boldsymbol{\theta}_0$, is set to zero. Vector $\boldsymbol{\theta}$ under $\mathcal{H}_1$, denoted by $\boldsymbol{\theta}_1$, is replaced by its MLE $\boldsymbol{\theta}_1^*$. Vector $\boldsymbol{\theta}_0$ is set to zero because the out of band frequency content is assumed to be zero under $\mathcal{H}_0$. Vector $\boldsymbol{\theta}_1$ needs to be estimated fr4om the data since its filtered (with $\mathbf{U}_{0,K}^c$) frequency content is not necessarily zero (this estimate will not contain knowledge about $\mathcal{G}_1$). Finally, the GLRT is compared with a threshold γ, which is determined by a fixed false alarm rate P_{FA}, and a decision is made. More formally:

$$\begin{aligned} \boldsymbol{\theta}_0 &= \mathbf{0} \\ \boldsymbol{\theta}_1^* &= \arg\max_{\boldsymbol{\theta}} \quad p(\hat{\mathbf{y}}_K^c; \boldsymbol{\theta}, \mathcal{H}_1) = \hat{\mathbf{y}}_K^c. \end{aligned} \tag{3.5}$$

$$L_G(\hat{\mathbf{y}}_K^c) = \frac{p(\hat{\mathbf{y}}_K^c; \boldsymbol{\theta}_1^*, \mathcal{H}_1)}{p(\hat{\mathbf{y}}_K^c; \boldsymbol{\theta}_0, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \tag{3.6}$$

Working out (3.6), gives the BTCD (the derivations are reported in appendix A):

$$T_{\text{BTCD}}(\hat{\mathbf{y}}_K^c) = \hat{\mathbf{y}}_K^{cH} \boldsymbol{\Sigma}^{-1} \hat{\mathbf{y}}_K^c. \tag{3.7}$$

In Figure 3.2 a block scheme is given where $\mathbf{U}_{0,K}^c$ is depicted as a high pass filter:

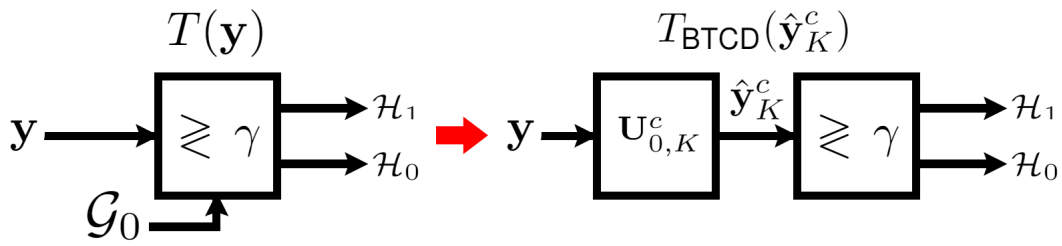


Figure 3.1: Block scheme of the BTCD. Matrix $\mathbf{U}_{0,K}^c$ is depicted here as a high pass filter.

The BTCD, basically determines the out of band energy of an observation signal $\mathbf{y}$. Based on this measured energy, a decision is made between $\mathcal{H}_0$ and $\mathcal{H}_1$. Due to the quadratic nature of the BTCD, specified by $\hat{\mathbf{y}}_K^c$, the following holds for the statistics of the BTCD (see (3.2) for the distribution of $\hat{\mathbf{y}}_K^c$):

$$\begin{aligned} \mathcal{H}_0: &\quad T_{\text{BTCD}}(\hat{\mathbf{y}}_K^c) \sim \chi^2_{N\text{-}K} \\ \mathcal{H}_1: &\quad T_{\text{BTCD}}(\hat{\mathbf{y}}_K^c) \sim \chi^2_{N\text{-}K}(\delta) \end{aligned} \tag{3.8}$$

which means that the test statistic $T_{\text{BTCD}}(\hat{\mathbf{y}}_K^c)$ follows a Chi squared distribution $\chi^2_{N\text{-}K}$ with N-K degrees of freedom under $\mathcal{H}_0$, and under $\mathcal{H}_1$, a non-central Chi squared distribution $\chi^2_{N\text{-}K}(\delta)$ again with N-K degrees of freedom but now with non-centrality parameter δ which is defined by:

$$\delta = \boldsymbol{\theta}_1^H \boldsymbol{\Sigma}^{-1} \boldsymbol{\theta}_1. \tag{3.9}$$

The latter will of course affect the theoretical derivations for the probability of detection P_D but not those for the probability of false alarm P_{FA}.

This affirms that graph topological changes are reflected in the statistics of the observed graph signal data. The non-centrality parameter δ depends on $\boldsymbol{\theta}$ which confirms the previous statement. The non-centrality parameter δ indicates whether the out of band energy is present or not, i.e., whether graph topological changes have occurred or not. The test statistic measures the out of band energy and based on this, a decision is made. If the BTCD is compared with a predefined threshold γ', the probability of false alarm $P_{FA} = P(\mathcal{H}_1; \mathcal{H}_0)$ is given by the right tail function of the central Chi squared distribution, i.e., $Q_{\chi^2_{N\text{-}K}}$. The threshold γ' is then given by the inverse right tail function, i.e., $Q^{-1}_{\chi^2_{N\text{-}K}}$:

$$\begin{aligned} P_{FA} = P(\mathcal{H}_1; \mathcal{H}_0) &= Pr\{T_{\mathsf{BTCD}}(\hat{\mathbf{y}}^c_K) > \gamma'; \mathcal{H}_0\} = Q_{\chi^2_{N\text{-}K}}(\gamma') \\ \Rightarrow \quad \gamma' &= Q^{-1}_{\chi^2_{N\text{-}K}}(P_{FA}). \end{aligned} \tag{3.10}$$

The probability of detection $P_D = P(\mathcal{H}_1; \mathcal{H}_1)$ is given by the right tail function of the non-central chi squared distribution, i.e., $Q_{\chi^2_{N\text{-}K}(\delta)}$:

$$\begin{aligned} P_D = P(\mathcal{H}_1; \mathcal{H}_1) &= Pr\{T_{\mathsf{BTCD}}(\hat{\mathbf{y}}^c_K) > \gamma'; \mathcal{H}_1\} = Q_{\chi^2_{N\text{-}K}(\delta)}(\gamma') \\ &= Q_{\chi^2_{N\text{-}K}(\delta)}\Big(Q^{-1}_{\chi^2_{N\text{-}K}}(P_{FA})\Big). \end{aligned} \tag{3.11}$$

It should be noticed that for the evaluation of the test statistic, the MLE of $\boldsymbol{\theta}_1$ (i.e., $\boldsymbol{\theta}^*_1$) is needed into the expression of the non centrality parameter (3.9). This estimate is then denoted by δ^*:

$$\delta^* = \hat{\mathbf{y}}^{c\,H}_K \boldsymbol{\Sigma}^{-1} \hat{\mathbf{y}}^c_K. \tag{3.12}$$

3.2. Simple matched subspace detector

The simple matched subspace detector (SMSD) is the detector described by [11]. The derivation of this detector starts from the detection problem given by (2.16) where both graph structures are assumed to be known. Instead of using $\mathbf{U}_{0,K}^{c}$ (as for the BTCD), the matrices $\mathbf{U}_{0,K}$ and $\mathbf{U}_{1,K}$ are used. These matrices contain the in band eigenvectors of the Laplacians of $\mathcal{G}_0$ and $\mathcal{G}_1$, respectively. These eigenvectors correspond to the in band frequencies ($\lambda_i \leq \lambda_K$). By using $\mathbf{U}_{0,K}$ and $\mathbf{U}_{1,K}$, the filtered signals $\hat{\mathbf{y}}_{0,K}$ and $\hat{\mathbf{y}}_{1,K}$ are obtained:

$$\begin{aligned} \hat{\mathbf{y}}_{0,K} &= \mathbf{U}_{0,K}^{H}\mathbf{y} \\ \hat{\mathbf{y}}_{1,K} &= \mathbf{U}_{1,K}^{H}\mathbf{y}. \end{aligned} \tag{3.13}$$

These matrices consist now of low pass filtering on the graphs. In contrast to the BTCD, where the band K was found with (3.4), the band in [11] is fixed. In their work, parameter K was assumed to be known. Finally, the SMSD is defined as the energy difference between $\hat{\mathbf{y}}_{0,K}$ and $\hat{\mathbf{y}}_{1,K}$:

$$T_{\text{SMSD}}(\hat{\mathbf{y}}_{0,K}, \hat{\mathbf{y}}_{1,K}) = \hat{\mathbf{y}}_{0,K}^{H}\hat{\mathbf{y}}_{0,K} - \hat{\mathbf{y}}_{1,K}^{H}\hat{\mathbf{y}}_{1,K}. \tag{3.14}$$

Hence, the SMSD determines a difference in projection energy on the first K graph Laplacian eigenvectors of each hypothetical graph [11]. The block scheme of the SMSD is given below:

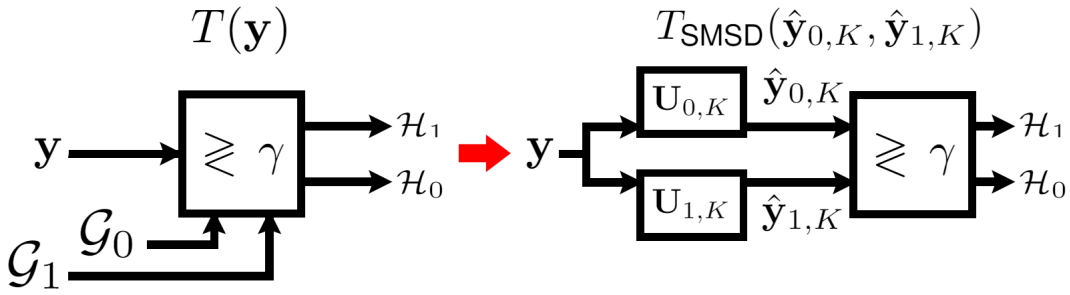


Figure 3.2: Block scheme of the SMSD. Matrices $\mathbf{U}_{0,K}$ and $\mathbf{U}_{1,K}$ are depicted here as low pass filters.

With numerical simulations it has been verified that it is valid to make the following approximation for the distribution of T_{SMSD}:

$$\begin{aligned} \mathcal{H}_0: &\quad T_{\text{SMSD}}(\hat{\mathbf{y}}_{0,K}, \hat{\mathbf{y}}_{1,K}) \sim \mathcal{N}(\boldsymbol{\mu}_0, \sigma^2\mathbf{I}) \\ \mathcal{H}_1: &\quad T_{\text{SMSD}}(\hat{\mathbf{y}}_{0,K}, \hat{\mathbf{y}}_{1,K}) \sim \mathcal{N}(\boldsymbol{\mu}_1, \sigma^2\mathbf{I}). \end{aligned} \tag{3.15}$$

This means that the statistics of the SMSD can be approximated using normal distributions. This approximation only holds if the number of nodes N is large enough [21]. It is remarkable that derivation (3.15) is not obtained from [11] but in this work. With this approximation, the threshold γ and probability of detection P_D can be specified by the following analytic expressions specified by the right tail and inverse right tail functions of a normal distribution, i.e., $Q_{\mathcal{N}}$ and $Q_{\mathcal{N}}^{-1}$:

$$\begin{aligned} P_{FA} &= P(\mathcal{H}_1; \mathcal{H}_0) = Pr\{T_{\text{SMSD}}(\mathbf{y}) > \gamma; \mathcal{H}_0\} = Q_{\mathcal{N}}(\gamma) \\ \Rightarrow \quad \gamma &= Q_{\mathcal{N}}^{-1}(P_{FA}). \end{aligned} \tag{3.16}$$

$$\begin{aligned} P_D &= P(\mathcal{H}_1; \mathcal{H}_1) = Pr\{T_{\text{SMSD}}(\mathbf{y}) > \gamma; \mathcal{H}_1\} = Q_{\mathcal{N}}(\gamma) \\ &= Q_{\mathcal{N}}\Big(Q_{\mathcal{N}}^{-1}(P_{FA})\Big). \end{aligned} \tag{3.17}$$

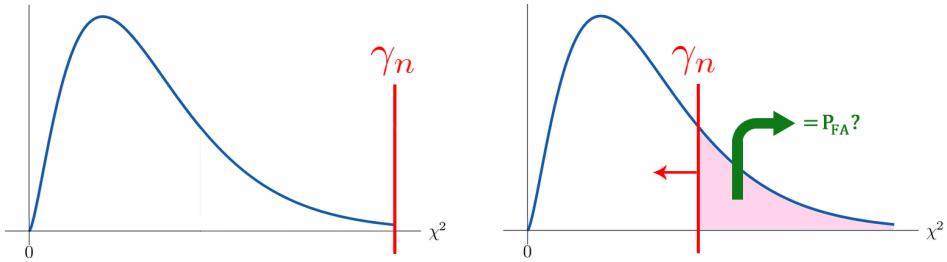
3.3. Numerical evaluation

In this section, the simulation scenario is firstly discussed. In Figure 3.4, a flow chart is given describing the process. A similar procedure used in [11] is adopted to mimic the same simulation scenario for a fair comparison. Two for-loops (red and blue) are depicted, which form the basis of the simulation process. The first for-loop (red), is used to average over $Z = 100$ runs. This loop is used to account for the randomness in the graph models and graph signals (generated in the blue for-loop), which is explained in Section 3.3.1. After exiting this loop, an average performance is determined using a sample mean ROC curve, i.e., $\overline{ROC}$ defined as:

$$\overline{ROC} = \frac{1}{Z} \sum_{i_z=1}^{Z} [ROC]_{i_z}. \tag{3.18}$$

In the second for-loop (blue), $M = 10^5$ noise realizations are generated. Prior to these realizations, the (stochastic) graph models ($\mathcal{G}_0$ and $\mathcal{G}_1$) and spectral signal $\hat{\mathbf{x}}$ are generated. After exiting the second for-loop, the numerical and theoretical detection thresholds γ_n and γ_t are determined. The detection threshold γ_n is determined numerically where the following steps were taken to achieve this (the last two steps are also shown in Figure 3.3):

- First, a histogram of the test statistic under $\mathcal{H}_0$ is created using the M noise realizations.
- Secondly, an initial detection threshold γ_n is placed on the end of the right tail of the histogram.
- Finally, threshold γ_n is moved (with a fixed small step size) towards the left tail of the histogram. This is done until the desired probability of false alarm P_{FA} is captured.



(a) An initial detection threshold γ_n is placed on the end of the right tail of the histogram.

(b) Threshold γ_n is moved towards the left tail of the histogram.

Figure 3.3: Graphic overview of finding the detection threshold γ_n numerically.

As mentioned in Section 2.2, the thresholds are found by keeping the probability of false alarm P_{FA} fixed. Threshold γ_t is determined using (3.10) where the inverse right tail function of the central Chi squared distribution is used. Also two probabilities of detection, $P_{D,n}$ and $P_{D,t}$, are determined at the end of this loop. The probability of detection $P_{D,n}$ is determined by simply counting how many times a detection threshold (γ_n or γ_t) is exceeded [8]. The probability of detection $P_{D,t}$ is determined theoretically using the right tail function of the Chi squared distribution (3.11) with either γ_n or γ_t as its argument.

Table 3.1 is given to summarize the evaluation possibilities in terms of detection probabilities. Also the observed relative error sensitivity is provided (this is provided to give a sense of the simulation errors which might corrupt the results). The relative error sensitivity is based on the potential error that can be made during the numerical procedures in finding the numerical threshold γ_n and numerical detection probability $P_{D,n}$:

Table 3.1: Relative error sensitivity of probability of detection (numerical of theoretical) using the detection threshold (numerical or theoretical)

probability of detection P_D	relative error sensitivity
$P_{D,t}(\gamma_t)$	low
$P_{D,t}(\gamma_n)$	medium
$P_{D,n}(\gamma_t)$	medium
$P_{D,n}(\gamma_n)$	high

Regarding Figure 3.4, also some internal steps concerning the noise realizations are shown on the right (green). The first step specifies that the noise is modeled as $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. Thereafter, how observation signals under both hypotheses $\mathbf{y}_0$ and $\mathbf{y}_1$ are generated as specified in (2.16). The values T_0 and T_1 are the detector outcomes determined by (3.7) with $\mathbf{y}_0$ and $\mathbf{y}_1$ as the arguments, respectively.

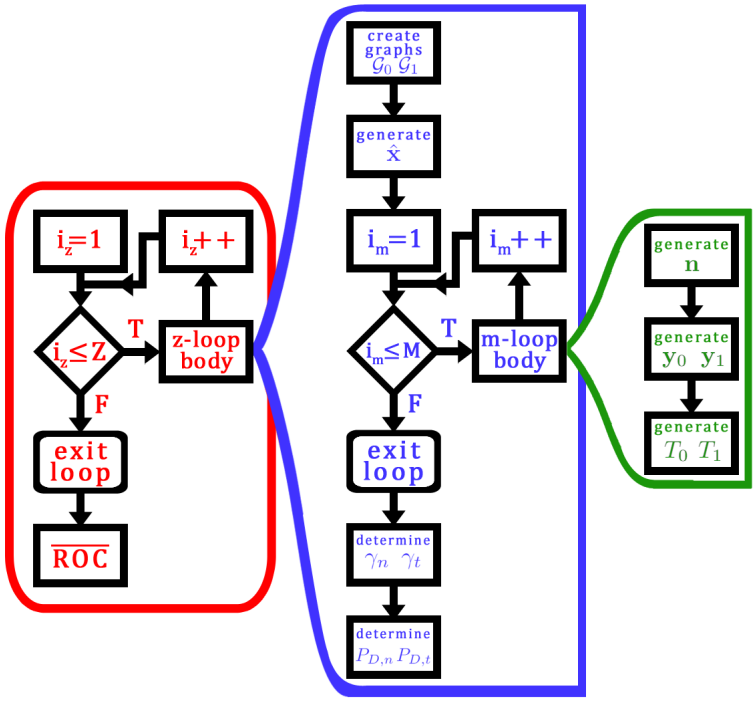


Figure 3.4: Flow chart of the simulation process: it contains three for-loops describing the simulation process. The loop on the left (red) describes the for-loop for the $Z = 100$ runs for averaging. The loop in the middle (blue) represents the for-loop regarding the $M = 10^5$ noise realizations. In this loop, also the graphs, $\mathcal{G}_0$, $\mathcal{G}_0$, and the graph signal spectrum $\hat{\mathbf{x}}$ are generated. The steps on the right (green), give a more detailed view on the noise realizations. The observation signals $\mathbf{y}_0$ and $\mathbf{y}_1$ are generated according to (2.16). The values for T_0 and T_1 are created according (3.7) with $\mathbf{y}_0$ and $\mathbf{y}_1$ as their arguments, respectively.

The remaining part of this section is organized as follows: in Section 3.3.1, synthetic data are used to evaluate the detectors. The SMSD is used as a reference for the evaluation of the BTCD. This section consists of two parts. The first part considers the scenario where the graph models, are the same as in [11]. In the second part of this section, smaller topological changes are considered. The reason why smaller topological changes are considered, is the following: when random graph models are used to model $\mathcal{G}_0$ and $\mathcal{G}_1$, their mutual differences can be too large which could lead to a biased result. Detecting graph topological changes , since the graphs may differ entirely from eachother.

In Section 3.3.2, real data are used to evaluate and to compare the detectors. The data set contains recorded meteorological data in France[1] and is also used in [15]. The locations of the weather stations are used for the node set. Links between these nodes are placed according to graph learning methods [22]. Links are place in such a way, such that the spectrum stays as smooth as possible. The temperature progressions per weather station are used as graph signals. This data set was used to see whether bandlimitedness still acts as a property with sufficient discriminative power.

As a final remark: in this chapter, where the BTCD is considered, all analytic expressions were available, e.g., the expressions for the probability of detection and thresholds given by the right tail fucntions. Expressions (3.10) and (3.11) were used to get the detection threshold and probability of detection for the BTCD. For the SMSD, the expressions (3.16) and (3.17) were used. In the remainder of this section, "estimated P_D" means that $P_{D,n}(\gamma_t)$ is used and "theoretical P_D" means that $P_{D,t}(\gamma_t)$ is used (**these notions are used in the plots/results**).

[1] Access to the raw data is possible directly through
$https://donneespubliques.meteofrance.fr/donnees_libres/Hackathon/RADOMEH.tar.gz$

3.3.1. Synthetic graphs

In [11], both graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ are modeled as random small-world graph. The authors of [23] have shown how the interdependent connectivity of a small-world graph relates to cortical connectivity derived from anatomical studies. This knowledge is also exploited in [11]. Small-world graphs are modeled as follows: $\mathcal{G} = \mathcal{G}_{SW}(N, \bar{d}, p_r)$. Parameter N stands for the number of nodes, $\bar{d}$ for the average link degree per node and p_r for the probability for random rewiring of links [24]. The parameters that were used to create the small world graphs, are summarized in Table 3.2.

Table 3.2: Small world graph parameters

	N	$\bar{d}$	p_r
$\mathcal{G}_{SW,0}$	40	12	0.1
$\mathcal{G}_{SW,1}$	40	20	0.4

In [11], the frequency content $\hat{\mathbf{x}}$ is modelled as a realization of a normal distribution:

$$\hat{\mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\hat{x}}). \tag{3.19}$$

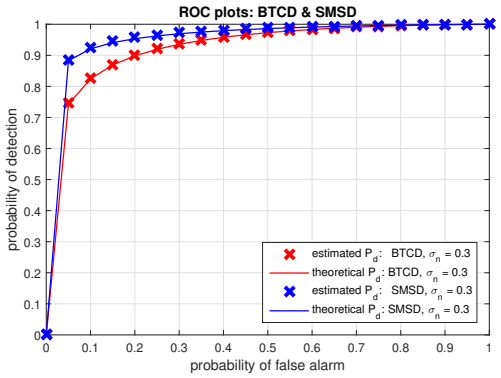
The GFT coefficient distribution of $\hat{\mathbf{x}}$ followed either a step function:

$$[\mathbf{\Sigma}_{\hat{x}}]_{ii} = \begin{cases} 1 & \forall \quad i \in \{1, 2, \cdots, K\} \\ 0.1 & \forall \quad i \in \{K{+}1, K{+}2, \cdots, N\}, \end{cases} \tag{3.20}$$

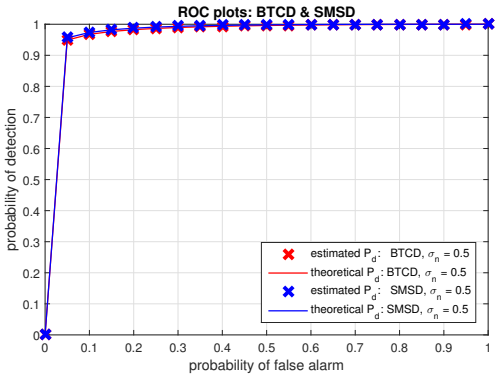
or an exponential decay:

$$[\mathbf{\Sigma}_{\hat{x}}]_{ii} = \exp(-0.2 \cdot i) \quad \forall \quad i \in \{1, 2, \cdots, N\}. \tag{3.21}$$

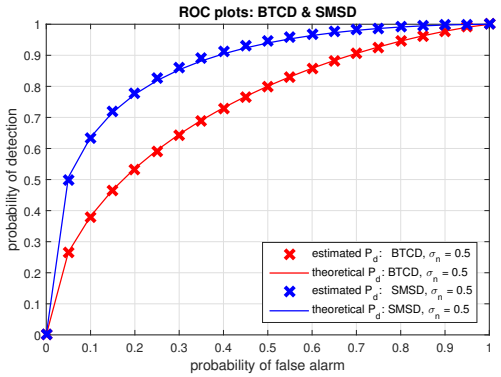
The bandwidth K was set to 12. To analyze the effect of the noise, two standard deviation values were used i.e. $\sigma \in \{0.3, 0.5\}$. In Figure 3.5, the performance of the BTCD versus the SMSD is given. Each sub figure represents one simulation scenario described by [11]. The parameters for each simulation scenario are listed under each ROC plot. The figures show four averaged $\overline{ROC}$ plots. The estimated P_D is $P_{D,n}(\gamma_t)$ (depicted as crosses) and the theoretical P_D is $P_{D,t}(\gamma_t)$ (depicted as a continuous line) as defined in section 3.3.



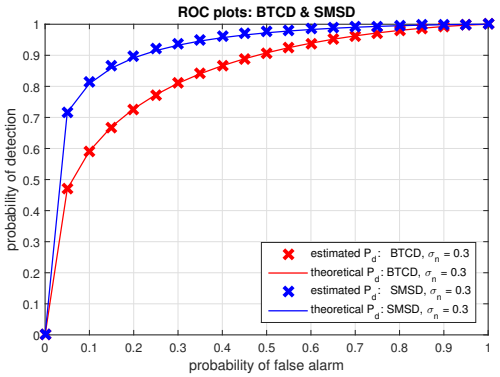
(a) $\mathcal{G}_0 = \mathcal{G}_{SW,0}$, $\mathcal{G}_1 = \mathcal{G}_{SW,1}$, exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.3$



(b) $\mathcal{G}_0 = \mathcal{G}_{SW,0}$, $\mathcal{G}_1 = \mathcal{G}_{SW,1}$, step $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.5$



(c) $\mathcal{G}_0 = \mathcal{G}_{SW,0}$, $\mathcal{G}_1 = \mathcal{G}_{SW,1}$, exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.5$



(d) $\mathcal{G}_0 = \mathcal{G}_{SW,0}$, $\mathcal{G}_1 = \mathcal{G}_{SW,0}$, exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.3$

Figure 3.5: Simulation settings described per each scenario [11]. Beneath each sub figure, the simulation settings are listed. The models used for $\mathcal{G}_0$ and $\mathcal{G}_1$ are given ($\mathcal{G}_{SW,0}$ or $\mathcal{G}_{SW,1}$). Also which covariance matrix is used to generate $\hat{\mathbf{x}}$ (exp. $\mathbf{\Sigma}_{\hat{x}}$ or step $\mathbf{\Sigma}_{\hat{x}}$) and finally, which noise standard deviation σ ($\sigma = 0.3$ or $\sigma = 0.5$).

From these $\overline{ROC}$ plots, the following can be concluded. Due to the fact that the BTCD only uses information of graph $\mathcal{G}_0$, a performance deterioration is seen. Only in Figure 3.5b an equivalent performance is attained. This plot shows that blind detection has a lot of potential as it seems to track the detector of [11]. In this simulation, the covariance matrix $\mathbf{\Sigma}_{\hat{x}}$ is modeled as step function. As result of this, a perfect band limited graph signal is modeled. This means that the performance of the BTCD depends on the band limitedness of graph signals. When $\mathbf{\Sigma}_{\hat{x}}$ is modeled as an exponential decay, the energy is smeared out and the graph signals become less band limited which directly affects the performance of the BTCD.

However, it is observed that the total ignorance of the graph $\mathcal{G}_1$ leads to a performance degradation up to 20% ($P_{FA} = 0.1$) in Figure 3.5c and 3.5d and a degradation up to 15% ($P_{FA} = 0.1$) in Figure 3.5a. Additionally, observe that the theoretical derivations for the BTCD and SMSD match perfectly with the numerical performance.

In the remaining part of this section, smaller changes concerning the graph topology are considered. To capture smaller graph topological changes, the following method was followed. The nominal graph $\mathcal{G}_0$ is used to create $\mathcal{G}_1$. The node set is kept the same but in the edge set existing links are randomly removed with a link loss probability P_{LL} (in this work only link removal is considered and new link additions were kept untested as they are expected to give a similar result):

$$\mathcal{G}_1 \rightarrow \mathcal{G}_1(\mathcal{G}_0, P_{LL}). \tag{3.22}$$

The same model was used for $\hat{\mathbf{x}}$ as given in (3.19). Only the exponential decay model for $\mathbf{\Sigma}_{\hat{x}}$ is analyzed since having out of band energy is a common case (using the step $\mathbf{\Sigma}_{\hat{x}}$, models a perfect bandlimited spectrum). In Figure 3.6, ROC plots are shown with increasing noise standard deviation, i.e., $\sigma = \{0.3,\ 0.5,\ 0.9\}$ while the link loss probability was kept constant as $P_{LL} = 0.25$.

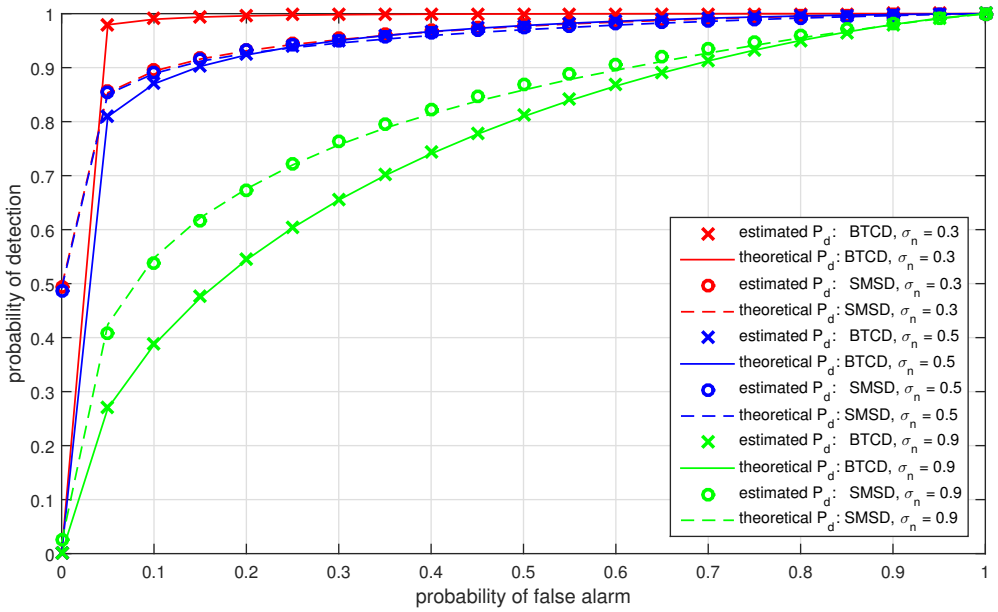


Figure 3.6: Simulation settings smaller changes: exp. $\mathbf{\Sigma}_{\hat{x}}$, $P_{LL} = 0.25$.

From Figure 3.6, the following conclusions can be drawn. The performance of both the BTCD and the SMSD tend to deteriorate as the noise power increases. The reason for this is that both detectors depend on projection energy which gets affected as the noise power increases. Furthermore, it should be noticed that there is a critical noise level where the BTCD tends to perform better than the SMSD. The BTCD uses the out of band energy for the decision making in contrast to the SMSD which uses the difference in the in band energy. For low noise levels, it seems that using out of band energy is more useful to detect topological changes leading the proposed approach (BTCD) to perform better.

In the last simulation the effect of the link loss probability P_{LL} is analyzed. In Figure 3.7, ROC plots are shown with increasing link loss probabilities, i.e., $P_{LL} = \{0.1,\ 0.25,\ 0.5\}$ while the noise standard deviation was kept constant to $\sigma = 0.5$.

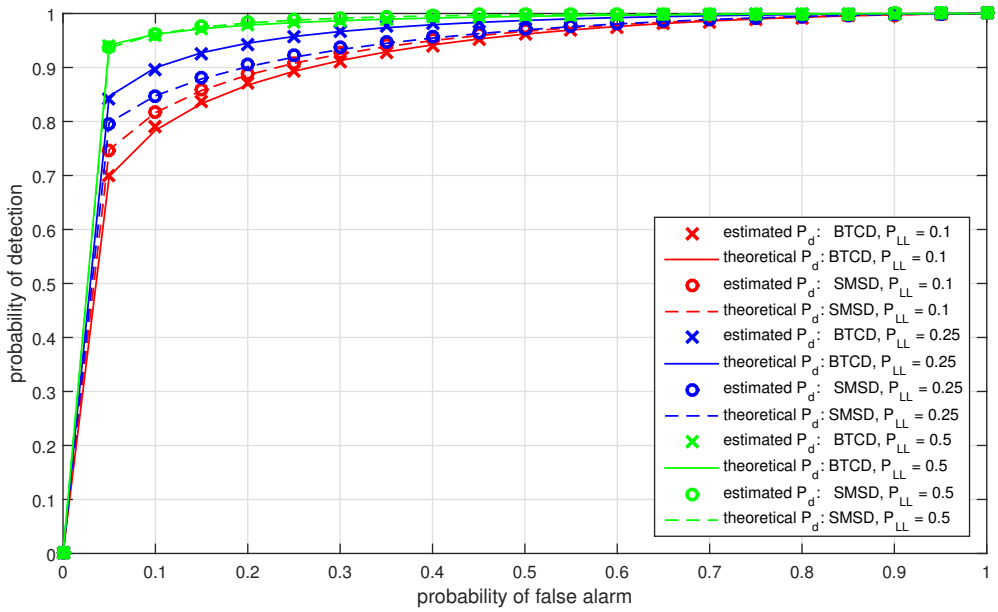


Figure 3.7: Simulation settings smaller changes: exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.5$.

From these plots, it can be seen that the performance of the BTCD and SMSD seems to be proportional to the amount of change in the graph topology. With P_{LL} close to zero, no links are selected for removal and $\mathcal{G}_1$ will be exactly the same as $\mathcal{G}_0$. This should be seen as the worst case scenario with the worst performance. On the other hand, if P_{LL} is close to one, then all links in $\mathcal{G}_0$ will be selected for removal and $\mathcal{G}_1$ will differ entirely w.r.t. $\mathcal{G}_0$. This is the best case scenario with the best performance. Furthermore, it can be seen that even with small topological changes, the proposed approach still performs well.

It is remarkable that for most of the P_{FA} values, the performance of the two detectors is approximately the same. This highlights the fact that information about $\mathcal{G}_1$ might not be needed to detect small graph changes.

3.3.2. Real graphs

As specified in Section 3.3, also real weather station data is used to evaluate the performance. The nominal graph $\mathcal{G}_0$ is obtained with graph learning methods [22]. The alternative graph $\mathcal{G}_1$ is again obtained by randomly removing existing links with a link loss probability P_{LL} as given in (3.22). The graph signal $\mathbf{x}$ in this case, are the weather progressions in time per node. In this case, the weather progression signal will not match $\mathcal{G}_1$ (this has been done intentionally) and the challenge now is to detect this mismatch.

In Figures 3.8 and 3.9 again the effects of increasing σ_n and P_{LL} are analyzed respectively. For the first simulation, the noise standard deviation again varied i.e. $\sigma_n = \{0.3,\ 0.5,\ 0.9\}$ while the link loss probability was kept constant as $P_{LL} = 0.25$.

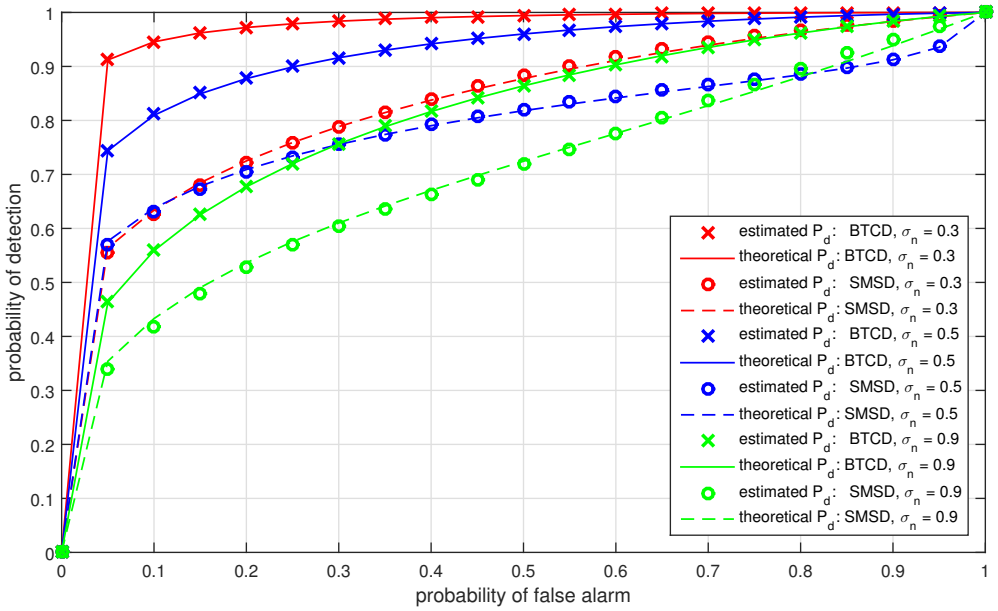


Figure 3.8: Simulation real data: $P_{LL} = 0.25$.

From these plots, it is again visible that the performance of both detectors is inversely proportional to the magnitude of the noise standard deviation. Furthermore, it should be noticed that the BTCD tend to perform better for all different noise levels. Especially, for $\sigma_n = 0.3$ where a big step in terms of P_D is obtained w.r.t. the SMSD approach. This can again be attributed to the fact that out of band energy is considered in the BTCD in contrast to the SMSD.

For the second simulation, the link loss probability P_{LL} is changed. The following values for the link loss probability are used: $P_{LL} = \{0.1,\ 0.25,\ 0.5\}$ while the noise standard deviation was kept constant to $\sigma_n = 0.5$.

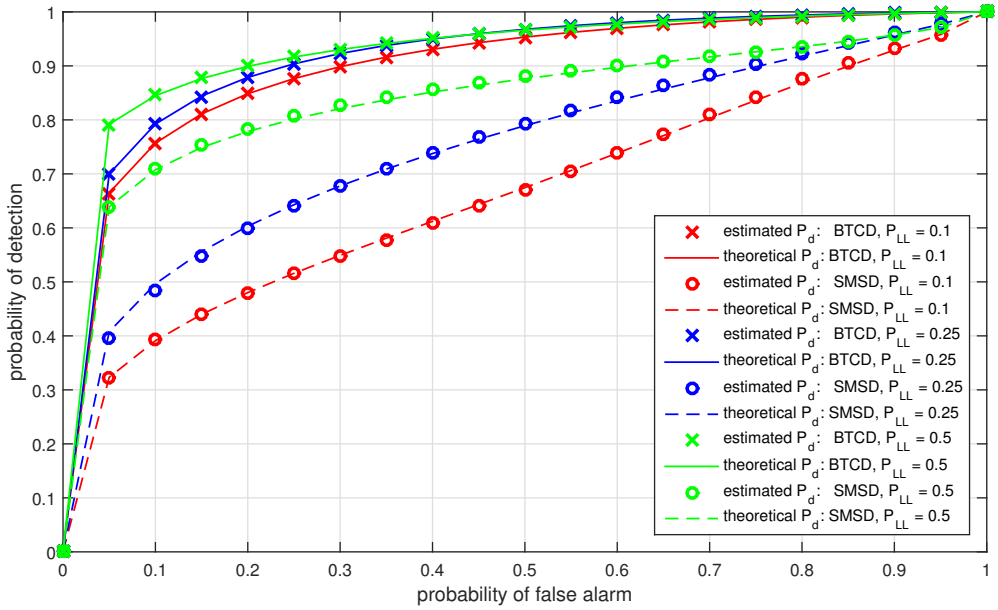


Figure 3.9: Simulation real data: $\sigma_n = 0.5$.

Also for this simulation, the same trend as given for earlier simulations is visible. As the probability for link loss increases, the performance of both detectors also tend to increase. Moreover, the performance of the BTCD tend to perform better for all different link loss probabilities. Furthermore, the performance of the SMSD seems to proportional to the amount of change. If the amount of change gets less, the worse the SMSD is performing. This phenomenon is not seen for the BTCD. The fact that the BTCD is performing better w.r.t. the SMSD, can also be attributed to the fact that out of band energy is used in the BTCD while the SMSD is using an in band energy difference. The in band energy difference does not change much. ... .. low pass so the performance ...

4

Exploiting smooth signals

This chapter is dedicated to study the blind detection possibilities by exploiting the smoothness of graph signals. This chapter is organized as follows: Section 4.1 first gives a short introduction on how graph signal smoothness can be exploited. Thereafter, a detector derived in this perspective. In Section 4.2, the detector derived by [11] is given where the smoothness of the graph signal is exploited as well. Their detector is again used as a reference to compare the performance in Section 4.3.

4.1. Constrained blind topology change detector

In this chapter, graph signal smoothness of $\hat{\mathbf{x}}$ is exploited and the constrained blind topology change detector (CTCD) is derived. This is done by using the smoothness constraint (2.6) as given in Section 2.1.2. First assume that the smoothness constraint $C_0(\hat{\mathbf{x}})$ is bounded under hypothesis $\mathcal{H}_0$ with a upper bound r:

$$\mathcal{H}_0: \quad C_0(\hat{\mathbf{x}}) = \frac{\hat{\mathbf{x}}^H \mathbf{\Lambda}_0 \hat{\mathbf{x}}}{\hat{\mathbf{x}}^H \hat{\mathbf{x}}} = \frac{\mathbf{x}^H \mathbf{L}_0 \mathbf{x}}{\mathbf{x}^H \mathbf{x}} \leq r. \tag{4.1}$$

Usually an eigenvalue of the Laplacian of the graph under consideration is chosen for r as this bounds the smoothness [2]. As said previously, this quantity is small when $\hat{\mathbf{x}}$ (in the vertex domain) tends to be (approximately) constant in adjacent vertices. When graph $\mathcal{G}_1$ is not known, the smoothness of $\hat{\mathbf{x}}$ can only be tested w.r.t. $\mathcal{G}_0$. Under $\mathcal{H}_1$, the observation signal can be: $\mathbf{y} \neq \mathbf{U}_0\hat{\mathbf{x}} + \mathbf{n}$ and/or $C_0(\hat{\mathbf{x}}) > r$. For this reason, the following hypothesis testing problem is formulated:

$$\begin{aligned} \mathcal{H}_0: &\quad \mathbf{y} = \mathbf{U}_0\hat{\mathbf{x}} + \mathbf{n} \quad \text{and} \quad C_0(\hat{\mathbf{x}}) = \frac{\hat{\mathbf{x}}^H \mathbf{\Lambda}_0 \hat{\mathbf{x}}}{\hat{\mathbf{x}}^H \hat{\mathbf{x}}} \leq r \\ \mathcal{H}_1: &\quad \text{otherwise.} \end{aligned} \tag{4.2}$$

The noise $\mathbf{n} \in \mathbb{R}^N$ for this problem is again modeled as: $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. For the sake of ease, define the vector $\boldsymbol{\theta} \in \mathbb{R}^N$ as:

$$\boldsymbol{\theta} = \mathbf{U}_0\hat{\mathbf{x}}. \tag{4.3}$$

The smoothness constraint $C_0(\hat{\mathbf{x}})$ in (4.1), is expressed in terms of $\boldsymbol{\theta}$ as:

$$C_0(\boldsymbol{\theta}) = \frac{\boldsymbol{\theta}^H \mathbf{L}_0 \boldsymbol{\theta}}{\boldsymbol{\theta}^H \boldsymbol{\theta}} \leq r. \tag{4.4}$$

Before deriving the CTCD, some elaboration is needed concerning the smoothness upper bound r. This upper bound is adaptive in the data $\mathbf{y}$ and is motivated by the following: first, the clean frequency signal $\hat{\mathbf{x}}$ itself is not known. Secondly, the spread of the signal energy of $\hat{\mathbf{x}}$ (in its spectrum), is also not known in advance. However, in this chapter, the bound on the smoothness of $\hat{\mathbf{x}}$ under $\mathcal{H}_0$ is exploited and forms an essential part of the CTCD. For this reason, the (known) upper bound r on the smoothness is found with the following minimization rule using the data $\mathbf{y}$:

$$r(\mathbf{y}) = \min \left\{ \lambda_{0,i} \in \lambda(\mathbf{L}_0) \quad \middle| \quad \frac{\mathbf{y}^H \mathbf{L}_0 \mathbf{y}}{\mathbf{y}^H \mathbf{y}} \leq \lambda_{0,i} \right\}, \tag{4.5}$$

where $\lambda(\mathbf{L}_0)$ denotes the set with the eigenvalues of $\mathbf{L}_0$ as its elements. This means that r is the smallest eigenvalue of $\mathbf{L}_0$ for which the smoothness of $\hat{\mathbf{x}}$ stays bounded w.r.t. $\mathcal{G}_0$.

The distributions of $\mathbf{y}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ are both modeled with parametric models which contain unknown parameters specified by $\boldsymbol{\theta}$. These distributions are denoted by $p(\mathbf{y}; \boldsymbol{\theta}, \mathcal{H}_0)$ and $p(\mathbf{y}; \boldsymbol{\theta}, \mathcal{H}_1)$ respectively. Using the definition of $\boldsymbol{\theta}$, the distribution $p(\mathbf{y}; \boldsymbol{\theta}, \mathcal{H}_0)$, can be expressed as:

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}. \tag{4.6}$$

The vector $\boldsymbol{\theta}$ is not known due to the unknown signal $\hat{\mathbf{x}}$ under $\mathcal{H}_0$. Under $\mathcal{H}_1$, the vector $\boldsymbol{\theta}$ is unknown due to $\hat{\mathbf{x}}$ and the graph topology. Define $\boldsymbol{\theta}_0$ as $\boldsymbol{\theta}$ under $\mathcal{H}_0$ and $\boldsymbol{\theta}_1$ as $\boldsymbol{\theta}$ under $\mathcal{H}_1$. The parametric models of these distributions are then known except for parameter vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. With this in mind, a similar GLRT $L_G(\mathbf{y})$ is proposed to derive the detector [8]. Because $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are not known, they are replaced by their MLEs, i.e., $\boldsymbol{\theta}_0^*$ and $\boldsymbol{\theta}_1^*$. For $\boldsymbol{\theta}_1^*$, the MLE is simply:

$$\boldsymbol{\theta}_1^* = \arg\max_{\boldsymbol{\theta}} \quad p(\mathbf{y}; \boldsymbol{\theta}, \mathcal{H}_1) = \mathbf{y}. \tag{4.7}$$

Since the smoothness of $\hat{\mathbf{x}}$ is constrained under $\mathcal{H}_0$, the following constrained MLE problem is proposed to find $\boldsymbol{\theta}_0^*$:

$$\begin{aligned} \boldsymbol{\theta}_0^* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \quad & p(\mathbf{y}; \boldsymbol{\theta}, \mathcal{H}_0) \\ \text{s.t.} \quad & C_0(\boldsymbol{\theta}) \quad \leq \quad r, \end{aligned} \tag{4.8}$$

which is equivalent to the following minimization problem (see Appendix B.1 for derivation):

$$\begin{aligned} \boldsymbol{\theta}_0^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \quad & (\mathbf{y} - \boldsymbol{\theta})^H (\mathbf{y} - \boldsymbol{\theta}) \\ \text{s.t.} \quad & \boldsymbol{\theta}^H \mathbf{L}_0 \boldsymbol{\theta} - r \cdot \boldsymbol{\theta}^H \boldsymbol{\theta} \quad \leq \quad 0. \end{aligned} \tag{4.9}$$

The solution of the problem given by (4.9) was found using the Karush-Kuhn-Tucker (KKT) conditions. These are necessary conditions for a solution of such an (constrained) optimization problem [25]. The expressions for the solution of $\boldsymbol{\theta}_0^*$ are found in Appendix B.2:

$$\boldsymbol{\theta}_0^* = [\mathbf{I} + \mu \sigma^2 \mathbf{L}_{0,r}]^{-1} \mathbf{y}. \tag{4.10}$$

The GLRT is again compared with a threshold γ to decide which hypothesis is true:

$$L_G(\mathbf{y}) = \frac{p(\mathbf{y}; \boldsymbol{\theta}_1^*, \mathcal{H}_1)}{p(\mathbf{y}; \boldsymbol{\theta}_0^*, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{4.11}$$

Working out the GLRT with a similar approach as given in the previous chapter, the CTCD is found (further derivations can be found in appendix B):

$$\begin{aligned} T_{\text{CTCD}}(\mathbf{y}) &= \mathbf{y}^H \boldsymbol{\Sigma}^{-1} [\mathbf{y} - \boldsymbol{\theta}_0^*] \\ &= \mathbf{y}^H \boldsymbol{\Sigma}^{-1} \mathbf{y} \quad - \quad \mathbf{y}^H \boldsymbol{\Sigma}^{-1} \boldsymbol{\theta}_0^* \end{aligned} \tag{4.12}$$

In Figure 4.1, a block scheme is given for the CTCD with the constraint $C_0(\boldsymbol{\theta})$ depicted in an MLE block. Vector $\boldsymbol{\theta}_0^*$ is given as the output of this block:

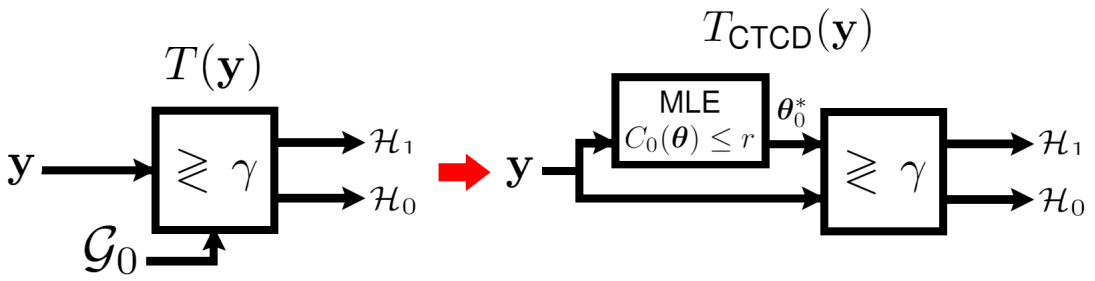


Figure 4.1: Block scheme of the CTCD with constraint $C_0(\boldsymbol{\theta})$ depicted in the MLE block. The MLE $\boldsymbol{\theta}_0^*$ is generated at the output of this block.

Using the CTCD (4.12), an energy difference is calculated between the unconstrained energy and the constrained energy of $\mathbf{y}$. This energy difference discriminates the two hypotheses. For the statistics of the CTCD, expression (B.15) is used for $\boldsymbol{\theta}_0^*$. With some rearrangement, the CTCD can then be expressed as follows:

$$T_{\text{CTCD}}(\mathbf{y}) = \mathbf{y}^H \mathbf{U}_0 \left[\frac{1}{\sigma^2} \left[\mathbf{I} - [\mathbf{I} + \mu\sigma^2 \boldsymbol{\Lambda}_{0,r}]^{-1} \right] \right] \mathbf{U}_0^H \mathbf{y}. \tag{4.13}$$

By defining $\hat{\mathbf{y}}_0 = \mathbf{U}_0^H \mathbf{y}$, it is easy to see that $\hat{\mathbf{y}}_0$ is distributed as: $\hat{\mathbf{y}}_0 \sim \mathcal{N}(\mathbf{U}_0^H \boldsymbol{\theta}, \sigma^2 \mathbf{I})$. The CTCD is then given by $T_{\text{CTCD}}(\hat{\mathbf{y}}_0)$:

$$T_{\text{CTCD}}(\hat{\mathbf{y}}_0) = \hat{\mathbf{y}}_0^H \left[\frac{1}{\sigma^2} \left[\mathbf{I} - [\mathbf{I} + \mu\sigma^2 \boldsymbol{\Lambda}_{0,r}]^{-1} \right] \right] \hat{\mathbf{y}}_0, \tag{4.14}$$

and in scalar form this becomes:

$$T_{\text{CTCD}}(\hat{\mathbf{y}}_0) = \sum_{i=1}^{N} \left(1 - \frac{1}{1 + \mu\sigma^2(\lambda_{0,i} - r)} \right) \frac{\hat{y}_{0,i}^2}{\sigma^2}. \tag{4.15}$$

With the distribution of $\hat{\mathbf{y}}_0$ in place, the following can be concluded for $\frac{1}{\sigma^2} \cdot \hat{y}_{0,i}^2$:

$$\frac{\hat{y}_{0,i}^2}{\sigma^2} \sim \chi_1^2(\delta_i) \quad \text{with} \quad \delta_i = \frac{1}{\sigma^2} \theta_i^2. \tag{4.16}$$

This means that the statistics of $T_{\text{CTCD}}(\mathbf{y})$ are given by a linear combination of non central Chi squared random variables with one degree of freedom and non centrality parameter δ_i. The non centrality parameter is different under each hypothesis because it is a function of the unknown vector $\boldsymbol{\theta}$. Once again, it can be concluded that graph topological changes are reflected in the statistics of the CTCD. The non-centrality parameter in this scenario indicates whether the energy of the observation signal $\mathbf{y}$ is constrained or not, i.e., whether graph topological changes have occurred or not. This can be said because $\boldsymbol{\theta}$ changes if topology changes occur. Under $\mathcal{H}_1$, (4.7) was used for $\boldsymbol{\theta}$ and (4.10) for $\boldsymbol{\theta}$ under $\mathcal{H}_0$. With this in place, the statistics of the CTCD (4.12) can be summarized as follows:

$$\begin{aligned} \mathcal{H}_0: \quad & T_{\text{CTCD}}(\mathbf{y}) \sim \quad \text{linear combination of} \quad \chi_1^2(\delta_i) \quad \text{with} \quad \delta_i = \frac{1}{\sigma^2} \theta_{0,i}^{*2} \\ \mathcal{H}_1: \quad & T_{\text{CTCD}}(\mathbf{y}) \sim \quad \text{linear combination of} \quad \chi_1^2(\delta_i) \quad \text{with} \quad \delta_i = \frac{1}{\sigma^2} \theta_{1,i}^{*2} = \frac{1}{\sigma^2} y_i^2. \end{aligned} \tag{4.17}$$

It is well known from the literature that a closed form expression for the distribution of a linear combination of Chi squared variables is challenging to achieve [21]. As a consequence, there are no closed form expressions for the detection threshold γ' and/or probability of detection P_D. These, however, can be approximated using proper approximation methods for the distribution of a linear combination of Chi squared random variables [26]. The procedure for finding the statistics of the CTCD for this thesis is described in Section 4.3.

4.2. Constrained simple matched subspace detector

In [11], the constrained simple matched subspace detector (CMSD) is derived. For the problem discussed by [11], the signal is considered to be smooth under both hypotheses in contrast to the scenario considered in this work, i.e.:

$$\mathcal{H}_0: \quad C_0(\hat{\mathbf{x}}) = \frac{\hat{\mathbf{x}}^H \mathbf{\Lambda}_0 \hat{\mathbf{x}}}{\hat{\mathbf{x}}^H \hat{\mathbf{x}}} \leq r_0 \quad \text{and} \quad \mathcal{H}_1: \quad C_1(\hat{\mathbf{x}}) = \frac{\hat{\mathbf{x}}^H \mathbf{\Lambda}_1 \hat{\mathbf{x}}}{\hat{\mathbf{x}}^H \hat{\mathbf{x}}} \leq r_1, \tag{4.18}$$

where the matrix $\mathbf{\Lambda}_i$ stands for the diagonal matrix containing the eigenvalues of the Laplacian of graph $\mathcal{G}_i$, $i \in \{0, 1\}$. Constants r_0 and r_1 stand for the smoothness bounds for the graph frequency content $\hat{\mathbf{x}}$ w.r.t. $\mathcal{G}_0$ and $\mathcal{G}_1$ respectively. First, define $\hat{\mathbf{y}}_0$ and $\hat{\mathbf{y}}_1$ as the projections of the observation signal $\mathbf{y}$ on the known graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ respectively:

$$\begin{aligned} \hat{\mathbf{y}}_0 &= \mathbf{U}_0^H \mathbf{y} \\ \hat{\mathbf{y}}_1 &= \mathbf{U}_1^H \mathbf{y}. \end{aligned} \tag{4.19}$$

Regarding to the problem given by (2.16), it is now only necessary to estimate $\hat{\mathbf{x}}$ under both hypotheses instead of $\boldsymbol{\theta}$ as this was done for the CTCD. The MLEs of $\hat{\mathbf{x}}$ were found by solving the following constrained MLE problems which are specified below [11]:

$$\begin{aligned} \hat{\mathbf{x}}_0^* = \underset{\hat{\mathbf{x}}}{\text{argmax}} \quad & p(\mathbf{y}; \hat{\mathbf{x}}, \mathcal{H}_0) \\ \text{s.t.} \quad & C_0(\hat{\mathbf{x}}) \quad \leq \quad r_0 \end{aligned} \tag{4.20}$$

$$\begin{aligned} \hat{\mathbf{x}}_1^* = \underset{\hat{\mathbf{x}}}{\text{argmax}} \quad & p(\mathbf{y}; \hat{\mathbf{x}}, \mathcal{H}_1) \\ \text{s.t.} \quad & C_1(\hat{\mathbf{x}}) \quad \leq \quad r_1 \end{aligned} \tag{4.21}$$

where $C_0(\hat{\mathbf{x}})$ and $C_1(\hat{\mathbf{x}})$ are specified by (4.18). In contrast to the proposed approach, where the smoothness bound r is made adaptive on the data $\mathbf{y}$, the smoothness bound in [11] is fixed. The following values are chosen for the bounds:

$$r(\mathbf{y}) = r_0 = \lambda_{0,20} \qquad \text{and} \qquad r(\mathbf{y}) = r_1 = \lambda_{1,20} \tag{4.22}$$

where $\lambda_{j,i}$ is the i^{th} eigenvalue of the Laplacian of graph $\mathcal{G}_j$. These specific choices for the bounds were chosen for simulation purposes [11]. Based on this constrained energy difference, a decision is made in the binary hypothesis testing problem given by (2.16). The constrained simple matched subspace detector (CMSD), described by [11], is given as follows:

$$T_{CMSD}(\hat{\mathbf{y}}_0, \hat{\mathbf{y}}_1) = (\hat{\mathbf{y}}_0 - \hat{\mathbf{x}}_0^*)^H (\hat{\mathbf{y}}_0 - \hat{\mathbf{x}}_0^*) - (\hat{\mathbf{y}}_1 - \hat{\mathbf{x}}_1^*)^H (\hat{\mathbf{y}}_1 - \hat{\mathbf{x}}_1^*) \tag{4.23}$$

The CMSD determines a difference in constrained projection energy of the observation $\mathbf{y}$ on each hypothetical graph. The constrained projection energy is determined by correcting each GFT coefficient in $\hat{\mathbf{y}}_0$ and $\hat{\mathbf{y}}_1$ with the GFT coefficients of the MLEs $\hat{\mathbf{x}}_0^*$ and $\hat{\mathbf{x}}_1^*$ [11]. In Figure 4.2 a block scheme of the CMSD is given:

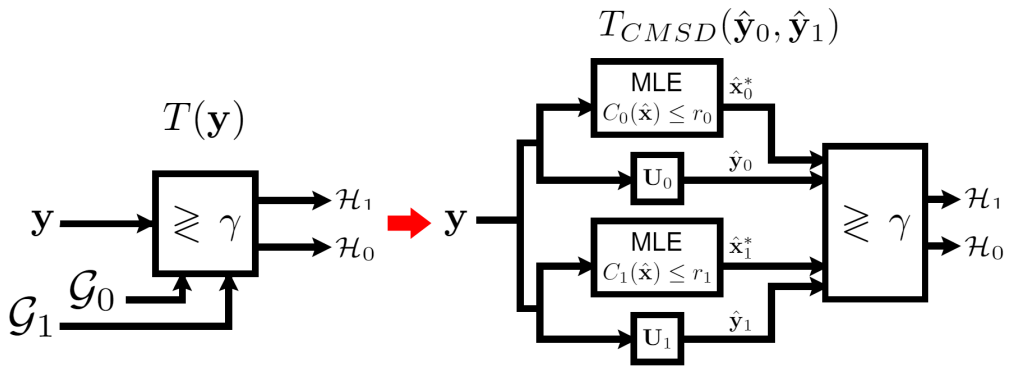


Figure 4.2: Block scheme of the CMSD with constraints $C_0(\hat{\mathbf{x}})$ and $C_1(\hat{\mathbf{x}})$ depicted in the MLE blocks. The MLEs $\hat{\mathbf{x}}_0^*$ and $\hat{\mathbf{x}}_1^*$ are generated at the output of these blocks.

For the distribution of the CMSD, no analytic expression was given by [11]. As a consequence, the distribution was approximated numerically in order to evaluate its performance. The procedure for finding the statistics of the CMSD is also described in Section 4.3.

4.3. Numerical evaluation

In this section, the performance of the CTCD is evaluated. A similar simulation process is used as presented in Section 3.3. In Figure 3.4, a flow chart is given describing this process. For the same reason as given previously, this specific simulation process is used to mimic the a similar simulation scenario as given by [11]. Again, $Z = 100$ runs were used to account for the randomness in the graph models and signals. Also $M = 10^5$ noise realizations were generated to reveal the statistics of the CTCD in order to determine the detection thresholds and detection probabilities. Compared to the flowchart that was used for the BTCD, an extra step was involved in the evaluation of the CTCD. In the M-body loop (green), an extra step was needed for finding $\boldsymbol{\theta}_0^*$. This adjustment is seen in Figure 4.3. For the final evaluation of the CTCD, again an average performance was determined using $\overline{ROC}$ given by (3.18).

The remaining part of this section is organized as follows: in Section 4.3.1, synthetic data are used to evaluate the detectors. This section consists of two parts. The first part considers the scenario where the same graph models, as given in [11], are used. In the second part, smaller topological changes are considered.

In Section 4.3.2, real data are used to evaluate and to compare the detectors. The data set contains recorded meteorological data in France[1] and is also used in [15]. The graph is build is the same way as given in section 3.3. The temperature progressions per weather station are again used as graph signals. This data set was used to see whether signal smoothness still acts as a property with sufficient discriminative power.

As a final remark, in this chapter, where the CTCD is considered, **no** analytic expressions were available. This means that all expressions for the detection probabilities and for the detection thresholds must be determined numerically. This means that they are found using the procedure described in Section 3.3. The detection thresholds for the CTCD and the CSMD was found numerically and is denoted by γ_n. These thresholds were then used for finding the detection probabilities numerically, i.e., $P_{D,n}(\gamma_n)$. Referring back to Table 3.1, it must be stated that this combination ($P_{D,n}(\gamma_n)$) gives the highest relative error sensitivity. What is meant by this is that the combination of using a numerically determined P_D and γ gives the highest risk of possible errors. In the remainder of this section, "estimated P_D" means that $P_{D,n}(\gamma_n)$ is used (**these notions are used in the plots/results**).

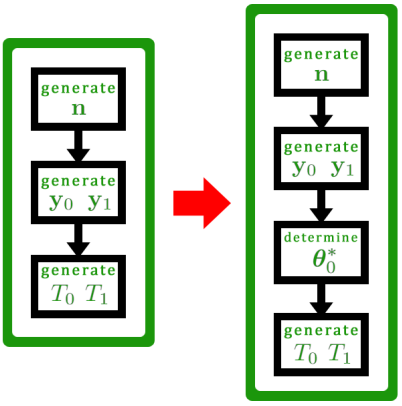
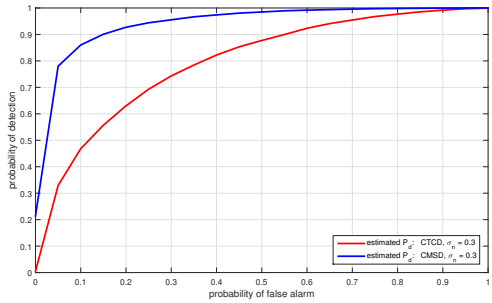


Figure 4.3: Flow chart adjustment: an extra step was included in the M-body loop for finding $\boldsymbol{\theta}_0^*$

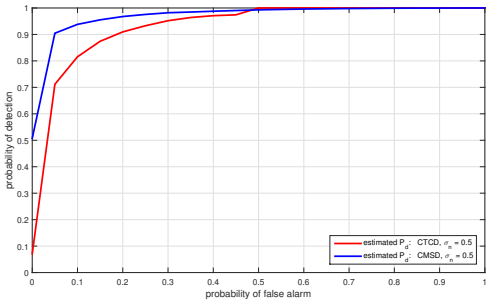
[1] Access to the raw data is possible directly through
$https://donneespubliques.meteofrance.fr/donnees_libres/Hackathon/RADOMEH.tar.gz$

4.3.1. Synthetic graphs

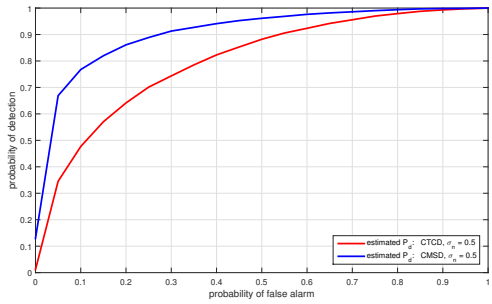
As said earlier, in [11], both graphs are modeled as random small-world graphs i.e. $\mathcal{G} = \mathcal{G}_{SW}(N, \bar{d}, p_r)$. The same parameters were used to create the small world graph and are summarized in Table 3.2. In Figure 4.4, four sub figures are seen, where each sub figure represents one simulation scenario described by [11].



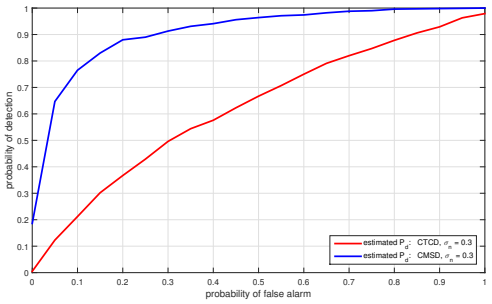
(a) $\mathcal{G}_0 = \mathcal{G}_{SW,0}$, $\mathcal{G}_1 = \mathcal{G}_{SW,1}$, exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.3$



(b) $\mathcal{G}_0 = \mathcal{G}_{SW,0}$, $\mathcal{G}_1 = \mathcal{G}_{SW,1}$, step $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.5$



(c) $\mathcal{G}_0 = \mathcal{G}_{SW,0}$, $\mathcal{G}_1 = \mathcal{G}_{SW,1}$, exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.5$



(d) $\mathcal{G}_0 = \mathcal{G}_{SW,0}$, $\mathcal{G}_1 = \mathcal{G}_{SW,0}$, exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma = 0.3$

Figure 4.4: Simulation settings described per each scenario [11]. Beneath each sub figure, the simulation settings are listed. The models used for $\mathcal{G}_0$ and $\mathcal{G}_1$ are given ($\mathcal{G}_{SW,0}$ and $\mathcal{G}_{SW,1}$). Also the used covariance matrix to generate $\hat{\mathbf{x}}$ (exp. $\mathbf{\Sigma}_{\hat{x}}$ or step $\mathbf{\Sigma}_{\hat{x}}$) is provided along with the noise standard deviations σ ($\sigma = 0.3$ or $\sigma = 0.5$).

From these simulations, the effect of not knowing the alternative graph $\mathcal{G}_1$ is again highly visible. The best performance of the CTCD is again attained in Figure 4.4b. This testing scenario should be seen as the easiest case since the covariance matrix of $\hat{\mathbf{x}}$ was modeled as step (see (3.20)). Furthermore, the worst performance is seen in Figure 4.4d, where both graphs were modeled as small-world graphs with the same parameters. Detection based on signal smoothness seems to be less discriminative as expected when the mutual differences between both graphs is too small.

In comparison with the BTCD, where detection was based on the bandlimitedness of graph signal $\hat{\mathbf{x}}$, a similar performance is attained except for the fact that the CTCD is more sensitive to little mutual differences (see Figure 3.5d).

Lastly, the CMSD seems to attain the same performance as presented in [11]. Therefore, it must be mentioned that although a numerical method was used to find the probability of detection, one can verify that the numerical approach was a correct one.

For the remaining simulations, smaller changes were considered as the one described in Section 3.3.1. A similar performance of the CTCD with respect to the CMSD was attained because the amount of change in between the two topologies under both hypotheses was large. This gave reason to investigate the effect of having small topological changes. In Figure 4.5, $\overline{ROC}$ plots are shown with increasing noise standard deviation i.e. $\sigma_n = \{0.3,\ 0.5,\ 0.9\}$ while the link loss probability was kept constant i.e. $P_{LL} = 0.25$.

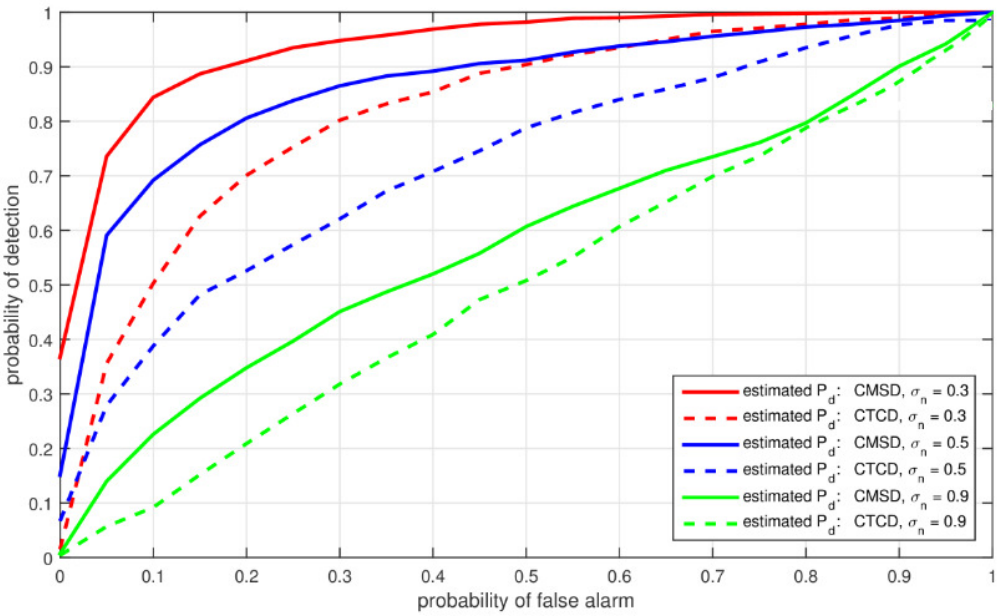


Figure 4.5: Simulation settings smaller changes: exp. $\mathbf{\Sigma}_{\hat{x}}$, $P_{LL} = 0.25$

From Figure 4.5, the following conclusions can be drawn. The performance of both the CTCD and the CMSD tend to deteriorate as the noise standard deviation increases. The reason for this is that both detectors depend on projection energy just as the BTCD. The main difference w.r.t. the BTCD is that the CTCD also considers how the signal energy is distributed over its spectrum, i.e., the energy decay.

For $\sigma_n = 0.3$ and $\sigma_n = 0.5$, the CTCD still shows detection capabilities. For these two noise levels, it can be concluded that energy decay stays bounded with the given bound r. For $\sigma = 0.9$, the CTCD performs the worst. The extra energy added by the noise affects the signal smoothness to such an extend, such that signal smoothness is not discriminative anymore.

Lastly, it is seen that the CMSD performs better than the CTCD under all noise levels which is the benefit of knowing the topology of the alternative graph. Detecting small graph topology changes seems to be more challenging for the CTCD w.r.t. the BTCD.

For the second simulation considering small topology changes, a slightly different set of link loss probabilities was used w.r.t. the BTCD. As given previously, the CTCD tend to perform bad when very small changes are considered. For this reason, the following values were chosen: $P_{LL} = \{0.1,\ 0.5,\ 0.9\}$ while the noise standard deviation was kept constant to $\sigma_n = 0.1$.

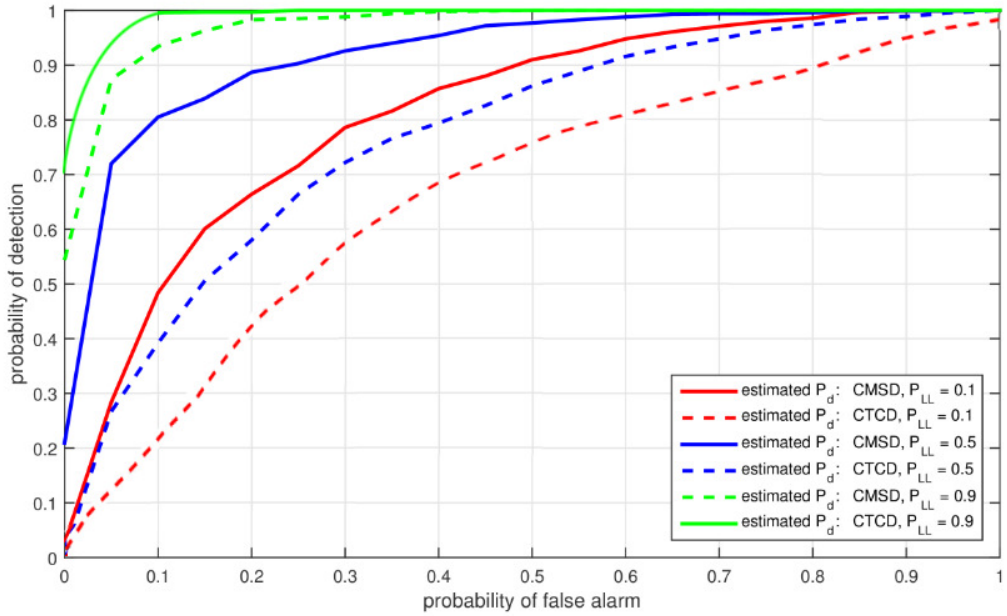


Figure 4.6: Simulation settings smaller changes: exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma_n = 0.1$

From Figure 4.6, it can be seen that the performance of the CTCD and CMSD seems again to be proportional to the amount of change in the graph topology. With P_{LL} close to zero, the worst case scenario is again attained. On the other hand, if P_{LL} is close to one, best case scenario is attained. One should learn form these plots, that the blind approach with the smoothness constraint is preferred for high topology changes only.

For this experiment, the noise level was kept low to see how the performance reacts to different link loss probabilities. In earlier experiments, it was shown how the smoothness get affected when the additive noise energy is high (see Figure 4.5). This is the reason why $\sigma_n = 0.1$ is used in this experiment.

4.3.2. Real graphs

In this section, real weather station data are used to evaluate the performance. The nominal graph $\mathcal{G}_0$ is obtained by graph learning methods. The alternative graph $\mathcal{G}_1$ is again obtained by randomly removing existing links with a link loss probability P_{LL}. In Figures 4.7 and 4.8 again the effects of increasing σ_n and P_{LL} are analyzed respectively. For the first simulation (Figure 4.7), the noise standard deviation again varied i.e. $\sigma = \{0.3,\ 0.5,\ 0.9\}$, while the link loss probability was kept constant to $P_{LL} = 0.5$.

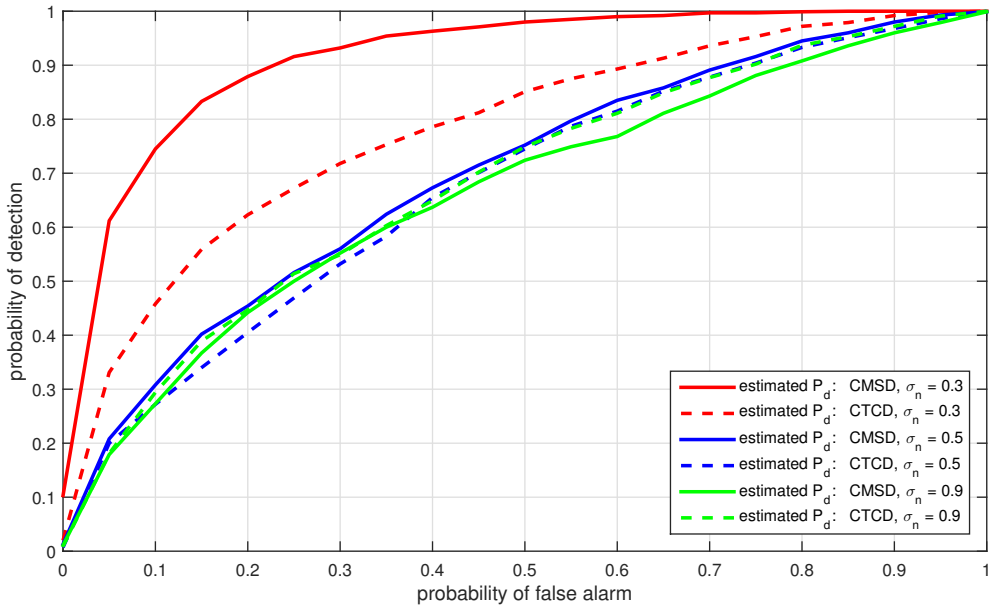


Figure 4.7: Simulation real data: exp. $\mathbf{\Sigma}_{\hat{x}}$, $P_{LL} = 0.5$

From Figure 4.7, it is again visible that the performance of both detectors is inversely proportional to the magnitude of the noise standard deviation. The link loss probability was kept constant with $P_{LL} = 0.5$ for the same reason as given for the experiments with synthetic graphs (more change were generated).

Furthermore, from this experiment it can be seen that CTCD and the CMSD have a similar performance. As long as the amount of topology change is high enough, signal smoothness can be decisive. Although, this seems promising, this is still a point open for discussion as for both detectors, the probability of detection was determined with a high P_{LL}.

For the second simulation, the link loss probability varied i.e. $P_{LL} = \{0.1,\ 0.5,\ 0.9\}$, while the noise standard deviation was kept constant, i.e., $\sigma_n = 0.1$.

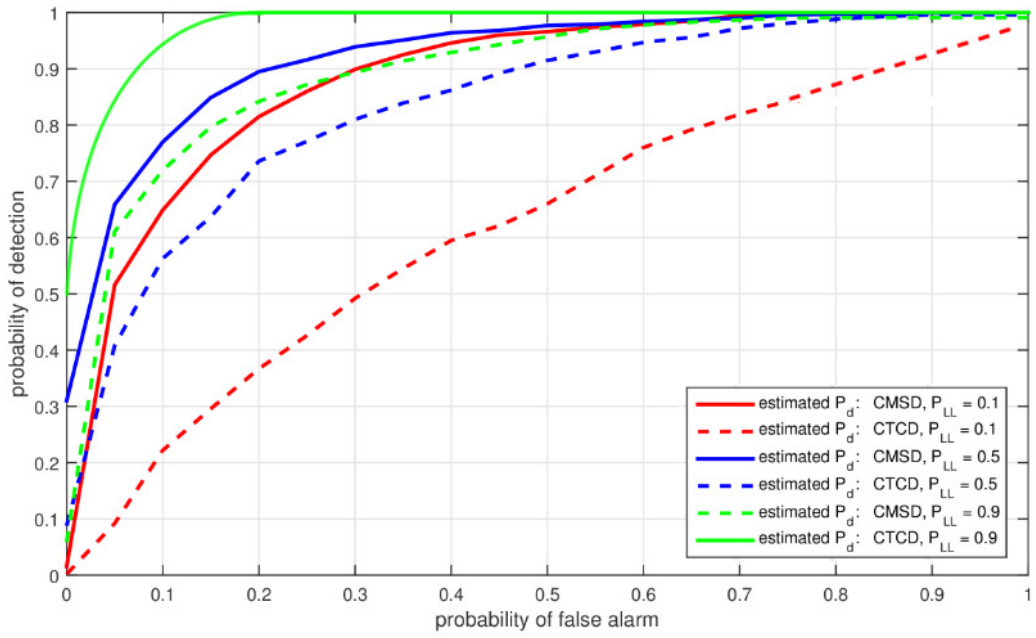


Figure 4.8: Simulation real data: exp. $\mathbf{\Sigma}_{\hat{x}}$, $\sigma_n = 0.1$

Also for this simulation, the same trend as for earlier simulations is visible. As the probability for link loss increases, the performance of both detectors tend to increase. Again the CMSD performs better for all link loss probabilities even for real graphs.

The main conclusion that can be drawn for the CTCD are the following. Only if the amount of change is high enough, the CTCD is able to detect topological changes. The signal smoothness is a sensitive property for synthetic and real graphs This property gets affected significantly for high noise levels and low link loss probabilities.

5

Future work

This chapter is dedicated to the study of the blind detection possibilities exploiting the stationarity of graph signals. This chapter should be seen as a suggestion for future work. First, in Section 5.1 a glance is given on stationarity in the context of GSP. Thereafter, an explanation is given on how this property can be exploited.

5.1. Exploiting stationary signals

In this section, the detection problem is considered when the graph signal is stationary (to a certain degree) under hypothesis $\mathcal{H}_0$. Under $\mathcal{H}_1$ it is not known whether the graph signal is stationary as this property might get lost due to unknown changes in the topology of the graph under the alternative hypothesis. Again define vector $\boldsymbol{\theta} \in \mathbb{R}^N$ as follows:

$$\boldsymbol{\theta} = \mathbf{U}_0 \hat{\mathbf{x}}, \tag{5.1}$$

such that the observation signal is given by $\mathbf{y} = \boldsymbol{\theta} + \mathbf{n}$ with $\mathbf{n} \in \mathbb{R}^N$ as the noise vector. The vector $\boldsymbol{\theta}$ in this chapter is modeled as a realization of a zero mean normal distribution, i.e., $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\theta)$. The matrix $\mathbf{C}_\theta \in \mathbb{R}^{N \times N}$ is the covariance matrix of $\boldsymbol{\theta}$ and is assumed to be unknown due to $\hat{\mathbf{x}}$ under $\mathcal{H}_0$. Under $\mathcal{H}_1$, matrix $\mathbf{C}_\theta$ is unknown due to $\hat{\mathbf{x}}$ and unknown topological changes. In this problem, $\boldsymbol{\theta}$ is considered to be stationary w.r.t. the graph under $\mathcal{H}_0$. Referring back to subsection 2.1.3, this means that the covariance matrix of $\boldsymbol{\theta}$ can be written as follows: $\mathbf{C}_\theta = \mathbf{U}_0 \cdot \text{diag}(\mathbf{p}_\theta) \cdot \mathbf{U}_0^H$ where $\mathbf{p}_\theta$ is the PSD of $\boldsymbol{\theta}$. In other words, it means that $\mathbf{C}_\theta$ can be diagonalized with the eigenvector matrix of the graph Laplacian under $\mathcal{H}_0$. This property that is then exploited to discriminate between the two hypotheses. The noise in this problem is again modeled as: $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. With this in place, the observation signal $\mathbf{y}$ is distributed as:

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_y) \quad \text{with} \quad \mathbf{C}_y = \mathbf{C}_\theta + \sigma^2\mathbf{I} \tag{5.2}$$

Since $\boldsymbol{\theta}$ is stationary w.r.t. the graph under $\mathcal{H}_0$, and because matrix $\mathbf{U}_0$ is unitary (by definition), the covariance matrix $\mathbf{C}_y$ can be written as:

$$\begin{aligned} \mathbf{C}_y &= \mathbf{U}_0 \cdot \text{diag}(\mathbf{p}_\theta) \cdot \mathbf{U}_0^H + \sigma^2 \mathbf{U}_0 \mathbf{U}_0^H \\ &= \mathbf{U}_0 \cdot \text{diag}(\mathbf{p}_\theta + \sigma^2 \mathbf{1}) \cdot \mathbf{U}_0^H \\ &= \mathbf{U}_0 \cdot \text{diag}(\mathbf{p}_y) \cdot \mathbf{U}_0^H, \end{aligned} \tag{5.3}$$

where $\mathbf{p}_y$ denotes the PSD of $\mathbf{y}$ under $\mathcal{H}_0$. Consequently, the observation signal $\mathbf{y}$ (which is a contaminated version of $\boldsymbol{\theta}$) is still stationary w.r.t. the graph under $\mathcal{H}_0$ since $\mathbf{C}_y$ is still diagonalizable with eigenvector matrix $\mathbf{U}_0$. The PSD $\mathbf{p}_y$ is given by: $\mathbf{p}_y = \mathbf{p}_\theta + \sigma^2 \mathbf{1}$ where $\mathbf{1}$ is a vector of ones of proper size.

If the unknown graph topological changes occur under $\mathcal{H}_1$, the covariance matrix $\mathbf{C}_y$ is expected to be no longer diagonalizable with matrix $\mathbf{U}_0$ and one can conclude that the stationary property is (partially) lost. In terms of stationarity level (see (2.8)), this means that: $s(\mathbf{C}_y, \mathbf{U}_0) < 1$. Besides this, also the clean frequency graph signal $\hat{\mathbf{x}}$ is unknown. Under $\mathcal{H}_0$, the stationarity level $s(\mathbf{C}_y, \mathbf{U}_0)$ of $\mathbf{y}$ is expected to be upper bounded by s. With this in mind, the following binary hypothesis testing problem is formulated:

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathbf{y} = \mathbf{U}_0\hat{\mathbf{x}} + \mathbf{n} \quad \text{and} \quad s(\mathbf{C}_y, \mathbf{U}_0) \leq s \\ \mathcal{H}_1 : &\quad \text{otherwise.} \end{aligned} \tag{5.4}$$

What differs this problem even more, regarding the problems that were discussed in the previous chapters, is that batch data are needed. The assumption is that we have an observation matrix $\mathbf{Y} \in \mathbb{R}^{N \times N}$ where N is also the number of $\mathbf{y}_i \in \mathbb{R}^N$ observations in the batch (parameter N was also defined as the number of nodes in the graph). The data are needed to estimate the covariance matrices, which are used to detect stationarity losses. In [14] methods are presented to estimate the covariance matrix with its so called empirical estimate $\hat{\mathbf{C}}_y$:

$$\hat{\mathbf{C}}_y = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^H. \tag{5.5}$$

Two methods can be used to cope with this specific problem. The first method is by neglecting the NP approach and by using this estimate directly in the stationarity measure, i.e., $s(\hat{\mathbf{C}}_y, \mathbf{U}_0)$. The stationarity measure is then used as the detector itself:

$$s(\hat{\mathbf{C}}_y, \mathbf{U}_0) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} s. \tag{5.6}$$

For the second method, an NP approach can be used. Again let $\boldsymbol{\theta}_0$ be $\boldsymbol{\theta}$ under $\mathcal{H}_0$ and $\boldsymbol{\theta}_1$ be $\boldsymbol{\theta}$ under $\mathcal{H}_1$. Denote the distribution of $\mathbf{y}$ under $\mathcal{H}_0$ as $p(\mathbf{y}; \boldsymbol{\theta}_0, \mathcal{H}_0)$ and the distribution of $\mathbf{y}$ under $\mathcal{H}_1$ as $p(\mathbf{y}; \boldsymbol{\theta}_1, \mathcal{H}_1)$. These parametric models are then known except for parameter vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. These vectors then need to be replaced by their MLEs, i.e., $\boldsymbol{\theta}_0^*$ and $\boldsymbol{\theta}_1^*$. Under $\mathcal{H}_1$, a conventional MLE problem has to be solved and under $\mathcal{H}_0$, a constrained MLE problem. The constraint is then defined by the stationarity measure which has to be bounded by s under $\mathcal{H}_0$ (since this was assumed). The covariance matrix $\mathbf{C}_\theta$ should then be used as the argument since $p(\mathbf{y}; \boldsymbol{\theta}, \mathcal{H}_0)$ is maximized over $\boldsymbol{\theta}$. If $\mathbf{C}_\theta$ is not known, it has to be estimated using the observation matrix $\mathbf{Y}$ and a proper estimation technique to get $\hat{\mathbf{C}}_\theta$. With this in place, the MLE estimators for $\boldsymbol{\theta}$ are:

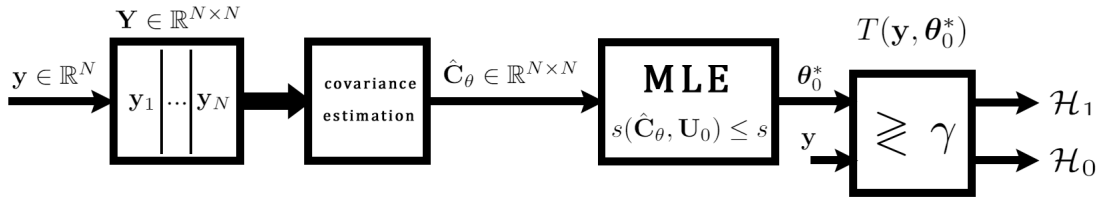
$$\boldsymbol{\theta}_1^* = \arg\max_{\boldsymbol{\theta}} \quad p(\mathbf{y}; \boldsymbol{\theta}, \mathcal{H}_1) = \mathbf{y}. \tag{5.7}$$

$$\begin{aligned} \boldsymbol{\theta}_0^* = \operatorname*{argmax}_{\boldsymbol{\theta}} \quad & p(\mathbf{y}; \boldsymbol{\theta}, \mathcal{H}_0) \\ \text{s.t.} \quad & s(\hat{\mathbf{C}}_\theta, \mathbf{U}_0) \quad \leq \quad s, \end{aligned} \tag{5.8}$$

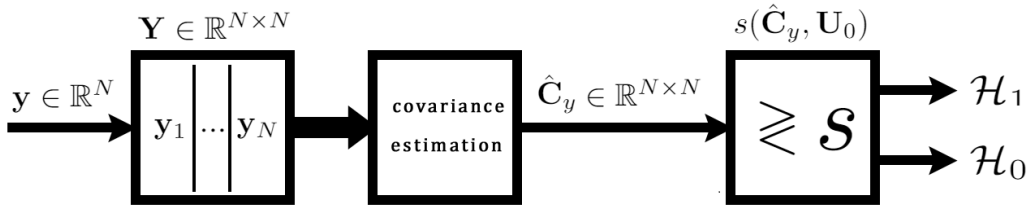
where the bound s has to be found with a proper rule using the data $\mathbf{y}$. With this in place, a GLRT $L_G(\mathbf{y})$ can be proposed to derive a detector:

$$L_G(\mathbf{y}) = \frac{p(\mathbf{y}; \boldsymbol{\theta}_1^*, \mathcal{H}_1)}{p(\mathbf{y}; \boldsymbol{\theta}_0^*, \mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{5.9}$$

In Figure 5.1, two block schemes are given of possible detectors exploiting the stationarity of $\hat{\mathbf{x}}$. Further analysis using the stationarity of graph signals is left for more interested readers.



(a) With this scheme, an NP approach is used. The covariance matrix of $\boldsymbol{\theta}$ is estimated and used for a constrained MLE problem to find $\boldsymbol{\theta}_0^*$.



(b) With this scheme, the constrained $s(\hat{\mathbf{C}}_y, \mathbf{U}_0)$ is used as the detector. The covariance matrix of $\mathbf{y}$ is estimated and then plugged in the constraint detector.

Figure 5.1: Possible detection block schemes when the stationarity of $\hat{\mathbf{x}}$ is exploited.

6

Conclusion

This chapter concludes the thesis by briefly revisiting the research question considered at the start of this project. The previous findings will be shortly summarized once again. Furthermore, some suggestions and areas of improvements will be given concerning blind detection.

The scope of this thesis was to study the possibilities with blind detection (i.e. when there is no information about the alternative graph) and to evaluate detectors that are designed according to this principle. This study contributes the a more realistic scenario where graph topology changes are not known in advance. Two detectors, the BTCD and the CTCD, were designed which respectively exploited the signal bandlimitedness and smoothness. This study was compared with the detectors of [11] as a reference. In their work, they assume to know the nominal and alternative graphs. This implies that (potential) topological changes are known a priori. Furthermore, the relation between the statistics of graphs signals/detectors and topological changes was studied.

In Chapter 3, the BTCD was derived. The BTCD basically applies a high pass filter on the frequency content of the observation graph signal. Under the nominal hypothesis, most of the energy of the observation signal is expected to be bandlimited in the lower frequencies. Under the alternative hypothesis, the out of band energy can be present and this can be attributed to topological changes and/or noise energy. Promising results were attained as the BTCD was able to track the SMSD (one of the detectors derived in [11]) in most of the test scenarios. In one of the test scenarios given by [11], a perfect bandlimited graph signal was modeled. For this case, the BTCD performed as good as the SMSD. When a graph signal is not bandlimited, a degradation is seen. The signal energy is spread over its spectrum and the ability to detect topological changes deteriorates. When smaller topological changes were considered, again a good performance was observed. The SMSD still outperforms the BTCD, but the BTCD still posses detection capabilities eventhough without information about the alternative graph. For the real graph experiment, an interesting observation was made. Although less information is used for the BTCD, its performance was better w.r.t. the SMSD. The SMSD calculates the in band projection energy difference. This may cause the noise level to double and could explain why the BTCD performs better w.r.t. the SMSD in this specific scenario.

In Chapter 4, the CTCD was derived. This detector determines a difference between unconstrained and constrained energy in the observation signal. The constrained energy is found by exploiting the signal smoothness under the nominal hypothesis. Under the alternative hypothesis, the signal energy cannot be constrained as the energy decay is not known due to the topology changes. The obtained results have shown that exploiting signals smoothness is not as powerful as exploiting the bandlimitedness. Regarding the result that were obtained, it can be seen that the CSMD (also one of the detectors derived in [11]) outperforms the CTCD in most of the cases. As long the amount of change is high enough, the CTCD is still capable to detect changes. The smoothness of a signal also tends to be less discriminative when the noise power is high. Energy decay becomes less bounded and therefore it becomes harder to detect.

A

Derivation blind topology change detector

In this section, the derivations for the BTCD in Chapter 3 are given. Working out (3.6) gives the following:

$$\frac{\frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\hat{\mathbf{y}}_K^c-\boldsymbol{\theta}_1^*)^H\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{y}}_K^c-\boldsymbol{\theta}_1^*)\right)}{\frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}\hat{\mathbf{y}}_K^{cH}\boldsymbol{\Sigma}^{-1}\hat{\mathbf{y}}_K^c\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{A.1}$$

By taking the logarithm on both sides, the following expression is found:

$$-\frac{1}{2}\left((\hat{\mathbf{y}}_K^c-\boldsymbol{\theta}_1^*)^H\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{y}}_K^c-\boldsymbol{\theta}_1^*)-\hat{\mathbf{y}}_K^{cH}\boldsymbol{\Sigma}^{-1}\hat{\mathbf{y}}_K^c\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln(\gamma) \tag{A.2}$$

which reduces to:

$$-\frac{1}{2}\left(-2\hat{\mathbf{y}}_K^{cH}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_1^*+\boldsymbol{\theta}_1^{*H}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_1^*\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln(\gamma). \tag{A.3}$$

The final expression is obtained by substituting the MLE specified by (3.5):

$$\hat{\mathbf{y}}_K^{cH}\boldsymbol{\Sigma}^{-1}\hat{\mathbf{y}}_K^c \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma'. \tag{A.4}$$

This finally led to the BTCD:

$$T_{\text{BTCD}}(\hat{\mathbf{y}}_K^c)=\hat{\mathbf{y}}_K^{cH}\boldsymbol{\Sigma}^{-1}\hat{\mathbf{y}}_K^c. \tag{A.5}$$

B

Derivation constrained blind topology change detector

In this section, the derivations for the CMSD in Chapter 4 are given. Working out the GLRT in (4.11) with a similar approach as given in Chapter 3, the CTCD is found:

$$\frac{\frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\theta}_1^*)^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\theta}_1^*)\right)}{\frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\theta}_0^*)^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\theta}_0^*)\right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{B.1}$$

Taking the logarithm on both sides results in the following expression:

$$-\frac{1}{2}\Big((\mathbf{y}-\boldsymbol{\theta}_1^*)^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\theta}_1^*)-(\mathbf{y}-\boldsymbol{\theta}_0^*)^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\theta}_0^*)\Big) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln(\gamma). \tag{B.2}$$

This reduces to:

$$-\frac{1}{2}\Big(2\mathbf{y}^H\boldsymbol{\Sigma}^{-1}[\boldsymbol{\theta}_0^*-\boldsymbol{\theta}_1^*]+\boldsymbol{\theta}_1^{*H}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_0^{*H}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_0^*\Big) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \ln(\gamma). \tag{B.3}$$

Now by absorbing the energy terms $\boldsymbol{\theta}_1^{*H}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_0^{*H}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_0^*$ in the threshold, the following expression is obtained:

$$\mathbf{y}^H\boldsymbol{\Sigma}^{-1}[\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_0^*] \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma'. \tag{B.4}$$

Now by plugging MLE $\boldsymbol{\theta}_1^*$ (4.7) back, the CTCD is found:

$$\begin{aligned} T_{\text{CTCD}} &= \mathbf{y}^H\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\boldsymbol{\theta}_0^*] \\ &= \mathbf{y}^H\boldsymbol{\Sigma}^{-1}\mathbf{y} \quad - \quad \mathbf{y}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_0^*. \end{aligned} \tag{B.5}$$

B.1. Equivalent optimization problems

In this section, the equivalence of (4.8) and (4.9) is proven. First of all, one should notice that the cost function in (4.8), denoted by $p(\mathbf{y};\boldsymbol{\theta},\mathcal{H}_0)$, models a normal distribution since the noise $\mathbf{n}$ considered in this is case, is modeled as: $\mathbf{n} \sim \mathcal{N}(\mathbf{0},\sigma^2\mathbf{I})$. This means the following for $p(\mathbf{y};\boldsymbol{\theta},\mathcal{H}_0)$:

$$p(\mathbf{y};\boldsymbol{\theta},\mathcal{H}_0) = \frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\theta})^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\theta})\right) \tag{B.6}$$

Now if $p(\mathbf{y};\boldsymbol{\theta},\mathcal{H}_0)$ is maximized over $\boldsymbol{\theta}$, one can easily see that only the exponential needs to be maximized since $\exp(\bullet)$ is an increasing function in its argument. However, its argument is negative which means that maximizing $p(\mathbf{y};\boldsymbol{\theta},\mathcal{H}_0)$ over $\boldsymbol{\theta}$ is equivalent to minimizing: $(\mathbf{y}-\boldsymbol{\theta})^H(\mathbf{y}-\boldsymbol{\theta})$.

B.2. Solution constrained optimization problem

The solution of the constrained optimization problem, specified by (4.7), follows next:

$$\begin{aligned} \boldsymbol{\theta}_0^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \quad & (\mathbf{y}-\boldsymbol{\theta})^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\theta}) \\ \text{s.t.} \quad & \boldsymbol{\theta}^H\mathbf{L}_0\boldsymbol{\theta} - r\cdot\boldsymbol{\theta}^H\boldsymbol{\theta} \quad \leq \quad 0 \end{aligned} \tag{B.7}$$

Let $\mathbf{L}_{0,r}$ be a matrix which is defined as follows:

$$\mathbf{L}_{0,r} = \mathbf{L}_0 - r\mathbf{I} \tag{B.8}$$

The constraint is then reformulated as:

$$\boldsymbol{\theta}^H\mathbf{L}_0\boldsymbol{\theta} - r\cdot\boldsymbol{\theta}^H\boldsymbol{\theta} \leq 0 \quad \Rightarrow \quad \boldsymbol{\theta}^H\mathbf{L}_{0,r}\boldsymbol{\theta} \leq 0 \tag{B.9}$$

Then by KKT, an optimal solution for $\boldsymbol{\theta}$ has to satisfy the following conditions:

$$\frac{\partial}{\partial\boldsymbol{\theta}}\Big[(\mathbf{y}-\boldsymbol{\theta})^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\theta}) + \mu\cdot\boldsymbol{\theta}^H\mathbf{L}_{0,r}\boldsymbol{\theta}\Big] = 0 \tag{B.10}$$

$$\mu\cdot\boldsymbol{\theta}^H\mathbf{L}_{0,r}\boldsymbol{\theta} = 0 \tag{B.11}$$

$$\mu \geq 0 \tag{B.12}$$

where μ is known as the KKT multiplier. Working out (B.10) gives the following expression for $\boldsymbol{\theta}$:

$$\begin{aligned} -2\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\theta}) + 2\mu\mathbf{L}_{0,r}\boldsymbol{\theta} &= 0 \\ [\boldsymbol{\Sigma}^{-1}+\mu\mathbf{L}_{0,r}]\boldsymbol{\theta} &= \boldsymbol{\Sigma}^{-1}\mathbf{y} \\ \Rightarrow \qquad \boldsymbol{\theta} &= [\boldsymbol{\Sigma}^{-1}+\mu\mathbf{L}_{0,r}]^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{y} \end{aligned} \tag{B.13}$$

Since it is known that $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \sigma^2\mathbf{I}$, expression (B.13) can be rewritten as:

$$\boldsymbol{\theta} = [\mathbf{I} + \mu\sigma^2\mathbf{L}_{0,r}]^{-1}\mathbf{y} \tag{B.14}$$

and by using the eigenvalue decomposition of the laplacian, $\boldsymbol{\theta}$ can also be expressed as:

$$\boldsymbol{\theta} = \mathbf{U}_0[\mathbf{I} + \mu\sigma^2\mathbf{\Lambda}_{0,r}]^{-1}\mathbf{U}_0^H\mathbf{y} \tag{B.15}$$

where $\mathbf{\Lambda}_{0,r} = \mathbf{\Lambda}_0 - r\mathbf{I}$. Now by substituting (B.15) in (B.11), a solution for the KKT multiplier μ can be found (notice how the property of taking powers of diagonal matrices is exploited):

$$\begin{aligned} \mu \cdot \mathbf{y}^H\mathbf{U}_0[\mathbf{I} + \mu\sigma^2\mathbf{\Lambda}_{0,r}]^{-2}\mathbf{\Lambda}_{0,r}\mathbf{U}_0^H\mathbf{y} = 0 \\ \mu \cdot \hat{\mathbf{y}}_0^H[\mathbf{I} + \mu\sigma^2\mathbf{\Lambda}_{0,r}]^{-2}\mathbf{\Lambda}_{0,r}\hat{\mathbf{y}}_0 = 0 \end{aligned} \tag{B.16}$$

where $\hat{\mathbf{y}}_0 = \mathbf{U}_0\mathbf{y}$ is the projection of the observation signal on graph $\mathcal{G}_0$ with the following statistics: $\hat{\mathbf{y}}_0 \sim \mathcal{N}(\hat{\mathbf{x}}, \sigma^2\mathbf{I})$. In scalar form, the expression for μ is:

$$\mu \cdot \sum_{i=1}^{N} \frac{(\lambda_{0,i} - r)\hat{y}_{0,i}^2}{(1 + \mu\sigma^2(\lambda_{0,i} - r))^2} = 0 \tag{B.17}$$

when the square in de denominator is worked out, this is equivalent to:

$$\begin{aligned} \mu \cdot \sum_{i=1}^{N} \frac{(\lambda_{0,i} - r)\hat{y}_{0,i}^2}{1 + 2\mu\sigma^2(\lambda_{0,i} - r) + \mu^2\sigma^4(\lambda_{0,i} - r)^2} = 0 \\ \mu \cdot \sum_{i=1}^{N} \frac{\frac{1}{\sigma^2}\left(\frac{1}{\sigma^2(\lambda_{0,i} - r)}\right)\hat{y}_{0,i}^2}{\left(\frac{1}{\sigma^4(\lambda_{0,i} - r)^2}\right) + 2\mu\left(\frac{1}{\sigma^2(\lambda_{0,i} - r)}\right) + \mu^2} = 0 \end{aligned} \tag{B.18}$$

Let $a_i = (\sigma^2(\lambda_{0,i} - r))^{-1}$, then:

$$\begin{aligned} \frac{\mu}{\sigma^2} \cdot \sum_{i=1}^{N} \frac{a_i\hat{y}_{0,i}^2}{a_i^2 + 2\mu a_i + \mu^2} = 0 \\ \frac{\mu}{\sigma^2} \cdot \sum_{i=1}^{N} \frac{a_i\hat{y}_{0,i}^2}{(\mu + a_i)^2} = 0 \end{aligned} \tag{B.19}$$

Bibliography

[1] P. J. Carrington, J. Scott, and S. Wasserman, Models and methods in social network analysis. Cambridge university press, 2005, vol. 28.

[2] 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 Processing Magazine, vol. 30, no. 3, pp. 83–98, 2013.

[3] M. F. Othman and K. Shazali, "Wireless sensor network applications: A study in environment monitoring system," Procedia Engineering, vol. 41, pp. 1204–1210, 2012.

[4] A. Venkitaraman, S. Chatterjee, and P. Handel, "On hilbert transform of signals on graphs," Proc. Sampling Theory Applications, 2015.

[5] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," IEEE transactions on signal processing, vol. 61, no. 7, pp. 1644–1656, 2013.

[6] Y. Xirouhakis, A. Tirakis, and A. Delopoulos, "An efficient graph representation for image retrieval based on color composition," 1999.

[7] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," IEEE transactions on signal processing, vol. 61, no. 7, pp. 1644–1656, 2013.

[8] S. M. Kay, Fundamentals of statistical signal processing. Prentice Hall PTR, 1993.

[9] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," Data Mining and Knowledge Discovery, vol. 29, no. 3, pp. 626–688, 2015.

[10] F. D. V. Fallani, J. Richiardi, M. Chavez, and S. Achard, "Graph analysis of functional brain networks: practical issues in translational neuroscience," Phil. Trans. R. Soc. B, vol. 369, no. 1653, p. 20130521, 2014.

[11] C. Hu, L. Cheng, J. Sepulcre, G. El Fakhri, Y. M. Lu, and Q. Li, "Matched signal detection on graphs: theory and application to brain network classification." in IPMI, 2013, pp. 1–12.

[12] B. A. Miller, N. T. Bliss, P. J. Wolfe, and M. S. Beard, "Detection theory for graphs," Lincoln Laboratory Journal, vol. 20, no. 1, pp. 10–30, 2013.

[13] S. P. Chepuri and G. Leus, "Subgraph detection using graph signals." in ACSSC, 2016, pp. 532–534.

[14] G. S. Marques, Antonio & Segarra & Leus, "Stationary graph processes and spectral estimation," 2016.

[15] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," IEEE Transactions on Signal Processing, vol. 65, no. 13, pp. 3462–3477, 2017.

[16] B. Girault, "Stationary graph signals using an isometric graph translation," in Signal Processing Conference (EUSIPCO), 2015 23rd European. IEEE, 2015, pp. 1516–1520.

[17] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis." IEEE Trans. Signal Processing, vol. 62, no. 12, pp. 3042–3054, 2014.

[18] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," IEEE Transactions on Signal Processing, vol. 64, no. 14, pp. 3775–3789, 2016.

[19] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," IEEE transactions on signal processing, vol. 63, no. 24, pp. 6510–6523, 2015.

[20] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," IEEE Transactions on Signal Processing, vol. 65, no. 2, pp. 274–288, 2017.

[21] R. B. Davies, "The distribution of a linear combination of x2 random variables," Applied Statistics, vol. 29, no. 3, pp. 323–333, 1980.

[22] D. Dong, Xiaowen & Thanou, "Learning laplacian matrix in smooth graph signal representations," 2016.

[23] D. S. Bassett and E. Bullmore, "Small-world brain networks," The neuroscientist, vol. 12, no. 6, pp. 512–523, 2006.

[24] D. J. Watts and S. H. Strogatz, "Collective dynamics of'small-world'networks," nature, vol. 393, no. 6684, p. 440, 1998.

[25] Z. Wen, R. Zhang, and K. Ramamohanarao, "Enabling precision/recall preferences for semi-supervised svm training," in Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, 2014, pp. 421–430.

[26] H.-T. Ha and S. B. Provost, "An accurate approximation to the distribution of a linear combination of non-central chi-square random variables," REVSTAT Stat. J, vol. 11, no. 3, pp. 231–254, 2013.