Optimal Power Allocation and Optimal Linear Encoding for Parameter Estimation in the Presence of a Smart Eavesdropper

Erfan Mehdipour Abadi, Cagri Goken, Cuneyd Ozturk, and Sinan Gezici, *Senior Member, IEEE*

***Abstract*—In this paper, we consider secure transmission of a deterministic vector parameter from a transmitter to an intended receiver in the presence of a smart eavesdropper. The aim is to determine the optimal power allocation and optimal linear encoding strategies at the transmitter to maximize the estimation performance at the intended receiver under constraints on the estimation performance at the eavesdropper and on the transmit power. First, the A-optimality criterion is adopted by utilizing the Cramér-Rao lower bound (CRLB) as the estimation performance metric, and the optimal power allocation and optimal linear encoding strategies are characterized theoretically. Then, corresponding to the D-optimality criterion, the determinant of the Fisher information matrix is considered as the estimation performance metric. It is shown that the optimal linear encoding and optimal power allocation strategies lead to the same solution for this criterion. In addition, extensions of the theoretical results are provided to cases with statistical knowledge of systems parameters. Numerical examples are provided to investigate the optimal power allocation and optimal linear encoding strategies in different scenarios.**

***Index Terms*—Cramér-Rao lower bound (CRLB), estimation, Fisher information, parameter encoding, power allocation, secrecy.**

I. Introduction

A. Literature Review

Eavesdropping is one of the most common threats for security of communication systems [1]. In many applications such as internet of things (IoT), smart homes and cities, and self-driving cars, it is crucial to secretly send information from a transmitter to an intended receiver in the presence of an eavesdropper. In the literature, a common precaution against eavesdropping is to employ key-based cryptographic approaches [2], [3]. However, key generation and distribution can be costly and challenging in heterogenous and dynamic networks with large numbers of connections [4], [5]. In addition, for low-cost and battery operated devices with stringent power, bandwidth, and/or latency constraints, cryptographic approaches may not be well-suited to provide security [6]. In such scenarios, physical layer secrecy can be considered as an alternative or complementary approach to design secure communication systems. Physical layer secrecy exploits varying characteristics of wireless channels related to the intended receiver and the eavesdropper for ensuring secure communications [7]. In order to quantify the amount of achieved secrecy, various types of metrics, such as information, detection, or estimation theoretic metrics, can be used.

Information theoretic metrics for quantifying secrecy are commonly based on mutual information, secrecy rate, or capacity [8]–[18]. For example, it is shown in [8] that when the channel between the transmitter and the eavesdropper is a degraded version of the channel between the transmitter and the intended receiver, a non-zero secrecy rate can be achieved between the transmitter and the intended receiver while providing zero information to the eavesdropper. In [9] and [10], secure information transmission problems are investigated by considering communications over fading channels, and secrecy capacities are characterized in various scenarios. Also, physical layer secrecy is studied from a number of information theoretic perspectives for Gaussian wiretap, broadcast, and interference channels in [12]–[18]. In addition to information theoretic metrics, the secrecy outage probability is adopted as a secrecy metric in various studies such as [19] and [20]. Moreover, [21]–[23] specify the secrecy level based on the signal-to-noise ratio (SNR) metric in a quality-of-service (QoS) framework, while [24] utilizes the Bayesian and Neyman-Pearson frameworks for investigating secrecy constrained distributed detection problems.

Estimation theoretic metrics, such as Fisher information and mean-squared error (MSE), have been employed in a multitude of studies related to secure transmission of information. In [25], Gaussian interference channels with vector parameters are considered in the presence of eavesdroppers by assuming Gaussian prior distribution for the vector parameters in a Bayesian estimation framework. The aim is to minimize the total minimum mean-squared error (MMSE) at the intended receivers under a constraint on the MMSE at the eavesdroppers by using joint artificial noise and linear encoding schemes. In [26], by adopting a nonrandom (i.e., non-Bayesian) parameter estimation framework, the Fisher information is employed as a metric of privacy in a smart-grid network in which adversary parties try to estimate energy consumption based on data gathered from smart meters. Secrecy in a distributed inference framework is considered in [27] and [28], where the information coming to a fusion center from various sensor nodes are also observed by eavesdroppers. In particular, [27] focuses on the estimation of a single point Gaussian source under a Bayesian estimation framework in the presence of an eavesdropper. Optimal transmit power allocation policies are presented for minimizing the average MSE for the parameter of interest while guaranteeing a target MSE at the eavesdropper. In [28], the asymptotic secrecy and estimation problem

E. Mehdipour Abadi and S. Gezici are with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara, Turkey, Tel: +90 (312) 290-3139 (e-mails: {abadi,gezici}@ee.bilkent.edu.tr).

C. Goken is with the Department of Communications and Information Technologies, Aselsan Inc., Ankara 06800, Turkey (e-mail: cgoken@aselsan.com.tr).

C. Ozturk is with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208, USA (e-mail: cuneyd.ozturk@northwestern.edu).

is investigated in a non-Bayesian framework when the sensor measurements are quantized and the channel between sensors and receivers are assumed to be binary symmetric channels. In this setting, sensor quantization thresholds are designed to achieve perfect secrecy for an asymptotically large number of sensors.

One of the most common estimation theoretic metrics is the Cramér-Rao lower bound (CRLB), which is based on the Fisher information matrix (FIM) and provides a lower bound on MSEs of unbiased estimators [29]. In [30], the CRLB is employed as a performance metric for analyzing the secure inference problem for deterministic parameters in IoT systems under spoofing and man-in-the-middle-attack. The secure estimation of a random parameter in the presence of an eavesdropper is investigated in Bayesian estimation settings in [31] and [32]. Specifically, the optimal deterministic encoding of a random scalar parameter is proposed in [31] based on the minimization of expectation of the conditional Cramér-Rao bound (ECRB) at the intended receiver while keeping the estimation error of the linear MMSE (LMMSE) estimator at the eavesdropper above a certain threshold. For the same setting, [32] develops a robust parameter encoding approach by employing the worst-case CRLB of the parameter as the performance metric at the intended receiver. The work in [33] extends the results in [31] to vector parameter estimation problems. In [34], estimation of a deterministic vector parameter is considered by utilizing the CRLB metric at the intended receiver and the MSE of the maximum likelihood (ML) estimator at the eavesdropper. An algorithm is proposed to perform optimal power allocation for secure estimation of multiple deterministic parameters under a total power constraint. The common assumption in [31]–[34] is that the eavesdropper is unaware of the encoding function at the transmitter. The FIM or MSE based performance metrics are also employed in numerous other studies, such as [35] and [36], in the absence of eavesdropping. For example, the distributed estimation of a vector parameter is studied in [35] for wireless sensor networks (WSNs) in the Bayesian framework. By modeling the prior distribution of the vector parameter as Gaussian, Bayesian FIM based performance metrics, namely, the trace and the log-determinant of the Bayesian FIM, are used to optimize the transmit powers of the sensors in the WSN. In [36], linear estimation of correlated Gaussian parameters is investigated and achievable power-distortion regions are derived by considering individual distortion constraints and an average MSE distortion constraint. In addition, optimal power allocation strategies that achieve the minimum total transmission power are obtained under the proposed distortion criteria.

Stochastic encoding and encryption can also be utilized as a defense mechanism against eavesdropping for estimation theoretic secrecy [37]–[41]. For example, stochastic encryption is performed in [38] based on the 1-bit quantized version of a noisy sensor measurement of a deterministic parameter to realize secure communication. It is shown that biased estimation and large errors can be induced at the eavesdropper via symmetric and asymmetric bit flipping strategies which are unknown to the eavesdropper. In [39], the binary stochastic encryption approach in [38] is extended to non-binary stochastic encryption for facilitating estimation of vector parameters. In [41], a randomized mapping between two one-to-one and continuous functions is employed for encoding a random scalar parameter at the transmitter for estimation theoretic secure transmission. The aim is to minimize the estimation error at the intended receiver under a secrecy constraint at the eavesdropper, which is fully aware of the encoding strategy at the transmitter. By considering linear MMSE estimation for small numbers of observations and the ECRB metric for large numbers of observations, optimal encoder randomization strategies are developed.

B. Contributions

Although optimal parameter encoding problems are investigated for secure transmission of *random* parameters with known prior distributions in [31]–[33], [41], a power allocation approach for optimal transmission of multiple *deterministic* parameters has recently been developed in [34] by considering an eavesdropper that is unaware of the power allocation strategy at the transmitter. In this work, we propose optimal encoding problems for secure transmission of multiple *deterministic* parameters in the presence of a *smart* eavesdropper that is aware of the encoding function at the transmitter. Also, we consider two scenarios in which the encoding is via either *power allocation* or *linear encoding*. In addition, we adopt two Fisher information based optimality criteria for quantifying the estimation performance at the intended receiver and the eavesdropper. In particular, the trace of the inverse FIM, namely, the CRLB, and the determinant of the FIM are considered as two alternative performance metrics, which are also referred to as A-optimality and D-optimality criteria [42]. In both scenarios and for both optimality criteria, the optimal power allocation and linear encoding solutions are characterized theoretically in the presence of constraints on the estimation performance of the eavesdropper and on the transmit power. Also, extensions are provided in the presence of statistical knowledge of system parameters. The main contributions and novelty of this paper can be summarized as follows:

- For secure transmission of a deterministic vector parameter, we explicitly characterize the optimal power allocation strategy that minimizes the CRLB at an intended receiver under constraints on the CRLB at a smart eavesdropper and on the transmit power (Proposition 1). Although a similar problem was analyzed in [34], the eavesdropper was modeled to be unaware of the power allocation strategy at the transmitter in that work, which leads to a different solution as the CRLB metric cannot be utilized to quantify secrecy in that setting. (From a practical perspective, the problem considered in this paper corresponds to a worst-case scenario for the estimation system since the eavesdropper is smart and can learn the power allocation strategy instantly.)
- For the first time in the literature, we perform the optimal linear encoding of a deterministic vector parameter under a transmit power limit by minimizing the CRLB at an intended receiver while constraining the CRLB at a smart eavesdropper (Proposition 2). We also show that optimal linear encoding can provide significant performance improvements over the optimal power allocation

approach in some cases. Even though generic encoding operations were considered in [33], the vector parameter was modeled as a random vector with a known prior distribution and the eavesdropper was modeled as unaware of the encoding (i.e., not smart). Accordingly, ECRB and LMMSE metrics were utilized in [33], leading to different formulations.

- We propose optimal power allocation and linear encoding problems for secure transmission of deterministic vector parameters according to the D-optimality criterion for the first time in the literature, and show that these problems admit the same equal power allocation solution (Proposition 3).
- We show that when system parameters are not known perfectly, all the theoretical results can still be applied if there exists statistical knowledge of system parameters.

In addition, numerical examples are presented to illustrate and compare the proposed optimal solutions in various settings.

C. Organization

The remainder of the paper is organized as follows. In Section II, the system model is described and the problem formulations are introduced. In Section III, optimal power allocation and linear encoding approaches are developed according to the A-optimality criterion. Then, Section IV employs the D-optimality criterion and presents the solution of the optimal power allocation and the linear encoding problem. The theoretical results are extended in Section V to cases with statistical knowledge of system parameters. Finally, various numerical examples are presented in Section VI, and concluding remarks are made in Section VII.

II. System Model and Problem Formulation

Consider a vector of unknown deterministic parameters represented by $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_k]^T \in \mathbb{R}^k$ with $k \geq 2$. Measurements are obtained at an intended receiver and an eavesdropper via the following linear models [34]–[36]:

$$\mathbf{Y}_r = \mathbf{F}_r^T \mathbf{P} \boldsymbol{\theta} + \mathbf{N}_r \tag{1}$$

$$\mathbf{Y}_e = \mathbf{F}_e^T \mathbf{P} \boldsymbol{\theta} + \mathbf{N}_e \tag{2}$$

where $\mathbf{Y}_r \in \mathbb{R}^{n_r}$ and $\mathbf{Y}_e \in \mathbb{R}^{n_e}$ denote the measurements at the intended receiver and the eavesdropper, respectively, $\mathbf{F}_r$ and $\mathbf{F}_e$ are, respectively, $k \times n_r$ and $k \times n_e$ real matrices with full row ranks ($k \leq n_r$ and $k \leq n_e$), which are assumed to be known, $\mathbf{N}_r \in \mathbb{R}^{n_r}$ and $\mathbf{N}_e \in \mathbb{R}^{n_e}$ are the additive Gaussian noise vectors at the intended receiver and the eavesdropper, respectively, which are distributed according to $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_r)$ and $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e)$ with $\boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_e \succ \mathbf{0}$, and $\mathbf{P}$ is a $k \times k$ symmetric positive definite matrix, which is to be optimized.[1] Other than the preceding specifications, there are no additional assumptions on $\mathbf{F}_r$, $\mathbf{F}_e$, $\boldsymbol{\Sigma}_r$, and $\boldsymbol{\Sigma}_e$.

To quantify the estimation performance at the intended receiver and the eavesdropper, we utilize the FIMs for the measurements $\mathbf{Y}_r$ and $\mathbf{Y}_e$ with respect to the parameter vector $\boldsymbol{\theta}$, which are given by [34], [43], [44, Lemma 5]

$$\mathbf{I}(\mathbf{Y}_r; \boldsymbol{\theta}) = \mathbf{P} \mathbf{F}_r \boldsymbol{\Sigma}_r^{-1} \mathbf{F}_r^T \mathbf{P} \tag{3}$$

$$\mathbf{I}(\mathbf{Y}_e; \boldsymbol{\theta}) = \mathbf{P} \mathbf{F}_e \boldsymbol{\Sigma}_e^{-1} \mathbf{F}_e^T \mathbf{P} \tag{4}$$

Based on the FIM, two popular and useful performance metrics are the CRLB and the determinant of the FIM, which are referred to as the A-optimality and D-optimality criteria, respectively [42].

The CRLB provides a lower bound on covariance matrices of unbiased estimators as follows [29]:

$$\mathrm{Cov}\big(\widehat{\boldsymbol{\theta}}(\mathbf{Y}_r)\big) \geq \mathbf{I}^{-1}(\mathbf{Y}_r; \boldsymbol{\theta}) \tag{5}$$

$$\mathrm{Cov}\big(\widehat{\boldsymbol{\theta}}(\mathbf{Y}_e)\big) \geq \mathbf{I}^{-1}(\mathbf{Y}_e; \boldsymbol{\theta}) \tag{6}$$

where $\widehat{\boldsymbol{\theta}}(\mathbf{Y}_r)$ and $\widehat{\boldsymbol{\theta}}(\mathbf{Y}_e)$ denote any unbiased estimators of $\boldsymbol{\theta}$ based on measurements $\mathbf{Y}_r$ and $\mathbf{Y}_e$, respectively. Since $\mathrm{Cov}\big(\widehat{\boldsymbol{\theta}}(\mathbf{Y}_r)\big) = \mathrm{E}\big[(\widehat{\boldsymbol{\theta}}(\mathbf{Y}_r) - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}(\mathbf{Y}_r) - \boldsymbol{\theta})^T\big]$ and $\mathrm{Cov}\big(\widehat{\boldsymbol{\theta}}(\mathbf{Y}_e)\big) = \mathrm{E}\big[(\widehat{\boldsymbol{\theta}}(\mathbf{Y}_e) - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}(\mathbf{Y}_e) - \boldsymbol{\theta})^T\big]$ due to unbiasedness, the lower bounds on the MSEs of the vector parameter can be obtained from (5) and (6) as follows:

$$\mathrm{E}\big[\|\widehat{\boldsymbol{\theta}}(\mathbf{Y}_r) - \boldsymbol{\theta}\|^2\big] \geq \mathrm{tr}\{\mathbf{I}^{-1}(\mathbf{Y}_r; \boldsymbol{\theta})\} \tag{7}$$

$$\mathrm{E}\big[\|\widehat{\boldsymbol{\theta}}(\mathbf{Y}_e) - \boldsymbol{\theta}\|^2\big] \geq \mathrm{tr}\{\mathbf{I}^{-1}(\mathbf{Y}_e; \boldsymbol{\theta})\} \tag{8}$$

According to the CRLB metric (i.e., A-optimality criterion), we aim to design the optimal $\mathbf{P}$ at the transmitter that minimizes the CRLB at the intended receiver subject to constraints on the CRLB at the eavesdropper and on the average power. From (7) and (8), this problem is formulated as follows:

$$\min_{\mathbf{P}} \quad \mathrm{tr}\{\mathbf{I}^{-1}(\mathbf{Y}_r; \boldsymbol{\theta})\} \tag{9a}$$

$$\text{s.t.} \quad \mathrm{tr}\{\mathbf{P}\mathbf{P}^T\} \leq P_\Sigma \tag{9b}$$

$$\mathrm{tr}\{\mathbf{I}^{-1}(\mathbf{Y}_e; \boldsymbol{\theta})\} \geq \eta \tag{9c}$$

where $0 < P_\Sigma < \infty$ is the power constraint, and $0 \leq \eta < \infty$ specifies the secrecy constraint for the eavesdropper.

On the other hand, for the D-optimality criterion, the aim is to maximize the determinant of the FIM for the intended receiver [42], [43], [45, Sec. 7.5.2], [46]. This corresponds to minimizing the volume of the ellipsoid representing the maximum confidence region for the ML estimate of the unknown parameters [42], [45, Sec. 7.5.2], [47, Sec. III-C]. Accordingly, the following constrained optimization problem is proposed for the D-optimality criterion:

$$\max_{\mathbf{P}} \quad \det\left(\mathbf{I}(\mathbf{Y}_r; \boldsymbol{\theta})\right) \tag{10a}$$

$$\text{s.t.} \quad \mathrm{tr}\{\mathbf{P}\mathbf{P}^T\} \leq P_\Sigma \tag{10b}$$

$$\det\left(\mathbf{I}(\mathbf{Y}_e; \boldsymbol{\theta})\right) \leq \tilde{\eta} \tag{10c}$$

where $0 < \tilde{\eta} < \infty$ specifies the secrecy constraint for the eavesdropper.

The main motivations behind the use of the A-optimality and D-optimality criteria can be summarized as follows: (i) As FIM based metrics are employed in these criteria, generic approaches can be obtained with no dependence on specific estimator structures. (ii) The use of these metrics facilitates theoretical analyses, leading to intuitive explanations. (iii) The CRLB used in the A-optimality framework corresponds to the

[1] It is noted that $\mathbf{F}_r$ in (1) (and $\mathbf{F}_e$ in (2)) can represent the combined effects of pre-processing at the transmitter (if any) and the effects of channel. Via matrix $\mathbf{P}$, the parameter vector (data) is transformed into another vector of the same size without changing any other blocks at the transmitter.

MSE of the ML estimator for linear systems models with additive Gaussian noise as in (1) and (2) [29], [48].[2] Also, the determinant of the FIM used in the D-optimality framework is related to the volume of the ellipsoid representing the maximum confidence region for the ML estimator [42]. (Please see [47, Sec. III-C] for motivations behind the use of the D-optimality criterion.)

We investigate the proposed problems in (9) and (10) in two different scenarios as follows:

- **Scenario 1:** $\mathbf{P}$ is assumed to be a diagonal matrix as in [34], which is given by $\mathbf{P} = \mathrm{diag}\{\sqrt{p_1}, \ldots, \sqrt{p_k}\}$, where $p_i \geq 0$ for $i \in \{1, \ldots, k\}$. We can consider the problem in this scenario as the *optimal power allocation* problem for parameter estimation [34].
- **Scenario 2:** $\mathbf{P}$ is assumed to be a symmetric matrix which is positive definite.[3] We regard the problem in this scenario as the *optimal linear encoding* problem for parameter estimation [33].

Since Scenario 1 can be considered as a special case of Scenario 2, the performance achieved in Scenario 2 is always superior or equal to that in Scenario 1. Investigation of these two scenarios is useful to determine whether the more general linear encoding approach has advantages over the power allocation approach according to the A-optimality and D-optimality criteria.

Remark 1: In this work, it is assumed that the intended receiver knows $\mathbf{F}_r$, $\mathbf{\Sigma}_r$ and $\mathbf{P}$, the eavesdropper knows $\mathbf{F}_e$, $\mathbf{\Sigma}_e$ and $\mathbf{P}$, and the transmitter knows $\mathbf{F}_r$, $\mathbf{F}_e$, $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_e$. It is practical to assume that the intended receiver knows $\mathbf{\Sigma}_r$, and can learn $\mathbf{F}_r$ and $\mathbf{P}$ as it is in collaboration with the transmitter. However, it is challenging for the eavesdropper to learn $\mathbf{F}_e$ and $\mathbf{P}$, which may require obtaining some prior knowledge about the transmitter location and the channel model, and/or eavesdropping of messages (signal exchanges) between the transmitter and the intended receiver. By assuming the knowledge of $\mathbf{F}_e$, $\mathbf{\Sigma}_e$, and $\mathbf{P}$ at the eavesdropper, we effectively consider a worst-case scenario; i.e., a smart eavesdropper, since the estimation performance of any eavesdropper is bounded by that of the smart eavesdropper. Related to the knowledge of $\mathbf{F}_r$, $\mathbf{F}_e$, $\mathbf{\Sigma}_r$, and $\mathbf{\Sigma}_e$ at the transmitter, $\mathbf{\Sigma}_r$ and $\mathbf{F}_r$ can be learned via feedback from the intended receiver. However, it can be challenging for the transmitter to learn $\mathbf{\Sigma}_e$ and $\mathbf{F}_e$, which may require prior knowledge related to the estimator employed at the eavesdropper and the location of the eavesdropper (and a suitable channel model). If the transmitter does not have accurate knowledge of $\mathbf{\Sigma}_e$ and $\mathbf{F}_e$, this inaccuracy can be modeled by statistical knowledge as in Section V and the optimal power allocation and optimal linear encoding can be performed in the presence of statistical knowledge.

The assumptions in Remark 1 are similar to those in [34] except that the eavesdropper is modeled to be unaware of $\mathbf{P}$ in [34]. In this paper, we consider a smart eavesdropper that also knows $\mathbf{P}$ and investigate the optimal design of $\mathbf{P}$ when it is known by the eavesdropper, as well.

III. Power Allocation and Linear Encoding based on A-Optimality

In this section, we consider the CRLB metric, i.e., the A-optimality criterion, and focus on the problem in (9). For convenience of notation, system dependent matrices can be defined as $\mathbf{A}_r \triangleq \left(\mathbf{F}_r \mathbf{\Sigma}_r^{-1} \mathbf{F}_r^T\right)^{-1}$ and $\mathbf{A}_e \triangleq \left(\mathbf{F}_e \mathbf{\Sigma}_e^{-1} \mathbf{F}_e^T\right)^{-1}$, which are assumed to be positive definite matrices. Then, the inverses of the FIMs in (3) and (4) can be stated as

$$\mathbf{I}^{-1}(\mathbf{Y}_r; \boldsymbol{\theta}) = \mathbf{P}^{-1}\mathbf{A}_r\mathbf{P}^{-1} \tag{11}$$

$$\mathbf{I}^{-1}(\mathbf{Y}_e; \boldsymbol{\theta}) = \mathbf{P}^{-1}\mathbf{A}_e\mathbf{P}^{-1} \tag{12}$$

In order to eliminate scenarios in which the design of matrix $\mathbf{P}$ becomes trivial, it is assumed that $\mathbf{A}_r \neq \zeta \mathbf{A}_e$ for any $\zeta \in \mathbb{R}$; that is, $\mathbf{A}_r$ is not a scaled version of $\mathbf{A}_e$. Since $\mathbf{A}_r$ and $\mathbf{A}_e$ depend on the channels related to the intended receiver and the eavesdropper, respectively, this assumption holds in most practical cases.

Based on (11) and (12), we can express the problem in (9) as follows:

$$\begin{aligned} \min_{\mathbf{P}} \quad & \mathrm{tr}\{\mathbf{P}^{-1}\mathbf{A}_r\mathbf{P}^{-1}\} && \text{(13a)} \\ \text{s.t.} \quad & \mathrm{tr}\{\mathbf{P}\mathbf{P}^T\} \leq P_\Sigma && \text{(13b)} \\ & \mathrm{tr}\{\mathbf{P}^{-1}\mathbf{A}_e\mathbf{P}^{-1}\} \geq \eta && \text{(13c)} \end{aligned}$$

where $\mathbf{P}$ is a diagonal matrix in Scenario 1 and a positive definite matrix in Scenario 2. We investigate the problem in (13) under Scenario 1 and Scenario 2 in the following sections.

A. Scenario 1: A-Optimal Power Allocation

In this section, $\mathbf{P}$ is assumed to be diagonal as $\mathbf{P} = \mathrm{diag}\{\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_k}\}$ with $p_i \geq 0$ for $i \in \{1, \ldots, k\}$. Then, the problem in (13) reduces to

$$\begin{aligned} \min_{\{p_i\}_{i=1}^k} \quad & \sum_{i=1}^k \frac{\alpha_i}{p_i} && \text{(14a)} \\ \text{s.t.} \quad & \sum_{i=1}^k p_i \leq P_\Sigma && \text{(14b)} \\ & \sum_{i=1}^k \frac{\beta_i}{p_i} \geq \eta && \text{(14c)} \\ & p_i \geq 0\,, \quad i = 1, \ldots, k && \text{(14d)} \end{aligned}$$

where α_i and β_i are defined as the ith diagonal elements of $\mathbf{A}_r$ and $\mathbf{A}_e$, respectively; that is, $\alpha_i \triangleq [\mathbf{A}_r]_{ii}$ and $\beta_i \triangleq [\mathbf{A}_e]_{ii}$. Since $\mathbf{A}_r$ and $\mathbf{A}_e$ are assumed to be positive definite, $\alpha_i > 0$ and $\beta_i > 0$ for all $i \in \{1, \ldots, k\}$ in (14a) and (14c).

As it is assumed that $\mathbf{A}_r$ is not a scaled version of $\mathbf{A}_e$, it is known that $\frac{\alpha_1}{\beta_1}, \ldots, \frac{\alpha_k}{\beta_k}$ are not all the same. Then, the solution of (14) is specified by the following proposition.

Proposition 1: *If* $\sum_{i=1}^k \frac{\beta_i \sum_{j=1}^k \sqrt{\alpha_j}}{P_\Sigma \sqrt{\alpha_i}} \geq \eta$*, then the solution of* (14) *is*

$$p_i^* = \frac{P_\Sigma \sqrt{\alpha_i}}{\sum_{j=1}^k \sqrt{\alpha_j}}\,, \quad i = 1, \ldots, k \tag{15}$$

[2] Namely, the CRLB in (9a) is the MSE of the ML estimator for $\boldsymbol{\theta}$ based on $\mathbf{Y}_r$, and the CRLB in (9c) is the MSE of the ML estimator for $\boldsymbol{\theta}$ based on $\mathbf{Y}_e$.

[3] We consider symmetric matrices in order to have a lower number of design parameters. Similarly, positive definite matrices are assumed for obtaining a closed-form (hence, low-complexity) solution in the A-optimality framework.

Otherwise, the solution of (14) *is given by*

$$p_i^* = \frac{P_\Sigma \sqrt{\alpha_i - \mu^* \beta_i}}{\sum_{j=1}^k \sqrt{\alpha_j - \mu^* \beta_j}}\,, \quad i = 1, \ldots, k \tag{16}$$

where μ^* *is the unique solution of*

$$\sum_{i=1}^k \sqrt{\alpha_i - \mu \beta_i} \sum_{i=1}^k \frac{\beta_i}{\sqrt{\alpha_i - \mu \beta_i}} = \eta P_\Sigma \tag{17}$$

for $\mu \in [0, \min_{i \in \{1,\ldots,k\}} \alpha_i / \beta_i)$.

Proof: Please see Appendix A.

Proposition 1 characterizes the solution of (14) explicitly, and also illustrates that the optimal power allocation approach utilizes all the available power; i.e., $\sum_{i=1}^k p_i^* = P_\Sigma$ under the assumption that $\frac{\alpha_1}{\beta_1}, \ldots, \frac{\alpha_k}{\beta_k}$ are not all the same. It can be noted from [43, Eq. (20)] that the solution in (15) in Proposition 1 corresponds to the situation in which the secrecy constraint in (14c) is not effective.[4] When the secrecy constraint becomes effective, the solution of (14) is given by (16), which requires a one-dimensional search to obtain μ^* from (17). Since the expression on the left-hand-side of (17) is monotone increasing (as shown in the proof of Proposition 1), the bisection algorithm [50] can be implemented to solve (17) rapidly.

B. Scenario 2: A-Optimal Linear Encoding

In this scenario, $\mathbf{P}$ is assumed to be a positive definite matrix, and (13) is stated as

$$\min_{\mathbf{P} \in \mathcal{M}^+} \quad \mathrm{tr}\{\mathbf{P}^{-1} \mathbf{A}_r \mathbf{P}^{-1}\} \tag{18a}$$

$$\text{s.t.} \quad \mathrm{tr}\{\mathbf{P}\mathbf{P}^T\} \leq P_\Sigma \tag{18b}$$

$$\mathrm{tr}\{\mathbf{P}^{-1} \mathbf{A}_e \mathbf{P}^{-1}\} \geq \eta \tag{18c}$$

where $\mathcal{M}^+$ denotes the set of positive definite matrices.

Let $\mathbf{P}$ be expressed as $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, where the columns of $\mathbf{V}$ are orthonormal eigenvectors of $\mathbf{P}$, and $\mathbf{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues of $\mathbf{P}$; i.e., $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_k\}$. Then, (18) can be transformed into the following problem after some manipulation:

$$\min_{\mathbf{V}, \{\lambda_i^2\}_{i=1}^k} \quad \mathrm{tr}\{\mathbf{V}^T \mathbf{A}_r \mathbf{V} \mathbf{\Lambda}^{-2}\} \tag{19a}$$

$$\text{s.t.} \quad \sum_{i=1}^k \lambda_i^2 \leq P_\Sigma \tag{19b}$$

$$\mathrm{tr}\{\mathbf{V}^T \mathbf{A}_e \mathbf{V} \mathbf{\Lambda}^{-2}\} \geq \eta \tag{19c}$$

$$\mathbf{V}^T \mathbf{V} = \mathbf{I} \tag{19d}$$

Furthermore, to express the problem in (19) in an alternative form, we define $\mathbf{D}$ as $\mathbf{D} \triangleq \mathbf{\Lambda}^{-2}$. Then, (19) becomes

$$\min_{\mathbf{V}, \mathbf{D}} \quad \mathrm{tr}\{\mathbf{V}^T \mathbf{A}_r \mathbf{V} \mathbf{D}\} \tag{20a}$$

$$\text{s.t.} \quad \mathrm{tr}\big\{\mathbf{D}^{-1}\big\} \leq P_\Sigma \tag{20b}$$

$$\mathrm{tr}\{\mathbf{V}^T \mathbf{A}_e \mathbf{V} \mathbf{D}\} \geq \eta \tag{20c}$$

$$\mathbf{V}^T \mathbf{V} = \mathbf{I} \tag{20d}$$

First, we present the following lemma, which will be useful for finding the solution of (20).

Lemma 1: *Let* $\mu_{\min}$ *be the minimum value of* $\mu \geq 0$ *such that the minimum eigenvalue of* $\mathbf{A}_r - \mu \mathbf{A}_e$ *is equal to zero. Then,* $h(\mu) \triangleq \mathrm{tr}\left\{(\mathbf{A}_r - \mu \mathbf{A}_e)^{1/2}\right\} \mathrm{tr}\left\{\mathbf{A}_e (\mathbf{A}_r - \mu \mathbf{A}_e)^{-1/2}\right\}$ *is a continuous and monotone increasing function of* μ *for* $\mu \in [0, \mu_{\min})$. *In addition, the derivative of* $h(\mu)$ *with respect to* μ *is equal to zero if and only if* $\mathbf{A}_r$ *is a scaled version of* $\mathbf{A}_e$.

Proof: Please see Appendix B.

Then, we provide the following proposition, which specifies the solution of (20) by utilizing Lemma 1.

Proposition 2: *Let* $\mathbf{V}^*$ *and* $\mathbf{\Phi}^*$ *represent the solution of* $\mathbf{A}_r \mathbf{V} = \mathbf{V} \mathbf{\Phi}$ *such that the columns of* $\mathbf{V}^*$ *are orthonormal eigenvectors of* $\mathbf{A}_r$, *and* $\mathbf{\Phi}^*$ *is a diagonal matrix with the corresponding eigenvalues of* $\mathbf{A}_r$ *in its diagonals. Also, let* $\mathbf{D}^*$ *be given by*

$$\mathbf{D}^* = \left(\frac{P_\Sigma (\mathbf{\Phi}^*)^{1/2}}{\mathrm{tr}\big\{(\mathbf{\Phi}^*)^{1/2}\big\}} \right)^{-1}. \tag{21}$$

If $\mathrm{tr}\{(\mathbf{V}^*)^T \mathbf{A}_e \mathbf{V}^* \mathbf{D}^*\} \geq \eta$, *then* $\mathbf{V}^*$ *and* $\mathbf{D}^*$ *are the solution of* (20). *Otherwise, the solution is given by* $\mathbf{V}^\star$ *and* $\mathbf{D}^\star$, *where*

$$\mathbf{D}^\star = \left(\frac{P_\Sigma (\mathbf{\Psi}_{\mu^\star})^{1/2}}{\mathrm{tr}\big\{(\mathbf{\Psi}_{\mu^\star})^{1/2}\big\}} \right)^{-1}, \tag{22}$$

and the columns of $\mathbf{V}^\star$ *are orthonormal eigenvectors of* $(\mathbf{A}_r - \mu^\star \mathbf{A}_e)$. *Here,* $\mathbf{\Psi}_{\mu^\star}$ *is the diagonal matrix consisting of the eigenvalues of* $(\mathbf{A}_r - \mu^\star \mathbf{A}_e)$ *and* $\mu^\star$ *denotes the unique solution of*

$$\mathrm{tr}\left\{(\mathbf{A}_r - \mu \mathbf{A}_e)^{1/2}\right\} \mathrm{tr}\left\{\mathbf{A}_e (\mathbf{A}_r - \mu \mathbf{A}_e)^{-1/2}\right\} = \eta P_\Sigma \tag{23}$$

for $\mu \in [0, \mu_{\min})$, *where* $\mu_{\min}$ *is as defined in Lemma 1.*

Proof: Please see Appendix C.

Once the optimal $\mathbf{V}$ and $\mathbf{D}$ are determined as described in Proposition 2, the optimal $\mathbf{\Lambda}$ is obtained as the square-root of $\mathbf{D}^{-1}$; that is, $\mathbf{\Lambda} = \mathbf{D}^{-1/2}$, and the optimal linear encoding matrix (i.e., the solution of (18)) can be calculated as $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$. It is noted that the solutions in Proposition 2 satisfy the constraint in (20b) with equality; hence, the full power utilization property is observed (as in Section III-A) under the assumption that $\mathbf{A}_r$ is not a scaled version of $\mathbf{A}_e$.

By comparing the results in Proposition 1 and Proposition 2, it is noted that A-optimal linear encoding provides a more generic approach than A-optimal power allocation since the former considers all the information in matrices $\mathbf{A}_r$ and $\mathbf{A}_e$ whereas the latter depends only on the diagonal elements of $\mathbf{A}_r$ and $\mathbf{A}_e$. As $\mathbf{A}_r$ and $\mathbf{A}_e$ depend on the channel matrices and the noise covariance matrices (namely, $\mathbf{A}_r = \left(\mathbf{F}_r \mathbf{\Sigma}_r^{-1} \mathbf{F}_r^T\right)^{-1}$ and $\mathbf{A}_e = \left(\mathbf{F}_e \mathbf{\Sigma}_e^{-1} \mathbf{F}_e^T\right)^{-1}$), A-optimal linear encoding is expected to outperform A-optimal power allocation unless the noise components are uncorrelated and channel matrices are diagonal (i.e., no interference among different channels).

[4] The solution in (15) has a similar intuition to the water-filling solution [49] that more power is allocated for a parameter when the quality of the channel related to that parameter is higher.

IV. Power Allocation and Linear Encoding based on D-Optimality

According to the D-optimality criterion, the determinant of the FIM is considered as the performance metric, which is related to the volume of the ellipsoid that represents the maximum confidence region for the ML estimate of the unknown parameters [42]. As argued in [47], the D-optimality criterion also characterizes the estimation performance similarly to the CRLB. In the D-optimality framework, the optimal linear encoding problem can be formulated from (3), (4), and (10) as follows:

$$\max_{\mathbf{P}\in\mathcal{M}^+} \quad \det\left(\mathbf{P}\mathbf{A}_r^{-1}\mathbf{P}\right) \tag{24a}$$

$$\text{s.t.} \quad \text{tr}\{\mathbf{P}\mathbf{P}^T\} \leq P_\Sigma \tag{24b}$$

$$\det\left(\mathbf{P}\mathbf{A}_e^{-1}\mathbf{P}\right) \leq \tilde{\eta} \tag{24c}$$

where $\mathcal{M}^+$ denotes the set of positive definite matrices, $\mathbf{A}_r^{-1} = \mathbf{F}_r\mathbf{\Sigma}_r^{-1}\mathbf{F}_r^T$, and $\mathbf{A}_e^{-1} = \mathbf{F}_e\mathbf{\Sigma}_e^{-1}\mathbf{F}_e^T$ (as defined in Section II).

As in Section III-B, let $\mathbf{P}$ be expressed as $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, where the columns of $\mathbf{V}$ are the orthonormal eigenvectors of $\mathbf{P}$, and $\mathbf{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues of $\mathbf{P}$; i.e., $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_k\}$. Since $\mathbf{V}$ is an orthonormal matrix, the determinant of $\mathbf{P}$ can be calculated as follows: $\det(\mathbf{P}) = \det(\mathbf{V}\mathbf{V}^T)\det(\mathbf{\Lambda}) = \det(\mathbf{\Lambda}) = \prod_{i=1}^k \lambda_i$. Then, (24) can be stated as

$$\max_{\{\lambda_i\}_{i=1}^k} \quad \frac{1}{\det(\mathbf{A}_r)}\left(\prod_{i=1}^k \lambda_i\right)^2 \tag{25a}$$

$$\text{s.t.} \quad \sum_{i=1}^k \lambda_i^2 \leq P_\Sigma \tag{25b}$$

$$\frac{1}{\det(\mathbf{A}_e)}\left(\prod_{i=1}^k \lambda_i\right)^2 \leq \tilde{\eta} \tag{25c}$$

$$\lambda_i \geq 0\,, \quad i = 1, \ldots, k \tag{25d}$$

Since the selection of the orthonormal eigenvectors, i.e., matrix $\mathbf{V}$, does not affect the optimization problem in (25), we can select $\mathbf{V} = \mathbf{I}$ without loss of generality. Hence, for the D-optimality criterion, the optimal linear encoding problem in (24) reduces to the optimal power allocation problem in (25); i.e., $\mathbf{P}$ becomes a diagonal matrix as $\mathbf{P} = \mathbf{\Lambda}$, and Scenario 1 and Scenario 2 become identical.

In the following proposition, the solution of (25) is presented.

Proposition 3: *The solution of* (25) *is given by*

$$\lambda_i^* = \min\left\{\sqrt{\frac{P_\Sigma}{k}}, (\tilde{\eta}\det(\mathbf{A}_e))^{\frac{1}{2k}}\right\} \tag{26}$$

for $i = 1, \ldots, k$.

Proof: The Lagrangian function for the problem in (25) can be expressed as follows:

$$\mathcal{L}\left(\{\lambda_i\}_{i=1}^k, \{\varsigma_i\}_{i=1}^k, \nu, \mu\right) = \frac{-1}{\det(\mathbf{A}_r)}\left(\prod_{i=1}^k \lambda_i\right)^2 - \sum_{i=1}^k \varsigma_i\lambda_i + \nu\left(\sum_{i=1}^k \lambda_i^2 - P_\Sigma\right) + \mu\left(\frac{1}{\det(\mathbf{A}_e)}\left(\prod_{i=1}^k \lambda_i\right)^2 - \tilde{\eta}\right) \tag{27}$$

where $\nu \geq 0$, $\mu \geq 0$ and $\varsigma_i \geq 0$ for $i \in \{1, \ldots, k\}$ are the Lagrange multipliers. Based on (27), the stationarity and complementary slackness conditions can be derived as follows:

Stationarity conditions:

$$\frac{\partial\mathcal{L}}{\partial\lambda_i} = \frac{-2}{\det(\mathbf{A}_r)}\left(\prod_{j=1}^k \lambda_j\right)^2 \frac{1}{\lambda_i} - \varsigma_i + 2\nu\lambda_i + \frac{2\mu}{\det(\mathbf{A}_e)}\left(\prod_{j=1}^k \lambda_j\right)^2 \frac{1}{\lambda_i} = 0\,, \quad i = 1, \ldots, k \tag{28}$$

Complementary slackness conditions:

$$\varsigma_i\lambda_i = 0\,, \quad i \in \{1, \ldots, k\} \tag{29}$$

$$\nu\left(\sum_{i=1}^k \lambda_i^2 - P_\Sigma\right) = 0 \tag{30}$$

$$\mu\left(\frac{1}{\det(\mathbf{A}_e)}\left(\prod_{j=1}^k \lambda_j\right)^2 - \tilde{\eta}\right) = 0 \tag{31}$$

For the maximization of the objective function in (25a), none of the λ_i terms should be zero, i.e., $\lambda_i > 0$ for all $i \in \{1, \ldots, k\}$. Therefore, (29) implies that $\varsigma_i = 0$ for all $i \in \{1, \ldots, k\}$. Hence, the stationary conditions in (28) become

$$\left(\frac{1}{\det(\mathbf{A}_r)} - \frac{\mu}{\det(\mathbf{A}_e)}\right)\left(\prod_{j=1}^k \lambda_j\right)^2 = \nu\lambda_i^2\,, \quad i = 1, \ldots, k \tag{32}$$

which implies that $\lambda_1 = \cdots = \lambda_k$. Consequently, the solution of (25) can be obtained as in (26) by considering the constraints in (25b) and (25c). It can be verified that the solution in (26) is feasible and satisfies all the KKT conditions. ■

Since $\mathbf{P} = \mathbf{\Lambda}$ as discussed above, Proposition 3 implies that the solution of the D-optimal linear encoding problem in (24) (equivalently, the solution of the D-optimal power allocation problem) is given by

$$\mathbf{P}^* = \min\left\{\sqrt{\frac{P_\Sigma}{k}}, (\tilde{\eta}\det(\mathbf{A}_e))^{\frac{1}{2k}}\right\}\mathbf{I} \tag{33}$$

It is noted that the same power level is assigned to all the parameters in the D-optimality framework, which considers the volume of the ellipsoid that represents the maximum confidence region for the ML estimate of the unknown parameters. On the other hand, for the A-optimality criterion, which considers the MSE of unbiased estimators, different power levels are assigned to parameters in general and linear encoding provides a more general approach than power allocation.

V. Extensions to Cases with Statistical Knowledge of System Parameters

In the previous sections, it is assumed that $\mathbf{A}_r$ and $\mathbf{A}_e$ are known exactly at the transmitter. In the presence of statistical information about $\mathbf{A}_r$ and $\mathbf{A}_e$, we can extend the theoretical results as follows. Suppose that $\mathbf{A}_r$ takes M_r possible values and $\mathbf{A}_e$ takes M_e possible values with known probabilities. In particular, $\mathbf{A}_r = \mathbf{A}_r^{(j)}$ with probability $\rho^{(j)}$ for $j = 1, \ldots, M_r$, and $\mathbf{A}_e = \mathbf{A}_e^{(j)}$ with probability $\kappa^{(j)}$ for $j = 1, \ldots, M_e$. In this setup, for the A-optimality criterion, we can consider the average CRLBs as the performance metrics for both the intended receiver and the eavesdropper, and update the expressions in (13a) and (13c) as follows:

$$\sum_{j=1}^{M_r} \rho^{(j)} \mathrm{tr}\left\{\mathbf{P}^{-1}\mathbf{A}_r^{(j)}\mathbf{P}^{-1}\right\} = \mathrm{tr}\left\{\mathbf{P}^{-1}\overline{\mathbf{A}}_r\mathbf{P}^{-1}\right\} \tag{34}$$

$$\sum_{j=1}^{M_e} \kappa^{(j)} \mathrm{tr}\left\{\mathbf{P}^{-1}\mathbf{A}_e^{(j)}\mathbf{P}^{-1}\right\} = \mathrm{tr}\left\{\mathbf{P}^{-1}\overline{\mathbf{A}}_e\mathbf{P}^{-1}\right\} \tag{35}$$

where $\overline{\mathbf{A}}_r \triangleq \sum_{j=1}^{M_r} \rho^{(j)}\mathbf{A}_r^{(j)}$ and $\overline{\mathbf{A}}_e \triangleq \sum_{j=1}^{M_e} \kappa^{(j)}\mathbf{A}_e^{(j)}$. Since (34) and (35) are in the form of (13a) and (13c), respectively, the results in Section III-A and Section III-B can also be employed for this setup by replacing $\mathbf{A}_r$ with $\overline{\mathbf{A}}_r$ and $\mathbf{A}_e$ with $\overline{\mathbf{A}}_e$, and assuming that $\overline{\mathbf{A}}_r$ is not a scaled version of $\overline{\mathbf{A}}_e$.

If $\mathbf{A}_r$ and $\mathbf{A}_e$ have continuous distributions with probability density functions $f_r(\cdot)$ and $f_e(\cdot)$, respectively, then (34) and (35) can be updated as

$$\int f_r(\mathbf{A}_r)\mathrm{tr}\left\{\mathbf{P}^{-1}\mathbf{A}_r\mathbf{P}^{-1}\right\} d\mathbf{A}_r = \mathrm{tr}\left\{\mathbf{P}^{-1}\overline{\mathbf{A}}_r\mathbf{P}^{-1}\right\} \tag{36}$$

$$\int f_e(\mathbf{A}_e)\mathrm{tr}\left\{\mathbf{P}^{-1}\mathbf{A}_e\mathbf{P}^{-1}\right\} d\mathbf{A}_e = \mathrm{tr}\left\{\mathbf{P}^{-1}\overline{\mathbf{A}}_e\mathbf{P}^{-1}\right\} \tag{37}$$

where $\overline{\mathbf{A}}_r \triangleq \int \mathbf{A}_r f(\mathbf{A}_r) d\mathbf{A}_r$ and $\overline{\mathbf{A}}_e \triangleq \int \mathbf{A}_e f(\mathbf{A}_e) d\mathbf{A}_e$. Hence, the structure of the problem remains the same. To cover the cases of both discrete and continuous distributions, $\overline{\mathbf{A}}_r$ and $\overline{\mathbf{A}}_e$ in (34)–(37) can be stated as $\overline{\mathbf{A}}_r = \mathrm{E}\{\mathbf{A}_r\}$ and $\overline{\mathbf{A}}_e = \mathrm{E}\{\mathbf{A}_e\}$.

Regarding the D-optimality criterion, we can consider the average values for the determinants of the FIMs as the performance metrics, and modify the expressions in (24a) and (24c) as follows:

$$\sum_{j=1}^{M_r} \rho^{(j)} \det\left(\mathbf{P}\left(\mathbf{A}_r^{(j)}\right)^{-1}\mathbf{P}\right) = \left(\prod_{i=1}^{k}\lambda_i\right)^2 \sum_{j=1}^{M_r} \frac{\rho^{(j)}}{\det\left(\mathbf{A}_r^{(j)}\right)} \tag{38}$$

$$\sum_{j=1}^{M_e} \kappa^{(j)} \det\left(\mathbf{P}\left(\mathbf{A}_e^{(j)}\right)^{-1}\mathbf{P}\right) = \left(\prod_{i=1}^{k}\lambda_i\right)^2 \sum_{j=1}^{M_e} \frac{\kappa^{(j)}}{\det\left(\mathbf{A}_e^{(j)}\right)} \tag{39}$$

Then, by following similar steps to those in the proof of Proposition 3, the D-optimal linear encoding (equivalently, power allocation) can be obtained as follows:

$$\mathbf{P}^* = \min\left\{\sqrt{\frac{P_\Sigma}{k}}, \left(\frac{\tilde{\eta}}{\sum_{j=1}^{M_e} \frac{\kappa^{(j)}}{\det\left(\mathbf{A}_e^{(j)}\right)}}\right)^{\frac{1}{2k}}\right\}\mathbf{I} \tag{40}$$

For the case of $\mathbf{A}_r$ and $\mathbf{A}_e$ with continuous distributions, similar derivations can be performed to obtain the D-optimal linear encoding (equivalently, power allocation) as

$$\mathbf{P}^* = \min\left\{\sqrt{\frac{P_\Sigma}{k}}, \left(\frac{\tilde{\eta}}{\mathrm{E}\{1/\det(\mathbf{A}_e)\}}\right)^{\frac{1}{2k}}\right\}\mathbf{I} \tag{41}$$

where $\mathrm{E}\{1/\det(\mathbf{A}_e)\} = \int f_e(\mathbf{A}_e)/\det\left(\mathbf{A}_e\right)d\mathbf{A}_e$.

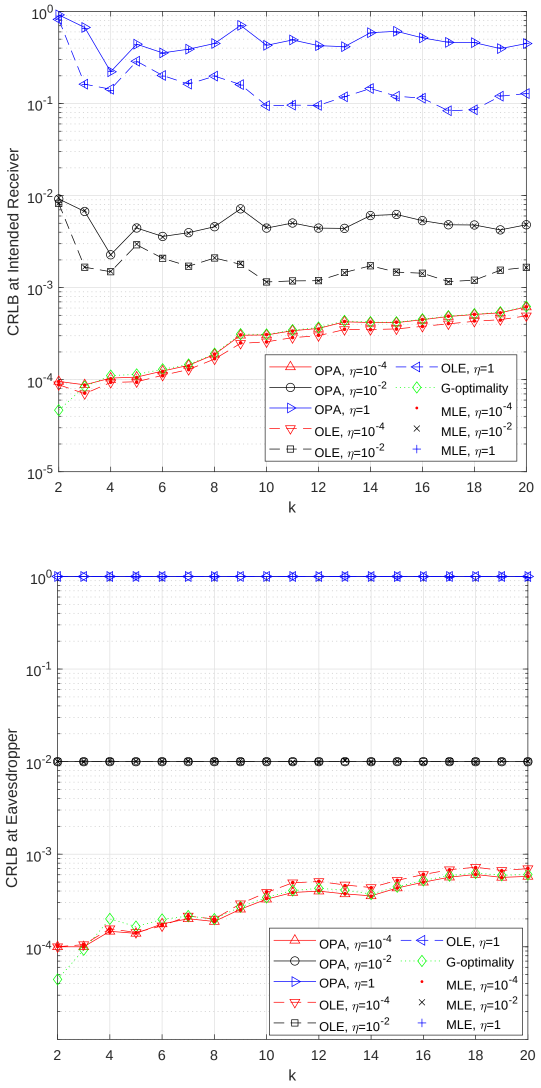

Fig. 1. CRLBs achieved by the OPA and OLE approaches versus k for various values of η, where $P_\Sigma = 10$. Also, the performance of the G-optimality based power allocation is presented for comparison purposes. In addition, the MSEs of the ML estimators for the OPA and OLE approaches are illustrated.

VI. Numerical Results

In this section, we scrutinize the theoretical results based on various numerical examples. As in [34], the system matrices for the intended receiver and the eavesdropper, i.e., $\mathbf{F}_r$ in (1) and $\mathbf{F}_e$ in (2), consist of i.i.d. uniform random variables over $[-0.1, 0.1]$, which are generated as a single realization in MATLAB with seed 1. Also, the additive noise vectors at the intended receiver and the eavesdropper, i.e., $\mathbf{N}_r$ and $\mathbf{N}_e$ in (1) and (2), are modeled as zero-mean and independent Gaussian random vectors with i.i.d. components, where each component has a variance of 10^{-6}. In addition, the numbers of measurements in (1) and (2) are set to $n_e = n_r = 2k$ with k denoting the number of parameters [34].[5]

First, the A-optimality criterion is considered, and the optimal power allocation (OPA) approach in Proposition 1 and the optimal linear encoding (OLE) approach in Proposition 2

[5] The following MATLAB code is used for generating $\mathbf{F}_r$ and $\mathbf{F}_e$: "rng(1); F=(rand(20,80)-0.5)/5; Fr=F(1:k,1:nr); Fe=F(1:k,(nr+1):(nr+ne));".

are evaluated. Fig. 1 presents the CRLBs at the intended receiver and at the eavesdropper that are achieved by the OPA and OLE approaches versus the number of parameters, k, for three different values of the secrecy constraint η and for $P_\Sigma = 10$ (see (9)). It is noted that OLE can provide lower CRLBs at the intended receiver than OPA while satisfying the same secrecy constraints. This is expected as OLE is a more general approach than OPA, and the system matrices are not diagonal in the considered scenario (even though the noise components are i.i.d.). In addition, when $\eta = 10^{-4}$, the secrecy constraint becomes ineffective for the OPA approach when $k \geq 4$ and for the OLE approach when $k \geq 3$. In these situations, the first cases in Propositions 1 and 2 (i.e., (15) and (21)) become valid. In all other cases, the secrecy constraint is effective as noted from Fig. 1, and the solutions are obtained as described in the second cases in Propositions 1 and 2. For comparison purposes, we also present a power allocation approach according to the G-optimality criterion [42], where the aim is to minimize the largest diagonal entry of the CRLB at the intended receiver under the average power constraint. (Namely, $\min_{\{p_i\}_{i=1}^k} \max_{i\in\{1,\ldots,k\}} \alpha_i/p_i$ such that $\sum_{i=1}^k p_i \leq P_\Sigma$ and $p_i \geq 0$, $i = 1,\ldots,k$ (cf. (14)).) From Fig. 1, it is noted that whenever the secrecy constraint is active (that is, when $\eta = 10^{-4}$ and $k \in \{4, 5, \ldots, 20\}$), the proposed OLE approach outperforms the power allocation approach based on G-optimality. Furthermore, we implement the ML estimators for the OPA and OLE approaches (based on 5000 Monte-Carlo trials) and present their MSEs in Fig. 1. As expected, the MSEs coincide with the CRLBs due to the consideration of linear systems models with additive Gaussian noise [29].

For the same setting, Fig. 2 illustrates the CRLBs at the intended receiver and at the eavesdropper achieved by the OPA and OLE approaches versus the secrecy constraint η, where $k = 5, 10, 20$ and $P_\Sigma = 10$. For small values of η, the secrecy constraint is not effective, and the minimum CRLB is achieved at the intended receiver under the average power limit. In this region, the CRLBs are highest for $k = 20$ and lowest for $k = 5$ in accordance with Fig. 1 (see $\eta = 10^{-4}$). However, after a certain value of η, the secrecy constraint becomes effective, and the CRLBs increase in order to satisfy the secrecy constraint. In that regime, the CRLB at the eavesdropper is always equal to η, and the CRLB at the receiver depends on the system dependent matrices $\mathbf{A}_e$ and $\mathbf{A}_r$, which are determined by $\mathbf{F}_r$, $\mathbf{F}_e$, and the covariance matrices of the noise components. For example, it is noted that for large values of η, the CRLBs achieved by OLE are lowest for $k = 10$ and highest for $k = 5$. This behavior is in compliance with Fig. 1 (see 'OLE, $\eta = 10^{-2}$' and OLE, $\eta = 1$), and it is due to the random generation of the system matrices $\mathbf{F}_r$ and $\mathbf{F}_e$. Fig. 2 also shows that the OLE approach achieves lower CRLBs at the intended receiver than the OPA approach for all values of η. For comparison purposes, the performance of the G-optimality based power allocation is also presented in Fig. 2, which corresponds to a constant value for each k due to the omission of the secrecy constraint. Moreover, the MSEs of the ML estimators for the OPA and OLE approaches are illustrated in the figure, which coincide with the CRLBs as expected.

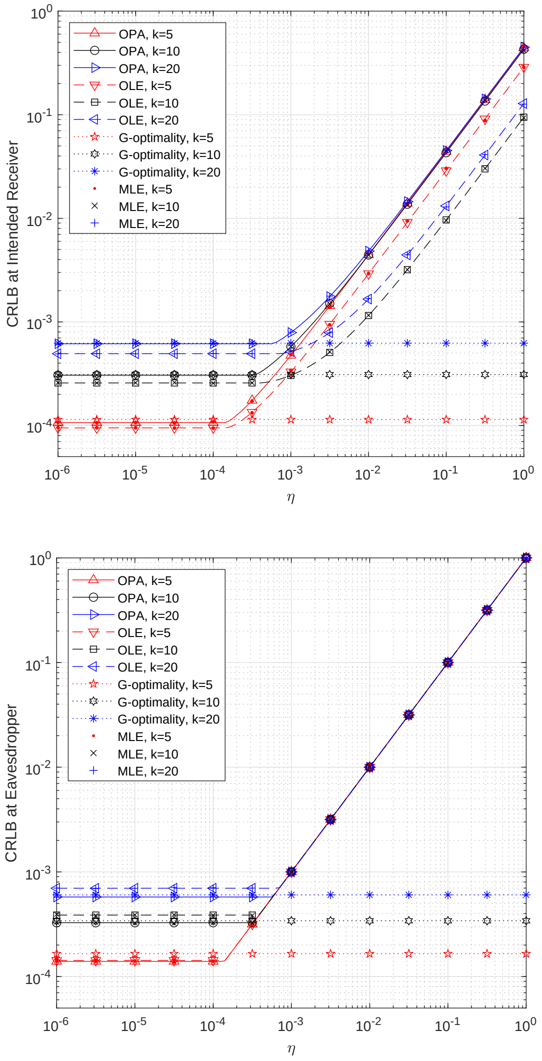

Fig. 2. CRLBs achieved by the OPA and OLE approaches versus η for various values of k, where $P_\Sigma = 10$. Also, the performance of the G-optimality based power allocation is presented for comparison purposes. In addition, the MSEs of the ML estimators for the OPA and OLE approaches are illustrated.

Next, the CRLBs at the intended receiver and the eavesdropper achieved by the OPA and OLE approaches are plotted versus the average power constraint, P_Σ, in Fig. 3 for various values of η and for $k = 10$. As P_Σ increases, the secrecy constraint becomes effective and the CRLB at the receiver cannot be reduced below certain levels after some values of P_Σ. It is also noted that the performance difference between OLE and OPA is more significant for high average power constraints. Moreover, the performance of the G-optimality based power allocation is presented for comparison purposes. Furthermore, as in the previous scenarios, the MSEs of the ML estimators for the OPA and OLE approaches coincide with the CRLBs.

Finally, the D-optimality criterion is considered, and the optimal approach in (33) (see Proposition 3) is evaluated. (Optimal linear encoding and optimal power allocation are equivalent for this criterion.) Fig. 4 presents the determinants of FIM at the intended receiver and at the eavesdropper that are achieved by the optimal approach versus the number of parameters, k, for four different values of the secrecy

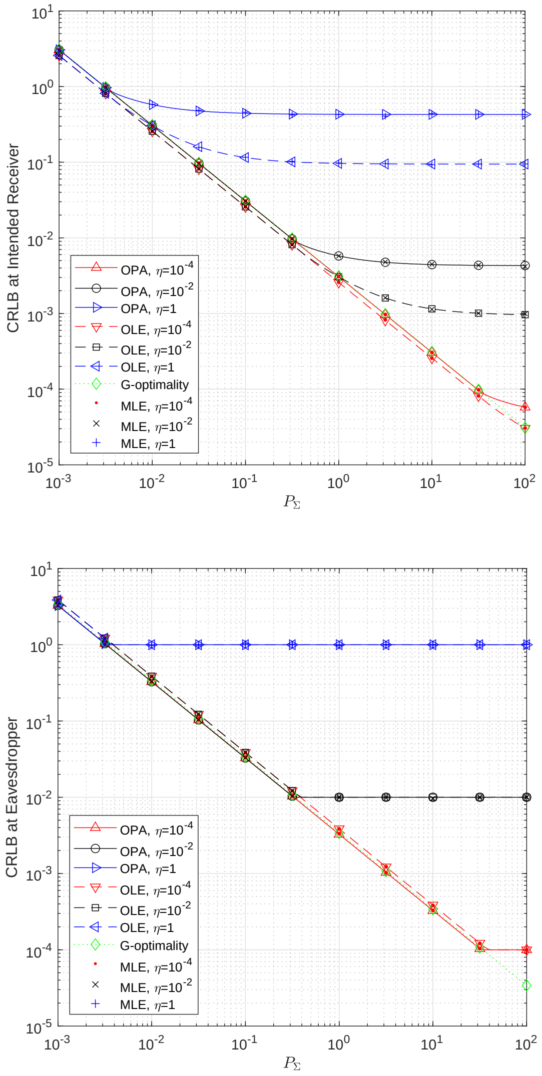

Fig. 3. CRLBs achieved by the OPA and OLE approaches versus P_Σ for various η values, where $k = 10$. Also, the performance of the G-optimality based power allocation is presented for comparison purposes. In addition, the MSEs of the ML estimators for the OPA and OLE approaches are illustrated.

constraint $\tilde{\eta}$ and for $P_\Sigma = 0.01$ (see (10)). It is noted that higher determinants of FIM are obtained as $\tilde{\eta}$ increases; i.e., as the secrecy constraint is relaxed. Also, the secrecy constraint is effective at all points in the figure except for $\tilde{\eta} = 10^6$ with $k = 2$ and $k = 3$ and for $\tilde{\eta} = 10^4$ with $k = 2$. In addition, the fluctuations in the determinants of FIM at the intended receiver are due to the random generation of the system matrices $\mathbf{F}_r$ and $\mathbf{F}_e$, as mentioned previously. For comparison purposes, the optimal solution in the absence of the secrecy constraint (e.g., similar to [35, Eq. (4)]) is also presented in Fig. 4 (labeled as 'Insecure'), which leads to $p_i = P_\Sigma/k$ for $i = 1, \ldots, k$. It is noted that the determinant of the FIM increases with k for both the intended receiver and the eavesdropper in this case, resulting in a violation of the secrecy constraint. In Fig. 5, the determinants of FIM at the intended receiver and at the eavesdropper are plotted versus P_Σ by considering various values of $\tilde{\eta}$ and k. It is observed that the secrecy constraint becomes effective as P_Σ increases, and the determinants of FIM at the intended receiver are larger for $k = 5$ than those for $k = 10$ in compliance with the results in Fig. 4. Again, the optimal solution in the absence of the secrecy constraint is also presented for comparison purposes.

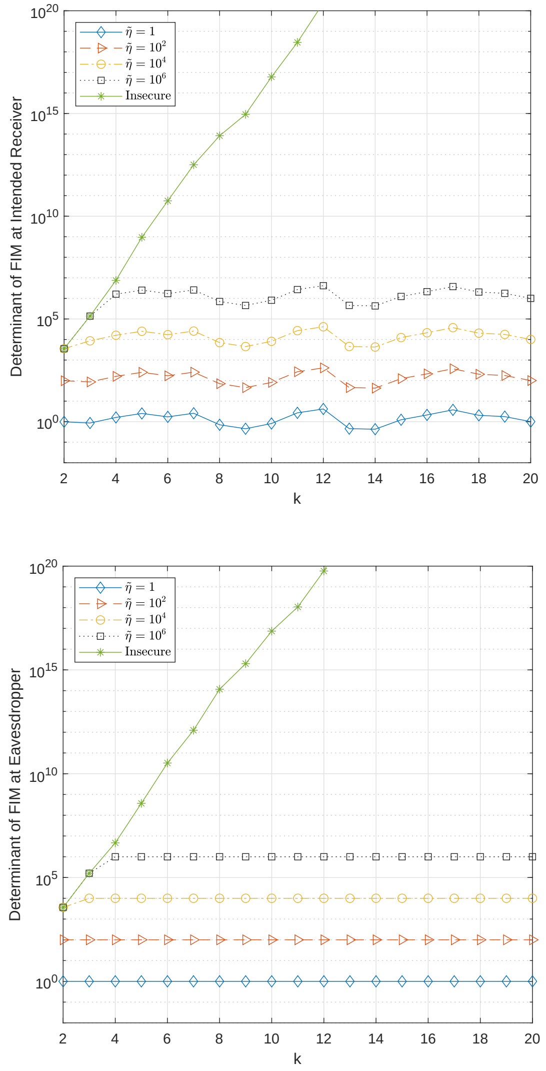

Fig. 4. Determinant of FIM achieved by the optimal approach versus k for various values of $\tilde{\eta}$, where $P_\Sigma = 0.01$. Also, the solution in the absence of the secrecy constraint (labeled as 'Insecure') is presented for comparison purposes.

By comparing the results for the A-optimality and D-optimality criteria, it is deduced that the performance difference between the intended receiver and the eavesdropper can be made more significant in the A-optimality framework in most cases. This is due to the optimization of all the elements of matrix $\mathbf{P}$ in (1) and (2) by the OLE approach employed for the A-optimality criterion. On the other hand, the D-optimality criterion always results in a diagonal $\mathbf{P}$ with equal diagonal elements, leading to limited flexibility to achieve improved performance over the eavesdropper. Hence, it is more challenging to perform secure estimation according to the D-optimality criterion.

VII. Concluding Remarks

The secure transmission of deterministic vector parameters has been investigated in the presence of a smart eavesdropper

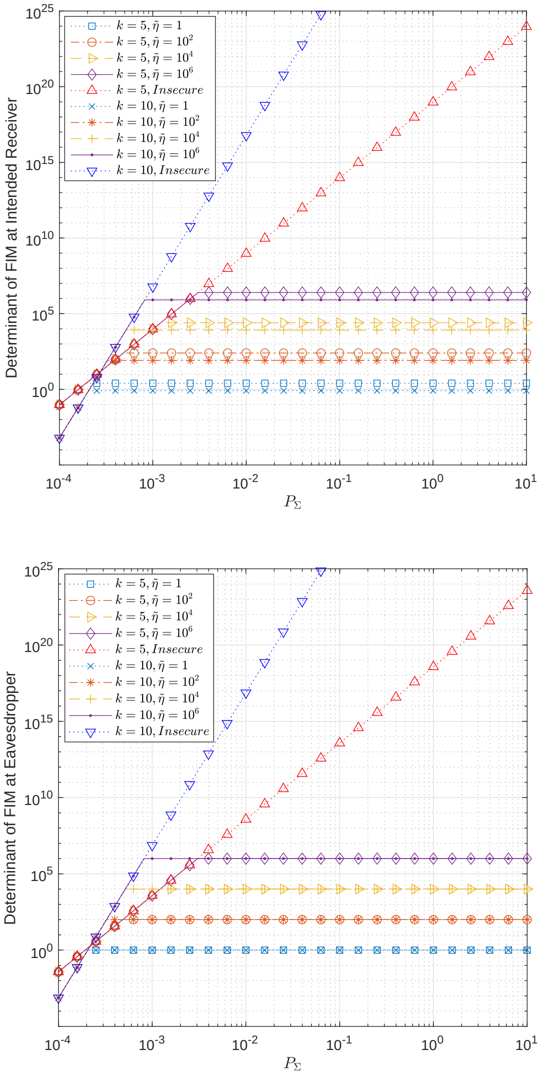

Fig. 5. Determinant of FIM achieved by the optimal approach versus P_Σ for various $\tilde{\eta}$ and k values. Also, the solution in the absence of the secrecy constraint (labeled as 'Insecure') is presented for comparison purposes.

according to the A-optimality and D-optimality criteria. First, the optimal power allocation and optimal linear encoding strategies have been characterized theoretically for the A-optimality criterion. It has been stated that the optimal linear encoding can provide improved estimation performance compared to the optimal power allocation in general. Then, it has been shown that the optimal linear encoding and the optimal power allocation lead to the same equal power allocation solution for the D-optimality criterion. In addition, extensions have been provided to cases with statistical knowledge of systems parameters. Via numerical examples, the optimal power allocation and optimal linear encoding strategies have been investigated in different scenarios.

Although the presence of a single eavesdropper is considered in this work, the case of multiple eavesdroppers can also be treated based on similar theoretical approaches. In particular, by considering an individual secrecy constraint for each eavesdropper, a straightforward extension can be performed in the D-optimality framework. On the other hand, in the A-optimality framework, the complexity of the solutions of the A-optimal linear encoding and power allocation problems depends on which of the M secrecy constraints are active, and the computational complexity can get high in the presence of many eavesdroppers.

As another extension, worst-case design approaches can be considered in the presence of statistical information about $\mathbf{A}_r$ and $\mathbf{A}_e$ instead of the average performance based approach in Section V. In particular, for the A-optimality criterion, the worst-case design problem becomes

$$\min_{\mathbf{P}} \; \max_{\mathbf{A}_r\in\mathcal{S}_r} \operatorname{tr}\{\mathbf{P}^{-1}\mathbf{A}_r\mathbf{P}^{-1}\} \tag{42a}$$

$$\text{s.t.} \quad \operatorname{tr}\{\mathbf{P}\mathbf{P}^T\} \le P_\Sigma \tag{42b}$$

$$\min_{\mathbf{A}_e\in\mathcal{S}_e} \operatorname{tr}\{\mathbf{P}^{-1}\mathbf{A}_e\mathbf{P}^{-1}\} \ge \eta \tag{42c}$$

where $\mathcal{S}_r = \{\mathbf{A}_r^{(1)},\dots,\mathbf{A}_r^{(M_r)}\}$ and $\mathcal{S}_e = \{\mathbf{A}_e^{(1)},\dots,\mathbf{A}_e^{(M_e)}\}$. Similarly, for the D-optimality criterion, the worst case design problem can be formulated as follows:

$$\max_{\mathbf{P}} \; \min_{\mathbf{A}_r\in\mathcal{S}_r} \det\left(\mathbf{P}\mathbf{A}_r^{-1}\mathbf{P}\right) \tag{43a}$$

$$\text{s.t.} \quad \operatorname{tr}\{\mathbf{P}\mathbf{P}^T\} \le P_\Sigma \tag{43b}$$

$$\max_{\mathbf{A}_e\in\mathcal{S}_e} \det\left(\mathbf{P}\mathbf{A}_e^{-1}\mathbf{P}\right) \le \tilde{\eta} \tag{43c}$$

Obtaining a closed-form solution for (42) is difficult in general, which is considered as a possible direction for future work.[6] On the other hand, the problem in (43) can easily be solved since it is equivalent to the following problem, which is in the same form as (24):

$$\max_{\mathbf{P}} \quad \det\left(\mathbf{P}(\mathbf{A}_r^*)^{-1}\mathbf{P}\right) \tag{44a}$$

$$\text{s.t.} \quad \operatorname{tr}\{\mathbf{P}\mathbf{P}^T\} \le P_\Sigma \tag{44b}$$

$$\det\left(\mathbf{P}(\mathbf{A}_e^*)^{-1}\mathbf{P}\right) \le \tilde{\eta} \tag{44c}$$

where $\mathbf{A}_r^* \triangleq \underset{\mathbf{A}_r}{\operatorname{argmax}} \det(\mathbf{A}_r)$ and $\mathbf{A}_e^* \triangleq \underset{\mathbf{A}_e}{\operatorname{argmin}} \det(\mathbf{A}_e)$.

Appendix

A. Proof of Proposition 1

To obtain the relations that a solution of (14) must satisfy, the KKT conditions can be considered. To that aim, the Lagrangian function for (14) is expressed as follows:

$$\mathcal{L}\left(\{p_i\}_{i=1}^k, \{\varsigma_i\}_{i=1}^k, \nu, \mu\right) = \sum_{i=1}^k \frac{\alpha_i}{p_i} - \sum_{i=1}^k \varsigma_i p_i + \nu\left(\sum_{i=1}^k p_i - P_\Sigma\right) + \mu\left(\eta - \sum_{i=1}^k \frac{\beta_i}{p_i}\right) \tag{45}$$

where $\nu \ge 0$, $\mu \ge 0$ and $\varsigma_i \ge 0$ for $i \in \{1,\dots,k\}$ are the Lagrange multipliers. From (45), the stationarity and complementary slackness conditions can be derived as follows:

Stationarity conditions:

$$\frac{\partial \mathcal{L}}{\partial p_i} = -\frac{\alpha_i}{p_i^2} - \varsigma_i + \nu + \mu\frac{\beta_i}{p_i^2} = 0\,, \quad i = 1,\dots,k \tag{46}$$

[6] If the average CRLB performance is considered for the intended receiver and the worst-case CRLB is employed for the eavesdropper, the formulation corresponds to the A-optimal linear encoding problem in the presence of multiple eavesdroppers.

Complementary slackness conditions:

$$\varsigma_i p_i = 0\,, \quad i \in \{1, \ldots, k\} \tag{47}$$

$$\nu \left(\sum_{i=1}^{k} p_i - P_\Sigma \right) = 0 \tag{48}$$

$$\mu \left(\eta - \sum_{i=1}^{k} \frac{\beta_i}{p_i} \right) = 0 \tag{49}$$

Since $\alpha_i > 0$ for all $i \in \{1, \ldots, k\}$, p_i cannot be zero for minimizing the objective function in (14a); hence, $p_i > 0$. As a result, according to (47), it is concluded that $\varsigma_i = 0$ for all $i \in \{1, \ldots, k\}$. Therefore, the stationary conditions in (46) become

$$\frac{\alpha_i - \mu\beta_i}{p_i^2} = \nu\,, \quad i = 1, \ldots, k \tag{50}$$

Then, the following two cases are investigated:

<u>**Case 1** ($\mu = 0$):</u>

In this case, (50) simplifies to $\nu = \alpha_i / p_i^2$, implying that $\nu > 0$. Hence, (48) leads to $\sum_{i=1}^{k} p_i = P_\Sigma$. Accordingly, the following relation is obtained:

$$\sqrt{\nu} = \frac{1}{P_\Sigma} \sum_{i=1}^{k} \sqrt{\alpha_i} \tag{51}$$

which results in (15) since $p_i = \sqrt{\alpha_i}/\sqrt{\nu}$. The solution in (15) valid whenever $\sum_{i=1}^{k} \beta_i / p_i^* \geq \eta$, leading to the statement at the beginning of Proposition 1 by considering p_i^* in (15).

<u>**Case 2** ($\mu > 0$):</u>

In this case, (49) leads to the condition of $\sum_{i=1}^{k} \beta_i / p_i = \eta$. At this point, two sub-cases can be investigated as follows:

<u>**Case 2-a:**</u> Suppose that $\sum_{i=1}^{k} p_i < P_\Sigma$. Then, based on (48), $\nu = 0$ holds. Therefore, the stationarity conditions in (50) become

$$\alpha_i = \mu\beta_i\,, \quad i = 1, \ldots, k \tag{52}$$

implying that $\frac{\alpha_1}{\beta_1}, \ldots, \frac{\alpha_k}{\beta_k}$ are all the same. However, this is not possible under the assumption that $\mathbf{A}_r$ is not a scaled version of $\mathbf{A}_e$, as stated before Proposition 1. (In other words, $\sum_{i=1}^{k} p_i < P_\Sigma$ is possible only when $\frac{\alpha_i}{\beta_i}$'s are all the same. In that case, the expressions in (14a) and (14c) are always scaled version of each other, and the optimization problem becomes trivial and achieves the minimum objective value of $\mu\,\eta$.)

<u>**Case 2-b:**</u> Suppose that $\sum_{i=1}^{k} p_i = P_\Sigma$. Then, (48) implies that $\nu \geq 0$. Therefore, the stationarity conditions in (50) can be utilized to conclude that $\alpha_i - \mu\beta_i \geq 0$ for all $i \in \{1, \ldots, k\}$. Hence, μ can be bounded from above as $\mu \leq \alpha_i / \beta_i$ for all $i \in \{1, \ldots, k\}$, which can also be written as

$$\mu \leq \min_{i \in \{1, \ldots, k\}} \frac{\alpha_i}{\beta_i} \tag{53}$$

Besides, p_i^* can be calculated from (50) as

$$p_i^* = \frac{\sqrt{\alpha_i - \mu\beta_i}}{\sqrt{\nu}}\,, \quad i = 1, \ldots, k \tag{54}$$

From (54) and the condition of $\sum_{i=1}^{k} p_i = P_\Sigma$, ν can be calculated as

$$\sqrt{\nu} = \frac{1}{P_\Sigma} \sum_{i=1}^{k} \sqrt{\alpha_i - \mu\beta_i} \tag{55}$$

In addition, from (54) and the condition of $\sum_{i=1}^{k} \beta_i / p_i = \eta$, ν can also be stated as

$$\sqrt{\nu} = \frac{\eta}{\sum_{i=1}^{k} \frac{\beta_i}{\sqrt{\alpha_i - \mu\beta_i}}} \tag{56}$$

Then, equating the expressions in (55) and (56), the following condition is obtained:

$$\eta P_\Sigma = \sum_{i=1}^{k} \sqrt{\alpha_i - \mu\beta_i} \sum_{i=1}^{k} \frac{\beta_i}{\sqrt{\alpha_i - \mu\beta_i}} \triangleq g(\mu) \tag{57}$$

The expression in (57) leads to unique solution for μ, as proved in the following. First, $g(0) = \sum_{i=1}^{k} \sqrt{\alpha_i} \sum_{i=1}^{k} \beta_i / \sqrt{\alpha_i} \leq \eta P_\Sigma$ due to the condition of $\sum_{i=1}^{k} \frac{\beta_i \sum_{j=1}^{k} \sqrt{\alpha_j}}{P_\Sigma \sqrt{\alpha_i}} \leq \eta$. (Otherwise, the solution in Case 1 would be valid.) Also, for the maximum value of μ specified in (53), it can be shown from (57) that $g(\min_{i \in \{1, \ldots, k\}} \alpha_i / \beta_i) = \infty$. In addition, $g(\mu)$ is a continuously differentiable function for $\mu \in [0, \min_{i \in \{1, \ldots, k\}} \alpha_i / \beta_i)$ and its derivative can be obtained from (57) as

$$\begin{aligned} \frac{dg(\mu)}{d\mu} &= -\frac{1}{2} \left(\sum_{i=1}^{k} \frac{\beta_i}{\sqrt{\alpha_i - \mu\beta_i}} \right)^2 \\ &+ \frac{1}{2} \left(\sum_{i=1}^{k} \sqrt{\alpha_i - \mu\beta_i} \right) \left(\sum_{i=1}^{k} \frac{\beta_i^2}{(\alpha_i - \mu\beta_i)^{3/2}} \right) \end{aligned} \tag{58}$$

From Cauchy-Schwarz inequality, it can be shown that $\frac{dg(\mu)}{d\mu} > 0$ for all $\mu \in [0, \min_{i \in \{1, \ldots, k\}} \alpha_i / \beta_i)$. (The equality condition in Cauchy-Schwarz inequality, i.e., $\frac{dg(\mu)}{d\mu} = 0$, holds if $\beta_i = \tilde{K}(\alpha_i - \mu\beta_i)$ for each i, which leads to $\alpha_i = (\mu + 1/\tilde{K})\beta_i$. However, this is in contrary to the assumption that $\frac{\alpha_i}{\beta_i}$'s are not all the same.) Therefore, $g(\mu)$ is a continuous and monotone increasing function of μ from $\mu = 0$ to $\mu = \min_{i \in \{1, \ldots, k\}} \alpha_i / \beta_i$. Since it starts from a value less than or equal to ηP_Σ and goes to infinity, it is guaranteed that (57) has a unique solution denoted by μ^*, as specified in (17) in Proposition 1. Once μ^* is obtained from (57) (i.e., (17)), the corresponding ν^* can be calculated from (55). Then, inserting this ν^* into (54) yields the solution in (16) of Proposition 1.

It should be emphasized that even though the problem in (14) is not convex (due to the constraint in (14c)), the KKT conditions become both necessary and sufficient for the minimizer since they lead to a unique structure and the problem admits a minimizer over the feasible region.

B. Proof of Lemma 1

As $\mathbf{A}_r - \mu\mathbf{A}_e$ is positive definite for $\mu \in [0, \mu_{\min})$, it can be diagonalized as follows:

$$\mathbf{A}_r - \mu\mathbf{A}_e = \sum_{i=1}^{k} \psi_i(\mu) \mathbf{q}_i(\mu) \mathbf{q}_i(\mu)^T \tag{59}$$

where $\psi_i(\mu)$ and $\mathbf{q}_i(\mu)$ denote the ith eigenvalue and the ith eigenvector of $\mathbf{A}_r - \mu\mathbf{A}_e$, respectively, and the eigenvectors are chosen as orthonormal vectors. From (59), $h(\mu)$ defined in the lemma can be expressed as

$$h(\mu) = \sum_{j=1}^{k} \sqrt{\psi_j(\mu)} \sum_{i=1}^{k} \frac{\mathrm{tr}\{\mathbf{A}_e \mathbf{q}_i(\mu) \mathbf{q}_i(\mu)^T\}}{\sqrt{\psi_i(\mu)}} \tag{60}$$

To compute the derivative of $h(\mu)$, we need to calculate the derivatives of $\psi_i(\mu)$ and $\mathbf{q}_i(\mu)$ with respect to μ. To this aim, the orhonormality condition of $\mathbf{q}_i(\mu)^T\mathbf{q}_i(\mu) = 1$ is employed first in order to obtain the following relation:

$$\frac{\partial \mathbf{q}_i(\mu)^T\mathbf{q}_i(\mu)}{\partial\mu} = 2\mathbf{q}_i(\mu)^T\frac{\partial \mathbf{q}_i(\mu)}{\partial\mu} = 0. \tag{61}$$

Hence, $\mathbf{q}_i(\mu)^T\frac{\partial \mathbf{q}_i(\mu)}{\partial\mu} = 0$ for any $i \in \{1,\ldots,k\}$. Also, by taking the derivative of both sides of the equation $(\mathbf{A}_r - \mu\mathbf{A}_e)\mathbf{q}_i(\mu) = \psi_i(\mu)\mathbf{q}_i(\mu)$, we obtain

$$(\mathbf{A}_r-\mu\mathbf{A}_e)\frac{\partial \mathbf{q}_i(\mu)}{\partial\mu}-\mathbf{A}_e\mathbf{q}_i(\mu) = \frac{\partial\psi_i(\mu)}{\partial\mu}\mathbf{q}_i(\mu)+\frac{\partial \mathbf{q}_i(\mu)}{\partial\mu}\psi_i(\mu) \tag{62}$$

After multiplying both sides of (62) with $\mathbf{q}_i(\mu)^T$ and using (61), the following relation is derived:

$$\frac{\partial\psi_i(\mu)}{\partial\mu} = -\mathbf{q}_i(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu) \tag{63}$$

Moreover, if we multiply both sides of (62) with $\mathbf{q}_j(\mu)^T$ for any $j \neq i$ and employ the fact that $\mathbf{q}_j(\mu)^T\mathbf{q}_i(\mu) = 0$ for $j \neq i$, the following equation is reached:

$$\left(\psi_j(\mu)-\psi_i(\mu)\right)\mathbf{q}_j(\mu)^T\frac{\partial \mathbf{q}_i(\mu)}{\partial\mu} = \mathbf{q}_j(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu) \tag{64}$$

When $i \neq j$, (64) can be interpreted as follows:

$$\psi_i(\mu) = \psi_j(\mu) \implies \mathbf{q}_j(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu) = 0 \tag{65}$$

$$\psi_i(\mu) \neq \psi_j(\mu) \implies \mathbf{q}_j(\mu)^T\frac{\partial \mathbf{q}_i(\mu)}{\partial\mu} = \frac{\mathbf{q}_j(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu)}{\psi_j(\mu)-\psi_i(\mu)} \tag{66}$$

It is noted that since $\{\mathbf{q}_i(\mu)\}_{i=1}^k$ form an orthonormal basis, there exists $\{a_{ij}\}_{i,j} \in \mathbb{R}$ such that

$$\frac{\partial \mathbf{q}_i(\mu)}{\partial\mu} = \sum_{j=1}^k a_{ij}\mathbf{q}_j(\mu) \tag{67}$$

It is known that $a_{ii} = 0$ by (61), and if $\psi_i(\mu)$ and $\psi_j(\mu)$ are different, a_{ij} is given by (66). From (67), the derivative of $\mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T$ with respect to μ can be expressed as

$$\frac{\partial \mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T}{\partial\mu} = \sum_{j=1}^k a_{ij}\left(\mathbf{q}_j(\mu)\mathbf{q}_i(\mu)^T + \mathbf{q}_i(\mu)\mathbf{q}_j(\mu)^T\right) \tag{68}$$

Based on (60), the derivative of $h(\mu)$ with respect to μ is expressed as

$$\begin{aligned}\frac{\partial h(\mu)}{\partial\mu} &= \sum_{j=1}^k\sqrt{\psi_j(\mu)}\sum_{i=1}^k\frac{1}{\sqrt{\psi_i(\mu)}}\frac{\partial\mathrm{tr}\{\mathbf{A}_e\mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T\}}{\partial\mu}\\ &+\sum_{j=1}^k\sum_{i=1}^k\frac{\partial\gamma_{i,j}(\mu)}{\partial\mu}\mathrm{tr}\{\mathbf{A}_e\mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T\}\end{aligned} \tag{69}$$

where $\gamma_{i,j}(\mu) \triangleq \sqrt{\psi_j(\mu)/\psi_i(\mu)}$. From (68), the derivative of $\mathrm{tr}\{\mathbf{A}_e\mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T\}$ can be calculated as

$$\begin{aligned}&\frac{\partial}{\partial\mu}\mathrm{tr}\{\mathbf{A}_e\mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T\}\\ &= \sum_{j=1}^k a_{ij}\mathrm{tr}\left\{\mathbf{A}_e\left(\mathbf{q}_j(\mu)\mathbf{q}_i(\mu)^T + \mathbf{q}_i(\mu)\mathbf{q}_j(\mu)^T\right)\right\}\\ &= \sum_{j=1}^k 2a_{ij}\mathbf{q}_i(\mu)^T\mathbf{A}_e\mathbf{q}_j(\mu)\end{aligned} \tag{70}$$

From the results given in (65) and (66), (70) can be rewritten as follows:

$$\begin{aligned}\sum_{j=1}^k 2a_{ij}\mathbf{q}_i(\mu)^T\mathbf{A}_e\mathbf{q}_j(\mu) &= 2\sum_{j\in\mathcal{S}_i}a_{ij}\mathbf{q}_i(\mu)^T\mathbf{A}_e\mathbf{q}_j(\mu)\\ &= 2\sum_{j\in\mathcal{S}_i}\frac{\left(\mathbf{q}_j(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu)\right)^2}{\psi_j(\mu)-\psi_i(\mu)}\end{aligned} \tag{71}$$

where $\mathcal{S}_i \triangleq \{j : \psi_j(\mu) \neq \psi_i(\mu)\}$ for $i \in \{1,\ldots,k\}$.

Based on (70) and (71), the summation in the first term of (69) can be calculated as follows:

$$\sum_{i=1}^k\frac{1}{\sqrt{\psi_i(\mu)}}\frac{\partial}{\partial\mu}\mathrm{tr}\{\mathbf{A}_e\mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T\} \tag{72}$$

$$= 2\sum_{i=1}^k\sum_{j\in\mathcal{S}_i}\frac{\left(\mathbf{q}_j(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu)\right)^2}{\sqrt{\psi_i(\mu)}\left(\psi_j(\mu)-\psi_i(\mu)\right)} \tag{73}$$

In addition, it is observed that when $\psi_j(\mu) \neq \psi_i(\mu)$, the following inequality must hold:

$$\begin{aligned}&\frac{\left(\mathbf{q}_j(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu)\right)^2}{\sqrt{\psi_i(\mu)}\left(\psi_j(\mu)-\psi_i(\mu)\right)} + \frac{\left(\mathbf{q}_i(\mu)^T\mathbf{A}_e\mathbf{q}_j(\mu)\right)^2}{\sqrt{\psi_j(\mu)}\left(\psi_i(\mu)-\psi_j(\mu)\right)}\\ &= \frac{\left(\mathbf{q}_j(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu)\right)^2}{\psi_j(\mu)-\psi_i(\mu)}\left(\frac{1}{\sqrt{\psi_i(\mu)}}-\frac{1}{\sqrt{\psi_j(\mu)}}\right)\\ &= \frac{\left(\mathbf{q}_j(\mu)^T\mathbf{A}_e\mathbf{q}_i(\mu)\right)^2}{\sqrt{\psi_j(\mu)}+\sqrt{\psi_i(\mu)}}\frac{1}{\sqrt{\psi_i(\mu)\psi_j(\mu)}} \geq 0\end{aligned} \tag{74}$$

The relations in (73) and (74) imply that

$$\sum_{i=1}^k\frac{1}{\sqrt{\psi_i(\mu)}}\frac{\partial}{\partial\mu}\mathrm{tr}\{\mathbf{A}_e\mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T\} \geq 0 \tag{75}$$

Hence, it is concluded that the first term in (69) is non-negative; i.e.,

$$\sum_{j=1}^k\sqrt{\psi_j(\mu)}\sum_{i=1}^k\frac{1}{\sqrt{\psi_i(\mu)}}\frac{\partial}{\partial\mu}\mathrm{tr}\{\mathbf{A}_e\mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T\} \geq 0 \tag{76}$$

Regarding the the second term in (69), the derivative of $\gamma_{i,j}(\mu)$ with respect to μ can easily be computed as follows:

$$\frac{\partial\gamma_{i,j}(\mu)}{\partial\mu} = \frac{\sqrt{\psi_i(\mu)}\left(\tau_i\psi_j(\mu)-\tau_j\psi_i(\mu)\right)}{2\sqrt{\psi_j(\mu)}\psi_i(\mu)^2} \tag{77}$$

where $\tau_i \triangleq \mathbf{q}_i(\mu)^T \mathbf{A}_e \mathbf{q}_i(\mu)$ for $i \in \{1, \ldots, k\}$. Thus, we can write the following chain of equations:

$$\sum_{j=1}^{k}\sum_{i=1}^{k} \frac{\partial \gamma_{i,j}(\mu)}{\partial \mu} \mathrm{tr}\{\mathbf{A}_e \mathbf{q}_i(\mu)\mathbf{q}_i(\mu)^T\} \tag{78}$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{k} \frac{\partial \gamma_{i,j}(\mu)}{\partial \mu} \tau_i \tag{79}$$

$$= \frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{k} \frac{\sqrt{\psi_i(\mu)}\,(\tau_i \psi_j(\mu) - \tau_j \psi_i(\mu))}{\sqrt{\psi_j(\mu)}\psi_i(\mu)^2} \tau_i \tag{80}$$

$$= \frac{1}{2}\sum_{j=1}^{k}\sqrt{\psi_j(\mu)}\sum_{i=1}^{k}\frac{\tau_i^2}{\psi_i(\mu)^{3/2}} - \frac{1}{2}\left(\sum_{i=1}^{k}\frac{\tau_i}{\sqrt{\psi_i(\mu)}}\right)^2 \geq 0 \tag{81}$$

where (80) follows from (77) and (81) is due to the Cauchy Schwarz inequality. By combining (69), (76), and (81), it is proved that

$$\frac{\partial h(\mu)}{\partial \mu} \geq 0\,. \tag{82}$$

Equality Case: To determine when the derivative in (82) becomes zero, we investigate the conditions under which the inequalities in (74) and (81) are satisfied with equality. By considering (74), it is noted that whenever $\psi_i(\mu) \neq \psi_j(\mu)$, we must have $\mathbf{q}_i(\mu)^T \mathbf{A}_e \mathbf{q}_j(\mu) = 0$. Also, by (65), it is known that when $i \neq j$ and $\psi_i(\mu) \neq \psi_j(\mu)$, we have $\mathbf{q}_i(\mu)^T \mathbf{A}_e \mathbf{q}_j(\mu) = 0$. That is, (74) is satisfied with equality if and only if

$$\mathbf{q}_i(\mu)^T \mathbf{A}_e \mathbf{q}_j(\mu) = 0 \tag{83}$$

for any $i \neq j$. On the other hand, to satisfy the inequality in (81) with equality, $\tau_i/\psi_i(\mu)$ must be a constant for each $i \in \{1, \ldots, k\}$. Let that constant be denoted by $C \in \mathbb{R}$. In other words, (81) is satisfied with equality if and only if

$$\mathbf{q}_i(\mu)^T \mathbf{A}_e \mathbf{q}_i(\mu) = C \psi_i(\mu) \tag{84}$$

for any $i \in \{1, \ldots, k\}$. Therefore, by combining (83) and (84), it is obtained that

$$\frac{\partial h(\mu)}{\partial \mu} = 0 \iff \mathbf{q}_i(\mu)^T \mathbf{A}_e \mathbf{q}_j(\mu) = \begin{cases} 0, & \text{if } i \neq j \\ C\psi_i(\mu), & \text{if } i = j \end{cases} \tag{85}$$

If the condition in (85) is satisfied, by multiplying both sides of the equation $(\mathbf{A}_r - \mu \mathbf{A}_e)\mathbf{q}_i(\mu) = \psi_i(\mu)\mathbf{q}_i(\mu)$ with $\mathbf{q}_j(\mu)^T$, we can reach the following equation:

$$\mathbf{q}_i(\mu)^T \mathbf{A}_r \mathbf{q}_j(\mu) = \begin{cases} 0, & \text{if } i \neq j \\ (1+\mu C)\psi_i(\mu), & \text{if } i = j \end{cases} \tag{86}$$

As $\{\mathbf{q}_i(\mu)\}_{i=1}^k$ form an orthonormal basis, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$, there exist $\{x_i\}_{i=1}^k, \{y_i\}_{i=1}^k \in \mathbb{R}$ such that

$$\mathbf{x} = \sum_{i=1}^{k} x_i \mathbf{q}_i(\mu) \text{ and } \mathbf{y} = \sum_{i=1}^{k} y_i \mathbf{q}_i(\mu) \tag{87}$$

Then, the following chain of equations must be true:

$$\mathbf{x}^T \mathbf{A}_e \mathbf{y} = C \sum_{i=1}^{k} x_i y_i \psi_i(\mu) \tag{88}$$

$$\mathbf{x}^T \mathbf{A}_r \mathbf{y} = (1+\mu C) \sum_{i=1}^{k} x_i y_i \psi_i(\mu) \tag{89}$$

This means that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$,

$$\mathbf{x}^T \mathbf{A}_e \mathbf{y} = \frac{C}{1+\mu C} \mathbf{x}^T \mathbf{A}_r \mathbf{y} \tag{90}$$

By taking $\mathbf{x} = \mathbf{e}_\ell$ and $\mathbf{y} = \mathbf{e}_m$, it is evident that $[\mathbf{A}_e]_{\ell,m} = C/(1+\mu C)[\mathbf{A}_r]_{\ell,m}$, for any ℓ, m, where $\mathbf{e}_\ell$ and $\mathbf{e}_m$ are unit-norm vectors with only the ℓth and the mth elements being equal to one, respectively. Therefore, it is shown that $\mathbf{A}_r$ is a scaled version of $\mathbf{A}_e$ in this cases. Hence, it is proved that

$$\frac{\partial h(\mu)}{\partial \mu} = 0 \iff \mathbf{A}_r \text{ is a scaled version of } \mathbf{A}_e. \tag{91}$$

C. *Proof of Proposition 2*

For the problem in (20), the Lagrangian can be expressed as [51]

$$\begin{aligned} \mathcal{L}(\mathbf{V}, \mathbf{D}, \mu, \nu, \mathbf{C}) = \mathrm{tr}\{\mathbf{V}^T \mathbf{A}_r \mathbf{V}\mathbf{D}\} + \mu(\eta - \mathrm{tr}\{\mathbf{V}^T \mathbf{A}_e \mathbf{V}\mathbf{D}\}) \\ + \nu(\mathrm{tr}\{\mathbf{D}^{-1}\} - P_\Sigma) - \mathrm{tr}\{\mathbf{C}^T(\mathbf{V}^T\mathbf{V} - \mathbf{I})\} \end{aligned} \tag{92}$$

with $\mathbf{C}$ being a diagonal matrix, where $\mu \geq 0$, $\nu \geq 0$, and the diagonal elements of $\mathbf{C}$ are the Lagrange multipliers. Considering the KKT conditions, we can express the complementary slackness conditions for (20) as [45]

$$\mu(\eta - \mathrm{tr}\{\mathbf{V}^T \mathbf{A}_e \mathbf{V}\mathbf{D}\}) = 0 \tag{93}$$

$$\nu(\mathrm{tr}\{\mathbf{D}^{-1}\} - P_\Sigma) = 0 \tag{94}$$

In addition, the stationarity conditions can be stated as [45]

$$\frac{\partial \mathcal{L}}{\partial \mathbf{V}} = \mathbf{0}\,, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{D}} = \mathbf{0} \tag{95}$$

which, after some manipulation of (92), lead to

$$\frac{\partial \mathcal{L}}{\partial \mathbf{V}} = 2\mathbf{A}_r\mathbf{V}\mathbf{D} - 2\mu\mathbf{A}_e\mathbf{V}\mathbf{D} - 2\mathbf{V}\mathbf{C} = \mathbf{0} \tag{96}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{D}} = \mathbf{V}^T\mathbf{A}_r\mathbf{V} - \mu\mathbf{V}^T\mathbf{A}_e\mathbf{V} - \nu\mathbf{D}^{-2} = \mathbf{0} \tag{97}$$

The expressions in (96) and (97) can be stated, respectively, as

$$(\mathbf{A}_r - \mu\mathbf{A}_e)\mathbf{V} = \mathbf{V}\mathbf{C}\mathbf{D}^{-1} \tag{98}$$

$$\mathbf{V}^T(\mathbf{A}_r - \mu\mathbf{A}_e)\mathbf{V} = \nu\mathbf{D}^{-2} \tag{99}$$

Since $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, (98) and (99) also imply that $\mathbf{C}\mathbf{D}^{-1} = \nu\mathbf{D}^{-2}$; that is,

$$\mathbf{C}\mathbf{D} = \nu\mathbf{I} \tag{100}$$

It is observed that a solution of (20) cannot satisfy $\mu = 0$ and $\nu = 0$ simultaneously (which would imply $\mathbf{V}^T\mathbf{A}_r\mathbf{V} = \mathbf{0}$ due to (99); i.e., $\mathbf{A}_r = \mathbf{0}$). Therefore, we can investigate the following two cases:

<u>**Case 1** ($\mu = 0$):</u>

In this case, $\nu > 0$; hence, (94) implies that $\mathrm{tr}\{\mathbf{D}^{-1}\} = P_\Sigma$. In addition, the stationarity conditions in (98)–(100) become $\mathbf{A}_r\mathbf{V} = \mathbf{V}\mathbf{C}\mathbf{D}^{-1}$ and $\mathbf{C}\mathbf{D} = \nu\mathbf{I}$. Since $\nu > 0$ and $\mathbf{C}$ is a diagonal matrix, it can be inferred from (100) that $\mathbf{D}$ must be a diagonal matrix.[7] Let $\mathbf{\Phi}$ be defined as $\mathbf{\Phi} \triangleq \mathbf{C}\mathbf{D}^{-1}$, which is also a diagonal matrix. Then, it is noted that $\mathbf{A}_r\mathbf{V} = \mathbf{V}\mathbf{\Phi}$ holds, meaning that the columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{A}_r$ and the diagonal elements of $\mathbf{\Phi}$ are the corresponding eigenvalues of $\mathbf{A}_r$. Since $\mathbf{A}_r$ is symmetric, the eigenvectors can be chosen to be orthonormal to satisfy (20d). Let $\mathbf{V}^*$ denote the solution of $\mathbf{A}_r\mathbf{V} = \mathbf{V}\mathbf{\Phi}$ such that the columns of $\mathbf{V}^*$ are orthonormal eigenvectors of $\mathbf{A}_r$. Also, let $\phi_1, \ldots, \phi_k$ denote the diagonal elements of $\mathbf{\Phi}$; i.e., the eigenvalues of $\mathbf{A}_r$; that is, $\mathbf{\Phi} = \mathrm{diag}\{\phi_1, \ldots, \phi_k\}$. Similarly, $\mathbf{C}$ and $\mathbf{D}$ can be expressed as $\mathbf{C} = \mathrm{diag}\{c_1, \ldots, c_k\}$ and $\mathbf{D} = \mathrm{diag}\{d_1, \ldots, d_k\}$. Then, we can summarize the relations in this case as follows:

$$\mathrm{tr}\{\mathbf{D}^{-1}\} = P_\Sigma \implies \sum_{j=1}^{k} \frac{1}{d_j} = P_\Sigma \tag{101}$$

$$\mathbf{C}\mathbf{D} = \nu\mathbf{I} \implies c_j d_j = \nu\,,\ j \in \{1, \ldots, k\} \tag{102}$$

$$\mathbf{\Phi} = \mathbf{C}\mathbf{D}^{-1} \implies \phi_j = \frac{c_j}{d_j}\,,\ j \in \{1, \ldots, k\} \tag{103}$$

By combining (102) and (103), we first obtain $1/d_j = \sqrt{\phi_j/\nu}$ for $j \in \{1, \ldots, k\}$. (Since $\nu > 0$ and $\mathbf{A}_r$ is assumed to be positive definite, the square roots of ϕ_j/ν exist.) Then, we utilize (101) and obtain $\sum_{j=1}^{k} \sqrt{\phi_j/\nu} = P_\Sigma$, which leads to $\sqrt{\nu} = (1/P_\Sigma)\sum_{j=1}^{k} \sqrt{\phi_j}$. Therefore, optimal d_j's are calculated as

$$d_j^* = \frac{\sum_{i=1}^{k} \sqrt{\phi_i}}{P_\Sigma \sqrt{\phi_j}} \tag{104}$$

for $j \in \{1, \ldots, k\}$. Hence, $\mathbf{D}^*$ in (21) of Proposition 2 is obtained.

Overall, the optimal solution of (20) in Case 1 is given by $\mathbf{V}^*$ and $\mathbf{D}^*$, where the columns of $\mathbf{V}^*$ are orthonormal eigenvectors of $\mathbf{A}_r$, and $\mathbf{D}^*$ is a diagonal matrix, the jth diagonal element of which is given by (104) with ϕ_j denoting the jth eigenvalue of $\mathbf{A}_r$. As long as $\mathrm{tr}\{(\mathbf{V}^*)^T\mathbf{A}_e\mathbf{V}^*\mathbf{D}^*\} \geq \eta$, $\mathbf{V}^*$ and $\mathbf{D}^*$ form a valid solution.

Case 2 ($\mu > 0$):

As $\mu > 0$ in this case, the complementary slackness condition in (93) results in $\mathrm{tr}\{\mathbf{V}^T\mathbf{A}_e\mathbf{V}\mathbf{D}\} = \eta$. Also, the stationarity condition in (99) can be expressed as

$$(\mathbf{A}_r - \mu\mathbf{A}_e)\mathbf{V} = \mathbf{V}\big(\nu\mathbf{D}^{-2}\big) \tag{105}$$

Since it is assumed that $\mathbf{A}_r$ is not a scaled version of $\mathbf{A}_e$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, ν cannot be zero in (105); hence, $\nu > 0$ is obtained. Therefore, (100) implies that $\mathbf{D}$ is a diagonal matrix. In addition, as $\nu > 0$, $\mathrm{tr}\{\mathbf{D}^{-1}\} = P_\Sigma$ due to (94). Moreover, (105) indicates that the columns of $\mathbf{V}$ must be orthonormal eigenvectors of $(\mathbf{A}_r - \mu\mathbf{A}_e)$ and the diagonal matrix $\nu\mathbf{D}^{-2}$ must contain the corresponding eigenvalues in its diagonals. For a given $\mu > 0$, let $\mathbf{V}_\mu$ denote a matrix with its columns being orthonormal eigenvectors of $(\mathbf{A}_r - \mu\mathbf{A}_e)$, and let $\mathbf{\Psi}_\mu \triangleq \nu\mathbf{D}^{-2}$ represent a diagonal matrix, the diagonal elements of which are the corresponding eigenvalues of $(\mathbf{A}_r - \mu\mathbf{A}_e)$. Then, by combining the conditions of $\mathrm{tr}\{\mathbf{V}^T\mathbf{A}_e\mathbf{V}\mathbf{D}\} = \eta$ and $\mathrm{tr}\{\mathbf{D}^{-1}\} = P_\Sigma$, the following relation is obtained:

$$P_\Sigma\,\eta = \mathrm{tr}\left\{\frac{1}{\sqrt{\nu}}(\mathbf{\Psi}_\mu)^{1/2}\right\}\mathrm{tr}\left\{\sqrt{\nu}\,\mathbf{V}_\mu^T\mathbf{A}_e\mathbf{V}_\mu(\mathbf{\Psi}_\mu)^{-1/2}\right\} \tag{106}$$

The expression in (106) can be shown to be equivalent to

$$P_\Sigma\,\eta = \mathrm{tr}\left\{(\mathbf{A}_r - \mu\mathbf{A}_e)^{1/2}\right\}\mathrm{tr}\left\{\mathbf{A}_e(\mathbf{A}_r - \mu\mathbf{A}_e)^{-1/2}\right\} \triangleq h(\mu) \tag{107}$$

To show that there exists a unique value of μ that satisfies (107), we first note that $h(0) \leq P_\Sigma\,\eta$ since Case 1 (with $\mu = 0$) would give the optimal solution otherwise. Also, it can be deduced from (106) that $h(\mu_{\min}) = \infty$, where $\mu_{\min}$ is as defined in Lemma 1. In addition, since $h(\mu)$ is a continuous and monotone increasing function of μ for $\mu \in [0, \mu_{\min})$ according to Lemma 1, it is concluded that there exists a unique solution of (107). Let $\mu^\star$ represent the unique value of μ that satisfies (107) for $\mu \in [0, \mu_{\min})$. (It is noted that μ cannot be larger than $\mu_{\min}$ due to (99).) Then, from $\mathbf{\Psi}_\mu = \nu\mathbf{D}^{-2}$ and $\mathrm{tr}\{\mathbf{D}^{-1}\} = P_\Sigma$, the solution of (20) can be specified as

$$\mathbf{D}^\star = \left(\frac{P_\Sigma(\mathbf{\Psi}_{\mu^\star})^{1/2}}{\mathrm{tr}\big\{(\mathbf{\Psi}_{\mu^\star})^{1/2}\big\}}\right)^{-1}. \tag{108}$$

Overall, the optimal solution of (20) in Case 2 is given by $\mathbf{V}^\star$ and $\mathbf{D}^\star$, where the columns of $\mathbf{V}^\star$ are orthonormal eigenvectors of $(\mathbf{A}_r - \mu^\star\mathbf{A}_e)$, and $\mathbf{D}^\star$ is given by (108) with $\mathbf{\Psi}_{\mu^\star}$ denoting the diagonal matrix consisting of the eigenvalues of $(\mathbf{A}_r - \mu^\star\mathbf{A}_e)$, where $\mu^\star$ is obtained by solving (107).[8]

Acknowledgement: The authors would like to thank Prof. Orhan Arikan and Prof. Tolga Mete Duman from Bilkent University for their insightful comments.

References

[1] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9, pp. 1727–1765, Sep. 2016.

[2] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. 22, no. 6, pp. 644–654, Nov. 1976.

[3] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, Feb. 1978.

[4] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. Di Renzo, "Safeguarding 5G wireless communication networks using physical layer security," vol. 53, no. 4, pp. 20–27, Apr. 2015.

[5] A. Mukherjee, "Physical-layer security in the Internet of Things: Sensing and communication confidentiality under resource constraints," *Proc. IEEE*, vol. 103, no. 10, pp. 1747–1761, Oct. 2015.

[6] Y. Wang, G. Attebury, and B. Ramamurthy, "A survey of security issues in wireless sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 8, no. 2, pp. 2–23, Apr.–June 2006.

[7] A. Yener and S. Ulukus, "Wireless physical-layer security: Lessons learned from information theory," *Proc. IEEE*, vol. 103, no. 10, pp. 1814–1825, Oct. 2015.

[8] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[7] A generic $\mathbf{D}$ is considered at the beginning of the proof but then the optimal $\mathbf{D}$ is shown to be a diagonal matrix, in compliance with the model.

[8] Although the problem in (20) is not convex, the KKT conditions become both necessary and sufficient for the minimizer since they lead to a unique structure and the problem admits a minimizer over the feasible region.

[9] Y. Liang, H. V. Poor, and S. Shamai, "Secure communication over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2470–2492, June 2008.

[10] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.

[11] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2453–2469, June 2008.

[12] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, July 1978.

[13] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[14] Y. Liang, A. Somekh-Baruch, H. V. Poor, S. Shamai, and S. Verdu, "Capacity of cognitive interference channels with and without secrecy," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 604–619, Feb. 2009.

[15] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[16] J. Zhu, J. Mo, and M. Tao, "Cooperative secret communication with artificial noise in symmetric interference channel," *IEEE Commun. Lett.*, vol. 14, no. 10, pp. 885–887, Oct. 2010.

[17] R. Bassily, E. Ekrem, X. He, E. Tekin, J. Xie, M. R. Bloch, S. Ulukus, and A. Yener, "Cooperative security at the physical layer: A summary of recent advances," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 16–28, Sep. 2013.

[18] Y.-W. P. Hong, P.-C. Lan, and C.-C. J. Kuo, "Enhancing physical-layer secrecy in multiantenna wireless systems: An overview of signal processing approaches," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 29–40, Sep. 2013.

[19] Y. Liu, Z. Su, and Y. Wang, "Artificial noise-assisted beamforming and power allocation for secure D2D-enabled V2V communications," in *IEEE 94th Vehicular Tech. Conf. (VTC 2021-Fall)*, 2021, pp. 1–5.

[20] D. Do, A. Le, and S. Mumtaz, "Secure performance analysis of RIS-aided wireless communication systems," in *IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6.

[21] H. Shen, W. Xu, and C. Zhao, "QoS constrained optimization for multi-antenna AF relaying with multiple eavesdroppers," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2224–2228, Dec. 2015.

[22] J. Yang, I.-M. Kim, and D. I. Kim, "Optimal cooperative jamming for multiuser broadcast channel with multiple eavesdroppers," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 6, pp. 2840–2852, June 2013.

[23] W.-C. Liao, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "QoS-based transmit beamforming in the presence of eavesdroppers: An optimized artificial-noise-aided approach," *IEEE Trans. on Signal Process.*, vol. 59, no. 3, pp. 1202–1216, March 2011.

[24] J. Guo, U. Rogers, X. Li, and H. Chen, "Secrecy constrained distributed detection in sensor networks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 2, pp. 378–391, June 2018.

[25] A. Ozcelikkale and T. M. Duman, "Cooperative precoding and artificial noise design for security over interference channels," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2234–2238, Dec. 2015.

[26] F. Farokhi and H. Sandberg, "Fisher information as a measure of privacy: Preserving privacy of households with smart meters using batteries," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4726–4734, Sep. 2018.

[27] X. Guo, A. S. Leong, and S. Dey, "Estimation in wireless sensor networks with security constraints," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 544–561, Apr. 2017.

[28] J. Guo, H. Chen, and U. Rogers, "Asymptotic perfect secrecy in distributed estimation for large sensor networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 3336–3340.

[29] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer-Verlag, 1994.

[30] J. Zhang, R. S. Blum, and H. V. Poor, "Approaches to secure inference in the Internet of Things: Performance bounds, algorithms, and effective attacks on IoT sensor networks," *IEEE Signal Proc. Mag.*, vol. 35, no. 5, pp. 50–63, Sep. 2018.

[31] C. Goken and S. Gezici, "ECRB-based optimal parameter encoding under secrecy constraints," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3556–3570, July 2018.

[32] ——, "Optimal parameter encoding based on worst case Fisher information under a secrecy constraint," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1611–1615, Nov. 2017.

[33] C. Goken, S. Gezici, and O. Arikan, "Estimation theoretic optimal encoding design for secure transmission of multiple parameters," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4302–4316, Aug 2019.

[34] D. Gurgunoglu, C. Goken, and S. Gezici, "Optimal power allocation for secure estimation of multiple parameters," *IEEE Signal Process. Lett.*, vol. 28, pp. 1784–1788, 2021.

[35] M. Shirazi and A. Vosoughi, "On Bayesian Fisher information maximization for distributed vector estimation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 4, pp. 628–645, Dec. 2019.

[36] I. Bahceci and A. K. Khandani, "Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation," *IEEE Transactions on Communications*, vol. 56, no. 7, pp. 1146–1156, July 2008.

[37] B. Kailkhura, V. S. S. Nadendla, and P. K. Varshney, "Distributed inference in the presence of eavesdroppers: A survey," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 40–46, June 2015.

[38] T. C. Aysal and K. E. Barner, "Sensor data cryptography in wireless sensor networks," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 2, pp. 273–289, June 2008.

[39] A. N. Samudrala and R. S. Blum, "Asymptotic analysis of a new low complexity encryption approach for Internet of Things, smart cities and smart grid," in *Proc. IEEE Int. Conf. on Smart Grid and Smart Cities (ICSGSCS)*, July 2017, pp. 200–204.

[40] ——, "On the estimation and secrecy capabilities of stochastic encryption for parameter estimation in IoT," in *52nd Annual Conf. on Inf. Sci. and Syst. (CISS)*, Mar. 2018, pp. 1–6.

[41] C. Goken and S. Gezici, "Estimation theoretic secure communication via encoder randomization," *IEEE Trans. on Signal Process.*, vol. 67, no. 23, pp. 6105–6120, 2019.

[42] A. F. Emery and A. V. Nenarokomov, "Optimal experiment design," *Measurement Science and Technology*, vol. 9, no. 6, pp. 864–876, June 1998.

[43] D. Gurgunoglu, B. Dulek, and S. Gezici, "Power adaptation for vector parameter estimation according to Fisher information based optimality criteria," *Signal Process.*, vol. 192, no. 108390, Mar. 2022.

[44] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1246–1250, May 1998.

[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. USA: Cambridge University Press, 2004.

[46] K. Dogancay, "Online optimization of receiver trajectories for scan-based emitter localization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 1117–1125, July 2007.

[47] T. C. Xygkis, G. N. Korres, and N. M. Manousakis, "Fisher information-based meter placement in distribution grids via the D-optimal experimental design," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1452–1461, March 2018.

[48] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, Inc., 1993.

[49] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[50] R. L. Burden and J. D. Faires, *Numerical Analysis*, 3rd ed. Boston, MA: Prindle, Weber & Schmidt, Oct. 1985.

[51] B. Ghojogh, F. Karray, and M. Crowley, "Eigenvalue and generalized eigenvalue problems: Tutorial," *ArXiv Preprint, arXiv:1903.11240*, Mar. 2019.