Network Theoretic Analysis of Maximum a Posteriori Detectors for Sensor Analysis and Design☆

Rajasekhar Anguluri[a], Vaibhav Katewa[a], Sandip Roy[b], Fabio Pasqualetti[a]

[a]*Department of Mechanical Engineering, University of California, Riverside, CA, USA*
[b]*School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA*

Abstract

In this paper we characterize the performance of a class of *maximum-a-posteriori* (MAP) detectors for network systems driven by unknown stochastic inputs, as a function of the location of the sensors and the topology of the network. We consider two scenarios: one in which the changes occurs in the mean of the input, and the other where the changes are allowed to happen in the covariance (or power) of the input. In both the scenarios, to detect the changes, we associate suitable MAP detectors for a given set of sensors, and study its detection performance as function of the network topology, and the graphical distance between the input nodes and the sensors location. When the input and measurement noise follow a Gaussian distribution, we show that, as the number of measurements goes to infinity, the detectors' performance can be studied using the input to output gain of the transfer function of the network system. Using this characterization, we derive conditions under which the detection performance obtained when the sensors are located on a network cut is not worse (resp. not better) than the performance obtained by measuring all nodes of the subnetwork induced by the cut and not containing the input nodes. Our results provide structural insights into the sensor placement from a detection-theoretic viewpoint. Finally, we illustrate our findings via numerical examples.

Keywords: Statistical hypotheses testing, mean detection, covariance detection, network systems, sensor placement

1. Introduction

Security of cyber-physical networks is of timely and utmost importance [1]. In recent years, researchers have proposed a number of model-based and heuristic approaches for detecting and mitigating attacks against the actuators and the sensors in the network (see [2] and the references therein). Despite the success of these studies in revealing the performance and the limitations of attack detection mechanisms, several challenges remain, particularly in distinguishing malicious signals from ambient data, selecting optimal sensor locations to maximize the detection performance [3, 4], and deriving simple graphical rubrics to readily evaluate and optimize network security [5, 6].

This study contributes to a growing research effort on characterizing dynamic properties of network systems, including observability, estimation, detection, and input-output behaviors [3, 7, 8, 9]. In particular, we provide network theoretic insights into the detection of changes in the statistical properties of a stationary stochastic input driving certain network nodes, with the ultimate objective of informing the placement of sensors for detection.

Our work is also associated with the recent studies on constrained sensor selection and actuator placement in network systems [10, 11, 12]. It is also aligned with network-theoretic studies that consider metrics for detection and estimation [7, 10, 9, 13, 14, 15]. Compared to these works, we pursue an explicit characterization of the relationships between the detection performance of a set of sensors and the graphical structure of the system.

Contributions:[1] The main contributions of this work are as follows. First, we consider a binary hypothesis testing problem for a discrete time Gaussian process driving the linear network dynamics through certain network nodes. We primarily consider the scenario where hypothesis on either the mean or the covariance of the input process must be detected using the measurements (possibly corrupted with white Gaussian noise) collected from output nodes that are at least at a specified distance apart from the input nodes. We characterize the *maximum a posteriori* (MAP) detector, and quantify its performance as a function of the gain of the input-output transfer matrix of the network system. These results are significant in

---

[1]In a preliminary version of this paper [6], we considered a SISO system, and studied the MAP detector's performance for changes in mean of the stochastic input, assuming noiseless measurements. Instead, in this paper, we consider a MIMO system, and study the detector's performance for changes occurring in both mean and covariance, under noisy and noiseless measurements. In addition, this paper also includes results on networks with non-negative edge weights.

their own rights. For instance, besides our results, there are only limited works related to detecting changes in the covariance of unknown input signals, and this problem is highly relevant in the context of cyber-physical security.

Second, we study the MAP detector's performance as a function of the sensors location. In the absence of noise, regardless of the network structure and edge weights, we show that the performance of the detector associated with a set of sensors forming a cut of the network (nodes on the cut shall be referred as to cutset nodes) is as good as the performance obtained by measuring all nodes of the subnetwork identified by the cut and not containing the nodes affected by the input nodes (referred as partitioned set nodes). Instead, in the presence of noise, depending upon the transfer matrix gain between the cutset nodes and the partitioned nodes, we show that the detection performance of sensors on the cutset nodes may be better or worse than those of sensors on the partitioned nodes. Finally, we demonstrate our theoretical findings on Toeplitz line networks and some illustrative numerical examples.

Our analysis leads to the interesting results that, depending on the network weights and structure, and the intensity of sensor noise, the detection performance may improve as the graphical distance between the input nodes and the sensors location increases. In fact, our results (i) inform the optimal positioning of sensors for the detection of failure of system components or malicious tampering modeled by unknown stochastic inputs, (ii) allow for the detection of unexpected changes of the system structure, because such changes would modify the original detection profile, and (iii) provide network design guidelines to facilitate or prevent measurability of certain network signals.

Mathematical notation: The cardinality of a set A is denoted by $\mathrm{card}(A)$. The set of natural numbers, real numbers, and complex numbers are denoted as $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$, respectively. The eigenspectrum and the spectral radius of a matrix $M \in \mathbb{C}^{n \times n}$ are denoted by $\mathrm{spec}(M)$ and $\bar{\lambda}(M)$, resp. A symmetric positive (resp. semi) definite matrix M is denoted as $M \succ 0$ (resp. $M \succeq 0$). Instead, a non-negative matrix M is denoted as $M \geq 0$. Let $M_1, \ldots, M_n$ be matrices of different dimensions, then $\mathrm{diag}(M_1, \cdots, M_N)$ represents a block diagonal matrix. The Kronecker product of M_1 and M_2 is denoted by $M_1 \otimes M_2$. An $n \times n$ identity matrix is denoted by I or I_n. The norm on the Banach space of matrix-valued functions, that are essentially bounded on the unit circle $\{z \in \mathbb{C} : |z| = 1\}$, is defined as $\|F(z)\|_\infty := \mathrm{ess} \sup \|F(z)\|_2$ [16]. All finite dimensional vectors are denoted by bold faced symbols. The set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ denotes the standard basis vectors of $\mathbb{R}^n$. Let $\mathbf{x} = (x_1, \ldots, x_n)^\mathsf{T}$ and $\mathbf{y} = (y_1, \ldots, y_n)^\mathsf{T}$. Then, $\mathbf{x} \leq \mathbf{y}$ if and only if $x_i \leq y_i$, $i = 1, \ldots, n$. The probability of an event $\mathcal{E}$ is denoted by $\Pr[\mathcal{E}]$. The conditional mean and covariance of a random variable (vector) X is denoted by $\mathbb{E}[X|\mathcal{E}]$ and $\mathrm{Cov}[X|\mathcal{E}]$, respectively. For $Z \sim \mathcal{N}(0,1)$, $Q_\mathcal{N}(\tau)$ denotes $\Pr[Z \geq \tau]$. For $Y \sim \chi^2(p)$, central chi-squared distribution with q degrees of freedom, $Q_{\chi^2}(p, \tau)$ denotes $\Pr[Y \geq \tau]$.

2. Preliminaries and problem setup

Consider a network represented by the digraph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} := \{1, \ldots, n\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ are the node and edge sets. Let $g_{ij} \in \mathbb{R}$ be the weight assigned to the edge $(i, j) \in \mathcal{E}$, and define the *weighted adjacency matrix* of $\mathcal{G}$ as $G := [g_{ij}]$, where $g_{ij} = 0$ whenever $(i, j) \notin \mathcal{E}$. Let $\mathcal{K} := \{k_1, \ldots, k_r\} \subseteq \mathcal{V}$ be the set of input nodes, which receive r inputs. Let $w(i, j)$ denote a path on $\mathcal{G}$ from node i to j, and let $|w(i, j)|$ be the number of edges of $w(i, j)$. Define the distance between input node set $\mathcal{K}$ and a set of nodes $\mathcal{S} \subseteq \mathcal{V}$ as $\mathrm{dist}(\mathcal{K}, \mathcal{S}) := \min\{|w(i, j)| : i \in \mathcal{K}, j \in \mathcal{S}\}$.

We associate to each node i a state $x_i \in \mathbb{R}$, and let the network evolve with discrete linear dynamics

$$\mathbf{x}[k + 1] = G\mathbf{x}[k] + \Pi\mathbf{w}[k], \tag{1}$$

where $\mathbf{x} = [x_1 \cdots x_n]^T \in \mathbb{R}^n$ contains the states of the nodes at time $k \in \mathbb{N}$, $\mathbf{x}[0] \sim \mathcal{N}(\mathbf{0}, \Sigma_0)$ is the initial state, and $\mathbf{w}[k] \in \mathbb{R}^r$ is the input vector. The input matrix $\Pi = [\mathbf{e}_{k_1}, \ldots, \mathbf{e}_{k_r}]$ indicates the location of the input nodes. The input $\mathbf{w}[k]$ be governed by one of the following two competing statistical hypotheses:

$$\begin{aligned} H_1 : \quad \mathbf{w}[k] &\overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1), \quad k = 0, 1, \ldots, N, \\ H_2 : \quad \mathbf{w}[k] &\overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2), \quad k = 0, 1, \ldots, N, \end{aligned} \tag{2}$$

where the moments $\boldsymbol{\mu}_i \in \mathbb{R}^r$ and $\Sigma_i \in \mathbb{R}^{r \times r}(\succ 0)$, $i \in \{1, 2\}$, are completely known. In other words, the competing hypotheses are simple. However, the true hypothesis is assumed to be unknown over the interval $k = 0, 1, \ldots, N$. We are concerned with detecting the true hypothesis on the input signal, using measurements from the sensors that are not collocated with the input nodes.

We assume that the nodes $\mathcal{J} := \{j_1, \ldots, j_m\} \subseteq \mathcal{V}$ are accessible for sensor placement (one sensor for each node), if $\mathrm{dist}(\mathcal{K}, \mathcal{J}) \geq d$, where $d \in \mathbb{N}$. We refer to $\mathcal{J}$ as the sensor set. The output of these sensors is given by

$$\mathbf{y}_\mathcal{J}[k] = C\mathbf{x}[k] + \mathbf{v}[k], \tag{3}$$

where $C = [\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_m}]^\mathsf{T}$ and $\mathbf{v}[k] \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 I)$. Let the process $\{\mathbf{x}[0], \mathbf{w}[0], \mathbf{v}[0], \mathbf{w}[1], \mathbf{v}[1], \ldots\}$ be uncorrelated. To detect the true hypothesis, we task sensors with a detector, which maps the following time aggregated measurements

$$\mathbf{Y}_\mathcal{J}^\mathsf{T} = \begin{bmatrix} \mathbf{y}_\mathcal{J}^\mathsf{T}[1] & \mathbf{y}_\mathcal{J}^\mathsf{T}[2] & \cdots \mathbf{y}_\mathcal{J}^\mathsf{T}[N] \end{bmatrix}, \tag{4}$$

to a detected hypothesis $\widehat{H}$. We will consider the *maximum a posteriori probability* (MAP) detector, which is given by the following decision rule:

$$\Pr(\{H_2 \text{ is true}\}|\mathbf{Y}_\mathcal{J}) \overset{\widehat{H}=H_2}{\underset{\widehat{H}=H_1}{\gtrless}} \Pr(\{H_1 \text{ is true}\}|\mathbf{Y}_\mathcal{J}). \tag{5}$$

For a predetermined set of input nodes $\mathcal{K}$, the focus of our analysis is to characterize the performance of the detector (5), in terms of the network's adjacency matrix
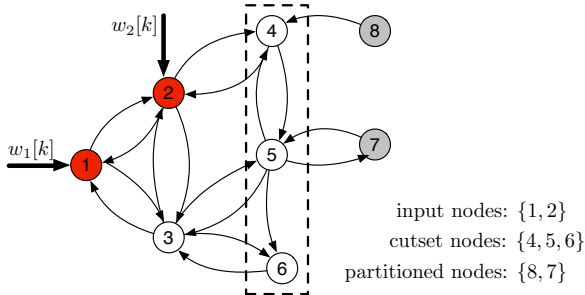
input nodes: $\{1, 2\}$
cutset nodes: $\{4, 5, 6\}$
partitioned nodes: $\{8, 7\}$

Figure 1: Illustration of network partitions induced by a node cutset.

G. The performance of the detector (5) is measured by its error probability, which is given by

$$\mathbb{P}_e(\mathcal{J}) = \sum_{i \in \{1,2\}} \Pr(\widehat{H} \neq H_i | \{H_i \text{ is true}\}) \pi_i. \quad (6)$$

where $\pi_i = \Pr(\{H_i \text{ is true}\})$ is the prior probability.

For any sensor set $\mathcal{J}$ that satisfies $\text{dist}(\mathcal{K}, \mathcal{J}) \geq d$, one expects that the MAP detector's performance (6) is maximum when $\text{dist}(\mathcal{K}, \mathcal{J}) = d$. However, for certain network configurations, studies have shown that the gain of transfer function, which is closely related to the *signal-to-noise* ratio (SNR) of a detector, is maximum when the input and output nodes are farther apart [17]. Hence, it remains unclear whether the closeness of the sensors to the input nodes improves the performance of the detector.

In this paper, we show that the graphical proximity indeed modulate the MAP detector's performance, for certain classes of the detection problems (2). In particular, we characterize networks for which the detection performance obtained when sensors are located on a *node cutset* is better (or worse) than the performance obtained when sensors are placed on nodes of the subnetwork induced by the *node cutset* that does not contain the input nodes (precise statements are provided in Section 4). See Fig 1 for an illustration of *node cutset* and the subnetwork (partitioned nodes) induced by it.

Throughout the paper, we will distinguish the performance of sensors with measurement noise $\sigma_v^2 = 0$ and without noise $\sigma_v^2 > 0$. The essential reason, as we will show later, is that only in the presence of noise, we have a scenario where the performance of a MAP detector associated with the cutset nodes may be worse than that of the partitioned nodes. We end this section with an example that shows the impact of noise on the detection performance.

Example 1. Let $H_i : w[k] \sim \mathcal{N}(\mu_i, \sigma^2)$ and $y[k] = \alpha w[k] + v[k]$, where α is the tuning parameter and $v(k) \sim \mathcal{N}(0, \sigma_v^2)$. For $\pi_i = 0.5$, the MAP detector's error probability, based on N samples of y, is $0.5\, Q_{\mathcal{N}}(0.5\,\eta)$. Here $\eta^2 = N(\mu_1 - \mu_2)^2 / (\sigma^2 + \alpha^{-2} \sigma_v^2)$ denotes the SNR, and $Q_{\mathcal{N}}(0.5\,\eta)$ is decreasing in η. For $\sigma_v^2 = 0$, the error probability does not depend on α, and is lesser than the case where $\sigma^2 > 0$. However, when $\sigma_v^2 > 0$, $Q_{\mathcal{N}}(\cdot)$ depends on α. In particular, when α is small, $Q_{\mathcal{N}}(\cdot)$ is high, and vice versa. Thus,

in the presence of noise, the error probability can be reduced by properly tuning α. $\qquad \square$

3. Detection performance of the MAP detector

In this section, we derive the algebraic expressions of the MAP decision rules and their error probabilities for two special cases of the hypotheses in (2). The first case is the *mean shift model*, in which the covariance matrices in (2) are equal, but the mean vectors are different. The MAP detector (see (9) for the decision rule) for this case is the optimal detector in the (Bayesian) sense that, for the simple hypotheses H_1 and H_2 defined as in (2), no other detector outperforms the detection performance of (9). The second case is the *covariance shift model* in which the mean vectors in (2) are equal, but the covariance matrices are different. For this latter case, we will rely on the sub optimal LD-MAP detector (see below) for deciding the hypothesis. The reason for working with these models is twofold: (i) the error probability expressions are analytically tractable and (ii) these models are widely used for detection and classification problems that arise in practice [18, 19]. The probability expressions derived in this section will be used for the network analysis of the MAP detector's performance (Section 4). Our work draws on the extensive literature on hypothesis testing using Multivariate Gaussian data, but uses a specific simplified detector which allows development of network-theoretic results. Extension of our framework and corresponding results to more general setting is mentioned in Remarks 2 and 3.

The results mentioned in the section are obtained from a standard application of hypothesis testing tools for linear models. Yet, to the best of our knowledge, the asymptotic characterization of the error probability of the MAP detector (Lemma 3.4) is novel, and it serves as the starting point for the results presented in the subsequent sections.

Definition 1. (Linear discriminant function-based MAP detector: LD-MAP) *A LD-MAP detector is as in (5) with $\mathbf{Y}_{\mathcal{J}}$ (4) replaced by the discriminant function $y = \mathbf{b}^{\mathsf{T}} \mathbf{Y}_{\mathcal{J}}$, where the vector* [2] *$\mathbf{b} \in \mathbb{R}^{mN}$ is the maximizer of the following information divergence criterion:*

$$I = \pi_1 \mathbb{E}\left[\ln \frac{f_{H_1}(y)}{f_{H_2}(y)} \,\Big|\, H_1\right] + \pi_2 \mathbb{E}\left[\ln \frac{f_{H_2}(y)}{f_{H_1}(y)} \,\Big|\, H_2\right], \quad (7)$$

where $f_{H_i}(y)$ is the density of y given H_i and $I > 0$.

Remark 1. (Optimal discriminant vector) *For any arbitrary vector $\mathbf{b}$, the I-divergence measure (7) indicates how well a LD-MAP detector is performing in deciding between H_1 and H_2. Thus by maximizing (7), we are finding a best detector among the class of LD-MAP detectors parameterized by $\mathbf{b}$ [19].* $\qquad \square$

---

[2] In the literature of pattern recognition and communications, $\mathbf{b}$ is commonly referred as to the Fisher's discriminant and optimal SINR beam former, respectively [20, 21].

We now state a lemma that provides us with the algebraic expressions of the MAP detectors associated with the mean shift model, and the LD-MAP detector associated with the covariance shift model.

Proposition 3.1. *(Mean and covariance of* $\mathbf{Y}_{\mathcal{J}}$*)* *Let* $\mathbf{Y}_{\mathcal{J}}$ *and* H_i *be defined as in (4) and (2), resp. Then,*

$$\overline{\boldsymbol{\mu}}_i \triangleq \mathbb{E}[\mathbf{Y}_{\mathcal{J}}|H_i] = \mathcal{F}(\mathbf{1}_N \otimes \boldsymbol{\mu}_i) \ \ and$$
$$\overline{\Sigma}_i \triangleq \mathrm{Cov}[\mathbf{Y}_{\mathcal{J}}|H_i] = \mathcal{O}\Sigma_0\mathcal{O}^\mathsf{T} + \mathcal{F}(I_N \otimes \Sigma_i)\mathcal{F}^\mathsf{T} + \sigma_v^2 I, \quad (8)$$

where, the observability and impulse response matrices are

$$\mathcal{O} = \begin{bmatrix} CG \\ CG^2 \\ \vdots \\ CG^N \end{bmatrix} \ and \ \mathcal{F} = \begin{bmatrix} C\Pi & 0 & \dots & 0 \\ CG\Pi & C\Pi & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CG^{N-1}\Pi & CG^{N-2}\Pi & \dots & C\Pi \end{bmatrix}.$$

Lemma 3.2. *(MAP detectors)* *Let* π_1 *and* π_2 *be nonzero priors, and define* $\gamma = \ln(\pi_1/\pi_2)$*. Let* $\mathbf{Y}_{\mathcal{J}}$ *be as in (4), and let* $(\boldsymbol{\mu}_i, \Sigma_i)$ *and* $(\overline{\boldsymbol{\mu}}_i, \overline{\Sigma}_i)$ *be as in (2) and (8), resp.*

(i) *The MAP detector associated with the mean shift model* $(\Sigma_1 = \Sigma_2 \text{ but } \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2)$ *is given by:*

$$\left(2\,\overline{\boldsymbol{\mu}}_\Delta^\mathsf{T}\overline{\Sigma}_c^{-1}\right)\mathbf{Y}_{\mathcal{J}} \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} 2\gamma + \overline{\boldsymbol{\mu}}_\Delta^\mathsf{T}\overline{\Sigma}_c^{-1}\left(\overline{\boldsymbol{\mu}}_1 + \overline{\boldsymbol{\mu}}_2\right), \quad (9)$$

where $\overline{\boldsymbol{\mu}}_\Delta = \overline{\boldsymbol{\mu}}_2 - \overline{\boldsymbol{\mu}}_1$ *and* $\overline{\Sigma}_c \triangleq \overline{\Sigma}_1 = \overline{\Sigma}_1$*.*

(ii) *The LD-MAP detector associated with the covariance shift model* $(\Sigma_1 \neq \Sigma_2 \text{ but } \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2)$ *is given by:*

$$\ln\left(\frac{d_1}{d_2}\right) - 2\gamma \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} (y - \mathbf{b}^\mathsf{T}\overline{\boldsymbol{\mu}}_c)^2 \left[\frac{1}{d_2} - \frac{1}{d_1}\right], \quad (10)$$

where $y = \mathbf{b}^\mathsf{T}\mathbf{Y}_{\mathcal{J}}$*,* $d_i = \mathbf{b}^\mathsf{T}\overline{\Sigma}_i\mathbf{b}$*, and* $\overline{\boldsymbol{\mu}}_c \triangleq \overline{\boldsymbol{\mu}}_1 = \overline{\boldsymbol{\mu}}_2$*.*

The detectors (9) and (10) are functions of the sufficient statistics $2\,\overline{\boldsymbol{\mu}}_\Delta^\mathsf{T}\overline{\Sigma}_c^{-1}\mathbf{Y}_{\mathcal{J}}$ and $y - \mathbf{b}^\mathsf{T}\overline{\boldsymbol{\mu}}_c$, respectively. This means that, given these statistics, other information in $\mathbf{Y}_{\mathcal{J}}$ is not needed for deciding between H_1 and H_2. In order to characterize the error probabilities of the detectors in Lemma 3.2, we make the following assumption:

Assumption 3.3. *The LTI system (1) is stable. Further,*

(i) *for the mean shift model,* $\lim_{N\to\infty} N\|\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\|_2 = c$*, where* $0 < c < \infty$*, and* $G^k = 0$ *for some* $k \in \mathbb{N}$*, and*

(ii) *for the covariance shift model,* $\Sigma_1 \succ 0$ *and* $\Sigma_2 = 0$*.*

Lemma 3.4. *(Error probability: infinite horizon)* *Let* $\pi_1 = \pi_2 = 0.5$ *and* $\mathbf{x}[0] = 0$*. Let* $T(z) = C(zI - G)^{-1}\Pi$*, where* $z \notin spec(G)$*. The error probability of the MAP detector (9) and the LD-MAP detector (10) as* $N \to \infty$ *are*

$$\mathbb{P}_{e_m}(\mathcal{J}) = 0.5\,Q_{\mathcal{N}}(0.5\,\eta) \ \ and \quad (11)$$
$$\mathbb{P}_{e_v}(\mathcal{J}) = 0.5\left[1 - Q_{\chi^2}(1, \tau)\right] + 0.5\,Q_{\chi^2}(1, \tau R), \quad (12)$$

respectively, where $\tau = \ln R/(R-1)$*. The SNRs are*

$$\eta^2 = N\widetilde{\boldsymbol{\mu}}_\Delta^\mathsf{T}\left([L^\mathsf{T}L + \sigma_v^2 I]^{-1}L^\mathsf{T}L\right)\widetilde{\boldsymbol{\mu}}_\Delta \ \ and \quad (13)$$
$$R = 1 + \sigma_v^{-2}\|T(z)\Sigma_1^{\frac{1}{2}}\|_\infty^2, \quad (14)$$

where $L = T(1)\Sigma_c^{\frac{1}{2}}$ *and* $\widetilde{\boldsymbol{\mu}}_\Delta = \Sigma_c^{-\frac{1}{2}}[\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1]$*, and* $\Sigma_c^{\frac{1}{2}}$ *and* $\Sigma_1^{\frac{1}{2}}$ *are the positive square roots of* Σ_c *and* Σ_1*, respectively.*

The assumptions $\pi_i = 0.5$ and $\mathbf{x}[0] = 0$ are for the ease of presentation, and the probability expressions can be easily adjusted to include other priors and initial conditions. The assumption $N\|\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\|_2 \to c$ ensures that $\mathbb{P}_{e_m}(\mathcal{J}) < 0.5$. Instead, the assumption $G^k = 0$ is to eliminate the remainder terms in the computation of η. We emphasize that the only restriction on k is that it should be finite, but can be arbitrarily large. We now state a corollary to the above lemma in which we do not assume $\Sigma_2 = 0$ in the covariance shift model (see Remark 2).

Corollary 3.5. *(SNRs: identical input statistics)* *Let* H_i *in (2) be* $\mathbf{w}[k] \sim \mathcal{N}(\mu_i \mathbf{1}, \sigma_i^2 D)$*, where* μ_i *and* σ_i^2 *are scalars, and* $D > 0$*. For the covariance shift model let* $\sigma_1^2 > \sigma_2^2$*. Then,*

$$\eta_s^2 = \left(N\mu_\Delta^2\right)\mathbf{1}^\mathsf{T}[\sigma_c^2 L^\mathsf{T}L + \sigma_v^2 I]^{-1}L^\mathsf{T}L\mathbf{1}, \quad (15)$$
$$R_s = \frac{\sigma_1^2\|T(z)D^{\frac{1}{2}}\|_\infty + \sigma_v^2}{\sigma_2^2\|T(z)D^{\frac{1}{2}}\|_\infty + \sigma_v^2}, \quad (16)$$

where $\mu_c = \mu_i$*,* $\sigma_c = \sigma_i$*,* $L = T(1)D^{\frac{1}{2}}$*, and* $\mu_\Delta = \mu_2 - \mu_1$*.*

The error probabilities for the identical statistics case can be obtained by substituting η_s and R_s to η and R in (11) and (12), respectively. The effect of sensor noise is also evident from the SNR expressions in the above corollary. In particular, by setting $\sigma_v^2 = 0$ in (15) and (16), the probabilities do not depend on the network matrix G.

Notice that the expressions of $\mathbb{P}_{e_m}(\mathcal{J})$ and $\mathbb{P}_{e_m}(\mathcal{J})$ in above lemma are valid even when N is finite. However, in this case, η and R are complicated functions of the adjacency matrix G. Instead, the elegance of SNRs in Lemma 3.4 and Corollary 3.5 is that they depend on the adjacency matrix G through the well understood transfer matrix $T(z)$. Thus, when $N \to \infty$, one can easily understand the impact of network structure on the detection performance by analyzing $T(z)$. By interpreting the quadratic function in η (or η_s) and $\|\cdot\|_\infty$ in R (or R_s) as a measure of gain, one expects that higher gains results in minimum error probabilities. This intuition is made precise in the following proposition:

Proposition 3.6. $\mathbb{P}_{e_m}(\mathcal{J})$ *and* $\mathbb{P}_{e_v}(\mathcal{J})$ *are decreasing in the SNRs* η *(or* η_s*) and* R *(or* R_s*), respectively.*

The above proposition also helps us to compare the performance of the MAP and LD-MAP detectors associated with different sensor sets. This fact will be exploited greatly in the next section.

Remark 2. (LD-MAP detector's error probability for other covariance matrix structures) *We now comment on extending $\mathbb{P}_{e_v}(\mathcal{J})$ (12) for including other covariance matrices. The case $\Sigma_1 = 0$ and $\Sigma_2 > 0$ can be handled using the proof of Lemma 3.4. For the scenario where neither of Σ_1 or Σ_2 is zero, if we have $N < \infty$ and $\lambda_{max}(\overline{\Sigma}_1 \overline{\Sigma}_2^{-1}) > \lambda_{min}(\overline{\Sigma}_1 \overline{\Sigma}_2^{-1})$, then $\mathbb{P}_{e_v}(\mathcal{J})$ remains the same as in (12), with $R = \lambda_{max}(\overline{\Sigma}_1 \overline{\Sigma}_2^{-1})$. For other cases we refer the reader to [19]. However, the main difficulty in analyzing any of these error probabilities lies in the fact that resulting expressions of SNRs (R) are not amenable to analysis. If one assumes $\overline{\Sigma}_1$ and $\overline{\Sigma}_2$ to be simultaneously diagonalizable, as is the case with Corollary 3.5, an expression of R similar to (16) may be obtained.* $\square$

4. Network analysis of the MAP detector

In this section, we characterize networks for which the MAP detector's performance associated with the sensors that are close to the input nodes is better (or worse) than those of sensors that are farther apart. We distinguish two separate cases when the sensors are without noise ($\sigma_v^2 > 0$) and with noise ($\sigma_v^2 = 0$). To make the notion of closeness precise, we introduce the notion of a node cutset.

Definition 2. (Node cutset) *For the graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ with input nodes $\mathcal{K}$, the nodes $\mathcal{C}_d \subseteq \mathcal{V}$, with $d > 1$, form a node cutset if there exist a non empty source set $\mathcal{S} \subseteq \mathcal{V}$ and a non empty partitioned set $\mathcal{P} \subseteq \mathcal{V}$ such that $\mathcal{V} = \mathcal{S} \sqcup \mathcal{C}_d \sqcup \mathcal{P}$, where $\sqcup$ denotes the disjoint union, and*

(i) $\mathcal{K} \subseteq \mathcal{S}$ and $dist(\mathcal{K}, \mathcal{C}_d) \geq d$, and

(ii) every path from $\mathcal{S}$ to $\mathcal{P}$ contains a node in $\mathcal{C}_d$.

The requirement (i) ensures that the node cutset is at least d edges away from the input nodes. To illustrate Definition 2, consider the network in Fig 1. For the input nodes $\mathcal{K} = \{1, 2\}$, the nodes $\mathcal{C}_1 = \{4, 5, 6\}$ forms a node cutset. However, the nodes $\{5, 6, 7\}$ ceases to form a node cutset, since they failed to satisfy requirement (ii) in the above definition.

4.1. Noiseless measurements

In this section, we state our results on network theoretic characterization of the MAP detectors assuming that the measurement noise in (1) is negligible, i.e., $\sigma_v^2 = 0$. It should be noted that, if a result holds true for the general detection problem (2), we do not state the analogous result for the *mean* and *covariance shift models*.

Theorem 4.1. (Performance of sensors on the node cutset vs the partitioned set: noiseless measurements) *Consider the general detection problem (2). Let $\mathcal{C}_d$ and $\mathcal{P}$ be as in Definition 2, and assume that the measurements from both these node sets are noiseless ($\sigma_v^2 = 0$). Associated with these measurements, let $\mathbb{P}_e(\mathcal{C}_d)$ and $\mathbb{P}_e(\mathcal{P})$ be the respective error probabilities that are computed using (6). Then, $\mathbb{P}_e(\mathcal{C}_d) \leq \mathbb{P}_e(\mathcal{P})$.*

This comparison result is a mere consequence of the following well known result in the binary hypotheses detection problem, known as *theorem of irrelevance* [22] and the *invariance of MAP decision rule* [23].

Lemma 4.2. (Error probability of the MAP detector: dependent measurements) *Let M_1 and M_2 be any two arbitrary simple hypotheses with non-zero priors. Let δ_1 be the error probability of a MAP detector relying on the measurement $\mathbf{Y} \in \mathbb{R}^{p_1}$, and δ_2 be such a quantity associated with the measurement $\mathbf{Z} = g(\mathbf{Y}) + \mathbf{v}$, where $g(.) : \mathbb{R}^{p_1} \to \mathbb{R}^{p_2}$ and $\mathbf{v}$ is stochastically independent of the hypotheses. Then, $\delta_1 \leq \delta_2$.*

From Lemma 4.2, it also follows that Theorem 4.1 holds true even (i) for the case of non-Gaussian input and measurements (provided that the joint density exists), and (ii) if the set $\mathcal{P}$ is replaced with $\mathcal{P} \cup \widetilde{\mathcal{C}}_d$, where $\widetilde{\mathcal{C}}_d \subseteq \mathcal{C}_d$.

Theorem 4.1 implies that, in the absence of noise, nodes near the input location achieve better detection performance compared to those far away from the inputs, *irrespective of the edge weights in the adjacency matrix G and the measurement horizon N. Here, the notion of closeness is to be understood in the sense of node cutsets*, since, $d \leq dist(\mathcal{K}, \mathcal{C}_d) < dist(\mathcal{K}, \mathcal{P})$. Thus, if node cutsets exist in a graph and the measurements are noiseless, one should always place sensors on the cutsets. Thus, if a budget is associated with the sensor placement, it makes sense to find a cutset $\mathcal{C}_d$ of minimum cardinality.

Proposition 4.3. (Error probability of the oracle detector) *Consider the general detection problem (2), and let δ_1 be the error probability of a MAP detector which can directly access the inputs $\mathbf{w}[k]$, $k = 0, \ldots, N$. For any sensor set $\mathcal{J}$, let δ_2 and δ_3 be the error probabilities associated with the noiseless ($\sigma_v^2 = 0$) and noisy ($\sigma_v^2 > 0$) measurements $\mathbf{Y}_{\mathcal{J}}$ (4), respectively. Then, $\delta_1 \leq \delta_2 \leq \delta_3$.*

Proposition 4.3 states that sensor noise degrades the performance of the MAP detector (this fact is also illustrated in Example 1). It also implies that measuring the inputs directly is always better than measuring the noisy/noiseless states (dynamics) of the nodes. Of course, given this fact, it is always beneficial to place the sensors at the input nodes, rather than dealing with the *node cutsets* and the *partitioned sets*.

4.2. Noisy measurements

We now consider the case of noisy measurements ($\sigma_v^2 > 0$). Notice that our results will be specific to the MAP and LD-MAP detectors associated with the *mean* and *covariance shift models*, respectively. Possible extensions to the general detection problem (2) are mentioned in the remarks. We now introduce some additional notation. For a cutset $\mathcal{C}_d$, let $\mathbf{x}_c[k]$, $\mathbf{x}_s[k]$, and $\mathbf{x}_p[k]$ denote the states of the node sets $\mathcal{C}_d$, $\mathcal{S}$, and $\mathcal{P}$, respectively. Let M be a permutation matrix such that $\mathbf{x}[k] = M[\mathbf{x}_s[k]^\intercal, \mathbf{x}_c[k]^\intercal, \mathbf{x}_p[k]^\intercal]^\intercal$,

where $\mathbf{x}[k]$ is the state vector of (1). Then, from (1) it also follows that

$$\begin{bmatrix} \mathbf{x}_s[k+1] \\ \mathbf{x}_c[k+1] \\ \mathbf{x}_p[k+1] \end{bmatrix} = \underbrace{\begin{bmatrix} G_{ss} & G_{sc} & 0 \\ G_{cs} & G_{cc} & G_{cp} \\ 0 & G_{pc} & G_{pp} \end{bmatrix}}_{M^{-1}GM} \begin{bmatrix} \mathbf{x}_s[k] \\ \mathbf{x}_c[k] \\ \mathbf{x}_p[k] \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{w}_s[k] \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{M^{-1}\Pi\mathbf{w}[k]}. \quad (17)$$

From the above relation, note that the states of $\mathcal{C}_d$ serve as an input for the states of partitioned nodes set $\mathcal{P}$, i.e.,

$$\mathbf{x}_p[k+1] = G_{pp}\mathbf{x}_p[k] + G_{pc}\mathbf{x}_c[k]. \quad (18)$$

Based on the transfer function matrix of subsystem (18), we now state a result that is analogous to Theorem 4.1, for the case $\sigma_v^2 > 0$.

Theorem 4.4. (Performance of sensors on the node cutset vs the partitioned set: noisy measurements) Let G_{pp} and G_{pc} be as in (17), and assume that $\mathrm{spec}(G_{pp}) \cap \{z \in \mathbb{C} : |z| = 1\} = \phi$. Let $\overline{\rho}(z)$ and $\underline{\rho}(z)$ be the maximum and minimum singular values of $T_s(z) = (zI - G_{pp})^{-1}G_{pc}$, respectively. Let $\mathbb{P}_{e_m}(\mathcal{C}_d)$ in (11) and $\mathbb{P}_{e_v}(\mathcal{C}_d)$ in (12) be the error probabilities obtained using the noisy measurements ($\sigma_v^2 > 0$) from the cutset $\mathcal{C}_d$. Instead, let $\mathbb{P}_{e_m}(\mathcal{P})$ and $\mathbb{P}_{e_v}(\mathcal{P})$ be the error probabilities associated with the partitioned set $\mathcal{P}$. Then we have:

1a) If $\overline{\rho}(1) \leq 1$, then $\mathbb{P}_{e_m}(\mathcal{C}_d) \leq \mathbb{P}_{e_m}(\mathcal{P})$.
1b) If $\underline{\rho}(1) > 1$, then $\mathbb{P}_{e_m}(\mathcal{C}_d) > \mathbb{P}_{e_m}(\mathcal{P})$.
2a) If $\sup_{|z|=1} \overline{\rho}(z) \leq 1$ then $\mathbb{P}_{e_v}(\mathcal{C}_d) \leq \mathbb{P}_{e_v}(\mathcal{P})$.
2b) If $\inf_{|z|=1} \underline{\rho}(z) > 1$, then $\mathbb{P}_{e_v}(\mathcal{C}_d) > \mathbb{P}_{e_v}(\mathcal{P})$.

Hence, in the presence of noise, depending upon the entries in the matrix $[G_{pp}\,G_{pc}]$, measuring the cutset $\mathcal{C}_d$ might not be always optimal for the purposes of the detection. Instead, in the noiseless case, Theorem 4.1 states that measuring the cutset is always optimal, irrespective of the entries in G. We now explain the reason behind this contrasting behaviors.

Notice that, the quantities sup and inf of $\overline{\rho}(z)$ and $\underline{\rho}(z)$ in Theorem 4.4, respectively, are the maximum and minimum input to output gains of the transfer function matrix $T_s(z)$, associated with the system (18). Theorem 4.4 says that, if the gain between the states $\mathbf{x}_c[k]$ and the states $\mathbf{x}_p[k]$ is high (low), the detection performance with sensors in $\mathcal{P}$ should be better (worse) that that of $\mathcal{C}_d$. In fact, recall from Lemma 3.4 that the detectors associated with the noisy measurements of $\mathcal{C}_d$ and $\mathcal{P}$, respectively, depends on the SNRs of $\mathbf{x}_c[k]$ and $\mathbf{x}_c[k]$ (plus the sensor noise), respectively. Since $\mathbf{x}_p[z] = T_s(z)\mathbf{x}_c[z]$, it is clear that the SNRs are influenced by the gains of $T_s(z)$. In particular, a higher gain increases the SNR of the detector associated with $\mathcal{P}$, which results in a better performance compared to the detector associated with that of $\mathcal{C}_d$.

The above reasoning also holds in the case of noiseless measurements, however, the transfer function gain do not influence MAP detector's performance. In fact, this

gain gets canceled in the error probability computations (this can be clearly seen in Example 1 by interpreting α as the gain). Theorem 4.4 provides conditions for placing sensors on or away from the cutset nodes. For general adjacency matrix, one needs to rely on the software (based on LMI based inequalities) to validate those conditions. However, for non-negative adjacency matrices, the conditions for placing (or not) sensors on the cutset nodes can be stated based on algebraic conditions on the entries of the adjacency matrix. In fact, we have the following result:

Lemma 4.5. (Non-negative adjacency matrix) Let the matrix G in (1) be non-negative, and $\widetilde{G} = [G_{pp}\,G_{pc}] \in \mathbb{R}^{m_1 \times n_1}$, where G_{pp} and G_{pc} are defined in (18).

(i) If $\|\widetilde{G}\|_\infty \leq 1/\sqrt{m_1}$, then we have $\mathbb{P}_{e_m}(\mathcal{C}_d) \leq \mathbb{P}_{e_m}(\mathcal{P})$ and $\mathbb{P}_{e_v}(\mathcal{C}_d) \leq \mathbb{P}_{e_v}(\mathcal{P})$.

(ii) If $n_1 = 1$, and all row sums of $\widetilde{G}$ are greater than one, then $\mathbb{P}_{e_m}(\mathcal{C}_d) \geq \mathbb{P}_{e_m}(\widetilde{\mathcal{P}})$ and $\mathbb{P}_{e_v}(\mathcal{C}_d) \geq \mathbb{P}_{e_v}(\widetilde{\mathcal{P}})$, where $\widetilde{\mathcal{P}} \subseteq \mathcal{P}$.

The inequality $\mathbb{P}_{e_m}(\mathcal{C}_d) \leq \mathbb{P}_{e_m}(\mathcal{P})$ can be obtained even without the non-negativity assumption on G. However, this might not be true for the case of $\mathbb{P}_{e_v}(\cdot)$. Thus, by ensuring that the maximum row sum of $\widetilde{G}$ is bounded by $1/\sqrt{m_1}$ (here m_1 refers to the cardinality of the partitioned set $\mathcal{P}$), one can guarantee that the detection performance of sensors on the cutset is always superior than that of the sensors on the partitioned nodes. The assumption $n_1 = 1$ in part 2) of above lemma implies that $\mathrm{card}(\mathcal{C}_d) = 1$. For arbitrary n_1, the condition row sums of $\widetilde{G}$ greater than one may not be sufficient, and more assumptions on G are required to handle this case. For instance, when G is a diagonally dominant matrix, required sufficient conditions can be obtained using the lower bounds in [24]. Finally, we notice that the bounds presented in Lemma 4.5 depends on the cardinality of the node sets, and hence, our results on networks with non-negative edge weights may be conservative when these cardinalities are large.

The network-theoretic analysis of the MAP and the LD-MAP detectors developed in this section can also be used to inform the placement of sensors for detection. The results show that sensors placed close to the stochastic inputs are effective for batch detection. More precisely, measurements on separating cutsets of the network necessarily outperform downstream sensing strategies. Thus, a strategy for the placement of few sensors is to find small node cutsets that isolate the input nodes. The design of these algorithms falls outside the scope of this paper and is left as the subject of future research.

Remark 3. (Extension of network theoretic results to the other detectors: noisy measurements) *In the cases where the analytical error probability calculation is difficult, eg., the general Gaussian or non-Gaussian detection problem and the covariance shift model with arbitrary covariance matrix structures, one relies on the Chernoff*
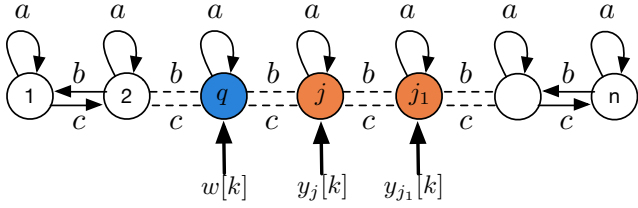
Figure 2: Toeplitz line network with n nodes. The q-th node is injected with the input, and the j-th node represents the cutset node.

type bounds (for eg., see [18]) to quantify the detection performance. In both the cases, i.e., evaluating the performance directly or via bounds, Theorem 4.4 holds true for any detector whose performance (resp. bounds) is monotonically increasing in $\|T(z) = C(zI-G)^{-1}\Pi\|_M$, for some suitable $M \succ 0$. For instance, the Chernoff bounds on the error probability of the general Gaussian detection problem (2) depend on the moment generating function (mgf) of the sufficient statistic of the MAP detector, which ultimately depends on the filtered mean and covariance matrices (8), and our analysis becomes applicable. In the non-Gaussian case, the mgf might depend on other moments as well, and extending our analysis to this case will be challenging. $\square$

4.3. Single input single output (SISO) line networks

In this section, we validate our cutset based results, that we presented in previous section, for the case of line networks by explicitly expressing the error probabilities as a function of the entires of G, and then compare the performance of sensors on $\mathcal{C}_d$ versus sensors on $\mathcal{P}$. We restrict our attention to the SISO systems.

We assume that a stochastic input enters the network through a fixed node $q \in \{1, \ldots, n\}$, and we allow any node $l \in \{1, \ldots, n\}$ with $\text{dist}(l, q) \geq d$ for sensor placement. For this setup, we assume that probabilities $\mathbb{P}_{e_m}(l)$ and $\mathbb{P}_{e_v}(l)$ are obtained by substituting the SNRs η_s (15) and R_s (16) in the expressions of (11) and (12), respectively. Notice that, in contrast to the previous analysis, in which we assume $\Sigma_2 = 0$ (see Assumption 3.3), in this section we do not assume $\sigma_2^2 = 0$ in R_s. For the ease of presentation, we assume the cutset to be a singleton set, i.e., $\mathcal{C}_d = \{j\}$. The following proposition is an extension of Lemma 4.5 for our SISO system setup with the revised error probabilities.

Proposition 4.6. *Let $\widetilde{G}$ be as in Lemma 4.5, and $\sigma_v^2 > 0$. Let $\{j\}$ and $\mathcal{P}$ be the cutset and partitioned sets, resp. If $\|\widetilde{G}\|_\infty \leq 1$, then for any $j_1 \in \mathcal{P}$, we have $\mathbb{P}_{e_m}(j) \leq \mathbb{P}_{e_m}(j_1)$ and $\mathbb{P}_{e_v}(j) \leq \mathbb{P}_{e_v}(j_1)$. The opposite inequality holds true if all row sums of $\widetilde{G}$ are greater than one.*

The proof of above proposition is similar to the proof of Lemma 4.5 and hence, the details are omitted. By not resorting to any proof techniques, i.e, the functional dependence arguments, that we used in previous section, we now validate assertions in above proposition by expressing the error probability in terms of the entries in the matrix

G. To this aim, we consider a line network (see Fig. 2), whose adjacency matrix is given by the following matrix:

$$G = \begin{bmatrix} a & b & 0 & \cdots & 0 & 0 \\ c & a & b & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a & b \\ 0 & 0 & 0 & \cdots & c & a \end{bmatrix}_{n \times n}, \tag{19}$$

where, $a, b, c \in \mathbb{R}_{\geq 0}$. We let the cutset node j be located on the right of the input node q, i.e., $1 \leq q < j < n$ (see Fig 2). The case when j is to the left side of the input node q follows similar analysis. Thus, we have the partitioned set $\mathcal{P} = \{j + 1, \ldots, n\}$. We now show that, for any $l \in \mathcal{P}$, the error probabilities $\mathbb{P}_{e_m}(l)$ and $\mathbb{P}_{e_v}(l)$ are greater or smaller than those of the cutset node j. The following proposition helps us achieve the required goal:

Proposition 4.7. *Let G be as in (19) and $\overline{\lambda}(G) < 1$. Let $|(I-G)^{-1}_{l,q}|$ be the absolute value of (l, q)-th entry of $(I-G)^{-1}$. Let $\widetilde{G}$ be as in Lemma 4.5. Then, we have:*

i) If $\|\widetilde{G}\|_\infty < 1$, then $|(I-G)^{-1}_{q,q}| \geq |(I-G)^{-1}_{q+1,q}| \cdots \geq |(I-G)^{-1}_{n,q}|$.

ii) If all row sums of $\widetilde{G}$ are greater than one, then $|(I-G)^{-1}_{lq}| \geq |(I-G)^{-1}_{qq}|$ for all $q < l \leq n$. If $b = 0$, we have $|(I-G)^{-1}_{q+1,q}| \geq |(I-G)^{-1}_{q+2,q}| \cdots \geq |(I-G)^{-1}_{n,q}|$.

For a fixed input q, above proposition characterizes the qualitative behavior of the input-to-output transfer function gains associated with different output nodes. This fact can be easily seen by expressing $|(I-G)^{-1}_{lq}|$ as $|\mathbf{e}_l^\mathsf{T}(I-G)^{-1}\mathbf{e}_q|$. For the case of Toeplitz line networks, the assertion in Proposition 4.6 is now an easy consequence of Proposition 3.6 and 4.7. In particular, if $b = 0$ and $a + c > 1$, Proposition 4.7 also implies that, the node that is farthest from the input has better detection performance than any other node, including the cutset node. Similarly, assertion in Theorem 4.1 can be verified by letting $\sigma_v^2 = 0$.

The procedure illustrated above, evaluating the error probabilities via the entries of $(I-G)^{-1}$, becomes tedious and might not be even possible for arbitrary network structures. In such situations, one can use the proof techniques presented in Section 4 for understanding the detection performance of sensors on networks.

5. Simulation results

In this section, we present numerical simulations to validate the effectiveness of our cutset based characterization of MAP detection performance on networks, for the case of noisy measurements.

(Detection performance of sensors on the partitioned nodes is better than that of the sensors on the cutset nodes): For this scenario, consider the network in Fig 3. The network has 10 nodes, with 1 and 2 being the input nodes,
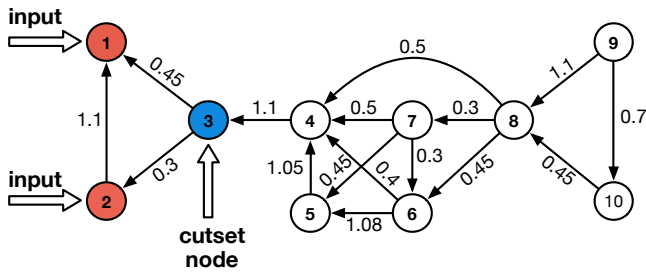
Figure 3: The graph of a network consisting of 10 nodes. The nodes that are to the right of the cutset node {3} form the partitioned set. Instead, nodes 1 and 2 form the source set.

(a) mean model: short memory

(b) covariance model: short memory

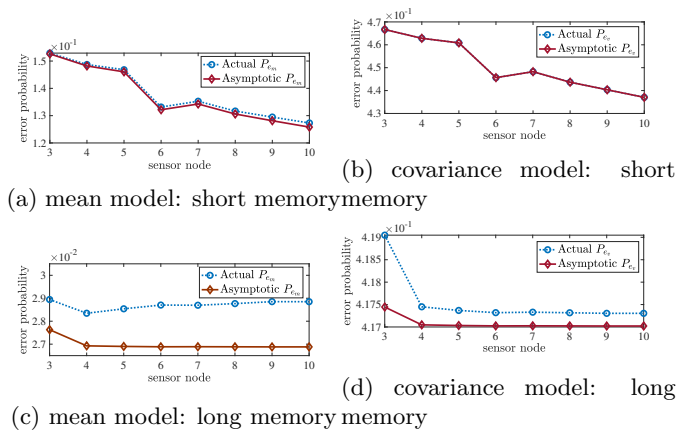(c) mean model: long memory

(d) covariance model: long memory

Figure 4: Actual and asymptotic error probabilities (Lemma 3.4) of the MAP and LD-MAP detectors associated with various nodes of the network shown in Fig. 3. The panels (a) and (b) corresponds to the adjacency matrix that results in the shorter memory of the network dynamics (1). Instead, panels (c) and (d) are associated with an adjacency matrix that results in the longer memory of the network dynamics. The error probability associated with each node in the partitioned set $\mathcal{P} = \{4, \ldots, 10\}$ is less than that of the cutset node $\mathcal{C}_d = \{3\}$. This result is consistent with Lemma 4.5, because all row sums of submatrix $\widetilde{G}$ are greater than one.

$\mathcal{C}_d = \{3\}$ is the cutset node, and $\mathcal{P} = \{4, \ldots, 10\}$ is the partitioned node set. The adjacency matrix of this network is nilpotent, and as a result, system (1) evolving on this network will have a short memory (in fact $G^{10} = 0$). By short (resp. long) memory, we mean that the current state of the network depends on few (resp. several) past states. For the mean shift model, the input $\mathbf{w}_i[k] \sim \mathcal{N}(\mu_i \mathbf{1}, \sigma_i^2 I_{2 \times 2})$, where $\mu_1 = 2$, $\mu_2 = 1$, and $\sigma_2^2 = \sigma_2^2 = 1.5$. Instead, for the covariance shift model, the input[3] $\mathbf{w}_i[k] \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 I_{2 \times 2})$, where $\sigma_1^2 = 2.0$ and $\sigma_2^2 = 1.0$. In both the models, $N = 200$ and the sensor noise variance $\sigma_v^2 = 1.2$.

Fig. 4a and Fig. 4b illustrates the actual and asymptotic error probabilities of the mean and covariance shift models, respectively. The error probabilities are computed using the formulas in Lemma 3.4. In particular, for the asymptotic case, we use the SNRs in Corollary 3.5. In

---

[3]the choice of zero mean is arbitrary, since, the LD-MAP detector's error probability do not depend on the mean; see Lemma 3.4.
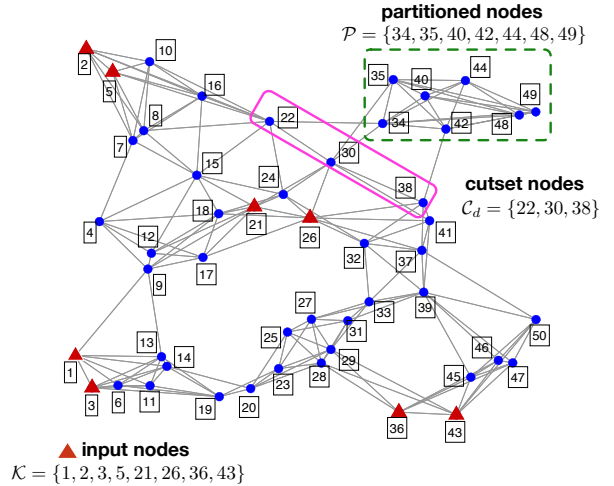

Figure 5: Graph associated with a randomly generated network consisting of 50 nodes [25]. A total of 8 nodes are subjected to stochastic inputs. Instead, sensors are placed on the cutset nodes and the partitioned nodes that are not collocated with the input nodes.

both figures, the error probability associated with the cutset node is greater than that of any node in the partitioned set. This must be the case since $G \geq 0$, and the row sums of the submatrix $\widetilde{G}$ are greater than one (see Lemma 4.5).

The error between the asymptotic and actual error probabilities in Fig. 4a and Fig. 4b is almost negligible, even when N is not large. This is because the adjacency matrix G is a nilpotent matrix, and as a result, the difference between the actual and asymptotic SNRs is minimum. However, this might not be the case when G has long memory, i.e., $G^k \approx 0$ only for a very large k. For $N = 800$, Fig. 4c and Fig. 4d illustrate this scenario for the network that is obtained by modifying some edges of the network in Fig. 3, such that $G^k \approx 0$ for very large k.

(Detection performance of sensors on the cutset nodes is better than that of the sensors on the partitioned nodes):

Consider the network shown in Fig 5. The network has 50 nodes among which $\mathcal{K} = \{1, 2, 3, 5, 21, 26, 36, 43\}$ are the input nodes. The cutset $\mathcal{C}_d = \{22, 30, 38\}$ separates $\mathcal{K}$ from the partitioned set $\mathcal{P} = \{34, 35, 40, 42, 44, 48, 49\}$. For the mean shift model, the input $\mathbf{w}_i[k] \sim \mathcal{N}(\mu_i \mathbf{1}, \sigma_i^2 I_8)$, where $\mu_1 = 2$, $\mu_2 = 1$, and $\sigma_2^2 = \sigma_2^2 = 1.5$, and $\sigma_v^2 = 1.2$. Instead, for the covariance shift model, the input $\mathbf{w}_i[k] \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 I_8)$, where $\sigma_1^2 = 25.0$, $\sigma_2^2 = 0.1$, and $\sigma_v^2 = 0.5$. In both the models, $N = 200$.

Consider all possible subsets of $\mathcal{C}_d \sqcup \mathcal{P}$ whose cardinalities are same as that of the cutset $\mathcal{C}_d$. It is easy to see that there are 120 such sets. For each of these sets, we associate a label $\mathcal{J}_{\text{ind}}$, where ind $\in \{1, \ldots, 120\}$. The labels are given based on a decreasing order of the error probabilities associated with the subsets. In Fig. 6a and Fig. 6b, we show the actual and asymptotic error probabilities of the mean and covariance shift models, respectively. In both figures, the error probability associated with the $\mathcal{C}_d$ is lesser than that of any $\mathcal{J}_{\text{ind}}$. This must be the case because $G \geq 0$, and

the row sums of the submatrix $\|\widetilde{G}\|_\infty < 1/\sqrt{7} = 0.3780$ (see Lemma 4.5).
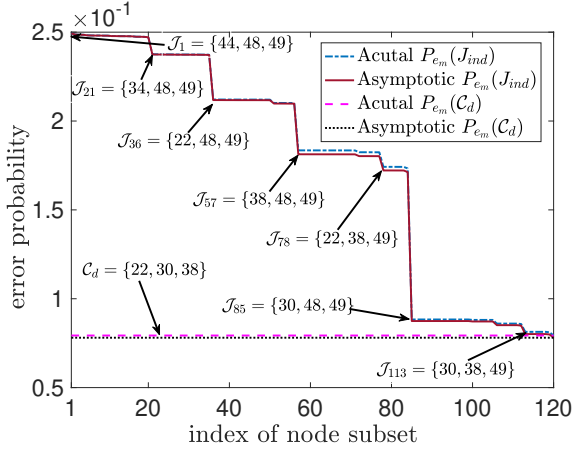
(a) mean shift model.
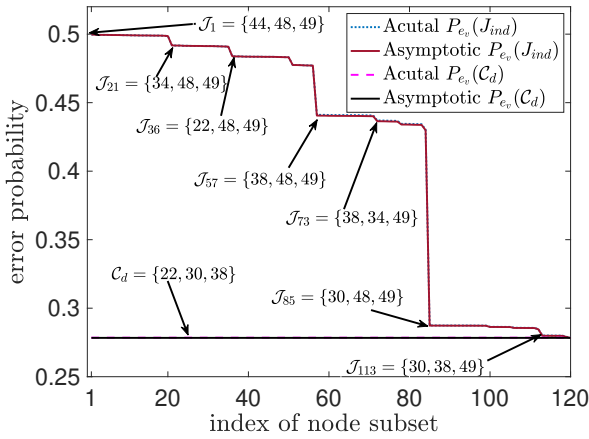
(b) covariance shift model.

Figure 6: Actual and asymptotic error probabilities (Lemma 3.4) of the MAP and LD-MAP detectors associated with the node cutset $\mathcal{C}_d$ and all possible 3 node subsets of $\mathcal{C}_d \sqcup \mathcal{P}$ of the network shown in Fig. 5.The error probabilities of the detectors associated with cutset nodes is lower than that of the detectors associated with any subset of the nodes in the partitioned set. This result is consistent with Lemma 4.5, because the submatrix $\widetilde{G}$ row sums of the adjacency matrix G are less than $1/\sqrt{m_1}$ ($m_1 = 7$).

6. Conclusion

In this paper we formulate mean and covariance detection placement problems for linear dynamical systems defined over networks with unknown stochastic inputs. The main technical contribution of the paper is to identify graphical conditions that predict the performance of MAP and LD-MAP detectors based on the distance between the employed sensors and the stochastic inputs. For networks with non-negative edge weights, we also show that the performance of a detector can be independent of the graphical distance between the sensors and the stochastic inputs.

APPENDIX

Proof of proposition 3.1: From the network dynamics (1) and sensor measurements (3), $\mathbf{Y}_\mathcal{J}$ (4) can be expanded as

$$\mathbf{Y}_\mathcal{J} = \mathcal{O}\mathbf{x}[0] + \mathcal{F}\mathbf{w}_{0:N-1} + \mathbf{v}_{1:N}, \qquad (20)$$

where the vectors $\mathbf{w}_{0:N-1} = [\mathbf{w}[0]^\mathsf{T}, \ldots, \mathbf{w}[N-1]^\mathsf{T}]^\mathsf{T}$ and $\mathbf{v}_{1:N} = [\mathbf{v}[1]^\mathsf{T}, \ldots, \mathbf{v}[N]^\mathsf{T}]^\mathsf{T}$, respectively. The matrices $\mathcal{O}$ and $\mathcal{F}$ are defined in the statement of the proposition. The expressions of $\overline{\boldsymbol{\mu}}_i$ and $\overline{\Sigma}_i$ in (8) follows by taking the expectation and covariance of $\mathbf{Y}_\mathcal{J}$, respectively. $\square$

Proof of Lemma 3.2: Let ζ and z are the realizations of $\mathbf{Y}_\mathcal{J}$ and y, respectively. Since the input and measurement noises follows a Gaussian distribution, the probability density functions of $\mathbf{Y}_\mathcal{J}$ (4) and $y = \mathbf{b}^\mathsf{T}\mathbf{Y}_\mathcal{J}$ are

$$f(\zeta|H_i) \propto \frac{1}{\sqrt{|\overline{\Sigma}_i|}} \exp\left[-\frac{1}{2}(\zeta - \overline{\boldsymbol{\mu}}_i)^T \overline{\Sigma}_i^{-1}(\zeta - \overline{\boldsymbol{\mu}}_i)\right] \text{ and}$$

$$g(z|H_i) \propto \frac{1}{\sqrt{\mathbf{b}^\mathsf{T}\overline{\Sigma}_i\mathbf{b}}} \exp\left[-\frac{(z - \mathbf{b}^\mathsf{T}\overline{\boldsymbol{\mu}}_i)^2}{2\,\mathbf{b}^\mathsf{T}\overline{\Sigma}_i\mathbf{b}}\right], \qquad (21)$$

respectively, where $|\cdot|$ denotes the determinant. Define the log likelihood ratios $\Psi(\zeta) = \ln(f(\zeta|H_2)/f(\zeta|H_1))$ and $\widehat{\Psi}(z) = \ln(f(z|H_2)/f(z|H_1))$. Then, from the mixed Bayes formula [23], the MAP decision rules based on ζ and z, respectively, are given by

$$\Psi(\zeta) \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} \gamma \text{ and } \widehat{\Psi}(z) \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} \gamma. \qquad (22)$$

part 1) Since $\Sigma_1 = \Sigma_2$ and $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, from (8), it follows that $\overline{\Sigma}_1 = \overline{\Sigma}_2$ and $\overline{\boldsymbol{\mu}}_1 \neq \overline{\boldsymbol{\mu}}_2$. Invoking this observation in $f(\zeta|H_i)$, yields the following expression for $\psi(\zeta)$:

$$\Psi(\zeta) = -0.5\,\overline{\boldsymbol{\mu}}_\Delta^\mathsf{T} \overline{\Sigma}_2^{-1}\overline{\boldsymbol{\mu}}_\Delta + (y - \overline{\boldsymbol{\mu}}_1)^\mathsf{T}\overline{\Sigma}_2^{-1}\overline{\boldsymbol{\mu}}_\Delta. \qquad (23)$$

Substitute (23) in the first decision rule of (22) and simplify the resulting expression to obtain the MAP decision rule (9) for ζ. Finally, replacing ζ with $\mathbf{Y}_\mathcal{J}$ yields the required expression.

part 2) In this case we have $\overline{\boldsymbol{\mu}}_1 = \overline{\boldsymbol{\mu}}_2$ and $\overline{\Sigma}_1 \neq \overline{\Sigma}_2$. A similar procedure, as in part 1), based on $g(z|H_i)$ (22) and the second decision rule in (21), yields the LD-MAP detector's expression (9). Details are left to the reader. $\square$

Proof of Lemma 3.4: We divide the proof into two parts. In part 1) we derive the expressions (11) and (13) Instead, in part 2) we derive the expressions (12) and (14).

part 1) Under the assumption that $N < \infty$, let $\widehat{\mathbb{P}}_{e_m}(\mathcal{J})$ be the error probability of (9).. Then, from (9), we have

$$\Pr\left(\widehat{H} = H_2|H_1\right) = \Pr\left(s > \overline{\boldsymbol{\mu}}_\Delta^\mathsf{T}\overline{\Sigma}_c^{-1}\left(\overline{\boldsymbol{\mu}}_1 + \overline{\boldsymbol{\mu}}_2\right)|H_1\right) \text{ and}$$
$$\Pr\left(\widehat{H} = H_1|H_2\right) = \Pr\left(s < \overline{\boldsymbol{\mu}}_\Delta^\mathsf{T}\overline{\Sigma}_c^{-1}\left(\overline{\boldsymbol{\mu}}_1 + \overline{\boldsymbol{\mu}}_2\right)|H_2\right),$$

where $s = 2\,\overline{\boldsymbol{\mu}}_\Delta^\mathsf{T}\overline{\Sigma}_c^{-1}\mathbf{Y}_\mathcal{J}$ follows $\mathcal{N}(\overline{\boldsymbol{\mu}}_\Delta^\mathsf{T}\overline{\Sigma}_c^{-1}\overline{\boldsymbol{\mu}}_1, 4\overline{\boldsymbol{\mu}}_\Delta^\mathsf{T}\overline{\Sigma}_c^{-1}\overline{\boldsymbol{\mu}}_\Delta)$ under H_i, because s is a linear transform of $\mathbf{Y}_\mathcal{J}|H_i$, which

follows a Gaussian distribution. Define $\widehat{\eta}^2 = \overline{\boldsymbol{\mu}}_\Delta^\mathsf{T} \overline{\Sigma}_c^{-1} \overline{\boldsymbol{\mu}}_\Delta$, and notice that $\Pr(\widehat{H} = H_2|H_1) = Q_\mathcal{N}(0.5\,\widehat{\eta})$ and $\Pr(\widehat{H} = H_1|H_2) = 1 - Q_\mathcal{N}(0.5\,\widehat{\eta})$. Finally, from (6), we have $\widehat{\mathbb{P}}_{e_m}(\mathcal{J}) = 0.5\,Q_\mathcal{N}(\widehat{\eta})$. Define $\mathbb{P}_{e_m}(\mathcal{J}) = \lim_{N\to\infty} \widehat{\mathbb{P}}_{e_m}(\mathcal{J})$, and note the following:

$$\mathbb{P}_{e_m}(\mathcal{J}) = \lim_{N\to\infty} 0.5\,Q_\mathcal{N}(0.5\,\widehat{\eta}) = 0.5\,Q_\mathcal{N}\left(0.5 \lim_{N\to\infty} \widehat{\eta}\right)$$

where the final equality follows because $\widehat{\eta}$ is increasing in N (see Proposition A.1). We now show that $\lim_{N\to\infty} \widehat{\eta} = \eta$. From (8), it follows that

$$\widehat{\eta}^2 = (\mathcal{F}\mathbf{m})^\mathsf{T} \overline{\Sigma}_c^{-1} (\mathcal{F}\mathbf{m}), \qquad (24)$$

where $\mathbf{m} = \mathbf{1}_N \otimes \boldsymbol{\mu}_\Delta$ and $\boldsymbol{\mu}_\Delta = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$. Let $l = 1, 2, \ldots$, and define $K(l) = CG^l\Pi$ and $S(i) = \sum_{l=0}^{i-1} K(l)$. With these definitions and the assumption $\overline{\lambda}(G) < 1$, we have $\lim_{i\to\infty} S(i) = C(I-G)^{-1}\Pi \triangleq \overline{K}$, and

$$\mathcal{F}_J\mathbf{m} = \underbrace{\begin{bmatrix} S(1) - \overline{K} \\ S(2) - \overline{K} \\ \vdots \\ S(N) - \overline{K} \end{bmatrix}}_{S_N} \boldsymbol{\mu}_\Delta + \left[\mathbf{1}_N \otimes \overline{K}\right]\boldsymbol{\mu}_\Delta. \qquad (25)$$

Let $t(S_N) = \boldsymbol{\mu}_\Delta^\mathsf{T}\left[S_N^\mathsf{T}\overline{\Sigma}_c^{-1}S_N + 2S_N^\mathsf{T}\Sigma_c^{-1}\left[\mathbf{1}_N \otimes \overline{K}\right]\right]\boldsymbol{\mu}_\Delta$. By substituting (25) in (24), we have

$$\widehat{\eta}^2 = \boldsymbol{\mu}_\Delta^\mathsf{T} \underbrace{\left[\mathbf{1}_N \otimes \overline{K}\right]^\mathsf{T} \overline{\Sigma}_c^{-1} \left[\mathbf{1}_N \otimes \overline{K}\right]}_{F} \boldsymbol{\mu}_\Delta + t(S_N). \qquad (26)$$

Consider the first term of (26). Since $\mathbf{x}[0] = 0$, from (8), it follows that $\overline{\Sigma}_c = \left[\mathcal{F}(I_N \otimes \Sigma_c)\mathcal{F}^\mathsf{T} + \sigma_v^2 I\right]$. Further,

$$\left[\mathbf{1}_N \otimes \overline{K}\right]^\mathsf{T} \overline{\Sigma}_c = \left[\overline{K}^\mathsf{T}\overline{K}\Sigma_c + \sigma_v^2 I\right]\left[\mathbf{1}_N \otimes \overline{K}\right]^\mathsf{T} + \overline{K}^\mathsf{T}\underbrace{\left[\widetilde{S}_N^\mathsf{T}(I \otimes \Sigma_c)\mathcal{F}^\mathsf{T} + \overline{K}\Sigma_c S_N^\mathsf{T}\right]}_{\widetilde{M}}, \qquad (27)$$

where $\widetilde{S}_N$ is obtained by permuting, bottom to top, the block matrices of S_N (25). Right multiplying either sides of (27) with $\overline{\Sigma}_c^{-1}\left[\mathbf{1}_N \otimes \overline{K}\right]$ gives us:

$$N\overline{K}^\mathsf{T}\overline{K} = \left[\overline{K}^\mathsf{T}\overline{K}\Sigma_c + \sigma_v^2 I\right]F + P, \qquad (28)$$

where $P = \widetilde{M}\overline{\Sigma}_c^{-1}[\mathbf{1}_N \otimes \overline{K}]$ and F is defined in (26). Since $\overline{K}^\mathsf{T}\overline{K}\Sigma_c \succeq 0$, from (28), it follows that $F = [\overline{K}^\mathsf{T}\overline{K}\Sigma_c + \sigma_v^2 I]^{-1}[N\overline{K}^\mathsf{T}\overline{K} - P]$. Substituting F in (26) yields

$$\widehat{\eta}^2 = N\boldsymbol{\mu}_\Delta^\mathsf{T}([\overline{K}^\mathsf{T}\overline{K}\Sigma_c + \sigma_v^2 I]^{-1}\overline{K}^\mathsf{T}\overline{K})\boldsymbol{\mu}_\Delta + \epsilon(N),$$

where $\epsilon(N) = -\boldsymbol{\mu}_\Delta^\mathsf{T}[\overline{K}^\mathsf{T}\overline{K}\Sigma_c + \sigma_v^2 I]^{-1}P\boldsymbol{\mu}_\Delta + t(S_N)$. Finally, substituting $K = L\Sigma_c^{-\frac{1}{2}}$ in the above expression, and manipulating the terms will give us

$$\widehat{\eta}^2 = N\widetilde{\boldsymbol{\mu}}_\Delta^\mathsf{T}\left([L^\mathsf{T}L + \sigma_v^2 I]^{-1}L^\mathsf{T}L\right)\widetilde{\boldsymbol{\mu}}_\Delta + \epsilon(N). \qquad (29)$$

We claim that $\lim_{N\to\infty} \epsilon(N) = 0$. To see this, rewrite $\epsilon(N)$ as $\boldsymbol{\mu}_\Delta^\mathsf{T}(Q_1(N) + Q_2(N) + Q_3(N))\boldsymbol{\mu}_\Delta$, where

$$Q_1(N) = S_N^\mathsf{T}\left[\overline{\Sigma}_c^{-1}S_N + 2\Sigma_c^{-1}\left[\mathbf{1}_N \otimes \overline{K}\right]\right]$$
$$Q_2(N) = [\overline{K}^\mathsf{T}\overline{K}\Sigma_c + \sigma_v^2 I]^{-1}\overline{K}^\mathsf{T}\widetilde{S}_N^\mathsf{T}(I \otimes \Sigma_c)\mathcal{F}^\mathsf{T}$$
$$Q_3(N) = [\overline{K}^\mathsf{T}\overline{K}\Sigma_c + \sigma_v^2 I]^{-1}\overline{K}^\mathsf{T}\overline{K}\Sigma_c S_N^\mathsf{T}.$$

From part 1) of Assumption 3.3, there exist a $k \in \mathbb{N}$ such that for all $m \in \{k, k+1, \ldots, N\}$, $S(m) - \overline{K} = 0$. Thus, all but finite rows of S_N (25) are zeros, i.e., we can express S_N^T as $[F_1^\mathsf{T}\ 0^\mathsf{T}]$ and $\widetilde{S}_N^\mathsf{T}$ as $[0^\mathsf{T}\ F_2^\mathsf{T}]$, where the dimension of F_1 and F_2 depends only k. Thus, for all $N > k$, $Q_i(N)$ is a constant matrix, say Q_i, and we may conclude that

$$\|\boldsymbol{\mu}_\Delta\|_2^2 \sum_{i=1}^3 \lambda_{\min}(Q_i + Q_i^\mathsf{T}) \leq 2\sum_{i=1}^3 \boldsymbol{\mu}_\Delta^\mathsf{T}Q_i\boldsymbol{\mu}_\Delta$$
$$\leq \|\boldsymbol{\mu}_\Delta\|_2^2 \sum_{i=1}^3 \lambda_{\max}(Q_i + Q_i^\mathsf{T}),$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ are the maximum and minimum eigenvalues. Since $\lim_{N\to\infty} N\|\boldsymbol{\mu}_\Delta\|_2 = c$ (Assumption 3.3), it follows that $\lim_{N\to\infty} \|\boldsymbol{\mu}_\Delta\|_2 = 0$. Hence, $\lim_{N\to\infty} \epsilon(N) = 0$ and $\lim_{N\to\infty} \widehat{\eta} = \eta$ (13).

part 2) Under the assumption that $N < \infty$, let $\widehat{\mathbb{P}}_{e_v}(\mathcal{J})$ be the error probability of (10). Then, from (9), we have

$$\Pr\left(\widehat{H} = H_2|H_1\right) = \Pr\left(\ln(\widehat{R}) > \left[\frac{Z^2}{\mathbf{b}^\mathsf{T}\overline{\Sigma}_2\mathbf{b}} - \frac{Z^2}{\mathbf{b}^\mathsf{T}\overline{\Sigma}_1\mathbf{b}}\right]|H_1\right),$$
$$\Pr\left(\widehat{H} = H_1|H_2\right) = \Pr\left(\ln(\widehat{R}) < \left[\frac{Z^2}{\mathbf{b}^\mathsf{T}\overline{\Sigma}_2\mathbf{b}} - \frac{Z^2}{\mathbf{b}^\mathsf{T}\overline{\Sigma}_1\mathbf{b}}\right]|H_2\right),$$

where $Z = \mathbf{b}^\mathsf{T}[\mathbf{Y}_J - \overline{\boldsymbol{\mu}}_c]$ and $\widehat{R} = (\mathbf{b}^\mathsf{T}\overline{\Sigma}_1\mathbf{b}/(\mathbf{b}^\mathsf{T}\overline{\Sigma}_2\mathbf{b}) > 1$ (since $\Sigma_2 = 0$; Assumption 3.3). Let $U \sim \mathcal{N}(0,1)$. Then, $Z|H_i \overset{d}{=} (\sqrt{\mathbf{b}^\mathsf{T}\overline{\Sigma}_i\mathbf{b}})U$, where $\overset{d}{=}$ means equality in the distribution. From this fact, we now have $\Pr(\widehat{H} = H_2|H_1) = \Pr(\widehat{\tau} > U^2)$ and $\Pr(\widehat{H} = H_1|H_2) = \Pr(U^2 > \widehat{\tau}\widehat{R})$, where $\widehat{\tau} = \ln(\widehat{R})/(\widehat{R} - 1)$. Since $U^2 \sim \chi^2(1)$, we finally have

$$\widehat{\mathbb{P}}_{e_v}(\mathcal{J}) = 0.5\left[1 - Q_{\chi^2}(1, \widehat{\tau})\right] + 0.5\,Q_{\chi^2}(1, \widehat{\tau}\widehat{R}).$$

To simplify $\widehat{R}$, note the following: since $\mathbf{b}$ is the maximizer of I-divergence (7), from [19], we can also express $\widehat{R}$ as

$$\widehat{R} = \frac{\mathbf{b}^\mathsf{T}\overline{\Sigma}_1\mathbf{b}}{\mathbf{b}^\mathsf{T}\overline{\Sigma}_2\mathbf{b}} = \max_{\mathbf{d} \in \mathbb{R}^{mN}} \frac{\mathbf{d}^\mathsf{T}\overline{\Sigma}_1\mathbf{d}}{\mathbf{d}^\mathsf{T}\overline{\Sigma}_2\mathbf{d}}.$$

Let $\mathbf{c} = \overline{\Sigma}_2^{1/2}\mathbf{d}$, and note the following:

$$\widehat{R} = \max_{\mathbf{c} \in \mathbb{R}^{mN}} \left(\frac{\mathbf{c}}{\|\mathbf{c}\|_2}\right)^\mathsf{T} \overline{\Sigma}_2^{-1/2}\overline{\Sigma}_1\overline{\Sigma}_2^{-1/2}\left(\frac{\mathbf{c}}{\|\mathbf{c}\|_2}\right)$$
$$= \lambda_{\max}\left(\overline{\Sigma}_2^{-1/2}\overline{\Sigma}_1\overline{\Sigma}_2^{-1/2}\right) = \lambda_{\max}\left(\overline{\Sigma}_1\overline{\Sigma}_2^{-1}\right).$$

Since $\widehat{R}$ is an increasing sequence, with respect to N (see Proposition A.1), the limits $R = \lim_{N\to\infty} \widehat{R}$, $\tau = \lim_{N\to\infty} \widehat{\tau}$ and $\lim_{N\to\infty} \widehat{\tau}\widehat{R} = \tau R$ are well defined. Now, consider

$$
\begin{aligned}
\mathbb{P}_{e_v}(\mathcal{J}) &= \lim_{N\to\infty} \widehat{\mathbb{P}}_{e_v}(\mathcal{J}) \\
&= \lim_{N\to\infty} 0.5\left[1 - Q_{\chi^2}\left(1, \widehat{\tau}\right)\right] + 0.5\, Q_{\chi^2}(1, \widehat{\tau}\widehat{R}) \\
&= 0.5\left[1 - Q_{\chi^2}\left(1, \tau\right)\right] + 0.5\, Q_{\chi^2}(1, \tau R),
\end{aligned}
$$

where the last equality follows because $\widehat{\tau}$ and $\widehat{\tau}\widehat{R}$ are decreasing and increasing in N (Proposition A.1), resp.

We now show that R is given by (14). Since $\Sigma_2 = 0$ and $\mathbf{x}[0] = 0$, we have $\overline{\Sigma}_2 = \sigma_v^2 I$ and $\overline{\Sigma}_1 = FF^{\mathsf{T}} + \sigma_v^2 I$, where $F = \mathcal{F}(I_N \otimes \Sigma_1^{\frac{1}{2}})$ and $\Sigma_1^{\frac{1}{2}}$ satisfies $\Sigma_1 = \Sigma_1^{\frac{1}{2}}\Sigma_1^{\frac{1}{2}}$. From these observations, we may conclude that

$$
\begin{aligned}
R = \lim_{N\to\infty} \widehat{R} &= \lim_{N\to\infty} \frac{\lambda_{\max}(FF^{\mathsf{T}} + \sigma_v^2 I)}{\sigma_v^2} \\
&= 1 + \sigma_v^{-2} \lim_{N\to\infty} \lambda_{\max}(FF^{\mathsf{T}}). \quad (30)
\end{aligned}
$$

It now suffices to evaluate $\lim_{N\to\infty} \lambda_{\max}(FF^{\mathsf{T}})$. Since $\overline{\lambda}(G) < 1$, we may define the following matrix valued function [26]:

$$
A(\omega) = \sum_{l=0}^{\infty} K(l)\Sigma_1^{1/2} e^{jk\omega} \quad \omega \in [0, 2\pi],
$$

where $K(l) = CG^l\Pi$ and $j = \sqrt{-1}$. Since the coefficients $K(l)\Sigma_1^{1/2}$ are absolutely summable, for any $l \in \mathbb{N}$, these coefficients can also be recovered as [26]:

$$
K(l)\Sigma_1^{1/2} = \frac{1}{2\pi}\int_0^{2\pi} A(\omega)e^{-jl\omega}d\omega.
$$

Let $\overline{z}$ be the conjugate of $z \in \mathbb{C}$. Then, from [27, Chapter 6.4], we have

$$
\begin{aligned}
\lim_{N\to\infty} \lambda_{\max}^{1/2}(FF^{\mathsf{T}}) &= \operatorname*{ess\,sup}_{\omega\in[0,2\pi]} \|A(\omega)\|_2 \\
&= \operatorname*{ess\,sup}_{\omega\in[0,2\pi]} \left\|C\left(\sum_{l=0}^{\infty} G^l e^{jl\omega}\right)\Pi\Sigma_1^{1/2}\right\|_2 \\
&= \operatorname*{ess\,sup}_{\omega\in[0,2\pi]} \left\|C\left(I - Ge^{jw}\right)^{-1}\Pi\Sigma_1^{1/2}\right\|_2 \\
&= \operatorname*{ess\,sup}_{\{z\in\mathbb{C}:|z|=1\}} \left\|C\left(\overline{z}I - G\right)^{-1}\Pi\Sigma_1^{1/2}\right\|_2 \\
&\overset{(a)}{=} \operatorname*{ess\,sup}_{\{z\in\mathbb{C}:|z|=1\}} \left\|C\left(zI - G\right)^{-1}\Pi\Sigma_1^{1/2}\right\|_2 \\
&= \operatorname*{ess\,sup}_{\{z\in\mathbb{C}:|z|=1\}} \left\|T(z)\Sigma_1^{\frac{1}{2}}\right\|_2 = \|T(z)\Sigma_1^{\frac{1}{2}}\|_\infty.
\end{aligned}
$$

where (a) follows because, for any $A \in \mathbb{C}^{N\times N}$ with A^* denoting its complex conjugate transpose, $\|A\|_2 = \|A^{\mathsf{T}}\|_2 = \|A^*\|_2$. Substituting $\lim_{N\to\infty} \lambda_{\max}^{1/2}(FF^{\mathsf{T}})$ in (30) gives us $R = 1 + \sigma_v^{-2}\|T^*(z)\|_\infty^2$. $\qquad\square$

Proof of Theorem 4.1 Let $\mathbf{y}_{\mathcal{P}}[k]$ and $\mathbf{y}_{\mathcal{S}}[k]$ denote the measurements of associated with the sensor sets $\mathcal{P}$ and $\mathcal{C}$, respectively. Since $\sigma_v^2 = 0$, from (18), we have

$$
\mathbf{y}_{\mathcal{P}}[k+1] = G_{pp}\mathbf{y}_{\mathcal{P}}[k] + B\mathbf{y}_{\mathcal{C}}[k], \quad (31)
$$

where $B = G_{pc}$. From (31), it follows that

$$
\underbrace{\begin{bmatrix}\mathbf{y}_{\mathcal{P}}[1]\\ \mathbf{y}_{\mathcal{P}}[2]\\ \vdots \\ \mathbf{y}_{\mathcal{P}}[N]\end{bmatrix}}_{\mathbf{Y}_{\mathcal{P}}} = \underbrace{\begin{bmatrix}G_{pp} & B \\ G_{pp}^2 & G_{pp}B \\ \vdots & \vdots \\ G_{pp}^N & G_{pp}^{N-1}B\end{bmatrix}}_{M} \underbrace{\begin{bmatrix}\mathbf{y}_{\mathcal{P}}[0]\\ \mathbf{y}_{\mathcal{C}}[0]\end{bmatrix}}_{\widehat{\mathbf{Y}}[0]}
$$

$$
+ \underbrace{\begin{bmatrix}0 & 0 & \cdots & 0 & 0\\ B & 0 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ G_{pp}^{N-2}B & G_{pp}^{N-3}B & \cdots & B & 0\end{bmatrix}}_{\widehat{M}} \underbrace{\begin{bmatrix}\mathbf{y}_{\mathcal{C}}[1]\\ \mathbf{y}_{\mathcal{C}}[2]\\ \vdots\\ \mathbf{y}_{\mathcal{C}}[N]\end{bmatrix}}_{\mathbf{Y}_{\mathcal{C}}}.
$$

Since $\widehat{\mathbf{Y}}[0]$ is independent of H_i, the assertion of the theorem follows from Lemma 4.2. $\qquad\square$

Proof of Lemma 4.2 We shall prove the result assuming that $\mathbf{Y}$ and $\mathbf{Z} = g(\mathbf{Y}) + \mathbf{v}$ admits density functions. With the expense of notation, the given proof can be adapted to handle random variables that do not have densities. Let $\mathbf{L} = [\mathbf{Y}^{\mathsf{T}}, \mathbf{Z}^{\mathsf{T}}]^{\mathsf{T}}$. Consider the following log likelihood ratio (LR) based on $\mathbf{L}$:

$$
\begin{aligned}
\frac{f(l|M_2)}{f(l|M_1)} &= \frac{f(y, g(y) + v|M_2)}{f(y, g(y) + v|M_1)} \\
&= \frac{f(y, g(y) + v|y, M_2)f(y|M_2)}{f(y, g(y) + v|y, M_1)f(y|M_1)} \\
&\overset{(a)}{=} \frac{f(y, g(y) + v|y)f(y|M_2)}{f(y, g(y) + v|y)f(y|M_1)} = \frac{f(y|M_2)}{f(y|M_1)},
\end{aligned}
$$

where (a) follows because $\mathbf{v}$ is independent of M_i. Since LRs of $\mathbf{L}$ and $\mathbf{Y}$ are equal, the error probabilities associated with their MAP rules should be the same. Instead, the error probability of the MAP rule based on $\mathbf{L}$ is always superior to that of $\mathbf{Y}$ or $\mathbf{Z}$ alone. Thus $\delta_1 \leq \delta_2$. $\qquad\square$

Proof of Theorem 4.4 Consider the following deterministic analogue of (1): $\mathbf{x}[k+1] = G\mathbf{x}[k] + \Pi\mathbf{u}$, where $\mathbf{u}$ is arbitrary. Recall that $\mathbf{x}_p[k+1] = G_{pp}\mathbf{x}_p[k] + G_{pc}\mathbf{x}_c[k]$ (18). Since $\mathbf{x}[0] = 0$, for $z \notin \operatorname{spec}(G) \cup \operatorname{spec}(G_{pp})$, we have

$$
\mathbf{x}[z] = (zI - G)^{-1}\Pi\mathbf{u} \text{ and} \quad (32a)
$$

$$
\mathbf{x}_p[z] = (zI - G_{pp})^{-1}G_{pc}\mathbf{x}_c[z] = T_s(z)\mathbf{x}_c[z]. \quad (32b)
$$

From (32b), the following inequalities are obvious

$$
\underline{\rho}(z)\|\mathbf{x}_c[z]\|_2 \leq \|\mathbf{x}_p[z]\|_2 \leq \overline{\rho}(z)\|\mathbf{x}_c[z]\|_2. \quad (33)
$$

Let C_1 and C_2 be the sensor matrices associated with $\mathcal{C}$ and $\mathcal{P}$, respectively. Then,

$$
\mathbf{x}_c[z] = C_1\mathbf{x}[z] \text{ and } \mathbf{x}_p[z] = C_1\mathbf{x}[z]. \quad (34)
$$

part 1) We now consider the cases 1a) and 1b). Let $L_i = C_i(I - G)^{-1}\Pi\Sigma_c^{\frac{1}{2}}$, where $\Sigma_c = \Sigma_c^{\frac{1}{2}}\Sigma_c^{\frac{1}{2}}$ is defined in Lemma 3.2. Let $z = 1$. Then, from (34) note that

$$\|\mathbf{x}_c[1]\|_2^2 = \|C_1\mathbf{x}[1]\|_2^2 = \mathbf{u}^\mathsf{T} L_1^\mathsf{T} L_1 \mathbf{u} \text{ and}$$
$$\|\mathbf{x}_p[1]\|_2^2 = \|C_2\mathbf{x}[1]\|_2^2 = \mathbf{u}^\mathsf{T} L_2^\mathsf{T} L_2 \mathbf{u}.$$

From (33) and above identities, it follows that

$$\begin{aligned} \overline{\rho}(1) < 1 &\implies L_1^\mathsf{T} L_1 + \sigma_v^2 I \succ L_2^\mathsf{T} L_2 + \sigma_v^2 I \text{ and} \\ \underline{\rho}(1) > 1 &\implies L_2^\mathsf{T} L_2 + \sigma_v^2 I \succ L_1^\mathsf{T} L_1 + \sigma_v^2 I. \end{aligned} \quad (35)$$

Let $\mathbf{u} = \widetilde{\boldsymbol{\mu}}_\Delta$, where $\widetilde{\boldsymbol{\mu}}_\Delta$ is defined in the statement of Lemma 3.4. Let η_1 and η_2 be the SNRs of $\mathbb{P}_{e_m}(\mathcal{C})$ and $\mathbb{P}_{e_m}(\mathcal{P})$, respectively. Then from (13), we have

$$\eta_i^2 = N\widetilde{\boldsymbol{\mu}}_\Delta^\mathsf{T} \left([L_i^\mathsf{T} L_i + \sigma_v^2 I]^{-1} L_i^\mathsf{T} L_i\right) \widetilde{\boldsymbol{\mu}}_\Delta.$$

Using the identity $[L_i^\mathsf{T} L_i + \sigma_v^2 I]^{-1} L_i^\mathsf{T} L_i = I - \sigma_v^2 [L_i^\mathsf{T} L_i + \sigma_v^2 I]^{-1}$, we can also express η_i^2 as

$$\eta_i^2 = \widetilde{\boldsymbol{\mu}}_\Delta^\mathsf{T}\widetilde{\boldsymbol{\mu}}_\Delta - \sigma_v^2 \widetilde{\boldsymbol{\mu}}_\Delta^\mathsf{T} \left[L_i^\mathsf{T} L_i + \sigma^2 I\right]^{-1} \widetilde{\boldsymbol{\mu}}_\Delta^\mathsf{T}. \quad (36)$$

Finally, from (36) and (35), and Proposition 3.6, we have

$$\begin{aligned} \overline{\rho}(1) < 1 &\implies \eta_1^2 \geq \eta_2^2 \implies \mathbb{P}_{e_m}(\mathcal{C}_d) \leq \mathbb{P}_{e_m}(\mathcal{P}) \text{ and} \\ \underline{\rho}(1) > 1 &\implies \eta_1^2 \leq \eta_2^2 \implies \mathbb{P}_{e_m}(\mathcal{C}_d) \geq \mathbb{P}_{e_m}(\mathcal{P}). \end{aligned}$$

part 2) We now consider the cases 2a) and 2b). Let $T_i(z) = C_i(zI - G)^{-1}$. Let $\mathbf{u} = \Sigma_1^{1/2}\mathbf{d}$, where $\Sigma_1^{1/2}$ is defined in the statement of Lemma 3.4. From (34) and (32a), we have $\mathbf{x}_c[k] = T_1(z)\Sigma_1^{1/2}\mathbf{d}$ and $\mathbf{x}_c[k] = T_2(z)\Sigma_1^{1/2}\mathbf{d}$. By invoking these two facts in (33), we may now conclude that

$$\sup_{|z|=1} \overline{\rho}(z) < 1 \implies \|T_2(z)\Sigma_1^{\frac{1}{2}}\mathbf{d}\|_2 \leq \|T_1(z)\Sigma_1^{\frac{1}{2}}\mathbf{d}\|_2 \text{ and}$$

$$\inf_{|z|=1} \underline{\rho}(z) > 1 \implies \|T_2(z)\Sigma_1^{\frac{1}{2}}\mathbf{d}\|_2 \geq \|T_1(z)\Sigma_1^{\frac{1}{2}}\mathbf{d}\|_2,$$

for all z that satisfies $|z| = 1$. Let R_1 and R_2 be the SNRs of $\mathbb{P}_{e_v}(\mathcal{C})$ and $\mathbb{P}_{e_v}(\mathcal{P})$, respectively. Then, from (14)

$$R_i - 1 = \frac{\|T_i(z)\Sigma_1^{\frac{1}{2}}\|_\infty^2}{\sigma_v^2} = \left(\operatorname*{ess\,sup}_{\{z\in\mathbb{C}:|z|=1\}} \|T_2(z)\Sigma_1^{1/2}\mathbf{d}\|_2\right)^2.$$

From Proposition 3.6, it follows that

$$\begin{aligned} \sup_{|z|=1} \overline{\rho}(z) < 1 &\implies R_1 \geq R_2 \implies \mathbb{P}_{e_v}(\mathcal{C}_d) \leq \mathbb{P}_{e_v}(\mathcal{P}) \text{ and} \\ \inf_{|z|=1} \underline{\rho}(z) > 1 &\implies R_1 \leq R_2 \implies \mathbb{P}_{e_v}(\mathcal{C}_d) \geq \mathbb{P}_{e_v}(\mathcal{P}). \quad \square \end{aligned}$$

Proof of Corollary 4.5 We shall prove part 1) of the corollary, and part 2) can be derived using similar analysis (the details are omitted). The idea of the proof is to show that $\|\widetilde{G}\|_\infty \leq 1/\sqrt{m} \implies \overline{\rho}(1) < 1$, and there upon invoking Theorem 4.4 yields the desired assertion.

step 1) For $G \geq 0$, it follows that $\sup_{|z|=1} \overline{\rho}(z) = \overline{\rho}(1)$, where $\overline{\rho}(z)$ is $\|(zI - G_{pp})^{-1}G_{pc}\|_2$. To see this, note the following: For any $\mathbf{d} \in \mathbb{C}^{n_1}$, let $|\mathbf{d}| = (|d_1|,\ldots,|d_{n_1}|)^\mathsf{T}$. Then, for any $l \in \mathbb{N}$ and z that satisfies $|z| = 1$, we have

$$|(\overline{z}G_{pp})^l G_{pc}\mathbf{d}| = |(G_{pp})^l G_{pc}\mathbf{d}| \leq (G_{pp})^l G_{pc}|\mathbf{d}|,$$

where the inequality, to be understood coordinate wise, follows because $[G_{pp}\ G_{pc}] \geq 0$. From the above inequality, and the fact $|\mathbf{y} + \mathbf{z}| \leq |\mathbf{y}| + |\mathbf{z}|$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{C}^p$, we have

$$\left|\sum_{l=0}^{\infty} (\overline{z}G_{pp})^l G_{pc}\mathbf{d}\right| \leq \sum_{l=0}^{\infty} |(\overline{z}G_{pp})^l G_{pc}\mathbf{d}| \leq \sum_{l=0}^{\infty} (G_{pp})^l G_{pc}|\mathbf{d}|.$$

Since G_{pp} is a submatrix of $G \geq$, which is a non-negative matrix, we have $|\lambda_{\max}(\overline{z}G_{pp})| = |\lambda_{\max}(G_{pp})| \leq |\lambda_{\max}(G)| < 1$ [28], the above inequality can also be expressed as

$$\left|(I - \overline{z}G_{pp})^{-1}G_{pc}\mathbf{d}\right| \leq (I - G_{pp})^{-1}G_{pc}|\mathbf{d}|.$$

Taking 2-norm on both sides of the inequality yields us:

$$\left\|\,|(I - \overline{z}G_{pp})^{-1}G_{pc}\mathbf{d}|\,\right\|_2 \leq \left\|\,(I - G_{pp})^{-1}G_{pc}|\mathbf{d}|\,\right\|_2.$$

Since the above inequality holds for any vector $\mathbf{d} \in \mathbf{R}^{n_1}$, using the identity $\|\,|\mathbf{x}|\,\|_2 = \|\mathbf{x}\|_2$ for any $\mathbf{x} \in \mathbb{C}^p$, the following inequality is now obvious:

$$\sup_{|z|=1} \sup_{\|\mathbf{d}\|_2=1} \|(I - \overline{z}G_{pp})^{-1}G_{pc}\mathbf{d}\|_2 \leq \left\|\,(I - G_{pp})^{-1}G_{pc}\,\right\|_2,$$

which can be expressed as $\sup_{|z|=1} \overline{\rho}(z) \leq \overline{\rho}(1)$. The equality is attained at $z = 1$.

step 2) Since $\sup_{|z|=1} \overline{\rho}(z) = \overline{\rho}(1)$, from Theorem 4.4 it readily follows that, both $\mathbb{P}_{e_m}(\mathcal{C}_d) \leq \mathbb{P}_{e_m}(\mathcal{P})$ and $\mathbb{P}_{e_v}(\mathcal{C}_d) \leq \mathbb{P}_{e_v}(\mathcal{P})$ holds true whenever $\overline{\rho}(1) < 1$. We now show that $\|\widetilde{G}\|_\infty = \|[G_{pp}\ G_{pc}]\|_\infty < 1/\sqrt{m_1}$ guarantees $\overline{\rho}(1) < 1$. Let $\mathbf{1}$ denote the all ones vector, and note the following identity:

$$\begin{bmatrix} G_{pp} & G_{pc} \\ 0 & I \end{bmatrix}^k \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} G_{pp}^k & \sum_{l=0}^{k-1} G_{pp}^l G_{pc} \\ 0 & I \end{bmatrix}^k \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}. \quad (37)$$

Since $\|[G_{pp}\ G_{pc}]\|_\infty < 1/\sqrt{m_1}$, for any $k \in \mathbb{N}$, we also have

$$\begin{bmatrix} G_{pp} & G_{pc} \\ 0 & I \end{bmatrix}^k \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \leq \begin{bmatrix} \frac{1}{\sqrt{m_1}}\mathbf{1} \\ \mathbf{1} \end{bmatrix}.$$

From the above inequality and (37), it follows that

$$\lim_{k\to\infty} \begin{bmatrix} G_{pp}^k & \sum_{l=0}^{k-1} G_{pp}^l G_{pc} \\ 0 & I \end{bmatrix}^k \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} \leq \begin{bmatrix} \frac{1}{\sqrt{m_1}}\mathbf{1} \\ \mathbf{1} \end{bmatrix}.$$

Since $\overline{\lambda}(G_pp) < 1$, as $k \to \infty$, it follows that $G_{pp}^k \to 0$ and $\sum_{l=0}^{k-1} G_{pp}^l G_{pc} \to (I-G_{pp})^{-1}G_{pc}$. Thus $(I-G_{pp})^{-1}G_{pc}\mathbf{1} = \|(I - G_{pp})^{-1}G_{pc}\|_\infty < 1/\sqrt{m_1}$, and hence, $\overline{\rho}(1) = \|(I - G_{pp})^{-1}G_{pc}\|_2 < \sqrt{m_1}\|(I - G_{pp})^{-1}G_{pc}\|_\infty < 1$. $\square$

Proof of Proposition 3.6 Since $Q_\mathcal{N}(x)$ is decreasing function of x, $\mathbb{P}_{e_m}(\mathcal{J})$ (11) is decreasing in SNR η, given by

either (13) or (15). For $\mathbb{P}_{e_v}(\mathcal{J})$ (12) note the following: first, observe that $R > 1$ in both (14) and (16). Thus

$$\frac{d\tau}{dR} = \frac{\left(\frac{R-1}{R}\right) - \ln R}{(R-1)^2} < 0, \text{ and}$$
$$\frac{d(\tau R)}{dR} = \frac{(R-1) - \ln R}{(R-1)^2} > 0. \tag{38}$$

Hence, we conclude that τ is deceasing in R. Instead, τR is increasing in R. From this observation and the fact that the $Q_{\chi^2}(1, z) = \Pr[Z \geq z]$, where $Z \sim \chi^2(1)$, is decreasing in z, it follows that $\mathbb{P}_{e_v}(\mathcal{J})$ is decreasing in R. $\square$

Proof of Proposition 4.7 From (19), and the fact that $1 \leq q < j < \ldots < n$, where $\mathcal{C}_d = \{j\}$ and $\mathcal{P} = \{j+1, \ldots, n\}$, the row sums of $\widetilde{G}$ takes values in the set $\{a+c, a+b+c\}$. Let $|\overline{G}_{l,q}| = |(I-G)^{-1}_{l,q}|$. Using the principle of backward induction, we shall show that, when $\|\widetilde{G}\|_\infty = a+b+c < 1$, $\{|\overline{G}_{lq}|\}^n_{l=q}$ is monotonically decreasing. The proof of part (ii) is left to the reader as an exercise.

Let $\widetilde{a} = 1 - a$, $\widetilde{b} = -b$, $\widetilde{c} = -c$. If $\widetilde{a} \neq 0$, then $\overline{G}_{l,q}$ of $(I-G)^{-1}$ are given by the following expressions [29]:

$$\overline{G}_{l,q} = \frac{1}{\theta_n} \begin{cases} (-1)^{l+q}\widetilde{b}^{q-l}\theta_{l-1}\phi_{q+1} & q \geq l \\ (-1)^{l+q}\widetilde{c}^{l-q}\theta_{q-1}\phi_{l+1} & q < l \end{cases} \tag{39}$$

where $l, q \in \{1, \ldots, n\}$, and θ_k and ϕ_k are governed by

$$\theta_k = \widetilde{a}\theta_{k-1} - \widetilde{b}\widetilde{c}\theta_{k-2} \quad \text{for } k = 2, \ldots, n$$
$$\phi_k = \widetilde{a}\phi_{k+1} - \widetilde{b}\widetilde{c}\phi_{k+2} \quad \text{for } k = n-1, \ldots, 1 \tag{40}$$

where $\theta_0 = 1$, $\theta_1 = \widetilde{a}$, $\phi_n = \widetilde{a}$, $\phi_{n+1} = 1$ and $\theta_n = \det(I - G)$. Let $\mathcal{L} = \{q+1, \ldots, n\}$. Then, for any $l \in \mathcal{L} \cup \{q\}$,

$$|\overline{G}_{lq}| \triangleq |(I-G)^{-1}_{lq}| = \left| \frac{\theta_{q-1}\widetilde{c}^{-q}}{\theta_n} \right| \left| \widetilde{c}^l \phi_{l+1} \right|,$$

Let $l \in \mathcal{L}$, and define $\zeta(l) = |\overline{G}_{l,q}|/|\overline{G}_{l-1,q}|$. Since $\phi_{n+1} = 1$ and $\phi_n = \widetilde{a}$, for $l = n$ (base step), it follows that

$$\zeta(n) = \frac{|\overline{G}_{nq}|}{|\overline{G}_{n-1,q}|} = \frac{|\widetilde{c}^n\phi_{n+1}|}{|\widetilde{c}^{n-1}\phi_n|} = \frac{|\widetilde{c}|}{|\widetilde{a}|} = \frac{c}{1-a} \overset{(i)}{<} 1,$$

where (i) follows because $a, b, c > 0$, and $a+b+c < 1$. Let $q < l < n$ and $\zeta(l+1) < 1$ (inductive step). Then,

$$\zeta(l) = \frac{|\overline{G}_{l,q}|}{|\overline{G}_{l-1,q}|} = \frac{|\widetilde{c}|\,|\phi_{l+1}|}{|\phi_l|} \overset{(40)}{=} \frac{c}{\left| \widetilde{a} - \widetilde{b}\widetilde{c}\left(\frac{\phi_{l+2}}{\phi_{l+1}}\right) \right|} < 1,$$

To see the last inequality, consider the following:

$$b + c < 1 - a \overset{(ii)}{\implies} b\left(\frac{|\widetilde{c}|\,|\phi_{l+2}|}{|\phi_{l+1}|}\right) + c < 1 - a$$
$$\implies b\left(\frac{c\phi_{l+2}}{\phi_{l+1}}\right) + c < 1 - a$$
$$\implies \frac{c}{\left| (1-a) - bc\left(\frac{\phi_{k+2}}{\phi_{k+1}}\right) \right|} < 1$$
$$\overset{(iii)}{\implies} \frac{|\widetilde{c}|}{\left| \widetilde{a} - \widetilde{b}\widetilde{c}\left(\frac{\phi_{l+2}}{\phi_{l+1}}\right) \right|} < 1,$$

where (ii) follows because the hypothesis $\zeta(l+1) < 1$ implies that $|\widetilde{c}|(|\phi_{l+2}|/|\phi_{l+1}|) < 1$, and (iii) from the fact that $\widetilde{a} = 1 - a$, $\widetilde{b} = -b$, and $\widetilde{c} = -c$. From the principle of finite induction, for all $l \in \mathcal{L}$, we have $\zeta(l) < 1$. Hence, $\{|\overline{G}_{lq}|\}^n_{l=q}$ is a decreasing sequence. $\square$

Proposition A.1. *Let $\widehat{\eta}^2 = \overline{\boldsymbol{\mu}}^\mathsf{T}_\Delta \overline{\Sigma}^{-1}_c \overline{\boldsymbol{\mu}}_\Delta$, $\widehat{R} = \lambda_{max}(\overline{\Sigma}_1 \overline{\Sigma}^{-1}_2)$ and $\widehat{\tau} = \ln(\widehat{R})/(\widehat{R} - 1)$, where $(\overline{\boldsymbol{\mu}}_\Delta, \overline{\Sigma}_c, \overline{\Sigma}_1, \overline{\Sigma}_2)$ are defined in the statement of Lemma 3.2. Then, $\widehat{\eta}$, $\widehat{R}$, and τ are increasing in N. However, $\widehat{\tau}\widehat{R}$ is decreasing in N.*

Proof: Let $N < \infty$. Then, from Proposition 3.1, we have $\overline{\boldsymbol{\mu}}_\Delta = \mathbb{E}[\mathbf{Y}_\mathcal{J}|H_2] - \mathbb{E}[\mathbf{Y}_\mathcal{J}|H_1]$, $\overline{\Sigma}_c = \text{Cov}[\mathbf{Y}_\mathcal{J}|H_1] = \text{Cov}[\mathbf{Y}_\mathcal{J}|H_2]$. For clarity, we drop the existing subscripts and replace them with the total number of measurements. Let $N_2 = N_1 + k$, $k \in \mathbb{N}$, and consider $\mathbf{Y}^\mathsf{T}_{N_2} = [\mathbf{Y}^\mathsf{T}_{N_1}, \mathbf{Z}^\mathsf{T}_k]$, where $\mathbf{Z}_k$ are the measurements collected after N_1. Then,

$$\overline{\boldsymbol{\mu}}_{N_2} = \begin{bmatrix} \overline{\boldsymbol{\mu}}_{N_1} \\ \mathbf{m}_k \end{bmatrix} \text{ and } \overline{\Sigma}_{N_2} = \left[\begin{array}{c|c} \overline{\Sigma}_{N_1} & D \\ \hline D^\mathsf{T} & M \end{array} \right],$$

where $\mathbf{m}_k = \mathbb{E}[\mathbf{Z}_k|H_2] - \mathbb{E}[\mathbf{Z}_k|H_1]$, $M = \text{Cov}[\mathbf{Z}_k|H_1] > 0$, and $D = \text{Cov}[\mathbf{Y}_k, \mathbf{Z}_k|H_1]$. Further, using the Schur complement, $\overline{\Sigma}^{-1}_{N_2}$ can be expressed as

$$\overline{\Sigma}^{-1}_{N_2} = \left[\begin{array}{c|c} \overline{\Sigma}_{N_1} & D \\ \hline D^\mathsf{T} & M \end{array} \right]^{-1} = \left[\begin{array}{c|c} \overline{\Sigma}^{-1}_{N_1} & 0 \\ \hline 0^\mathsf{T} & 0 \end{array} \right] + \underbrace{F}_{>0}.$$

From the above identity, it follows that

$$\widehat{\eta}_{N_2} = \left(\overline{\boldsymbol{\mu}}^\mathsf{T}_{N_2} \overline{\Sigma}^{-1}_{N_2} \overline{\boldsymbol{\mu}}_{N_2} \right)^{\frac{1}{2}} = \left(\overline{\boldsymbol{\mu}}^\mathsf{T}_{N_1} \overline{\Sigma}^{-1}_{N_1} \overline{\boldsymbol{\mu}}_{N_1} + \overline{\boldsymbol{\mu}}^\mathsf{T}_{N_2} F \overline{\boldsymbol{\mu}}_{N_2} \right)^{\frac{1}{2}}$$
$$\geq \left(\overline{\boldsymbol{\mu}}^\mathsf{T}_{N_1} \overline{\Sigma}^{-1}_{N_1} \overline{\boldsymbol{\mu}}_{N_1} \right) = \widehat{\eta}_{N_1}.$$

Hence, we may conclude that $\widehat{\eta}$ is increasing in N. Instead, from the eigenvalue interlacing property for the symmetric matrix pencils [30], it follows that $\widehat{R} = \lambda_{\max}(\overline{\Sigma}_1 \overline{\Sigma}^{-1}_2)$ is increasing in N. Finally, from (38), it follows that $\widehat{\tau}$ and $\widehat{\tau}\widehat{R}$ are decreasing and increasing in N, respectively. $\blacksquare$

References

[1] F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, Nov 2013.

[2] H.S. Sánchez, D. Rotondo, T. Escobet, V. Puig, and J. Quevedo. Bibliographical review on cyber attacks from a control oriented perspective. *Annual Reviews in Control*, 48:103 – 128, 2019.

[3] S. Roy, J. Abed Torres, and M. Xue. Sensor and actuator placement for zero-shaping in dynamical networks. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 1745–1750, Dec 2016.

[4] L. Ye, S. Roy, and S. Sundaram. On the complexity and approximability of optimal sensor selection for kalman filtering. In *2018 Annual American Control Conference (ACC)*, pages 5049–5054, June 2018.

[5] R. Dhal, J. Abad Torres, and S. Roy. Detecting link failures in complex network processes using remote monitoring. *Physica A: Statistical Mechanics and its Applications*, 437:36 – 54, 2015.

[6] R. Anguluri, R. Dhal, S. Roy, and F. Pasqualetti. Network invariants for optimal input detection. In *2016 American Control Conference (ACC)*, pages 3776–3781, July 2016.

[7] S. Roy, M. Xue, and S. Sundaram. Graph-theoretic analysis of estimators for stochastically-driven diffusive network processes. In *2018 Annual American Control Conference (ACC)*, pages 1796–1801, June 2018.

[8] Y. Liu, J. J. Slotine, and A. L. Barabási. Observability of complex systems. *Proceedings of the National Academy of Sciences*, 110(7):2460–2465, 2013.

[9] A. Vosughi, C. Johnson, M. Xue, S. Roy, and S. Warnick. Target control and source estimation metrics for dynamical networks. *Automatica*, 100:412–416, 2019.

[10] T. H. Summers, F. L. Cortesi, and J. Lygeros. On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems*, 3(1):91–101, March 2016.

[11] H. Zhang, R. Ayoub, and S. Sundaram. Sensor selection for kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms. *Automatica*, 78:202–210, 2017.

[12] J. Abad Torres, S. Roy, and Y. Wan. Sparse resource allocation for linear network spread dynamics. *IEEE Transactions on Automatic Control*, 62(4):1714–1728, April 2017.

[13] F. Pasqualetti, S. Zampieri, and F. Bullo. Controllability metrics, limitations and algorithms for complex networks. *IEEE Transactions on Control of Network Systems*, 1(1):40–52, March 2014.

[14] S. Zhao and F. Pasqualetti. Discrete-time dynamical networks with diagonal controllability gramian. *IFAC-PapersOnLine"*, 50(1):8297–8302, 2017. 20th IFAC World Congress.

[15] H. J. Van Waarde, M. K. Camlibel, and H. L. Trentelman. A distance-based approach to strong target control of dynamical networks. *IEEE Transactions on Automatic Control*, 62(12):6266–6277, Dec 2017.

[16] I. Gohberg, A. M. Kaashoek, and S. Goldberg. Block toeplitz operators. In *Classes of Linear Operators Vol. II*. Birkhäuser Basel, 1993.

[17] J. Abed Torres and S. Roy. Graph-theoretic analysis of network input-output processes: Zero structure and its implications on remote feedback control. *Automatica*, 61:73–79, 2015.

[18] H. L. Van Trees. *Detection, Estimation, and Modulation Theory. Part I:*. Wiley Press, 2004.

[19] L. L. Scharf. *Statistical Signal Processing*. Reading, MA: Addition-Wesley, 1991.

[20] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. New York: Wiley, 2000.

[21] Y. Bresler, V. U. Reddy, and T. Kailath. Optimum beamforming for coherent signal and interferences. *IEEE Trans. on Acou., Speech, and Signal Process.*, 36(6):833–843, June 1988.

[22] J. M. Wozencraft and I. M. Jacobs. *Principles of Communication Engineering*. New York: Wiley, 1965.

[23] J. L. Melsa and D. L. Cohn. *Decision and Estimation Theory*. McGraw-Hill, 1978.

[24] J. M. Varah. A lower bound for the smallest singular value of a matrix. *Linear Algebra and its Applications*, 11(1):3–5, 1975.

[25] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, and D. K. Hammond. Gspbox: A toolbox for signal processing on graphs. *ArXiv e-prints*, Aug. 2014.

[26] R. M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory*, 2(3):155–239, 2006.

[27] A. Böttcher and B. Silbermann. Block toeplitz matrices. In *Introduction to Large Truncated Toeplitz Matrices*, pages 185–219. Springer New York, 1999.

[28] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics, 1994.

[29] J. W. Lewis. Inversion of tridiagonal matrices. *Numerische Mathematik*, 38(3):333–345, Oct 1982.

[30] R. C. Li. Rayleigh quotient based optimization methods for eigenvalue problems. In *Matrix Functions and Matrix Equations*, pages 76–108. World Scientific, 2015.