Observation-Driven Scheduling for Remote Estimation of Two Gaussian Random Variables

Marcos M. Vasconcelos and Urbashi Mitra, *Fellow, IEEE*

***Abstract*—Joint estimation and scheduling for sensor networks is considered in a system formed by two sensors, a scheduler, and a remote estimator. Each sensor observes distinct Gaussian random variables, which may be correlated. This system can be analyzed as a team decision problem with two agents: the scheduler and the remote estimator. The scheduler observes the output of both sensors and chooses which of the two is revealed to the remote estimator. The goal is to jointly design scheduling and estimation policies that minimize a mean-squared estimation error criterion. The person-by-person optimality of a policy pair called "max-scheduling/mean-estimation" is established, where the measurement with the largest absolute value is revealed to the estimator, which uses a corresponding conditional mean operator. This result is obtained for independent Gaussian random variables, and correlated Gaussian random variables with symmetric variances. Finally, the joint design of scheduling and linear estimation policies for any two Gaussian random variables with an arbitrary correlation structure is considered. In this case, the optimization problem is recast as a difference-of-convex program, and locally optimal solutions can be found using a simple numerical procedure.**

***Index Terms*—Decision theory, estimation, multiagent systems, networked control systems, optimization, quantization.**

I. Introduction

THE MULTIPLE components of cyber-physical systems are often interconnected by shared communication links of limited bandwidth [1]. One way to model this bandwidth constraint is to assume that, at any time instant, a single packet can be reliably transmitted over a communication link to its destination [2]. Therefore, the system designer must come up with rules/algorithms that allocate shared communication resources among multiple transmitting nodes. This paper introduces a new class of remote estimation problems, where the communication resources are allocated *dynamically* based on the observations at the sensors, rather than based purely on the statistical description of the information sources.

Manuscript received September 17, 2018; revised September 25, 2018 and December 28, 2018; accepted January 30, 2019. Date of publication February 21, 2019; date of current version March 18, 2020. This work was supported in part by the Office of Naval Research under Grant N00014-15-1-2550, in part by the National Science Foundation under Grant CNS-1213128, Grant CCF-1718560, Grant CCF-1410009, and Grant CPS-1446901, and in part by the Air Force Office of Scientific Research under Grant FA9550-12-1-0215. Recommended by Associate Editor S. Dey. *(Corresponding author: Marcos M. Vasconcelos.)*

The authors are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: mvasconc@usc.edu; ubli@usc.edu).

Digital Object Identifier 10.1109/TCNS.2019.2900864

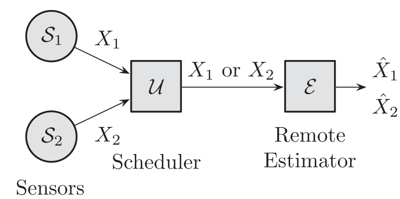

Fig. 1. Block diagram of the equivalent observation-driven sensor scheduling problem.

The basic framework considered here is shown in Fig. 1. Two sensors, making possibly correlated observations, report their measurements to a scheduler. The role of the scheduler is to select one of the observations and transmit it to a remote estimator. Finally, the remote estimator forms estimates of both measurements. Our goal is to jointly design scheduling and estimation policies that minimize a mean-squared estimation error criterion. Alternatively, this problem can be understood as one of dimensionality reduction, where an encoder–decoder pair is designed to minimize the expected distortion between the original and reconstructed vectors, with the constraint that a scalar (versus a vector) is transmitted or stored [3].

A. Motivation

One motivating application for the framework considered herein is in systems known as wireless body area networks [4]. In such systems, multiple biometric sensors are deployed on a person and report the collected measurements to an intermediate node, e.g., a mobile phone (scheduler), which, in turn, sends the data to a remote healthcare provider (estimator) over a wireless network. Due to the constraint imposed by the network, at a given time, the mobile phone can only send information coming from a single sensor at a time. The overall goal of the system is to enable real-time monitoring such that the remote healthcare provider can assist its patients with minimal delay.

Our framework can also be used to model communication constraints in other research fields such as organizational economics (see, e.g., [5]). Consider, for example, a firm where a manager (scheduler) has access to accurate information about different divisions of the company and reports to the company's chief executive officer (CEO) (estimator). Due to the CEO's limited and costly time, the manager cannot communicate everything and has to judiciously choose *what* to report (communication constraint). The CEO must then make a decision based on its belief (estimate) of the overall state of the firm.

B. Related Literature

The problem of selecting a subset among a larger set of sensors with the purpose of transmission over a bandwidth constrained network and subsequent estimation is generally referred to as *sensor scheduling*. With applications spanning many areas in engineering such as networked control, sensor networks, target tracking, and remote estimation [6]–[8], sensor scheduling has had a long and rich history. In general, sensor scheduling is a hard combinatorial optimization problem [9]. However, this computational complexity may be circumvented by suboptimal pruning of decision trees [10] or by using convex relaxations [11]. In certain cases, it is possible to show that the solution to these relaxed problems yields optimal scheduling schemes, which are periodic and therefore admit simple implementations [12]–[14]. In a related line of work, a framework for sensor selection with a cost function augmented with a sparsity promoting term is introduced. This framework has the goal of trading-off complexity versus performance [15].

Sending more important packets at the expense of others can greatly improve the performance of networked control systems [16]. Indeed, [16] was the first work to compare static (open-loop) versus dynamic (closed-loop) scheduling policies, where the sensing node with the maximum error is given priority to transmit over others. However, unlike [16], the results herein arise as person-by-person optimal solutions to an optimization problem, whereas in [16], the scheduling policy is specified *a priori* by the system designer. The problem of sequential transmission over a shared channel for discrete Markov sources was studied in [17]. Even though their motivation and problem formulation is similar to ours, their assumptions preclude signaling through decision variables. Moreover, they do not provide structural results on the optimal scheduling policies. The results obtained herein hold for a one-shot problem, but unlike [17], we do not avoid signaling, which is a central feature of our framework.

A sensor scheduling problem of multiple mutually independent linear time-invariant systems with an average cost criterion was considered in [18], where event-based policies must be designed such that only the most informative sensor is allowed to transmit at a time. Due to the inherent complexity of the problem, the optimal solution is difficult to obtain. A greedy scheduling scheme is then used to provide a quantifiable performance gap between the optimal solution and other suboptimal scheduling schemes. Unlike their work, we obtain a scheduling policy with a specific structure, which holds for independent and symmetrically correlated Gaussian observations. A more realistic scheduling problem for dynamical systems than the one addressed herein was studied in [19]; however, this generality comes at the expense of a lack of structural results, as observed in [19]. In order to obtain structural results, further simplifying assumptions are incorporated in [19]. Indeed, our work reinforces this point by showing that even in the one-shot Gaussian case, the resulting joint optimization problem is intractable. However, we are able to provide some additional insight on the design of scheduling and estimation (in the one-shot case) based on the concept of person-by-person optimality, which is not considered in [19].

Our approach to the scheduling problem is aligned with the work of Xia *et al.* [20], where the decision is made based on the realizations of the measurements themselves. The idea is to design and exploit *event-triggers* [21]–[23] for the transmission of one of the variables over the other, which allows for implicit communication via signaling [24]. In a way, the problem we address here is an *observation selection* problem [25], and the techniques we use are reminiscent of quantization in task-driven sensing, where the observation space is partitioned in regions where certain decisions are made [26]. Despite the simple description of our framework, the resulting optimization problem is nontrivial. The fundamental challenge is that the joint design of scheduling and estimation policies is entangled by *signaling*, and the transmitted variable serves as side information for estimation of the nontransmitted one. Since the action of the scheduler directly affects what the estimator observes, this is a *team-decision problem with a nonclassical information structure* [27]. Such problems are known to be NP-hard in general, and systematic methods for their solution are still unknown. Typically, this difficulty is either overlooked, or simplifying assumptions are made leading to suboptimal designs.

Finally, the problem studied in this paper is closely related to the problem of estimating random variables observed by individual sensors, which independently decide to transmit over a collision channel [2]. Here, the inclusion of a scheduler has the goal of completely avoiding collisions. In a sense, the problem considered here is a "centralized" version of the problem in [2] and can be used to lower bound the performance of the decentralized system.

C. Contributions and Organization

The main contributions of this paper are the following.

1) We establish the *person-by-person optimality* of the max-scheduling/mean-estimation policy pair for sensors making independent Gaussian observations. One remarkable feature of this result is that the structure of the scheduling policy is completely independent of the variances of the observations. The mean-estimation policy, in this case, is a piecewise linear function of the received packet at the remote estimator.
2) We establish the *person-by-person optimality* of the max-scheduling/soft-thresholding estimation policy pair in the case when the observations are *correlated* but have equal variances. In this case, the soft-thresholding policy is a nonlinear function of the received packet at the remote estimator. The proof of this result depends on certain symmetry and monotonicity properties related to the soft-thresholding nonlinear estimator induced by the max-scheduling policy.
3) For two Gaussian random variables with an arbitrary correlation structure, we provide a numerical procedure that finds locally optimal solutions to the nonconvex optimization problem obtained when the estimators are constrained to the class of piecewise linear functions.
4) Finally, we extend the person-by-person optimality result to account for any number of sensors observing independent zero-mean Gaussian sources.

Preliminary versions of Theorems 1 and 2 presented here have appeared previously in [28], where certain key technical aspects of the proofs were either conjectured or omitted. The proofs of the results reported here are detailed and precise. Additionally, we provide several new results, which have not appeared elsewhere in Theorems 3–7.

D. Notation

We adopt the following notation: random variables and random vectors are represented using upper case letters, such as X. Realizations of random variables and random vectors are represented by the corresponding lower case letter, such as x. The probability density function (pdf) of a continuous random variable X, provided that it is well defined, is denoted by f_X. Functions and functionals are denoted using calligraphic letters such as $\mathcal{F}$. We use $\mathcal{N}(m,\sigma^2)$ to represent the Gaussian probability distribution of mean m and variance σ^2, respectively. The n-dimensional Euclidean space is denoted by $\mathbb{R}^n$. Sets are represented in blackboard bold font, such as $\mathbb{A}$. The probability of an event $\mathfrak{E}$ is denoted by $\mathbf{P}(\mathfrak{E})$; the expectation of a random variable Z is denoted by $\mathbf{E}[Z]$. If two random variables X and Y are independent, this relationship is denoted by $X \perp\!\!\!\perp Y$. The indicator function of a statement $\mathfrak{S}$ is defined as follows:

$$\mathbf{1}(\mathfrak{S}) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } \mathfrak{S} \text{ is true} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

II. Problem Formulation

Consider the system in Fig. 1 comprised of two sensors labeled $\mathcal{S}_1$ and $\mathcal{S}_2$. Each sensor observes a Gaussian random variable with known mean and variance. Without loss of generality, we assume that sensor $\mathcal{S}_i$ observes X_i, where

$$X_i \sim \mathcal{N}(0,\sigma_i^2), \ \ i \in \{1,2\}. \tag{2}$$

The correlation coefficient between X_1 and X_2 is defined as

$$\rho \stackrel{\text{def}}{=} \frac{\mathbf{E}[X_1X_2]}{\sigma_1 \cdot \sigma_2}. \tag{3}$$

The observations X_1 and X_2 must be communicated to a remote estimator over a communication link, where a single packet is transmitted to the remote estimator at a time. A scheduler decides which of the random variables is chosen to be transmitted over the channel. The scheduler's decision variable, denoted by U, is computed according to a *scheduling policy*, which is a measurable function $\mathcal{U}:\mathbb{R}^2 \rightarrow \{1,2\}$ such that

$$U = \mathcal{U}(X_1,X_2). \tag{4}$$

The set of all admissible scheduling policies is denoted by $\mathbb{U}$. The scheduler's decision U determines what the remote estimator observes as follows:

$$Y = (U, X_U). \tag{5}$$

The vector Y belongs to the set $\mathbb{Y} \stackrel{\text{def}}{=} \{1,2\}\times\mathbb{R}$.

Remark 1: Notice that the scheduler effectively sends a packet containing the index U in addition to the real number X_U. The reason behind this assumption is to let the estimator know the origin of the packet before forming its estimates. The presence of an identification number on a packet is a standard assumption in data networks [29].

Upon observing Y, the remote estimator forms estimates of the observations at both sensors X_1 and X_2, denoted by $\hat{X}_1$ and $\hat{X}_2$, respectively. This is done according to an *estimation policy* $\mathcal{E}:\mathbb{Y}\rightarrow\mathbb{R}^2$ as follows:

$$(\hat{X}_1, \hat{X}_2) = \mathcal{E}(Y). \tag{6}$$

The set of all admissible estimation policies is denoted by $\mathbb{E}$.

Our goal is to solve the following optimization problem.

Problem 1: Given the variances $\sigma_1^2, \sigma_2^2 > 0$ and the correlation coefficient $\rho \in [0,1)$, find a scheduling and estimation policy pair $(\mathcal{U},\mathcal{E}) \in \mathbb{U}\times\mathbb{E}$ that jointly minimizes the following cost:

$$\mathcal{J}(\mathcal{U},\mathcal{E}) \stackrel{\text{def}}{=} \mathbf{E}\left[\left(X_1-\hat{X}_1\right)^2 + \left(X_2-\hat{X}_2\right)^2\right]. \tag{7}$$

A. Notions of Optimality

1) Global Optimality: A pair of scheduling and estimation strategies $(\mathcal{U}^\star,\mathcal{E}^\star)\in\mathbb{U}\times\mathbb{E}$ is *globally optimal* if

$$\mathcal{J}(\mathcal{U}^\star,\mathcal{E}^\star) \leq \mathcal{J}(\mathcal{U},\mathcal{E}), \ \ (\mathcal{U},\mathcal{E})\in\mathbb{U}\times\mathbb{E}. \tag{8}$$

2) Person-by-Person Optimality: A pair of scheduling and estimation strategies $(\mathcal{U}^\star,\mathcal{E}^\star)\in\mathbb{U}\times\mathbb{E}$ is *person-by-person optimal* if

$$\mathcal{J}(\mathcal{U}^\star,\mathcal{E}^\star) \leq \mathcal{J}(\mathcal{U},\mathcal{E}^\star), \ \ \mathcal{U}\in\mathbb{U} \tag{9}$$

$$\mathcal{J}(\mathcal{U}^\star,\mathcal{E}^\star) \leq \mathcal{J}(\mathcal{U}^\star,\mathcal{E}), \ \ \mathcal{E}\in\mathbb{E}. \tag{10}$$

Remark 2: The concept of person-by-person optimal solution is weaker than the notion of global optimality [27]. However, there are no known systematic approaches to solve Problem 1 for a globally optimal solution. Even the characterization of person-by-person optimal solutions in closed form for Problem 1 is a challenging task and must be done on a case-by-case basis.

III. Main Results

The main contribution of this work is to establish the person-by-person optimality of several pairs of scheduling and estimation policies for Problem 1 for different structures of correlation between the observations X_1 and X_2. Before formally stating the results, we first define the max-scheduling, mean-estimation, and soft-thresholding estimation policies.

Definition 1 (Max-scheduling policy): Let $x\in\mathbb{R}^2$. The max-scheduling policy is defined as

$$\mathcal{U}^{\max}(x) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } |x_1| \geq |x_2| \\ 2, & \text{otherwise.} \end{cases} \tag{11}$$

Definition 2 (Mean-estimation policy): Let $\xi\in\mathbb{R}$ and $i\in\{1,2\}$. The mean-estimation policy is defined as

$$\mathcal{E}^{\text{mean}}(i,\xi) = \begin{cases} [\xi \ \ 0]^\mathsf{T}, & \text{if } i=1 \\ [0 \ \ \xi]^\mathsf{T}, & \text{if } i=2. \end{cases} \tag{12}$$

The reason why the policy above is called *mean-estimation* is that the estimator outputs the *mean* of the unobserved random variable as its estimate. In other words, the side information provided by observing $X_i = \xi$ is irrelevant for estimating X_j, $i \neq j$. In this case, since the random variables X_1 and X_2 are assumed to be zero-mean, the mean-estimation policy takes the form above.

Definition 3 (Soft-thresholding estimation policy): Let $\xi \in \mathbb{R}$ and $i \in \{1, 2\}$. The soft-thresholding estimation policy is defined as

$$\mathcal{E}^{\text{soft}}(i, \xi) = \begin{cases} [\xi \;\; \eta(\xi)]^{\mathsf{T}}, & \text{if } i = 1 \\ [\eta(\xi) \;\; \xi]^{\mathsf{T}}, & \text{if } i = 2 \end{cases} \tag{13}$$

where $\eta(\xi)$ is a nonlinear soft-thresholding function with parameters $\sigma^2 > 0$ and $\rho \in [0, 1)$ defined as

$$\eta(\xi) \stackrel{\text{def}}{=} \frac{\int_{-|\xi|}^{|\xi|} \tau \exp\left(-\frac{(\tau-\rho\xi)^2}{2\sigma^2(1-\rho^2)}\right) d\tau}{\int_{-|\xi|}^{|\xi|} \exp\left(-\frac{(\tau-\rho\xi)^2}{2\sigma^2(1-\rho^2)}\right) d\tau}. \tag{14}$$

Our main results are stated in the following theorems.

Theorem 1: If $\rho = 0$, the policy pair $(\mathcal{U}^{\max}, \mathcal{E}^{\text{mean}})$ is a person-by-person optimal solution for the cost $\mathcal{J}(\mathcal{U}, \mathcal{E})$ in (7).

Theorem 2: If $\rho \in [0, 1)$ and $\sigma_1^2 = \sigma_2^2$, the policy pair $(\mathcal{U}^{\max}, \mathcal{E}^{\text{soft}})$ is a person-by-person optimal solution for the cost $\mathcal{J}(\mathcal{U}, \mathcal{E})$ in (7).

Remark 3: Theorems 1 and 2 present candidates for globally optimal scheduling and estimation policy pairs for Problem 1. We conjecture that these pairs are globally optimal. An alternative way to interpret this result is from the perspective from game theory, as Theorems 1 and 2 say that the pairs $(\mathcal{U}^{\max}, \mathcal{E}^{\text{mean}})$ and $(\mathcal{U}^{\max}, \mathcal{E}^{\text{soft}})$ constitute Nash-equilibrium solutions for a noncooperative game played by the scheduler and the estimator [30].

IV. Independent Observations

We start with the simpler case, where the sensors observe independent random variables. Let X_1 and X_2 be uncorrelated scalar Gaussian random variables, i.e., the correlation coefficient $\rho = 0$. We will now state two necessary optimality conditions reminiscent of quantization theory [31]. The first property pertains to the optimality of an optimal estimation policy for an arbitrarily fixed scheduling policy $\mathcal{U} \in \mathbb{U}$.

Lemma 1 (Optimal estimator): For a fixed scheduling policy $\mathcal{U} \in \mathbb{U}$, the estimation policy that minimizes the mean-squared error cost in (7) is the following:

$$\mathcal{E}^{\star}_{\mathcal{U}}(y) = \mathbf{E}\left[X \mid Y = y\right]. \tag{15}$$

Proof citation: This is the classical nonlinear filtering result. Its proof is found in many texts, such as [32, p. 143]. ■

Remark 4: There are two noteworthy facts about Lemma 1: 1) the optimal estimation policy is always a function of the scheduling policy. This coupling leads to the lack of convexity of Problem 1; and 2) the scheduling policy creates a coupling between the random variables X_1 and X_2 even when they are independent, which means that no matter what is received by the remote estimator, it should be used as side information for forming the optimal estimates $\hat{X}_1$ and $\hat{X}_2$.

Lemma 2 (Identity structure): The search for optimal estimation policies can be constrained to the set of policies $\mathcal{E}$ that satisfy the following identity property:

$$\mathcal{E}(1, \xi) = \begin{bmatrix} \xi \\ \eta_2(\xi) \end{bmatrix} \text{ and } \mathcal{E}(2, \xi) = \begin{bmatrix} \eta_1(\xi) \\ \xi \end{bmatrix} \tag{16}$$

where $\xi \in \mathbb{R}$ and $\eta_i : \mathbb{R} \to \mathbb{R}$, $i \in \{1, 2\}$.

Proof: Let $i, j \in \{1, 2\}$ such that $i \neq j$ and $\xi \in \mathbb{R}$; then, for any fixed scheduling policy $\mathcal{U} \in \mathbb{U}$, the event $\{Y = (i, \xi)\}$ is identical to the event $\{U = i, X_i = \xi\}$. Therefore, we have

$$\mathbf{E}[X_i \mid Y = (i, \xi)] = \xi. \tag{17}$$

Similarly, we have

$$\mathbf{E}[X_j \mid Y = (i, \xi)] = \int_{\mathbb{R}} x_j f_{X_j | U=i, X_i=\xi}(x_j) dx_j \tag{18}$$

$$\stackrel{\text{def}}{=} \eta_j(\xi). \tag{19}$$

■

For the remainder of this paper, every admissible estimator $\mathcal{E} \in \mathbb{E}$ satisfies the identity property in Lemma 2 and is completely specified by *representation functions* denoted by η_1 and η_2. This fact will be used to establish a necessary optimality condition for the optimal scheduling policy, for a given estimation policy $\mathcal{E} \in \mathbb{E}$.

Lemma 3 (Generalized nearest neighbor condition): For a fixed estimation policy $\mathcal{E} \in \mathbb{E}$ parameterized by representation functions η_1 and η_2, the following scheduling policy minimizes the cost in (7):

$$\mathcal{U}^{\star}_{\mathcal{E}}(x) = \begin{cases} 1, & \text{if } |x_1 - \eta_1(x_2)| \geq |x_2 - \eta_2(x_1)| \\ 2, & \text{otherwise.} \end{cases} \tag{20}$$

Proof: Using the law of total expectation, we write

$$\mathcal{J}(\mathcal{U}, \mathcal{E}) = \mathbf{E}\left[\|X - \hat{X}\|^2 \mid U = 1\right] \mathbf{P}(U = 1) + \mathbf{E}\left[\|X - \hat{X}\|^2 \mid U = 2\right] \mathbf{P}(U = 2). \tag{21}$$

Due to the identity structure in Lemma 2, the following holds:

$$\mathcal{J}(\mathcal{U}, \mathcal{E}) = \int_{\mathbb{R}^2} (x_2 - \eta_2(x_1))^2 \mathbf{1}(\mathcal{U}(x) = 1) f_X(x) dx + \int_{\mathbb{R}^2} (x_1 - \eta_1(x_2))^2 \mathbf{1}\left(\mathcal{U}(x) = 2\right) f_X(x) dx. \tag{22}$$

For fixed representation functions η_1 and η_2, we can construct a scheduling policy that minimizes the expression above. Let $\mathbb{Q}_i$ be defined as

$$\mathbb{Q}_i \stackrel{\text{def}}{=} \{x \in \mathbb{R}^2 \mid \mathcal{U}(x) = i\}, \quad i \in \{1, 2\}. \tag{23}$$

As a function of the sets $\mathbb{Q}_1$ and $\mathbb{Q}_2$, the cost in (22) becomes

$$\mathcal{J}(\mathcal{U}, \mathcal{E}) = \int_{\mathbb{Q}_1} (x_2 - \eta_2(x_1))^2 f_X(x) dx + \int_{\mathbb{Q}_2} (x_1 - \eta_1(x_2))^2 f_X(x) dx. \tag{24}$$

Since $\mathbb{Q}_1$ and $\mathbb{Q}_2$ form a partition of $\mathbb{R}^2$, we can equivalently write (24) as

$$\mathcal{J}(\mathcal{U},\mathcal{E}) = \int_{\mathbb{Q}_1} \left[(x_2-\eta_2(x_1))^2 - (x_1-\eta_1(x_2))^2\right] f_X(x)dx + \int_{\mathbb{R}} (x_1-\eta_1(x_2))^2 f_X(x)dx. \tag{25}$$

Finally, notice that in order to minimize (25), we assign to $\mathbb{Q}_1$ every point $x \in \mathbb{R}^2$ such that the argument of the integral is nonpositive, i.e., we assign to $\mathbb{Q}_1$ every point $x \in \mathbb{R}^2$ that satisfies the following inequality:

$$(x_2-\eta_2(x_1))^2 \leq (x_1-\eta_1(x_2))^2. \tag{26}$$

The remaining points are assigned to $\mathbb{Q}_2$. This minimizing choice is unique up to sets of measure zero. ∎

Remark 5: Notice that Lemma 3 is completely independent of the joint pdf f_X.

We are now equipped to prove Theorem 1.

Proof of Theorem 1: Let the estimation policy be $\mathcal{E} = \mathcal{E}^{\text{mean}}$. The associated representation functions are given by

$$\eta_i(\xi) = 0, \quad \xi \in \mathbb{R}, \quad i \in \{1,2\}. \tag{27}$$

From Lemma 3, an optimal scheduling policy for $\mathcal{E}^{\text{mean}}$ is

$$\mathcal{U}^\star_{\mathcal{E}^{\text{mean}}}(x) = \begin{cases} 1, & \text{if } |x_1| \geq |x_2| \\ 2, & \text{otherwise} \end{cases} \tag{28}$$

which is equal to the max-scheduling policy $\mathcal{U}^{\max}$.

Conversely, assume that the scheduling policy $\mathcal{U} = \mathcal{U}^{\max}$. Lemma 1 implies that the optimal estimator is given by

$$\mathcal{E}^\star_{\mathcal{U}^{\max}}(i,\xi) = \mathbf{E}[X \mid Y = (i,\xi)] \tag{29}$$

where $i \in \{1,2\}$ and $\xi \in \mathbb{R}$.

If $i = 1$, then

$$\mathbf{E}[X \mid Y = (1,\xi)] = \begin{bmatrix} \xi \\ \eta_2(\xi) \end{bmatrix} \tag{30}$$

where

$$\eta_2(\xi) = \frac{\int_{\mathbb{R}} x_2 \mathbf{1}\left(\mathcal{U}^{\max}(\xi,x_2)=1\right) f_{X_2|X_1=\xi}(x_2)dx_2}{\int_{\mathbb{R}} \mathbf{1}\left(\mathcal{U}^{\max}(\xi,x_2)=1\right) f_{X_2|X_1=\xi}(x_2)dx_2}. \tag{31}$$

Since $\rho = 0$, then the conditional pdf is equal to

$$f_{X_2|X_1=\xi}(x_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{x_2^2}{2\sigma_2^2}\right). \tag{32}$$

Therefore, the representation function η_2 can be explicitly computed as

$$\eta_2(\xi) = \frac{\int_{-|\xi|}^{|\xi|} x_2 \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{x_2^2}{2\sigma_2^2}\right) dx_2}{\int_{-|\xi|}^{|\xi|} \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{x_2^2}{2\sigma_2^2}\right) dx_2}. \tag{33}$$

Due to the even symmetry of the marginal Gaussian density around zero, we have

$$\eta_2(\xi) = 0, \quad \xi \in \mathbb{R}. \tag{34}$$

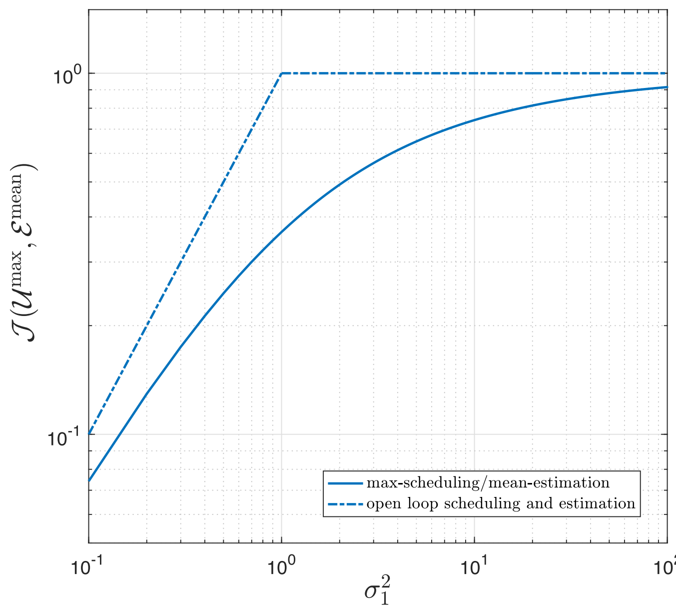

Fig. 2. Performance of the max-scheduling and mean-estimation policy computed as a function of the variance σ_1^2, while keeping $\sigma_2^2 = 1$ fixed. The dashed curve corresponds to the performance of the open-loop scheduling policy, where the variable with the largest variance is always transmitted to the remote estimator.

Repeating the same steps for $i = 2$ leads to

$$\eta_1(\xi) = 0, \quad \xi \in \mathbb{R}. \tag{35}$$

Therefore, we have

$$\mathcal{E}^\star_{\mathcal{U}^{\max}} = \mathcal{E}^{\text{mean}}. \tag{36}$$

∎

A. Illustrative Example

For two independent Gaussian observations $X_1 \sim \mathcal{N}(0,\sigma_1^2)$ and $X_2 \sim \mathcal{N}(0,\sigma_2^2)$, the performance of the person-by-person optimal pair of policies $(\mathcal{U}^{\max},\mathcal{E}^{\text{mean}})$ is given by the following formula:

$$\mathcal{J}(\mathcal{U}^{\max},\mathcal{E}^{\text{mean}}) = \mathbf{E}\left[\min\{X_1^2, X_2^2\}\right]. \tag{37}$$

Fig. 2 shows the performance of this pair of policies as a function of σ_1^2 while keeping $\sigma_2^2 = 1$.

We compare the performance of the pair of policies $(\mathcal{U}^{\max},\mathcal{E}^{\text{mean}})$ with an "open-loop" scheduling scheme, where the variable with the largest variance is transmitted, i.e.,

$$\mathcal{U}^{\text{open}}(x) \overset{\text{def}}{=} \arg\max_{i\in\{1,2\}} \sigma_i^2. \tag{38}$$

The optimal estimator for the policy $\mathcal{U}^{\text{open}}$ defined above is

$$\mathcal{E}^\star_{\mathcal{U}^{\text{open}}} = \mathcal{E}^{\text{mean}}. \tag{39}$$

Therefore, the performance of the open-loop scheme is given by the following expression:

$$\mathcal{J}(\mathcal{U}^{\text{open}},\mathcal{E}^{\text{mean}}) = \min\{\sigma_1^2,\sigma_2^2\}. \tag{40}$$

Fig. 2 shows the performance of the two schemes and the improvement achieved by the max-scheduling policy, which schedules the transmissions among sensors dynamically.

V. Symmetric Correlated Case

The two properties that enabled us with a simple proof for Theorem 1 were: 1) the fact that the two random variables X_1 and X_2 are independent; and 2) the fact that the conditional Gaussian pdfs are symmetric about the mean (which is zero in this case). In the correlated case, these two properties no longer hold.

We proceed by exploring the case when the variances are equal, but the observations are correlated, i.e., the covariance matrix is

$$\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{41}$$

In this case, for a fixed $\xi \in \mathbb{R}$, the conditional density of $X_i \mid X_j = \xi$ is

$$\mathcal{N}\left(\rho\xi, \sigma^2(1-\rho^2)\right). \tag{42}$$

Let us define the optimal nonlinear representation functions induced by the max-scheduling policy when the observations are symmetrically correlated. Let $i, j \in \{1,2\}$ such that $i \neq j$ and $\xi \in \mathbb{R}$. Then, under the max-scheduling policy, we have

$$\mathbf{E}\left[X_i \mid Y = (j,\xi)\right] = \frac{\int_{-|\xi|}^{|\xi|} x_i f_{X_i|X_j=\xi}(x_i)dx_i}{\int_{-|\xi|}^{|\xi|} f_{X_i|X_j=\xi}(x_i)dx_i}. \tag{43}$$

Notice that, due to the symmetric variances, the two nonlinear estimates corresponding to $i = 1, 2$ given by the expression above are equal. This leads to the nonlinear soft-thresholding representation function $\eta(\xi)$ defined in (14). The representation function $\eta(\xi)$ is shown in Fig. 3. It is straightforward to show that η has odd symmetry. We state this fact without proof as a lemma.

Lemma 4 (Odd symmetry of the nonlinear soft-thresholding representation function): The function η defined in (14) satisfies

$$\eta(-\xi) = -\eta(\xi), \quad \xi \in \mathbb{R}. \tag{44}$$

In the proof of Theorem 2, we will make extensive use of two auxiliary functions, $\mathcal{P}$ and $\mathcal{T}$.

Definition 4 (Auxiliary functions): Let $\mathcal{P}$ and $\mathcal{T}$ be defined as follows:

$$\mathcal{P}(\xi) \overset{\text{def}}{=} \xi - \eta(\xi), \quad \xi \in \mathbb{R} \tag{45}$$

and

$$\mathcal{T}(\xi) \overset{\text{def}}{=} \xi + \eta(\xi), \quad \xi \in \mathbb{R}. \tag{46}$$

Lemma 5 (Monotonicity of $\mathcal{P}$ and $\mathcal{T}$): Let $\xi_1, \xi_2 \in \mathbb{R}$. For all ξ_1 and ξ_2 such that $\xi_1 \leq \xi_2$, we have

$$\mathcal{P}(\xi_1) \leq \mathcal{P}(\xi_2) \tag{47}$$

$$\mathcal{T}(\xi_1) \leq \mathcal{T}(\xi_2). \tag{48}$$

Proof: See Appendix A. ∎

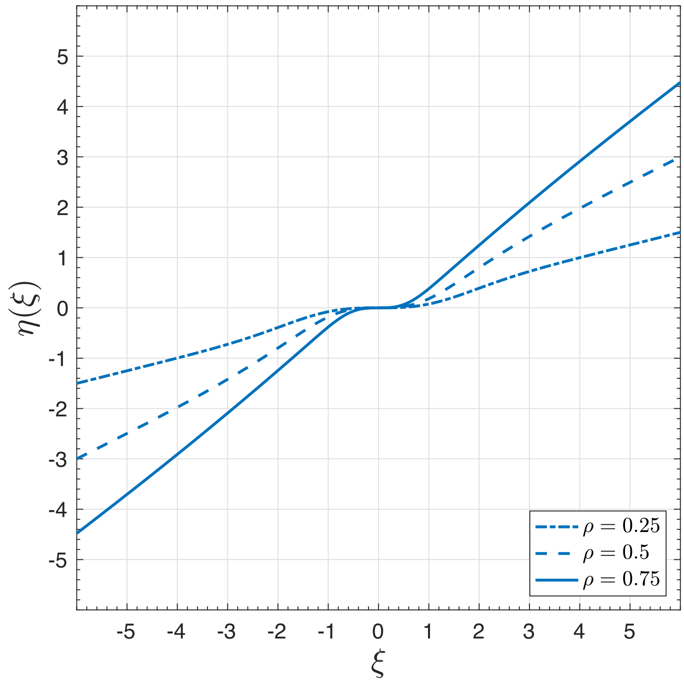

Fig. 3. Nonlinear soft-thresholding estimator induced by the max-scheduling policy for symmetric Gaussian sources. In this figure, the variance is normalized to 1.

The fact that $\mathcal{P}$ and $\mathcal{T}$ satisfy this monotonicity property is of paramount importance in the proof of Theorem 2. We are now equipped to prove Theorem 2.

Proof of Theorem 2: Assuming that $\mathcal{U} = \mathcal{U}^{\max}$, due to the symmetry of the pdf, Lemma 2 implies that the optimal estimator $\mathcal{E}^\star_{\mathcal{U}^{\max}}$ is given by

$$\mathcal{E}^\star_{\mathcal{U}^{\max}}(1,\xi) = \begin{bmatrix} \xi \\ \eta(\xi) \end{bmatrix} \text{ and } \mathcal{E}^\star_{\mathcal{U}^{\max}}(2,\xi) = \begin{bmatrix} \eta(\xi) \\ \xi \end{bmatrix} \tag{49}$$

where the representation function $\eta(\xi)$ is defined in (14). Therefore, we have

$$\mathcal{E}^\star_{\mathcal{U}^{\max}} = \mathcal{E}^{\text{soft}}. \tag{50}$$

We will show that this choice of estimation policy implies, via Lemma 3, the optimality of the max-scheduling policy. Define the function $\mathcal{H} : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathcal{H}(x) \overset{\text{def}}{=} \left(x_2 - \eta(x_1)\right)^2 - \left(x_1 - \eta(x_2)\right)^2. \tag{51}$$

The above function can be rewritten using the two auxiliary functions $\mathcal{P}$ and $\mathcal{T}$ from (45) and (46) as follows:

$$\mathcal{H}(x) = \left[\mathcal{T}(x_2) - \mathcal{T}(x_1)\right] \times \left[\mathcal{P}(x_2) + \mathcal{P}(x_1)\right]. \tag{52}$$

Lemma 3 implies the optimal scheduling policy given by

$$\mathcal{U}^\star_{\mathcal{E}^\star_{\mathcal{U}^{\max}}}(x) = \begin{cases} 1, & \text{if } \mathcal{H}(x) \leq 0 \\ 2, & \text{otherwise.} \end{cases} \tag{53}$$

Partition $\mathbb{R}^2$ into eight subsets $\{\mathbb{A}_1, \ldots, \mathbb{A}_8\}$ depicted in Fig. 4. Let $x \in \mathbb{A}_1$, which is characterized by $x_1 \geq 0$, $x_2 \geq 0$ and

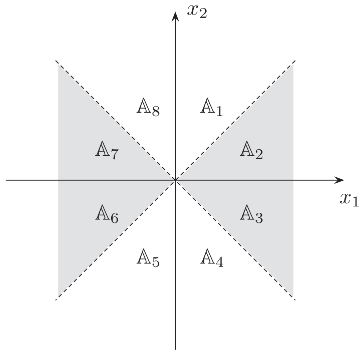

Fig. 4. Partition of the observation space used in the proof of Theorem 2.

$x_1 \le x_2$. Lemma 5 implies that

$$\mathcal{P}(x_i) \ge \mathcal{P}(0) = 0, \quad i \in \{1,2\}. \tag{54}$$

Therefore, we have

$$\mathcal{P}(x_1) + \mathcal{P}(x_2) \ge 0. \tag{55}$$

Since $x_1 \le x_2$, Lemma 5 also implies that

$$\mathcal{T}(x_2) - \mathcal{T}(x_1) \ge 0. \tag{56}$$

Together, (55) and (56) imply

$$\mathcal{H}(x) \ge 0. \tag{57}$$

Similarly, if $x \in \mathbb{A}_2$, we have

$$\mathcal{P}(x_1) + \mathcal{P}(x_2) \ge 0. \tag{58}$$

On the other hand, we have

$$\mathcal{T}(x_2) - \mathcal{T}(x_1) \le 0. \tag{59}$$

Therefore, we have

$$\mathcal{H}(x) \le 0. \tag{60}$$

Proceeding in a similar way, making use of Lemmas 4 and 5, we can cover all the eight regions. For each of the eight regions, all the points either satisfy $\mathcal{H}(x) \ge 0$ or $\mathcal{H}(x) \le 0$. Moreover, if $x \in \mathbb{A}_1 \cup \mathbb{A}_4 \cup \mathbb{A}_5 \cup \mathbb{A}_8$, then $\mathcal{H}(x) \ge 0$, and if $x \in \mathbb{A}_2 \cup \mathbb{A}_3 \cup \mathbb{A}_6 \cup \mathbb{A}_7$, then $\mathcal{H}(x) \le 0$. Thus, we have

$$\mathcal{U}^\star_{\mathcal{E}^{\text{soft}}} = \mathcal{U}^{\max}. \tag{61}$$

■

Remark 6: There exists a connection between Theorems 1 and 2, although neither one implies the other in general. When $\rho = 0$ in Theorem 2, the result follows immediately from Theorem 1. Similarly, when $\sigma_1^2 = \sigma_2^2$ as in Theorem 1, the result follows from Theorem 2. This particular case of Theorem 1 can be understood as a Corollary to Theorem 2, and vice versa.

VI. Decorrelating Transformation Approach

In this section, we propose a person-by-person optimal solution to the scheduling of two arbitrarily correlated Gaussian random variables by using pre- and postprocessing blocks on the observations and the estimates. The idea is to "decorrelate" the two observations using an invertible linear transformation, use the max-scheduling/mean-estimation policies on the transformed random variables, and then to "correlate" the estimates using the inverse transformation. This strategy is depicted in the block diagram of Fig. 5.

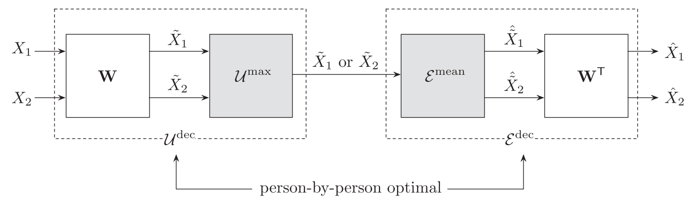

Fig. 5. System architecture for the arbitrarily correlated case. The preprocessing block implements a decorrelating linear transformation $\mathbf{W}$ obtained from the eigendecomposition of the covariance matrix $\mathbf{\Sigma}$. The postprocessing block implements the inverse of the decorrelating transformation $\mathbf{W}^\mathsf{T}$.

Consider the eigendecomposition of the covariance matrix $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^\mathsf{T} \tag{62}$$

where $\mathbf{W}\mathbf{W}^\mathsf{T} = \mathbf{I}$, and $\mathbf{\Lambda}$ is a diagonal matrix. Using the matrix $\mathbf{W}$, define the following scheduling and estimation policies:

$$\mathcal{U}^{\text{dec}}(x) \overset{\text{def}}{=} \mathcal{U}^{\max}\left(\mathbf{W}x\right), \quad x \in \mathbb{R}^2 \tag{63}$$

and

$$\mathcal{E}^{\text{dec}}(i,\xi) \overset{\text{def}}{=} \mathbf{W}^\mathsf{T}\mathcal{E}^{\text{mean}}(i,\xi), \quad i \in \{1,2\}, \quad \xi \in \mathbb{R}. \tag{64}$$

Theorem 3: Let $X \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a symmetric positive-definite covariance matrix. The pair $(\mathcal{U}^{\text{dec}}, \mathcal{E}^{\text{dec}})$ is a person-by-person optimal solution to Problem 1.

Proof: Let $\mathbf{W}$ be computed from the eigendecomposition of $\mathbf{\Sigma}$, and denote

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}. \tag{65}$$

Let $\tilde{x} \in \mathbb{R}^2$ be defined as $\tilde{x} \overset{\text{def}}{=} \mathbf{W}x$. Assuming that the estimator uses policy $\mathcal{E}^{\text{dec}}$, the optimal scheduling decision rule is to transmit $\tilde{x}_1$ if

$$(x_1 - w_{11}\tilde{x}_1)^2 + (x_2 - w_{12}\tilde{x}_1)^2 \le (x_1 - w_{21}\tilde{x}_2)^2 + (x_2 - w_{22}\tilde{x}_2)^2. \tag{66}$$

Recalling that $x = \mathbf{W}^\mathsf{T}\tilde{x}$, (66) is equivalent to

$$(w_{21}^2 + w_{22}^2)\tilde{x}_2^2 \le (w_{11}^2 + w_{12}^2)\tilde{x}_1^2. \tag{67}$$

Since $\mathbf{W}$ is a unitary matrix, the above inequality is equivalent to

$$|\tilde{x}_2| \le |\tilde{x}_1|. \tag{68}$$

Therefore, we have

$$\mathcal{U}^\star_{\mathcal{E}^{\text{dec}}} = \mathcal{U}^{\text{dec}}. \tag{69}$$

Conversely, assume that the scheduler uses $\mathcal{U}^{\text{dec}}$. Let $i, j \in \{1,2\}$ such that $i \neq j$. Then, we have

$$\mathcal{E}^\star_{\mathcal{U}^{\text{dec}}}(i,\tilde{x}_i) = \mathbf{W}^\mathsf{T}\mathbf{E}\big[\tilde{X} \mid Y = (i,\tilde{x}_i)\big] \tag{70}$$

where

$$\tilde{X} \overset{\text{def}}{=} \mathbf{W}X. \tag{71}$$

Computing the above expectation gives

$$\mathbf{E}\big[\tilde{X}_i \mid Y = (i, \tilde{x}_i)\big] = \tilde{x}_i \tag{72}$$

and, for $j \neq i$, we have

$$\mathbf{E}\big[\tilde{X}_j \mid Y = (i, \tilde{x}_i)\big] = \frac{\int_{-|\tilde{x}_i|}^{|\tilde{x}_i|} \tilde{x}_j f_{\tilde{X}_j|\tilde{X}_i=\tilde{x}_i}(\tilde{x}_j) d\tilde{x}_j}{\int_{-|\tilde{x}_i|}^{|\tilde{x}_i|} f_{\tilde{X}_j|\tilde{X}_i=\tilde{x}_i}(\tilde{x}_j) d\tilde{x}_j}. \tag{73}$$

Since $\tilde{X}_i \perp\!\!\!\perp \tilde{X}_j$, and $f_{\tilde{X}_j}$ is an even function, we have

$$\mathbf{E}\big[\tilde{X}_j \mid Y = (i, \tilde{x}_i)\big] = 0, \;\; \tilde{x}_i \in \mathbb{R}. \tag{74}$$

Therefore, we have

$$\mathcal{E}^\star_{\mathcal{U}^{\text{dec}}} = \mathcal{E}^{\text{dec}}. \tag{75}$$

■

Remark 7: Despite the fact that $(\mathcal{U}^{\text{dec}}, \mathcal{E}^{\text{dec}})$ is person-by-person optimal for Problem 1, we will show later that this is a suboptimal solution in general. For example, for a symmetrically correlated Gaussian random variables, the pair $(\mathcal{U}^{\max}, \mathcal{E}^{\text{soft}})$ yields a smaller cost, albeit the difference in performance is not large. We conjecture that, in general, the globally optimal estimation policy is nonlinear. However, Theorem 3 is a useful result because it leads to person-by-person optimal policies for systems with n sensors measuring correlated Gaussian random variables, as we will formally state in Section IX.

VII. Normalization Approach

In view of Theorem 3, one may be inclined to think that another possible person-by-person optimal solution to Problem 1 can be constructed for the general correlated case by means of the *normalization approach*. The normalization approach consists of applying a linear transformation to the random variables X_1 and X_2 such that they are symmetrically correlated with unit variances.

Define the following scheduling and estimation policies:

$$\mathcal{U}^{\text{norm}}(x) \overset{\text{def}}{=} \mathcal{U}^{\max}(\mathbf{N}x), \;\; x \in \mathbb{R}^2 \tag{76}$$

where

$$\mathbf{N} \overset{\text{def}}{=} \begin{bmatrix} 1/\sigma_1 & 0 \\ 0 & 1/\sigma_2 \end{bmatrix} \tag{77}$$

and

$$\mathcal{E}^{\text{norm}}(i, \xi) \overset{\text{def}}{=} \mathbf{N}^{-1} \cdot \mathcal{E}^{\text{soft}}(i, \xi), \;\; i \in \{1, 2\}, \;\; \xi \in \mathbb{R}. \tag{78}$$

Notice that when using this scheme, what is transmitted to the estimator over the channel are the normalized random variables, i.e., $\tilde{X}_1$ or $\tilde{X}_2$, where

$$\tilde{X}_i \overset{\text{def}}{=} \frac{X_i}{\sigma_i}, \;\; i \in \{1, 2\}. \tag{79}$$

Theorem 4: The policy pair $(\mathcal{U}^{\text{norm}}, \mathcal{E}^{\text{norm}})$ is not person-by-person optimal for Problem 1.

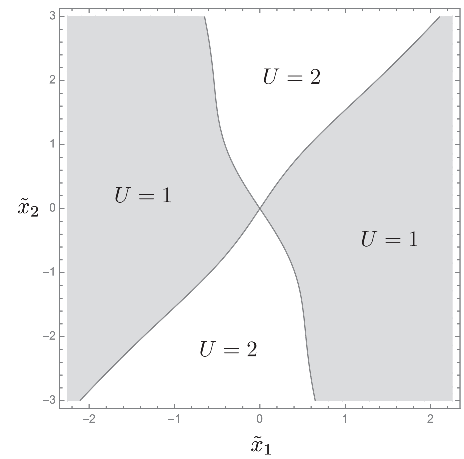

Fig. 6. Scheduling decision region induced by the normalization approach.

Proof: Assume by contradiction that $(\mathcal{U}^{\text{norm}}, \mathcal{E}^{\text{norm}})$ is person-by-person optimal.

Assuming that $\mathcal{U} = \mathcal{U}^{\text{norm}}$, it is straightforward to show that

$$\mathcal{E}^\star_{\mathcal{U}^{\text{norm}}}(1, \tilde{x}_1) = \begin{bmatrix} \sigma_1 \cdot \tilde{x}_1 \\ \sigma_2 \cdot \eta^{\text{soft}}(\tilde{x}_1) \end{bmatrix} \tag{80}$$

and

$$\mathcal{E}^\star_{\mathcal{U}^{\text{norm}}}(2, \tilde{x}_2) = \begin{bmatrix} \sigma_1 \cdot \eta^{\text{soft}}(\tilde{x}_2) \\ \sigma_2 \cdot \tilde{x}_2 \end{bmatrix}. \tag{81}$$

Therefore, we have

$$\mathcal{E}^\star_{\mathcal{U}^{\text{norm}}} = \mathcal{E}^{\text{norm}}. \tag{82}$$

However, we will now show that $\mathcal{U}^\star_{\mathcal{E}^{\text{norm}}} \neq \mathcal{U}^{\text{norm}}$. Assuming that the estimator uses policy $\mathcal{E}^{\text{norm}}$, then the optimal scheduling decision rule is to transmit $\tilde{x}_1$ if

$$\big|x_2 - \sigma_2 \cdot \eta^{\text{soft}}(\tilde{x}_1)\big| \leq \big|x_1 - \sigma_1 \cdot \eta^{\text{soft}}(\tilde{x}_2)\big| \tag{83}$$

which is equivalent to

$$\sigma_2 \cdot \big|\tilde{x}_2 - \eta^{\text{soft}}(\tilde{x}_1)\big| \leq \sigma_1 \cdot \big|\tilde{x}_1 - \eta^{\text{soft}}(\tilde{x}_2)\big|. \tag{84}$$

Consider an instance of Problem 1 with $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, and $\rho = 0.3$. The decision region induced by (84) for this set of parameters must be obtained numerically and is shown in Fig. 6. As we can readily see, the decision region is characterized by two nonlinear curves and differs significantly from the desired scheduling decision region defined by the inequality $|\tilde{x}_1| \geq |\tilde{x}_2|$. Therefore, we have

$$\mathcal{U}^\star_{\mathcal{E}^{\text{norm}}} \neq \mathcal{U}^{\text{norm}} \tag{85}$$

which implies that $(\mathcal{U}^{\text{norm}}, \mathcal{E}^{\text{norm}})$ is not person-by-person optimal for Problem 1. ■

VIII. Linear MMSE Estimators

In this section, we will consider the design of jointly optimal scheduling and estimation policies when the estimation policies are constrained to belong to the parameterizable class of *piecewise linear* estimation policies.

Definition 5 (Class of admissible piecewise linear estimation policies): Let $a \in \mathbb{R}^2$, $\xi \in \mathbb{R}$, and $i \in \{1, 2\}$. An admissible estimation policy $\mathcal{E}_a^{\text{linear}} \in \mathbb{E}$ is piecewise linear if it has the following structure:

$$\mathcal{E}_a^{\text{linear}}(i,\xi) = \begin{cases} [\xi \quad a_2\xi]^\mathsf{T}, & \text{if } i=1 \\ [a_1\xi \quad \xi]^\mathsf{T}, & \text{if } i=2. \end{cases} \tag{86}$$

The set of all admissible piecewise linear estimation policies is denoted by $\mathbb{E}^{\text{linear}}$.

A. Linear MMSE Estimation of Symmetrically Correlated Gaussian Random Variables

Within the class of piecewise linear estimators, Problem 1 admits a unique solution when the random variables have equal variance. Before stating this result in Theorem 5, we show that the search for linear MMSE estimators can be performed by solving a finite-dimensional optimization problem.

Proposition 1: Consider Problem 1 with the additional constraint that $\mathcal{E} \in \mathbb{E}^{\text{linear}}$. Then, the problem is equivalent to the following finite-dimensional nonconvex optimization problem:

$$\underset{a\in\mathbb{R}^2}{\text{minimize}} \quad \mathcal{J}_q(a) \tag{87}$$

where the objective function $\mathcal{J}_q : \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$\mathcal{J}_q(a) \stackrel{\text{def}}{=} \mathbf{E}\left[\min\left\{(X_1 - a_1X_2)^2, (X_2 - a_2X_1)^2\right\}\right]. \tag{88}$$

Proof: Recall the cost functional

$$\mathcal{J}(\mathcal{U},\mathcal{E}) = \mathbf{E}\left[(X_1 - \hat{X}_1)^2 + (X_2 - \hat{X}_2)^2\right]. \tag{89}$$

If $\mathcal{E} \in \mathbb{E}^{\text{linear}}$, then the cost can be rewritten in integral form as

$$\mathcal{J}(\mathcal{U},\mathcal{E}) = \int_{\mathbb{R}^2} (x_1 - a_1x_2)^2\, \mathbf{1}(\mathcal{U}(x)=2) f_X(x)dx + \int_{\mathbb{R}^2} (x_2 - a_2x_1)^2 \mathbf{1}(\mathcal{U}(x)=1) f_X(x)dx. \tag{90}$$

For an arbitrarily fixed $a \in \mathbb{R}^2$, the optimal scheduling policy $\mathcal{U}_a^\star$ is given by

$$\mathcal{U}_a^\star(x) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } (x_2 - a_2x_1)^2 \le (x_1 - a_1x_2)^2 \\ 2, & \text{otherwise.} \end{cases} \tag{91}$$

Therefore, we may, without loss of optimality, define a new cost solely in terms of $a \in \mathbb{R}^2$ as follows:

$$\mathcal{J}_q(a) \stackrel{\text{def}}{=} \mathcal{J}(\mathcal{U}_a^\star, \mathcal{E}) \tag{92}$$

which is equal to the expression in (88). ■

Theorem 5: Consider two symmetrically correlated Gaussian random variables with variance σ^2 and correlation coefficient $\rho \in [0, 1)$. Constraining the estimator to belong to the class of piecewise linear functions, the policy pair $(\mathcal{U}^{\max}, \mathcal{E}_{a^\star}^{\text{linear}})$ is globally optimal for Problem 1, where

$$a^\star = \frac{\rho \cdot \sigma^2}{2 \cdot \int_{\mathbb{R}^2} x_1^2 \mathbf{1}(|x_1| \ge |x_2|) f_X(x)dx} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{93}$$

Proof: See Appendix B. ■

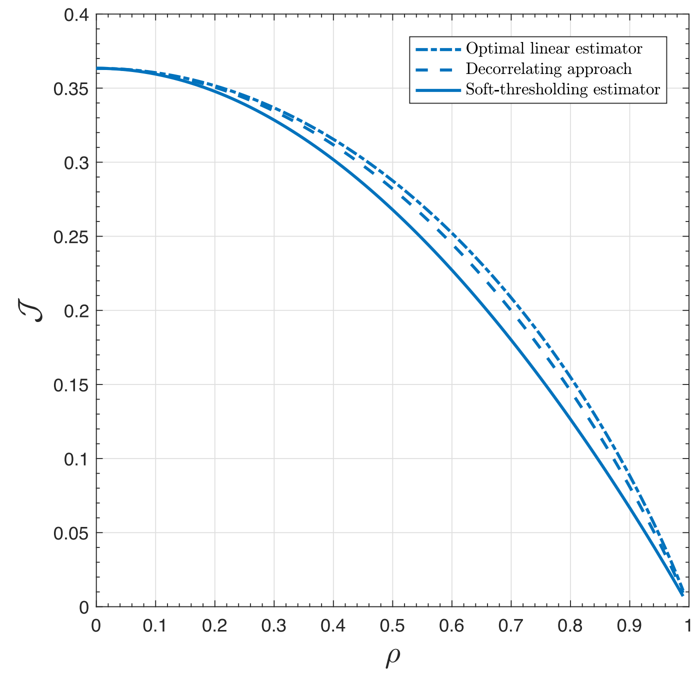

Fig. 7. Performance of three different scheduling and estimation schemes for symmetric correlated Gaussian random variables with variance $\sigma^2 = 1$.

Example 1: The performance of the scheduling/estimation schemes of Sections V, VI, and VIII-A is displayed in Fig. 7 for symmetrically correlated Gaussian random variables with variance $\sigma^2 = 1$. The system implementing max-scheduling and nonlinear soft-thresholding estimation of Theorem 2 has the best performance. We conjecture that $(\mathcal{U}^{\max}, \mathcal{E}^{\text{soft}})$ is indeed the globally optimal solution in this case. The performance of the strategy based on the decorrelating transformation approach has the second best performance, and as we can see, it shows that a person-by-person optimal solution is not necessarily globally optimal. Finally, the worst performance is of max-scheduling followed by the optimal linear estimator of Theorem 5. Despite being suboptimal for Problem 1, this is a globally optimal solution among the class of piecewise linear estimators.

We also compare the same scheduling/estimation schemes by evaluating their performances for symmetrically correlated Gaussian random variables with a fixed correlation coefficient $\rho = 0.6$ as a function of the variance σ^2. Fig. 8 shows that the scheme based on the decorrelating transformation approach performs slightly better than the scheme based on the optimal linear estimator. The scheme based on max-scheduling and soft-thresholding estimation outperforms both schemes for all values of σ^2. Finally, we observe that the performance of the three schemes scale linearly with σ^2 for every $\rho \in [0, 1)$. This fact can be established analytically, and its proof is omitted for brevity.

B. Optimization via the Convex–Concave Procedure

Notice that the equivalent optimization problem stated in Proposition 1, although finite dimensional, is a nonconvex stochastic program. Unlike the symmetric case, a closed-form solution to this problem for the general case is not known.

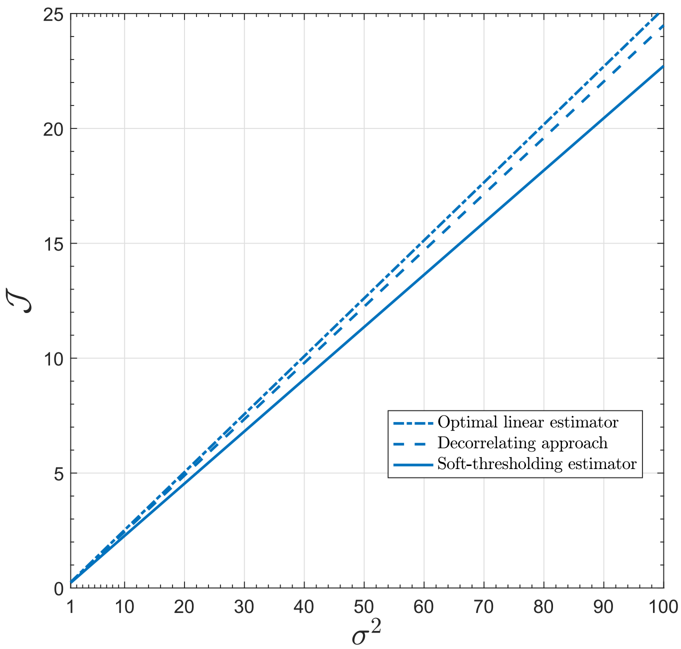

Fig. 8. Performance of three different scheduling and estimation schemes for symmetrically correlated Gaussian random variables with correlation coefficient $\rho = 0.6$.

However, in our subsequent analysis, we will decompose the cost into a difference-of-convex functions and derive a numerical optimization algorithm to compute locally optimal solutions using the *convex–concave procedure* (CCP) [33].

Define the functions $\mathcal{F}, \mathcal{G}\colon \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathcal{F}(a) \overset{\text{def}}{=} (1 + a_2^2)\sigma_1^2 + (1 + a_1^2)\sigma_2^2 - 2\rho\sigma_1\sigma_2(a_1 + a_2) \quad (94)$$

and

$$\mathcal{G}(a) \overset{\text{def}}{=} \mathbf{E}\left[\max\left\{(X_1 - a_1X_2)^2, (X_2 - a_2X_1)^2\right\}\right]. \quad (95)$$

The cost function in (88) can be expressed as a difference of convex functions as follows:

$$\mathcal{J}_q(a) = \mathcal{F}(a) - \mathcal{G}(a). \quad (96)$$

The CCP for minimizing $\mathcal{J}_q$ is defined as

$$a^{(k+1)} = \arg\min_{a \in \mathbb{R}^2}\left\{\mathcal{F}(a) - \mathcal{G}_{\text{affine}}(a; a^{(k)})\right\} \quad (97)$$

where

$$\mathcal{G}_{\text{affine}}\big(a; a^{(k)}\big) \overset{\text{def}}{=} \mathcal{G}\big(a^{(k)}\big) + \big(a^{(k)}\big)^{\mathsf{T}}\big(a - a^{(k)}\big) \quad (98)$$

and $g(a)$ is a subgradient of $\mathcal{G}(a)$. We solve the optimization problem in (97), by using the unique solution of the first-order optimality condition

$$\nabla\mathcal{F}(a^\star) - g\big(a^{(k)}\big) = 0. \quad (99)$$

Since the function $\mathcal{F}$ is differentiable in both of its arguments, its gradient can be explicitly computed as

$$\nabla\mathcal{F}(a) = \begin{bmatrix} 2a_1\sigma_2^2 - 2\rho\sigma_1\sigma_2 \\ 2a_2\sigma_1^2 - 2\rho\sigma_1\sigma_2 \end{bmatrix} \quad (100)$$

which leads to the following dynamical system:

$$\begin{bmatrix} a_1^{(k+1)} \\ a_2^{(k+1)} \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sigma_2^2} & 0 \\ 0 & \frac{1}{2\sigma_1^2} \end{bmatrix} g(a^{(k)}) + \begin{bmatrix} \rho\sigma_1/\sigma_2 \\ \rho\sigma_2/\sigma_1 \end{bmatrix}. \quad (101)$$

The sequence $\{a_k\}_{k=1}^{\infty}$ defined by the above system always converges to a critical point of $\mathcal{J}_q$ [34]. In order to compute a subgradient $g(a)$, we use the rules of (weak) subgradient calculus [35].

Proposition 2: The map $g : \mathbb{R}^2 \to \mathbb{R}^2$ defined as

$$g(a) \overset{\text{def}}{=} -2 \cdot \mathbf{E} \begin{bmatrix} (X_1 - a_1X_2) \cdot X_2 \cdot \mathbf{1}(|X_1 - a_1X_2| \geq |X_2 - a_2X_1|) \\ (X_2 - a_2X_1) \cdot X_1 \cdot \mathbf{1}(|X_1 - a_1X_2| < |X_2 - a_2X_1|) \end{bmatrix} \quad (102)$$

is a subgradient of $\mathcal{G}(a)$ defined in (95).

Proof: Let the function $\mathcal{G}(a; x)$ be defined as

$$\mathcal{G}(a; x) \overset{\text{def}}{=} \max\{\mathcal{G}_1(a; x), \mathcal{G}_2(a; x)\} \quad (103)$$

where

$$\mathcal{G}_1(a; x) \overset{\text{def}}{=} (x_1 - a_1x_2)^2 \quad (104)$$

and

$$\mathcal{G}_2(a; x) \overset{\text{def}}{=} (x_2 - a_2x_1)^2. \quad (105)$$

Hence, we have

$$\mathcal{G}(a) = \mathbf{E}[\mathcal{G}(a; X)]. \quad (106)$$

In order to construct a subgradient of $\mathcal{G}(a)$, we first find a subgradient $g(a; x)$ of $\mathcal{G}(a; x)$ and take its expectation with respect to x.

At the points where $\mathcal{G}_1(a; x) > \mathcal{G}_2(a; x)$, we have

$$g(a; x) \overset{\text{def}}{=} \nabla\mathcal{G}_1(a; x) = \begin{bmatrix} -2(x_1 - a_1x_2)x_2 \\ 0 \end{bmatrix}. \quad (107)$$

Similarly, at the points where $\mathcal{G}_1(a; x) < \mathcal{G}_2(a; x)$, we have

$$g(a; x) \overset{\text{def}}{=} \nabla\mathcal{G}_2(a; x) = \begin{bmatrix} 0 \\ -2(x_2 - a_2x_1)x_1 \end{bmatrix}. \quad (108)$$

When $\mathcal{G}_1(a; x) = \mathcal{G}_2(a; x)$, either one of the gradients above can be chosen as a subgradient of $\mathcal{G}(a; x)$. The following choice is a valid subgradient of $\mathcal{G}(a; x)$:

$$g(a; x) = \nabla\mathcal{G}_1(a; x)\mathbf{1}\left(\mathcal{G}_1(a; x) \geq \mathcal{G}_2(a; x)\right) + \nabla\mathcal{G}_2(a; x)\mathbf{1}\left(\mathcal{G}_1(a; x) < \mathcal{G}_2(a; x)\right). \quad (109)$$

Finally, we let

$$g(a) \overset{\text{def}}{=} \mathbf{E}\left[g(a; X)\right]. \quad (110)$$

■

Remark 8: The CCP does not guarantee that the solutions found through it are globally optimal. Table I shows the solutions found by the CCP algorithm for random variables with

TABLE I
BEST KNOWN COST FOR SCHEDULING POLICIES INDUCED BY PIECEWISE LINEAR ESTIMATION POLICIES FOR GAUSSIAN RANDOM VARIABLES WITH $\sigma_1^2 = 5$ AND $\sigma_2^2 = 7$

ρ	$\mathcal{J}_q^\star$	$a_1^\star$	$a_2^\star$
0	2.1271	0.0007	0.0012
0.1	2.1099	0.0552	0.0678
0.2	2.0579	0.1131	0.1330
0.3	1.9704	0.1709	0.2023
0.4	1.8457	0.2292	0.2772
0.5	1.6815	0.2936	0.3513
0.6	1.4741	0.3612	0.4345
0.7	1.2179	0.4336	0.5314
0.8	0.9038	0.5189	0.6426
0.9	0.5149	0.6255	0.7897

variances $\sigma_1^2 = 5$ and $\sigma_2^2 = 7$ and several values of the correlation coefficient ρ. Notice that the CCP is completely independent of the joint density f_X. One advantage of this numerical scheme is that it can be used for jointly designing schedulers and piecewise linear estimators for any pair of random variables, regardless of their joint distribution.

IX. EXTENSIONS

Theorem 1 can be extended to any number of sensors observing independent zero-mean Gaussian random variables. This is a significant generalization of the two-sensor case considered in Section IV. Let $x \in \mathbb{R}^n$ and consider the following generalization to the max-scheduling and mean-estimation strategies for $n \geq 2$:

$$\mathcal{U}^{\max}(x) \overset{\text{def}}{=} \arg \max_{i \in \{1,\ldots,n\}} |x_i| \tag{111}$$

and

$$\mathcal{E}^{\text{mean}}(i, \xi) \overset{\text{def}}{=} \xi \cdot \mathbf{e}_i \tag{112}$$

where $\mathbf{e}_i$ is the ith standard basis vector in $\mathbb{R}^n$, $i \in \{1, 2, \ldots, n\}$, and $\xi \in \mathbb{R}$.

Theorem 6: If $X \sim \mathcal{N}(0, \mathbf{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2))$, then $(\mathcal{U}^{\max}, \mathcal{E}^{\text{mean}})$ is a person-by-person optimal solution to Problem 1.

Theorem 6 is useful for the scheduling of n sensors with arbitrary correlation matrix provided that we use the pre- and postprocessing using the decorrelating transformation approach. We present the generalization of Theorem 3 to an arbitrarily correlated n-dimensional Gaussian random vector.

Theorem 7: Let $X \sim \mathcal{N}(0, \mathbf{\Sigma})$ and the unitary matrix $\mathbf{W}$ be obtained from the eigendecomposition of the covariance matrix as $\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^\mathsf{T}$. Let $x \in \mathbb{R}^n$. Define

$$\mathcal{U}^{\text{dec}}(x) \overset{\text{def}}{=} \mathcal{U}^{\max}(\mathbf{W}x) \tag{113}$$

and

$$\mathcal{E}^{\text{dec}}(i, \xi) \overset{\text{def}}{=} \mathbf{W}^\mathsf{T}\mathcal{E}^{\text{mean}}(i, \xi) \tag{114}$$

where $\mathcal{U}^{\max}$ and $\mathcal{E}^{\text{mean}}$ are given by (111) and (112), respectively, with $i \in \{1, 2, \ldots, n\}$ and $\xi \in \mathbb{R}$. The pair $(\mathcal{U}^{\text{dec}}, \mathcal{E}^{\text{dec}})$ is a person-by-person optimal solution to Problem 1.

Remark 9: The proofs of Theorems 6 and 7 are similar to the proofs of Theorems 1 and 3 and are omitted for brevity.

X. CONCLUSION

We have introduced a new scheduling problem, where a centralized agent observes the realization of a bivariate Gaussian vector and chooses a single component to be transmitted to a remote estimator. The motivation for this problem comes from information constraints commonly found in networked control and estimation, where a single packet can be reliably transmitted over a communication link at a time. The design of globally optimal scheduling and estimation policies is elusive due to the lack of convexity of the overall optimization problem. However, we can establish the person-by-person optimality of specific pairs of policies in two important special cases: independent observations and symmetrically correlated observations. In the independent case, the person-by-person optimality result can be extended to any number of sensors. We also showed how to use the first person-by-person result to obtain suboptimal policies for the general correlated Gaussian case. Finally, we have considered the joint design of scheduling and estimation policies for a bivariate Gaussian source when the estimator is constrained to the class of piecewise linear estimators. In this case, we obtained locally optimal solutions to the resulting nonconvex optimization problem by using the CCP.

There are many open research questions stemming from this work. One important problem is to prove the conjecture that the pair $(\mathcal{U}^{\max}, \mathcal{E}^{\text{mean}})$ is globally optimal for the two main cases in the first part of this paper. We believe that the proof of global optimality will involve results from information theory, such as the data-processing inequality and rate-distortion function for Gaussian vectors. A second topic for future work is to extend the person-by-person optimality result for n-dimensional correlated Gaussian random vectors under suitable symmetry assumptions on the covariance matrix. Finally, another line of work is to investigate more general jointly distributed random variables, and to possibly account for situations where the joint distribution is unknown and needs to be estimated from data.

APPENDIX A
MONOTONICITY OF $\mathcal{P}$ AND $\mathcal{T}$

The proof of Theorem 2 relies on the fact that the functions $\mathcal{P}$ and $\mathcal{T}$ defined in (45) and (46) are nondecreasing. Here, we present a proof of Lemma 5.

Proof of Lemma 5: Consider the function $\mathcal{P}$. Recall that

$$\mathcal{P}(\alpha) \overset{\text{def}}{=} \alpha - \eta(\alpha) \tag{115}$$

where η is defined in (14). The function $\mathcal{P}$ can be alternatively expressed as

$$\mathcal{P}(\alpha) = \mathbf{E}\left[\alpha - X \;\middle|\; -|\alpha| \leq X \leq |\alpha|\right] \tag{116}$$

where $X \sim \mathcal{N}(\rho\alpha, \sigma^2(1-\rho^2))$. Since $\mathcal{P}$ is an odd function, we can constrain our analysis to $\alpha \geq 0$, without loss of generality. Therefore, we assume that

$$\mathcal{P}(\alpha) = \mathbf{E}\left[\alpha - X \;\middle|\; -\alpha \leq X \leq \alpha\right]. \tag{117}$$

Notice that, when conditioned on $\{-\alpha \leq X \leq \alpha\}$, the following inequality holds:

$$\alpha - X \geq 0 \text{ a.s.} \tag{118}$$

Therefore, we can rewrite $\mathcal{P}$ as

$$\mathcal{P}(\alpha) = \int_0^\infty \mathbf{P}\left(\alpha - X \geq t \,\Big|\, -\alpha \leq X \leq \alpha\right) dt. \tag{119}$$

Define the following function:

$$\mathcal{W}(\alpha, t) \overset{\text{def}}{=} \mathbf{P}\left(X \leq \alpha - t \,\Big|\, -\alpha \leq X \leq \alpha\right) \tag{120}$$

and notice that

$$\mathcal{W}(\alpha, t) = 0, \;\; t \geq 2\alpha. \tag{121}$$

For any fixed $t \in [0, 2\alpha]$, the function $\mathcal{W}(\alpha, t)$ is nondecreasing in α. In order to show this, consider

$$\mathcal{W}(\alpha, t) = 1 - \frac{\mathbf{P}\left(\alpha - t \leq X \leq \alpha\right)}{\mathbf{P}\left(-\alpha \leq X \leq \alpha\right)}. \tag{122}$$

Let

$$f_t(\alpha) \overset{\text{def}}{=} \mathbf{P}\left(\alpha - t \leq X \leq \alpha\right). \tag{123}$$

The fact that $t \in [0, 2\alpha]$ implies that the derivative of f with respect to α satisfies

$$f_t'(\alpha) \leq 0. \tag{124}$$

We proceed to define g as

$$g(\alpha) \overset{\text{def}}{=} \mathbf{P}\left(-\alpha \leq X \leq \alpha\right). \tag{125}$$

It can be easily verified that g is nondecreasing. Therefore, we have

$$g'(\alpha) \geq 0. \tag{126}$$

Since

$$f_t'(\alpha) g(\alpha) \leq 0 \leq f_t(\alpha) g'(\alpha) \tag{127}$$

we have

$$\left(\frac{f_t(\alpha)}{g(\alpha)}\right)' \leq 0 \tag{128}$$

which implies that $\mathcal{W}(\alpha, t)$ is a nondecreasing function of α for all $t \in [0, 2\alpha]$. Since

$$\mathcal{P}(\alpha) = \int_0^{2\alpha} \mathcal{W}(\alpha, t) dt \tag{129}$$

is a superposition of nondecreasing functions, the function $\mathcal{P}(\alpha)$ is also nondecreasing. The proof of monotonicity of $\mathcal{T}(\alpha)$ follows the same sequence of steps and is omitted for brevity. ■

Appendix B
Proof of Theorem 5

Proof: Under the assumption that the random variables are symmetrically correlated, i.e., $\sigma_1 = \sigma_2$, we may constrain the optimization of $\mathcal{J}_q$ to $a = [\theta \;\; \theta]^\mathsf{T}$ with $\theta \in \mathbb{R}$, without loss of optimality. Therefore, we have

$$\mathcal{E}_a^{\text{linear}}(1, x) = \begin{bmatrix} x \\ \theta x \end{bmatrix} \text{ and } \mathcal{E}_a^{\text{linear}}(2, x) = \begin{bmatrix} \theta x \\ x \end{bmatrix}. \tag{130}$$

Lemma 2 implies that for all $\theta < 1$, the policy $\mathcal{U}^{\max}$ is optimal for $\mathcal{E}_a^{\text{linear}}$. Evaluating the cost for the pair $(\mathcal{U}^{\max}, \mathcal{E}_a^{\text{linear}})$, we obtain

$$\mathcal{J}(\mathcal{U}^{\max}, \mathcal{E}_a^{\text{linear}}) = \mathbf{E}\left[(X_2 - \theta X_1)^2 \mid U = 1\right] \mathbf{P}(U = 1) + \mathbf{E}\left[(X_1 - \theta X_2)^2 \mid U = 2\right] \mathbf{P}(U = 2). \tag{131}$$

The symmetry of the sources implies that the first-order derivative with respect to θ is

$$\frac{\partial}{\partial \theta} \mathcal{J}(\mathcal{U}^{\max}, \mathcal{E}_a^{\text{linear}}) = -2\rho \cdot \sigma^2 + 4\theta \cdot \mathbf{E}\left[X_1^2 \mid U = 1\right] \mathbf{P}(U = 1). \tag{132}$$

Therefore, the first-order optimality condition implies that

$$\theta^\star = \frac{\rho \cdot \sigma^2}{2 \cdot \int_{\mathbb{R}^2} x_1^2 \mathbf{1}(|x_1| \geq |x_2|) f_X(x) dx}. \tag{133}$$

It can be shown that $\theta^\star$ computed above satisfies $\theta^\star < 1$. ■

References

[1] P. J. Antsaklis, "Goals and challenges in cyber-physical systems research," *IEEE Trans. Autom. Control*, vol. 59, no. 12, pp. 3117–3119, Dec. 2014.

[2] M. M. Vasconcelos and N. C. Martins, "Optimal estimation over the collision channel," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 321–336, Jan. 2017.

[3] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[4] U. Mitra *et al.*, "KNOWME: A case study in wireless body area sensor network design," *IEEE Commun. Mag.*, vol. 50, no. 5, pp. 116–125, May 2012.

[5] K. Arrow, "The economics of agency," Inst. Math. Stud. Social Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep. No. 451, Oct. 1984.

[6] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proc. IEEE*, vol. 95, no. 1, pp. 138–162, Jan. 2007.

[7] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, pp. 393–422, 2002.

[8] J. Le Ny, E. Feron, and M. A. Dahleh, "Scheduling continuous-time Kalman filters," *IEEE Trans. Autom. Control*, vol. 56, no. 6, pp. 1381–1394, Jun. 2011.

[9] J. Moon and T. Basar, "Static optimal sensor selection via linear integer programming: The orthogonal case," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 953–957, Jul. 2017.

[10] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," *Automatica*, vol. 48, pp. 2482–2493, 2012.

[11] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[12] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47, no. 7, pp. 1330–1338, Jul. 2011.

[13] L. Shi and H. Zhang, "Scheduling two Gauss-Markov systems: An optimal solution for remote state estimation under bandwidth constraint," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2038–2042, Apr. 2012.

[14] S. Liu, M. Fardad, E. Masazade, and P. K. Varshney, "Optimal periodic sensor scheduling in networks of dynamical systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3055–3068, Jun. 2014.

[15] A. Zare, H. Mohammadi, N. K. Dhingra, M. R. Jovanović, and T. T. Georgiou, "Proximal algorithms for large-scale statistical modeling and optimal sensor/actuator selection," 2018, *arXiv:1807.01739*.

[16] G. C. Walsh and H. Ye, "Scheduling of networked control systems," *IEEE Control Syst. Mag.*, vol. 21, no. 1, pp. 57–65, Feb. 2001.

[17] M. Mannan and A. Mahajan, "Simultaneous real-time transmission of multiple Markov sources over a shared channel," in *Proc. IEEE Int. Symp. Inf. Theory*, 2014, pp. 2356–2360.

[18] D. Han, J. Wu, Y. Mo, and L. Xie, "On stochastic sensor network scheduling for multiple processes," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6633–6640, Dec. 2017.

[19] A. S. Leong, S. Dey, and D. E. Quevedo, "Sensor scheduling in variance based event triggered estimation with packet drops," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1880–1895, Apr. 2017.

[20] M. Xia, V. Gupta, and P. J. Antsaklis, "Networked state estimation over a shared communication medium," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1729–1741, Apr. 2017.

[21] J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, "Event-based sensor data scheduling: Trade-off between communication rate and estimation quality," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 1041–1046, Apr. 2013.

[22] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2661–2675, Oct. 2015.

[23] S. Weerakkody, Y. Mo, B. Sinopoli, D. Han, and L. Shi, "Multi-sensor scheduling for state estimation with event-based stochastic triggers," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2695–2701, Sep. 2016.

[24] Y.-C. Ho, M. Kastner, and E. Wong, "Teams, signaling, and information theory," *IEEE Trans. Autom. Control*, vol. AC-23, no. 2, pp. 305–312, Apr. 1978.

[25] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *Proc. 22nd Nat. Conf. Artif. Intell.*, 2007, pp. 1650–1654.

[26] R. M. Gray, "Quantization in task-driven sensing and distributed processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, p. V.

[27] S. Yuksel and T. Basar, *Stochastic Networked Control Systems*. New York, NY, USA: Springer, 2013.

[28] M. M. Vasconcelos and U. Mitra, "Observation-driven sensor scheduling," in *Proc. IEEE Int. Conf. Commun.*, 2017, pp. 1–6.

[29] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1992.

[30] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory* (ser. Classics in Applied Mathematics). Philadelphia, PA, USA: SIAM, 1999.

[31] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. New York, NY, USA: Springer, 1992.

[32] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY, USA: Springer, 1998.

[33] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

[34] T. Lipp and S. Boyd, "Variations and extensions of the convex-concave procedure," *Optim. Eng.*, vol. 17, no. 2, pp. 263–287, Jun. 2016.

[35] S. Boyd, J. Duchi, and L. Vandenberghe, Subgradients. 2018. [Online]. Available: https://stanford.edu/class/ee364b/lectures/subgradients_notes.pdf

Marcos M. Vasconcelos received the B.Sc. and M.Sc. degrees in electrical engineering from the Federal University of Pernambuco, Recife, Brazil, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2016.

He is currently a Postdoctoral Research Associate with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA. His research interests include networked control and estimation, multiagent optimization, information theory, and systems biology.

Urbashi Mitra (F'07) received the B.S. and M.S. degrees from the University of California at Berkeley, CA, USA, in 1987 and 1989, and the Ph.D. degree from Princeton University, Princeton, NJ, USA.

She is currently the Gordon S. Marshall Professor in Engineering with the University of Southern California, Los Angeles, CA, with appointments in Electrical Engineering and Computer Science. Her research interests include wireless communications, communication and sensor networks, biological communication systems, detection and estimation, and the interface of communication, sensing, and control.

Dr. Mitra is the inaugural Editor-in-Chief of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS. She was a member of the IEEE Information Theory Society's Board of Governors from 2002 to 2007 and from 2012 to 2017, the IEEE Signal Processing Society's Technical Committee on Signal Processing for Communications and Networks from 2012 to 2017, and the IEEE Signal Processing Society's Awards Board from 2017 to 2018, and the Vice Chair of the IEEE Communications Society's Communication Theory Working Group (2017-2018). She is the recipient of the 2017 IEEE Communications Society Women in Communications Engineering Technical Achievement Award, a 2016 U.K. Royal Academy of Engineering Distinguished Visiting Professorship, a 2016 U.S. Fulbright Scholar Award, a 2016–2017 U.K. Leverhulme Trust Visiting Professorship, a 2015–2016 IEEE Communications Society Distinguished Lecturer, the 2012 Globecom Signal Processing for Communications Symposium Best Paper Award, the 2012 U.S. National Academy of Engineering Lillian Gilbreth Lectureship, the 2009 DCOSS Applications and Systems Best Paper Award, the Texas Instruments Visiting Professor (Fall 2002, Rice University), the 2001 Okawa Foundation Award, the 2000 Ohio State University (OSU) College of Engineering Lumley Award for Research, the 1997 OSU College of Engineering MacQuigg Award for Teaching, and a 1996 National Science Foundation CAREER Award.