Automatica 141 (2022) 110277

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Network theoretic analysis of maximum a posteriori detectors for optimal input detection☆

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

[a] Department of Mechanical Engineering, University of California, Riverside, CA 92521, USA
[b] Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, 560012, India
[c] School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99163, USA

ARTICLE INFO

Article history:
Received 4 February 2020
Received in revised form 21 January 2022
Accepted 15 February 2022
Available online 13 April 2022

Keywords:
Statistical hypotheses testing
Mean detection
Covariance detection
Network systems
Sensor placement

ABSTRACT

We study maximum-a-posteriori detectors to detect changes in the constant mean vector and the covariance matrix of a Gaussian stationary stochastic input driving a few nodes in a network, using remotely located sensor measurements. We show that the detectors' performance can be analyzed using specific input-to-output gain of the network system's transfer function matrix and the input statistics and sensor noise in the asymptotic measurement regime. Using this result, we study the detector's performance using node cutsets that separate the nodes containing inputs from a partitioned set of nodes not containing inputs. In the absence of noise, we show that the detectors' performance is no better for sensors on a partitioned set than those on the cutset. Instead, in the presence of noise, we show that the detectors' performance can be better for sensors on a partitioned set than those on the cutset for certain choices of edge weights. Our results quantify the extent to which input and sensor nodes' distance modulates detection performance via separating cutsets, and have potential applications in sensor placement problems. Finally, we complement the theory with simulations.

1. Introduction

Detecting abrupt changes in inputs using remote measurements is essential for the safety and security of network systems, including power, transportation, and sensor networks (Giraldo, Sarkar, Cardenas, Maniatakos, & Kantarcioglu, 2017). These changes may result from a faulty component injecting unwanted signals or an adversary intentionally compromising inputs. Designing detectors to track input changes using tools ranging from geometric control theory (Chen & Patton, 1999) to statistical signal processing (Basseville & Nikiforov, 1993) has a rich history in the controls community. However, studying detectors' performance in terms of network structure and noise statistics received limited attention in contrast to the research on estimation and control for network systems (Zhang, Han, & Yu, 2016).

In non-networked RADAR systems, a typical maximum-a-posteriori (MAP) detector's performance deteriorates by increasing the input-to-sensor[1] nodes' distance (Richards, Scheer, & Holm, 2019). However for network systems, in our previous work (Anguluri, Dhal, Roy, & Pasqualetti, 2016; Anguluri & Pasqualetti, 2021), we numerically showed that certain MAP detectors' performance improves with increased input-to-sensor nodes' distance. Interestingly, this behavior manifests only when the measurements are noisy. We provide a theoretical justification for such counter-intuitive behavior of MAP detectors for network systems. Our MAP detectors discriminate between the hypotheses $H_1 : \mathbf{w}[k] \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1)$ and $H_2 : \mathbf{w}[k] \sim \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2)$, where $\mathbf{w}[k]$ is the input exciting a few network nodes, and $\boldsymbol{\mu}_i$ and Σ_i are independent of k for all $i \in 1, 2$. We consider two cases: (a) change in mean vectors ($\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ but $\Sigma_1 = \Sigma_2$) and (b) change in covariance matrices ($\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ but $\Sigma_1 \neq \Sigma_2$). These cases capture various faults/attacks studied in the literature (Giraldo et al., 2017). For instance, we can model integrity attacks by setting $\boldsymbol{\mu}_1 = \mathbf{0}$ (nominal) and $\boldsymbol{\mu}_2 \neq \mathbf{0}$ (attack) in the case (a).

☆ This material is based upon work supported in part by ARO award W911NF-18-1-0213 and in part by UCOP award LFR-18-548175, AFOSR FA9550-19-1-0235. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Claudio Altafini under the direction of Editor Christos G. Cassandras.

* Corresponding author.
E-mail addresses: rangu003@ucr.edu (R. Anguluri), vkatewa@iisc.ac.in (V. Katewa), sroy@eecs.wsu.edu (S. Roy), fabiopas@engr.ucr.edu (F. Pasqualetti).

[1] Throughout, we interchange the words sensor and output.

https://doi.org/10.1016/j.automatica.2022.110277
0005-1098/

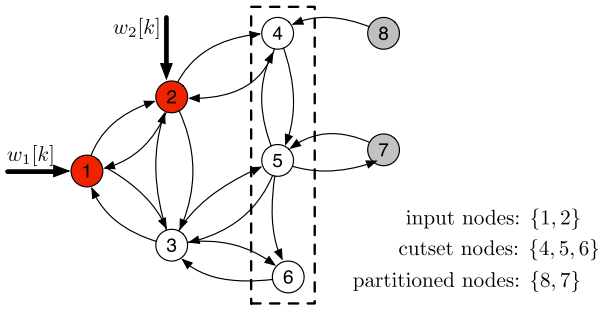

Fig. 1. Network partitions induced by a network cut (or cutset).

For attacks modeled using covariance matrices, see Anguluri and Pasqualetti (2021) and Guo, Shi, Johansson, and Shi (2018).

Related Literature: The problem of discriminating multiple hypothesis for stationary Gaussian signals in noise using MAP detectors has a long history (Kailath & Poor, 1998; Schweppe, 1973; Van Trees & Bell, 2013). We study MAP detectors for network systems driven by stationary Gaussian signals and characterize their qualitative behavior with respect to input-to-sensor distance using network cuts akin to the works that focused on control and estimation problems (Liu, Slotine, & Barabási, 2013; Roy, Xue, & Sundaram, 2018; Vosughi, Johnson, Xue, Roy, & Warnick, 2019). For network dynamics with only process noise, Roy et al. (2018) showed that the estimation performance worsens as the sensor location is moved away from the input. Instead, Vosughi et al. (2019) showed that the control energy is greater for target nodes far away from the input. We show that this behavior also holds in the context of input detection; however, in the presence of sensor noise, for certain class of networks, for e.g., non-normal networks (Baggio, Rutten, Hennequin, & Zampieri, 2020), we show that the detection performance improves as the input-to-sensor distance increases. Finally, our work adds to the literature on deriving exact asymptotic expressions for the error probabilities of detection performance for stationary Gaussian signals (Sung, Tong, & Poor, 2006).

Contributions: The contribution of this work is two-fold: First, using sensor measurements, we develop closed-form expressions for MAP detectors concerning mean change and covariance change problems. We then characterize detectors' performance as a function of network edge weights, input-to-sensor nodes' distance, and input and noise statistics. For large measurement horizon, we provide simpler expressions for detectors' performance that depend on the transfer function matrix and input and noise statistics.

Second, for fixed input nodes set, we study detectors' performance using non-trivial node cutsets[2] that separate the nodes containing inputs from a partitioned set; see Fig. 1 and Definition 2. For noise-less measurements, we show that the detectors' performance is no better for sensors on a partitioned set than those on the cutset for any network. Instead, for the noisy case, we show that the detection performance is better for sensors on a partitioned set than those on the cutset for specific networks—for e.g., non-negative networks with edge weights satisfying certain algebraic conditions (see Section 4). For such networks, we derive simple conditions based on edge weights to check if the performance is better or worse for sensors on a cutset.

[2] A set of vertices $\mathcal{C}$ in the graph $\mathcal{G}$ is referred to as cut or cutset, if all directed paths between input and output nodes in $\mathcal{G}$ pass through a vertex in $\mathcal{C}$; see Fig. 1 for an illustration.

Potential practical applications of our results include (i) enforcing improved constraints on sensor placement optimization algorithms to improve numerical algorithms' efficiency, and (ii) providing design guidelines to facilitate or prevent measurability of specific nodes.

Notation: We denote the finite dimensional vectors by bold faced symbols. The set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ denotes the canonical basis vectors of $\mathbb{R}^n$. For $\mathbf{x} = (x_1, \dots, x_n)^\mathsf{T}$ and $\mathbf{y} = (y_1, \dots, y_n)^\mathsf{T}$, $\mathbf{x} \preceq \mathbf{y}$ iff $x_i \le y_i$, $i = 1, \dots, n$. Let $\mathrm{spec}(M)$ and $\bar{\lambda}(M)$ denote the eigenspectrum and spectral radius of $M \in \mathbb{C}^{n\times n}$. Denote by $M_1 \otimes M_2$ the Kronecker product of two matrices M_1 and M_2. Denote by I or I_n the $n \times n$ identity matrix. Denote by $M \succ 0$ ($M \succeq 0$) the symmetric positive (semi) definite matrix. Denote by $M \ge 0$ the non-negative matrix. Denote by $\mathrm{diag}(M_1 \cdots M_m)$ the block diagonal matrix formed by matrices $M_1 \dots M_m$. Define the infinity norm on the space of matrix-valued functions as $\|F(z)\|_\infty := \operatorname{ess\,sup} \|F(z)\|_2$, where $\{z \in \mathbb{C} : |z| = 1\}$. For $Z_1 \sim \mathcal{N}(0,1)$ (standard normal), the right tail probability is $Q_{\mathcal{N}}(\tau) := \Pr[Z_1 \ge \tau]$. For $Z_2 \sim \chi^2(p)$ (central chi-squared distribution with q degrees of freedom), let $Q_{\chi^2}(p,\tau) := \Pr[Z_2 \ge \tau]$.

2. Preliminaries and problem setup

Consider a network represented by the digraph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} := \{1, \dots, n\}$ 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 $\mathcal{K} := \{k_1, \dots, k_r\} \subseteq \mathcal{V}$ 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. The input matrix $\Pi = [\mathbf{e}_{k_1}, \dots, \mathbf{e}_{k_r}]$ indicates location of the input nodes, and the input $\mathbf{w}[k] \in \mathbb{R}^r$ is governed by one of the following two competing statistical hypotheses:

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

where $\boldsymbol{\mu}_i \in \mathbb{R}^r$ and $\Sigma_i \in \mathbb{R}^{r\times r} (\succ 0)$, $i \in \{1,2\}$, are known; however, the true hypothesis is unknown. For simplicity, we focus on the binary hypothesis testing problem. By obvious modifications, we can extend our results to multiple (finite) hypothesis testing problem, i.e., H_i with $i \ge 2$ (Poor, 1994).

As discussed in the introduction, we focus on two cases: (i) *mean shift* (MS) model: $\boldsymbol{\mu}_1 \ne \boldsymbol{\mu}_2$ but $\Sigma_1 = \Sigma_2 \triangleq \Sigma_c$; and (ii) *covariance shift* (CS) model: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \triangleq \boldsymbol{\mu}_c$ but $\Sigma_1 \ne \Sigma_2$. However, our network theoretic results on MAP detectors (see Section 4) can be extended to the case of inputs with unknown distributions; see Remark 1.

To detect the input hypothesis (H_1 or H_2), our MAP detectors (see below) rely on the sensor measurements

$$\mathbf{Y}_\mathcal{J}^\mathsf{T} = \begin{bmatrix} \mathbf{y}_\mathcal{J}^\mathsf{T}[1] & \mathbf{y}_\mathcal{J}^\mathsf{T}[2] & \cdots \mathbf{y}_\mathcal{J}^\mathsf{T}[N] \end{bmatrix} \tag{3}$$

recorded from the nodes set $\mathcal{J} := \{j_1, \dots, j_m\} \subseteq \mathcal{V}$ such that $\mathrm{dist}(\mathcal{K}, \mathcal{J}) \ge d$. Here, N is the detection time horizon,

$$\mathbf{y}_\mathcal{J}[k] = C\mathbf{x}[k] + \mathbf{v}[k], \tag{4}$$

$C = [\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_m}]^\mathsf{T}$, and $\mathbf{v}[k] \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 I)$ is the sensor noise. Further, $\mathbf{x}[0]$, $\mathbf{w}[k]$, $\mathbf{v}[k]$ are uncorrelated for all k.

For the MS and CS models, we consider a MAP and a linear discriminant (LD) based MAP detector, resp:

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

$$\text{LD-MAP}: \quad \Pr(H_2|\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{\mathcal{J}}) \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} \Pr(H_1|\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{\mathcal{J}}), \tag{6}$$

where (5) implies that the detected hypothesis $\widehat{H}=H_i$ if $\Pr(\widehat{H}=H_i|\mathbf{Y}_{\mathcal{J}}) > \Pr(\widehat{H}=H_{\neq i}|\mathbf{Y}_{\mathcal{J}})$, for $i\in\{1,2\}$. Similar interpretation holds for (6). We discuss the role of vector $\mathbf{b}$ in (6) in Section 3.

Error probabilities of (5) and (6) are computed using

$$\mathbb{P}_e(\mathcal{J}) = \sum_{i\in\{1,2\}} \Pr(\widehat{H}\neq H_i|H_i)\pi_i, \tag{7}$$

where $\pi_i = \Pr(H_i)$ is the prior probability. The smaller the $\mathbb{P}_e(\mathcal{J})$, the better is the detection performance.

In Sections 3 and 4, we rigorously characterize MAP and LD-MAP detectors' performance from two view points:

(1) Algebraic aspects of detectors: for the input nodes set $\mathcal{K}$, we characterize the detection performance (7)—in terms of the adjacency matrix G, the input and noise statistics, and the sensor locations $\mathcal{J}$.

(2) Network analysis of detectors: for the input nodes set $\mathcal{K}$, we study the qualitative behavior of the detectors with respect to input-to-sensor nodes distance.

3. Algebraic analysis of MAP detectors

We address our first objective; that is, to develop algebraic expressions for MAP and LD-MAP detectors' decision rules and their corresponding error probabilities. We prefer LD-MAP detector for the CS model because its error probability is easy to characterize. Consequently, the LD-MAP detector's performance may be sub-optimal; see Remark 2. The results in this section are based on the standard results of hypothesis testing using Gauss–Markov models. The asymptotic characterization of MAP and LD-MAP detectors' error probability (Lemma 3.4) is novel.

Definition 1 (*Optimal LD-MAP Detector*)**.** An LD-MAP detector (6) is optimal if $\mathbf{b}$ is replaced by $\mathbf{b}^*$, where

$$\mathbf{b}^* = \underset{\mathbf{b}\in\mathbb{R}^{mN}}{\arg\max} \underbrace{\pi_1\mathbb{E}\left[\ln\frac{f_{H_1}(y)}{f_{H_2}(y)}\middle| H_1\right] + \pi_2\mathbb{E}\left[\ln\frac{f_{H_2}(y)}{f_{H_1}(y)}\middle| H_2\right]}_{I(H_1;H_2)},$$

where, for $i\in\{1,2\}$, π_i denotes the prior and $f_{H_i}(y)$ is the density of $y=\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{\mathcal{J}}$ given the hypothesis H_i.

Given a vector $\mathbf{b}$, the divergence measure $I(H_1;H_2)$ indicates how well an LD-MAP detector performs in discriminating between H_1 and H_2. By maximizing $I(H_1;H_2)$ over $\mathbf{b}$, we are indeed finding a best linear detector among the class of LD-MAP detectors parameterized by $\mathbf{b}$. With a slight abuse of notation, denote $\mathbf{b}^*$ by $\mathbf{b}$, and refer to the optimal LD-MAP detector as the LD-MAP detector.

Proposition 3.1. *For $\mathbf{Y}_{\mathcal{J}}$ in (3) and H_i in (2), we have*

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

where the observability and impulse response matrices are

$$\mathcal{O} = \begin{bmatrix} CG\\ CG^2\\ \vdots\\ CG^N\end{bmatrix}; \mathcal{F} = \begin{bmatrix} C\Pi & 0 & \dots & 0\\ CG\Pi & C\Pi & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ CG^{N-1}\Pi & CG^{N-2}\Pi & \dots & C\Pi\end{bmatrix}. \tag{9}$$

Proof. See Appendix. ■

Lemma 3.2 (*Decision Rules of MAP and LD-MAP Detectors*)**.** *Let $\pi_1,\pi_2\neq 0$ and define $\gamma = \ln(\pi_1/\pi_2)$. Let $(\boldsymbol{\mu}_i,\Sigma_i)$ and $(\overline{\boldsymbol{\mu}}_i,\overline{\Sigma}_i)$ be as in (2) and (8). Then,*

(i) *The MAP detector of the MS model is given by:*

$$\left(2\,\overline{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}}\overline{\Sigma}_c^{-1}\right)\mathbf{Y}_{\mathcal{J}} \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} 2\gamma + \overline{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}}\overline{\Sigma}_c^{-1}\left(\overline{\boldsymbol{\mu}}_1+\overline{\boldsymbol{\mu}}_2\right), \tag{10}$$

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 of the CS Model is given by:*

$$\ln\left(\frac{d_1}{d_2}\right) - 2\gamma \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} (y-\mathbf{b}^{\mathsf{T}}\overline{\boldsymbol{\mu}}_c)^2\left[\frac{1}{d_2}-\frac{1}{d_1}\right], \tag{11}$$

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

Proof. See Appendix. ■

Using (7) and Lemma 3.2, we obtain tractable error probability expressions for MAP and LD-MAP detectors.

Assumption 3.3. System (1) is asymptotically stable, i.e., $\rho(G)<1$, $\mathbf{x}[0]=0$, and $\pi_1=\pi_2=0.5$. Further,

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

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

Assumptions $\mathbf{x}[0]=0$ and $\pi_i=0.5$ are for the ease of presentation. Instead, assumption $\sqrt{N}\|\boldsymbol{\mu}_2-\boldsymbol{\mu}_1\|_2\to c$ holds when $\boldsymbol{\mu}_i = O(c/\sqrt{N})$, and it ensures that $\mathbb{P}_{e_m}(\mathcal{J}) < \pi_i$. This assumption is standard in the theory of local asymptotically normality for obtaining limiting expressions of the decision rules (van der Vaart, 1998, Chapter 7.5). Assumption $G^k=0$ helps suppress the error term in computing η (14). If $G^k\neq 0$, η should be replaced with $\widehat{\eta}$ (27). Error probabilities for networks satisfying $G^k\neq 0$ are visualized in Section 5.

Lemma 3.4 (*Probability of Error*)**.** *Let $T(z)=C(zI-G)^{-1}\Pi$, where $z\notin spec(G)$. As $N\to\infty$, the error probabilities of MAP (10) and LD-MAP detectors (11) are*

$$\mathbb{P}_{e_m}(\mathcal{J}) = Q_{\mathcal{N}}(0.5\,\eta) \tag{12}$$

$$\mathbb{P}_{e_v}(\mathcal{J}) = 0.5\left[1-Q_{\chi^2}(1,\tau)\right] + 0.5\,Q_{\chi^2}(1,\tau R), \tag{13}$$

where $\tau = \ln R/(R-1)$,

$$\eta^2 = N\widetilde{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}}\left([L^{\mathsf{T}}L+\sigma_v^2 I]^{-1}L^{\mathsf{T}}L\right)\widetilde{\boldsymbol{\mu}}_{\Delta}. \tag{14}$$

$$R = 1+\sigma_v^{-2}\|T(z)\Sigma_1^{\frac{1}{2}}\|_{\infty}^2, \tag{15}$$

Furthermore, $L=T(1)\Sigma_c^{\frac{1}{2}}$ 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.

Proof. See Appendix. ■

Corollary 3.5 (*Spatially Identical Input Statistics*)**.** *Let H_i in (2) be $\mathbf{w}[k]\sim\mathcal{N}(\mu_i\mathbf{1},\sigma_i^2 I)$, where μ_i and σ_i^2 are scalars. For the MS model, with $\sigma_c \triangleq \sigma_1=\sigma_2$, we have*

$$\eta_s^2 = \left(N\mu_{\Delta}^2\right)\mathbf{1}^{\mathsf{T}}[\sigma_c^2 L^{\mathsf{T}}L+\sigma_v^2 I]^{-1}L^{\mathsf{T}}L\mathbf{1}, \tag{16}$$

where $\mu_{\Delta}=\mu_2-\mu_1(\neq 0)$, and $L=T(1)$. For the CS model with $\sigma_1^2>\sigma_2^2$ and $\mu_1=\mu_2$, we have

$$R_s = \frac{\sigma_1^2\|T(z)\|_{\infty}+\sigma_v^2}{\sigma_2^2\|T(z)\|_{\infty}+\sigma_v^2}. \tag{17}$$

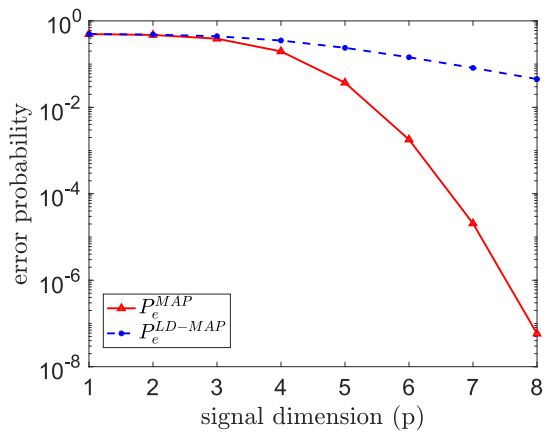

Fig. 2. Suboptimality of the LD-MAP detector.

Proof. Similar to the proof of Lemma 3.4 ■

The error probabilities for the setting in Corollary 3.5 can be obtained by replacing η_s with η in (12) and R_s with R in (13). In the above corollary, we do need $\Sigma_2 = 0$.

The error probability expressions given in Lemma 3.4 and Corollary 3.5 are valid even when N is finite. However, in this case, η and R are complicated functions of G, albeit can be computed numerically (see Section 5). Instead, as we let $N \to \infty$, η depends on G via a quadratic form of the transfer function matrix $T(z)$, and R depend on G via the $\|T(z)\|_\infty$-norm. Thus, we can interpret η and R as a measure of system gain due to the input, and the gain is inversely related to the error probability as shown below.

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

Proof. See Appendix. ■

Proposition 3.6 helps study the MAP and LD-MAP detectors' performance as a function of sensor nodes. This fact will be exploited greatly in the next section.

Remark 1 (*LD-MAP Detector for Other Covariance Matrix Forms*)**.** By adapting the proof of Lemma 3.4, we can easily derive error probability expression for the case $\Sigma_1 = 0$ and $\Sigma_2 \succ 0$. For $\Sigma_1 \neq 0$ and $\Sigma_2 \neq 0$, see Scharf (1991). If $\overline{\Sigma}_1$ and $\overline{\Sigma}_2$ in (8) are simultaneously diagonalizable, an expression of R similar to (17) may be obtained. □

Remark 2 (*Suboptimal LD-MAP Detectors*)**.** In general, LD-MAP detectors are sub-optimal. To see this consider the following example: Let $H_1 : \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 I_N)$ and $H_2 : \mathbf{y} \sim \mathcal{N}(\mathbf{0}, K(p) + \sigma_v^2 I_N)$, where $K(p) = \sigma^2 \sum_{j=1}^{p} \mathbf{e}_j \mathbf{e}_j^{\mathsf{T}}$ and $1 \leq p \leq N$. Then, from (5) and (7)

$$\mathbb{P}_e^{MAP} = 0.5\left[1 - Q_{\chi^2}(p, p\tau_1)\right] + 0.5\, Q_{\chi^2}(p, p\tau_2)\,, \tag{18}$$

where $\tau_1 = (\sigma_v^2/\sigma^2)\ln(1 + \sigma^2/\sigma_v^2)$ and $\tau_2 = (1 + \sigma_v^2/\sigma^2)\ln(1 + \sigma^2/\sigma_v^2)$. To obtain $\mathbb{P}_e^{\text{LD-MAP}}$, we set $p = 1$ in (18). Fix σ^2 and σ_v^2 and note that $\mathbb{P}_e^{MAP}$ is decreasing in p. Instead, p do not effect $\mathbb{P}_e^{\text{LD-MAP}}$ unless σ^2 or σ_v^2 depends on p. Fig. 2 illustrates this behavior for $\sigma = 2^p$ and $\sigma_v = 1$. □

4. Network analysis of MAP detectors

This section addresses our second objective: how does the input-to-output nodes' distance modulate the detection performance? To this aim, we introduce the notion of node cutsets (see below) that help provide sufficient conditions to determine if the detection performance is better (or worse) for the sensor nodes closer to the input nodes.

Definition 2 (*Node Cutset*)**.** For graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ with input nodes $\mathcal{K}$, the nodes $\mathcal{C}_d \subseteq \mathcal{V}$, with $d \geq 1$, 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 $\text{dist}(\mathcal{K}, \mathcal{C}_d) \geq d$, and
(ii) every path from $\mathcal{S}$ to $\mathcal{P}$ contains a node in $\mathcal{C}_d$.

We visualize Definition 2 for the network in Fig. 1. For the input nodes $\{1, 2\}$, a cutset is $\mathcal{C}_1 = \{4, 5, 6\}$. However, $\{5, 6, 7\}$ is not a cutset because it fails to satisfy (ii).

We state the first network-theoretic result for our detectors assuming noise-less measurements ($\sigma_v^2 = 0$).

Theorem 4.1 (*Detection Performance of Noise Less Sensors on the Cutset Vs the Partitioned Set*)**.** *Suppose that $\mathcal{F}$ (9) is full row rank. Let $\sigma_v^2 = 0$ and consider the general detection problem in (2). Let $\mathbb{P}_e(\mathcal{C}_d)$ and $\mathbb{P}_e(\mathcal{P})$ be evaluated using (7). Then, $\mathbb{P}_e(\mathcal{C}_d) \leq \mathbb{P}_e(\mathcal{P})$.*

Proof. See the extended version (Anguluri, Katewa, Roy, & Pasqualetti, 2020). ■

The full row-rank assumption ensures that $\overline{\Sigma_i}^{-1}$ (8) exists when $\sigma_v^2 = 0$; hence, under H_1 and H_2, the probability densities of measurements are well-defined. The preceding theorem is generic in the sense that we do need the measurements to be Gaussian. Thus, the assertion in Theorem 4.1 holds for MS and CS models. Finally, Theorem 4.1 is valid even when $\mathcal{P}$ is replaced with $\mathcal{P} \cup \widetilde{\mathcal{C}}_d$, where $\widetilde{\mathcal{C}}_d \subseteq \mathcal{C}_d$.

Theorem 4.1 says that sensor nodes near the input nodes achieve better detection performance than those far away from the inputs, *irrespective of adjacency matrix G entries and the measurement horizon N*. The closeness is understood in the sense of *node cutsets*, since $d \leq \text{dist}(\mathcal{K}, \mathcal{C}_d) < \text{dist}(\mathcal{K}, \mathcal{P})$. Thus if cutsets of size at most r (the number of inputs) exist in a network and $\sigma_v^2 = 0$, we should place sensors on the cutsets.

We consider the case of noisy measurements ($\sigma_v^2 > 0$). Unlike the noise less case, results for the noisy case will be specific to the MAP and LD-MAP detectors. Extensions to the more general cases are mentioned in Remark 3. For the cutset $\mathcal{C}_d$, let $\mathbf{x}_c[k]$, $\mathbf{x}_s[k]$, and $\mathbf{x}_p[k]$ denote the states of $\mathcal{C}_d$, $\mathcal{S}$, and $\mathcal{P}$, resp. Then, (1) can also be expressed as

$$\begin{bmatrix} \mathbf{x}_s[k+1] \\ \mathbf{x}_c[k+1] \\ \mathbf{x}_p[k+1] \end{bmatrix} = \begin{bmatrix} G_{ss} & G_{sc} & 0 \\ G_{cs} & G_{cc} & G_{cp} \\ 0 & G_{pc} & G_{pp} \end{bmatrix} \begin{bmatrix} \mathbf{x}_s[k] \\ \mathbf{x}_c[k] \\ \mathbf{x}_p[k] \end{bmatrix} + \begin{bmatrix} \mathbf{w}_s[k] \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{19}$$

From (19), see that the states of $\mathcal{C}_d$ drive the states of $\mathcal{P}$:

$$\mathbf{x}_p[k+1] = G_{pp}\mathbf{x}_p[k] + G_{pc}\mathbf{x}_c[k]. \tag{20}$$

Theorem 4.2 (*Performance of Noisy Sensors on the Node Cutset Vs the Partitioned Set*)**.** *Let $\sigma_v^2 = 0$. Let G_{pp} and G_{pc} in (20) satisfy $\text{spec}(G_{pp}) \cap \{z \in \mathbb{C} : |z| = 1\} = \phi$. Let $\overline{\rho}(z)$ and $\underline{\rho}(z)$ be the maximum and minimum singular values of $T_s(z) = (zI - G_{pp})^{-1}G_{pc}$. Let $\mathbb{P}_{e_m}(\mathcal{C}_d)$ (12) and $\mathbb{P}_{e_v}(\mathcal{C}_d)$ (13) be the error probabilities associated 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*

(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* $\mathbb{P}_{e_v}(\mathcal{C}_d) > \mathbb{P}_{e_v}(\mathcal{P})$.

Proof. See Appendix. ■

Notice that the sup and inf of $\overline{\rho}(z)$ and $\underline{\rho}(z)$, respectively, are the maximum and minimum input–output gains of $T_s(z)$. Since $\mathbf{x}_c[k]$ and $\mathbf{x}_p[k]$ are linearly related (see (20)), Theorem 4.2 says that the smaller the gain of $T_s(z)$, the better the detection performance of sensors on $\mathcal{C}_d$, and vice versa. Thus, depending upon the entries of submatrix $[G_{pp}\, G_{pc}]$, measuring the cutset $\mathcal{C}_d$ may not be optimal for input detection. However, this is not the case for noiseless measurements because the input–output gain gets cancel in the error probability computations. To see this set $\sigma_v^2 = 0$ in (16) and (17) and note that η and R, which govern the error probabilities, do not depend on G.

For networks that satisfy $G \geq 0$, the sufficient conditions for placing sensors on or away from the cutset $\mathcal{C}_d$ can be stated directly in terms of entries in G.

Lemma 4.3 (*Error Probability: Non-Negative Adjacency Matrix*)**.** *Let* $G \geq 0$, *and define* $\widetilde{G} = [G_{pp}\, G_{pc}] \in \mathbb{R}^{m_1 \times n_1}$, *where* G_{pp} *and* G_{pc} *are defined in* (20).

(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* $\widetilde{G}\mathbf{1} \geq \mathbf{1}$, *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* $\widetilde{\mathcal{P}} \subseteq \mathcal{P}$.

Proof. See the extended version (Anguluri et al., 2020). ■

Notice that m_1 and n_1 denote the cardinalities of $\mathcal{C}_d$ and $\mathcal{P}$. We need to impose more restrictions on G to tackle the case $n_1 > 1$ in part (ii). For instance, if G is diagonally dominant (Varah, 1975), we can show that (ii) holds for $n_1 > 1$.

Example 1 (*Networks Satisfying Hypothesis of Lemma 4.3*)**.** Consider a Toeplitz network with $n > 3$ nodes and the adjacency matrix

$$G_{n\times n} = \begin{bmatrix} a & b & 0 & \cdots & 0 & 0 \\ c & a & b & \cdots & 0 & 0 \\ 0 & c & a & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & a & b \\ 0 & 0 & 0 & \cdots & c & a \end{bmatrix},$$

(G_{pc}: second column; G_{pp}: remaining columns)

where $(a, b, c) \geq 0$. Let $\mathcal{C}_d = \{3\}$ and note that $n_1 = 1$ and $m_1 = n - 2$. The hypothesis of Lemma 4.3 (i) and (ii) holds if $a + b + c \leq 1/\sqrt{n-2}$, and $a, b = 0$ and $c > 1$, respectively. Toeplitz networks with weights $a = b = 0$, and $c > 1$ belong to the class of non-normal networks (Baggio et al., 2020).

We remark that our network theoretic results become relevant when solving certain sensor placement problems, especially when node cutsets can be easily identified from the connectivity properties.

Remark 3 (*Extension of Network Theoretic Results to Arbitrary Input Distributions*)**.** We outline the steps required to extend Theorem 4.2 to the general Gaussian detection problem and non-Gaussian input distribution. For the former (i.e., both the mean and covariance change), obtaining tractable error probability is hard. However, we can study qualitative behavior of the MAP detector via cutsets using Chernoff type upper bound (Van Trees & Bell, 2013). A careful analysis shows that this bound is a quadratic function of the filtered mean and covariance matrix (8), and as $N \to \infty$, it depends on the spectrum of $T(z) = C(zI - G)^{-1}\Pi$.

For the non-Gaussian input case, obtaining closed form expressions for Chernoff bounds might not be easy. Instead, we can consider proxy detectors, such as the deflection-based detector (Anguluri & Pasqualetti, 2021), that optimize a signal-to-noise criterion instead of the error probability. In our recent work (Anguluri & Pasqualetti, 2021), we showed that the deflection-based detector's performance depends on the singular values of $T(z)$. In either cases, for Theorem 4.2 to hold, we need to only show that the detection performance (or the upper bound) is increasing or decreasing in $\|T(z)\|_M$, for $M \succ 0$.

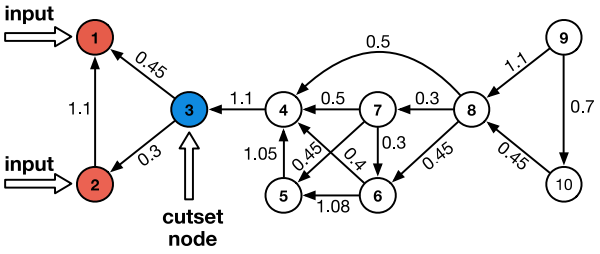

Fig. 3. The graph of a directed network of size $n = 10$. The nodes that are to the right of the cutset node {3} form the partitioned set.

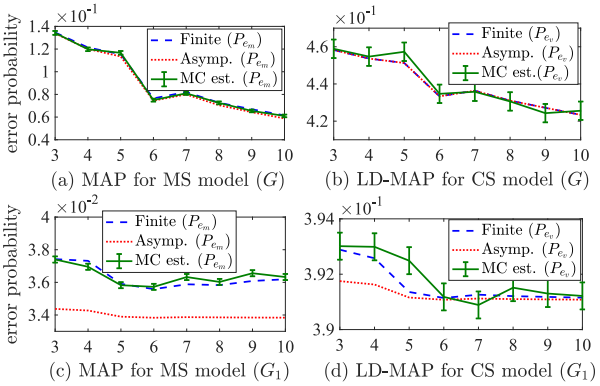

Fig. 4. Finite, asymptotic, and estimated error probabilities of the MAP and LD-MAP detectors for network in Fig. 3. Error probability of each node in the partitioned set $\mathcal{P} = \{4, \ldots, 10\}$ is less than that of the cutset $\mathcal{C}_d = \{3\}$. This result is consistent with Lemma 4.3.

5. Simulation results

We validate the theoretical results of our detectors on small- and large-scale networks, for noisy measurements.

(Detectors' performance on the partitioned nodes is better than that of the cutset): Consider the network in Fig. 3 comprising 10 nodes with the input nodes set $\mathcal{K} = \{1, 2\}$; cutset $\mathcal{C}_d = \{3\}$; and partitioned nodes set $\mathcal{P} = \{4, \ldots, 10\}$. For MS model, under the hypothesis H_i, the input $\mathbf{w}_i[k] \sim \mathcal{N}(\boldsymbol{\mu}_i, \sigma_c^2 I)$, where $\boldsymbol{\mu}_1 = [2, 2]^\mathsf{T}$, $\boldsymbol{\mu}_2 = [1, 1]^\mathsf{T}$, and $\sigma_c^2 = 1.5$. For the CS model, the input $\mathbf{w}_i[k] \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 I_{2\times 2})$, where $\sigma_1^2 = 2.0$ and $\sigma_2^2 = 1.0$. We set $N = 200$ (measurement horizon for the finite case) and $\sigma_v^2 = 1.2$ (noise variance).

Fig. 4 illustrates the error probabilities of the detectors for G (entries shown in Fig. 3) and G_1, obtained by adding edges $\{3 \to 9, 6 \to 10, 5 \to 6\}$ with weights $(0.25, 0.3, 0.3)$ to the network in Fig. 3. Here, $G^N \approx 0$ but $G_1^N \neq 0$. We use (12) and

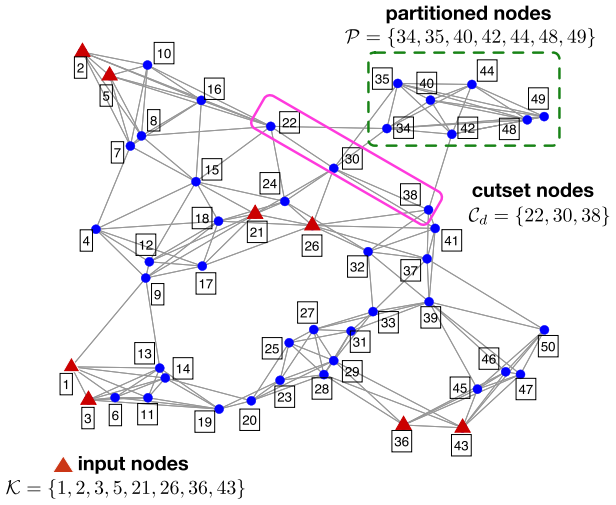

Fig. 5. A random network on 50 nodes (Perraudin et al., 2014).

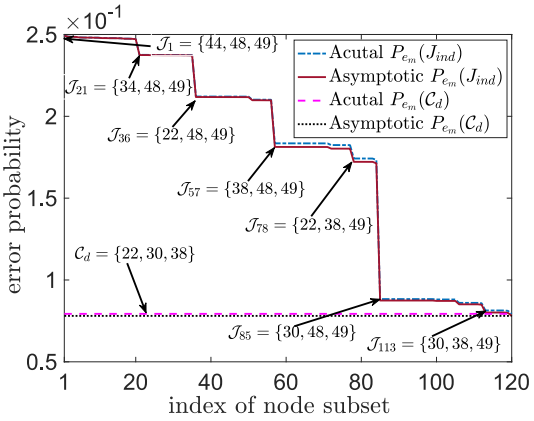

Fig. 6. Finite and asymptotic error probabilities of the MAP detector associated with the MS model for the network in Fig. 5. The error probability of cutset $\mathcal{C}_d$ is less than that of any subset of partitioned set. This behavior is predicted by Lemma 4.3(i).

(13) to compute the error probabilities of MS and CS models. In the asymptotic case, the error probability parameters η and R are given by Corollary 3.5. Instead, in the finite case, the parameters are given by (25) and (30).

We compute the Monte Carlo estimate (MC est.) of the finite error probability numerically for each sensor node, over 10000 and 50000 instances for upper and lower panels in Fig. 4. We require more instances in the lower panel because $G_1^N \neq 0$, which results in significant variance profile (due to long range dependence) among coordinates of each sample of $\mathbf{Y}_{\mathcal{J}}$. Thus requiring us to collect more number dependent samples.

In the upper panel of Fig. 4, note that the error between the finite and asymptotic error probabilities is negligible. This behavior is predicted by Lemma 3.4 because $G^N \approx 0$. Instead, in the lower panel, the error is large for G_1 because $G_1^N \neq 0$. However, in either cases, the qualitative behavior of the error probability agrees with Lemma 4.3.

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

Consider the network in Fig. 5 with $G \geq 0$ For the MS model, we let the input $\mathbf{w}_i[k] \sim \mathcal{N}(\mu_i \mathbf{1}, 1.5 I_{8\times 8})$, where $\mu_1 = 2$, $\mu_2 = 1$, and $\sigma_v^2 = 1.2$. We set $N = 200$.

Let $\mathcal{J} \subseteq \mathcal{C}_d \sqcup \mathcal{P}$ be such that $|\mathcal{J}| = |\mathcal{C}_d| = 3$. We label 120 subsets that satisfy this requirement as $\mathcal{J}_1, \ldots, \mathcal{J}_{120}$, such that $\mathbb{P}_{e_m}(\mathcal{J}_k) \leq \mathbb{P}_{e_m}(\mathcal{J}_l)$, for $k < l \leq 120$. In Fig. 6, we plot the finite and asymptotic error probabilities as a function $\mathcal{J}_i$'s for the MS model. The error probability of the cutset $\mathcal{C}_d = \{22, 30, 38\}$ is lesser than that of any $\mathcal{J}_i$. This result agrees with Lemma 4.3(i) because $\|\widetilde{G}\|_\infty < 1/\sqrt{7} = 0.3780$. A plot similar to Fig. 6 for the CS model is provided in the extended version (Anguluri et al., 2020).

6. Conclusion

We consider a MAP and an LD-MAP detector, respectively, to detect changes in the mean vectors and the covariance matrices of a stationary Gaussian input driving a network. For large measurement horizon, we characterize the detectors' performance in terms of network transfer function and input and senor noise statistics. We then study the detection performance as a function of input-to-sensor nodes' distance using the notion of node cutsets. Surprisingly, in the absence of sensor noise, we show that detectors' performance is better if the sensors are placed away from the input nodes. Our numerical experiments reinforce the theoretical results both when we do and not have access to the input and noise statistics. For the latter scenario, we perform Monte Carlo simulations. The closed form expressions of our detectors' performance (Lemma 3.4) in terms of cutsets leave door to analyze optimal placement algorithms using the H_2 and H_∞ based optimization methods (Münz, Pfister, & Wolfrum, 2014) and the recent randomized methods (Bopardikar, 2021). We leave these topics for future research.

Acknowledgments

We thank the anonymous reviewers for their constructive comments and suggestions that lead to several improvements in presenting our results in this paper.

Appendix.[3]

Proof of Proposition 3.1. From (1) and (4), the measurement vector in $\mathbf{Y}_{\mathcal{J}}$ (3) can be expanded as

$$\mathbf{Y}_{\mathcal{J}} = \mathcal{O}\mathbf{x}[0] + \mathcal{F}\mathbf{w}_{0:N-1} + \mathbf{v}_{1:N}, \tag{21}$$

where $\mathbf{w}_{0:N-1} = [\mathbf{w}[0]^\mathsf{T}, \ldots, \mathbf{w}[N-1]^\mathsf{T}]^\mathsf{T}$ and $\mathbf{v}_{1:N} = [\mathbf{v}[1]^\mathsf{T}, \ldots, \mathbf{v}[N]^\mathsf{T}]^\mathsf{T}$, and $\mathcal{O}$ and $\mathcal{F}$ are defined in the statement of the proposition. Taking expectation and covariance of $\mathbf{Y}_{\mathcal{J}}$ in (21), under the hypothesis H_i, yields us $\overline{\boldsymbol{\mu}}_i$ and $\overline{\Sigma}_i$ in (8). □

Proof of Lemma 3.2. Let ζ and z be the realizations of $\mathbf{Y}_{\mathcal{J}}$ and y, respectively. Since the input and measurement noises follow a Gaussian distribution, the probability density functions of $\mathbf{Y}_{\mathcal{J}}$ (3) and $y = \mathbf{b}^\mathsf{T}\mathbf{Y}_{\mathcal{J}}$ are

$$f(\zeta|H_i) \propto \frac{1}{\sqrt{|\overline{\Sigma}_i|}} \exp\left[-\frac{1}{2}(\zeta - \overline{\boldsymbol{\mu}}_i)^T \overline{\Sigma}_i^{-1}(\zeta - \overline{\boldsymbol{\mu}}_i)\right] \text{ and}$$
$$g(z|H_i) \propto \frac{1}{\sqrt{\mathbf{b}^\mathsf{T}\overline{\Sigma}_i\mathbf{b}}} \exp\left[-\frac{(z - \mathbf{b}^\mathsf{T}\overline{\boldsymbol{\mu}}_i)^2}{2\,\mathbf{b}^\mathsf{T}\overline{\Sigma}_i\mathbf{b}}\right], \tag{22}$$

respectively, where $|\cdot|$ is the determinant. Define $\Psi(\zeta) = \ln(f(\zeta|H_2)/f(\zeta|H_1))$ and $\widehat{\Psi}(z) = \ln(f(z|H_2)/f(z|H_1))$. Then, from

[3] **Note**: We provide only short proofs for most of our results. For detailed proofs, see our extended paper (Anguluri et al., 2020).

the mixed Bayes formula (Melsa & Cohn, 1978), the MAP decision rules based on ζ and z, respectively, are given by

$$\Psi(\zeta) \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} \gamma \text{ and } \widehat{\Psi}(z) \underset{\widehat{H}=H_1}{\overset{\widehat{H}=H_2}{\gtrless}} \gamma. \tag{23}$$

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 $\psi(\zeta)$:

$$\Psi(\zeta) = -0.5\, \overline{\boldsymbol{\mu}}_\Delta^\mathsf{T} \overline{\Sigma}_2^{-1} \overline{\boldsymbol{\mu}}_\Delta + (y - \overline{\boldsymbol{\mu}}_1)^\mathsf{T} \overline{\Sigma}_2^{-1} \overline{\boldsymbol{\mu}}_\Delta. \tag{24}$$

Substitute (24) in the first decision rule of (23) and simplify the resulting expression to obtain the MAP decision rule (10) for ζ. Finally, replace ζ with $\mathbf{Y}_\mathcal{J}$.

part (2) Notice that $\overline{\boldsymbol{\mu}}_1 = \overline{\boldsymbol{\mu}}_2$ and $\overline{\Sigma}_1 \neq \overline{\Sigma}_2$. A similar procedure, as in part (1), based on $g(z|H_i)$ (23) and the rule in (22), yields the LD-MAP detector's expression (10). □

Proof of Lemma 3.4. We divide the proof into two parts. In part (1) we derive the expressions (12) and (14) Instead, in part (2) we derive the expressions (13) and (15).

part (1) Let $N < \infty$, and $\widehat{\mathbb{P}}_{e_m}(\mathcal{J})$ be the error probability of the MAP detector (10). Then, it follows that

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

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

where $s = 2\, \overline{\boldsymbol{\mu}}_\Delta^\mathsf{T} \overline{\Sigma}_c^{-1} \mathbf{Y}_\mathcal{J}$ is a linear transform of a Gaussian random vector $\mathbf{Y}_\mathcal{J}$. Thus, under H_i, it follows that $s \sim \mathcal{N}(2\overline{\boldsymbol{\mu}}_\Delta^\mathsf{T} \overline{\Sigma}_c^{-1} \overline{\boldsymbol{\mu}}_i, 4\widehat{\eta}^2)$, where $\widehat{\eta}^2 = \overline{\boldsymbol{\mu}}_\Delta^\mathsf{T} \overline{\Sigma}_c^{-1} \overline{\boldsymbol{\mu}}_\Delta$. We can show that $\Pr(\widehat{H} = H_2|H_1) = \Pr(\widehat{H} = H_1|H_2) = Q_\mathcal{N}(0.5\,\widehat{\eta})$. Thus, for finite N, from (7), we have $\widehat{\mathbb{P}}_{e_m}(\mathcal{J}) = 0.5\, Q_\mathcal{N}(\widehat{\eta})$.

Consider the following asymptotic error probability:

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

where (a) follows because $\widehat{\eta}$ is increasing in N (see Proposition A.1). We show that $\lim_{N\to\infty} \widehat{\eta}$ exists and equals η (14). Let $\mathbf{m} = \mathbf{1}_N \otimes \boldsymbol{\mu}_\Delta$ and $\boldsymbol{\mu}_\Delta = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$, and from (8), notice that $\widehat{\eta}^2 = (\mathcal{F}\mathbf{m})^\mathsf{T} \overline{\Sigma}_c^{-1} (\mathcal{F}\mathbf{m})$. Define $K(l) = CG^l\Pi$, where $l \in \mathbb{N}$, and $S(i) = \sum_{l=0}^{i-1} K(l)$. Since $\rho(G) < 1$, we have $\lim_{i\to\infty} S(i) = C(I - G)^{-1}\Pi \triangleq \overline{K}$. Let $S_N^\mathsf{T} = \left[(S(1) - \overline{K})^\mathsf{T} \quad \cdots \quad (S(N) - \overline{K})^\mathsf{T}\right]$. Then, from Proposition 3.1 we note that $\mathcal{F}\mathbf{m} = S_N \boldsymbol{\mu}_\Delta + \left[\mathbf{1}_N \otimes \overline{K}\right] \boldsymbol{\mu}_\Delta$. Thus,

$$\widehat{\eta}^2 = \boldsymbol{\mu}_\Delta^\mathsf{T} \underbrace{\left[\mathbf{1}_N \otimes \overline{K}\right]^\mathsf{T} \overline{\Sigma}_c^{-1} \left[\mathbf{1}_N \otimes \overline{K}\right]}_{F} \boldsymbol{\mu}_\Delta + \underbrace{\boldsymbol{\mu}_\Delta^\mathsf{T} [S_N^\mathsf{T} \overline{\Sigma}_c^{-1} S_N + 2 S_N^\mathsf{T} \Sigma_c^{-1} \left[\mathbf{1}_N \otimes \overline{K}\right]] \boldsymbol{\mu}_\Delta}_{t(S_N)}, \tag{25}$$

Substituting $\overline{\Sigma}_c = \mathcal{F}(I_N \otimes \Sigma_c)\mathcal{F}^\mathsf{T} + \sigma_v^2 I$ in F yields us

$$F = [\overline{K}^\mathsf{T} \overline{K} \Sigma_c + \sigma_v^2 I]^{-1} [N \overline{K}^\mathsf{T} \overline{K} - P], \tag{26}$$

where $P = \widetilde{M} \overline{\Sigma}_c^{-1} [\mathbf{1}_N \otimes \overline{K}]$; $\widetilde{M} = \overline{K}^\mathsf{T} [\widetilde{S}_N^\mathsf{T} (I \otimes \Sigma_c) \mathcal{F}^\mathsf{T} + \overline{K} \Sigma_c S_N^\mathsf{T}]$; and $\widetilde{S}_N$ is given by permuting sub matrices of S_N from bottom to top. Thus, from (25) and (26),

$$\widehat{\eta}^2 = N \widetilde{\boldsymbol{\mu}}_\Delta^\mathsf{T} \left([L^\mathsf{T} L + \sigma_v^2 I]^{-1} L^\mathsf{T} L\right) \widetilde{\boldsymbol{\mu}}_\Delta + \epsilon_N, \tag{27}$$

where $L = \Sigma^{\frac{1}{2}} \overline{K}$ and $\epsilon_N = -\boldsymbol{\mu}_\Delta^\mathsf{T} [\overline{K}^\mathsf{T} \overline{K} \Sigma_c + \sigma_v^2 I]^{-1} P \boldsymbol{\mu}_\Delta + t(S_N)$.

We show that $\lim_{N\to\infty} \epsilon_N = 0$. To this aim, rewrite $\epsilon_N = \boldsymbol{\mu}_\Delta^\mathsf{T} (Q_N^{(1)} + Q_N^{(2)} + Q_N^{(3)}) \boldsymbol{\mu}_\Delta$, where

$$Q_N^{(1)} = S_N^\mathsf{T} \left[\overline{\Sigma}_c^{-1} S_N + 2\Sigma_c^{-1} \left[\mathbf{1}_N \otimes \overline{K}\right]\right],$$

$$Q_N^{(2)} = [\overline{K}^\mathsf{T} \overline{K} \Sigma_c + \sigma_v^2 I]^{-1} \overline{K}^\mathsf{T} \widetilde{S}_N^\mathsf{T} (I \otimes \Sigma_c) \mathcal{F}^\mathsf{T}, \text{ and}$$

$$Q_N^{(3)} = [\overline{K}^\mathsf{T} \overline{K} \Sigma_c + \sigma_v^2 I]^{-1} \overline{K}^\mathsf{T} \overline{K} \Sigma_c S_N^\mathsf{T}.$$

We can bound ϵ_N using the following inequality:

$$2\epsilon_N \leq \|\boldsymbol{\mu}_\Delta\|_2^2 \sum_{i=1}^3 \lambda_{\max}(Q_N^{(i)} + (Q_N^{(i)})^\mathsf{T}), \tag{28}$$

where $\lambda_{\max}(\cdot)$ is the max eigenvalue. Note that $\lambda_{\max}(Q_N^{(i)} + (Q_N^{(i)})^\mathsf{T}) < \infty$ for $N \gg N$ because $G^k = 0$, for some $k \in \mathbb{N}$ (see Assumption 3.3(i)). Thus, as required, $\lim_{N\to\infty} \epsilon(N) = 0$ because $\lim_{N\to\infty} \|\boldsymbol{\mu}_\Delta\|_2 = 0$ (see Assumption 3.3(i)).

part (2) Let $N < \infty$, and $\widehat{\mathbb{P}}_{e_v}(\mathcal{J})$ be the error probability of (11). Then, it follows that

$$\Pr\left(\widehat{H} = H_2|H_1\right) = \Pr\left(\ln(\widehat{R}) > \left[\frac{Z^2}{c_2} - \frac{Z^2}{c_1}\right]|H_1\right) \text{ and}$$

$$\Pr\left(\widehat{H} = H_1|H_2\right) = \Pr\left(\ln(\widehat{R}) < \left[\frac{Z^2}{c_2} - \frac{Z^2}{c_1}\right]|H_2\right),$$

where $c_i = \mathbf{b}^\mathsf{T} \overline{\Sigma}_i \mathbf{b}$, for $i \in \{1, 2\}$, $Z = \mathbf{b}^\mathsf{T}[\mathbf{Y}_\mathcal{J} - \overline{\boldsymbol{\mu}}_c]$ and $\widehat{R} = c_1/c_2 > 1$ (since $\Sigma_2 = 0$; see Assumption 3.3(ii)). Notice that, under H_i, the distributions of Z and $(\sqrt{\mathbf{b}^\mathsf{T} \overline{\Sigma}_i \mathbf{b}})U$ are equal, where $U \sim \mathcal{N}(0, 1)$. Thus, $\Pr(\widehat{H} = H_2|H_1) = \Pr\left(\widehat{\tau} > U^2\right)$ and $\Pr(\widehat{H} = H_1|H_2) = \Pr(U^2 > \widehat{\tau}\widehat{R})$, where $\widehat{\tau} = \ln(\widehat{R})/(\widehat{R} - 1)$. Finally, using the fact that $U^2 \sim \chi^2(1)$, for finite N, from (7) it follows that

$$\widehat{\mathbb{P}}_{e_v}(\mathcal{J}) = 0.5\left[1 - Q_{\chi^2}(1, \widehat{\tau})\right] + 0.5\, Q_{\chi^2}(1, \widehat{\tau}\widehat{R}). \tag{29}$$

We simplify $\widehat{R}$ in (29) by substituting $\mathbf{b}$ that maximizes the I-divergence (see Definition 1). Recall from Scharf (1991, Chapter 4) the following equality:

$$\mathbf{b}^* = \arg\max_{\mathbf{b}\in\mathbb{R}^d} I(H_1; H_2) = \arg\max_{\mathbf{b}\in\mathbb{R}^d} \widehat{R} = \arg\max_{\mathbf{b}\in\mathbb{R}^d} \frac{\mathbf{b}^\mathsf{T} \overline{\Sigma}_1 \mathbf{b}}{\mathbf{b}^\mathsf{T} \overline{\Sigma}_2 \mathbf{b}},$$

where $\doteq mN$. Let $\mathbf{c} = \overline{\Sigma}_2^{1/2} \mathbf{b}$, and note the following:

$$\mathbf{b}^* = \arg\max_{\mathbf{c}\in\mathbb{R}^d} \left(\frac{\mathbf{c}}{\|\mathbf{c}\|_2}\right)^\mathsf{T} \overline{\Sigma}_2^{-1/2} \overline{\Sigma}_1 \overline{\Sigma}_2^{-1/2} \left(\frac{\mathbf{c}}{\|\mathbf{c}\|_2}\right) \text{ and}$$

$$\widehat{R}^* = \lambda_{\max}\left(\Sigma_2^{-1/2} \overline{\Sigma}_1 \overline{\Sigma}_2^{-1/2}\right) = \lambda_{\max}\left(\overline{\Sigma}_1 \overline{\Sigma}_2^{-1}\right). \tag{30}$$

By abuse of notation, we denote $\mathbf{b}^*$ by $\mathbf{b}$ and $\widehat{R}^*$ by $\widehat{R}$.

Consider the asymptotic case. As $N \to \infty$, we have $\tau \to \widehat{\tau}$, $R \to \widehat{R}$, and $\tau R \to \widehat{\tau}\widehat{R}$. The limits are well defined because $\widehat{R}$ is monotone in N (see Proposition A.1). Thus,

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

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

We obtain the expression of R in (15). From Assumption 3.3, we have $\Sigma_2 = 0$ and $\mathbf{x}[0] = 0$. Thus, $\overline{\Sigma}_2 = \sigma_v^2 I$ and $\overline{\Sigma}_1 = FF^\mathsf{T} + \sigma_v^2 I$, where $F = \mathcal{F}(I_N \otimes \Sigma_1^{\frac{1}{2}})$ and $\Sigma_1 = \Sigma_1^{\frac{1}{2}} \Sigma_1^{\frac{1}{2}}$. From these observations and (30), we may conclude that

$$\begin{aligned}R \triangleq \lim_{N\to\infty} \widehat{R} &= \lim_{N\to\infty} \frac{\lambda_{\max}(FF^\mathsf{T} + \sigma_v^2 I)}{\sigma_v^2} \\ &= 1 + \sigma_v^{-2} \lim_{N\to\infty} \lambda_{\max}(FF^\mathsf{T}) \\ &= 1 + \sigma_v^{-2} \|T(z)\Sigma_1^{\frac{1}{2}}\|_\infty.\end{aligned} \tag{31}$$

The last equality follows from the standard results on the limiting expressions of the upper singular values of block Toeplitz matrices; see Böttcher and Silbermann (1999, Chapter 6.4). For detailed steps, please see Appendix section in Anguluri et al. (2020). □

Proof of Theorem 4.2. Consider the deterministic system: $\mathbf{x}[k+1] = G\mathbf{x}[k] + \Pi\mathbf{u}$, where $\mathbf{u}$ is a constant vector. Since $\mathbf{x}[0] = 0$ and $z \notin \text{spec}(G) \cup \text{spec}(G_{pp})$, it follows that

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

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

where G_{pp} and G_{pc} are given in (20). From (32b), we have

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

Let $C_{\mathcal{C}}$ and $C_{\mathcal{P}}$ be the sensor matrices of $\mathcal{C}_d$ and $\mathcal{P}$. Then,

$$\mathbf{x}_c[z] = C_{\mathcal{C}}\mathbf{x}[z] \text{ and } \mathbf{x}_p[z] = C_{\mathcal{P}}\mathbf{x}[z]. \tag{34}$$

part (1) Consider parts (1a) and (1b). First, we compute $\eta_{\mathcal{C}}$ and $\eta_{\mathcal{P}}$ using (14), and $\mathbb{P}_{e_m}(\mathcal{C}_d)$ and $\mathbb{P}_{e_m}(\mathcal{P})$ using (12). Then, from Proposition 3.6, it follows that $\mathbb{P}_{e_m}(\mathcal{C}_d) \le \mathbb{P}_{e_m}(\mathcal{P})$ if $\eta_{\mathcal{C}}^2 \ge \eta_{\mathcal{P}}^2$ and $\mathbb{P}_{e_m}(\mathcal{C}_d) \ge \mathbb{P}_{e_m}(\mathcal{P})$ if $\eta_{\mathcal{C}}^2 \le \eta_{\mathcal{P}}^2$. Thus to prove (1a) and (1b), we need to show that $\overline{\rho}(1) < 1 \implies \eta_{\mathcal{C}}^2 \ge \eta_{\mathcal{P}}^2$ and $\underline{\rho}(1) > 1 \implies \eta_{\mathcal{C}}^2 \le \eta_{\mathcal{P}}^2$. We begin with (1a).

Let $L_{\mathcal{C}} = C_{\mathcal{C}}(I-G)^{-1}\Pi\,\Sigma_c^{\frac{1}{2}}$ and $L_{\mathcal{P}} = C_{\mathcal{P}}(I-G)^{-1}\Pi\,\Sigma_c^{\frac{1}{2}}$. Then, from (14) and the fact that $[L^{\mathsf{T}}L+\sigma_v^2 I]^{-1}L^{\mathsf{T}}L = I-\sigma_v^2[L^{\mathsf{T}}L+\sigma_v^2 I]^{-1}$, we can express $\eta_{\mathcal{C}}^2$ and $\eta_{\mathcal{P}}^2$ as

$$\eta_{\mathcal{C}}^2 = \widetilde{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}}\widetilde{\boldsymbol{\mu}}_{\Delta} - \sigma_v^2\,\widetilde{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}}\left[L_{\mathcal{C}}^{\mathsf{T}}L_{\mathcal{C}} + \sigma^2 I\right]^{-1}\widetilde{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}} \text{ and}$$
$$\eta_{\mathcal{P}}^2 = \widetilde{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}}\widetilde{\boldsymbol{\mu}}_{\Delta} - \sigma_v^2\,\widetilde{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}}\left[L_{\mathcal{P}}^{\mathsf{T}}L_{\mathcal{P}} + \sigma^2 I\right]^{-1}\widetilde{\boldsymbol{\mu}}_{\Delta}^{\mathsf{T}},$$

respectively. We show that $\overline{\rho}(1) < 1 \implies L_{\mathcal{P}}^{\mathsf{T}}L_{\mathcal{P}} \preceq L_{\mathcal{C}}^{\mathsf{T}}L_{\mathcal{C}}$, which in turn implies $\eta_{\mathcal{C}}^2 \ge \eta_{\mathcal{P}}^2$. To this aim, let $z = 1$, and from (34), (32a), and (32b) note the following:

$$\|\mathbf{x}_c[1]\|_2^2 = \|C_{\mathcal{C}}\mathbf{x}[1]\|_2^2 = \mathbf{u}^{\mathsf{T}}\left(\Sigma_c^{-\frac{1}{2}}L_{\mathcal{C}}^{\mathsf{T}}L_{\mathcal{C}}\,\Sigma_c^{-\frac{1}{2}}\right)\mathbf{u},$$
$$\|\mathbf{x}_p[1]\|_2^2 = \|C_{\mathcal{P}}\mathbf{x}[1]\|_2^2 = \mathbf{u}^{\mathsf{T}}\left(\Sigma_c^{-\frac{1}{2}}L_{\mathcal{P}}^{\mathsf{T}}L_{\mathcal{P}}\,\Sigma_c^{-\frac{1}{2}}\right)\mathbf{u}.$$

Notice that the above identities hold for any $\mathbf{u}$. Thus, from (33), we have $\overline{\rho}(1) < 1 \implies L_{\mathcal{P}}^{\mathsf{T}}L_{\mathcal{P}} \preceq L_{\mathcal{C}}^{\mathsf{T}}L_{\mathcal{C}}$ as required. We can use similar reasoning to prove (1b). The details are omitted.

part (2) Consider parts (2a) and (2b). Let $R_{\mathcal{C}}$ and $R_{\mathcal{P}}$ be evaluated using (15) and $\mathbb{P}_{e_v}(\mathcal{C}_d)$ and $\mathbb{P}_{e_v}(\mathcal{P})$ using (13). Then, from Proposition 3.6, we have $\mathbb{P}_{e_m}(\mathcal{C}_d) \le \mathbb{P}_{e_m}(\mathcal{P})$ if $R_{\mathcal{C}} \ge R_{\mathcal{P}}$ and $\mathbb{P}_{e_m}(\mathcal{C}_d) \ge \mathbb{P}_{e_m}(\mathcal{P})$ if $R_{\mathcal{C}} \le R_{\mathcal{P}}$. Thus to prove (2a), it suffices to show that $\sup_{|z|=1}\overline{\rho}(z) < 1$ implies $R_{\mathcal{C}} \ge R_{\mathcal{P}}$. Instead, to prove (2b) it suffices to show that $\inf_{|z|=1}\underline{\rho}(z) < 1$ implies $R_{\mathcal{C}} \le R_{\mathcal{P}}$. We begin with (2a).

Let $T_{\mathcal{C}}(z) = C_{\mathcal{C}}(zI-G)^{-1}$ and $T_{\mathcal{P}}(z) = C_{\mathcal{P}}(zI-G)^{-1}$. Let $\mathbf{u} = \Sigma_1^{1/2}\mathbf{d}$, where $\Sigma = \Sigma_1^{1/2}\Sigma_1^{1/2}$. Then, from (15),

$$\sqrt{\sigma_v^2(R_{\mathcal{C}}-1)} = \|T_{\mathcal{C}}(z)\Sigma_1^{\frac{1}{2}}\|_{\infty}$$
$$= \operatorname*{ess\,sup}_{|z|=1}\|T_{\mathcal{C}}(z)\Sigma_1^{1/2}\mathbf{d}\|_2 = \operatorname*{ess\,sup}_{|z|=1}\|\mathbf{x}_c[z]\|_2.$$

The last equality follows from (34) and (32a). A similar identity holds when we replace $R_{\mathcal{C}}$ with $R_{\mathcal{P}}$ and $\mathbf{x}_c[z]$ with $\mathbf{x}_p[z]$. Using these facts in conjunction with the inequality in (33) and the assumption $\sup_{|z|=1}\overline{\rho}(z) < 1$ yields us $R_{\mathcal{C}} \ge R_{\mathcal{P}}$. We can use similar reasoning to prove (2b). □

Proof of Proposition 3.6. First note that $Q_{\mathcal{N}}(x)$ is decreasing in x, and hence, $\mathbb{P}_{e_m}(\mathcal{J})$ (12) is decreasing in η (given by (14) and (16)). Second for $\mathbb{P}_{e_v}(\mathcal{J})$ (13), observe that $R > 1$ in both (15) and (17). Thus

$$\frac{d\tau}{dR} = \frac{\left(\frac{R-1}{R}\right) - \ln R}{(R-1)^2} < 0, \text{ and}$$
$$\frac{d(\tau R)}{dR} = \frac{(R-1) - \ln R}{(R-1)^2} > 0. \tag{35}$$

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

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

Proof. See the extended version (Anguluri et al., 2020). ■

References

Anguluri, R., Dhal, R., Roy, S., & Pasqualetti, F. (2016). Network invariants for optimal input detection. In *American control conference* (pp. 3776–3781).

Anguluri, R., Katewa, V., Roy, S., & Pasqualetti, F. (2020). Network theoretic analysis of maximum a posteriori detectors for optimal input detection. arXiv:2002.07673.

Anguluri, R., & Pasqualetti, F. (2016). Deflection-based attack detection for network systems. In *2021 American control conference* (pp. 3254–3259).

Baggio, G., Rutten, V., Hennequin, G., & Zampieri, S. (2020). Efficient communication over complex dynamical networks: The role of matrix non-normality. *Science Advances*, *6*(22).

Basseville, M., & Nikiforov, I. V. (1993). *Detection of abrupt changes: Theory and appl*. Englewood Cliffs, NJ: Prentice-Hall.

Bopardikar, S. D. (2021). A randomized approach to sensor placement with observability assurance. *Automatica*, *123*, Article 109340.

Böttcher, A., & Silbermann, B. (1999). Block Toeplitz matrices. In *Introduction to large truncated Toeplitz matrices* (pp. 185–219). Springer New York.

Chen, J., & Patton, R. J. (1999). *Robust model-based fault diagnosis for dynamic systems*. Hull, U.K.: Springer.

Giraldo, J., Sarkar, E., Cardenas, A. A., Maniatakos, M., & Kantarcioglu, M. (2017). Security & privacy in cyber-physical systems: A survey of surveys. *IEEE Design & Test*, *34*(4), 7–17.

Guo, Z., Shi, D., Johansson, K. H., & Shi, L. (2018). Worst-case stealthy innovation-based linear attack on remote state estimation. *Automatica*, *89*, 117–124.

Kailath, T., & Poor, H. V. (1998). Detection of stochastic processes. *IEEE Transaction on Information Theory*, *44*(6), 2230–2231.

Liu, Y., Slotine, J.-J., & Barabási, A.-L. (2013). Observability of complex systems. *Proceedings of the National Academy of Sciences*, *110*(7), 2460–2465.

Melsa, J. L., & Cohn, D. L. (1978). *Decision and estimation theory*. McGraw-Hill.

Münz, U., Pfister, M., & Wolfrum, P. (2014). Sensor and actuator placement for linear systems based on H_2 and H_∞ optimization. *IEEE Transactions on Automatic Control*, *59*(11), 2984–2989.

Perraudin, N., Paratte, J., Shuman, D., Martin, L., Kalofolias, V., Vandergheynst, P., et al. (2014). GSPBOX: A toolbox for signal processing on graphs. arXiv:1408.5781,

Poor, H. V. (1994). *An introduction to signal detection and estimation*. Springer-Verlag New York.

Richards, M. A., Scheer, J. A., & Holm, W. A. (2019). *Principles of modern RADAR: Volume I-basic principles*. Raleigh, NC, USA: Scitech Publishing.

Roy, S., Xue, M., & Sundaram, S. (2018). Graph-theoretic analysis of estimators for stochastically-driven diffusive network processes. In *American control conference* (pp. 1796–1801).

Scharf, L. L. (1991). *Statistical signal processing*. Reading, MA: Addition-Wesley.

Schweppe, F. C. (1973). *Uncertain dynamic systems*. Englewood Cliffs, NJ: Prentice-Hall.

Sung, Y., Tong, L., & Poor, H. V. (2006). Neyman-Pearson detection of Gauss-Markov signals in noise: Closed-form error exponent and properties. *IEEE Transaction on Information Theory*, *52*(4), 1354–1365.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

Van Trees, H. L., & Bell, K. L. (2013). *Detection estimation and modulation theory, Part 1*. John Wiley & Sons.

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

Vosughi, A., Johnson, C., Xue, M., Roy, S., & Warnick, S. (2019). Target control and source estimation metrics for dynamical networks. *Automatica*, *100*, 412–416.
Zhang, X., Han, Q., & Yu, X. (2016). Survey on recent advances in networked control systems. *IEEE Transactions on Industrial Informatics*, *12*(5), 1740–1752.

Rajasekhar Anguluri received the B.Tech. degree in electrical engineering from the National Institute of Technology Warangal, India, in 2013, and both the M.S. degree in statistics and the Ph.D. degree in mechanical engineering from the University of California at Riverside, CA, USA, in 2019. He is currently a postdoctoral research scholar with the School of Electrical, Computer, and Energy Engineering at Arizona State University, Tempe, AZ, USA. His current research interests include high-dimensional statistics, statistical signal processing, stochastic control, and power systems.

Vaibhav Katewa is an Assistant Professor in the Department of Electrical Communication Engineering and a full-time associate faculty member of the Robert Bosch Center for Cyber–Physical Systems at the Indian Institute of Science, Bangalore. He was a Postdoctoral Scholar in the department of Mechanical Engineering at the University of California, Riverside from 2017 to 2019. He received his M.S. and Ph.D. degrees from University of Notre Dame in 2012 and 2016, and his Bachelor's degree from Indian Institute of Technology, Kanpur in 2007, all in Electrical Engineering. His research interests include analysis and design of security and privacy methods for cyber–physical systems and complex networks, decentralized and sparse feedback control, and protocol design for networked control systems.

Sandip Roy received his B.S. in Electrical Engineering from University of Illinois at Urbana-Champaign (1998), and M.S. and Ph.D. in Electrical Engineering from Massachusetts Institute of Technology (2000, 2003). He is currently a Professor in the School of Electrical Engineering and Computer Science at Washington State University, and an affiliate faculty member in School of Global Animal Health at Washington State University. He also holds a joint appointment at the Pacific Northwest National Laboratory. His research is focused on the sparse monitoring and control of dynamical networks, and using these techniques to support wide-area management of disruptions in large-scale infrastructure networks, including electric power and air transportation networks.

Fabio Pasqualetti (SM'07, M'13) is an Associate Professor in the Department of Mechanical Engineering, University of California, Riverside. He completed a Doctor of Philosophy degree in Mechanical Engineering at the University of California, Santa Barbara, in 2012, a Laurea Magistrale degree (M.Sc. equivalent) in Automation Engineering at the University of Pisa, Italy, in 2007, and a Laurea degree (B.Sc. equivalent) in Computer Engineering at the University of Pisa, Italy, in 2004. His main research interests are in the areas of security for cyber–physical systems, distributed systems and networks, computational neuroscience, optimization, and robotic patrolling and persistent surveillance.