2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

The Mean Square Error in Kalman Filtering Sensor Selection is Approximately Supermodular

Luiz F. O. Chamon, George J. Pappas, and Alejandro Ribeiro

***Abstract*— This work considers the problem of selecting sensors in large scale system to minimize the state estimation mean-square error (MSE). More specifically, it leverages the concept of approximate supermodularity to derive near-optimality certificates for greedy solutions of this problem in the context of Kalman filtering. It also shows that in typical application scenarios, these certificates approach the typical $1/e$ guarantee. These performance bounds are important because sensor selection problems are in general NP-hard. Hence, their solution can only be approximated in practice even for moderately large problems. A common way of deriving these approximations is by means of convex relaxations. These, however, come with no performance guarantee. Another approach uses greedy search, although also in this case typical guarantees do not hold since the MSE is neither submodular nor supermodular. This issue is commonly addressed by using a surrogate supermodular figure of merit, such as the $\log\det$. Unfortunately, this is not equivalent to minimizing the MSE. This work demonstrates that no change to the original problem is needed to obtain performance guarantees.**

I. INTRODUCTION

We consider the problem of observing large scale systems in which sensing more than a reduced fraction of their output variables is impractical. Specifically, we are interested in finding a subset of the system outputs that minimizes the state estimation mean-square error (MSE) subject to some sensing budget. This problem is particularly critical in distributed system where power and communication constraints additionally limit the number of sensors available [1]–[3]. It can be found in applications such as target tracking, field monitoring, power allocation, and biological systems analysis [4]–[9].

In general, sensor selection is an NP-hard problem [10]–[16]. In fact, even finding the smallest subset of outputs that make the system observable is NP-hard [17]. Minimizing the state estimation MSE under a sensor budget constraint is therefore intractable even for moderately large systems. Thus, we must turn to approximate solutions.

A common approach is to cast sensor selection as an integer program and approximate its solution using a convex relaxation of its constraints [16], [18]–[23]. It typically includes a sparsity promoting regularization and obtains a sensing set by either rounding or some randomized procedure. Although these methods are practical and make it straightforward to include additional constraints [19], they do not have approximation guarantees.

Another avenue relies on approximation algorithms from discrete optimization [8], [24]. Here, greedy minimization remains ubiquitous due to its low complexity and near-optimal guarantees when the cost function is supermodular. This, however, is not the case for the MSE. Counter-examples in the context of control theory can be found in [14]–[16]. This issue is typically addressed by using a surrogate supermodular estimation error measure, such as the $\log\det$, or some information-theoretic measure relating the selected output set and its complement [10], [13], [20], [25]–[27]. Though effective, there is no direct relation between the estimation MSE and these cost functions, so it makes sense to take a page out of Tukey's book and try to solve the "right problem" [28].

This work therefore sets out to provide theoretical performance guarantees on greedily selecting sensors to minimize the state estimation MSE. To do so, we leverage the concept of α-supermodularity introduced in [29] and (i) derive bounds on the α-supermodularity of the state estimation MSE in the context of smoothing and Kalman filtering; (ii) use these bounds to provide a near-optimal certificate for greedy sensor selection in these applications; (iii) show that the state estimation MSE is almost supermodular in many cases of interest; and (iv) illustrate these results in numerical examples.

Notation: Lowercase boldface letters represent vectors ($\boldsymbol{x}$), uppercase boldface letters are matrices ($\boldsymbol{X}$), and calligraphic letters denote sets ($\mathcal{A}$). We write $|\mathcal{A}|$ for the cardinality of $\mathcal{A}$ and let $[p] = \{1, \dots, p\}$. Set subscripts are used to subset vectors, so that $\boldsymbol{x}_{\mathcal{A}}$ refers to the vector obtained by keeping only the elements with indices in $\mathcal{A}$. To say $\boldsymbol{X}$ is a positive semi-definite (PSD) matrix we write $\boldsymbol{X} \succeq 0$, so that for $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n \times n}$, $\boldsymbol{X} \preceq \boldsymbol{Y} \Leftrightarrow \boldsymbol{b}^T \boldsymbol{X} \boldsymbol{b} \leq \boldsymbol{b}^T \boldsymbol{Y} \boldsymbol{b}$, for all $\boldsymbol{b} \in \mathbb{R}^n$. Similarly, we write $\boldsymbol{X} \succ 0$ when $\boldsymbol{X}$ is positive definite. We use $\mathbb{R}_+$ to denote the set of non-negative real numbers and $\mathbb{S}_+^n$ for the set of $n \times n$ PSD matrices.

II. PROBLEM FORMULATION

Consider a dynamical system with states $\boldsymbol{x}_k \in \mathbb{R}^n$ and observations $\boldsymbol{y}_k \in \mathbb{R}^p$ indexed by a discrete time index k. The state evolution and observations follow the linear dynamics

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{F}\boldsymbol{x}_k + \boldsymbol{w}_k \\ \boldsymbol{y}_k &= \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{v}_k \end{aligned} \tag{1}$$

where $\boldsymbol{F} \in \mathbb{R}^{n \times n}$ is the state transition matrix and $\boldsymbol{H} \in \mathbb{R}^{p \times n}$ is the output matrix. The process noise $\boldsymbol{w}_k$ and measurement noise $\boldsymbol{v}_k$ are zero-mean vector-valued Gaussian random variables with covariances $\boldsymbol{R}_w = \mathbb{E}\, \boldsymbol{w}_k \boldsymbol{w}_k^T = \operatorname{diag}\left(\sigma_{w,i}^2\right)$ and $\boldsymbol{R}_v = \mathbb{E}\, \boldsymbol{v}_k \boldsymbol{v}_k^T = \operatorname{diag}\left(\sigma_{v,i}^2\right)$ for all k. We assume $\{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{w}_i, \boldsymbol{w}_j\}$ are independent for all $i \neq j$.

Electrical and Systems Engineering, University of Pennsylvania. {luizf, pappasg, aribeiro}@seas.upenn.edu

The initial state $\boldsymbol{x}_0 \sim \mathcal{N}(\bar{\boldsymbol{x}}_0, \boldsymbol{\Pi}_0)$ is assumed be Gaussian distributed with mean $\bar{\boldsymbol{x}}_0$ and covariance $\boldsymbol{\Pi}_0 \succ 0$.

The goal of sensor selection problems is to choose a subset of the outputs to be used for estimating the system states. We would naturally like for this subset to be small and lead to good estimation performance. Explicitly, we seek a sensing set $\mathcal{S}$ with at most s outputs that minimizes the MSE incurred from estimating the states $\boldsymbol{x}_k$ from the subset of outputs $(\boldsymbol{y}_k)_{\mathcal{S}}$. In particular, we are interested in studying two estimation scenarios:

(PI) **smoothing**, where we minimize the MSE incurred from estimating all states up to time ℓ based on past observations. Formally, we seek the state estimates $\tilde{\boldsymbol{x}}_k = \mathbb{E}\left[\boldsymbol{x}_k \mid \{(\boldsymbol{y}_j)_{\mathcal{S}}\}_{j\leq\ell}\right]$ for $k \leq \ell$ by solving

$$\begin{aligned} \underset{\mathcal{S}\subseteq[p]}{\text{minimize}} \quad & J_{\text{PI}}(\mathcal{S}) \\ \text{subject to} \quad & |\mathcal{S}| \leq s \end{aligned} \tag{PI}$$

where

$$J_{\text{PI}}(\mathcal{S}) = \min_{\{\tilde{\boldsymbol{x}}_k\}_{k\leq\ell}} \sum_{k=0}^{\ell} \mathbb{E}\left[\|\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_k\|_2^2 \mid \{(\boldsymbol{y}_j)_{\mathcal{S}}\}_{j\leq\ell}\right]. \tag{2}$$

(PII) **filtering**, where we minimize the MSE of estimating the current state using past observations, i.e., we seek $\hat{\boldsymbol{x}}_k = \mathbb{E}\left[\boldsymbol{x}_k \mid \{(\boldsymbol{y}_j)_{\mathcal{S}}\}_{j\leq k}\right]$. For generality, we can pose this problem for a m-steps window starting at $k = \ell$ as

$$\begin{aligned} \underset{\mathcal{S}\subseteq[p]}{\text{minimize}} \quad & \sum_{k=0}^{m-1} \theta_k \, J_{\text{PII},\ell+k}(\mathcal{S}) \\ \text{subject to} \quad & |\mathcal{S}| \leq s \end{aligned} \tag{PII}$$

where $\theta_i \geq 0$ and

$$J_{\text{PII},k}(\mathcal{S}) = \min_{\hat{\boldsymbol{x}}_k} \mathbb{E}\left[\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k\|_2^2 \mid \{(\boldsymbol{y}_j)_{\mathcal{S}}\}_{j\leq k}\right]. \tag{3}$$

Note that as opposed to PI, the state at each time k is only estimated based on observations that occurred up to time k.

It is worth noting that the sensing set $\mathcal{S}$ in PI is selected *a priori*, i.e., before the dynamical system operation. This is of interest, for instance, if sensors are to be installed on the selected outputs and then left to monitor the system. On the other hand, PII can be used both for *a priori* and dynamic sensor selection, where a fusion center chooses a limited subset of distributed sensors to activate during each time window. Moreover, it can accommodate different problems depending on the choice of m and θ_k. For instance, PII becomes a myopic sensor selection problem when $m = 1$ [25]. For arbitrary m, PII can optimize the final estimation MSE (for $\theta_k = 0$ for all $k < m-1$ and $\theta_{m-1} = 1$), the m-steps average MSE (for $\theta_k = 1$ for all k), or some weighted average of the error (e.g., $\theta_k = \rho^{m-1-k}$, $\rho < 1$).

Note that PI and PII are in fact joint optimization problems since their cost functions are themselves written as minimizations. However, it is possible to find closed-form expressions for (2) and (3), i.e., expressions for the estimation error given a fixed sensing set $\mathcal{S}$. In the next section, we derive these expressions before proceeding with the study of the sensor selection problems PI and PII.

A. Sensor selection for state estimation

We start by finding a closed-form expression for the MSE in (2). To do so, we cast smoothing as a stochastic estimation problem using lifting. Indeed, proceeding as in [26], note from (1) that estimating $\{\boldsymbol{x}_k\}_{k\leq\ell}$ is equivalent to estimating $\{\boldsymbol{x}_0, \boldsymbol{w}_k\}_{k\leq\ell-1}$. Therefore, defining the stacked vectors $\bar{\boldsymbol{y}}_\ell = \left[\; (\boldsymbol{y}_0)_{\mathcal{S}}^T \;\cdots\; (\boldsymbol{y}_\ell)_{\mathcal{S}}^T \;\right]^T$, $\bar{\boldsymbol{z}}_\ell = \left[\; \boldsymbol{x}_0^T \;\; \boldsymbol{w}_0^T \;\cdots\; \boldsymbol{w}_{\ell-1}^T \;\right]^T$, and $\bar{\boldsymbol{v}}_\ell = \left[\; (\boldsymbol{v}_0^T)_{\mathcal{S}} \;\cdots\; (\boldsymbol{v}_\ell^T)_{\mathcal{S}} \;\right]^T$, it is straightforward that [30]

$$\bar{\boldsymbol{y}}_\ell = \boldsymbol{O}_\ell(\mathcal{S})\bar{\boldsymbol{z}}_\ell + \bar{\boldsymbol{v}}_\ell, \tag{4}$$

with $\boldsymbol{O}_\ell(\mathcal{S}) = \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{H})\right]\boldsymbol{\Phi}_\ell$, where $\boldsymbol{S}$ is the $|\mathcal{S}| \times n$ selection matrix obtained by keeping only rows from the identity matrix that have indices in $\mathcal{S}$, $\otimes$ denotes the Kronecker product, and

$$\boldsymbol{\Phi}_k = \begin{bmatrix} \boldsymbol{I} & & & \\ \boldsymbol{F} & \boldsymbol{I} & & \\ \vdots & \vdots & \ddots & \\ \boldsymbol{F}^k & \boldsymbol{F}^{k-1} & \cdots & \boldsymbol{I} \end{bmatrix}.$$

Then, the following proposition holds:

Proposition 1: The value of the minimum in (2) is

$$J_{\text{PI}}(\mathcal{S}) = \operatorname{Tr}\left\{\left[\boldsymbol{C}^{-1} + \boldsymbol{Z}(\mathcal{S})\right]^{-1}\right\}, \tag{5}$$

where $\boldsymbol{Z}(\mathcal{A}) = \sum_{i\in\mathcal{A}} \sigma_{v,i}^{-2}\, \boldsymbol{\Phi}_\ell^T \left(\boldsymbol{I} \otimes \boldsymbol{h}_i\boldsymbol{h}_i^T\right)\boldsymbol{\Phi}_\ell$ and $\boldsymbol{h}_i^T$ is the i-th row of $\boldsymbol{H}$.

Proof: See appendix. ∎

For the filtering problem PII, we note from (3) that the state estimate only depends on past observations. Therefore, we are interested in recursively updating this estimate for each new observation using a Kalman filter. Here, a previous state estimate $\hat{\boldsymbol{x}}_{k-1}$ with error covariance matrix $\boldsymbol{P}_{k-1} = \mathbb{E}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^T$ is updated based on a new measurement $(\boldsymbol{y}_k)_{\mathcal{S}}$ using

$$\hat{\boldsymbol{x}}_k = \boldsymbol{F}\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{K}_k\left[(\boldsymbol{y}_k)_{\mathcal{S}} - \boldsymbol{S}\boldsymbol{H}\boldsymbol{F}\hat{\boldsymbol{x}}_{k-1}\right], \tag{6}$$

where $\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}^T\boldsymbol{S}^T\boldsymbol{R}_{e,k}^{-1}$ is the Kalman gain, $\boldsymbol{R}_{e,k} = \boldsymbol{S}(\boldsymbol{H}\boldsymbol{P}_{k|k-1}\boldsymbol{H}^T + \boldsymbol{R}_v)\boldsymbol{S}^T$ is the innovation covariance matrix, and $\boldsymbol{P}_{k|k-1} = \boldsymbol{F}\boldsymbol{P}_{k-1}\boldsymbol{F}^T + \boldsymbol{R}_w$ is the *a priori* error covariance matrix. Recall that $\boldsymbol{S}$ is the selection matrix of the set $\mathcal{S}$. The error covariance of the estimate in (6) is given by

$$\boldsymbol{P}_k(\mathcal{S}) = \left[\boldsymbol{P}_{k|k-1}^{-1} + \boldsymbol{H}^T\boldsymbol{S}^T\left(\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T\right)^{-1}\boldsymbol{S}\boldsymbol{H}\right]^{-1}, \tag{7}$$

with $\boldsymbol{P}_{0|-1} = \boldsymbol{\Pi}_0 \succ 0$ [30]. The estimation MSE is the trace of the estimation error covariance matrix (7), so that the solution of (3) can be written as

$$J_{\text{PII}}(\mathcal{S}, k) = \operatorname{Tr}\left[\left(\boldsymbol{P}_{k|k-1}^{-1} + \sum_{i\in\mathcal{S}} \sigma_{v,i}^{-2}\,\boldsymbol{h}_i\boldsymbol{h}_i^T\right)^{-1}\right], \tag{8}$$

where again $\boldsymbol{h}_i^T$ is the i-th row of $\boldsymbol{H}$ and we used the fact that, since $\boldsymbol{R}_v$ is diagonal, $\left(\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T\right)^{-1} = \boldsymbol{S}\boldsymbol{R}_v^{-1}\boldsymbol{S}^T$.

Using the closed-form expressions in (5) and (8), PI and PII become cardinality constrained set function optimization problems which are NP-hard in general [10], [13]–[16], [31]. Their solutions can therefore only be approximated in practice, which is typically done using either a (possibly sparsity-promoting) convex relaxation [16], [18]–[21], [23] or greedy algorithms [16]. Nevertheless, neither approach comes with performance guarantees. Indeed, though greedy search is near-optimal for the minimization of monotonically decreasing supermodular set functions [32], neither (5) nor (8) are supermodular in general [13]–[16], [31]. To recover suboptimality certificates, surrogate supermodular figures of merit are typically used instead of the MSE in PI and PII by for instance replacing the trace in (5) and (8) by the $\log\det$ [20], [25]–[27]. Although this cost function is related to the volume of the estimation error ellipsoid [19], [26], it introduces distortions in solution due to the *flatness* of the logarithm function.

To provide approximation certificates for the greedy minimization of the MSE in PI and PII, the following section presents and develops the theory of *approximately supermodular functions*, showing that members of this class can be near-optimally minimized (Section III). Then, we study a general class of *set trace functions* of which PI and PII are part of and give explicit suboptimality bounds for their greedy minimization (Section IV).

III. APPROXIMATE SUPERMODULARITY

Supermodularity (*submodularity*) encodes a "diminishing returns" property of certain set functions that allows suboptimality bounds on their greedy minimization (maximization) to be derived. Well-known representatives of this class include the rank or $\log\det$ of a sum of PSD matrices, the Shannon entropy, and the mutual information [31], [33]. Still, supermodularity is a stringent condition. In particular, it does not hold for the cost functions in (5) and (8) [13]–[15], [31].

The purpose of *approximate supermodularity* (*submodularity*) proposed in [29] is to allow certain levels of violations of the original "diminishing returns" property. The rationale is that if a function is "almost" supermodular, then it should behave similar to a supermodular function. In what follows, we formalize and quantify these statements.

Say a set function $f : 2^{\mathcal{V}} \to \mathbb{R}$ is α-*supermodular*, for some $\alpha \in \mathbb{R}_+$, if α is the largest number for which it holds that

$$f(\mathcal{A}) - f(\mathcal{A} \cup \{u\}) \geq \alpha \left[f(\mathcal{B}) - f(\mathcal{B} \cup \{u\})\right] \tag{9}$$

for all sets $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and all $u \in \mathcal{V} \setminus \mathcal{B}$. Explicitly,

$$\alpha = \min_{\substack{\mathcal{A}\subseteq\mathcal{B}\subseteq\mathcal{V} \\ u\in\mathcal{V}\setminus\mathcal{B}}} \frac{f(\mathcal{A}) - f(\mathcal{A}\cup\{u\})}{f(\mathcal{B}) - f(\mathcal{B}\cup\{u\})}. \tag{10}$$

We say f is α-submodular if $-f$ is α-supermodular. Notice that for $\alpha \geq 1$, (9) is equivalent to the traditional definition of supermodularity, in which case we refer to the function simply as *supermodular* [31], [33]. For $\alpha \in (0, 1)$, however, we say f is *approximately supermodular*. Notice that (9) always holds for $\alpha = 0$ if f is monotone decreasing. Indeed, a set function f is *monotone decreasing* if for all $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ it holds that $f(\mathcal{A}) \geq f(\mathcal{B})$.

Note that α in (10) is related to the *submodularity ratio* defined in [11]. It is, however, more amenable to give explicit (P-computable) bounds on its value because it is a ratio of rank one updates (see Section IV). In fact, the submodularity ratio bounds derived in [11] depend on the minimum sparse eigenvalue of a matrix, which is NP-hard to evaluate.

Before proceeding, we derive the following property of α-supermodular functions that will come in handy when studying J_{PII}:

Proposition 2: Let $f_i : 2^{\mathcal{V}} \to \mathbb{R}$ be α_i-supermodular functions. Then, $g = \sum_i \theta_i f_i$, $\theta_i \geq 0$, is $\min(\alpha_i)$-supermodular.

Proof: From the definition of α-supermodularity in (9), for $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and $u \notin \mathcal{B}$ it holds that

$$\begin{aligned} g(\mathcal{A}) - g(\mathcal{A} \cup \{u\}) &= \sum_i \theta_i \left[f_i(\mathcal{A}) - f_i(\mathcal{A} \cup \{u\})\right] \\ &\geq \sum_i \alpha_i \theta_i \left[f_i(\mathcal{B}) - f_i(\mathcal{B} \cup \{u\})\right]. \end{aligned}$$

Using the fact that $\alpha_i \geq \min(\alpha_i)$ we then have

$$\begin{aligned} &g(\mathcal{A}) - g(\mathcal{A} \cup \{u\}) \geq \\ &\quad \min(\alpha_i) \sum_i \theta_i \left[f_i(\mathcal{B}) - f_i(\mathcal{B} \cup \{u\})\right] \geq \\ &\quad\quad \min(\alpha_i) \left[g(\mathcal{B}) - g(\mathcal{B} \cup \{u\})\right]. \end{aligned}$$

■

Following a similar argument as in [32], we can now show that α-supermodular functions can be near-optimally minimized using greedy search:

Theorem 1: Let $f^\star = f(\mathcal{S}^\star)$ be the optimal value of the problem

$$\begin{aligned} &\underset{\mathcal{S}\subseteq\mathcal{V}}{\text{minimize}} && f(\mathcal{S}) \\ &\text{subject to} && |\mathcal{S}| \leq s \end{aligned} \tag{PIII}$$

and $\mathcal{G}_q$ be the q-th iteration of its greedy solution, obtained by taking $\mathcal{G}_0 = \emptyset$ and repeating for $j = 1, \ldots, q$

$$u = \operatorname{argmin}_{v\in\mathcal{V}} f(\mathcal{G}_{j-1} \cup \{v\}) \tag{11a}$$

$$\mathcal{G}_j = \mathcal{G}_{j-1} \cup \{u\} \text{ and } \mathcal{V} = \mathcal{V} \setminus \{u\} \tag{11b}$$

If f is (i) monotone decreasing and (ii) α-supermodular, then

$$\frac{f(\mathcal{G}_\ell) - f^\star}{f(\emptyset) - f^\star} \leq \left(1 - \frac{\alpha}{s}\right)^q \leq e^{-\alpha q/s}. \tag{12}$$

Proof: Since f is monotone decreasing, it holds for every set $\mathcal{G}_j$ that

$$\begin{aligned} f(\mathcal{S}^\star) &\geq f(\mathcal{S}^\star \cup \mathcal{G}_j) \\ &= f(\mathcal{G}_j) - \sum_{i=1}^{s} f(\mathcal{T}_{i-1}) - f(\mathcal{T}_{i-1} \cup u_i^\star), \end{aligned} \tag{13}$$

where $\mathcal{T}_0 = \mathcal{G}_j$, $\mathcal{T}_i = \mathcal{G}_j \cup \{u_1^\star, \dots, u_i^\star\}$, and $u_i^\star$ is the i-th element of $\mathcal{S}^\star$. Notice that (13) holds regardless of the order in which the $u_i^\star$ are taken. Using the fact that f is α-supermodular and $\mathcal{G}_j \subseteq \mathcal{T}_i$ for all i, the incremental gains in (13) can be bounded using (9) to get

$$f(\mathcal{S}^\star) \geq f(\mathcal{G}_j) - \alpha^{-1} \sum_{i=1}^{s} f(\mathcal{G}_j) - f(\mathcal{G}_j \cup u_i^\star). \qquad (14)$$

Finally, given that $\mathcal{G}_{j+1}$ is chosen as to maximize the gain in (14) [see (11)],

$$f(\mathcal{S}^\star) \geq f(\mathcal{G}_j) - \alpha^{-1} s \left[f(\mathcal{G}_j) - f(\mathcal{G}_{j+1})\right]. \qquad (15)$$

To obtain the expression in (12), suffices to solve the recursion in (15). To do so, let $\delta_j = f(\mathcal{G}_j) - f(\mathcal{S}^\star)$ so that (15) becomes

$$\delta_j \leq \alpha^{-1} k \left[\delta_j - \delta_{j+1}\right] \Rightarrow \delta_{j+1} \leq \left(1 - \frac{1}{\alpha^{-1} s}\right) \delta_j.$$

Noting that $\delta_0 = f(\emptyset) - f(\mathcal{S}^\star)$, we can find a direct expression for this recursion:

$$\frac{f(\mathcal{G}_\ell) - f(\mathcal{S}^\star)}{f(\emptyset) - f(\mathcal{S}^\star)} \leq \left(1 - \frac{\alpha}{s}\right)^q.$$

Using the fact that $1 - x \leq e^{-x}$ yields (12). ■

Theorem 1 bounds the relative suboptimality of the greedy solution of PIII when f is decreasing and α-supermodular. Under these conditions, it ensures a minimum improvement over the empty set. More importantly, (12) quantifies the effect of violating supermodularity in (9). Indeed, when $\alpha = 1$ (i.e., when f is supermodular) and the greedy search in (11) is repeated s times ($q = s$), we get the classical $e^{-1} \approx 0.37$ guarantee from [32]. On the other hand, if f is approximately supermodular ($\alpha < 1$), (12) shows that the same 37% guarantee is recovered if we greedily select a set of size s/α. Hence, α not only measures how much f violates supermodularity, but also gives a factor by which a greedy solution set must grow to maintain supermodular-like near-optimality. It is worth noting that, as with the original bound in [32], (12) is not tight and that better results are typically obtained in practice (see Section V).

Although Theorem 1 characterizes the loss in suboptimality incurred from violating supermodularity, its performance certificate depends on the value of α. Unfortunately, (10) reveals that finding α for a general function is a combinatorial problem. To give actual near-optimal guarantees for the greedy solution of PI and PII, the next section provides a bound on α for a class of set trace functions of which the cost functions (5) and (8) are particular cases. We then derive specific results on the α-supermodularity of these functions.

IV. AN α-SUPERMODULAR TRACE FUNCTION

Let $h : \mathbb{S}_+ \to \mathbb{R}$ be a *trace function* if it is of the form $h(\boldsymbol{M}) = \operatorname{Tr}\left[g\left(\boldsymbol{M}\right)\right]$ for some function g. Examples of trace functions include the $\log\det$ and the Von-Neumann entropy [34]. Here, we are interested in a specific family of trace functions defined on sets. Formally, we study *set trace functions* $h : 2^{\mathcal{V}} \to \mathbb{R}$ of the general form

$$h(\mathcal{A}) = \operatorname{Tr}\left[\left(\boldsymbol{M}_\emptyset + \sum_{i \in \mathcal{A}} \boldsymbol{M}_i\right)^{-1}\right], \qquad (16)$$

where $\mathcal{A} \subseteq \mathcal{V}$, $\boldsymbol{M}_\emptyset \succ 0$, and $\boldsymbol{M}_i \succeq 0$ for all $i \in \mathcal{V}$. Note that the cost functions of PI and PII are both of the form (16). Indeed, (5) takes $\boldsymbol{M}_\emptyset = \boldsymbol{C}^{-1}$ and $\boldsymbol{M}_i = \sigma_{v,i}^{-2} \boldsymbol{\Phi}_\ell^T \left(\boldsymbol{I} \otimes \boldsymbol{h}_i \boldsymbol{h}_i^T\right) \boldsymbol{\Phi}_\ell$, whereas (8) for $m = 1$ takes $\boldsymbol{M}_\emptyset = \boldsymbol{P}_{m|m-1}^{-1}$ and $\boldsymbol{M}_i = \sigma_{v,i}^{-2} \boldsymbol{h}_i \boldsymbol{h}_i^T$. We use Proposition 2 to study the $m > 1$ case in Section IV-B.

Functions as in (16) display neither submodular nor supermodular behavior in general. Indeed, counter-examples can be found in [12], [14], [15]. We can, however, show that they are monotonically decreasing and give a closed-form (P-computable) bound on their α-supermodularity.

Theorem 2: The set trace function h in (16) is (i) monotone decreasing and (ii) α-supermodular with

$$\alpha \geq \frac{\mu_{\min}}{\mu_{\max}} > 0, \qquad (17)$$

where

$$0 < \mu_{\min} \leq \lambda_{\min}\left[\boldsymbol{M}_\emptyset\right] \leq \lambda_{\max}\left[\boldsymbol{M}_\emptyset + \sum_{i \in \mathcal{V}} \boldsymbol{M}_i\right] \leq \mu_{\max}$$

and $\lambda_{\max}(\boldsymbol{M})$ and $\lambda_{\min}(\boldsymbol{M})$ denote the maximum and minimum eigenvalues of $\boldsymbol{M}$, respectively.

Remark 1: Although there exist examples in which the set trace function (16) is not supermodular, the general statement of Theorem 2 does not allow us to claim that $\alpha < 1$. A simple counter-example involves the case in which $\boldsymbol{M}_\emptyset = \mu_0 \boldsymbol{I}$ and $\boldsymbol{M}_i = \mu_i \boldsymbol{I}$, $\mu_0, \mu_i \geq 0$, so that (16) is effectively a scalar function of μ_0, μ_i. Since scalar convex functions of positive modular functions are supermodular [35], $\alpha \geq 1$ in this case.

Proof: The monotone decreasing nature of h [(i)] is a direct consequence of the fact that matrix inversion is an operator antitone function, i.e., that for $\boldsymbol{X}, \boldsymbol{Y} \succeq 0$, it holds that $\boldsymbol{X} \succeq \boldsymbol{Y} \Leftrightarrow \boldsymbol{X}^{-1} \preceq \boldsymbol{Y}^{-1}$ [36]. Indeed, let $\boldsymbol{Y}(\mathcal{X}) = \boldsymbol{M}_\emptyset + \sum_{i \in \mathcal{X}} \boldsymbol{M}_i$. Then, $\boldsymbol{M}_\emptyset, \boldsymbol{M}_i \succeq 0$ implies that

$$\mathcal{A} \subseteq \mathcal{B} \Leftrightarrow \boldsymbol{Y}(\mathcal{A}) \preceq \boldsymbol{Y}(\mathcal{B}) \Leftrightarrow \boldsymbol{Y}(\mathcal{A})^{-1} \succeq \boldsymbol{Y}(\mathcal{B})^{-1}. \qquad (18)$$

From the monotonicity of the trace [34] we have

$$\mathcal{A} \subseteq \mathcal{B} \Leftrightarrow h(\mathcal{A}) \geq h(\mathcal{B}).$$

To lower bound α [(ii)], we need to bound the gain incurred by adding element $i \notin \mathcal{A}$ to set $\mathcal{A}$, i.e., $\Delta_i(\mathcal{A}) = h\left(\mathcal{A} \cup \{i\}\right) - h\left(\mathcal{A}\right)$. Indeed, from the definition of α in (10) we can write

$$\alpha = \min_{\substack{\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V} \\ i \in \mathcal{V} \setminus \mathcal{B}}} \frac{\Delta_i(\mathcal{A})}{\Delta_i(\mathcal{B})}. \qquad (19)$$

To do so, note that $\boldsymbol{Y}(\mathcal{A})$ is an additive (modular) function of $\mathcal{A}$, so it holds that $h\left(\mathcal{A} \cup \{i\}\right) = \operatorname{Tr}\left[\left(\boldsymbol{Y}(\mathcal{A}) + \boldsymbol{M}_i\right)^{-1}\right]$. Since $\boldsymbol{Y}(\mathcal{A}) \succ 0$ because $\boldsymbol{M}_\emptyset \succ 0$ but $\boldsymbol{M}_i$ need not be invertible, we use an alternative formulation of the matrix inversion lemma [37] to obtain

$$h\left(\mathcal{A}\cup\{i\}\right)=\operatorname{Tr}\left[\boldsymbol{Y}(\mathcal{A})^{-1}-\boldsymbol{Y}(\mathcal{A})^{-1}\boldsymbol{M}_i\left[\boldsymbol{Y}(\mathcal{A})+\boldsymbol{M}_i\right]^{-1}\right].$$

From the linearity of the trace [38], the gain can be written as

$$\Delta_i(\mathcal{A})=\operatorname{Tr}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\boldsymbol{M}_i\left[\boldsymbol{Y}(\mathcal{A})+\boldsymbol{M}_i\right]^{-1}\right]. \tag{20}$$

The goal is now to explicitly lower bound (20) by exploiting spectral bounds. In particular, we will use the following lemma whose proof is sketched in the appendices:

Lemma 1: For all $\mathcal{A}\subseteq\mathcal{V}$ and $i\in\mathcal{V}\setminus\mathcal{A}$, it holds that

$$\begin{aligned}\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\right]\operatorname{Tr}\left[\boldsymbol{M}_i\left[\boldsymbol{Y}(\mathcal{A})+\boldsymbol{M}_i\right]^{-1}\right]&\leq\Delta_i(\mathcal{A})\\ \leq\lambda_{\max}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\right]\operatorname{Tr}\left[\boldsymbol{M}_i\left[\boldsymbol{Y}(\mathcal{A})+\boldsymbol{M}_i\right]^{-1}\right]&.\end{aligned} \tag{21}$$

Applying Lemma 1 to (25) yields

$$\alpha\geq\frac{\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\right]}{\lambda_{\max}\left[\boldsymbol{Y}(\mathcal{B})^{-1}\right]}\times\frac{\operatorname{Tr}\left[\boldsymbol{M}_i\left[\boldsymbol{Y}(\mathcal{A})+\boldsymbol{M}_i\right]^{-1}\right]}{\operatorname{Tr}\left[\boldsymbol{M}_i\left[\boldsymbol{Y}(\mathcal{B})+\boldsymbol{M}_i\right]^{-1}\right]}. \tag{22}$$

Note now that $\mathcal{A}\subseteq\mathcal{B}$ implies that $\left[\boldsymbol{Y}(\mathcal{A})+\boldsymbol{M}_i\right]^{-1}\succeq\left[\boldsymbol{Y}(\mathcal{B})+\boldsymbol{M}_i\right]^{-1}$, since $\boldsymbol{Y}^{-1}$ is a decreasing set function per (18). From the ordering of the PSD cone, the second term in (22) is therefore lower bounded by one. Since $\boldsymbol{Y}\succ 0$, it holds that $\sigma_i(\boldsymbol{Y})=\lambda_i(\boldsymbol{Y})$ and we get

$$\alpha\geq\frac{\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{B})\right]}{\lambda_{\max}\left[\boldsymbol{Y}(\mathcal{A})\right]}.$$

Finally, the lower bound in (17) is obtained by observing that, since $\boldsymbol{Y}$ is increasing, the following inequalities hold for any set $\mathcal{X}\subseteq\mathcal{V}$:

$$\lambda_{\min}\left(\boldsymbol{M}_\emptyset\right)\leq\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{X})\right]\leq\lambda_{\max}\left[\boldsymbol{Y}(\mathcal{X})\right]\leq\lambda_{\max}\left[\boldsymbol{Y}(\mathcal{V})\right].$$

■

Theorem 2 gives a deceptively simple bound on the α-supermodularity of set trace function (16) as a function of the spectrum of the $\boldsymbol{M}_\emptyset,\boldsymbol{M}_i$. Still, it is tighter than the one provided in [29].

Note that bound (17) is akin to the inverse of the condition number of the matrix-valued set function underlying h. Indeed, let $h(\mathcal{X})=\operatorname{Tr}\left[\boldsymbol{Y}(\mathcal{X})^{-1}\right]$ with $\boldsymbol{Y}(\mathcal{X})=\boldsymbol{M}_\emptyset+\sum_{i\in\mathcal{X}}\boldsymbol{M}_i$. Then, since $\boldsymbol{Y}^{-1}$ is symmetric, its condition number with respect to the spectral norm can be written as $\kappa[\boldsymbol{Y}^{-1}(\mathcal{X})]=\lambda_{\max}[\boldsymbol{Y}(\mathcal{X})]/\lambda_{\min}[\boldsymbol{Y}(\mathcal{X})]$ for any $\mathcal{X}\subseteq\mathcal{V}$. However, a more interesting geometrical interpretation of (17) can be given in terms of the "range" of $\boldsymbol{Y}^{-1}$. To see this, let the *numerical range* of $\boldsymbol{Y}^{-1}$ be

$$W_{\mathcal{V}}(\boldsymbol{Y}^{-1})=W\left[\bigoplus_{\mathcal{X}\subseteq\mathcal{V}}\boldsymbol{Y}(\mathcal{X})^{-1}\right], \tag{23}$$

where $\boldsymbol{A}\oplus\boldsymbol{B}=\operatorname{blkdiag}(\boldsymbol{A},\boldsymbol{B})$ is the direct sum of $\boldsymbol{A}$ and $\boldsymbol{B}$ and $W(\boldsymbol{M})=\{\boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x}\mid\|\boldsymbol{x}\|_2=1\}$ is the classical numerical range. Since the numerical range is a convex set [38], define its relative diameter as

$$\Delta=\max_{\mu,\eta\in W_{\mathcal{V}}(\boldsymbol{Y}^{-1})}\left|\frac{\mu-\eta}{\mu}\right|. \tag{24}$$

Then, the following holds:

Proposition 3: The set trace functions h in (16) is α-supermodular with

$$\alpha\geq\frac{\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{V})^{-1}\right]}{\lambda_{\max}\left[\boldsymbol{Y}(\emptyset)^{-1}\right]}=1-\Delta, \tag{25}$$

where Δ is the relative diameter of the numerical range of the underlying matrix-valued set function of h in (23).

Proof: Since $\boldsymbol{Y}\succ 0$, the numerical range in (23) is the bounded convex hull of the eigenvalues of $\boldsymbol{Y}(\mathcal{X})^{-1}$ for all $\mathcal{X}\subseteq\mathcal{V}$ [38]. We can therefore simplify (24) using the fact that it is monotonically increasing in μ and decreasing in η. Explicitly,

$$\Delta=\max_{\mathcal{X},\mathcal{Y}\subseteq\mathcal{V}}\left|\frac{\lambda_{\max}\left[\boldsymbol{Y}(\mathcal{Y})^{-1}\right]-\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{X})^{-1}\right]}{\lambda_{\max}\left[\boldsymbol{Y}(\mathcal{Y})^{-1}\right]}\right|.$$

From the antitonicity of matrix inversion [36], this maximum is achieved for

$$\Delta=\frac{\lambda_{\max}\left[\boldsymbol{Y}(\emptyset)^{-1}\right]-\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{V})^{-1}\right]}{\lambda_{\max}\left[\boldsymbol{Y}(\emptyset)^{-1}\right]},$$

which together with Theorem 2 yields

$$\alpha\geq\frac{\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{V})^{-1}\right]}{\lambda_{\max}\left[\boldsymbol{Y}(\emptyset)^{-1}\right]}=1-\Delta.$$

■

Therefore, (17) bounds how much h deviates from supermodularity (as quantified by α) in terms of the numerical range of its underlying function $\boldsymbol{Y}^{-1}$. The shorter the range of $\boldsymbol{Y}^{-1}$, the more supermodular-like (16) will be. As we show in the sequel, this is closely related in our applications to some measure of signal-to-noise ratio (SNR).

A. α-supermodular results for smoothing

From Theorem 2, we can bound the α-supermodularity of the cost function of PI as

$$\alpha\geq\frac{\lambda_{\min}(\boldsymbol{C}^{-1})}{\lambda_{\max}\left[\boldsymbol{C}^{-1}+\boldsymbol{\Phi}_\ell^T\left[\boldsymbol{I}\otimes\boldsymbol{H}^T\boldsymbol{R}_v^{-1}\boldsymbol{H}\right]\boldsymbol{\Phi}_\ell\right]}. \tag{26}$$

Although (26) depends only on the model description (1), it does not provide insight as to when α will be close to one. It is therefore worth studying the special case of $\boldsymbol{\Pi}_0=\boldsymbol{R}_w=\sigma_w^2\boldsymbol{I}$, $\boldsymbol{R}_v=\sigma_v^2\boldsymbol{I}$, and $\boldsymbol{H}=\boldsymbol{I}$, for which (26) can be simplified.

Indeed, recall that $\boldsymbol{C}$ is block diagonal and therefore $\lambda_{\min}\left(\boldsymbol{C}^{-1}\right)=\min[\lambda_{\max}(\boldsymbol{\Pi}_0)^{-1},\lambda_{\max}(\boldsymbol{R}_w)^{-1}]=\sigma_w^{-2}$. Also, from the spectral properties of the Kronecker product, it holds that $\lambda_{\max}(\boldsymbol{I}\otimes\boldsymbol{H}^T\boldsymbol{R}_v^{-1}\boldsymbol{H})=\lambda_{\max}(\boldsymbol{R}_v^{-1})=\sigma_v^{-2}$. Weyl's inequalities [38] and the spectral bounds on the product of PSD matrices from [39] therefore allow us to write

$$\alpha\geq\frac{1}{1+\sigma_v^{-2}\sigma_w^2\,\lambda_{\max}\left(\boldsymbol{\Phi}_\ell^T\boldsymbol{\Phi}_\ell\right)}.$$

To obtain a bound as an explicit function of $\boldsymbol{F}$, note that $\lambda_{\max}\left(\boldsymbol{\Phi}_\ell^T \boldsymbol{\Phi}_\ell\right) = \|\boldsymbol{\Phi}_\ell\|_2^2$, where $\|\boldsymbol{M}\|_2$ denotes the spectral norm of $\boldsymbol{M}$. The norm of $\boldsymbol{\Phi}_k$ can be bounded in terms of the norm of its blocks using the inequalities from [40]. Explicitly,

$$\alpha \geq \frac{1}{1 + \sigma_v^{-2}\sigma_w^2 \sum_{k=0}^{\ell} (\ell + 1 - k) \left\|\boldsymbol{F}^k\right\|_2^2}. \tag{27}$$

As we alluded to earlier, the ratio in the denominator of (27) plays the role of an SNR relating the state signal power (in terms of σ_w^2 and $\|\boldsymbol{F}\|_2$) and the measurement noise power (σ_v^2). When this SNR is high, i.e., the measurement noise level is small compared to the driving noise, the bound on α decreases weakening the guarantees from Theorem 1. This case is however of less practical value. Indeed, if the process noise dominates the estimation error, the system trajectory is mostly random and the choice of sensing subset has little impact on the estimation performance. In the limit, the best estimate of the system states is given by the instantaneous measurements [30], [41]. A similar argument holds for when the system is close to instability, i.e., the norm of $\boldsymbol{F}$ is close to one.

On the other hand, for low SNRs, the denominator of (27) decreases and Theorem 2 guarantees that the estimation MSE displays supermodular-like behavior. This occurs when the measurement noise is large compared to the process noise and the system has fast decaying modes ($\|\boldsymbol{F}\|_2 \ll 1$).

B. α-supermodular results for Kalman filtering

Using (6)–(8), the bound in (17) reads

$$\alpha_k \geq \frac{\lambda_{\min}(\boldsymbol{P}_{k|k-1}^{-1})}{\lambda_{\max}\left(\boldsymbol{P}_{k|k-1}^{-1} + \boldsymbol{H}^T \boldsymbol{R}_v^{-1} \boldsymbol{H}\right)},$$

for $k \geq 0$. From the Riccati equation (7), this simplifies to

$$\alpha_k \geq \frac{\lambda_{\min}(\boldsymbol{P}_{k|k-1}^{-1})}{\lambda_{\max}\left[\boldsymbol{P}_k([p])^{-1}\right]} = \frac{\lambda_{\min}\left[\boldsymbol{P}_k([p])\right]}{\lambda_{\max}\left(\boldsymbol{P}_{k|k-1}\right)}, \tag{28}$$

which is the ratio between the *a posteriori* error using all sensors and the *a priori* error. Using Proposition 2, the cost function in (8) is therefore α-supermodular with

$$\alpha \geq \min_{\ell \leq k \leq \ell+m-1} \alpha_\ell. \tag{29}$$

The bound (28) is large when the *a priori* and *a posteriori* error covariance matrices are similar. To get additional insight into when this occurs, we proceed as in Section IV-A and study the particular case in which $\boldsymbol{R}_w = \sigma_w^2 \boldsymbol{I}$ and $\boldsymbol{R}_v = \sigma_v^2 \boldsymbol{I}$. Then, from (6) and (7), the matrices of interest take the form

$$\boldsymbol{P}_{k|k-1} = \boldsymbol{F}\boldsymbol{P}_{k-1}\boldsymbol{F}^T + \sigma_w^2 \boldsymbol{I} \tag{30a}$$

$$\boldsymbol{P}_k([p]) = \left(\boldsymbol{P}_{k|k-1}^{-1} + \sigma_v^{-2}\boldsymbol{H}^T\boldsymbol{H}\right)^{-1} \tag{30b}$$

Note from (30a) that $\boldsymbol{P}_{k|k-1}$ depends directly on σ_w^2. Thus, from (30b), $\boldsymbol{P}_k([p]) \approx \boldsymbol{P}_{k|k-1}$ when $\sigma_v^2 \gg \sigma_w^2$. It is important to highlight that this is the scenario in which Kalman filters are most useful. Indeed, as in the smoothing case from Section IV-A, the filter tracking capability and stability are hindered if the driving noise power is much larger than the measurement noise power [41].

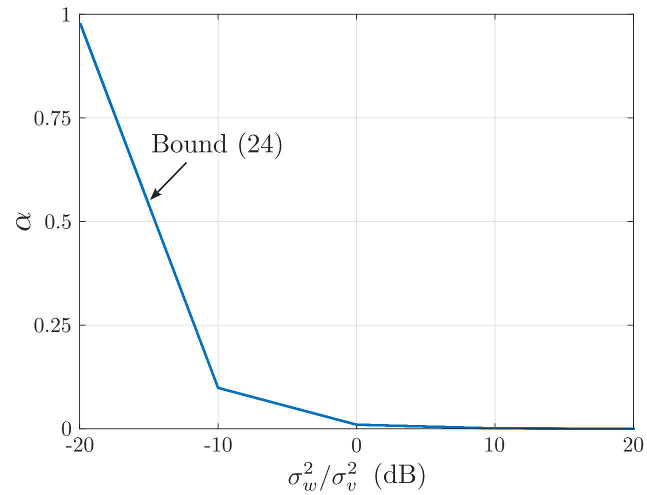

Fig. 1. α-supermodularity bound (26) for different process noise variances.

Additionally, the bound (28) will approach one only if $\boldsymbol{P}_{k|k-1}$ is well-conditioned. When σ_w^2 is large, this stems directly from (30a). On the other hand, if $\sigma_w^2 \ll 1$, $\boldsymbol{P}_{k|k-1} \approx \boldsymbol{F}\boldsymbol{P}_{k-1}\boldsymbol{F}^T$ and its condition number depends directly on that of $\boldsymbol{F}$, since $\kappa(\boldsymbol{P}_{k|k-1}) \leq \kappa(\boldsymbol{P}_{k-1})\kappa(\boldsymbol{F})^2$, where κ is the condition number with respect to the spectral norm [38]. Thus, as in the batch case, the guarantees from Theorem 1 will be stronger when $\sigma_v^2 \gg \sigma_w^2$ and when the system has fast decaying modes with similar rates ($\|\boldsymbol{F}\|_2 \ll 1$ and $\kappa(\boldsymbol{F}) \approx 1$).

V. SIMULATIONS

We start by evaluating the bounds derived in Section IV for different noise and system decay rates. Our goal is to illustrate the situations in which they are close to one, i.e., in which we can guarantee that the state estimation MSE is close supermodular. In Fig. 1 we show results for the smoothing problem, where $n = 100$, $k = 15$, $\boldsymbol{H} = \boldsymbol{I}$, $\boldsymbol{\Pi}_0 = 10^{-2}\boldsymbol{I}$, $\boldsymbol{R}_v = \boldsymbol{I}$, and $\boldsymbol{R}_w = \sigma_w^2 \boldsymbol{I}$ with $\sigma_w^2 \in [0.01, 100]$. The elements of $\boldsymbol{F}$ were drawn randomly from a standard Gaussian distribution and the matrix was normalized so that the magnitude of its largest eigenvalue is 0.3. We plot 20 system realizations. As noted in Section IV-A, the bound on α improves as σ_w^2/σ_v^2 decreases.

Similar results for the filtering problem are shown Fig. 2a, which considers the myopic problem by taking $m = 1$ and $\ell = 5$ in PII. All other parameters are as described above. In this setting, we also investigate the effect of the spectrum of the state transition matrix on bound (29). To do so, we fix $\sigma_w^2 = 10^{-2}$ and vary $\|\boldsymbol{F}\|_2$. The results are displayed in Fig. 2b and corroborate our observations that the presence of slowly decaying modes ($\|\boldsymbol{F}\|_2 \approx 1$) worsens the guarantees obtained from Theorem 2.

Although the α-supermodularity bound of Theorem 2 provides good guarantees for a reasonable range of parame-

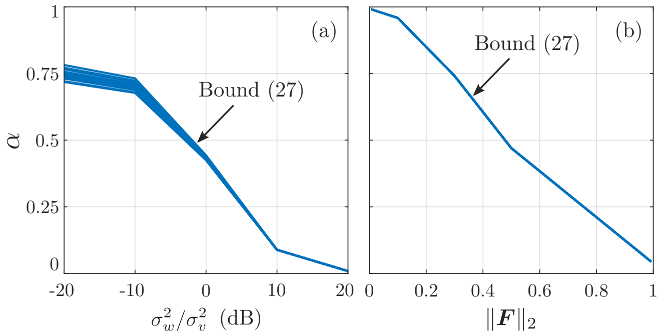

Fig. 2. α-supermodularity bound (29) for (a) different process noise variances and (b) different $\|\boldsymbol{F}\|_2$.

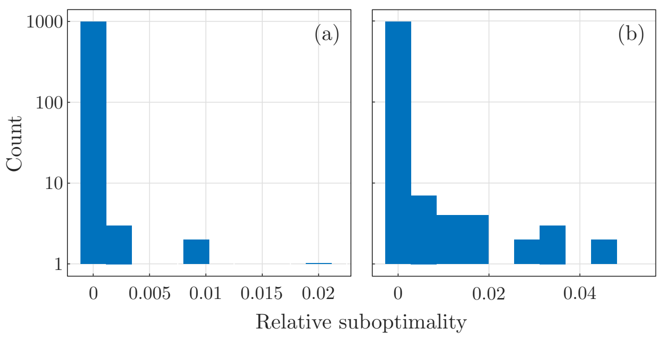

Fig. 3. Relative suboptimality of greedy sensor selection for 1000 system realizations: (a) smoothing and (b) filtering.

ters (see Fig. 1-2), we illustrate that greedy search typically gives better results than those predicted by Theorem 2. We do so by evaluating the relative suboptimality (12) of greedy sensing sets for both smoothing and filtering over 1000 system realizations. Explicitly, if $\mathcal{G}$ is the greedy solution and $\mathcal{S}^\star$ is the optimal solution of PI/PII, we evaluate

$$\nu^\star(\mathcal{G}) = \frac{f(\mathcal{G}) - f(\mathcal{S}^\star)}{f(\emptyset) - f(\mathcal{S}^\star)}.$$

Since $\nu^\star$ depends on the optimal sensing set, we restrict ourselves to small dynamical systems ($n = 10$ states, $p = 10$ outputs, and $s = 4$). This time, both $\boldsymbol{F}$ and $\boldsymbol{H}$ are random Gaussian matrices and $\boldsymbol{F}$ is normalized so that its norm is 0.9. Also, we draw the measurement noise variance at each output uniformly at random in $[10^{-2}, 1]$.

Results for PI are shown in Fig. 3a with $k = 15$. Note that greedy sensor selection finds the optimal sensing set in 99% of the realizations. Moreover, although the bound in (26) gives $\alpha \geq 0.21$ which only guarantees $\nu^\star \leq 0.81$, the relative suboptimality is never more than 0.02 in these simulations. The experiment is repeated for PII in Fig. 3b, except using $\ell = 0$, $m = 10$, and $\theta_i = 1$. In other words, we minimize the 10-steps average MSE. Here, (29) gives $\alpha \geq 0.24$ so that Theorem 1 guarantees $\nu^\star \leq 0.79$. The measured relative suboptimality is nevertheless considerably smaller.

VI. Conclusion

This work studied the Kalman filtering sensor selection problem and provided near-optimal guarantees to its greedy solution. To do so, it introduced and developed the concept of approximate supermodularity, giving bounds on the greedy minimization of this class of functions. This theory was then used to derive performance bounds for the state estimation MSE, which we showed approach the typical $1/e$ guarantee in typical application scenarios. This approach addresses the issue of giving near-optimal guarantees for sensor selection problems without relying on surrogate supermodular cost functions (e.g., the $\log\det$). We expect that the approximate supermodularity can be applied to provide bounds for other cost functions and used to solve cardinality constrained minimization problems in different contexts.

VII. Acknowledgments

The authors thank Vasileios Tzoumas for his kind and helpful reading of our manuscript and for spotting an important flaw in our original arguments. This work is supported by the ARL DCIST CRA W911NF-17-2-0181.

References

[1] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: Beyond the star topology," *IEEE Trans. Signal Process.*, vol. 56[7], pp. 2649–2661, 2008.
[2] Y. Mo, R. Ambrosino, and B. Sinopoli, "Sensor selection strategies for state estimation in energy constrained wireless sensor networks," *Automatica*, vol. 47[7], pp. 1330–1338, 2011.
[3] H. Rowaihy, S. Eswaran, M. Johnson, D. Verma, A. Bar-Noy, T. Brown, and T. La Porta, "A survey of sensor selection schemes in wireless sensor networks," *Proc. of the Int. Soc. for Optics and Photonics*, vol. 6562, 2007.
[4] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Process. Mag.*, vol. 19[2], pp. 61–72, 2002.
[5] H. Zhang, J.M.F. Moura, and B. Krogh, "Dynamic field estimation using wireless sensor networks: Tradeoffs between estimation error and communication cost," *IEEE Trans. Signal Process.*, vol. 57[6], pp. 2383–2395, 2009.
[6] A.O. Hero and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sensors J.*, vol. 11[12], pp. 3064–3075, 2011.
[7] L. Chen, N. Li, L. Jiang, and S.H. Low, *Optimal demand response: Problem formulation and deterministic case*, pp. 63–85, Springer New York, 2012.
[8] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Observability of complex systems," *Proc. of the National Academy of Sciences*, 2012.
[9] G. Thatte and U. Mitra, "Spatio-temporal sensor management for environmental field estimation," *Signal Process.*, vol. 128, pp. 369–381, 2016.
[10] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learning Research*, vol. 9, pp. 235–284, 2008.
[11] A. Das and D. Kempe, "Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection," in *Int. Conf. on Mach. Learning*, 2011.
[12] G. Sagnol, "Approximation of a maximum-submodular-coverage problem involving spectral functions, with application to experimental designs," *Discrete Appl. Math.*, vol. 161[1-2], pp. 258–276, 2013.
[13] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Trans. Signal Process.*, vol. 62[5], pp. 1135–1146, 2014.
[14] V. Tzoumas, M.A. Rahimian, G.J. Pappas, and A. Jadbabaie, "Minimal actuator placement with bounds on control effort," *IEEE Trans. Contr. Netw. Syst.*, vol. 3[1], pp. 67–78, 2016.
[15] A. Olshevsky, "On (non)supermodularity of average control energy," 2017, arXiv:1512.00037v4.
[16] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," *Automatica*, vol. 78, pp. 202–210, 2017.
[17] A. Olshevsky, "Minimal controllability problems," *IEEE Trans. Contr. Netw. Syst.*, vol. 1[3], pp. 249–258, 2014.

[18] J.E. Weimer, B. Sinopoli, and B.H. Krogh, "A relaxation approach to dynamic sensor selection in large-scale wireless networks," in *Int. Conf. on Distributed Computing Syst.*, 2008, pp. 501–506.
[19] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57[2], pp. 451–462, 2009.
[20] N. Moshtagh, L. Chen, and R. Mehra, "Optimal measurement selection for any-time Kalman filtering with processing constraints," in *Conf. on Decision and Contr.*, 2009, pp. 5074–5079.
[21] S.P. Chepuri and G. Leus, "Sensor selection for estimation, filtering, and detection," in *Int. Conf. on Signal Process. and Commun.*, 2014.
[22] N.K. Dhingra, M.R. Jovanović, and Z.Q. Luo, "An ADMM algorithm for optimal sensor and actuator selection," in *Conf. on Decision and Contr.*, 2014, pp. 4039–4044.
[23] S. Liu, S.P. Chepuri, M. Fardad, E. Maşazade, G. Leus, and P.K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Trans. Signal Process.*, vol. 64[13], pp. 3509–3522, 2016.
[24] S. Pequito, S. Kar, and A.P. Aguiar, "A framework for structural input/output and control configuration selection in large-scale systems," *IEEE Trans. Autom. Control*, vol. 61[2], pp. 303–318, 2016.
[25] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Conf. on Decision and Contr.*, 2010, pp. 2572–2577.
[26] V. Tzoumas, A. Jadbabaie, and G.J. Pappas, "Sensor placement for optimal Kalman filtering: Fundamental limits, submodularity, and algorithms," in *American Contr. Conf.*, 2016, pp. 191–196.
[27] T.H. Summers, F.L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Trans. Contr. Netw. Syst.*, vol. 3[1], pp. 91–101, 2016.
[28] G.E.P. Box, J.S. Hunter, and W.G. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery*, Springer, 2005.
[29] L.F.O. Chamon and A. Ribeiro, "Near-optimality of greedy set selection in the sampling of graph signals," in *Global Conf. on Signal and Inform. Process.*, 2016.
[30] T. Kailath, A.H. Sayed, and B. Hassibi, *Linear estimation*, Prentice-Hall, 2000.
[31] A. Krause and D. Golovin, "Submodular function maximization," in *Tractability: Practical Approaches to Hard Problems*. Cambridge University Press, 2014.
[32] G.L. Nemhauser, L.A. Wolsey, and M.L. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Mathematical Programming*, vol. 14[1], pp. 265–294, 1978.
[33] F. Bach, "Learning with submodular functions: A convex optimization perspective," *Foundations and Trends in Machine Learning*, vol. 6[2-3], pp. 145–373, 2013.
[34] E.A. Carlen, "Trace inequalities and quantum entropy: An introductory course," 2009, `http://www.mathphys.org/AZschool/material/AZ09-carlen.pdf`.
[35] L. Lovász, "Submodular functions and convexity," in *Mathematical Programming: The State of the Art*. Springer-Verlag, 1982.
[36] R. Bhatia, *Matrix analysis*, Springer, 1997.
[37] H.V. Henderson and S.R. Searle, "On deriving the inverse of a sum of matrices," *SIAM Review*, vol. 23[1], pp. 53–60, 1981.
[38] R.A. Horn and C.R. Johnson, *Matrix analysis*, Cambridge University Press, 2013.
[39] Boying Wang and Fuzhen Zhang, "Some inequalities for the eigenvalues of the product of positive semidefinite Hermitian matrices," *Linear Algebra and its Applications*, vol. 160, pp. 113–118, 1992.
[40] R. Bhatia and F. Kittaneh, "Norm inequalities for partitioned operators and an application," *Mathematische Annalen*, vol. 287[4], pp. 719–726, 1990.
[41] M.S. Grewal and A.P. Andrews, *Kalman filtering: Theory and practice using MATLAB*, Wiley, 2014.

APPENDIX

PROOF OF PROPOSITION 1

Proof: Since $\boldsymbol{O}_\ell$ and statistics of the signals are known, estimating $\bar{\boldsymbol{z}}$ from the linear model (4) is a stochastic estimation problem. Thus, the minimum MSE incurred from estimating $\bar{\boldsymbol{z}}$ from $\bar{\boldsymbol{y}}$ is given by the trace of [30]

$$\boldsymbol{K}(\mathcal{S}) = \left[\boldsymbol{C}^{-1} + \boldsymbol{O}_\ell(\mathcal{S})^T \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T)\right]^{-1} = \boldsymbol{O}_\ell(\mathcal{S})\right]^{-1}, \quad (31)$$

where $\boldsymbol{C} = \operatorname{blkdiag}(\boldsymbol{\Pi}_0, \boldsymbol{I} \otimes \boldsymbol{R}_w)$ and $\operatorname{blkdiag}(\boldsymbol{X}, \boldsymbol{Y})$ is the block diagonal matrix whose diagonal blocks are $\boldsymbol{X}$ and $\boldsymbol{Y}$.

To obtain the form in (5), notice from (31) that $\boldsymbol{K}$ depends on $\mathcal{S}$ only through $\boldsymbol{O}_\ell(\mathcal{S})^T \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T)\right]^{-1} \boldsymbol{O}_\ell(\mathcal{S})$. Using the fact that $\operatorname{blkdiag}(\boldsymbol{A}_i)^{-1} = \operatorname{blkdiag}(\boldsymbol{A}_i^{-1})$ and the definition of $\boldsymbol{O}_\ell(\mathcal{S})$ in (4) we obtain

$$\boldsymbol{O}_\ell(\mathcal{S})^T \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T)\right]^{-1} \boldsymbol{O}_\ell(\mathcal{S}) = \boldsymbol{\Phi}_\ell^T \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{H})\right]^T \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T)^{-1}\right] \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{H})\right] \boldsymbol{\Phi}_\ell.$$

From the mixed product property of the Kronecker product [38] we then get

$$\boldsymbol{O}_\ell(\mathcal{S})^T \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T)\right]^{-1} \boldsymbol{O}_\ell(\mathcal{S}) = \boldsymbol{\Phi}_\ell^T \left(\boldsymbol{I} \otimes \boldsymbol{H}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{R}_v^{-1}\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{H}\right) \boldsymbol{\Phi}_\ell, \quad (32)$$

where we also used the fact that $(\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T)^{-1} = \boldsymbol{S}\boldsymbol{R}_v^{-1}\boldsymbol{S}^T$ since $\boldsymbol{R}_v$ is a diagonal matrix. Letting $\boldsymbol{h}_i^T$ to be the i-th row of $\boldsymbol{H}$ in (32) and using the linearity of matrix products we obtain [38]

$$\boldsymbol{O}_\ell(\mathcal{S})^T \left[\boldsymbol{I} \otimes (\boldsymbol{S}\boldsymbol{R}_v\boldsymbol{S}^T)\right]^{-1} \boldsymbol{O}_\ell(\mathcal{S}) = \sum_{i \in \mathcal{S}} \boldsymbol{\Phi}_\ell^T \left(\boldsymbol{I} \otimes \sigma_{v,i}^{-2}\boldsymbol{h}_i\boldsymbol{h}_i^T\right) \boldsymbol{\Phi}_\ell. \quad (33)$$

Substituting (33) in (31) and taking its trace yields (5). ■

PROOF SKETCH OF LEMMA 1

Proof: Start by defining the perturbed gain as $\Delta_\epsilon = \operatorname{Tr}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\bar{\boldsymbol{M}}_i\left[\boldsymbol{Y}(\mathcal{A}) + \bar{\boldsymbol{M}}_i\right]^{-1}\right]$, for $\epsilon > 0$, where $\bar{\boldsymbol{M}}_i = \boldsymbol{M}_i + \epsilon\boldsymbol{I} \succ 0$. We omit the dependence on i for clarity. Note that, $\Delta_\epsilon \to \Delta$ as $\epsilon \to 0$. Using the invertibility of $\bar{\boldsymbol{M}}_i$, we obtain

$$\Delta_\epsilon(\mathcal{A}) = \operatorname{Tr}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\left[\boldsymbol{Y}(\mathcal{A})^{-1} + \bar{\boldsymbol{M}}_i^{-1}\right]^{-1}\boldsymbol{Y}(\mathcal{A})^{-1}\right].$$

Since $\boldsymbol{Y}(\mathcal{A})^{-1} \succ 0$, its square-root $\boldsymbol{Y}(\mathcal{A})^{-1/2}$ is well-defined and unique [38]. We can therefore use the circular commutation property of the trace to get

$$\Delta_\epsilon(\mathcal{A}) = \operatorname{Tr}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\boldsymbol{Z}\right], \quad (34)$$

with $\boldsymbol{Z} = \boldsymbol{Y}(\mathcal{A})^{-1/2}\left[\boldsymbol{Y}(\mathcal{A})^{-1} + \bar{\boldsymbol{M}}_i^{-1}\right]^{-1}\boldsymbol{Y}(\mathcal{A})^{-1/2}$. Since both matrices in (34) are positive definite, we can use the bound from [39] to obtain

$$\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\right]\operatorname{Tr}\left[\boldsymbol{Z}\right] \leq \Delta_\epsilon(\mathcal{A}) \leq \lambda_{\max}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\right]\operatorname{Tr}\left[\boldsymbol{Z}\right].$$

Reversing the manipulations used to get to (34) finally yields

$$\lambda_{\min}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\right]\operatorname{Tr}\left[\bar{\boldsymbol{M}}_i\left[\boldsymbol{Y}(\mathcal{A}) + \bar{\boldsymbol{M}}_i\right]^{-1}\right] \leq \Delta_\epsilon(\mathcal{A}) \leq \lambda_{\max}\left[\boldsymbol{Y}(\mathcal{A})^{-1}\right]\operatorname{Tr}\left[\bar{\boldsymbol{M}}_i\left[\boldsymbol{Y}(\mathcal{A}) + \bar{\boldsymbol{M}}_i\right]^{-1}\right].$$

The result in (21) is obtained by continuity as $\epsilon \to 0$. ■