Non-Myopic Sensor Scheduling for Linear Systems with Colored Noise

Kirsten Tuggle,* Donghae Kim,† Maruthi Akella,‡ and Takashi Tanaka§
The University of Texas at Austin, Austin, Texas 78712

https://doi.org/10.2514/1.G008324

This paper addresses the task of sensor selection over a finite time horizon for systems modeled via discrete-time, linear state-space representations. Our method for this general linear setting accommodates both spatial and temporal noise correlations. To our best knowledge, this is the first work to do so. Scheduling policies are designed to limit sensor usage and minimize a minimum-mean-square-error-based criterion with time-varying weights to accommodate different user scenarios (e.g., prioritizing certain state elements at certain times or performing linear quadratic Gaussian control). The approach is also nonmyopic since the effects of sensor activations on all time steps are incorporated. A new but algebraically equivalent formulation of the scheduling model is introduced that readily accounts for colored noise sequences. This lends a closed-form expression for the error covariance that is explicit in all scheduling variables. Such an expression had been considered intractable for filtering in both white and colored noise regimes. This expression is leveraged to develop a well-motivated surrogate objective function that is shown to be submodular, thus enabling the use of an efficient greedy algorithm accompanied by performance guarantees with respect to the surrogate objective. Numerical examples are provided to demonstrate the effectiveness of the proposed methodology.

I. Introduction

LARGE collections of sensors are ubiquitous in modern estimation and control endeavors, but the ability to leverage these resources under finite power, energy, and computing constraints is driven by practical and intelligent sensor management. Some applications include space object tracking [1], persistent surveillance, and wireless sensor networks. Although modern applications generally enjoy increased computational power, the amount of data available can quickly overwhelm the ability to locally process it as the problem complexity increases over time and the number of data sources. Sensor selection refers to the class of problems concerned with optimally selecting subsets of available sensing resources in order to best perform an estimation task.

This work considers sensor selection for discrete, linear systems over a finite horizon subject to additive Gaussian noises (which need not be white). Based on known measurement models, the user must select a subset of resources such that minimum-mean-square-error (MMSE) estimation is achieved given a cardinality constraint on the sensing subset. Linear combinations of the state may be the subject of the estimation task. This results in the ability to optimize the mean square error (MSE) with respect to specific time steps and/or linear functions of the state most relevant to the estimation or control application. This problem setting is most closely related to [2], but we importantly accommodate both spatially correlated (dependence across sensors) and temporally correlated (colored) noise sequences.

There exist few papers considering dependent measurement noise across sensors [2–5] and, to the authors' knowledge, no work that considers temporal dependence. Nontrivial correlations may be reasonably expected across large collections of sensors. In [3], it was demonstrated that schedulers that neglect strong correlations between sensors or approximate them as weak will show significantly degraded performance. Though correlation models may not be explicitly known across every sensor pair, those that are known may significantly improve scheduling performance if intelligently incorporated. Colored noises are similarly possible for practical distributed sensor networks [6,7]. Common positioning systems such as GPS are often subject to time-correlated noise. Certainly, methods are well known for adapting the standard Kalman filtering framework to account for colored noise, e.g., state augmentation [8] or measurement differencing [9,10]. These only apply once a set of sensing resources has been selected.

A key reason that nearly all existing sensor selection routines hinge on the assumption of uncorrelated noises [11–17] is the complicated evolution of the covariance matrices in terms of the schedules. It has been generally accepted that explicit expressions for the covariance (or information) matrices in terms of all schedules over time would be infeasible to obtain [3] in the presence of noise correlations. This is likely attributed to the tendency in the literature to work with the recursive Kalman filter, which by definition obscures the schedules at previous times within the recursion. For example, a recursive solution for covariance $P_{t|t}$ at time step t is an explicit function of covariance $P_{t-1|t-1}$ at time step $t-1$, measurement model matrices at time step t, and scheduling variables at time step t. All previous scheduling variables are implicitly represented within the term $P_{t-1|t-1}$. Efforts to expand the recursion for the purpose of exposing explicit dependence on all past schedules amount to reconstruction of the batch solution. This holds in the case of colored noise with the use of the augmented state Kalman filter. For this reason we choose to work with the equivalent, batch linear MMSE (LMMSE) equations. Though this means that the scheduling algorithm invokes terms from batch updates, a user is not required to operate a batch filter in execution. As a bonus feature, this solution easily lends itself to applications employing smoothers and those containing correlations between measurement and process noises. Note that batch estimation was considered in [15,16] but for a fundamentally different problem: estimation of the entire state history $[\boldsymbol{x}_1^T \ \ldots \ \boldsymbol{x}_t^T]^T$ using the entire measurement history $[\boldsymbol{y}_1^T \ \ldots \ \boldsymbol{y}_t^T]^T$ (smoothing). In addition to pivoting to the batch form, we also construct a new but equivalent update based on the binary scheduling variables. Without loss of generality,

Received 4 March 2024; accepted for publication 23 September 2024; published online 12 December 2024.

*Department of Aerospace Engineering and Engineering Mechanics; ktuggle@utexas.edu. Member AIAA.

†Department of Aerospace Engineering and Engineering Mechanics; donghae.kim@utexas.edu. Student Member AIAA.

‡Professor, Cockrell Family Chair in Engineering #19, Department of Aerospace Engineering and Engineering Mechanics; makella@mail.utexas.edu. Fellow AIAA.

§Associate Professor, Department of Aerospace Engineering and Engineering Mechanics; ttanaka@utexas.edu; also Associate Professor, The School of Aeronautics and Astronautics and the Elmore Family School of Electrical and Computer Engineering, Purdue University; tanaka16@purdue.edu. Member AIAA.

the schedules act on both the measurement model (what we process) and the filter itself (how we process). In Sec. III, these choices result in the closed-form expression for the covariance matrices thought to be unattainable.

This expression is crucial to our solution avenue for multiple reasons. We first use it to classify the problem as a mixed-integer semidefinite programming (MISDP) problem in Sec. IV. It is known that MSE-based scheduling problems are NP-hard, with no polynomial-time solution having a constant factor approximation [18,19]. In [2], a semidefinite programming (SDP) relaxation was adopted for the white noise case. Although potentially suitable for certain smaller-scale applications, SDP relaxation methods are recognized to scale poorly in both computational complexity and estimation performance as the number of available schedules increases [20]. A greedy-algorithm-based restriction to the prior information matrix was provided in [21]. We again leverage our analytical expression for MSE to pursue the greedy optimization avenue in Sec. V.

Our developed surrogate objective function is directly motivated by the explicit form of scheduling variables in the cost. The greedy algorithm may then be applied with well-known polynomial time guarantees [22] and performance bounds (with respect to the surrogate). This avenue is similar to that taken in Refs. [3,14,20,23,24]. All objectives for these approaches were proposed for specializations of the most general problem and designed without the benefit of the explicit MSE scheduling expression. In Sec. VII, a comparison with some of these methods is presented. Alternatively, in [20], the authors explicitly formulate the objective function using the Fisher information matrix and demonstrate a randomized greedy solution to alleviate computational burdens. The randomized greedy algorithm is further investigated in [24] to address the scenarios involving budget or performance constraints. However, these approaches are restricted to weak-modular objective functions with upper-bounded curvatures.

The contributions of this paper are as follows:

1) Few solutions exist for nonmyopic sensor scheduling with spatially correlated sensors. To our best knowledge, none exist for those with temporal correlation (i.e., colored noise). Difficulty has been attributed to the absence of a closed-form expression for the estimation error covariance in terms of the binary scheduling variables. We provide this result in Theorem 1, which enables a solution accommodating both spatially and temporally correlated noise.

2) Time- and sensor-specific weights are incorporated to address application priorities.

3) The performance of the proposed greedy algorithm is proved with respect to the surrogate objective. Performance with respect to the true objective is compared in simulation against direct extensions of popular methods for the independent noise case. Our solution compared favorably, particularly when application-specific weights are imposed.

4) This work may be summarized as advancing the colored noise case from omission to analogy with popular approaches (MISDP and submodular surrogate greedy) in the greatly studied independent noise case. Following the state of that area, we anticipate using Theorem 1 to develop a stochastic greedy approach and investigate weak submodularity of the pure greedy solution. Finally, we expect and hope this theorem's result enables other authors to make additional advancements.

II. Problem Statement

The n-dimensional state vector at time step t to be estimated, denoted $\boldsymbol{x}_t$, is assumed to be governed by the following general, discrete-time linear dynamics:

$$\boldsymbol{x}_1 \sim N(\boldsymbol{\mu}_{1|0}, P_{1|0}) \tag{1a}$$

$$\boldsymbol{x}_{t+1} = A_t\boldsymbol{x}_t + B_t\boldsymbol{u}_t + \boldsymbol{w}_t \tag{1b}$$

for $\forall t \in \{1, 2, \ldots, t_f - 1\}$. Matrices A_t are assumed to be invertible. This is commonly satisfied when the discrete dynamics in Eq. (1b) are derived from a continuous system. In that case, A_t is the state transition matrix [25] carrying the property of invertibility. Tracking scenarios, like that explored in Sec. VII.B, often fall in this category [26,27]. The notation $P_{i|j}$ indicates the error covariance matrix for the state at time step i using all measurements up to and including time step j, i.e., $P_{i|j} = \mathbb{E}\{[\boldsymbol{x}_i - \hat{\boldsymbol{x}}_{i|j}][\boldsymbol{x}_i - \hat{\boldsymbol{x}}_{i|j}]^T\}$, where $\hat{\boldsymbol{x}}_{i|j}$ is the estimate of $\boldsymbol{x}$ with the given measurements up to and including time step j at time step i. Note that hereafter we ignore control input $\boldsymbol{u}_t$ without loss of generality since the system is linear and only state covariance matrices and not state estimates themselves are optimized in scheduling. The process noise autocovariance is defined to be $\mathbb{E}\{\boldsymbol{w}_i\boldsymbol{w}_j^T\} = \delta_{i,j}W_i$. Here $\delta_{i,j}$ represents the Kronecker delta. The collection of all potential observations (measurements) at each time step, $\boldsymbol{y}_t$, is

$$\boldsymbol{y}_t = \begin{bmatrix} C_{t,1}\boldsymbol{x}_t \\ \vdots \\ C_{t,m_t}\boldsymbol{x}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{v}_{t,1} \\ \vdots \\ \boldsymbol{v}_{t,m_t} \end{bmatrix} = C_t\boldsymbol{x}_t + \boldsymbol{v}_t \tag{2}$$

for $\forall t \in \{1, \ldots, t_f\}$. Note that m_t is the total number of available sensors at time step $t \le t_f$. Matrix $C_{t,i}$ is the measurement mapping matrix for sensor i at time step t. An augmented matrix $C_t = [C_{t,1}^T \quad \cdots \quad C_{t,m_t}^T]^T$ is the collection of all available mapping matrices at time step t. Similar notation is used for measurement noises $\boldsymbol{v}_{t,i}$ and the augmented measurement noise vector $\boldsymbol{v}_t$. Note that vector observations are allowed. The measurement noise sequences are assumed to be zero-mean and Gaussian but not necessarily white or independent between "sensors" (which, as previously stated, may literally correspond to different sensors or rather different potential sensing placements or modes of operation). The collection of measurements noise is

$$\boldsymbol{v}_t \sim N(\boldsymbol{0}, V_t) \tag{3}$$

Note that matrix V_t is allowed to have off-diagonal entries, which capture any spatial correlations. The following common model [6] can be applied for any sensor or process exhibiting colored noise:

$$\boldsymbol{v}_{t+1,i} = \Psi_{t,i}^v\boldsymbol{v}_{t,i} + \boldsymbol{\xi}_{t,i}^v, \quad \boldsymbol{w}_{t+1} = \Psi_t^w\boldsymbol{w}_t + \boldsymbol{\xi}_t^w \tag{4}$$

where $\Psi_{t,i}^v$ and Ψ_t^w are any real, conformable matrices for all time steps $\forall t \in \{1, 2, \ldots, t_f - 1\}$ and sensors $\forall i \in \{1, 2, \ldots, m_t\}$. The process $\{\boldsymbol{\xi}_{t,i}\}_{t=1}^{t_f}$ is a Gaussian white noise sequence with autocovariance $\mathbb{E}\{\boldsymbol{\xi}_{j,i}\boldsymbol{\xi}_{k,i}^T\} = \delta_{j,k}R_j$ such that Eq. (3) is satisfied and $\mathbb{E}\{\boldsymbol{\xi}_{t+1,i}\boldsymbol{v}_{t,i}^T\} = \boldsymbol{0}$.

Our objective is to perform the following optimization:

$$\begin{aligned} \min_{\{\{d_{t,i}\}_{i=1}^{m_t}\}_{t=1}^{t_f}} \quad & \sum_{t=1}^{t_f} \operatorname{tr}\left(\Theta_t P_{t|t}\Theta_t^T\right) \\ \text{s.t.} \quad & d_{t,i} \in \{0, 1\} \\ & \sum_{t=1}^{t_f}\sum_{i=1}^{m_t} d_{t,i} \le n_s \end{aligned} \tag{5}$$

where the binary scheduling variables represent

$$d_{t,i} = \begin{cases} 0 & \text{The measurement at time-step } t \text{ for sensor } i \textbf{ is not} \text{ selected.} \\ 1 & \text{The measurement at time-step } t \text{ for sensor } i \textbf{ is} \text{ selected} \end{cases} \tag{6}$$

Weighting matrices Θ_t are determined by the user with no dependence on the schedules, and covariance matrices $P_{t|t}$ are implicit functions of the schedules—these functions will be made explicit in the next section. The natural number n_s indicates the maximum number of allowable sensor observations used. As shown in [2], this optimization problem addresses a large number of relevant applications, including linear-quadratic-Gaussian (LQG) control [28]. More applications are summarized in Table 1. Clearly, flexibility in the weighting matrices enables applicability to a wide variety of scheduling scenarios.

Table 1 Choice of weighting matrices for different applications

MSE cost function	Corresponding weight matrices
Average over time horizon	$\Theta_t = \frac{1}{\sqrt{t_f}} I_n \quad \forall t$
At final time	$\Theta_1 = \ldots = \Theta_{t_f-1} = 0_{n\times n}, \Theta_{t_f} = I_n$
LQG estimation and control	Solution of a backward Riccati recursion [28]
Total MSE for state element i over time	$\Theta_t = \boldsymbol{e}_i^T \forall t$ (ith unit vector in canonical basis) or $[\Theta_t]_{j,k} = \begin{cases} 1 & j=k=i \\ 0 & \text{otherwise} \end{cases} \quad \forall t$

III. Derivation of the Covariance Matrices as Explicit Functions of the Scheduling Variables

We provide an explicit expression for each $P_{t|t}$ in terms of all scheduling variables in Theorem 1. This conclusion not only aids the development of our algorithm but also is a standalone contribution.

Theorem 1 (Covariance Functions of All Schedules): The state covariance at each time step can be expressed in the following form:

$$P_{t|t} = P_{t|0} - (P_{t|0}H_t^T + F_t)G_t(H_tP_{t|0} + F_t^T) + (P_{t|0}H_t^T + F_t)G_t\left(G_t + \frac{1}{\alpha_t}D_t\right)^{-1}G_t(H_tP_{t|0} + F_t^T) \quad (7)$$

where the schedules only appear in diagonal matrix D_t that is populated by all schedules through time step t. Matrices H_t, F_t, G_t, and scalar α_t, which do not depend on the schedules, are defined in the proof. These matrices and others used in the final algorithm are computed recursively in Appendix B.

Proof:

1) *Batch formulation*. There is an equivalence between the batch and recursive forms of the LMMSE estimator, assuming that the user correctly modifies the filtering equations to handle colored noise (e.g., via state augmentation or measurement differencing). This property allows the use of the batch equations for the purpose of the proof without loss of generality. The batch form after substituting the dynamics in Eq. (16) is

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_t \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} C_{1,1}A_1^{-1}\ldots A_{t-1}^{-1} \\ \vdots \\ C_{1,m_1}A_1^{-1}\ldots A_{t-1}^{-1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} C_{t,1} \\ \vdots \\ C_{t,m_t} \end{bmatrix} \end{bmatrix} \boldsymbol{x}_t + \begin{bmatrix} \begin{bmatrix} \boldsymbol{v}_{1,1} - C_{1,1}(A_1^{-1}\ldots A_{t-1}^{-1}\boldsymbol{w}_{t-1} + \ldots + A_1^{-1}\boldsymbol{w}_1) \\ \vdots \\ \boldsymbol{v}_{1,m_1} - C_{1,m_1}(A_1^{-1}\ldots A_{t-1}^{-1}\boldsymbol{w}_{t-1} + \ldots + A_1^{-1}\boldsymbol{w}_1) \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \boldsymbol{v}_{t,1} \\ \vdots \\ \boldsymbol{v}_{t,m_t} \end{bmatrix} \end{bmatrix} \quad (8)$$

$$\text{or,} \quad \tilde{\boldsymbol{Y}}_t = H_t\boldsymbol{x}_t + \tilde{\boldsymbol{\nu}}_t \quad (9)$$

In the last line, the augmented vector $\tilde{\boldsymbol{Y}}_t = [\,\boldsymbol{y}_1^T \quad \cdots \quad \boldsymbol{y}_t^T\,]^T$ denotes the collection of all available measurements up to and including time step t. The augmented matrix H_t is the corresponding mapping matrix that relates all causal measurements to the state at the current time, and $\tilde{\boldsymbol{\nu}}_t$ is the corresponding augmented measurement noise vector. Note that colored measurement or process noise can easily be incorporated by substitution of Eq. (4) into Eq. (9), with the only effect being additional correlations within $\tilde{\boldsymbol{\nu}}_t$. By grouping and renaming terms, the process noise terms now appear as measurement noise. The (correlated) noise covariance now includes both V and W terms. An effect of process noise in batch solutions is measurement noise correlations between both individual measurement noise terms (no diagonal covariance) and the state (since $\boldsymbol{x}_t$ is a function of $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{t-1}$).

2) *Novel model for scheduling variables*. Using the selection variable $d_{t,i}$, we will now introduce the scheduling matrix. For convenience, let $\eta_{t,m}$ denote the number of rows of $C_{t,m}$. Now, define the scheduling matrix D_t as follows:

$$D_t = \text{blkdiag}\left(d_{1,1}I_{\eta_{1,1}}, d_{1,2}I_{\eta_{1,2}}, \ldots, d_{t,m_t-1}I_{\eta_{t,m_t-1}}, d_{t,m_t}I_{\eta_{t,m_t}}\right) \quad (10)$$

where blkdiag$(\cdot)$ represents the block-diagonal matrix composed of inputs. Furthermore, define the effective selection matrix $\tilde{D}_t$ by removing rows that only contain zero in D_t. Note that the two matrices satisfy the following relations:

$$\tilde{D}_t\tilde{D}_t^T = I_\eta, \tilde{D}_t^T\tilde{D}_t = D_t \quad (11)$$

where $\eta = \sum_{\tau=1}^{t}\sum_{m=1}^{m_t} d_{\tau,m}\eta_{\tau,m}$. Recall that the goal is the explicit appearance of all schedules in the covariance equation. Using a binary selection matrix $\tilde{D}_t$, the selected sensor output $\boldsymbol{Y}_t$ can be expressed as follows:

$$\boldsymbol{Y}_t = \tilde{D}_tH_t\boldsymbol{x}_t + \tilde{D}_t\tilde{\boldsymbol{\nu}}_t \quad (12)$$

For the notational brevity, let us define the selected noise vector $\boldsymbol{\nu}_t = \tilde{D}_t\tilde{\boldsymbol{\nu}}_t$. Now, the output model Eq. (12) comes with additional specifications:

$$\Sigma_{\boldsymbol{x}_t\boldsymbol{\nu}_t} = \mathbb{E}\{[\boldsymbol{x}_t - \bar{\boldsymbol{x}}_t][\boldsymbol{\nu}_t - \bar{\boldsymbol{\nu}}_t]^T\} = F_t\tilde{D}_t^T \quad (13)$$

$$\Sigma_{\boldsymbol{\nu}_t\boldsymbol{\nu}_t} = \mathbb{E}\{\boldsymbol{\nu}_t\boldsymbol{\nu}_t^T\} = \alpha_tI + \tilde{D}_tS_t\tilde{D}_t^T \quad (14)$$

where $\bar{\boldsymbol{x}}_t$ and $\bar{\boldsymbol{\nu}}_t$ represent the mean of $\boldsymbol{x}$ and $\boldsymbol{\nu}$, respectively. Matrix F_t equals $\Sigma_{\boldsymbol{x}_t\tilde{\boldsymbol{\nu}}_t}$ evaluated as if all measurements are included. Quantity $\alpha_t > 0$ and positive definite matrix S_t are defined such that $\alpha_tI + S_t = \Sigma_{\tilde{\boldsymbol{\nu}}_t\tilde{\boldsymbol{\nu}}_t}$. This linear algebra property ensures the invertibility of the innovation covariance for the model with schedules included, and it was also used in [3].

3) *LMMSE update*. We now perform the update for $P_{t|t}$ via the usual LMMSE equations found in §3.3.2 of [29].

$$P_{t|t} = P_{t|0} - \Sigma_{\boldsymbol{x}_t\boldsymbol{Y}_t}\Sigma_{\boldsymbol{Y}_t\boldsymbol{Y}_t}^{-1}\Sigma_{\boldsymbol{Y}_t\boldsymbol{x}_t} \quad (15)$$

We have

$$\Sigma_{\boldsymbol{x}_t\boldsymbol{Y}_t} = \mathbb{E}\{[\boldsymbol{x}_t - \bar{\boldsymbol{x}}_t][\tilde{D}_tH_t(\boldsymbol{x}_t - \bar{\boldsymbol{x}}_t) + \boldsymbol{\nu}_t]^T\} \quad (16)$$

$$= P_{t|0}H_t^T\tilde{D}_t^T + F_t\tilde{D}_t^T = (P_{t|0}H_t^T + F_t)\tilde{D}_t^T \quad (17)$$

$$\Sigma_{\boldsymbol{Y}_t\boldsymbol{Y}_t} = \mathbb{E}\{[\tilde{D}_tH_t(\boldsymbol{x}_t - \bar{\boldsymbol{x}}_t) + \boldsymbol{\nu}_t][\ldots]^T\} \quad (18)$$

$$= \tilde{D}_tH_tP_{t|0}H_t^T\tilde{D}_t^T + \tilde{D}_tH_tF_t\tilde{D}_t^T + \tilde{D}_tF_t^TH_t^T\tilde{D}_t^T + \alpha_tI + \tilde{D}_tS_t\tilde{D}_t^T \quad (19)$$

$$= \alpha_tI + \tilde{D}_tG_t^{-1}\tilde{D}_t^T \quad (20)$$

where

$$G_t = (H_t P_{t|0} H_t^T + H_t F_t + F_t^T H_t^T + S_t)^{-1} \tag{21}$$

to finally find

$$P_{t|t} = P_{t|0} - R_t \tilde{D}_t^T (\alpha_t I + \tilde{D}_t G_t^{-1} \tilde{D}_t^T)^{-1} \tilde{D}_t R_t^T \tag{22}$$

with $R_t = (P_{t|0} H_t^T + F_t)$. The Sherman–Woodbury–Morrison lemma states that, for conformable matrices A, B, C, D with invertible A,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \tag{23}$$

$$\implies B(C^{-1} + DA^{-1}B)^{-1}D = A - A(A + BCD)^{-1}A \tag{24}$$

Apply the result in Eq. (24) to the term $\tilde{D}_t^T (\alpha_t I + \tilde{D}_t G_t^{-1} \tilde{D}_t^T)^{-1} \tilde{D}_t$ in Eq. (22) to finally get

$$P_{t|t} = P_{t|0} - R_t \left\{ G_t - G_t \left(G_t + \frac{1}{\alpha_t} \tilde{D}_t^T \tilde{D}_t \right)^{-1} G_t \right\} R_t^T \tag{25}$$

$$= P_{t|0} - R_t G_t R_t^T + R_t G_t \left(G_t + \frac{1}{\alpha_t} D_t \right)^{-1} G_t R_t^T \tag{26}$$

□

IV. Classification of the Optimization Problem

Using the expression in Eq. (7), the optimization problem (5) can be reformulated by dropping some constant terms as follows:

$$\begin{aligned} \min_{\{\{d_{t,i}\}_{i=1}^{m_t}\}_{t=1}^{t_f}} \quad & \sum_{t=1}^{t_f} \operatorname{tr}\left[\tilde{\Theta}_t \left(G_t + \frac{1}{\alpha_t} D_t \right)^{-1} \tilde{\Theta}_t^T \right] \\ \text{s.t.} \quad & d_{t,i} \in \{0, 1\} \\ & \sum_{t=1}^{t_f} \sum_{i=1}^{m_t} d_{t,i} \le n_s \end{aligned} \tag{27}$$

where $R_t = (P_{t|0} H_t^T + F_t)$ and $\tilde{\Theta}_t = \Theta_t R_t G_t$. Note that the cost for each time step given in Eq. (27) is of a similar overall form, with the schedules entering only as a diagonal matrix within an inverse expression, as was found for the static problem [3,17]. Even though the overall form is the same, the parameters involved (G_t, $\tilde{\Theta}_t$, and α_t) are not the same. Directly extending the approaches outlined in [3,17] for the static case would not have led to Eq. (27) for the general case. However, due to the shared general form, as in these other works, we can recognize that our optimization problem in Eq. (5) is a member of the MISDP class of optimization problems.

Define matrix $D := D_{t_f}$ by Eq. (10). This becomes the diagonal matrix containing all schedules over all times. In addition, note that $D_t = N_t D N_t^T$, where N_t is the truncating matrix to remove scheduling variables at time step $t+1$ to t_f. When an auxiliary symmetric matrix Z_t is introduced, Eq. (5) becomes

$$\begin{aligned} \min_{D \ge 0, Z_t} \quad & \sum_{t=1}^{t_f} \operatorname{tr}(Z_t) \\ \text{s.t.} \quad & I - D \succeq 0 \\ & \operatorname{tr}(D) \le n_s \\ & Z_t \succeq \tilde{\Theta}_t \left(G_t + \frac{1}{\alpha_t} N_t D N_t^T \right)^{-1} \tilde{\Theta}_t^T \end{aligned} \tag{28}$$

for $\forall t \in \{1, 2, \ldots, t_f\}$. Note that the second constraint in Eq. (28) can be expressed as $\mathbf{1}^T \operatorname{vec}(D)$, where $\mathbf{1}$ represents a vector of 1 and $\operatorname{vec}(\cdot)$ is the operator vectorizing diagonal elements of an input matrix. Via Schur decomposition for the third constraint in Eq. (28), we may then express the optimization in Eq. (5) equivalently as

$$\begin{aligned} \min_{D \ge 0, Z_t} \quad & \sum_{t=1}^{t_f} \operatorname{tr}(Z_t) \\ \text{s.t.} \quad & I - D \succeq 0 \\ & \mathbf{1}^T \operatorname{vec}(D) \le n_s \\ & \begin{bmatrix} Z_t & \tilde{\Theta}_t \\ \tilde{\Theta}_t^T & G_t + \frac{1}{\alpha_t} N_t D N_t^T \end{bmatrix} \succeq 0 \end{aligned} \tag{29}$$

for $\forall t \in \{1, 2, \ldots, t_f\}$. One traditional approach to solve the SDP is relaxing the domain of schedules within $[0, 1]$, and recovering the solution in the binary domain. Some popular mapping routines are randomized rounding [30], semidefinite relaxation [31], and reweighted ℓ_1 minimization [32]. It is well known that this class of solutions scales poorly with the number of available sensors, in terms of both performance and computational complexity [20].

A more recent avenue pivots from optimization of real values to sets. Analogous to concavity/convexity in maximization/minimization of real-valued functions, submodularity/supermodularity imports efficient and provable solutions for maximization/minimization of set functions [12]. A seminal result in [33] established that if an optimization problem corresponds to submodular maximization over a uniform matroid, then the greedy algorithm applied to this problem is guaranteed to produce an objective value within at least $1 - 1/e$ of the true maximum. However, this problem involves maximization of the negative MSE, which is known to not be submodular [13]. This paper follows a solution avenue similar to that in Ref. [14], where a surrogate cost function is developed to attain this property. In doing so, we provide an algorithm that is much more efficient than MISDP solutions and comes with a performance guarantee with respect to the surrogate function. Though there is no similar guarantee with respect to the original objective function, this solution retains meaning based on the motivation of the surrogate function.

V. Development of a Surrogate Objective Function

As discussed in the previous section, a surrogate cost function provides a suboptimal solution of the exact optimization problem, as described in Eq. (5) or Eq. (29). In many references [13,14,16,34–36], the surrogate function chosen was $logdet$. There are a couple of impediments when applying this method to our particular application. First, the submodularity of $logdet$ has not been proven for the case of filtering problems over a time horizon with correlations between sensors and colored measurement noise. Second, $logdet$ does not handle the weighting matrices well. Suppose that $\Theta_{t|t}$ are square and nonsingular. Then,

$$J = \sum_{t=1}^{t_f} \left(\ell_n \left| \Theta_{t|t} P_{t|t} \Theta_{t|t}^T \right| \right) = \sum_{t=1}^{t_f} \left(\ell_n \left| \Theta_{t|t} \Theta_{t|t}^T \right| + \ell_n \left| P_{t|t} \right| \right) \tag{30}$$

which is equivalent to optimizing

$$\sum_{t=1}^{t_f} \ell_n \left| P_{t|t} \right| \tag{31}$$

The weighting matrices, which importantly make the problem relevant for certain applications, are effectively ignored. Before developing a different objective, let us restate the true objective in the set context. The optimization problem (5) is equivalent to

$$\begin{aligned} \max_{\mathcal{S} \subseteq E} \quad & -\sum_{t=1}^{t_f} \operatorname{tr}\left(\Theta_t P_{t|t} \Theta_t^T \right) \\ \text{s.t.} \quad & |\mathcal{S}| \le n_s \end{aligned} \tag{32}$$

where set E is the collection of all available sensors at all available times, i.e., $d_{i,t} \in E$ for $\forall i \in \{1, 2, \ldots, m_t\}$ and $\forall t \in \{1, 2, \ldots, t_f\}$. Note that this problem is known not to be submodular. Significantly, we can leverage Eq. (7) to directly examine all scheduling variable terms in the true objective, interpret their role, and provide a more precisely motivated surrogate objective. The true objective is equivalent to

$$\mathcal{F}(\mathcal{S}) = -\sum_{t=1}^{t_f} \mathrm{tr}\left[M_t G_t \left\{G_t + \frac{1}{\alpha_t} D_t(\mathcal{S})\right\}^{-1} G_t M_t^T\right] \tag{33}$$

$$= -\sum_{t=1}^{t_f} \mathrm{tr}\left[M_t \left\{G_t^{-1} + \frac{1}{\alpha_t} G_t^{-1} D_t(\mathcal{S}) G_t^{-1}\right\}^{-1} M_t^T\right] \tag{34}$$

by using Eq. (24), where $M_t = \Theta_t(P_{t|0} H_t^T + F_t)$. $D_t(\mathcal{S})$ is the block diagonal matrix from before but now with the binary scheduling variables set to 0 or 1 based on the selected subset $\mathcal{S}$. This is only a subtly different parameterization. Next we will inspect the objective at each time step.

Denote the orthogonal eigendecomposition of the positive semidefinite matrix as $M_t^T M_t = U_t \Lambda_t U_t^T$. The columns of U_t contain the normalized, orthogonal eigenvectors of $M_t^T M_t$, denoted $\boldsymbol{u}_{t,j}$. The diagonal of the purely diagonal matrix Λ_t contains the eigenvalues $\lambda_{t,j}$ of this matrix in descending order. Define $\mathrm{rank}(M_t^T M_t) = r_t$ and $\tilde{G}_t(\mathcal{S}) = \{G_t^{-1} + (1/\alpha_t) G_t^{-1} D_t(\mathcal{S}) G_t^{-1}\}^{-1}$ for the notational brevity. Then the objective at time step t is

$$\mathcal{F}_t(\mathcal{S}) = -\mathrm{tr}\left[U_t \Lambda_t U_t^T \tilde{G}_t(\mathcal{S})\right] = -\mathrm{tr}\left[\Lambda_t U_t^T \tilde{G}_t(\mathcal{S}) U_t\right] \tag{35}$$

$$= -\mathrm{tr}\left(\begin{bmatrix} \lambda_{t,1} \boldsymbol{u}_{t,1}^T \\ \vdots \\ \lambda_{t,q_t} \boldsymbol{u}_{t,q_t}^T \end{bmatrix} \tilde{G}_t(\mathcal{S}) \begin{bmatrix} \boldsymbol{u}_{t,1}^T \\ \vdots \\ \boldsymbol{u}_{t,q_t}^T \end{bmatrix}^T\right) \tag{36}$$

$$= -\sum_{j=1}^{r_t} \lambda_{t,j} \boldsymbol{u}_{t,j}^T \tilde{G}_t(\mathcal{S}) \boldsymbol{u}_{t,j} \tag{37}$$

We can recognize that the true objective at each time step is equivalent to a weighted sum of projections of the matrix $\tilde{G}_t(\mathcal{S})$ along the eigendirections of $M_t^T M_t$, with the weights being the associated eigenvalues. This type of projection corresponds to what is used in principal components analysis (PCA) [37] for the variance associated with a particular direction with respect to the uncertainty matrix $\tilde{G}_t(\mathcal{S})$. There, the projection is geometric in the sense that it may be interpreted as taking the collection of zero-mean data (which may be infinite) whose covariance is represented by this matrix, orthogonally projecting all data points in the particular direction, and evaluating the variance along that one dimension. Roughly, it is akin to restricting uncertainty only to our particular direction by geometrically collapsing uncertainty in all other directions. Note that though this uncertainty matrix is obviously not equal to covariance $P_{t|t}$, we have shown it to be the sole piece of the covariance matrix relevant to optimization. Our surrogate function will arise from simply utilizing a different version of projecting this uncertainty.

Our substitute notion of directional uncertainty corresponds to the squared length of the $1-\sigma$ ellipse for uncertainty matrix $\tilde{G}_t(\mathcal{S})$ along a particular direction. This is equivalent to the variance along the one dimension once the random variable has been conditioned on all orthogonal directions [38], and it is akin to restricting uncertainty only to our particular direction by probabilistically collapsing uncertainty in all other directions. This quantity is easily found via the definition of error ellipses and is, for direction $\boldsymbol{u}$,

$$\sigma_{\boldsymbol{u}}^2 = \frac{1}{\boldsymbol{u}^T \left(G_t^{-1} + \frac{1}{\alpha_t} G_t^{-1} D_t(\mathcal{S}) G_t^{-1}\right) \boldsymbol{u}} \tag{38}$$

Thus, our surrogate version is

$$\tilde{\mathcal{F}}_t(\mathcal{S}) = -\sum_{j=1}^{r_t} \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T \left\{G_t^{-1} + \frac{1}{\alpha_t} G_t^{-1} D_t(\mathcal{S}) G_t^{-1}\right\} \boldsymbol{u}_{t,j}} \tag{39}$$

Recall Eq. (21) and importantly notice that computing $\tilde{F}_t(\mathcal{S})$ does not require any matrix inversion. Though the batch terms are used, carrying out the batch solution (which would invert likely very large matrices) is not required. We do note the computational degradation for some large-scale applications due to solving eigendecompositions to produce the λ and u terms. In such a case, the computational load can be relieved by adjusting the scheduling time horizon.

Without loss of generality, we can add a constant that is independent of all scheduling variables to Eq. (39). In order to satisfy normalization in the next section, we will do so in the following way:

$$\tilde{F}_t(\mathcal{S}) = -\sum_{j=1}^{r_t} \left[\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T \left\{G_t^{-1} + \frac{1}{\alpha_t} G_t^{-1} D_t(\mathcal{S}) G_t^{-1}\right\} \boldsymbol{u}_{t,j}} - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1} \boldsymbol{u}_{t,j}}\right] \tag{40}$$

The complete objective function is then

$$\tilde{\mathcal{F}}(\mathcal{S}) = -\sum_{t=1}^{t_f} \sum_{j=1}^{r_t} \left[\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T \left\{G_t^{-1} + \frac{1}{\alpha_t} G_t^{-1} D_t(\mathcal{S}) G_t^{-1}\right\} \boldsymbol{u}_{t,j}} - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1} \boldsymbol{u}_{t,j}}\right] \tag{41}$$

so that our surrogate optimization is

$$\max_{\mathcal{S} \subseteq E} \quad \tilde{\mathcal{F}}(\mathcal{S}) \text{s.t.} \quad |\mathcal{S}| \leq n_s \tag{42}$$

The proposed procedure is enumerated in Algorithm 1.

Algorithm 1: Proposed algorithm

Compute $\lambda_{t,j}, \boldsymbol{u}_{t,j}, G_t^{-1}, \alpha_t$ for $j = 1, \ldots, r_t$ (all sensors at t) and all $t = 1, \ldots, t_f$. ▷ using App. B.
$\mathcal{S} \leftarrow \varnothing$
for $l = 1, \ldots, n_s$ **do**
  $s_l \leftarrow \underset{s \in \mathcal{V} \setminus \mathcal{S}}{arg\,min}\, \tilde{\mathcal{F}}(\mathcal{S} \cup s)$ ▷ using Eq. (41) for $\tilde{\mathcal{F}}$
  $\mathcal{S} \leftarrow \mathcal{S} \cup s_l$
end for
return $\mathcal{S}$

VI. Submodularity of the Surrogate Objective

The submodularity of an objective function enables proven performance for the greedy algorithm [33]:

Property 1: If the set function f is to be maximized subject to a uniform matroid constraint and f is normalized, nonnegative, monotone nondecreasing, and submodular, then the greedy algorithm applied to maximization of f achieves the approximation factor $(1 - 1/e)$.

Proof that the chosen surrogate function Eq. (42) satisfies Property 1 is a fairly simple exercise following [33]. Due to the importance of the property for our solution, it is established in Appendix A. Therefore, the greedy solution to Eq. (42) produces an objective function value that is no less than $(1 - 1/e)$ of the true maximum. This gap is with respect to the surrogate rather than the original objective, for which it has been shown that no such constant factor performance guarantee can be achieved.

While the approximation ratio for the original problem cannot be established, the conceptual framework of the surrogate problem can be discussed. The constant term introduced for normalization in Eq. (40) is equivalent to the prenormalized surrogate described in Eq. (39) when $\mathcal{S} = \varnothing$, indicating that no sensors are utilized in the measurement update. Consequently, the surrogate objective function in Eq. (41) relates to the original objective function as follows:

$$\tilde{\mathcal{F}}(\mathcal{S}) = -\sum_{t=1}^{t_f}\sum_{j=1}^{r_t}\left[\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T\left\{G_t^{-1} + \frac{1}{\alpha_t}G_t^{-1}D_t(\mathcal{S})G_t^{-1}\right\}\boldsymbol{u}_{t,j}} - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j}}\right] \tag{43}$$

$$\underset{\text{surrogate}}{\approx} -\sum_{t=1}^{t_f}\operatorname{tr}\left[M_t\left\{G_t^{-1} + \frac{1}{\alpha_t}G_t^{-1}D_t(\mathcal{S})G_t^{-1}\right\}^{-1}M_t^T\right] + \sum_{t=1}^{t_f}\operatorname{tr}\left[M_t G_t M_t^T\right] \tag{44}$$

Additionally, since the first two terms in Eq. (7) are independent of the scheduling variable $D_t(\mathcal{S})$, Eq. (44) can be rewritten as

$$\tilde{\mathcal{F}}(\mathcal{S}) \underset{\text{surrogate}}{\approx} \sum_{t=1}^{t_f}\operatorname{tr}\left[\Theta_t\{P_{t|t}(D_t(\varnothing)) - P_{t|t}(D_t(\mathcal{S}))\}\Theta_t^T\right] = \sum_{t=1}^{t_f}\operatorname{tr}\left[\Theta_t\{P_{t|0} - P_{t|t}(D_t(\mathcal{S}))\}\Theta_t^T\right] \tag{45}$$

Thus, the surrogate problem defined in Eq. (42) approximates the problem of optimizing the expected MSE decrement by scheduling in the original domain. From Eq. (45), the performance guarantee outlined in Property 1 provides an approximated bound for

$$\frac{\sum_{t=1}^{t_f}(\operatorname{tr}[\Theta_t\{P_{t|0} - P_{t|t}(D_t(\mathcal{S}))\}\Theta_t^T])}{\sum_{t=1}^{t_f}(\operatorname{tr}[\Theta_t\{P_{t|0} - P_{t|t}(D_t(\mathcal{S}^\star))\}\Theta_t^T])} \underset{\text{surrogate}}{\approx} \frac{\tilde{\mathcal{F}}(\mathcal{S})}{\tilde{\mathcal{F}}(\mathcal{S}^\star)} \geq 1 - \frac{1}{e} \tag{46}$$

where $\mathcal{S}$ and $\mathcal{S}^\star$ represent the solution obtained from the greedy algorithm and the optimal solution, respectively.

VII. Numerical Examples

A. Linear Quadratic Gaussian Control on a Randomized System

A six-dimensional ($n = 6$) state is to be simultaneously estimated and controlled to the origin by minimizing the standard linear quadratic Gaussian cost:

$$\sum_{t=1}^{t_f-1}\mathbb{E}\left[\boldsymbol{x}_t^T Q_t \boldsymbol{x}_t + \boldsymbol{u}_t^T R_t \boldsymbol{u}_t\right] + \mathbb{E}\left[\boldsymbol{x}_{t_f}^T Q_{t_f}\boldsymbol{x}_{t_f}\right] \tag{47}$$

where $Q_t = R_t = I_6$ and $t_f = 10$. Dynamics and control matrices $A_t, B_t \in \mathbb{R}^{6\times 6}$ are randomly generated as independent samples of the standard normal distribution but kept constant for each time step. The process noise covariance $W_t = 0.01I_6$, and the a priori covariance for the state at the initial time is a diagonal matrix:

$$P_{1|0} = \operatorname{diag}([5 \quad 1 \quad 1 \quad 2 \quad 0.4 \quad 0.4]) \tag{48}$$

where diag(·) represents the diagonal matrix with the input diagonal entries. Collections of five scalar measurements with mapping matrices C_t that are independently drawn from distribution $N(0, 1/n)$ are available at each time step. A total of 10 sensor activations ($n_s = 10$) were allowed. To coincide with the standard LQG problem, measurement noise in this example is taken to be white. The sensors are spatially distributed in a 50-unit square region with the 2-D location of sensor i defined as $\boldsymbol{\beta}_i$. The measurements are spatially correlated at each time based on distances between sensors. This model is $[V_t]_{i,j} = \sigma_\nu^2 e^{-\rho_{\text{spatial}}\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|}$, where we use $\rho_{\text{spatial}} = 0.1$ and $\sigma_\nu = 1$ [3]. The LQG problem can be expressed as the minimization of our original cost in Eq. (5), with weighting matrices $\Theta_t \in \mathbb{R}^{6\times 6}$ defined via [28]

$$\Phi_t = \Theta_t^T\Theta_t = K_t^T\left(B_t^T Z_t + R_t\right)K_t \tag{49}$$

$$K_t = -\left(B_t^T Z_t B_t + R_t\right)^{-1}B_t Z_t A_t \tag{50}$$

$$Z_t = \begin{cases} Q_t & t = t_f \\ Q_t + L_t & t \in \{1, \ldots, t_f - 1\} \end{cases} \tag{51}$$

$$L_t = A_t^T\{Z_t - Z_t B_t(B_t^T Z_t B_t + R_t)^{-1}B_t^T Z_t\}A_t \tag{52}$$

Both Algorithm 1 and the *logdet* greedy algorithm corresponding to Eq. (31) were applied with the following results. Table 2 lists the MSE costs for both methods. One might expect the observed worse performance for *logdet* approaches to this particular problem due to the weighting matrices mentioned in Sec. V. The weighting matrices here play a crucial role by placing special emphasis on relevant directions in the state space and relevant time steps over the time horizon with respect to the LQG objective. For intuition on the latter, magnitudes of the weighting matrices over time are shown in Fig. 1. Also, the MSEs for each state component from x_1 to x_6 over time are provided in Fig. 2.

Table 2 Mean-square-error performance for example A

Solution method	Resulting cost in Eq. (5)
Algorithm 1	120,954.56
logdet	152,483.24

Based on Fig. 1, one would expect that a successful scheduling policy would heavily penalize state errors in earlier time steps with decreasing importance on time steps toward the end of the horizon. Notice that the proposed solution does a far better job in this respect by roughly keeping uncertainties relatively low with growth to peaks only at the end of the time horizon. Since the *logdet* solution effectively ignores the weights, it is solely concerned with minimizing covariance "sizes" (specifically, volumes) with no penalization to any specific time step beyond its connection to covariance volumes at the other time steps. Additionally, $P_{1|0}$ is initially such that the covariance ellipsoid volume can be drastically reduced by favoring measurements that primarily reduce uncertainty in x_1. As seen in Fig. 2, these conditions can lead to the *logdet* algorithm, allowing high uncertainties in state components other than x_1 in time steps close to the initial time—the very time steps that should be most penalized for the LQG objective.

B. Two-Dimensional Multitarget Tracking

In the previous example, weight matrices for the LQG are demonstrated. Now, selective states and time-varying weight matrices in Table 1 will be tested in this multi-object tracking example. Comparisons will be made between greedy solutions to the proposed, *logdet* (info greedy) and exact MSE (pure greedy) objectives. It is assumed that five targets circle around the origin with a constant radius and velocity, as shown in Fig. 3. The discretized dynamics of each target are defined as follows:

$$\boldsymbol{x}_{t+1}^{(i)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & dt \\ 0 & 0 & 1 \end{bmatrix}\boldsymbol{x}_t^{(i)} + \boldsymbol{w}_t^{(i)} = A^{(i)}\boldsymbol{x}_t^{(i)} + \boldsymbol{w}_t^{(i)} \tag{53}$$

with an initial state $\boldsymbol{x}_0^{(i)} = [r_0^{(i)}, \theta_0^{(i)}, \omega_0^{(i)}]^T$ for $\forall t \in \{0, 1, \ldots, t_f - dt\}$ and $\forall i \in \{1, 2, \ldots, 5\}$. Here, $\boldsymbol{x}^{(i)}$ represents the state vector of target i at time step t, which consists of the radius r, azimuth angle θ, and

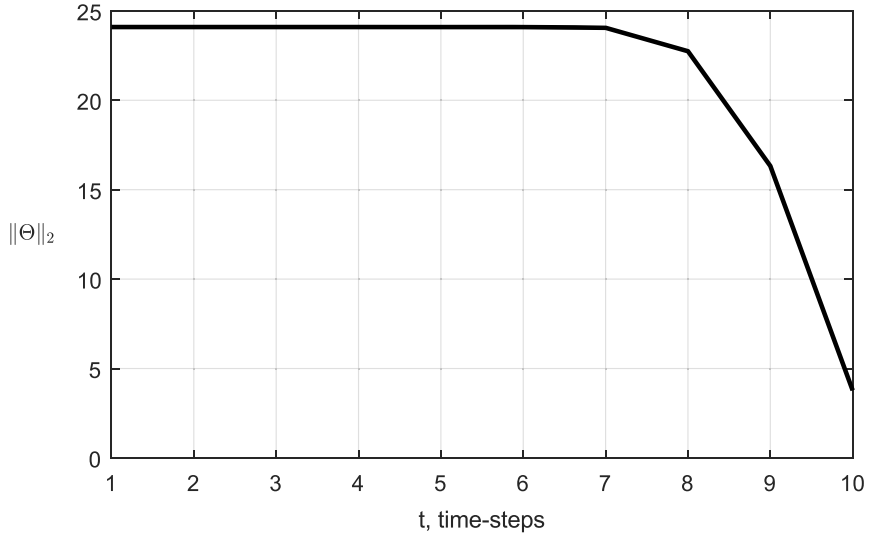

Fig. 1 The 2-norm of the LQG weighting matrices over time.

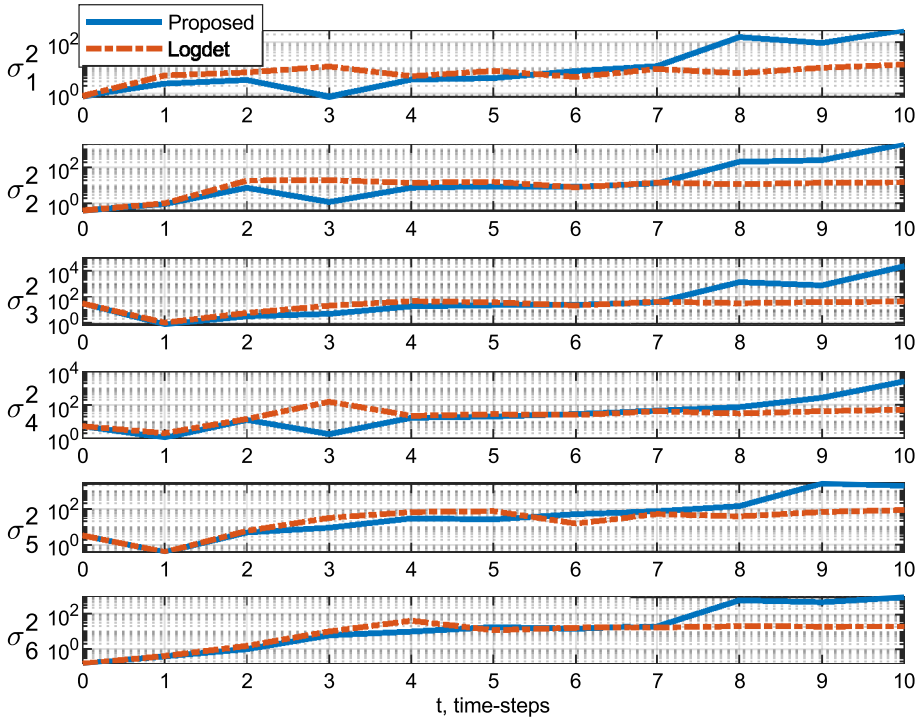

Fig. 2 State component mean square errors for the proposed vs *logdet* solution.

angular velocity ω. The dynamics of target i is denoted by $A^{(i)}$, and $\boldsymbol{w}_t^{(i)}$ is the process noise for target i at time step t.

The process $\{\boldsymbol{w}^{(i)}\}_{t=1}^{t_f}$ follows Gaussian white noise sequence with an autocovariance defined as $\mathbb{E}\{\boldsymbol{w}_j^{(i)}\boldsymbol{w}_k^{(i)T}\} = \delta_{j,k}Q^{(i)}$, where $Q^{(i)} = \text{diag}(Q_r, Q_\theta, Q_\omega)$. The state vector $\boldsymbol{x}_t$ and noise vector $\boldsymbol{w}_t$ encompass all targets, defined as $\boldsymbol{x}_t = [\boldsymbol{x}_t^{(1)T}, \ldots, \boldsymbol{x}_t^{(5)T}]^T$ and $\boldsymbol{w}_t = [\boldsymbol{w}_t^{(1)T}, \ldots, \boldsymbol{w}_t^{(5)T}]^T$, respectively. The overall dynamics matrix A naturally becomes a block-diagonal matrix of $A^{(i)}$ as follows:

$$A = \text{blkdiag}\left(A^{(1)}, \ldots, A^{(5)}\right) \tag{54}$$

It is important to note that the covariance of the process noise $\boldsymbol{w}_t$ is a constant diagonal matrix without any spatial or temporal correlation.

Multiple radar sensors are assumed to be installed at the origin. These sensors are either range radar or angle radar, capable of measuring either radius $r^{(i)}$ or azimuth angle $\theta^{(i)}$ of all targets. The output matrices of the range radar C_r and angle radar C_a at a specific time step t can be expressed as follows:

$$C_{r,t} = C_r = I_5 \otimes [1\,0\,0] \in \mathbb{R}^{5\times 15} \tag{55}$$

$$C_{a,t} = C_a = I_5 \otimes [0\,1\,0] \in \mathbb{R}^{5\times 15} \tag{56}$$

for $\forall t \in \{1, \ldots, t_f\}$, where $\otimes$ represents the Kronecker product. Both spatial and temporal correlations are assumed in the measurement noise. The same spatial correlation model described in the previous example is employed and applied to the output channels for the same target. Specifically, a spatial correlation coefficient of $\rho_{\text{spatial}} = 0.5$ is utilized separately for the range and angle radars. The coefficients $\boldsymbol{\beta}$ are randomly generated from a uniform distribution. Temporal correlation is also incorporated using $\Psi_t = 0.5$ to all sensors, as described in Eq. (4).

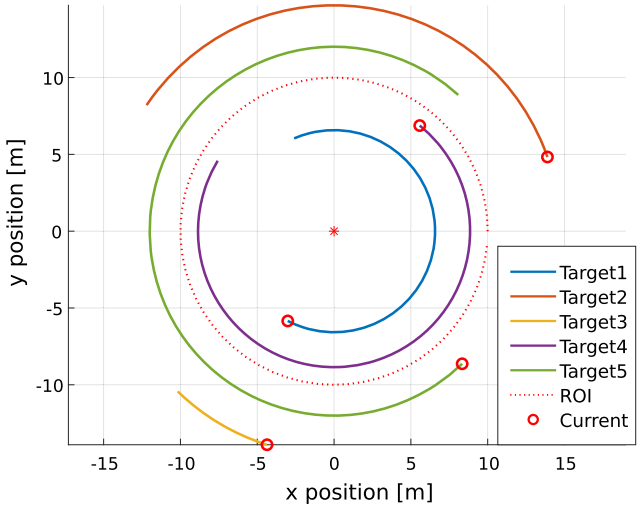

Fig. 3 Trajectory of all targets for 5 s.

The scheduling goal is to minimize MSE corresponding to the "orbit" states of each target. The radius of all targets is the primary state of interest, and the angular velocity is additionally considered if a target is inside the region of interest (ROI) indicated by the red dotted line in Fig. 3. In this example, the MSE of states $\omega^{(1)}$, $\omega^{(4)}$, and $r^{(i)}$ for $\forall i \in \{1, \ldots, 5\}$ are to be minimized. Thus, the default weight matrices $\bar{\Theta}$ is defined as follows:

$$\bar{\Theta} = \text{blkdiag}\Big(\text{diag}(1,0,1),\quad \text{diag}(1,0,0),\quad \text{diag}(1,0,0),\quad \text{diag}(1,0,1),\quad \text{diag}(1,0,0)\Big) \tag{57}$$

Two different weight matrices will be demonstrated in the following subsections. Simulation parameters are defined by $t_f = 5$, $n_s = 10$, $dt = 1$, and $n_r = n_a$, where n_r and n_a are the number of range radars and angle radars, respectively. The radius of ROI is 10 m. Prior, process noise, and measurement noise covariance are defined as follows:

$$P_{1|0}^{(i)} = \text{diag}\Big((1\text{ m})^2, (0.5\text{ rad})^2, (0.5\,\text{rad/s})^2\Big) \tag{58}$$

$$Q^{(i)} = \text{diag}\Big((0.5\text{ m})^2, (0.1\text{ rad})^2, (0.2\,\text{rad/s})^2\Big) \tag{59}$$

$$R_r^{(j)} = (1\text{ m})^2 \tag{60}$$

$$R_a^{(k)} = (0.2\text{ rad})^2 \tag{61}$$

where $i \in \{1, 2, \ldots, 5\}$, $j \in \{1, 2, \ldots, n_r\}$, and $k \in \{1, 2, \ldots, n_a\}$. Similar to the overall dynamics matrix A, overall covariance matrices can be generated by blkdiag$(\cdot)$ of Eqs. (58–61).

1. *Time-Weighted MSE*

First, time-weighted matrices Θ_t are assumed as follows:

$$\Theta_t = \frac{1}{t_f - t + 1}\bar{\Theta} \tag{62}$$

for $\forall t \in \{1, \ldots, t_f\}$. The weight matrices gradually grow as time step t approaches the final time t_f. Four different solutions are compared in Fig. 4 for $n_r = n_a \in \{1, 2, \ldots, 10\}$. Note that $logdet$ greedy applies the greedy algorithm to objective function (31). Pure and info greedy apply the greedy algorithm to the unmodified objective function (5) and inverse of the objective function, i.e., the Fisher information matrix, respectively. The proposed algorithm minimizes the time-weighted MSE more efficiently than $logdet$ for all cases. Though this is known to perform well in practice, theoretic performance bounds are

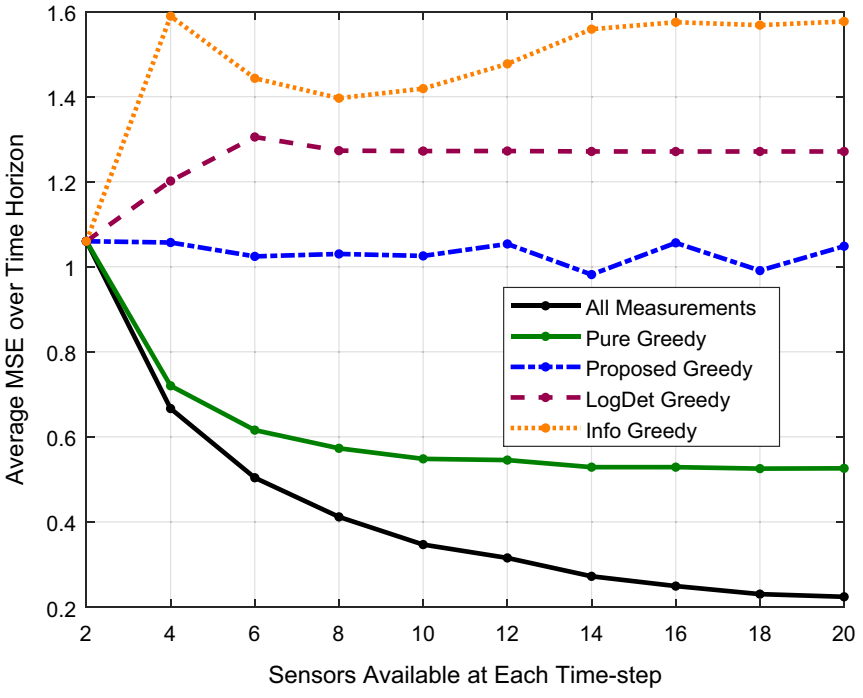

Fig. 4 Comparison of methods over a range of total sensors available at each time step with Θ_t in Eq. (62).

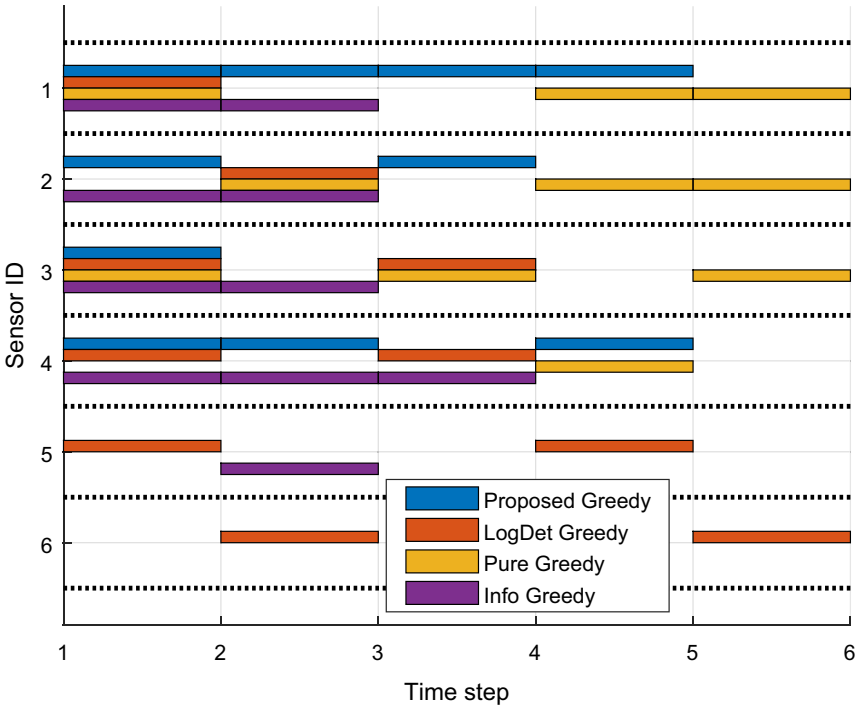

Fig. 5 Scheduling chart with six available sensors with Θ_t in Eq. (62).

lacking for cases beyond: restriction to a single time step with independent sensors [20] or multiple time steps but with spatially and temporally independent measurement noise [15].

The Gantt chart for the particular case of $n_r = n_a = 3$ is presented in Fig. 5. The first half of the sensor IDs represents range radars, while the second half represents angle radars. Since the default weight matrices $\bar{\Theta}$ described in Eq. (57) includes all radius states but the azimuth angle is included only for targets 1 and 4, it is expected that the optimal scheduler will select range radars more frequently. The scheduling chart shows that the proposed greedy algorithm assigns three task slots to angle radars, and the remaining to range radars. However, *logdet* greedy solution fails to weigh more on radial MSE due to its inability to account for the weight matrices, as shown in Eq. (31). By neglecting the weigh matrix Θ_t, the *logdet* greedy approach always chooses sensors in a scattered manner.

2. *Final Time MSE*

To observe the tasking behavior over time steps t instead of over sensors, more extreme weight matrices are assumed:

$$\Theta_t = \begin{cases} \bar{\Theta} & t = t_f \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{63}$$

In this case, it is natural for the optimal scheduler to select measurements near the final time t_f to minimize the MSE at the final time with the aforementioned weight matrices.

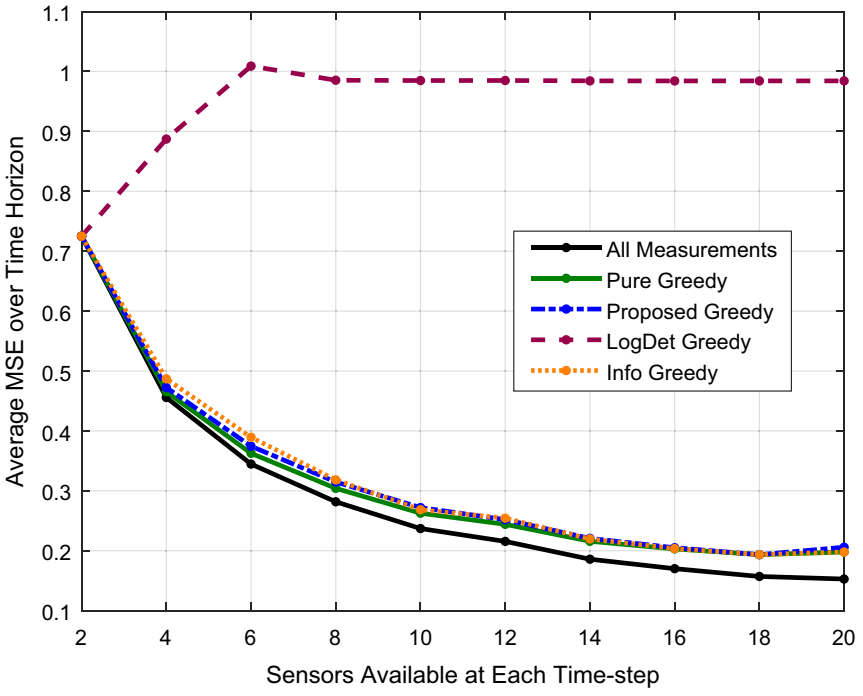

Fig. 6 Comparison of methods over a range of total sensors available at each time step with Θ_t in Eq. (63).

A similar comparison result is shown in Fig. 6. Except the *logdet* greedy solution, the remaining solutions exhibit parallel performance. Figure 7 presents the scheduling results for the particular case where $n_r = n_a = 3$. The first half of the sensors represents range radars, while the second half represents angle radars similarly. It is worth noting that the scheduling result of *logdet* greedy in Fig. 5 are identical to that in Fig. 7, as the weight matrices does not affect the objective function of the *logdet* surrogate problem. While the *logdet* greedy algorithm fails to adjust the scheduling with

Table 3 Mean-square-error performance for both Eq. (62) and Eq. (63) cases with $n_r = n_a = 2$

	Resulting cost in Eq. (8)	
Solution method	Case (1) Θ_t in Eq. (62)	Case (2) Θ_t in Eq. (63)
Optimum	3.5641	2.3262
Algorithm 1	5.2839	2.3606
logdet	6.0076	4.4354

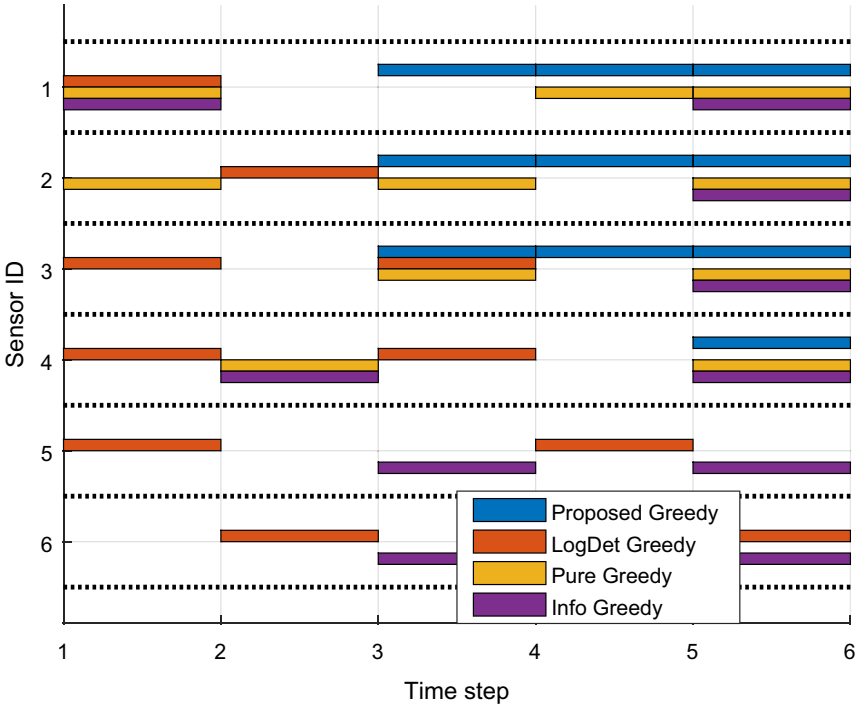

Fig. 7 Scheduling chart with six available sensors with Θ_t in Eq. (63).

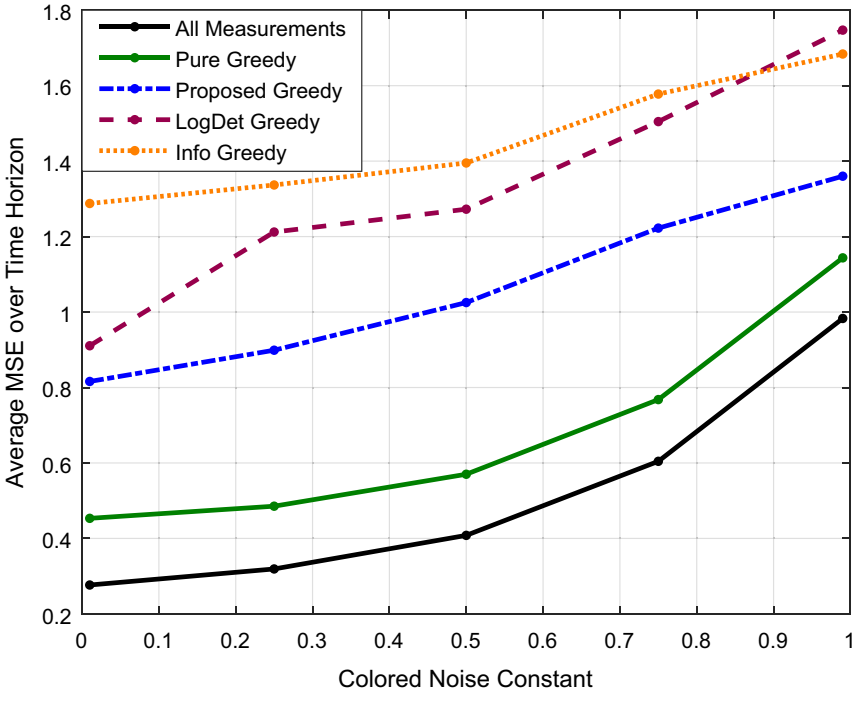

Fig. 8 Comparison of methods over a range of colored-noise constant Ψ_t.

respect to the weight matrices Θ_t, the proposed algorithm successfully reduces the MSE cost by distributing the schedules near the final time t_f, as well as among the first half of sensors, which corresponds to range radars.

The total cost of Eq. (5) using the *logdet* greedy and the proposed greedy, with $n_r = n_a = 2$, are presented in Table 3 for both of the weight matrices described in Eqs. (52) and (53). Compare to the *logdet* surrogate, the proposed algorithm reduces the total cost more efficiently by taking into account the time-varying weight matrices and achieves the cost values closer to the optimal solutions than *logdet* solutions.

3. *Effect of Temporal Correlation*

In this subsection, we demonstrate the effect of temporal correlation on the measurement noise. The target dynamics and measurement models remain identical to the previous example. The time-increasing weight matrices Θ_t is employed, as described in Eq. (52), and the number of range and that of angle sensors are assumed to be $n_r = n_a = 4$. We test temporal correlations over different values of $\Psi_t \in \{0.01, 0.25, 0.5, 0.75, 0.99\}$. The scheduling performance is illustrated in Fig. 8. In the figure, the first half of sensor IDs represents range radars, and the rest are angle radars similarly. Note that, as the temporal correlation gets stronger, the measurement uncertainty also

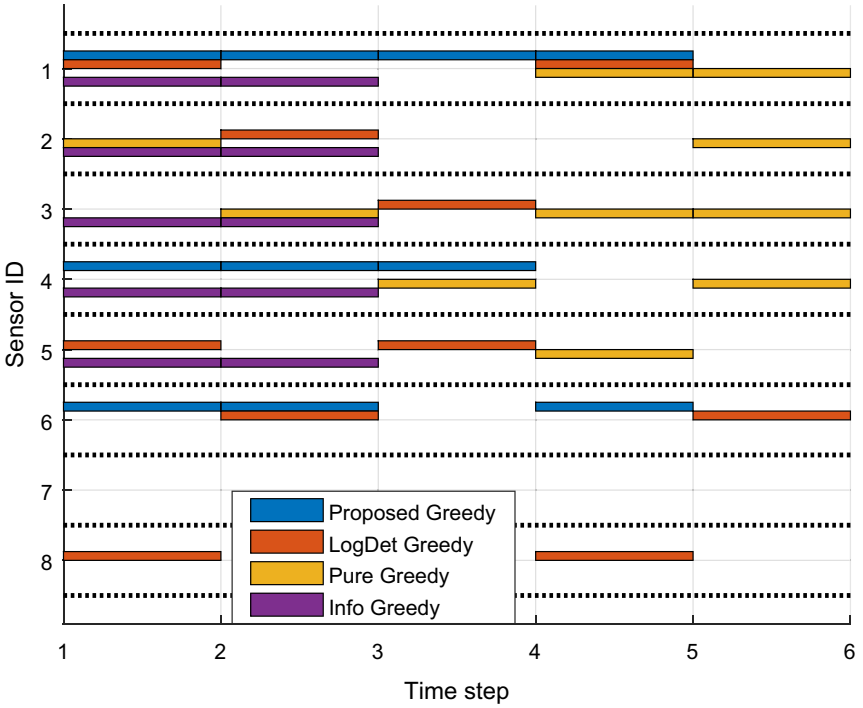

Fig. 9 Scheduling chart with eight available sensors and $\Psi_t = 0.01$.

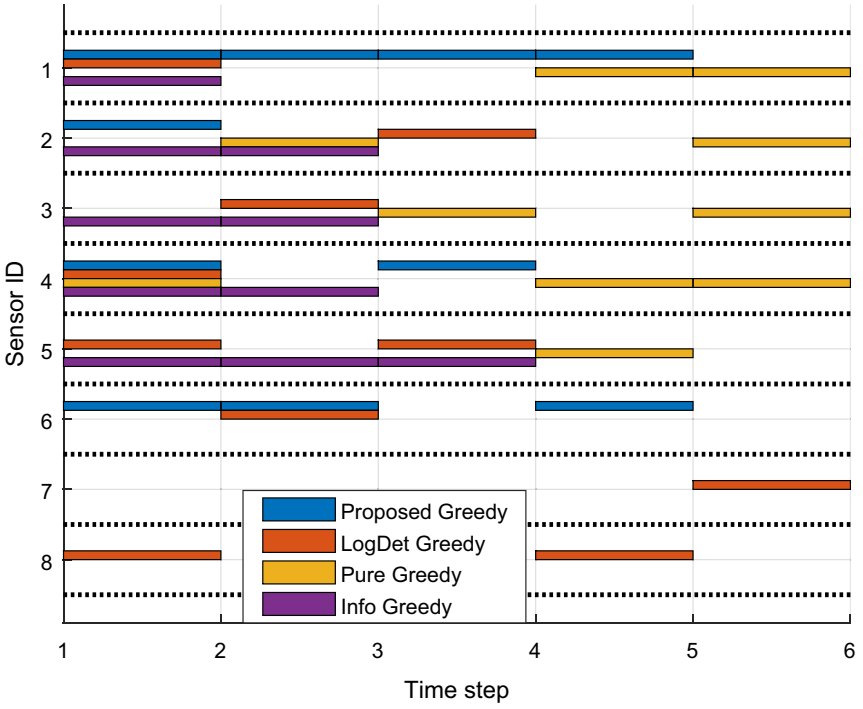

Fig. 10 Scheduling chart with eight available sensors and $\Psi_t = 0.5$.

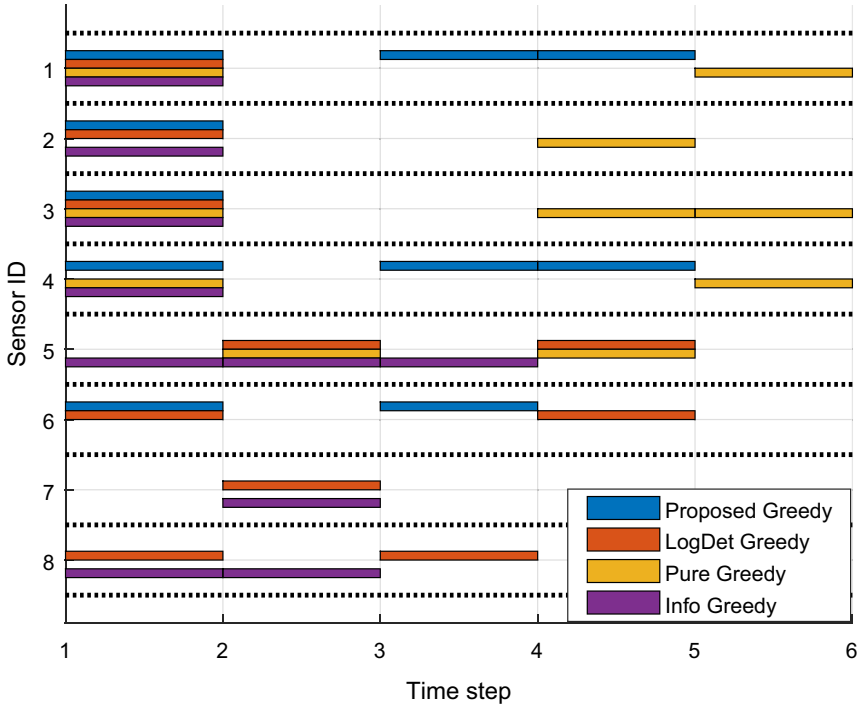

Fig. 11 Scheduling chart with eight available sensors and $\Psi_t = 0.99$.

increases as time step t approaches the final time t_f, resulting in overall performance degradation. In all five test cases, the proposed algorithm consistently outperforms the *logdet* solutions. Figures 9–11 present the scheduling results for the selective cases of $\Psi_t = 0.01$, 0.5, and 0.99, respectively. Unlike the previous example, the *logdet* solutions yield different sensor scheduling across all test cases. This discrepancy arises from the fact that the measurement covariance V_t varies along with Ψ_t though the weight matrices Θ_t remain the same for all cases.

As the temporal correlation Ψ_t increases, both the proposed algorithm and the *logdet* greedy solution exhibit a preference for selecting sensors near the initial time. The *logdet* solution tends to intensively select radial sensors near the initial time while scattering the remaining scheduling among angle radars. This behavior can be attributed to the default measurement variance of radius R_r being larger than that of angle R_a as shown in Eqs. (60) and (61). This difference grows as time elapses due to the colored noise model described in Eq. (4). Consequently, the measurement variance of range radars becomes too large to effectively reduce MSE, and the *logdet* greedy algorithm allocates more scheduling to angle radars, which can more efficiently reduce the overall cost. Similarly, the proposed algorithm chooses range radars near the initial time. However, it further allocates less scheduling to angle radars and favors the selection of additional radial sensors. This behavior arises from the capability of the proposed algorithm to adapt to time-varying weight matrices, unlike the *logdet* solution, which lacks this ability.

VIII. Conclusions

The problem of nonmyopic sensor selection in Kalman filtering applications subject to both spatially and temporally correlated measurement or process noises and problem-specific linear weightings of covariance matrices was considered. Linear estimation insights were employed to provide the first explicit, closed-form solution for the covariance matrices as functions of all schedules. With the benefit of this result, a surrogate objective was constructed by directly examining the role of the scheduling variables within the cost—a unique feature. Performance guarantees were proven for the application of the greedy algorithm to the surrogate objective. The effectiveness of the algorithm was demonstrated for a number of simulated examples. In future work, it is desired to use this explicit expression to provide guarantees for the greedy solution on the original, nonmyopic MSE objective. Furthermore, proofs for the single-step (myopic) formulation would be of interest for applications that are heavily computationally constrained.

Acknowledgments

This research was sponsored by the Army Research Office and was accomplished under Cooperative Agreement Number W911NF-19-2-0333 and the National Science Foundation (Award No. 1944318). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

Appendix A: Proof of Submodularity for the Surrogate Objective

Submodularity is established by verifying that Eq. (42) satisfies the conditions in Property 1. It is well known that the set domain in this sensor scheduling problem corresponds to a uniform matroid with respect to cardinality n_s. Also, $\tilde{F}$ is normalized by design in Eq. (40). It only remains to be shown that $\tilde{\mathcal{F}}$ is non-negative, monotone nondecreasing, and submodular.

1. Non-Negativity

Consider a single term in the double summation of $\tilde{\mathcal{F}}$, i.e., a single term at a single time step. This will be of the form

$$-\left[\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\boldsymbol{u}_{t,j}^T G_t^{-1} D_t(\mathcal{S}) G_t^{-1}\boldsymbol{u}_{t,j}} - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j}}\right]$$

$$= -\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j}}\left[\frac{1}{1 + \frac{\frac{1}{\alpha_t}\boldsymbol{u}_{t,j}^T G_t^{-1} D_t(\mathcal{S}) G_t^{-1}\boldsymbol{u}_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j}}} - 1\right] \geq 0 \quad \text{(A1)}$$

This follows from $\lambda_{t,j}$, $\alpha_t > 0$, positive definiteness of G_t, and positive semidefiniteness of $D_t(\mathcal{S})$.

2. Monotonicity

Let $\mathcal{A} = \{s_1, \ldots, s_l\}$ and $\mathcal{A} \subseteq \mathcal{B} = \{s_1, \ldots, s_l, s_{l+1}, \ldots, s_m\}$. Before, indices for different sensors had been defined by ordered pairs corresponding to time and element within the sensor collection at that time, e.g., (t, i). Each element in sets A and B implicitly defines selections of these ordered pairs. We set out to show that $\tilde{\mathcal{F}}(\mathcal{B}) - \tilde{\mathcal{F}}(\mathcal{A}) \geq 0$.

$$\begin{aligned}
&\tilde{\mathcal{F}}(\mathcal{B}) - \tilde{\mathcal{F}}(\mathcal{A}) \\
&= -\sum_{t=1}^{t_f}\sum_{j=1}^{r_t}\left[\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\boldsymbol{u}_{t,j}^T G_t^{-1} D_t(\mathcal{B}) G_t^{-1}\boldsymbol{u}_{t,j}} - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j}}\right] \\
&\quad + \sum_{t=1}^{t_f}\sum_{j=1}^{r_t}\left[\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\boldsymbol{u}_{t,j}^T G_t^{-1} D_t(\mathcal{A}) G_t^{-1}\boldsymbol{u}_{t,j}} - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j}}\right]
\end{aligned} \tag{A2}$$

$$\begin{aligned}
&= \sum_{t=1}^{t_f}\sum_{j=1}^{r_t}\left[\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\boldsymbol{u}_{t,j}^T G_t^{-1} D_t(\mathcal{A}) G_t^{-1}\boldsymbol{u}_{t,j}}\right. \\
&\quad \left. - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\boldsymbol{u}_{t,j}^T G_t^{-1} D_t(\mathcal{B}) G_t^{-1}\boldsymbol{u}_{t,j}}\right]
\end{aligned} \tag{A3}$$

Consider a single term at time step t. Since $\mathcal{A}$ and $\mathcal{B}$ can include sensors for times greater than t and our problem is causal, define sets $\tilde{\mathcal{A}} = \{\tilde{s}_1, \ldots, \tilde{s}_{\tilde{l}}\}$ and $\tilde{\mathcal{B}} = \{\tilde{s}_1, \ldots, \tilde{s}_{\tilde{l}}, \ldots, \tilde{s}_{\tilde{m}}\}$ to be the subsets of $\mathcal{A}$ and $\mathcal{B}$, respectively, containing all possible sensors relevant to time step t. Next, examine terms of the form $\boldsymbol{u}_{t,j}^T G_t^{-1} D_t(\cdot) G_t^{-1}\boldsymbol{u}_{t,j}$. Define $\tilde{\boldsymbol{u}}_{t,j} = G_t^{-1}\boldsymbol{u}_{t,j}$ so that the form is simply $\tilde{\boldsymbol{u}}_{t,j}^T D_t(\cdot)\tilde{\boldsymbol{u}}_{t,j}$. Since $D_t(\cdot)$ is diagonal, this value will just be a sum of squared elements of vector $\tilde{\boldsymbol{u}}_{t,j}$ for which the sensor schedules are nonzero. To define the notation that associates the elements of vector $\tilde{\boldsymbol{u}}_{t,j}$ to the sensors acting on it, let function $g(i)$ indicate the sensor schedule acting on the ith element, i.e., $[\tilde{\boldsymbol{u}}_{t,j}]_i$.

A single term for time step t is

$$\begin{aligned}
&\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\tilde{\boldsymbol{u}}_{t,j}^T D_t(\tilde{\mathcal{A}})\tilde{\boldsymbol{u}}_{t,j}} - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\tilde{\boldsymbol{u}}_{t,j}^T D_t(\tilde{\mathcal{B}})\tilde{\boldsymbol{u}}_{t,j}} \\
&= \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\sum_{g(i)\in\tilde{\mathcal{A}}}[\tilde{\boldsymbol{u}}_{t,j}]_i^2} - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\sum_{g(i)\in\tilde{\mathcal{B}}}[\tilde{\boldsymbol{u}}_{t,j}]_i^2}
\end{aligned} \tag{A4}$$

$$\begin{aligned}
&= \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\sum_{g(i)\in\tilde{\mathcal{A}}}[\tilde{\boldsymbol{u}}_{t,j}]_i^2} \\
&\quad - \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\left(\sum_{g(i)\in\tilde{\mathcal{A}}}[\tilde{\boldsymbol{u}}_{t,j}]_i^2 + \sum_{g(i)\in\{\tilde{\mathcal{B}}\setminus\tilde{\mathcal{A}}\}}[\tilde{\boldsymbol{u}}_{t,j}]_i^2\right)} \geq 0
\end{aligned} \tag{A5}$$

Recall that $\lambda_{t,j}, \alpha_t > 0$ and G_t is positive definite. Since all terms at all time steps are non-negative, we have that $\tilde{\mathcal{F}}(\mathcal{B}) - \tilde{\mathcal{F}}(\mathcal{A}) \geq 0$.

3. Submodularity

Use the above definitions for sets $\mathcal{A}, \mathcal{B}$, and let $s \in \mathcal{V} \setminus \mathcal{B}$. We set out to show that

$$\tilde{\mathcal{F}}(\mathcal{A} \cup s) - \tilde{\mathcal{F}}(\mathcal{A}) - \left(\tilde{\mathcal{F}}(\mathcal{B} \cup s) - \tilde{\mathcal{F}}(\mathcal{B})\right) \geq 0 \tag{A6}$$

For a single term in the double summation of $\tilde{\mathcal{F}}$, the left-hand side of the above expression is

$$\begin{aligned}
&-\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\tilde{\boldsymbol{u}}_{t,j}^T D_t(\mathcal{A}\cup s)\tilde{\boldsymbol{u}}_{t,j}} + \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\tilde{\boldsymbol{u}}_{t,j}^T D_t(\mathcal{A})\tilde{\boldsymbol{u}}_{t,j}} \\
&-\left(-\frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\tilde{\boldsymbol{u}}_{t,j}^T D_t(\mathcal{B}\cup s)\tilde{\boldsymbol{u}}_{t,j}} + \frac{\lambda_{t,j}}{\boldsymbol{u}_{t,j}^T G_t^{-1}\boldsymbol{u}_{t,j} + \frac{1}{\alpha_t}\tilde{\boldsymbol{u}}_{t,j}^T D_t(\mathcal{B})\tilde{\boldsymbol{u}}_{t,j}}\right)
\end{aligned} \tag{A7}$$

This is of the general form:

$$-\frac{1}{a+b} + \frac{1}{a} + \frac{1}{a+b+c} - \frac{1}{a+c} \tag{A8}$$

where $a, b, c \geq 0$. This is equivalent to

$$\begin{aligned}
&\frac{b}{a(a+b)} - \frac{b}{(a+c)(a+b+c)} \\
&\quad = b\left\{\frac{(a+c)(a+b+c) - a(a+b)}{a(a+b)(a+c)(a+b+c)}\right\} \geq 0
\end{aligned} \tag{A9}$$

since $[(a+c)(a+b+c)/a(a+b)] \geq [a+b+c/a+b] \geq 1$. Therefore, every term in the left-hand side of Eq. (A6) is non-negative. □

Appendix B: Computing Batch Terms

Let us define several dimension-relevant numbers for the following size of matrices; $\eta_{t,m}$ represents the dimension of the measurement from mth sensor at time step t. The total dimension of all measurements at time step t is denoted by $\eta_{\forall,t} = \sum_{k=1}^{k=m_t}\eta_{t,k}$. Then, the total dimension of all measurements from time step 1 to time step t becomes $N_{\forall,t} = \sum_{i=1}^{i=t}\eta_{\forall,i}$. Each state, $\boldsymbol{x}_t$, is of dimension n. The precomputed terms in Algorithm 1 are given by the following equations:

$$\alpha_t = \gamma \mathrm{eig}_{\min}\left(\Sigma_{\tilde{\nu}_t\tilde{\nu}_t^T}\right) > 0 \tag{B1}$$

Value $\gamma = 0.2$ was observed to work well in the presented examples.

$$G_t^{-1} = H_t P_{t|0} H_t^T + H_t F_t + F_t^T H_t^T + S_t \in \mathbb{R}^{N_{\forall,t}\times N_{\forall,t}} \tag{B2}$$

$$\text{where} \quad S_t = \Sigma_{\tilde{\nu}_t\tilde{\nu}_t^T} - \alpha_t I \in \mathbb{R}^{N_{\forall,t}\times N_{\forall,t}} \tag{B3}$$

Let $\lambda_{t,j}$ and $\boldsymbol{u}_{t,j}$ represent the singular values and right singular vectors of the following matrix M_t, respectively:

$$M_t = \Theta_t\left(P_{t|0}H_t^T + F_t\right) \in \mathbb{R}^{n\times N_{\forall,t}} \tag{B4}$$

It can be seen that each equation is a function of the batch terms $P_{t|0}$, H_t, F_t, and $\Sigma_{\tilde{\nu}_t\tilde{\nu}_t^T}$. Though these terms are ingredients of the batch processor, computing them for each time step does not amount to a full batch update at each time. Unlike the full update (when there exists colored noise), the individual terms can be computed via recursions over time. These recursions arise from rewriting the concatenated measurement vector in Eq. (9) as

$$\tilde{\boldsymbol{Y}}_t = C_t^{\text{big}} \begin{bmatrix} \Phi(1,t) \\ \vdots \\ \Phi(1,t-1) \\ I_n \end{bmatrix} \boldsymbol{x}_t + \begin{bmatrix} C_{t-1}^{\text{big}} \Pi_t \tilde{\boldsymbol{w}}_{\boldsymbol{t-1}} \\ \boldsymbol{0} \end{bmatrix} + \tilde{\boldsymbol{\nu}}_t \tag{B5}$$

where

$$C_t^{\text{big}} = \text{blkdiag}(C_{1,1}, \ldots, C_{1,m_1}, \ldots, C_{t,1}, \ldots, C_{1,m_t}) \in \mathbb{R}^{N_{\forall,t} \times nt} \tag{B6}$$

$$\tilde{\boldsymbol{w}}_{\boldsymbol{t-1}} = [\boldsymbol{w}_{\boldsymbol{1}}^T \ \ldots \ \boldsymbol{w}_{\boldsymbol{t-1}}^T]^T \in \mathbb{R}^{n(t-1)\times 1} \tag{B7}$$

$$\Phi(t,\tau) = A^{t-\tau} \in \mathbb{R}^{n \times n} \tag{B8}$$

$$\Pi_t = -\begin{bmatrix} \Phi(1,2) & \Phi(1,3) & \ldots & \Phi(1,t) \\ 0 & \Phi(2,3) & \ldots & \Phi(2,t) \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \Phi(t-1,t) \end{bmatrix} \in \mathbb{R}^{n(t-1)\times n(t-1)}$$

is an upper triangular matrix

(B9)

Then the necessary batch terms are then

$$H_1 = C_1 \in \mathbb{R}^{\eta_{\forall,1} \times n}, \qquad H_{t+1} = \begin{bmatrix} H_t \Phi(t,t+1) \\ C_{t+1} \end{bmatrix} \in \mathbb{R}^{N_{\forall,t+1} \times n} \tag{B10}$$

$$F_1 = \boldsymbol{0}_{n \times \eta_{\forall,1}},$$
$$F_{t+1} = \begin{bmatrix} \Phi(t+1,t)\Phi_{0,t}\Pi_t W_{t-1}^{\text{aug}} \Pi_t^t (C_{t-1}^{\text{big}})^T & \boldsymbol{0}_{n,\eta_{\forall,t}} \end{bmatrix} \in \mathbb{R}^{n \times N_{\forall,t+1}} \tag{B11}$$

$$\Sigma_{\tilde{\boldsymbol{\nu}},\tilde{\boldsymbol{\nu}}_t^T} = \begin{bmatrix} V_{t-1}^{\text{aug}} + C_{t-1}^{\text{big}} \Pi_t W_{t-1}^{\text{aug}} \Pi_t^t (C_t^{\text{big}})^T & \boldsymbol{0} \\ \boldsymbol{0} & V_t \end{bmatrix} \in \mathbb{R}^{N_{\forall,t} \times N_{\forall,t}} \tag{B12}$$

$$P_{t|0} = \Phi(t,1) P_{1|0} \Phi^T(t,1) + \Phi_{0,t} \Pi_t W_t^{\text{aug}} \Pi_t^T \Phi_{0,t}^T \in \mathbb{R}^{n \times n} \tag{B13}$$

$$\text{where it is defined} \quad \Phi_{0,t} = \begin{bmatrix} \Phi(t,1) & \boldsymbol{0}_{n \times n(t-2)} \end{bmatrix} \in \mathbb{R}^{n \times n(t-1)} \tag{B14}$$

Finally, matrices Π_t, W_t^{aug}, and V_t^{aug} are given by the following recursions. We notate concatenated $\Phi_t^{\text{big}} = [\Phi^T(1,t+1) \ \ldots \ \Phi^T(t-1,t+1)]$.

$$\Pi_2 = \Phi(1,2) \in \mathbb{R}^{n \times n}, \qquad \Pi_{t+1} = \begin{bmatrix} \Pi_t & (\Phi_t^{\text{big}})^T \\ \boldsymbol{0} & \Phi(t,t+1) \end{bmatrix} \in \mathbb{R}^{nt \times nt} \tag{B15}$$

The colored noise model in Eq. (4) is used to accommodate colored process noise. Similarly, notate concatenated

$$W_t^{\text{big}} = \begin{bmatrix} W_1 (\Psi_1^w)^T \ldots (\Psi_t^w)^T \\ W_2 (\Psi_2^w)^T \ldots (\Psi_t^w)^T \\ \vdots \\ W_t (\Psi_t^w)^T \end{bmatrix} \in \mathbb{R}^{nt \times n} \tag{B16}$$

$$W_1^{\text{aug}} = W_1 \in \mathbb{R}^{n \times n}, \quad W_{t+1}^{\text{aug}} = \begin{bmatrix} W_t^{\text{aug}} & W_t^{\text{big}} \\ (W_t^{\text{big}})^T & W_t \end{bmatrix} \in \mathbb{R}^{n(t+1) \times n(t+1)} \tag{B17}$$

Note that if the process noise is white, then all $\Psi_i^w = \boldsymbol{0}$, and all W_t^{aug} are block diagonal. Then $P_{t|0}$ can be expressed as

$$P_{t|0} = A_{t-1} \ldots A_1 P_{1|0} A_1 \ldots A_{t-1} + A_{t-1} \ldots A_2 W_1 A_2 \ldots A_{t-1} + A_{t-1} \ldots A_3 W_2 A_3 \ldots A_{t-1} + \ \ldots \ + W_{t-1} \tag{B18}$$

In this white process noise case, $P_{t|0}$ could also be computed recursively via the usual Kalman filter propagation equation $P_{t|0} = A_{t-1} P_{t-1|0} A_{t-1}^T + W_{t-1}$.

Colored measurement noise is also modeled via Eq. (4) but now with matrix $\Psi_i^v = \text{diag}([\Psi_{t,1} \ \ldots \ \Psi_{t,r_t}])$ containing colored noise transition values for each sensor at each time step. If sensor i has white noise, then $\Psi_{t,i} = \boldsymbol{0}$ for $\forall t \in \{1, \ldots, t_f\}$. If all sensors have white noise, then V_t^{aug} are block diagonal. Again we have

$$V_t^{\text{big}} = \begin{bmatrix} V_1 (\Psi_1^v)^T \ldots (\Psi_t^v)^T \\ V_2 (\Psi_2^v)^T \ldots (\Psi_t^v)^T \\ \vdots \\ V_t (\Psi_t^v)^T \end{bmatrix} \in \mathbb{R}^{N_{\forall,t} \times \eta_{\forall,t+1}} \tag{B19}$$

$$V_1^{\text{aug}} = V_1 \in \mathbb{R}^{\eta_{\forall,1} \times \eta_{\forall,1}}, \quad V_{t+1}^{\text{aug}} = \begin{bmatrix} V_t^{\text{aug}} & V_t^{\text{big}} \\ (V_t^{\text{big}})^T & V_{t+1} \end{bmatrix} \in \mathbb{R}^{N_{\forall,t+1} \times N_{\forall,t+1}} \tag{B20}$$

The procedure is then advanced through each time step, computing Eqs. (B10–B20) and corresponding Eqs. (B1–B4). This completes the precomputed values for Algorithm 1.

References

[1] Gualdoni, M. J., and DeMars, K. J., "Impartial Sensor Tasking via Forecasted Information Content Quantification," *Journal of Guidance, Control, and Dynamics*, Vol. 43, No. 11, 2020, pp. 2031–2045. https://doi.org/10.2514/1.G004757

[2] Mo, Y., Ambrosino, R., and Sinopoli, B., "Sensor Selection Strategies for State Estimation in Energy Constrained Wireless Sensor Networks," *Automatica*, Vol. 47, No. 7, 2011, pp. 1330–1338. https://doi.org/10.1016/j.automatica.2011.02.001

[3] Liu, S., Chepuri, S. P., Fardad, M., Maşazade, E., Leus, G., and Varshney, P. K., "Sensor Selection for Estimation with Correlated Measurement Noise," *IEEE Transactions on Signal Processing*, Vol. 64, No. 13, 2016, pp. 3509–3522. https://doi.org/10.1109/TSP.2016.2550005

[4] Yamada, K., Saito, Y., Nankai, K., Nonomura, T., Asai, K., and Tsubakino, D., "Fast Greedy Optimization of Sensor Selection in Measurement with Correlated Noise," *Mechanical Systems and Signal Processing*, Vol. 158, 2021, Paper 107619. https://doi.org/10.1016/j.ymssp.2021.107619

[5] Nagata, T., Yamada, K., Nonomura, T., Nankai, K., Saito, Y., and Ono, S., "Data-Driven Sensor Selection Method Based on Proximal Optimization for High-Dimensional Data with Correlated Measurement Noise," *IEEE Transactions on Signal Processing*, Vol. 70, 2022, pp. 5251–5264. https://doi.org/10.1109/TSP.2022.3212150

[6] Wang, J., Dong, P., Jing, Z., and Cheng, J., "Consensus-Based Filter for Distributed Sensor Networks with Colored Measurement Noise," *Sensors*, Vol. 18, No. 11, 2018, p. 3678. https://doi.org/10.3390/s18113678

[7] Zhou, J., Xu, Y., Li, J.-Y., and Li, Y., "Robust State Estimation for Discrete Time Systems with Colored Noises and Communication Constraints," *Journal of the Franklin Institute*, Vol. 355, No. 13, 2018, pp. 5790–5810. https://doi.org/10.1016/j.jfranklin.2018.05.043

[8] Bryson, A. E., and Johansen, D., "Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise," *IEEE Transactions on Automatic Control*, Vol. 10, No. 1, 1965, pp. 4–10. https://doi.org/10.1109/TAC.1965.1098063

[9] Bryson, A. E., and Henrikson, L., "Estimation Using Sampled Data Containing Sequentially Correlated Noise," *Journal of Spacecraft and Rockets*, Vol. 5, No. 6, 1968, pp. 662–665. https://doi.org/10.2514/3.29327

[10] Bryson, A. E., and Ho, Y.-C., *Applied Optimal Control: Optimization, Estimation and Control*, CRC Press, Boca Raton, FL, 1975. https://doi.org/10.1201/9781315137667

[11] Joshi, S., and Boyd, S., "Sensor Selection via Convex Optimization," *IEEE Transactions on Signal Processing*, Vol. 57, No. 2, 2008, pp. 451–462. https://doi.org/10.1109/TSP.2008.2007095

[12] Krause, A., and Golovin, D., *Submodular Function Maximization*, Cambridge Univ. Press, Cambridge, England, U.K., 2014.

[13] Jawaid, S. T., and Smith, S. L., "Submodularity and Greedy Algorithms in Sensor Scheduling for Linear Dynamical Systems," *Automatica*, Vol. 61, 2015, pp. 282–288. https://doi.org/10.1016/j.automatica.2015.08.022

[14] Shamaiah, M., Banerjee, S., and Vikalo, H., "Greedy Sensor Selection: Leveraging Submodularity," *49th IEEE Conference on Decision and Control (CDC)*, Inst. of Electrical and Electronics Engineers, New York, 2010, pp. 2572–2577. https://doi.org/10.1109/CDC.2010.5717225

[15] Chamon, L. F. O., Pappas, G. J., and Ribeiro, A., "Approximate Supermodularity of Kalman Filter Sensor Selection," *IEEE Transactions on Automatic Control*, Vol. 66, No. 1, 2021, pp. 49–63. https://doi.org/10.1109/TAC.2020.2973774

[16] Tzoumas, V., Jadbabaie, A., and Pappas, G. J., "Near-Optimal Sensor Scheduling for Batch State Estimation: Complexity, Algorithms, and Limits," *2016 IEEE 55th Conference on Decision and Control (CDC)*, Inst. of Electrical and Electronics Engineers, New York, 2016, pp. 2695–2702. https://doi.org/10.1109/CDC.2016.7798669

[17] Roy, V., Simonetto, A., and Leus, G., "Spatio-Temporal Sensor Management for Environmental Field Estimation," *Signal Processing*, Vol. 128, 2016, pp. 369–381. https://doi.org/10.1016/j.sigpro.2016.05.011

[18] Williamson, D. P., and Shmoys, D. B., *The Design of Approximation Algorithms*, Cambridge Univ. Press, Cambridge, England, U.K., 2011. https://doi.org/10.1017/CBO9780511921735

[19] Jünger, M., Liebling, T. M., Naddef, D., Nemhauser, G. L., Pulleyblank, W. R., Reinelt, G., Rinaldi, G., and Wolsey, L. A., *50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art*, Springer Science & Business Media, Berlin, 2009.

[20] Hashemi, A., Ghasemi, M., Vikalo, H., and Topcu, U., "Randomized Greedy Sensor Selection: Leveraging Weak Submodularity," *IEEE Transactions on Automatic Control*, Vol. 66, No. 1, 2021, pp. 199–212. https://doi.org/10.1109/TAC.2020.2980924

[21] Singh, P., Chen, M., Carlone, L., Karaman, S., Frazzoli, E., and Hsu, D., "Supermodular Mean Squared Error Minimization for Sensor Scheduling in Optimal Kalman Filtering," *2017 American Control Conference (ACC)*, Inst. of Electrical and Electronics Engineers, New York, 2017, pp. 5787–5794. https://doi.org/10.23919/ACC.2017.7963857

[22] Blum, A., Hopcroft, J., and Kannan, R., *Foundations of Data Science*, Cambridge Univ. Press, Cambridge, England, U.K., 2020. https://doi.org/10.1017/9781108755528

[23] Adurthi, N., Singla, P., and Majji, M., "Mutual Information Based Sensor Tasking with Applications to Space Situational Awareness," *Journal of Guidance, Control, and Dynamics*, Vol. 43, No. 4, 2020, pp. 767–789. https://doi.org/10.2514/1.G004399

[24] Hibbard, M., Hashemi, A., Tanaka, T., and Topcu, U., "Randomized Greedy Algorithms for Sensor Selection in Large-Scale Satellite Constellations," *2023 American Control Conference (ACC)*, Inst. of Electrical and Electronics Engineers, New York, Vol. 10, 2023, pp. 4276–4283. https://doi.org/10.23919/ACC55779.2023.10156009

[25] Crassidis, J. L., and Junkins, J. L., *Optimal Estimation of Dynamics Systems*, 2nd ed., Chapman and Hall/CRC, Boca Raton, FL, 2004. https://doi.org/10.1201/9780203509128

[26] Bar-Shalom, Y., Mallick, M., Chen, H., and Washburn, R., "One-Step Solution for the General Out-of-Sequence-Measurement Problem in Tracking," *Proceedings of the IEEE Aerospace Conference*, Inst. of Electrical and Electronics Engineers, New York, Vol. 4, 2002, pp. 1551–1559. https://doi.org/10.1109/AERO.2002.1036872

[27] Genovese, A. F., "The Interacting Multiple Model Algorithm for Accurate State Estimation of Maneuvering Targets," *Johns Hopkins APL Technical Digest*, Vol. 22, No. 4, 2001, pp. 614–623.

[28] Tanaka, T., and Sandberg, H., "SDP-Based Joint Sensor and Controller Design for Information-Regularized Optimal LQG Control," *2015 IEEE 54th Annual Conference on Decision and Control (CDC)*, Inst. of Electrical and Electronics Engineers, New York, 2015, pp. 4486–4491. https://doi.org/10.1109/CDC.2015.7402920

[29] Hajek, B., *Random Processes for Engineers*, Cambridge Univ. Press, Cambridge, England, U.K., 2015.

[30] Motwani, R., and Raghavan, P., *Randomized Algorithms*, Cambridge Univ. Press, Cambridge, England, U.K., 1995. https://doi.org/10.1017/CBO9780511814075

[31] Luo, Z.-Q., Ma, W.-K., So, A. M.-C., Ye, Y., and Zhang, S., "Semidefinite Relaxation of Quadratic Optimization Problems," *IEEE Signal Processing Magazine*, Vol. 27, No. 3, 2010, pp. 20–34. https://doi.org/10.1109/MSP.2010.936019

[32] Candes, E. J., Wakin, M. B., and Boyd, S. P., "Enhancing Sparsity by Reweighted l-1 Minimization," *Journal of Fourier Analysis and Applications*, Vol. 14, Nos. 5–6, 2008, pp. 877–905. https://doi.org/10.1007/s00041-008-9045-x

[33] Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L., "An Analysis of Approximations for Maximizing Submodular Set Functions – I," *Mathematical Programming*, Vol. 14, No. 1, 1978, pp. 265–294. https://doi.org/10.1007/BF01588971

[34] Tzoumas, V., Jadbabaie, A., and Pappas, G. J., "Sensor Placement for Optimal Kalman Filtering: Fundamental Limits, Submodularity, and Algorithms," *2016 American Control Conference (ACC)*, Inst. of Electrical and Electronics Engineers, New York, 2016, pp. 191–196. https://doi.org/10.1109/ACC.2016.7524914

[35] Tzoumas, V., Atanasov, N. A., Jadbabaie, A., and Pappas, G. J., "Scheduling Nonlinear Sensors for Stochastic Process Estimation," *2017 American Control Conference (ACC)*, Inst. of Electrical and Electronics Engineers, New York, 2017, pp. 580–585. https://doi.org/10.23919/ACC.2017.7963015

[36] Uciński, D., "D-Optimal Sensor Selection in the Presence of Correlated Measurement Noise," *Measurement*, Vol. 164, 2020, Paper 107873. https://doi.org/10.1016/j.measurement.2020.107873

[37] Wold, S., Esbensen, K., and Geladi, P., "Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, Vol. 2, Nos. 1–3, 1987, pp. 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

[38] Tuggle, K., and Zanetti, R., "Automated Splitting Gaussian Mixture Nonlinear Measurement Update," *Journal of Guidance, Control, and Dynamics*, Vol. 41, No. 3, 2017, pp. 725–734. https://doi.org/10.2514/1.G003109

Y. Cheng
Associate Editor