Optimal Data Rate Allocation for Dynamic Sensor Fusion over Resource Constrained Communication Networks

Hyunho Jung[1] Takashi Tanaka[2]

***Abstract*— We consider a dynamic sensor fusion problem where a large number of remote sensors observe a common Gauss-Markov process and the observations are transmitted to a fusion center over a resource constrained communication network. The design objective is to allocate an appropriate data rate to each sensor in such a way that the total data traffic to the fusion center is minimized, subject to a constraint on the fusion center's state estimation error covariance. We show that the problem can be formulated as a difference-of-convex program, to which we apply the convex-concave procedure (CCP) and the alternating direction method of multiplier (ADMM). Through a numerical study on a truss bridge sensing system, we observe that our algorithm tends to allocate zero data rate to unneeded sensors, implying that the proposed method is an effective heuristic for sensor selection.**

I. Introduction

The advancement of low-cost sensing technologies made a large amount of data easily accessible in control systems. While this is advantageous from the conventional control-theoretic viewpoints, engineers now face the issue of excessive data rate that often overwhelms systems' communication resources. Consequently, how to strategically discard superfluous sensor data is a relevant question to many applications.

In this paper, we consider a dynamic sensor fusion problem over a resource constrained communication network. Our primary focus is to optimize the allocation of scarce communication resources across a subset of different sensors. Our problem is closely related to several sensor selection problems that have been studied extensively in the literature. In [1], the sensor selection problem for minimizing the determinant of the estimation error covariance matrix is approached via a semidefinite programming (SDP) relaxation. In [2], energy constrained wireless network was considered and solved using the re-weighted ℓ_1 relaxation. The reference [3] applied stochastic dynamic programming to gather adequate information for multi-stage problem for control of a robotic assembly task. A structure sensor placement problem was considered in [4], where an iterative technique involving the Fisher information matrix (FIM) was developed. The work [5] also considered determinant of FIM and used genetic algorithm which selects set of sensor position that maximizes the determinant of the FIM. Structure sensor placement problems were also considered in [6] and [7]. In [8], mutual information (MI) was adopted as the information gain metric and was applied to a target location tracking problem using distributed sensors. The reference [9] introduced an approximation algorithm which estimates position of a target. The algorithm selects competitive sensors to guarantee estimation error within factor 2 of optimal choice under condition that the measurements are merged.

The problem considered in this paper is different in that we not only aim to select a subset of sensors, but also try to allocate an appropriate data rate to each sensor to minimize overall communication cost subject to a constraint on estimator accuracy. We first invoke basic results on entropy-coded data quantizers from the source coding literature to show that the communication data rate (measured in *bits*) from each sensor to the fusion center can be well-approximated by the MI between certain random variables. Based on this observation, we formulate an optimization problem (which is referred to as the *sensor resource allocation (SRA) problem* in the sequel) in which the sum of MI terms over all communication links is minimized subject to a constraint on the mean-square error (MSE) estimation performance achievable by the fusion center. The SRA problem is then formulated as a difference-of-convex program [10] to which we apply the heuristics of convex-concave procedure. Although our problem formulation is not combinatorial in nature, notably, the proposed mechanism is sparsity-promoting – the algorithm tends to identify unneeded sensors by allocating them zero data rates, and the number of unneeded sensors tends to increase as the constraint on the MSE performance is made less stringent. Therefore, the proposed method can be used as a new and effective heuristic for sensor selection.

This paper is organized as follows. In Section II, we formulate the SRA problem after reviewing the method of entropy-coded dithered-quantizers (ECDQ). In Section III, the SRA problem is reformulated as a difference-of-convex program. We propose practical solution approaches based on the CCP and the ADMM in Section IV. Numerical demonstrations on a truss bridge sensor placement problem are presented in Section V. We conclude in Section VI.

Notation: Lower case boldface symbols such as $\mathbf{x}$ are used to denote random variables. We use $\mathbf{x}_{1:t} = (\mathbf{x}_1, ..., \mathbf{x}_t)$ to denote the random process. We adopt standard notation for information-theoretic functions [11]: the entropy of a discrete random variable $\mathbf{x}$ is denoted by $H(\mathbf{x})$, while the differential entropy of a continuous random variable $\mathbf{x}$ is denoted by $h(\mathbf{x})$. The mutual information between $\mathbf{x}$ and $\mathbf{y}$ is denoted by $I(\mathbf{x};\mathbf{y})$, and the relative entropy is denoted by $D(\cdot\|\cdot)$. We use $\mathbb{S}^n$ for the set of symmetric matrices of size $n \times n$.

*This work is supported by DARPA Grant D19AP00078, FOA-AFRL-AFOSR-2019-0003 and NSF Award 1944318.

[1]H. Jung is with the Department of Mechanical Engineering, University of Texas at Austin, TX 78712, USA. jung.hyunho@utexas.edu

[2]T. Tanaka is with the Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, TX 78712, USA. ttanaka@utexas.edu

For $X \in \mathbb{S}^n$, $X \in \mathbb{S}^n_+$ or $X \succeq 0$ means that X is a positive semidefinite matrix, and $X \in \mathbb{S}^n_{++}$ or $X \succ 0$ means that X is a positive definite matrix.

II. Problem Formulation

We consider a remote estimation problem over a sensor network shown in Fig. 1. The random process to be estimated is a discrete-time, n-dimensional Gauss-Markov process

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + F\mathbf{w}_t, \quad \mathbf{w}_t \overset{i.i.d.}{\sim} \mathcal{N}(0, I), \quad t = 1, 2, ..., T \quad (1)$$

with $\mathbf{x}_1 \sim \mathcal{N}(0, P_{1|1})$, where matrices $A \in \mathbb{R}^{n\times n}$ and $P_{1|1} \in \mathbb{S}^n_+$ are given. There are M distributed sensors, each making a scalar-valued noiseless measurement $\mathbf{y}_{i,t} = C_i\mathbf{x}_t$, $i = 1, 2, ..., M$. For convenience, we write $\mathbf{y}_t = C\mathbf{x}_t$ where

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_{1,t} \\ \vdots \\ \mathbf{y}_{M,t} \end{bmatrix}, \quad C = \begin{bmatrix} C_1 \\ \vdots \\ C_M \end{bmatrix} \in \mathbb{R}^{M\times n}.$$

Throughout the paper, the pair (A, C) is assumed to be observable.

A. Data fusion over resource-constrained network

We consider the situation in which the output $\mathbf{y}_{i,t}$ of remote sensors must be transmitted to the data fusion center over a resource-limited communication network. The network model we introduce in this section assumes a CAN-bus-like communication system. All the nodes in the network operate synchronously in discrete-time. At every time step t, each sensor $i = 1, 2, ..., M$ encodes $\mathbf{y}_{i,t}$ into a packet $\mathbf{z}_{i,t}$. For each i and t, we assume that $\mathbf{z}_{i,t}$ is a uniquely decodable variable-length binary codeword with length $\ell_{i,t}$.[1] Packets $\mathbf{z}_{i,t}$ from sensors are received by the fusion center reliably (no packet losses, no bit flips) without delay. They are decoded by the fusion center where the least mean-square error (MSE) estimate $\hat{\mathbf{x}}_{t|t}$ of the source process (1) is computed. Based on $\hat{\mathbf{x}}_{t|t}$, the fusion center also computes a step-ahead prediction $\hat{\mathbf{x}}_{t+1|t} := A\hat{\mathbf{x}}_{t|t}$. Once $\hat{\mathbf{x}}_{t+1|t}$ is computed, we further assume that the fusion center broadcasts $\hat{\mathbf{x}}_{t+1|t}$ back to all the sensors (Fig. 1), which allows each sensor to apply the *predictive quantizer* at time step $t+1$.[2] We assume that the feedback channel delivers the $\mathbb{R}^n$-valued message $\hat{\mathbf{x}}_{t+1|t}$ reliably.[3]

[1] In reality, a packet frame in the CAN protocol contains a header and a tailer in addition to the data field. For simplicity, we assume $\ell_{i,t}$ only represents the lengths of the data field, ignoring the header and footer bits.

[2] To improve the coding efficiency, it is known to be more advantageous to quantize and encode the *innovation* $\mathbf{y}_{i,t+1} - C\hat{\mathbf{x}}_{t+1|t}$ than $\mathbf{y}_{i,t+1}$ itself. See, e.g., [12] [13].

[3] This assumption holds if the feedback message is given a sufficiently large codeword length so that the effect of quantization is negligible. Allowing feedback messages to have large blocklengths is a reasonable design since the backward channel delivers only one message per time step, whereas the forward channel delivers at most M messages per time step.

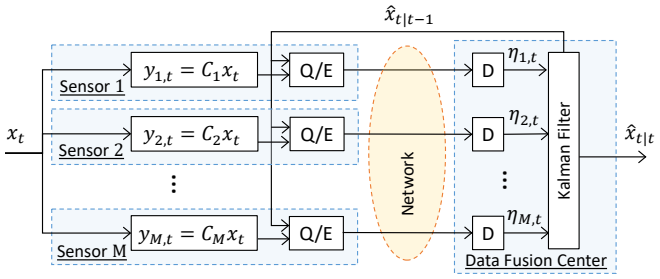

Fig. 1. Distributed sensors and data fusion center.

B. Entropy-coded dithered quantizer (ECDQ)

At each sensor node i, the continuous random variable $\mathbf{y}_{i,t}$ is first quantized into a discrete random variable, which is then encoded as a binary codeword $\mathbf{z}_{i,t}$. We assume that this process is implemented by the so-called *entropy-coded dithered quantizer (ECDQ)* mechanism [14] shown in Fig. 2(a). The ECDQ is easy to implement in practice, its mathematical analysis is relatively simple, and it attains a near-optimal performance (that is, the expected codeword length is close to the fundamental lower bound). As shown in Fig. 2(a), in time step t, the innovation signal $\boldsymbol{\theta}_{i,t} = \mathbf{y}_{i,t} - C_i\hat{\mathbf{x}}_{t|t-1}$ is first computed. It is then quantized by the dithered uniform quantizer with quantization step size $\Delta_{i,t}$:

$$Q_{\Delta_{i,t}}(\boldsymbol{\theta}_{i,t} + \boldsymbol{\xi}_{i,t}) = k\Delta_{i,t} \quad \text{if } (k - \tfrac{1}{2})\Delta_{i,t} \le \boldsymbol{\theta}_{i,t} + \boldsymbol{\xi}_{i,t} < (k + \tfrac{1}{2})\Delta_{i,t}.$$

Here, $\boldsymbol{\xi}_{i,t} \overset{i.i.d.}{\sim} \text{unif}[-\frac{\Delta_{i,t}}{2}, \frac{\Delta_{i,t}}{2}]$ is an artificial random variable called dither. The dither signal may not be necessary for practical implementations, but it simplifies the mathematical analysis of the communication system. The output $\mathbf{q}_{i,t}$ is then encoded into a binary codeword $\mathbf{z}_{i,t} \in \{0,1\}^{\boldsymbol{\ell}_{i,t}}$ using an entropy-based data-compression scheme (e.g., Huffman code, Shannon-Fano code). Notice that the codeword length $\boldsymbol{\ell}_{i,t}$ is a random variable. The codeword $\mathbf{z}_{i,t}$ is decoded losslessly as $\mathbf{q}_{i,t} = D(\mathbf{z}_{i,t})$ by the data fusion center. Then the dither signal is subtracted to compute $\boldsymbol{\eta}_{i,t} = \mathbf{q}_{i,t} - \boldsymbol{\xi}_{i,t}$, which is used for the belief update in the Kalman filter as shown in Fig. 2(a). Notice that the end-to-end effect of the ECDQ with input $\boldsymbol{\theta}_{i,t}$ and output $\boldsymbol{\eta}_{i,t}$ is

$$\boldsymbol{\eta}_{i,t} = Q_{\Delta_{i,t}}(\boldsymbol{\theta}_{i,t} + \boldsymbol{\xi}_{i,t}) - \boldsymbol{\xi}_{i,t}. \quad (2)$$

It can be shown [15] that (2) is mathematically equivalent to

$$\boldsymbol{\eta}_{i,t} = \boldsymbol{\theta}_{i,t} + \mathbf{v}_{i,t}, \quad \mathbf{v}_{i,t} \overset{i.i.d.}{\sim} \text{unif}[-\tfrac{\Delta_{i,t}}{2}, \tfrac{\Delta_{i,t}}{2}] \quad (3)$$

where the quantization noise $\mathbf{v}_{i,t}$ is independent of $\boldsymbol{\theta}_{i,1:t}$. The equivalence between (2) and (3) means that the channel models in Fig. 2(a) and Fig. 2(b) are equivalent, which simplifies the performance analysis.

C. Approximation of communication cost

We call the expected codeword length $R_i := \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\boldsymbol{\ell}_{i,t})$ the *rate* allocated to sensor i. The next lemma shows a relationship between the rate R_i and the mutual information $I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t})$:

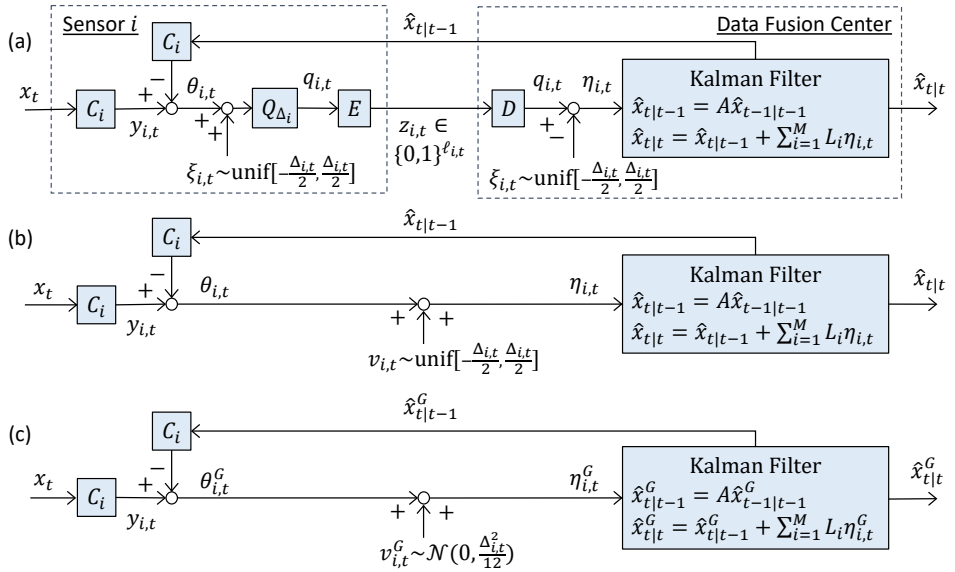

Fig. 2. (a) Channel model. (b) Equivalent channel model. (c) Simplified channel model.

Lemma 1: For every i and t, we have

$$I(\boldsymbol{\theta}_{i,t};\boldsymbol{\eta}_{i,t}) \le \mathbb{E}(\boldsymbol{\ell}_{i,t}) < I(\boldsymbol{\theta}_{i,t};\boldsymbol{\eta}_{i,t}) + 1$$

where the mutual information is evaluated under the joint distribution defined by the diagram in Fig. 2(b).

Proof: See Appendix A. ∎

Since $\boldsymbol{\theta}_{i,t}$ and $\boldsymbol{\eta}_{i,t}$ in Fig. 2(b) are not Gaussian random variables, it is difficult to evaluate $I(\boldsymbol{\theta}_{i,t};\boldsymbol{\eta}_{i,t})$ directly. A common approach (e.g., [16]) is to evaluate $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t})$ instead, where $\boldsymbol{\theta}^{\mathrm{G}}_{i,t}$ and $\boldsymbol{\eta}^{\mathrm{G}}_{i,t}$ are Gaussian random variables defined by the diagram in Fig. 2(c). In Fig. 2(c), the quantization noise $\mathbf{v}_{i,t} \overset{i.i.d.}{\sim} \text{unif}[-\frac{\Delta_{i,t}}{2},\frac{\Delta_{i,t}}{2}]$ is replaced by a Gaussian random variable $\mathbf{v}^{\mathrm{G}}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,V_{i,t})$, where we set $V_{i,t} = \frac{\Delta^2_{i,t}}{12}$ so that $\mathbf{v}_{i,t}$ and $\mathbf{v}^{\mathrm{G}}_{i,t}$ share the same covariance. Consequently, $\{\mathbf{x}^{\mathrm{G}}_{t|t}, \mathbf{x}^{\mathrm{G}}_{t|t-1}, \boldsymbol{\theta}^{\mathrm{G}}_t, \boldsymbol{\eta}^{\mathrm{G}}_t\}_{t=1,2,\dots,T}$ are jointly Gaussian random variables with the same mean and covariance as $\{\mathbf{x}_{t|t}, \mathbf{x}_{t|t-1}, \boldsymbol{\theta}_t, \boldsymbol{\eta}_t\}_{t=1,2,\dots,T}$. The following lemma provides an estimate of $I(\boldsymbol{\theta}_{i,t};\boldsymbol{\eta}_{i,t})$ in terms of $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t})$:

Lemma 2:

$$I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t}) \le I(\boldsymbol{\theta}_{i,t};\boldsymbol{\eta}_{i,t}) < I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t}) + \frac{1}{2}\log\frac{2\pi e}{12}.$$

Proof: See Appendix B. ∎

From Lemma 1 and Lemma 2, we obtain

$$I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t}) \le \mathbb{E}(\boldsymbol{\ell}_{i,t}) < I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t}) + \underbrace{1 + \frac{1}{2}\log\frac{2\pi e}{12}}_{\approx 1.254[\text{bits}]}. \quad (4)$$

This inequality implies that evaluating $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t})$ under the diagram Fig. 2(c) gives an estimate of the rate of the ECDQ under the architecture of Fig. 2(a) within the accuracy of 1.254 bits per time step. Notice that $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t})$ depends on $\Delta_{i,t}$ through the covariance of $\mathbf{v}^{\mathrm{G}}_{i,t} \sim \mathcal{N}(0,\Delta^2_{i,t}/12)$. Therefore, the rate R_i allocated to sensor i can be tuned by adjusting the quantizer step size $\Delta_{i,t}$ of the ECDQ.

D. Least MSE estimation

In Fig. 2(c), we assume that the Kalman filter block computes the least MSE estimates $\hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1} = \mathbb{E}(\mathbf{x}_t|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1})$ and $\hat{\mathbf{x}}^{\mathrm{G}}_{t|t} = \mathbb{E}(\mathbf{x}_t|\boldsymbol{\eta}^{\mathrm{G}}_{1:t})$ recursively by $\hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1} = A\hat{\mathbf{x}}^{\mathrm{G}}_{t-1|t-1}$ and $\hat{\mathbf{x}}^{\mathrm{G}}_{t|t} = \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1} + L_t\boldsymbol{\eta}^{\mathrm{G}}_t$ with the initial condition $\hat{\mathbf{x}}_{1|1} = 0$. Here,

$$L_t = P_{t|t-1}C^\top(CP_{t|t-1}C^\top + V)^{-1} \quad (5)$$

is the Kalman gain computed from the Riccati recursion

$$P^{-1}_{t|t} = P^{-1}_{t|t-1} + C^\top V^{-1}C \quad (6a)$$

$$P_{t+1|t} = AP_{t|t}A^\top + FF^\top. \quad (6b)$$

Matrices $P_{t|t} \in \mathbb{S}^n_{++}$ and $P_{t+1|t} \in \mathbb{S}^n_{++}$ represent the corresponding estimation error covariances

$$P_{t|t} = \text{Cov}(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t}), \quad P_{t+1|t} = \text{Cov}(\mathbf{x}_{t+1} - \hat{\mathbf{x}}^{\mathrm{G}}_{t+1|t}).$$

We assume that the same Kalman gains are used in Fig. 2(a) and (b) as well. Since random variables in Fig. 2(a) and (b) share the same second order statistics with random variables in Fig. 2(c), the MSE performance of the Kalman filter in Fig. 2(a) and (b) is identical to the MSE performance in Fig. 2(c). That is, we have

$$P_{t|t} = \text{Cov}(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}), \quad P_{t+1|t} = \text{Cov}(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}).$$

for Fig. 2(a) and (b). The next lemma provides an alternative expression of $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t})$ defined above.

Lemma 3: If L_t in Fig. 2(c) are chosen to be the Kalman gains defined by (5), then $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t};\boldsymbol{\eta}^{\mathrm{G}}_{i,t}) = I(\mathbf{x}_t;\boldsymbol{\eta}^{\mathrm{G}}_{i,t}|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1})$.

Proof: See Appendix C. ∎

E. Sensor Resource Allocation (SRA) Problem

We are now ready to state the main problem studied in this paper. For each $i = 1, 2, ..., M$, let $\alpha_i > 0$ be the cost of transmitting a binary value from the sensor i to the data fusion center. We seek the best allocation of the rate $R_i, i = 1, 2, ..., M$ in such a way that the total communication cost $\sum_{i=1}^{M} \alpha_i R_i$ is minimized subject to the constraint on the MSE estimation performance of the Kalman filter. Since (4) and Lemma 3 imply that R_i can be approximated by $I(\mathbf{x}_t; \boldsymbol{\eta}_{i,t}^{\mathrm{G}} | \boldsymbol{\eta}_{1:t-1}^{\mathrm{G}})$ evaluated under Fig. 2(c), in what follows, our analysis focuses on the system shown in Fig. 2(c). The SRA problem is formulated as:

$$\min \quad \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{M} \alpha_i I(\mathbf{x}_t; \boldsymbol{\eta}_{i,t}^{\mathrm{G}} | \boldsymbol{\eta}_{1:t-1}^{\mathrm{G}}) \tag{7a}$$

$$\text{s.t.} \quad \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}\|^2 \leq \beta. \tag{7b}$$

The decision variable is the noise covariance matrix $V_t = \mathrm{diag}(V_{1,t}, ..., V_{M,t}), t = 1, 2, ..., T$. If $\{V_t^*\}_{t=1}^T$ is the optimal solution to (7) and f^* is the corresponding optimal value, the argument above implies that one can construct the ECDQ-based communication system in Fig. 2(a) attaining the total network cost less than $f^* + 1.254 \times \sum_{i=1}^{M} \alpha_i$ by selecting the quantization step sizes $\Delta_{i,t}$ by $\frac{\Delta_{i,t}^2}{12} = V_{i,t}^*$. We are also interested in the infinite-horizon problem:

$$\min \quad \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{M} \alpha_i I(\mathbf{x}_t; \boldsymbol{\eta}_{i,t}^{\mathrm{G}} | \boldsymbol{\eta}_{1:t-1}^{\mathrm{G}}) \tag{8a}$$

$$\text{s.t.} \quad \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}\|^2 \leq \beta. \tag{8b}$$

III. Conversion to Convex-Concave Program

In this section, we reformulate (7) and (8) as more explicit optimization problems.

A. Reformulation of the SRA problem

As before, let $P_{t|t-1}$ be the estimation error covariance of $\mathbf{x}_t$ given $\boldsymbol{\eta}_{1:t-1}^{\mathrm{G}}$. Denote by $P_{t|t-1}^{(i)}$ the estimation error covariance of $\mathbf{x}_t$ given $\boldsymbol{\eta}_{1:t-1}^{\mathrm{G}}$ and $\boldsymbol{\eta}_{i,t}^{\mathrm{G}}$. They are related by

$$P_{t|t-1}^{(i)} = (P_{t|t-1}^{-1} + C_i^\top V_{i,t}^{-1} C_i)^{-1}.$$

1) Mutual information: The mutual information terms in (7a) can be written as

$$\begin{aligned}
&I(\mathbf{x}_t; \boldsymbol{\eta}_{i,t}^{\mathrm{G}} | \boldsymbol{\eta}_{1:t-1}^{\mathrm{G}}) = h(\mathbf{x}_t | \boldsymbol{\eta}_{1:t-1}^{\mathrm{G}}) - h(\mathbf{x}_t | \boldsymbol{\eta}_{1:t-1}^{\mathrm{G}}, \boldsymbol{\eta}_{i,t}^{\mathrm{G}}) \\
&= \tfrac{1}{2} \log \det P_{t|t-1} - \tfrac{1}{2} \log \det P_{t|t-1}^{(i)} \\
&= \tfrac{1}{2} \log \det P_{t|t-1} + \tfrac{1}{2} \log \det (P_{t|t-1}^{-1} + C_i^\top V_{i,t}^{-1} C_i) \\
&= \tfrac{1}{2} \log \det (I + P_{t|t-1}^{\frac{1}{2}} C_i^\top V_{i,t}^{-1} C_i P_{t|t-1}^{\frac{1}{2}}) \\
&= \tfrac{1}{2} \log (1 + V_{i,t}^{-\frac{1}{2}} C_i P_{t|t-1} C_i^\top V_{i,t}^{-\frac{1}{2}}) \\
&= \tfrac{1}{2} \log V_{i,t}^{-1} + \tfrac{1}{2} \log (V_{i,t} + C_i P_{t|t-1} C_i^\top).
\end{aligned}$$

Introducing changes of variables $Q_{t|t-1} := P_{t|t-1}^{-1}$, $Q_{t|t} := P_{t|t}^{-1}$ and $\delta_{i,t} := V_{i,t}^{-1}$,

$$\begin{aligned}
&I(\mathbf{x}_t; \boldsymbol{\eta}_{i,t}^{\mathrm{G}} | \boldsymbol{\eta}_{1:t-1}^{\mathrm{G}}) = \tfrac{1}{2} \log \delta_{i,t} - \tfrac{1}{2} \log (\delta_{i,t}^{-1} + C_i Q_{t|t-1}^{-1} C_i^\top)^{-1} \\
&= \begin{cases} \min_{\gamma_{i,t}} & \tfrac{1}{2} \log \delta_{i,t} - \tfrac{1}{2} \log \gamma_{i,t} \\ \text{s.t.} & \gamma_{i,t} \leq (\delta_{i,t}^{-1} + C_i Q_{t|t-1}^{-1} C_i \top)^{-1} \end{cases} \\
&= \begin{cases} \min_{\gamma_{i,t}} & \tfrac{1}{2} \log \delta_{i,t} - \tfrac{1}{2} \log \gamma_{i,t} \\ \text{s.t.} & \begin{bmatrix} \delta_{i,t} - \gamma_{i,t} & \delta_{i,t} C_i \\ C_i^\top \delta_{i,t} & Q_{t|t-1} + C_i^\top \delta_{i,t} C_i \end{bmatrix} \succeq 0. \end{cases}
\end{aligned} \tag{9}$$

The last equality is obtained by applying the matrix inversion lemma and the Schur complement formula to the constraint.

2) MSE: The MSE terms in (7b) can be written as

$$\begin{aligned}
&\mathbb{E} \|\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^{\mathrm{G}}\|^2 = \mathrm{Tr}(P_{t|t}) = \mathrm{Tr}(Q_{t|t}^{-1}) \\
&= \begin{cases} \min_{S_t} & \mathrm{Tr}(S_t) \\ \text{s.t.} & Q_{t|t}^{-1} \preceq S_t \end{cases} = \begin{cases} \min_{S_t} & \mathrm{Tr}(S_t) \\ \text{s.t.} & \begin{bmatrix} S_t & I \\ I & Q_{t|t} \end{bmatrix} \succeq 0. \end{cases}
\end{aligned} \tag{10}$$

3) Reformulation of (7)*:* From (9), (10) and (6), the SRA problem (7) can be written as

$$\min \quad \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{M} \frac{\alpha_i}{2} (\log \delta_{i,t} - \log \gamma_{i,t}) \tag{11a}$$

$$\text{s.t.} \quad \begin{bmatrix} \delta_{i,t} - \gamma_{i,t} & \delta_{i,t} C_i \\ C_i^\top \delta_{i,t} & Q_{t|t-1} + C_i^\top \delta_{i,t} C_i \end{bmatrix} \succeq 0, \tag{11b}$$

$$\begin{bmatrix} S_t & I \\ I & Q_{t|t} \end{bmatrix} \succeq 0, \ \frac{1}{T} \sum\nolimits_{t=1}^{T} \mathrm{Tr}(S_t) \leq \beta, \tag{11c}$$

$$Q_{t|t} = Q_{t|t-1} + \sum\nolimits_{i=1}^{M} \delta_{i,t} C_i^\top C_i, \tag{11d}$$

$$Q_{t|t-1}^{-1} = A Q_{t-1|t-1}^{-1} A^\top + F F^\top. \tag{11e}$$

with decision variables $(\delta_{i,t}, \gamma_{i,t})$ for $i = 1, ..., M$ and $t = 1, ..., T$, S_t for $t = 1, ..., T$ and $(Q_{t|t}, Q_{t|t-1})$ for $t = 2, ..., T$. The constraints (11b) and (11c) are imposed for $t = 1, 2, ..., T$ while the constraints (11d) and (11e) are imposed for $t = 2, ..., T$ with the boundary constraint $Q_{1|1} = P_{1|1}^{-1}$. Notice that (11b)-(11d) are convex constraints on the decision variables, while the last constraint (11e) is not. In the next proposition, we claim that (11e) can be replaced by a convex constraint without changing the nature of the problem. More precisely, introduce a new problem:

$$\min \quad \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{M} \frac{\alpha_i}{2} (\log \delta_{i,t} - \log \gamma_{i,t}) \tag{12a}$$

$$\text{s.t.} \quad \begin{bmatrix} \delta_{i,t} - \gamma_{i,t} & \delta_{i,t} C_i \\ C_i^\top \delta_{i,t} & Q_{t|t-1} + C_i^\top \delta_{i,t} C_i \end{bmatrix} \succeq 0, \tag{12b}$$

$$\begin{bmatrix} S_t & I \\ I & Q_{t|t} \end{bmatrix} \succeq 0, \ \frac{1}{T} \sum\nolimits_{t=1}^{T} \mathrm{Tr}(S_t) \leq \beta, \tag{12c}$$

$$Q_{t|t} = Q_{t|t-1} + \sum\nolimits_{i=1}^{M} \delta_{i,t} C_i^\top C_i, \tag{12d}$$

$$Q_{t|t-1}^{-1} \succeq A Q_{t-1|t-1}^{-1} A^\top + F F^\top \tag{12e}$$

which is different from (11) only in that the equality constraint (11e) is replaced by the inequality constraint (12e).

Proposition 1: Let J_1^* and J_2^* be the optimal values of (11) and (12), respectively. Then, $J_1^* = J_2^*$. Moreover, if $(\delta_{i,t}^*, \gamma_{i,t}^*, S_t^*, Q_{t|t}^*, Q_{t|t-1}^*)$ is an optimal solution to (12), then an optimal solution to (11) is given by $(\delta_{i,t}^*, \gamma_{i,t}^*, S_t^*, Q_{t|t}^{**}, Q_{t|t-1}^{**})$ where $Q_{t|t}^{**}$ and $Q_{t|t-1}^{**}$ are recursively defined by

$$Q_{t|t-1}^{**-1} = AQ_{t-1|t-1}^{**-1}A^\top + FF^\top \tag{13a}$$

$$Q_{t|t}^{**} = Q_{t|t-1}^{**} + \sum\nolimits_{i=1}^{M} \delta_{i,t}^* C_i^\top C_i \tag{13b}$$

with $Q_{1|1}^{**} = Q_{1|1}^*$.

Proof: See Appendix D. ∎

It is elementary to show that (12e) can be written as an equivalent linear matrix inequality (LMI) condition. Consequently, the SRA problem (7) can be equivalently written as

$$\min \quad \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{M}\frac{\alpha_i}{2}(\log\delta_{i,t} - \log\gamma_{i,t}) \tag{14a}$$

$$\text{s.t.} \quad \begin{bmatrix} \delta_{i,t}-\gamma_{i,t} & \delta_{i,t}C_i \\ C_i^\top\delta_{i,t} & Q_{t|t-1}+C_i^\top\delta_{i,t}C_i \end{bmatrix} \succeq 0, \tag{14b}$$

$$\begin{bmatrix} S_t & I \\ I & Q_{t|t} \end{bmatrix} \succeq 0, \quad \frac{1}{T}\sum\nolimits_{t=1}^{T}\mathrm{Tr}(S_t) \le \beta, \tag{14c}$$

$$Q_{t|t} = Q_{t|t-1} + \sum\nolimits_{i=1}^{M}\delta_{i,t}C_i^\top C_i. \tag{14d}$$

$$\begin{bmatrix} Q_{t|t-1} & Q_{t|t-1}A & Q_{t|t-1}F \\ A^\top Q_{t|t-1} & Q_{t-1|t-1} & 0 \\ F^\top Q_{t|t-1} & 0 & I \end{bmatrix} \succeq 0. \tag{14e}$$

Note that (14) is the problem of minimizing the difference of convex functions subject to convex constraints (14b)-(14e).

Before we proceed, we remark that the infinite-horizon, time-invariant counterpart (8) of the SRA problem can also be formulated as:

$$\min \quad \sum_{i=1}^{M}\frac{\alpha_i}{2}(\log\delta_i - \log\gamma_i) \tag{15a}$$

$$\text{s.t.} \quad \begin{bmatrix} \delta_i-\gamma_i & \delta_i C_i \\ C_i^\top\delta_i & \hat{Q}+C_i^\top\delta_i C_i \end{bmatrix} \succeq 0, \ \forall i = 1,...,M, \tag{15b}$$

$$Q = \hat{Q} + \sum\nolimits_{i=1}^{M}\delta_i C_i^\top C_i, \quad \mathrm{Tr}(S) \le \beta, \tag{15c}$$

$$\begin{bmatrix} S & I \\ I & Q \end{bmatrix} \succeq 0, \quad \begin{bmatrix} \hat{Q} & \hat{Q}A & \hat{Q}F \\ A^\top\hat{Q} & Q & 0 \\ F^\top\hat{Q} & 0 & I \end{bmatrix} \succeq 0. \tag{15d}$$

B. Nonconvexity of the SRA problem

Currently, it is not known to the authors if the SRA problem can be formulated as a convex optimization problem. Since the formulation (15) is nonconvex in its variables, it may admit multiple local minima. To see that multiple and distinct local minima can indeed occur, notice that (15) can be stated as a linear function minimization problem over the feasible rate region $\mathcal{R}_\beta$:

$$\min\nolimits_{(R_1,...,R_M)\in\mathcal{R}_\beta} \sum_{i=1}^{M}\alpha_i R_i \tag{16}$$

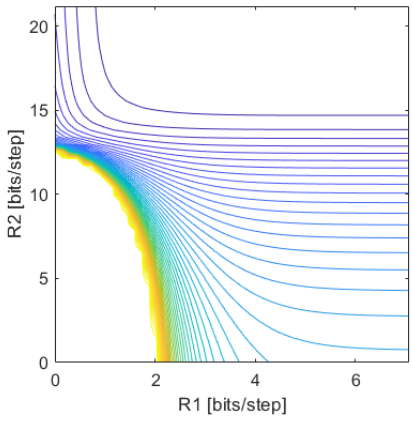

Fig. 3. The minimum MSE error β achievable under various rate assignments (R_1, R_2). Each sub-level set corresponds to the rate region $\mathcal{R}_\beta$. Clearly, they are nonconvex sets in general.

where $\mathcal{R}_\beta \in \mathbb{R}^M$ is the set of rate allocation under which (8b) (i.e., the MSE less than or equal to β) is achievable:

$$\mathcal{R}_\beta = \{(R_1,...,R_M) \in \mathbb{R}^M : \text{There exists } Q, \hat{Q}, S \text{ and } \{\delta_i,\gamma_i\}_{i=1}^M \text{ such that } \delta_i = 2^{2R_i}\gamma_i \text{ for } i = 1,...,M \text{ and (15b)-(15d) hold.}\}$$

For each β, the rate region $\mathcal{R}_\beta$ can be characterized by an SDP feasibility problem. Fig. 3 shows feasibility regions $\mathcal{R}_\beta$ for various β when system parameters are set to

$$A = \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } F = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$$

Fig. 3 shows an instance in which $\mathcal{R}_\beta$ is a nonconvex set.

IV. Algorithms

In this section, we propose two heuristic iterative algorithms to solve the SRA problem. Although we will focus on the infinite-horizon time-invariant case (15), the proposed approach is also applicable to the finite-horizon time-varying case (14). Despite the nonconvex nature of (15), we observe that stationary points obtained by the proposed algorithms often provide satisfactory solutions in practice.

A. Convex-concave procedure

Since $\log\delta_i$ is the only source of nonconvexity in (15), the class of convex-concave procedures [10] is applicable. Here, we consider the linear approximation of $\log\delta_i$ around a nominal point $\hat{\delta}_i$, which provides an upper bound, i.e., $\log\delta_i \le \frac{1}{\hat{\delta}_i}(\delta_i - \hat{\delta}_i) + \log\hat{\delta}_i$. Consequently, for any $\hat{\delta}_i > 0$, the value of the following convex optimization problem with decision variables $S, Q, \hat{Q}$ and $\{\delta_i,\gamma_i\}_{i=1}^M$ provides an upper bound to the value of (15):

$$\min \quad \sum_{i=1}^{M}\frac{\alpha_i}{2}(\delta_i/\hat{\delta}_i - 1 + \log\hat{\delta}_i - \log\gamma_i) \tag{17a}$$

$$\text{s.t.} \quad \text{(15b)-(15d)}. \tag{17b}$$

Proposed approach is summarized in Algorithm 1. Convergence of this class of algorithms is known [10].

Algorithm 1: Convex-Concave Procedure (CCP)

Initialize $f^{(0)} \leftarrow +\infty$; $\hat{\delta}_i \leftarrow 1$ for $i = 1, 2, ..., M$;
for $k = 1, 2, ...$ **do**
  Solve (17);
  $(\delta^k, \gamma^k, S^k, Q^k, \hat{Q}^k) \leftarrow$ Optimal solution to (17);
  $f^k \leftarrow$ Optimal value of (17);
  $\hat{\delta}_i \leftarrow \delta^k$ for $i = 1, 2, ..., M$;
  Break if $f^{k-1} - f^k$ is sufficiently small;

B. Alternating direction method of multipliers (ADMM)

In this subsection, we apply the ADMM [17] to (15). Notice that (15) can be expressed as the ADMM form as

$$\begin{aligned} \min \quad & f(j) + g(z) \\ \text{s.t.} \quad & \delta = \delta', \gamma = \gamma', \end{aligned}$$

where $j = \{\delta, \gamma\}$, $z = \{\delta', \gamma', Q, \hat{Q}, S\}$ and g is the indicator function for the convex set C characterized by (15b)-(15d). Setting $z' = \{\delta', \gamma'\}$ and $u = \{u_1, u_2\}$, the augmented Lagrangian is

$$L_\rho(j, z, u) = f(j) + g(z) + (\rho/2)\|j - z'^k + u^k\|_2^2$$

where ρ is a penalty parameter and u is the set of Lagrangian multipliers. The ADMM iterations for this problem are

$$\begin{aligned} j^{k+1} &:= \arg\min_j \{f(j) + (\rho/2)\|j - z'^k + u^k\|_2^2\}, \\ z^{k+1} &:= \Pi_c(j^{k+1} + u^k), \\ u^{k+1} &:= u^k + j^{k+1} - z'^{k+1}. \end{aligned}$$

Due to the nonconvexity of f, the j-update step involves a nonconvex optimization. Therefore, in the j-update step, we replaced $f(j)$ with its convex upper bound $\hat{f}(\hat{\delta}; j)$ by considering a linear approximation of the $\log \delta_i$ terms around the current iterate $\hat{\delta} = \delta^k$ in a similar fashion to (17). The projection operator Π_C is implemented by solving a Frobenius norm minimization problem subject to the convex constraints (15b)-(15d). Our proposed ADMM approach is presented in Algorithm 2.

Algorithm 2: The Alternating Direction Method of Multiplier (ADMM)

Initialize $f^{(0)} \leftarrow +\infty$;
Set initial value of j, z, and u;
for $k = 1, 2, ...$ **do**
  $j^{k+1} :=$ argmin $(\hat{f}(\hat{\delta}; j) + (\rho/2)\|j - z'^k + u^k\|_2^2)$;
  $z^{k+1} := \Pi_c(j^{k+1} + u^k)$;
  $u^{k+1} := u^k + j^{k+1} - z^{k+1}$;
  $f^k \leftarrow$ Current value of the objective function;
  $\hat{\delta} \leftarrow \delta^{k+1}$;
  Break if $f^k - f^{k+1}$ is sufficiently small;

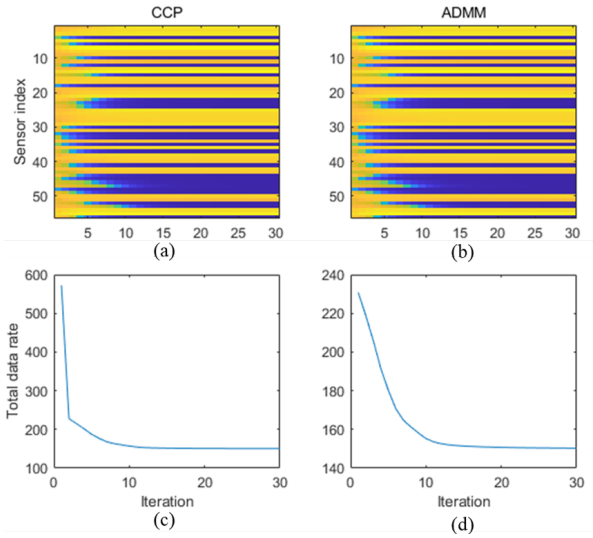

Fig. 4. CCP and ADMM data rate ($\beta = 0.1$). (a) Sensor data rate allocation under CCP, (b) sensor data rate allocation under ADMM, (c) total data rate under CCP, and (d) total data rate under ADMM.

V. Numerical Studies

In this section, we apply Algorithms 1 and 2 to a sensor selection problem in an undamped 2D truss bridge system. The system matrix of an undamped truss bridge is calculated by via (cf. [18]) $A = M^{-1}K$ where M is mass matrix and K is stiffness matrix. A method of generating the mass matrix [19] [20] and the stiffness matrix [21] is introduced in Appendix E. In this experiment, we developed a 14-node truss bridge model. Displacements and velocities of each node in both x- and y-coordinates are chosen as state variables. This results in a 56-dimensional state space, and we assume there are 56 sensors measuring individual state variables. In this study, we set $\alpha_i = 1$ for $i = 1, ..., 56$. Results for the CCP and ADMM algorithms are shown in Fig. 4. In each test, the same data rate is initially allocated to each sensors, which is updated as the iteration proceeds as color-coded in Fig. 4 (a) and (b). The total data rate is shown in Fig. 4 (c) and (d), respectively. We observe the both algorithms converge to similar solutions.

Fig. 5 presents allocated data rate to each sensor after a sufficient number of CCP iterations with $\beta = 0.1, 1$ and 10. We observe that the same subset of sensors is selected under $\beta = 0.1$ and 1, but the overall data rate is less under $\beta = 1$. As β is increased to 10, we observe more sensors are given zero data rate. However, we also observe that the selected set of sensors is not a subset of sensors selected under $\beta = 1$.

Fig. 6 shows the number of sensors allocated with nonzero data rate by CCP tested over a wide range of the β values. We observe a decrease tendency, although the relationship is not necessarily monotone. This plot exhibits a sparsity-promoting property of the proposed method, which is a reminiscent of the widely used ℓ_1 heuristics.

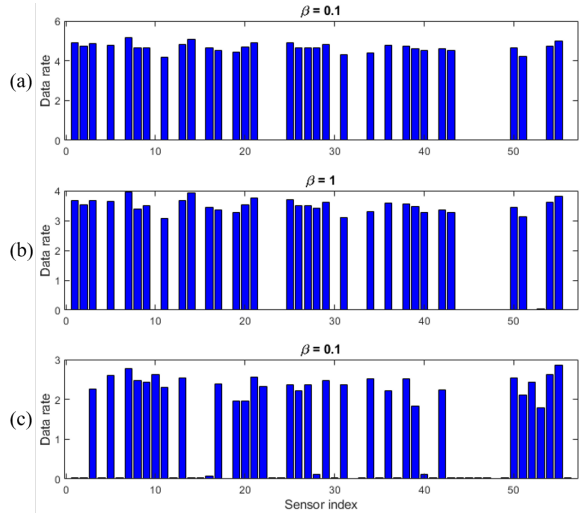

Fig. 5. Data rate allocation obtained by CCP with (a) $\beta = 0.1$ (b) $\beta = 1$ (c) $\beta = 10$.

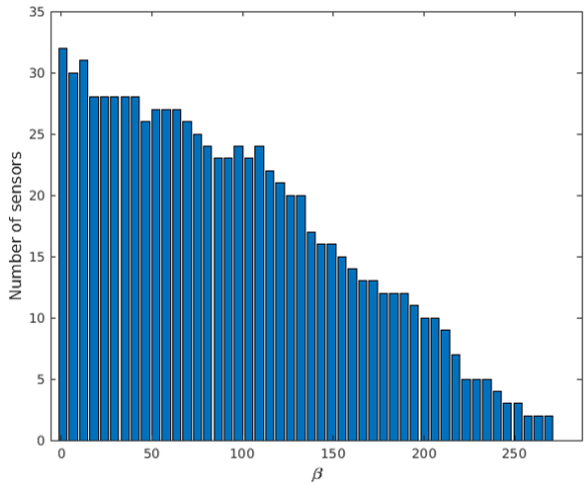

Fig. 6. Number of sensors allocated non-zero data rate by CCP tested over the range $1 \leq \beta \leq 280$.

VI. Conclusion

In this paper, we considered a dynamic sensor fusion problem over a resource constrained communication network. We formulated the optimal data rate assignment problem for remote sensors as the sensor resource allocation (SRA) problem, which was shown to be reformulated as a difference-of-convex program. The convex-concave procedure (CCP) and the alternating direction method of multipliers (ADMM) were applied. The algorithms were tested on a truss bridge sensor selection problem. The sparsity-promoting property of the proposed method was numerically confirmed, indicating the effectiveness of the proposed approach as a sensor selection heuristic. Future work includes the analysis of the nonconvexity of (15) (e.g., whether local minima can be severely suboptimal), scalable implementations of CCP and ADMM, and formal analyses of the sparsity-promoting property of the proposed method.

References

[1] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 451–462, 2009.

[2] 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, 2011.

[3] G. Hovland and B. J. McCarragher, "Dynamic sensor selection for robotic systems," *Proceedings of International Conference on Robotics and Automation*, vol. 1, pp. 272–277, 1997.

[4] D. C. Kammer, "Sensor placement for on-orbit modal identification and correlation of large space structures," *American Control Conference*, pp. 2984–2990, 1990.

[5] L. Yao, W. Sethares, and D. Kammer, "Sensor placement for on-orbit modal identification of large space structure via a genetic algorithm," *IEEE International Conference on Systems Engineering*, pp. 332–335, 1992.

[6] A. J. Walcker, "Model-based hybrid framework for live load carrying performance monitoring of bridges," 2017.

[7] F. Xiao, J. L. Husley, G. S. Chen, and Y. Xiang, "Optimal static strain sensor placement for truss bridges," *International Journal of Distributed Sensor Networks*, vol. 13, no. 5, 2017.

[8] F. Zhao and L. Guibas, *Wireless sensor networks: An information processing approach*. Morgan Kaufmann, 2004.

[9] V. Isler and R. Bajcsy, "The sensor selection problem for bounded uncertainty sensing models," *IEEE Transactions on Automation Science and Engineering*, vol. 3, no. 4, pp. 372–381, 2006.

[10] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optimization and Engineering*, vol. 17, no. 2, pp. 263–287, 2016.

[11] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[12] S. Yüksel, "Jointly optimal LQG quantization and control policies for multi-dimensional systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1612–1617, 2013.

[13] T. Tanaka, K. H. Johansson, T. Oechtering, H. Sandberg, and M. Skoglund, "Rate of prefix-free codes in LQG control systems," pp. 2399–2403, 2016.

[14] M. S. Derpich and J. Ostergaard, "Improved upper bounds to the causal quadratic rate-distortion function for gaussian stationary sources," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3131–3152, 2012.

[15] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 428–436, 1992.

[16] E. Silva, "A unified framework for the analysis and design of networked control systems," Ph.D. dissertation, University of Newcastle, 2009.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[18] L. S. Andrés. Undamped modal analysis of mdof systems. [Online]. Available: https://rotorlab.tamu.edu/TRIBGROUP/default.htm

[19] C. A. Felippa. (2004) Introduction to finite element methods. [Online]. Available: https://vulcanhammernet.files.wordpress.com/2017/01/ifem.pdf

[20] R. R. Craig Jr. and A. J. Kurdila, *Fundamentals of structural dynamics*. John Wiley & Sons, 2006.

[21] A. Kassimali, *Matrix analysis of structures*. Cengage Learning, 2012.

Appendix

A. Proof of Lemma 1

Since we assume entropy coding, the expected codeword length is approximated by the entropy of the source symbol $\mathbf{q}_{i,t}$ [11]:

$$H(\mathbf{q}_{i,t}|\boldsymbol{\xi}_{i,t}) \leq \mathbb{E}(\boldsymbol{\ell}_{i,t}) < H(\mathbf{q}_{i,t}|\boldsymbol{\xi}_{i,t}) + 1.$$

Here, the entropy terms are conditioned by $\boldsymbol{\xi}_{i,t}$ since the dither signal $\boldsymbol{\xi}_{i,t}$ is available for both the encoder and the

decoder. Since $\boldsymbol{\eta}_{i,t} = \mathbf{q}_{i,t} - \boldsymbol{\xi}_{i,t}$, clearly $H(\mathbf{q}_{i,t}|\boldsymbol{\xi}_{i,t}) = H(\boldsymbol{\eta}_{i,t}|\boldsymbol{\xi}_{i,t})$. Moreover, under Fig. 2(b), it can be shown that $H(\boldsymbol{\eta}_{i,t}|\boldsymbol{\xi}_{i,t}) = I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t})$. See [15] and [13, Lemma 1] for details. Therefore,

$$I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t}) \le \mathbb{E}(\boldsymbol{\ell}_{i,t}) < I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t}) + 1.$$

B. Proof of Lemma 2

To obtain an upper bound of $I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t})$, notice that

$$\begin{aligned} I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t}) &= h(\boldsymbol{\theta}_{i,t}) - h(\boldsymbol{\theta}_{i,t}|\boldsymbol{\eta}_{i,t}) \\ &= h(\boldsymbol{\theta}_{i,t}) - h(\boldsymbol{\eta}_{i,t} - \mathbf{v}_{i,t}|\boldsymbol{\eta}_{i,t}) \\ &= h(\boldsymbol{\theta}_{i,t}) - h(\mathbf{v}_{i,t}) \end{aligned}$$

However, we have $h(\boldsymbol{\theta}_{i,t}) \le h(\boldsymbol{\theta}^{\mathrm{G}}_{i,t})$ since entropy is maximized by Gaussian random variables when the covariances are the same. Moreover,

$$\begin{aligned} h(\mathbf{v}_{i,t}) &= h(\mathbf{v}^{\mathrm{G}}_{i,t}) - D(\mathbf{v}_{i,t} \| \mathbf{v}^{\mathrm{G}}_{i,t}) \\ &= h(\mathbf{v}^{\mathrm{G}}_{i,t}) - \frac{1}{2}\log\frac{2\pi e}{12}. \end{aligned}$$

Therefore,

$$\begin{aligned} I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t}) &\le h(\boldsymbol{\theta}^{\mathrm{G}}_{i,t}) - h(\mathbf{v}^{\mathrm{G}}_{i,t}) + \frac{1}{2}\log\frac{2\pi e}{12} \\ &= I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t}; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}) + \frac{1}{2}\log\frac{2\pi e}{12}. \end{aligned} \tag{18}$$

On the other hand, a lower bound $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t}; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}) \le I(\boldsymbol{\theta}_{i,t}; \boldsymbol{\eta}_{i,t})$ follows from Lemma C.1 of [16].

C. Proof of Lemma 3

To prove the claimed equality, we will establish $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t}; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}) \le I(\mathbf{x}_t; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1})$ and $I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t}; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}) \ge I(\mathbf{x}_t; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1})$ separately. First, recall that if random variables $(\mathbf{x}, \mathbf{y})$ are independent of $\mathbf{z}$, then

$$I(\mathbf{x}; \mathbf{y}|\mathbf{z}) = I(\mathbf{x}; \mathbf{y}). \tag{19}$$

Proof of "$\le$": This part follows as follows:

$$\begin{aligned} &I(\mathbf{x}_t; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}) \\ &= I(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}; C_i(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}) + \mathbf{v}^{\mathrm{G}}_{i,t}|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}) && \text{(20a)} \\ &= I(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}; C_i(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}) + \mathbf{v}^{\mathrm{G}}_{i,t}) && \text{(20b)} \\ &= I(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}) \\ &\ge I(C_i(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}); \boldsymbol{\eta}^{\mathrm{G}}_{i,t}) && \text{(20c)} \\ &= I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t}; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}). \end{aligned}$$

The equality (20a) holds because $\hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}$ is a deterministic constant when $\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}$ is given. The identity (19) is applicable in the step (20b) since $\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}$ and $\mathbf{v}^{\mathrm{G}}_{i,t}$ are independent of $\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}$ (the first independence is the consequence of the orthogonality principle for the least MSE estimator). The data processing inequality [11] is used in step (20c).

Proof of "$\ge$": This part follows as follows:

$$\begin{aligned} &I(\boldsymbol{\theta}^{\mathrm{G}}_{i,t}; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}) \\ &= I(C_i(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}); C_i(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}) + \mathbf{v}^{\mathrm{G}}_{i,t}) \\ &= I(C_i(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}); C_i(\mathbf{x}_t - \hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}) + \mathbf{v}^{\mathrm{G}}_{i,t}|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}) && \text{(21a)} \\ &= I(\mathbf{y}_{i,t}; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}) && \text{(21b)} \\ &\ge I(\mathbf{x}_t; \boldsymbol{\eta}^{\mathrm{G}}_{i,t}|\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}). && \text{(21c)} \end{aligned}$$

The identity (19) is used in step (21a). Equality (21b) holds since $\hat{\mathbf{x}}^{\mathrm{G}}_{t|t-1}$ is a deterministic function of $\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}$. The final step (21c) is the result of the data processing inequality since $\mathbf{x}_t \leftrightarrow \mathbf{y}_{i,t} \leftrightarrow \boldsymbol{\eta}^{\mathrm{G}}_{i,t}$ forms a Markov chain when $\boldsymbol{\eta}^{\mathrm{G}}_{1:t-1}$ is given, i.e., $\mathbf{x}_t$ and $\boldsymbol{\eta}^{\mathrm{G}}_{i,t}$ are conditionally independent given $(\mathbf{y}_{i,t}, \boldsymbol{\eta}^{\mathrm{G}}_{1:t-1})$.

D. Proof of Proposition 1

The proof is based on the following claim:

Claim 1: For each $t = 1, ..., T$, we have

$$Q^{**}_{t|t} \succeq Q^{*}_{t|t}. \tag{22a}$$

$$Q^{**}_{t+1|t} \succeq Q^{*}_{t+1|t}. \tag{22b}$$

Proof: Proof is by induction. When $t = 1$, (22a) is trivial since $Q^{**}_{1|1} = Q^{*}_{1|1}$ by construction. Notice also that $Q^{**-1}_{2|1} = AQ^{**-1}_{1|1}A^\top + FF^\top$ by construction, and $Q^{*-1}_{2|1} \succeq AQ^{*-1}_{1|1}A^\top + FF^\top$ because $(Q^{*}_{1|1}, Q^{*}_{2|1})$ is a feasible solution satisfying (12e). Thus $Q^{**-1}_{2|1} \preceq Q^{*-1}_{2|1}$ from which $Q^{**}_{2|1} \succeq Q^{*}_{2|1}$ follows. Therefore, the claim holds for $t = 1$. We now assume (22) holds for $t = k(\ge 1)$:

$$Q^{**}_{k|k} \succeq Q^{*}_{k|k}, \tag{23a}$$

$$Q^{**}_{k+1|k} \succeq Q^{*}_{k+1|k} \tag{23b}$$

and show that (22) also holds for $t = k+1$. By construction, we have

$$\begin{aligned} Q^{**}_{k+1|k+1} &= Q^{**}_{k+1|k} + \sum\nolimits_{i=1}^{M} \delta^{*}_{i,k+1} C_i^\top C_i \text{ and} \\ Q^{*}_{k+1|k+1} &= Q^{*}_{k+1|k} + \sum\nolimits_{i=1}^{M} \delta^{*}_{i,k+1} C_i^\top C_i \end{aligned}$$

Therefore

$$Q^{**}_{k+1|k+1} \succeq Q^{*}_{k+1|k+1} \tag{24}$$

follows immediately from (23b). Moreover, we have the following chain of inequalities:

$$\begin{aligned} Q^{**-1}_{k+2|k+1} &= AQ^{**-1}_{k+1|k+1}A^\top + FF^\top && \text{(25a)} \\ &\preceq AQ^{*-1}_{k+1|k+1}A^\top + FF^\top && \text{(25b)} \\ &\preceq Q^{*-1}_{k+2|k+1} && \text{(25c)} \end{aligned}$$

The equality (25a) holds by construction (13a). The second inequality (25b) is due to (24). The last inequality holds since $(Q^{*}_{k+1|k+1}, Q^{*}_{k+2|k+1})$ is a feasible solution satisfying (12e). It follows from (25) that

$$Q^{**}_{k+2|k+1} \succeq Q^{*}_{k+2|k+1}. \tag{26}$$

Inequalities (24) and (26) establish the claim for $t = k + 1$. ∎

The proof of Proposition 1 can be completed as follows. Let $(\delta_{i,t}^*, \gamma_{i,t}^*, S_t^*, Q_{t|t}^*, Q_{t|t-1}^*)$ be an optimal solution to (12) attaining the optimal value J_2^*. By Claim 1,

$$\begin{aligned} 0 &\preceq \begin{bmatrix} \delta_{i,t}^* - \gamma_{i,t}^* & \delta_{i,t}^* C_i \\ C_i^\top \delta_{i,t}^* & Q_{t|t-1}^* + C_i^\top \delta_{i,t}^* C_i \end{bmatrix} \\ &\preceq \begin{bmatrix} \delta_{i,t}^* - \gamma_{i,t}^* & \delta_{i,t}^* C_i \\ C_i^\top \delta_{i,t}^* & Q_{t|t-1}^{**} + C_i^\top \delta_{i,t}^* C_i \end{bmatrix}, \\ 0 &\preceq \begin{bmatrix} S_t^* & I \\ I & Q_{t|t}^* \end{bmatrix} \preceq \begin{bmatrix} S_t^* & I \\ I & Q_{t|t}^{**} \end{bmatrix}. \end{aligned}$$

Also, by construction, $(Q_{t|t}^{**}, Q_{t|t-1}^{**})$ satisfies the equality constraints (13). This implies that $(\delta_{i,t}^*, \gamma_{i,t}^*, S_t^*, Q_{t|t}^{**}, Q_{t|t-1}^{**})$ is a feasible solution to (11) and it attains the value J_2^*. This implies that $J_1^* \leq J_2^*$. However, since (12) has less stringent constraint than (11), $J_1^* \geq J_2^*$ must hold. Therefore, $J_1^* = J_2^*$.

E. Truss bridge

The mass and stiffness matrices are expressed as

$$M = \frac{\rho A l}{6} \begin{bmatrix} 2c_x c_x & 2c_x c_y & c_x c_x & c_x c_y \\ 2c_y c_x & 2c_y c_y & c_y c_x & c_y c_y \\ c_x c_x & c_x c_y & 2c_x c_x & 2c_x c_y \\ c_y c_x & c_y c_y & 2c_y c_x & 2c_y c_y \end{bmatrix},$$

$$K = \frac{AE}{l} \begin{bmatrix} c_x c_x & c_x c_y & -c_x c_x & -c_x c_y \\ c_y c_x & c_y c_y & -c_y c_x & -c_y c_y \\ -c_x c_x & -c_x c_y & c_x c_x & c_x c_y \\ -c_y c_x & -c_y c_y & c_y c_x & c_y c_y \end{bmatrix}$$

where ρ, A, l, and E are density, area, length, and modulus of elasticity for the member, respectively. c_x and c_y are the directional cosines with respect to the two-coordinates axis. Material of the truss bridge is assumed steel and density and elastic modulus values are applied based on this assumption. In addition, we consider area of each element is same and values are

$$E = 210\ Gpa,\ \rho = 7850\ kg/m^3,\ A = 0.9\ m^2$$