Sensor management with time, energy and communication constraints

Cristian Rusu, John Thompson and Neil M. Robertson

***Abstract*—In this paper we present new algorithms and analysis for the linear inverse sensor placement and scheduling problems over multiple time instances with power and communications constraints. The proposed algorithms, which deal directly with minimizing the mean squared and worst case errors (MSE and WCE), are based on the convex relaxation approach to address the binary optimization scheduling problems that are formulated in sensor network scenarios. We propose to balance the energy and communications demands of operating a network of sensors over time while we still guarantee a minimum level of estimation accuracy. We measure this accuracy by the MSE and WCE for which we provide tight average case and lower bounds analyses. We show experimentally how the proposed algorithm perform against state-of-the-art methods previously described in the literature.**

***Index Terms*—linear inverse problem, sensor placement, sensor scheduling, binary optimization, convex relaxation, energy constraints, communications constraints.**

I. Introduction

Sensor networks are often used to measure and monitor physical phenomena like temperature, humidity and concentration of certain pollutants in an area of interest over time [1]. Modern wireless sensor networks may be composed of a large number of heterogeneous sensors each with its own (possibly limited) power supply capable of performing measurements, processing the result and communicating it to other neighboring sensors in the network at regular times. In this paper we consider the situation where, without any particular assumptions on the parameters to be estimated, the measurements taken by the sensor network are used to solve a linear inverse problem. In this setting, the problem of selecting only a subset of the available sensors while ensuring a certain level of estimation accuracy has been extensively studied in the literature.

Sensor selection (or sensor placement) [2], [3] is of central importance when considering the classical problem of parameter estimation from a given set of linear measurements that describe an operational sensor network. Given a fixed set of potential locations, the sensor placement problem asks where the sensors should be placed in order to maximize on average the estimation accuracy of the network. Once the most informative locations are identified the sensors are placed in their locations for the whole lifetime of the network. If we now consider a network where each sensor has a particular energy and communication profile and is capable of performing a measurement with a particular quality, an interesting problem that arises is how to schedule each sensor over time such that the estimation accuracy of the network is never worse than a prescribed level while we also control the energy consumption. This is done for example to make sure that no sensor goes off-line due to overuse. We call this the sensor scheduling problem. In this paper we tackle both problems and propose new heuristics to address them and provide new theoretical insights into their behavior. Because the problems are combinatorial in nature (they amount to mixed-integer optimization problems) they are NP-hard to solve exactly in general. Therefore, following previous literature, we settle on proposing sub-optimal but numerically efficient algorithms and comparing them with previously proposed methods.

The sensor placement (and in general the sensor management) problems have been extensively studied in the past. Approaches to this problem include greedy methods [4] with submodularity based performance guarantees [5], [6], convex optimization for experimental design [7, Chapter 7.5] [8], [9], [10], [11], [12], [13], information theoretic criteria [14], [15], [16] or other search heuristics [17], [18] or full search branch-and-bound methods [19]. Several recent works have also considered non-linear sensor networks [12], sparsity-aware [13], [20], [21] and distributed sensing scenarios [22], [23], correlated noise [24], estimation of continuous variables [25] or energy constraints [26], [27] [28], [29] and scheduling [30], [31], [32] over the network.

The contribution of this paper is two-fold.

The first contribution is to propose algorithms for the two problems we have described: the sensor placement and the sensor scheduling problems. Both approaches are based on the previously introduced convex optimization approach. In the case of the sensor scheduling problem we are able to accommodate energy and communications constraints in order to balance the estimation accuracy of the sensor network over multiple time instances with its energy consumption. Solving the proposed convex optimization process is numerically efficient since it can be done in polynomial time by off-the-shelf solvers [33]. These approaches is slower than some previously proposed greedy methods but they have the advantage of accommodating the energy and communication constraints of the sensor network.

The second contribution is to provide an average analysis

C. Rusu and J. Thompson are with the Institute for Digital Communications, School of Engineering, The University of Edinburgh, UK. N. M. Robertson is with Queen's University of Belfast, UK. Email: {c.rusu, john.thompson}@ed.ac.uk, n.robertson@qub.ac.uk. Demo source code available online at http://udrc.eng.ed.ac.uk/sites/udrc.eng.ed.ac.uk/files/attachments/demo_sensor_management.zip.

The authors acknowledge support from the Engineering and Physical Sciences Research Council (EPSRC) [EP/K014277/1] and the MOD University Defence Research Collaboration (UDRC) on Signal Processing.

of the sensor placement problem in the case where the overall sensor network is described by a tight frame or measurement matrix. This analysis provides insights into the empirical performance past and currently proposed sensor placement algorithms. We also provide theoretical connections to other related research fields where bounds on the eigenvalues of sub-matrices of a given matrix have been developed.

Section II describes the measurement setup we consider; Section III presents the previously proposed methods in the literature for the sensor placement problem; Section IV proposes a new algorithms for the sensor placement and scheduling problems; Section V provides a theoretical analysis of the average estimation accuracy for the sensor placement problem and Section VI shows the numerical results where we compare the proposed method with the state-of-the-art methods from the literature.

II. The problem setup

Let us assume that we want to estimate a parameter vector $\mathbf{x}_t$ of size n that changes over $t = 1, \dots, T$, time instances from $n \leq k_t \leq m$ linear measurements that are given by

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{x}_t + \mathbf{n}_t, \tag{1}$$

where the noise vector $\mathbf{n}_t$ of size k_t is a zero-mean i.i.d. Gaussian vector with variance $\sigma^2 \mathbf{I}$. The rows of the measurement matrix $\mathbf{A}_t$ at time t are chosen from the rows of an overall full rank measurement matrix $\mathbf{A}$ of size $m \times n$. The i^{th} row of $\mathbf{A}$ corresponds to the linear measurement performed by the i^{th} sensor of the network. Therefore, the matrix $\mathbf{A}$ characterizes the full sensor network which is made up of m elements and the measurement matrices $\mathbf{A}_t$ are subsets of the rows from $\mathbf{A}$. Assuming $k_t \geq n$ sensors are used at time t, the least squares estimate are given by

$$\hat{\mathbf{x}}_t = \mathbf{A}_t^{\dagger} \mathbf{y}_t = (\mathbf{A}_t^T \mathbf{A}_t)^{-1} \mathbf{A}_t^T \mathbf{y}_t. \tag{2}$$

With the understanding that with increased noise levels the estimation accuracy decreases on average, in order to simplify the exposition of the results herein we only consider the fixed noise level $\sigma^2 = 1$. We focus on a full rank measurement matrix $\mathbf{A}_t \in \mathbb{R}^{k_t \times n}$, that represents the k_t sensors that perform linear measurements at time t, for which there are several ways to quantify its recovery performance in (2):

1) A-optimality: mean squared error (MSE)

$$\text{MSE}(\mathbf{A}_t) = \text{tr}((\mathbf{A}_t^T \mathbf{A}_t)^{-1}) = \sum_{i=1}^{n} \frac{1}{\lambda_i(\mathbf{A}_t^T \mathbf{A}_t)}. \tag{3}$$

2) E-optimality: worst case error (WCE)

$$\text{WCE}(\mathbf{A}_t) = \lambda_1((\mathbf{A}_t^T \mathbf{A}_t)^{-1}) = \frac{1}{\lambda_n(\mathbf{A}_t^T \mathbf{A}_t)}. \tag{4}$$

3) D-optimality: volume of the confidence ellipsoid (VCE)

$$\text{VCE}(\mathbf{A}_t) = \text{logdet}(\mathbf{A}_t^T \mathbf{A}_t) = \log\left(\prod_{i=1}^{n} \lambda_i(\mathbf{A}_t^T \mathbf{A}_t)\right). \tag{5}$$

We have denoted here $\lambda_i(\mathbf{A}_t^T \mathbf{A}_t)$ as the i^{th} eigenvalue of the symmetric positive-semidefinite matrix $\mathbf{A}_t^T \mathbf{A}_t$ and we assume without loss of generality the ordering $\lambda_1(\mathbf{A}_t^T \mathbf{A}_t) \geq \cdots \geq \lambda_n(\mathbf{A}_t^T \mathbf{A}_t) \geq 0$.

Notice that these performance indicators are related as we have $\text{MSE}(\mathbf{A}_t) \leq n\text{WCE}(\mathbf{A}_t)$, while maximizing the $\text{VCE}(\mathbf{A}_t)$ we also maximize the denominator of the $\text{MSE}(\mathbf{A}_t)$ in (3) – but we do not also control the numerator term in (3). Among all the measurement matrices of the same size, and for the same Frobenius norm, these performance measures are minimized for $\alpha-$tight frames, i.e, measurement matrices $\mathbf{A}_t \in \mathbb{R}^{k_t \times n}$ that obey $\mathbf{A}_t^T \mathbf{A}_t = \alpha \mathbf{I}$, for which we have the exact results

$$\begin{aligned} \text{MSE}(\mathbf{A}_t) = \frac{n}{\alpha} = \frac{n^2}{\|\mathbf{A}_t\|_F^2},\ \text{WCE}(\mathbf{A}_t) = \frac{1}{\alpha} = \frac{n}{\|\mathbf{A}_t\|_F^2}, \\ \text{VCE}(\mathbf{A}_t) = \frac{1}{\alpha^n} = \left(\frac{n}{\|\mathbf{A}_t\|_F^2}\right)^n, \end{aligned} \tag{6}$$

since we have that $\|\mathbf{A}_t\|_F^2 = n\alpha$. Higher Frobenius norm (and therefore also higher α) of the measurement matrix is equivalent to increasing on average the SNR of the measurements and therefore should lead to better recovery performance in general – for example $\text{MSE}(\beta \mathbf{A}_t) = \beta^{-2}\text{MSE}(\mathbf{A}_t),\ \beta \in \mathbb{R}$. Furthermore, since tight frames minimize these criteria then for any non-tight measurement matrix these values are lower bounds.

To achieve low error indicators, in terms of the eigenvalues of $\mathbf{A}_t^T \mathbf{A}_t$ our goal is twofold:

- first, increase the smallest eigenvalue

$$\lambda_n(\mathbf{A}_t^T \mathbf{A}_t) \gg 0, \tag{7}$$

- second, group all eigenvalues such that

$$\lambda_i(\mathbf{A}_t^T \mathbf{A}_t) \approx \lambda_j(\mathbf{A}_t^T \mathbf{A}_t),\ \forall i \neq j, \tag{8}$$

i.e., the measurement matrix $\mathbf{A}_t$ behaves approximately as a tight frame with high Frobenius norm.

Given a measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ that represents a sensor network of m elements, our goal is to choose a subset of measurements $\mathbf{A}_t \in \mathbb{R}^{k_t \times n}$ from $\mathbf{A}$ such that we optimize the $\text{MSE}(\mathbf{A}_t)$, $\text{WCE}(\mathbf{A}_t)$ or the $\text{VCE}(\mathbf{A}_t)$ over all the time instances $t = 1, \dots, T$, while we also balance the energy consumption of the network.

III. The sensor management problem

We defined now the sensor management problem for a single time instance, i.e., $T = 1$. Given a network of m sensors where each is capable of a linear measurement the sensor management problems asks which (and how many) sensors need to be activated in order to guarantee a fixed, given, performance measure (for example, the estimation accuracy in terms of MSE). An equivalent formulation can be made for example by fixing the number of sensor k to activate while we minimize any of the performance measures (3), (4) or (5).

We consider that the full network of m sensors is represented by $\mathbf{A} \in \mathbb{R}^{m \times n}$, i.e., each linear measurement is represented by a row $\mathbf{a}_i^T,\ i = 1, \dots, m$. The selected sensors are denoted in the measurement matrix

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{a}_{i_1} & \mathbf{a}_{i_2} & \dots & \mathbf{a}_{i_k} \end{bmatrix}^T \in \mathbb{R}^{k \times n}, \tag{9}$$

a subset of rows of $\mathbf{A}$ indexed in the set $\mathcal{I} = \{i_1, \ldots, i_k\}$ of size $k \geq n$, such its performance in terms of MSE($\mathbf{A}_1$) or WCE($\mathbf{A}_1$) is below a given threshold γ or, alternatively, a maximum given number of sensors k is activated. Notice that in order to optimize MSE($\mathbf{A}_1$), WCE($\mathbf{A}_1$) or VCE($\mathbf{A}_1$) we need to verify spectral properties of

$$\mathbf{A}_1^T \mathbf{A}_1 = \sum_{i \in \mathcal{I}} \mathbf{a}_i \mathbf{a}_i^T = \mathbf{A}^T \text{diag}(\mathbf{z}) \mathbf{A} \in \mathbb{R}^{n \times n}, \quad (10)$$

where $\mathbf{z} \in \{0,1\}^n$ with $z_i = 1$ if $i \in \mathcal{I}$ and zero otherwise.

There are several approaches in the literature to deal with the sensor management problem. Although there are algorithms for to the sensor management problem that use search techniques [17], [18] or cross-entropy optimization [16], we mainly distinguish two major approaches based on convex optimization and greedy methods and we discuss them separately in the next two subsections.

A. Convex relaxation approach

In this formulation, the sensor selection problem is relaxed to a convex optimization [7, Chapter 7.5][8] program as

$$\begin{aligned} \underset{\mathbf{z} \in [0,1]^m}{\text{maximize/minimize}} \quad & f(\mathbf{A}, \mathbf{z}) \\ \text{subject to} \quad & \mathbf{1}^T \mathbf{z} = k. \end{aligned} \quad (11)$$

The goal is to construct a binary solution $\mathbf{z}$ that selects k sensors such that z_i indicates whether the i^{th} sensor is selected or not. The objective function can be adapted to any of the performance measures in (3) (4) (5). For example, when considering the VCE($\mathbf{A}_1$) we maximize $f(\mathbf{A}, \mathbf{z}) = \text{logdet}(\mathbf{A}^T \text{diag}(\mathbf{z}) \mathbf{A})$ since the logarithm of the determinant is concave while for the MSE($\mathbf{A}_1$) we minimize $f(\mathbf{A}, \mathbf{z}) = \text{tr}((\mathbf{A}^T \text{diag}(\mathbf{z}) \mathbf{A})^{-1})$ since the trace of the inverse is convex. In order to reach a convex optimization problem the hard binary constraint $\mathbf{z} \in \{0,1\}^m$ is relaxed to $\mathbf{z} \in [0,1]^m$. Unfortunately, the problem in (11) does not produce binary solutions $\mathbf{z}$ in general, just sub-unitary entries as the relaxation dictates. A rounding procedure, and usually also a local search, follow.

A similar approach called SparSenSe is proposed in [12] where the same core optimization problem minimizes the MSE by selecting a few sensors given a maximum accepted level of MSE (not a fixed number of activated sensors). Again, a rounding procedure and potentially a local search can follow.

B. Greedy methods approach

FrameSense [6] proposes to activate the sensors in the network according to a greedy procedure. For example, to minimize MSE in this fashion, given a measurement matrix $\mathbf{A}_1$ as in (9) a greedy scheme asks how to choose a new measurement $\mathbf{a}_j^T$ from the given full set $\mathbf{A} \in \mathbb{R}^{m \times n}$ such that

$$j = \underset{j \in \{1,\ldots,m\} \setminus \mathcal{I}}{\arg\min} \text{MSE}(\tilde{\mathbf{A}}_1), \text{ with } \tilde{\mathbf{A}}_1^T = \begin{bmatrix} \mathbf{A}_1^T & \mathbf{a}_j \end{bmatrix}, \quad (12)$$

and then adds the measurement to the active set

$$\mathcal{I} \leftarrow \mathcal{I} \cup \{j\}. \quad (13)$$

As we have showed in (6), $\alpha-$tight frames $\mathbf{A}$ provide the best performance and therefore FrameSense uses a greedy procedure to remove sensors from the network one-by-one such that the resulting sensors behave as a tight frame. To do this, FrameSense minimizes the frame potential FP($\mathbf{A}$) $= \sum_{i=1}^m \sum_{j=1}^m (\mathbf{a}_i^T \mathbf{a}_j)^2 = \|\mathbf{A}^T \mathbf{A}\|_F^2$ which has been show to achieve its minimum value exactly for $\alpha-$tight frames [34] (in this case its minimum value is $n\alpha^2$). The authors choose to minimize this quantity because they are able to prove its submodularity, i.e., FP($\mathbf{A}$) is well suited for minimization with greedy approaches [35]. This approach is interesting because in general the performance indicators (3) and (4) are not submodular functions under the activation of new sensors and therefore greedy methods do not seem particularly well suited for the sensor management problem.

Since monitoring spatial phenomena can be modeled in the context of Gaussian processes, a greedy method with submodularity properties was proposed in [36] to solve the sensor placement problem with near-optimal results. Using the same information theoretic approach the authors propose a branch-and-bound method that guarantees the construction of the optimal solution.

Another proposed greedy strategy is to add each sensor one by one according to the maximal projection onto the minimum eigenspace of a defined dual observation matrix [4]. The method is computationally efficient and very successful in selecting the most informative sensors because it takes into account all eigenvalues of the observation matrix to encourage the two desirable properties (7) and (8).

Greedy approaches to the sensor selection problem face some difficulties. For example, as also pointed out in [4], we have some hard limitations for the eigenvalues of $\mathbf{A}^T \mathbf{A}$ when adding measurements one by one.

Result 1. Given a positive semidefinite matrix $\mathbf{A}^T \mathbf{A}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ then $\mathbf{A}^T \mathbf{A} + \mathbf{a}\mathbf{a}^T$ has eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ and we have the interlacing property

$$\mu_1 \geq \lambda_1 \geq \mu_2 \geq \lambda_2 \geq \cdots \geq \mu_n \geq \lambda_n. \quad (14)$$

Proof. A proof is given in [37, Chapter 4]. ■

This is one of the reasons given in [4] that the goal to increase all the eigenvalues λ_i with each new measurement. If the eigenvalues become too concentrated, greedy algorithms may run into some difficulties.

Result 2. Assume $\mathbf{A}^T \mathbf{A}$ has n eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ in arithmetic progression, i.e, $\lambda_i = \lambda_1 + (i-1)r,\ i = 2, \ldots, n$ with $r < 0$, then the largest eigenvalue μ_1 of $\mathbf{A}^T \mathbf{A} + \mathbf{a}\mathbf{a}^T$, where $\mathbf{a} \in \mathbb{R}^n$ is a new measurement, obeys

$$\mu_1 \geq (\lambda_1 + (n-1)r)(1 + \mathbf{a}^T (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{a}). \quad (15)$$

Proof. See Appendix A. ■

This result shows that as the eigenvalues become more concentrated (a highly desired property) they might exhibit a repelling property with regard to the largest eigenvalue when new measurements are added. This is turn means that though $\mathbf{A}^T \mathbf{A}$ behaves as a tight frame we have that $\mathbf{A}^T \mathbf{A} + \mathbf{a}\mathbf{a}^T$ may no longer behave as such and therefore, for the latter matrix, the performance criteria such as MSE or WCE do not decrease significantly.

IV. The proposed optimization techniques for sensor selection

Given a sensor network, we expect the best accuracy to be achieved if all its sensors are activated. Therefore, we will express the MSE and WCE performance of $\mathbf{A}_1$ relative to the overall performance of the full network $\mathbf{A}$.

Result 3. Given the measurements $\mathbf{A} \in \mathbb{R}^{m\times n}$ and any new measurement $\mathbf{a} \in \mathbb{R}^n$ then the new measurement matrix $\tilde{\mathbf{A}}_1^T = \begin{bmatrix}\mathbf{A}_1^T & \mathbf{a}\end{bmatrix} \in \mathbb{R}^{n\times(m+1)}$, improves all performance measures, i.e., $\text{MSE}(\tilde{\mathbf{A}}_1) \leq \text{MSE}(\mathbf{A}_1)$, $\text{WCE}(\tilde{\mathbf{A}}_1) \leq \text{WCE}(\mathbf{A}_1)$, $\text{VCE}(\tilde{\mathbf{A}}_1) \geq \text{VCE}(\mathbf{A}_1)$ and equality holds only when $\mathbf{a} = \mathbf{0}_{n\times 1}$. Equivalently, the performance measures are monotonically decreasing functions with the number of measurements.

Proof. See Appendix B. ■

Therefore, we introduce the reference (lowest MSE) performance of the full network as

$$\gamma_0 = \text{tr}((\mathbf{A}^T\mathbf{A})^{-1}). \tag{16}$$

We will impose estimation accuracy levels $\gamma = \rho\gamma_0$ where $\rho \geq 1$. Naturally, with larger ρ we will select fewer sensors from the network (allowing larger mean squared error) and vice-versa. When $\rho = 1$ the only feasible solution is to select the full sensor network. We next describe the proposed optimization methods for the minimization of the MSE (the extensions to the minimization of the WCE and maximization of the VCE follow immediately and they are omitted for brevity).

Algorithm 1 – Sensor management by ℓ_1 minimization.
Input: The sensing matrix of the network with m sensors $\mathbf{A} \in \mathbb{R}^{m\times n}$, the total number of time instances T, the maximum allowed error $\rho > 1$, the regularization parameter $\lambda > 0$, the vector of sensing costs $\mathbf{s} \in \mathbb{R}_+^m$ and the matrix of communication costs $\mathbf{C} \in \mathbb{R}_+^{m\times m}$.
Output: The scheduling table $\mathbf{Z} \in \{0,1\}^{m\times T}$ for the sensor activations at each time step such that the energy constraints are satisfied.

Initialization:

1. Set initial weights $\mathbf{w}_t = \mathbf{1}$ and initial all-zero solution $\mathbf{z}_t = \mathbf{0}_{m\times 1}$ for $t = 1,\ldots,T$, i.e., $\mathbf{Z} = \mathbf{0}_{m\times T}$.

2. Initialize sets $\mathcal{N} = \emptyset$ indexing sensors that are not selected and $\mathcal{K} = \emptyset$ indexing sensors that are selected.

3. Establish the best MSE performance γ_0 by (16).

Iterations:

1. Set $\mathbf{Z}_{\text{prev}} \leftarrow \mathbf{Z}$.

2. Update weights according to $w_{ij} = (z_{ij} + \epsilon)^{-1}$.

3. Solve (17) when $T = 1$ and (19) or (20) for $T > 1$ with $\rho\gamma_0$ to obtain the current estimate $\mathbf{Z}$.

4. Update the sets $\mathcal{N} = \{n \mid \mathbf{Z}(n) \leq \epsilon\}$ and $\mathcal{K} = \{k \mid \mathbf{Z}(k) \geq 1 - \epsilon\}$.

5. If iterative process has converged, i.e., $\|\mathbf{Z} - \mathbf{Z}_{\text{prev}}\|_F^2 \leq \epsilon$, then $\mathcal{K} = \mathcal{K} \cup \{\arg\max_k \mathbf{Z}(k),\ k \notin \mathcal{K}\}$.

6. If solution is binary, i.e., $|\mathcal{N}| + |\mathcal{K}| = mT$, then stop otherwise go to step 1 of the iterative process.

A. A single time instance

Since most of the previous work (including the papers discussed in Section III) deals with a single time instance, we propose the first optimization technique in the same scenario.

We assume we are given a network of sensors and the goal is to select the most informative subset of sensors from the network (i.e., the subset of sensors that achieves some level of accuracy or mean squared error). This formulation is equivalent to asking where sensors need to be placed (from a fixed set of possible locations) such that the resulting network achieves a minimum level of prescribed accuracy.

In order to select a reduced subset of sensors from a full sensor network such that the MSE is below a prescribed value, again based on convex optimization ideas, we propose to solve a series of convex problems in the spirit of iteratively reweighted ℓ_1 (IRL1) [38] as

$$\begin{aligned} \underset{\mathbf{z}\in[0,1]^m}{\text{minimize}} \quad & \mathbf{w}^T\mathbf{z} \\ \text{subject to} \quad & \text{tr}((\mathbf{A}^T\text{diag}(\mathbf{z})\mathbf{A})^{-1}) \leq \rho\gamma_0, \end{aligned} \tag{17}$$

where $w_i = (z_i + \epsilon)^{-1},\ i = 1,\ldots,m$, composes the weights vector $\mathbf{w} \in \mathbb{R}^m$. The performance level $\rho > 1$ is given and fixed.

B. Energy constraints over multiple time instances

In the previous formulation, a particular sensor in the network is either selected or not (or equivalently, we place a sensor in a particular place or not) for the whole lifetime of the network. In some situations this scenario is realistic while in others it may not be a proper approach. Consider for example a scenario where sensors are placed in an observation field where each sensor has its own energy supply and communications capabilities. If we choose the most informative sensors and disregard their other constraints we end up with a solution that will never activate certain sensors, which are wasted in the network.

In this section we also deal with a scenario where our goal is to schedule how sensor from a network are selected over multiple time instances such that at each time instant we guarantee a certain level of accuracy (e.g., the MSE is below a threshold) while we also balance the energy and communications constraints of the sensors.

We consider that we deal with m sensors to be scheduled over T time instances and therefore we introduce the binary scheduling table

$$\mathbf{Z} = \begin{bmatrix}\mathbf{z}_1 & \mathbf{z}_2 & \ldots & \mathbf{z}_T\end{bmatrix} \in \{0,1\}^{m\times T}, \tag{18}$$

and we denote the scheduler at time t by $\mathbf{z}_t \in \{0,1\}^m$, i.e., the columns of $\mathbf{Z}$, we denote z_{ij} the $(i,j)^{\text{th}}$ entry of $\mathbf{Z}$ and we denote z_i the entries of $\mathbf{z}_t$. We next propose two ways to construct this scheduling table.

Implicit energy constraints can be used to ensure that over the T time instances we do not selected the same sensors each time. Therefore, we propose the following regularized convex

optimization problem

$$\begin{aligned} \underset{\mathbf{Z}\in[0,1]^{m\times T}}{\text{minimize}} \quad & \sum_{t=1}^{T} \mathbf{w}_t^T \mathbf{z}_t + \lambda \max\left(\mathbf{W}\sum_{t=1}^{T}\mathbf{z}_t\right) \\ \text{subject to} \quad & \text{tr}((\mathbf{A}^T \text{diag}(\mathbf{z}_t)\mathbf{A})^{-1}) \leq \rho\gamma_0 \\ & \sum_{t=1}^{T}\mathbf{z}_t \geq \mathbf{1}, \end{aligned} \tag{19}$$

where $\mathbf{W} \in \mathbb{R}^{m\times m}$ is a diagonal matrix whose entries are positive and describe the cost of using any one sensor relative to the others. We also add an explicit constraint to ensure that each sensor is selected at least once over all T time instances. The regularization then penalizes the repeated use of the same sensor. For example, if sensor i has no energy constraints then we set $W_{ii} = 0$ while if $W_{ii} = 1$ and $W_{jj} = 2$ is interpreted as the fact that the j^{th} sensor has half of the energy supply of the i^{th} sensor.

An explicit energy constraint can be used to formulate the same problem when energy profiles of the sensors are available. Consider for example the following regularized convex optimization problem

$$\begin{aligned} \underset{\mathbf{e}\geq\mathbf{0},\ \mathbf{Z}\in[0,1]^{m\times T}}{\text{minimize}} \quad & \sum_{t=1}^{T} \mathbf{w}_t^T \mathbf{z}_t + \lambda g(\mathbf{e}) \\ \text{subject to} \quad & \text{tr}((\mathbf{A}^T \text{diag}(\mathbf{z}_t)\mathbf{A})^{-1}) \leq \rho\gamma_0 \\ & \sum_{t=1}^{T}\mathbf{z}_t \geq \mathbf{1},\ \text{diag}(\mathbf{s})\mathbf{Z}\mathbf{1} \leq \mathbf{e}_0 + \mathbf{e}, \end{aligned} \tag{20}$$

where $\mathbf{e}_0 \in \mathbb{R}^m$ with $(\mathbf{e}_0)_i$ denoting the reference amount of energy available to the i^{th} sensor and $\mathbf{s} \in \mathbb{R}^m_+$ denoting the cost of using the i^{th} sensor once (the cost of sensing and processing). The power profiles $\mathbf{c}$ and the reference upper power values $\mathbf{e}_0$ are supplied by the user. Deviations from $\mathbf{e}_0$ are penalized in the proposed optimization problem. The regularization function $g(\mathbf{e})$ can be chosen to be either an ℓ_2 penalty $g(\mathbf{e}) = \|\mathbf{e}\|_2$ or an ℓ_∞ penalty $g(\mathbf{e}) = \|\mathbf{e}\|_\infty$.

Finally, another energy constraint may deal with the cost of each sensor to transmit its measurement to a centralized node where the data from the full network is processed. Throughout this paper we have assumed this centralized model. In a simple scenario, each sensor is able to communicate directly to the central processing node. In this case, the cost of communication can be integrated into the cost of sensing denoted by $\mathbf{s}$ in (20). Otherwise, depending on the network topology, the energy constraint in (20) can be modified to

$$(\text{diag}(\mathbf{s}) + \mathbf{C})\mathbf{Z}\mathbf{1} \leq \mathbf{e}_0 + \mathbf{e}. \tag{21}$$

We have denoted by $\mathbf{C} \in \mathbb{R}^{m\times m}_+$ the communication cost of all the sensors. The entry $C_{ij} \geq 0$ expresses the cost incurred by the i^{th} sensor in order to convey data from the j^{th} sensor to the central node. When the i^{th} sensor has a direct link to the central processing node then on the i^{th} row of $\mathbf{C}$ has only one non-zero entry, namely C_{ii}.

To clearly illustrate this constraint, an example network topology of nine sensors is given in Fig. 1 and its associated

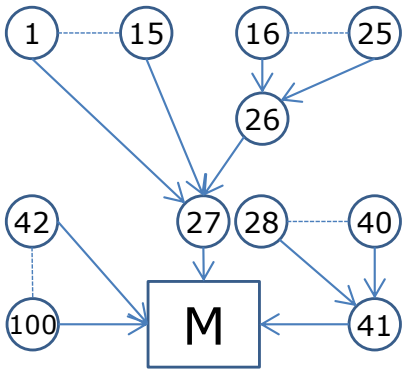

Fig. 1. An example of network topology consisting of $m = 100$ sensors and a master node M. In the scenarios we consider each sensor has a dual role: it can perform a linear measurement and/or forward data from other sensors. To avoid routing issues, we consider a single (possible multi-hop) predefined path from each sensor to the master.

cost matrix is $\mathbf{C} \in \mathbb{R}^{100\times 100}$ whose entries are for sensors 1 to 15:

$$\mathbf{C}^T = \begin{bmatrix} & \cdots & \\ \text{diag}(\mathbf{c}_{1:15}) & \mathbf{0}_{15\times 11} & c_{27}\mathbf{1}_{15\times 1} & \mathbf{0}_{15\times 73} \\ & \cdots & \end{bmatrix}, \tag{22}$$

while for sensors 16 to 25 we have:

$$\mathbf{C}^T = \begin{bmatrix} & \cdots & & \\ \mathbf{0}_{10\times 15} & \text{diag}(\mathbf{c}_{16:25}) & c_{26}\mathbf{1}_{10\times 1} & c_{27}\mathbf{1}_{10\times 1} & \mathbf{0}_{10\times 73} \\ & \cdots & & \end{bmatrix}, \tag{23}$$

and so on. The non-zero values C_{ij} are interpreted as the cost of transmitting data from the i^{th} to the j^{th} sensor. The i^{th} line of $\mathbf{C}$ denotes the total cost of the sensor network to transmit data from the i^{th} sensor to the master node.

C. The proposed optimization procedure

Based on the convex optimization problems described we now propose a method for the sensor selection problem. The full procedure is depicted in Algorithm 1. The method keeps track of two sets $\mathcal{N}$ and $\mathcal{K}$ that contain indices of the solution $\mathbf{Z}$ that are set to zero and one, respectively. The procedure terminates only when each entry of $\mathbf{Z}$ has been allocated to either $\mathcal{N}$ or $\mathcal{K}$ (see step 4 of Algorithm 1). Usually, convex optimization approaches to the sensor management problem relax the hard binary constraint and therefore the solutions are not binary in general. As such, a rounding procedure usually follows. In our case, the proposed method deals with the rounding problem internally in step 5 of Algorithm 1: if the algorithm has converged and the solution $\mathbf{Z}$ is not binary then take the largest entry from $\mathbf{Z}$ that is not in $\mathcal{K}$ and round it to one (and add it to $\mathcal{K}$).

V. Analysis of Sensor Selection

The proposed optimization problems are semidefinite programs (SDPs) with binary constraints used in an iterative fashion and therefore their analysis in terms of the optimality of the solution is difficult. In this section we explore several alternative ways to analyze (on average and worst case) the performance of sensor management solutions.

A. Results for a tight sensor network

For the purpose of understanding the behavior of the sensor selection problem, in this section we focus only on sensor networks that are characterized by tight measurement matrices $\mathbf{A}$. This choice is not restrictive in general, in fact, it is optimal if we consider also the possibility of using the full sensor network, as discussed in Section II.

Result 4. Assume we are given a sensor network represented by measurements in the $\alpha-$tight frame $\mathbf{A} \in \mathbb{R}^{m\times n}$. Selecting a subset of $n \leq k \leq m$ sensor measurements from $\mathbf{A}$ which we denote $\mathbf{A}_1 = \{\mathbf{a}_i^T\}_{i\in\mathcal{K}} \in \mathbb{R}^{k\times n}$ with $\mathbf{A}_1^T\mathbf{A}_1 = \sum_{i\in\mathcal{K}} \mathbf{a}_i\mathbf{a}_i^T$ where $\mathcal{K} = \{i_1, \ldots, i_k\}, |\mathcal{K}| = k$ then we have

$$\mathbb{E}[\text{MSE}(\mathbf{A}_1)]=\frac{mn}{(k-n+1)\alpha}, \mathbb{E}[\text{WCE}(\mathbf{A}_1)]\geq\frac{m}{(k-n+1)\alpha},$$
$$\mathbb{E}[\text{VCE}(\mathbf{A}_1)] = \log\left(n!\left(\frac{\alpha}{m}\right)^n\binom{k}{k-n}\right). \tag{24}$$

Proof. See Appendix C. ■

For example, when $k = m$ and $\alpha = m$ we know by (6) that $\text{MSE}(\mathbf{A}_1) = \text{MSE}(\mathbf{A}) = nm^{-1}$ but we have the different estimate $\mathbb{E}[\text{MSE}(\mathbf{A}_1)] = n(m-n+1)^{-1} > \text{MSE}(\mathbf{A})$. Still, the gap between the two decreases as m increases in a regime where $m \gg n$. For some choice of n, m and α, this effect can be seen in Fig. 2. It is clear from the figure that the largest differences are for low k (on the same order with n) while the gap closes for k approaching m. As we will also see from the results section, the largest differences between the performance of the methods we analyze are for low values of k. Indeed, even past research [4] has shown by numerical experimentation that most of the sensor selection methods proposed in the literature perform similarly in the regime $k \gg n$.

Also, Result 4 shows that the $\text{MSE}(\mathbf{A}_1)$ and $\text{WCE}(\mathbf{A}_1)$ decrease on average linearly with the number of selected sensors. Dependencies with the other dimensions are also linear and intuitive: increasing the number of parameters to estimate (n) and the total number of available sensors (m) leads to worse performance; increasing the energy, essentially the signal to noise ratio, of the measurement matrix (α) improves performance. Furthermore, the empirical standard deviation around the mean MSE decreases with k showing that the largest potential gains in MSE can be achieved only in the regime where $k \approx n$.

B. Relating sensor management to other problems

In this subsection we connect the sensor selection problem to other fields of research. Subset selection problems have been seen in many areas of research. For example, the problem of selecting a subset of column from a matrix such that some spectral guarantees are obeyed is well studied. Topics such as column subset selection [39] and the restricted invertibility problem [40] deal with constructing a matrix by selecting a subset of columns from a given matrix such that the new construction has the lowest singular value well above zero (i.e., the new matrix is well conditioned).

We next detail some research topics closely related to the sensor selection problem and discuss how these results apply in our case and to previous empirical observations from the literature.

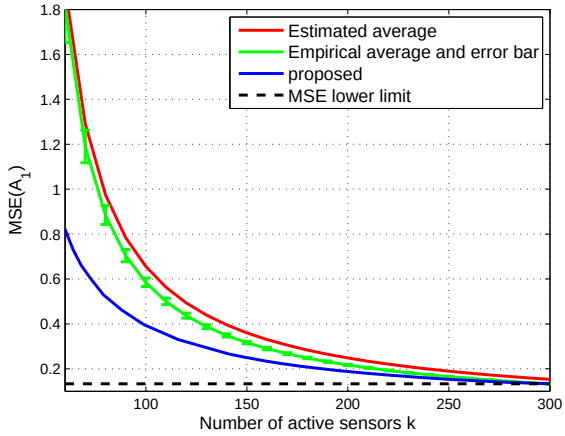

Fig. 2. Expected versus empirical values of $\text{MSE}(\mathbf{A}_1)$ constructed by selecting k sensors from a total of m belonging to random tight frame $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\alpha = m = 300$ and $n = 40$. Empirical results are averaged over 10^5 random sensor selections from $\mathbf{A}$. We show the lower limit of $\text{MSE}(\mathbf{A}_1)$ which is $\text{MSE}(\mathbf{A}) = nm^{-1}$ by (6) and is achieved by $\mathbf{A}_1$ when $k = m$.

Techniques that minimize the condition number [41] [42] can also be deployed for the sensor management problem. Given constants $s_0, t_0 \in \mathbb{R}_+$, consider the following semidefinite program:

$$\begin{aligned} \underset{t,s,\ \mathbf{z}\in[0,1]^m}{\text{minimize}} \quad & t - s + \lambda\mathbf{w}^T\mathbf{z} \\ \text{subject to} \quad & s\mathbf{I} \preceq \mathbf{A}^T\text{diag}(\mathbf{z})\mathbf{A} \preceq t\mathbf{I} \\ & s_0 \leq s \leq t \leq t_0. \end{aligned} \tag{25}$$

Sensor selection solutions provided by this optimization problem are well suited for our purposes since the constraints lead to the design of well invertible $\mathbf{A}^T\text{diag}(\mathbf{z})\mathbf{A}$ due to the threshold provided by the $s_0 \gg 0$ and a tight structure by the variables s and t. Therefore, the two desired properties are enforced (7) (8). Solving the problem in (25) (assuming also some rounding procedure to construct a binary solution $\mathbf{z}$) guarantees the following error bounds

$$\frac{n}{t_0} \leq \text{MSE}(\mathbf{A}_1) \leq \frac{n}{s_0}, \ \frac{1}{t_0} \leq \text{WCE}(\mathbf{A}_1) \leq \frac{1}{s_0}. \tag{26}$$

The upper bounds hold even after a rounding procedure is applied on the solution $\mathbf{z}$ of (25).

Compressed sensing [43] makes use of measurement matrices $\mathbf{A} \in \mathbb{R}^{m\times n}, m > n$, that obey the restricted isometry property

$$(1-\delta_k)\|\mathbf{z}\|_2^2 \leq \|\mathbf{A}^T\mathbf{z}\|_2^2 \leq (1+\delta_k)\|\mathbf{z}\|_2^2, \tag{27}$$

for a constant δ_k and any $k-$sparse vector $\mathbf{z} \in \mathbb{R}^m$. Let us denote by $\mathbf{A}_1 \in \mathbb{R}^{k\times n}$ any subset of $k \geq n$ rows from $\mathbf{A}$. Then, equivalently to (27) we have

$$1-\delta_k \leq \lambda_n(\mathbf{A}_1\mathbf{A}_1^T) \leq \lambda_1(\mathbf{A}_1\mathbf{A}_1^T) \leq 1+\delta_k. \tag{28}$$

In the case of random matrices obtained from a Gaussian distribution $\mathcal{N}(0, 1/n)$ we have that $\delta_k \approx 2\sqrt{kn^{-1}} + kn^{-1}$. Unfortunately, these bounds (28) on the eigenvalues hold when

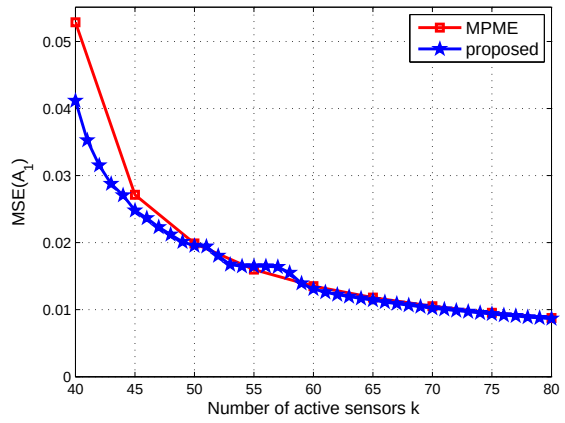

Fig. 3. Average mean squared error reached by MPME [4] and the proposed method for the estimation of a variable of size $n = 40$ with a sensor network of maximum $m = 100$ elements. The measurement matrix is random Gaussian and the results shown are averaged over 100 realizations.

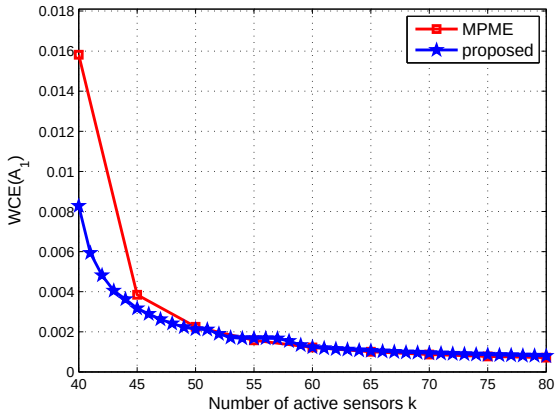

Fig. 4. Average worst case error reached by MPME [4] and the proposed method for the same experimental setup as Fig. 3.

$k/n \to 0$ while for the sensor management problem we are interested in the regime $k \geq n$. When $k \geq n$ we still have

$$\lambda_1(\mathbf{A}_1\mathbf{A}_1^T) \leq (1+\delta_k). \tag{29}$$

Therefore, it follows that

$$\begin{aligned} \text{MSE}(\mathbf{A}_1) &= \sum_{i=1}^{n} \frac{1}{\lambda_i(\mathbf{A}_1\mathbf{A}_1^T)} \geq \frac{n}{\lambda_1(\mathbf{A}_1\mathbf{A}_1^T)} \\ &\geq \frac{n}{1+2\sqrt{k/n}+k/n} = \frac{n^2}{(\sqrt{n}+\sqrt{k})^2} \approx \frac{n^2}{k}. \end{aligned} \tag{30}$$

This is better (lower) than the simple bound in (6) since $\mathbb{E}[\|\mathbf{A}_1\|_F^2] = k/n$ and therefore $\mathbb{E}[\text{MSE}(\mathbf{A}_1)] \geq n^3/k$ by (6). Alternatively, we can use the Marchenko-Pastur law to show that, asymptotically, for Gaussian random matrices whose entries are sampled from $\mathcal{N}(0,1)$ and with large k and n the MSE can be bounded. Results from non-asymptotic random matrix theory [44, Chapter 5] can also be used here to understand behavior of the extremal eigenvalues and bound the MSE with high probability.

The solution to the Kadison-Singer problem [45] shows that given a tight $\mathbf{A} \in \mathbb{R}^{m\times n}$, i.e., $\alpha = 1$, where $\delta = \max_{i=1,\ldots,m} \|\mathbf{a}_i^T\|_2^2$ is the maximum squared ℓ_2 norm of the rows of $\mathbf{A}$ there exists a partition of the rows into T sets $\{S_1,\ldots,S_T\}$ such that

$$\sigma_1(\mathbf{A}_{S_t}) = \sqrt{\lambda_1(\mathbf{A}_{S_t}^T\mathbf{A}_{S_t})} \leq \left(\frac{1}{\sqrt{T}}+\sqrt{\delta}\right)^2, \; t = 1,\ldots,T. \tag{31}$$

Therefore with $T = m/k$ for a tight $\mathbf{A}$ and denoting $\mathbf{A}_t = \mathbf{A}_{S_t} \in \mathbb{R}^{|S_t|\times n}$ the matrix composed of the rows from $\mathbf{A}$ indexed in the set S_t, with $|S_t| = k$, we have

$$\lambda_1(\mathbf{A}_t^T\mathbf{A}_t) \leq \left(\sqrt{\frac{k}{m}}+\sqrt{\delta}\right)^4, \; t = 1,\ldots,T. \tag{32}$$

The value δ for any measurement matrix $\mathbf{A}_t \in \mathbb{R}^{k\times n}$ can be estimated by using the Markov inequality and a union bound to show that

$$\begin{aligned} &\mathbb{P}\left(\max_{i=1,\ldots,k} \|\mathbf{a}_i^T\|_2^2 \geq c\mathbb{E}[\|\mathbf{a}_i^T\|_2^2]\right) \leq \\ &\qquad \sum_{i=1}^{k} \mathbb{P}(\|\mathbf{a}_i^T\|_2^2 \geq c\mathbb{E}[\|\mathbf{a}_i^T\|_2^2]) \leq \frac{k}{c}, \end{aligned} \tag{33}$$

where $\mathbb{E}[\|\mathbf{a}_i^T\|_2^2] = nm^{-1}$ and $c = c_1 k$, $c_1 \geq 1$. Therefore, with high probability $1 - c_1^{-1}$ we have that $\delta < c_1 knm^{-1}$. Still, empirically we observe that δ is within a constant factor of the expected value nm^{-1}, i.e., that $c = O(1)$, not $c = O(k)$.

Therefore, all measurement matrices $\mathbf{A}_t$ at each time instance $t = 1,\ldots,T$, obey

$$\begin{aligned} \text{MSE}(\mathbf{A}_t) &= \sum_{i=1}^{n} \frac{1}{\lambda_i(\mathbf{A}_t\mathbf{A}_t^T)} \geq \frac{n}{\lambda_1(\mathbf{A}_t\mathbf{A}_t^T)} \\ &\geq \frac{nm^2}{k^2}\left(1+\sqrt{\frac{\delta m}{k}}\right)^{-4} \\ &= \frac{nm^2}{(\sqrt{k}+\sqrt{c_1 n})^4} \approx \begin{cases} n\left(\frac{m}{k}\right)^2, & \text{if } k \gg n \\ \frac{1}{(1+\sqrt{c_1})^4}\frac{m^2}{k}, & \text{if } k \approx n. \end{cases} \end{aligned} \tag{34}$$

This result shows that potentially the MSE can exhibit a quadratic decrease with the number of selected sensors k. This is to be compared with (24) that shows a linear decrease of the expected MSE with k. When $k = m$ the bound in (34) matches the optimal value of MSE in (6) for $\alpha = 1$. These bounds are also reflected in the results from Fig. 2 where we can observe that for $k \approx n$ the decrease in MSE achieved by the proposed method, with the increased number of selected sensors k, is larger that of a random sensor selection algorithm when $k \gg n$. These insights confirm previous experimental results from the literature, like [4], where the methods proposed for sensors selection differ mostly when $k \approx n$ and are similar when $k \gg n$ (or $k \approx m$, equivalently) where even a random selection provides good estimation accuracy (seen as low MSE or WCE).

It is important to mention that constructing a set S_t such that (31) is always obeyed is still an open problem. Heuristics can be proposed similar to the approach presented in this paper

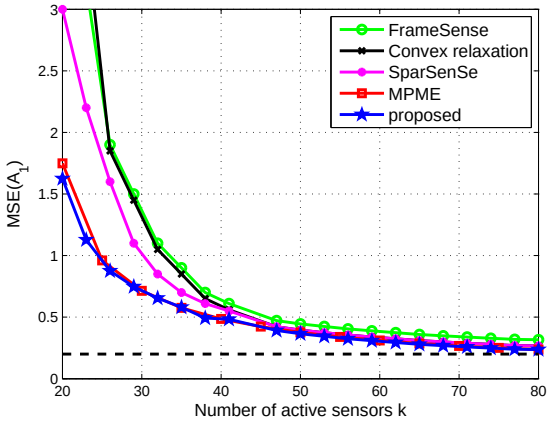

Fig. 5. Comparison of average mean squared error for the estimation of a variable of size $n = 20$ with a sensor network of $m = 100$ sensors. The measurement matrix is a random $\alpha-$tight with $\alpha = 100$ and the results are averaged over 100 realizations. We explicitly show the MSE lower bound value $n\alpha^{-1} = 0.2$.

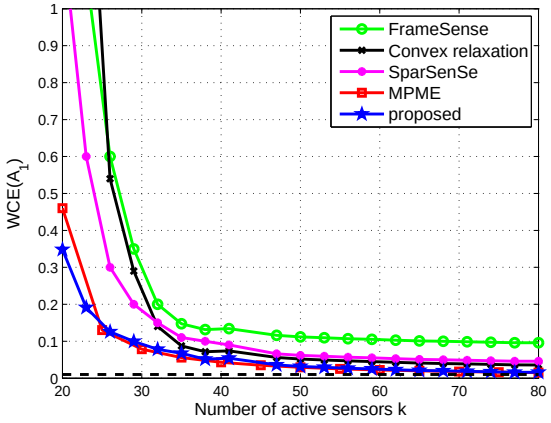

Fig. 6. Comparison of average worst case error for the same experimental setup as Fig. 5. We explicitly show the WCE lower bound value $\alpha^{-1} = 0.01$.

for binary optimization (especially because the bound in (32) is a convex constraint).

The solution to the Kadison-Singer problem is also useful in our discussion of operating a sensor network over multiple time instances. Intuitively, result (31) essentially states that a tight measurement matrix $\mathbf{A}$ can be partitioned such that the T partitions $\mathbf{A}_t$ themselves are also approximately tight, i.e., all the $\mathbf{A}_t,\ t = 1, \ldots, T$, obey (34). For example, this links with our objective in (19) of avoiding selecting the same sensors by partitioning the sensor set into disjoint subsets (quite a severe constraint) while ensuring that each subset still behaves well, i.e., similar estimation accuracy between the subsets according to MSE.

VI. Results

In this section we provide experimental numerical simulations to show the performance of the proposed methods and how they compare with state of the art approaches from the literature. We also present extensive numerical simulations to describe the performance of the proposed method to schedule a sensor network over time while also balancing power consumption.

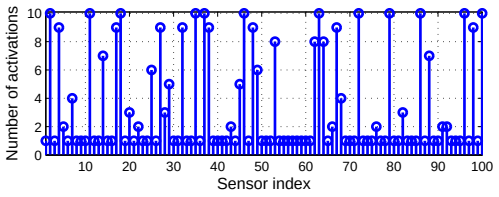

Fig. 7. Total number of activation of each sensor from the $m = 100$ elements of a sensing network described by a $\alpha-$tight measurement matrix of size 100×20 with $\alpha = 100$. The implicit energy constraint (19) runs with the regularization parameter $\lambda = 1$ and the optimization takes place of $T = 10$ time instances and the estimation accuracy is fixed to $\rho = 3$. Overall there and 342 sensors activations in the network.

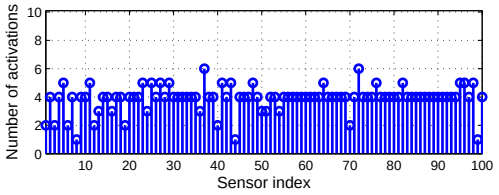

Fig. 8. Total number of activation of each sensor for the same experimental setup as in Fig. 7 with the regularization parameter $\lambda = 100$. Overall there and 389 sensors activations in the network.

A. Choosing how many sensors to activate

In the first experimental setting we provide numerical evidence on how the MSE and WCE evolve with the number of selected sensors given a fixed network. In this subsection we consider a sensor network whose measurements $\mathbf{A} \in \mathbb{R}^{m \times n}$ are described by a matrix with entries scaled i.i.d. random Gaussian from a zero mean distribution with variance one, i.e, $A_{ij} \sim \sqrt{m}\mathcal{N}(0,1),\ i = 1, \ldots, m,\ j = 1, \ldots, n$.

Results for a network of $m = 100$ sensors tasked to estimate an unknown of size $n = 40$ are shown in Fig. 3 and 4. We compare our proposed method with the state of the art approach MPME [4]. As previously noted by empirical simulations [4] and by the discussion in Section V, the performance of the two method in terms of MSE and WCE is similar when $k \gg n$ while their is a small gap in performance favoring the proposed method when $k \approx n$. To show the performance of the proposed method we evaluate it for $\rho \in (1, 10]$ on a fine grid. Fig. 3 and 4 provide an empirical practical way of choosing the number of sensor to activate while also balancing the MSE and WCE levels. Up to $k \approx 60$ sensors, the MSE and WCE shows significant decrease while after this level each new sensor activation has important diminishing returns. Also, to approach the performance of the full network a large number (close to m) of sensors need to be activated.

B. Comparisons with previous sensor selection algorithms

Following the experimental setup from [4], in this section we compare the proposed method with previously proposed methods from the literature. We choose to simulate a sensor network with $m = 100$ elements tasked to recover an unknown

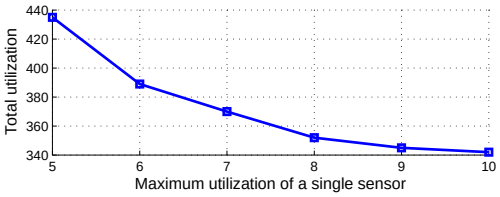

Fig. 9. Maximum sensor utilization $\max\left(\sum_{t=1}^{T}\mathbf{z}_t\right)$ versus total sensor utilization $\sum_{i=1}^{m}\sum_{j=1}^{T} z_{ij}$ for a sensor network of $m = 100$ elements over $T = 10$ time instances. The experimental setup is such that in all six points of the plot we have the same minimum accuracy (MSE) in an estimation task performed by the network. The point $(10, 342)$ corresponds to Fig. 7 and the point $(6, 389)$ to Fig. 8. The experimental setup is the same as in 7 but with various values for the regularization parameter λ.

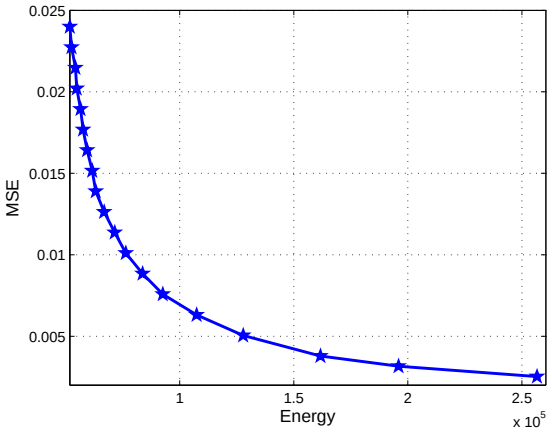

Fig. 10. Pareto curve of energy consumption versus MSE levels obtained by (20) with $\lambda = 10^3$ over $T = 10$ time instances for a fixed sensor network of $m = 100$ elements that estimate an unknown of size $n = 20$. The network topology we consider is shown in Fig. 1.

of size $n = 20$. Fig. 5 and 6 show the simulation results where we compare with FrameSense [6], convex relaxation (using the log determinant approach to minimize VCE) [8], SparSenSe [12] and MPME [4]. All measurement matrices used here are $\alpha-$tight with $\alpha = 100$. They were constructed after projecting random Gaussian matrices on the set of tight matrices (numerically this is done by taking the polar factor via a singular value decomposition).

MPME [4] and the proposed method perform best, with similar results. Just as in the previous section, the proposed method seems to outperform MPME slightly when the number of selected sensors k is close to n. The other methods perform significantly worse in this regime while the performance gaps mostly vanish when $k \gg n$. An exception to this observation is FrameSense [6], which exhibits higher MSE and WCE even when increasing k.

For similar MSE and WCE the computational complexities of the methods play an important role. Although the propose method runs in polynomial time (due to interior point solver like [33]) greedy methods are in general preferable in terms of computational complexity. As shown in [4], the MPME method is best in terms of computational complexity. Still, convex optimization approached have an edge when some extra constraints are added to the sensor selection problem (like, as we will see in the following section, energy and communications constraints). Also, all methods in this paper run without the local search mechanisms often deployed, like in [4] or [8], since these do not supply in general any great improvement.

C. Sensor selection with energy and communication constraints

To show the versatility of the convex optimization approach to the sensor selection problem we show how to deal with energy and communication constraints when scheduling the usage of a sensor network over multiple time instances.

First, we show the implicit energy constraint approach, i.e., without any explicit information about the energy profiles of the sensors our goal is to operate the sensor network over T time instances such that we do not activate the same sensors at each time. The ℓ_1/ℓ_∞ style optimization problem (19) balances between the estimation accuracy of the sensor network and making sure that the sensing is distributed more evenly between the network's sensors. Results are shown in Fig. 7 and 8. With the larger regularization parameter $\lambda = 100$ the results in Fig. 8 show o more balanced activation of the sensors, as opposed to the results in Fig. 7 that are obtained for a smaller $\lambda = 1$. With higher parameters λ the sensor scheduling problem is regularized to select less often the same sensors (for example in Fig. 8 most sensors are selected four, five or maximum six times as compared to Fig. 7 where several sensors are selected in all ten time instances), but at the cost of activating, overall, a larger number of sensors over the ten time instances.

The almost flat envelope of Fig. 8 is typical of solutions to convex optimization problems that involve ℓ_∞ regularized objective functions (for details see [7, Chapter 6]). The almost uniform activation of the sensors over time distributes the sensing workload of the network ensuring balanced power consumption together with increased robustness and fault tolerance in case of any particular sensor failure. Fig. 9 shows one of the side effects of the proposed optimization procedure: we can reduce the frequency with which one particular sensor is activated but at the cost of activating other (possibly many more) sensors from the network such that it operates with the same estimation accuracy.

Finally, we also show how the proposed algorithm can be applied to schedule a sensor network when absolute energy and communications costs are available. In terms of the sensing cost, we consider it proportional to the SNR like

$$s_i = O(\|\mathbf{a}_i^T\|_2^2),\ i = 1, \ldots, m, \tag{35}$$

meaning that the sensing cost is proportional to the quality of the measurement. We take the cost of the communications for the i^{th} sensor to be a fraction of the sensing cost

$$c_{ii} = O(s_i),\ i = 1, \ldots, m. \tag{36}$$

We take the sensor network topology in Fig. 1 to which we attach a Gaussian random measurement matrix. We will consider that no reference energy levels are available, i.e., $\mathbf{e}_0 = \mathbf{0}_{m\times 1}$, and that the energy penalty is reflected by the

cost function $g(\mathbf{e}) = \|\mathbf{e}\|_2^2$ in (20). In Fig. 10 we show the trade-off between the achievable MSE levels versus the energy consumption of the sensor network. To achieve the lowest levels of MSE we of course need (almost) all the sensors to be active (almost) all the time and therefore the energy consumption of the network is highest. Giving up some accuracy in the MSE has a positive impact on the energy consumption, especially at the limit of the best accuracy. Depending on the available energy supplies, Fig. 10 shows what levels estimation accuracy in terms of MSE are possible with the sensor network.

Regarding the running time, although it exhibits polynomial complexity, the proposed method is slower than some of the state-of-the-art methods from the literature, especially the ones based on greedy iterations. For example, for $m = 100$ and $n = 20$, averaged over 100 realizations the MPME [4] is computationally efficient with running times well below one second while the proposed method take about one minute to complete on a modern computing Intel i7® system. Therefore, the proposed method is now well suited for highly dynamical sensor network scheduling. As previously mentioned, the advantage of convex optimization based solutions is that they allow easy generalizations (like operating the sensor network over multiple time instances, without repetition of sensor selection) and allow the addition of extra constraints (like energy and communications).

VII. Conclusions

In this paper we describe a new algorithm based on a convex optimization approach to deal with the sensor placement and scheduling problems. Our method is competitive against state-of-the-art sensor management methods while it also allows to schedule the network operations over time and with energy and communication costs and constraints. We are also able to show that when the sensor network measurements are given by a tight measurement matrix then we can expect that the mean squared error of the estimation to decrease on average linearly with the number of active sensors. We also give a lower bound showing a potential quadratic decrease in the mean square error (in the best case scenario) with the number of active sensors.

Appendix A
Proof of Result 2

We use the determinant lemma $\det(\mathbf{A}^T\mathbf{A} + \mathbf{a}\mathbf{a}^T) = \det(\mathbf{A}^T\mathbf{A})(1 + \mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a})$ to reach

$$\begin{aligned}\mu_1 =& \frac{\lambda_1(\lambda_1+r)\dots(\lambda_1+(n-1)r)}{\mu_2\dots\mu_n}(1+\mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a})\\ \geq& \frac{\lambda_1(\lambda_1+r)\dots(\lambda_1+(n-1)r)}{\lambda_1\dots(\lambda_1+(n-2)r)}(1+\mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a})\\ =&(\lambda_1+(n-1)r)(1+\mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}).\end{aligned} \tag{37}$$

The inequality holds when $\mu_i = \lambda_{i-1},\ i = 2,\dots,n$ – the maximum values for μ_i according to (14). For example, when $r = 0$ all the eigenvalues of $\mathbf{A}^T\mathbf{A}$ are the same λ_1 and we have that $\mu_i = \lambda_1,\ i = 2,\dots,n$ while $\mu_1 = (1+\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a})\lambda_1$.

Appendix B
Proof of Result 3

This qualitative result follows straight forward from the fact that the eigenvalues of $\mathbf{A}^T\mathbf{A}$ are all smaller or equal than the eigenvalues of $\tilde{\mathbf{A}}^T\tilde{\mathbf{A}} = \mathbf{A}^T\mathbf{A} + \mathbf{a}\mathbf{a}^T$, by Result 1.

A quantitative analysis can also be made for the performance indicators. For the VCE we can use the determinant inversion lemma

$$\begin{aligned}\det(\mathbf{A}^T\mathbf{A} + \mathbf{a}\mathbf{a}^T) =& \det(\mathbf{A}^T\mathbf{A})(1+\mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a})\\ \geq& \det(\mathbf{A}^T\mathbf{A}), \forall\ \mathbf{a}\in\mathbb{R}^n.\end{aligned} \tag{38}$$

For the MSE we can use the matrix inversion lemma

$$\begin{aligned}\mathrm{tr}((\mathbf{A}^T\mathbf{A} + \mathbf{a}\mathbf{a}^T)^{-1}) =& \mathrm{tr}((\mathbf{A}^T\mathbf{A})^{-1}) - \frac{\|(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}\|_2^2}{1+\mathbf{a}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{a}}\\ \leq& \mathrm{tr}((\mathbf{A}^T\mathbf{A})^{-1}), \forall\ \mathbf{a}\in\mathbb{R}^n.\end{aligned} \tag{39}$$

In the case of the WCE we can bound the least singular value. Given a matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ and a row $\mathbf{a}\in\mathbb{R}^n$ then for extended matrix $\tilde{\mathbf{A}}^T = \begin{bmatrix}\mathbf{A}^T & \mathbf{a}\end{bmatrix} \in \mathbb{R}^{n\times(m+1)}$ we have

$$\left(\lambda_n^{-1/2}(\mathbf{A}^T\mathbf{A}) - \frac{\|(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\mathbf{a}\|_2^2}{1-\mathbf{a}^T(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\mathbf{a}}\right)^{-2} \geq \lambda_n(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}). \tag{40}$$

To show this we start by defining the smallest singular value

$$\sigma_{\min}^{-1}(\mathbf{A}) = \sqrt{\lambda_n^{-1}(\mathbf{A}^T\mathbf{A})} = \|\mathbf{A}^{-1}\|_2 = \sqrt{\lambda_1((\mathbf{A}^T\mathbf{A})^{-1})}. \tag{41}$$

We use the fact that $\mathbf{A}^T\mathbf{A} = \tilde{\mathbf{A}}^T\tilde{\mathbf{A}} - \mathbf{a}\mathbf{a}^T$ and use the Sherman-Morrison-Woodbury formula $(\mathbf{A}^T\mathbf{A})^{-1} = (\tilde{\mathbf{A}}^T\tilde{\mathbf{A}} - \mathbf{a}\mathbf{a}^T)^{-1} = (\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1} + \frac{(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\mathbf{a}\mathbf{a}^T(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}}{1-\mathbf{a}^T(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\mathbf{a}}$ to reach $\sigma_{\min}^{-1}(\tilde{\mathbf{A}}) \geq \sigma_{\min}^{-1}(\mathbf{A}) - \frac{\|(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\mathbf{a}\|_2^2}{1-\mathbf{a}^T(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\mathbf{a}}$. Result (40) follows directly from this last inequality.

In the special case of an invertible matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$ and a row $\mathbf{a}\in\mathbb{R}^n$ then for extended matrix $\tilde{\mathbf{A}}^T = \begin{bmatrix}\mathbf{A}^T & \mathbf{a}\end{bmatrix} \in \mathbb{R}^{n\times(n+1)}$ we have

$$\lambda_n(\mathbf{A}^T\mathbf{A}) \leq \lambda_n(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}) \leq \left(1+\|\mathbf{a}^T\mathbf{A}^{-1}\|_2^2\right)\lambda_n(\mathbf{A}^T\mathbf{A}). \tag{42}$$

To show this we start by developing

$$\tilde{\mathbf{A}} = \begin{bmatrix}\mathbf{A}\\ \mathbf{a}^T\end{bmatrix} = \begin{bmatrix}\mathbf{A}\\ \mathbf{a}^T\mathbf{A}^{-1}\mathbf{A}\end{bmatrix} = \begin{bmatrix}\mathbf{I}\\ \mathbf{a}^T\mathbf{A}^{-1}\end{bmatrix}\mathbf{A}. \tag{43}$$

We know that $\sigma_{\min}(\mathbf{X}\mathbf{Y}) \leq \|\mathbf{X}\|_2\sigma_{\min}(\mathbf{Y})$ by the Courant-Fischer-Weyl min-max principle. Coupled with the fact that the matrix $\begin{bmatrix}\mathbf{I} & (\mathbf{a}^T\mathbf{A}^{-1})^T\end{bmatrix}^T$ has all singular values 1 expect for the largest which is $\sqrt{1+\|\mathbf{a}^T\mathbf{A}^{-1}\|_2^2}$ we reach the result. The final inequality follows from

$$1 + \|\mathbf{a}^T\mathbf{A}^{-1}\|_2^2 \leq 1 + \sigma_{\min}^{-2}(\mathbf{A})\|\mathbf{a}\|_2^2. \tag{44}$$

The equality holds when we choose $\mathbf{a}$ to be any multiple of the right singular vector associated with $\sigma_{\min}(\mathbf{A})$. Overall the inequalities become $\sigma_{\min}(\mathbf{A}) \leq \sigma_{\min}(\tilde{\mathbf{A}}) \leq \left(\sqrt{1+\|\mathbf{a}^T\mathbf{A}^{-1}\|_2^2}\right)\sigma_{\min}(\mathbf{A}) \leq \sqrt{\sigma_{\min}^2(\mathbf{A}) + \|\mathbf{a}\|_2^2}$. Together with Result 1 we ultimately have that

$$\sigma_{\min}(\tilde{\mathbf{A}}) \leq \min\left\{\sqrt{\sigma_{\min}^2(\mathbf{A}) + \|\mathbf{a}\|_2^2},\ \sigma_{n-1}\right\}. \tag{45}$$

APPENDIX C
PROOF OF RESULT 4

We will use results presented in [46] to prove theorems about the expected characteristic polynomials of matrices that are changed by rank 1 updates. Consider first that for the $\alpha-$tight frame $\mathbf{A}^T = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_m\end{bmatrix} \in \mathbb{R}^{n\times m}$ and given any vector $\mathbf{u}$ such that $\|\mathbf{u}\|_2 = 1$ we have that

$$\begin{aligned}\mathbb{E}[(\mathbf{u}^T\mathbf{a}_{\text{avg}})^2] =& \frac{1}{m}\sum_{i=1}^{m}(\mathbf{u}^T\mathbf{a}_i)^2 \\ =& \frac{1}{m}\mathbf{u}^T\left(\sum_{i=1}^{m}\mathbf{a}_i\mathbf{a}_i^T\right)\mathbf{u} = \frac{\alpha\|\mathbf{u}\|_2^2}{m} = \frac{\alpha}{m},\end{aligned} \tag{46}$$

where we have defined the average frame vector $\mathbf{a}_{\text{avg}} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\mathbf{a}_i$.

If we denote by $p_{\mathbf{A}_1^T\mathbf{A}_1}(x)$ the characteristic polynomial of $\mathbf{A}_1^T\mathbf{A}_1$ then the characteristic polynomial of $\mathbf{A}_1^T\mathbf{A}_1 + \mathbf{a}\mathbf{a}^T$ is

$$\begin{aligned}p_{\mathbf{A}_1^T\mathbf{A}_1+\mathbf{a}\mathbf{a}^T}(x) &= \det(x\mathbf{I} - (\mathbf{A}_1^T\mathbf{A}_1 + \mathbf{a}\mathbf{a}^T)) \\ &= \det((x\mathbf{I} - \mathbf{A}_1^T\mathbf{A}_1) - \mathbf{a}\mathbf{a}^T) \\ &= \det(x\mathbf{I} - \mathbf{A}_1^T\mathbf{A}_1)(1 - \mathbf{a}^T(x\mathbf{I} - \mathbf{A}_1^T\mathbf{A}_1)^{-1}\mathbf{a}) \\ &= p_{\mathbf{A}_1^T\mathbf{A}_1}(x)\left(1 - \sum_{i=1}^{n}\frac{(\mathbf{u}_i^T\mathbf{a})^2}{x - \lambda_i(\mathbf{A}_1^T\mathbf{A}_1)}\right),\end{aligned} \tag{47}$$

where $\lambda_i(\mathbf{A}_1^T\mathbf{A}_1)$ are the eigenvalues of $\mathbf{A}_1^T\mathbf{A}_1$ corresponding to eigenvectors $\mathbf{u}_j$. If we choose any $\mathbf{a} \in \{\mathbf{a}_i\}_{i=1,i\notin\mathcal{K}}^m$ and use (46) to can show that

$$\begin{aligned}\mathbb{E}[p_{\mathbf{A}_1^T\mathbf{A}_1+\mathbf{a}\mathbf{a}^T}(x)] =& \mathbb{E}[p_{\mathbf{A}_1^T\mathbf{A}_1}(x)]\left(1 - \sum_{i=1}^{n}\frac{\alpha m^{-1}}{x - \lambda_i(\mathbf{A}_1^T\mathbf{A}_1)}\right) \\ =& \mathbb{E}[p_{\mathbf{A}_1^T\mathbf{A}_1}(x)] - \frac{\alpha}{m}\mathbb{E}[p_{\mathbf{A}_1^T\mathbf{A}_1}(x)]'.\end{aligned} \tag{48}$$

Starting from an empty (all zeros) matrix $\mathbf{A}_1^T\mathbf{A}_1$, i.e., $\mathcal{K} = \emptyset$, that has the characteristic polynomial $p^{(0)}_{\mathbf{A}_1^T\mathbf{A}_1}(x) = p_{\mathbf{0}_{n\times n}}(x) = x^n$ after adding k rank 1 updates of the type $\mathbf{a}\mathbf{a}^T$ leads to the matrix $\mathbf{A}_1^T\mathbf{A}_1$ with the expected characteristic polynomial is

$$\begin{aligned}\mathbb{E}[p^{(k)}_{\mathbf{A}_1^T\mathbf{A}_1}(x)] =& \mathbb{E}[p^{(k-1)}_{\mathbf{A}_1^T\mathbf{A}_1}(x)] - \frac{\alpha}{m}\mathbb{E}[p^{(k-1)}_{\mathbf{A}_1^T\mathbf{A}_1}(x)]' \\ =& a_n^{(k)}x^n + \cdots + a_1^{(k)}x + a_0^{(k)}.\end{aligned} \tag{49}$$

The results in (24) follow from Vieta's formulas that relate roots of polynomials to their coefficients. The expected value of the VCE($\mathbf{A}_1$) follows directly as the constant coefficient of the expected characteristic polynomial since it is the product of the roots while for the MSE($\mathbf{A}_1$) we have

$$\sum_{i=1}^{n}\frac{1}{\lambda_i(\mathbf{A}_1^T\mathbf{A}_1)} = \frac{\prod_{i\neq 1}\lambda_i(\mathbf{A}_1^T\mathbf{A}_1) + \cdots + \prod_{i\neq n}\lambda_i(\mathbf{A}_1^T\mathbf{A}_1)}{\prod_{i=1}^{n}\lambda_i(\mathbf{A}_1^T\mathbf{A}_1)}. \tag{50}$$

From this it follows that

$$\begin{aligned}\mathbb{E}[\text{MSE}(\mathbf{A}_1)] = -\frac{a_1^{(k)}}{a_0^{(k)}},\ \mathbb{E}[\text{WCE}(\mathbf{A}_1)] \geq -\frac{a_1^{(k)}}{a_0^{(k)}n}, \\ \mathbb{E}[\text{VCE}(\mathbf{A}_1)] = \log(a_0^{(k)}).\end{aligned} \tag{51}$$

Notice that for $k = n$ the expected characteristic polynomial is an associated Laguerre polynomial $n!L_n(x)$ [47] with coefficients

$$a_i^{(n)} = \frac{(-1)^i n!}{i!}\left(\frac{\alpha}{m}\right)^{n-i}\binom{n}{i},\ i = 0,\dots,n. \tag{52}$$

Then for $k \geq n$ we have the coefficients of interest

$$\begin{aligned}a_0^{(k)} &= n!\left(\frac{\alpha}{m}\right)^n\binom{k}{k-n}, \\ a_1^{(k)} &= -n!\left(\frac{\alpha}{m}\right)^{n-1}\binom{k}{k-n+1}.\end{aligned} \tag{53}$$

Since we are dealing with positive semidefinite matrices the expected values have to be positive and therefore once we have written the values for $a_0^{(k)}$ and $a_1^{(k)}$ results in (24) follow immediately from (53).

REFERENCES

[1] V. Roy, A. Simonetto, and G. Leus, "Spatio-temporal sensor management for environmental field estimation," *Signal Processing*, vol. 128, pp. 369–381, 2016.

[2] A. O. Hero III, D. A. Castanon, D. Cochran, and K. Kastella, *Foundations and Applications of Sensor Management*. Springer US, 2008.

[3] A. O. Hero III and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3064–3075, 2011.

[4] C. Jiang, Y. C. Soh, and H. Li, "Sensor placement by maximal projection on minimum eigenspace for linear inverse problems," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5595–5610, 2016.

[5] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *49th IEEE Conf. Decision Control*, 2010, pp. 2572–2577.

[6] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1135–1146, 2014.

[7] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

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

[9] S. Chepuri, G. Leus, and A.-J. van der Veen, "Sparsity-exploiting anchor placement for localization in sensor networks," in *21st Euro. Sig. Proc. Conf.*, 2013, pp. 1–5.

[10] S. Liu, A. Vempaty, M. Fardad, E. Masazade, and P. K. Varshney, "Energy-aware sensor selection in field reconstruction," *IEEE Signal Proc. Lett.*, vol. 21, no. 12, pp. 1476–1480, 2014.

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

[12] H. Jamali-Rad, A. Simonetto, and G. Leus, "Sparsity-aware sensor selection: Centralized and distributed algorithms," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 217–220, 2014.

[13] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Sig. Proc.*, vol. 63, no. 3, pp. 684–698, 2015.

[14] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Int. Symp. Inf. Proc. Sens. Netw.*, 2004, pp. 36–45.

[15] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placements in gaussian processes," in *Int. Conf. Mach. Learn.*, 2005, pp. 265–272.

[16] M. Naeem, S. Xue, and D. Lee, "Cross-entropy optimization for sensor selection problems," in *Int. Symp. Common. Info. Tec.*, 2009, pp. 396–401.

[17] L. Yao, W. A. Sethares, and D. C. Kammer, "Sensor placement for on-orbit modal identification via a genetic algorithm," *AIAA*, vol. 31, no. 10, pp. 1922–1928, 1993.

[18] S. Lau, R. Eichardt, L. D. Rienzo, and J. Haueisen, "Tabu search optimization of magnetic sensor systems for magnetocardiography," *IEEE Trans. Magn.*, vol. 44, no. 6, pp. 1442–1445, 2008.

[19] W. J. Welch, "Branch-and-bound search for experimental designs based on D-optimality and other criteria," *Technometrics*, vol. 24, no. 1, pp. 41–48, 1982.

[20] E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks," *IEEE Sig. Process. Lett.*, vol. 19, no. 12, pp. 845–848, 2012.

[21] S. P. Chepuri and G. Leus, "Sparsity-promoting adaptive sensor selection for non-linear filtering," in *Proc. IEEE ICASSP*, 2014, pp. 5080–5084.

[22] H. Jamali-Rad, A. Simonetto, X. Ma, and G. Leus, "Distributed sparsity-aware sensor selection," *IEEE Trans. Sig. Process.*, vol. 63, no. 22, pp. 595–5964, 2015.

[23] F. Altenbach, S. Corroy, G. Bocherer, and R. Mathar, "Strategies for distributed sensor selection using convex optimization," in *Proc. GLOBECOM*, 2012, pp. 2367–2372.

[24] S. Liu, S. P. Chepuri, M. Fardad, E. Masazade, G. Leus, and P. K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Trans. Sig. Process.*, vol. 64, no. 13, pp. 3509–3522, 2016.

[25] S. P. Chepuri and G. Leus, "Continuous sensor placement," *IEEE Sig. Process. Lett.*, vol. 22, no. 5, pp. 544–548, 2015.

[26] Y. Chen, C. Chuah, and Q. Zhao, "Sensor placement for maximizing lifetime per unit cost in wireless sensor networks," in *Proc. IEEE Military Commun. Conf.*, 2005, pp. 1097–1102.

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

[28] S. Liu, A. Vempaty, M. Fardad, E. Masazade, and P. K. Varshney, "Energy-aware sensor selection in field reconstruction," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1476–1480, 2014.

[29] S. Liu, F. Chen, A. Vempaty, M. Fardad, L. Shen, and P. K. Varshney, "Sparsity-promoting sensor management for estimation: an energy balance point of view," in *18th International Conference Information Fusion*, 2015, pp. 6–9.

[30] M. P. Vitus, W. Zhang, A. Abate, J. Hu, and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," *Automatica*, vol. 48, no. 10, pp. 2482–2493, 2012.

[31] V. Tzoumas, A. Jadbabaie, and G. J. Pappas, "Near-optimal sensor scheduling for batch state estimation: Complexity, algorithms, and limits," in *55th IEEE Conference on Decision and Control*, 2016.

[32] C. Rusu, J. Thompson, and N. M. Robertson, "Balanced sensor management across multiple time instances via l-1/l-infinity norm minimization," in *Proc. ICASSP*, 2017.

[33] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, 2014.

[34] J. J. Benedetto and M. Fickus, "Finite normalized tight frames," *Adv. Comput. Math.*, vol. 18, pp. 357–385, 2003.

[35] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of approximations for maximizing submodular set functionsi," *Math. Prog.*, vol. 14, pp. 265–294, 1978.

[36] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

[37] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[38] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[39] M. Broadbent, M. Brown, and K. Penner, "Subset selection algorithms: Randomized vs. deterministic," *SIAM Undergraduate Research Online*, vol. 3, pp. 50–71, 2010.

[40] D. A. Spielman and N. Srivastava, "An elementary proof of the restricted invertibility theorem," *Israel J. Math.*, vol. 190, no. 1, pp. 83–91, 2012.

[41] X. Chen, R. S. Womersley, and J. J. Ye, "Minimizing the condition number of a Gram matrix," *J. SIAM Optim.*, vol. 21, no. 1, pp. 127–148, 2011.

[42] Z. Lu and T. K. Pong, "Minimizing condition number via convex programming," *SIAM. J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1193–1211, 2011.

[43] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[44] Y. Eldar and G. Kutyniok, *Compressed Sensing*. Cambridge University Press, 2012.

[45] A. W. Marcus, D. A. Spielman, and N. Srivastava, "Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem," *Annals of Math.*, vol. 182, no. 1, pp. 327–350, 2015.

[46] J. Batson, D. A. Spielman, and N. Srivastava, "Twice-Ramanujan sparsifiers," in *41st Annual ACM Symposium on Theory of Computing*, 2009, pp. 255–262.

[47] I. Krasikov, "On extreme zeros of classical orthogonal polynomials," *J. Comp. and App. Math.*, vol. 193, no. 1, pp. 168–182, 2006.