2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

Feasibility Analysis on Optimal Sensor Selection in Cyber-physical Systems

Zhen Song, Chellury Ram Sastry, Nazif Cihan Taş
Siemens Corporate Research,
755 College Road East,
Princeton, NJ 08540
{zhen.song, chellury.sastry, nazif.tas}@siemens.com

YangQuan Chen
CSOIS, Electrical and Computer Engineering Department,
Utah State University,
Logan, UT 84321
yqchen@ieee.org

***Abstract*— A fundamental question for observation systems based on wireless sensor network (WSN) is to achieve the best tradeoff between estimation precision and many design factors. One challenge of this kind is to select a small number of sensors, if possible, to observe the environment, and thus conserve precious onboard battery energy by transmitting only valuable data to the base station. Although many sensor selection methods have been proposed, the analysis on this problem is relatively limited.**

Based on the theory of optimal experimental design and convex analysis, we present some feasibility analysis on the optimal sensor selection problem. Equipped with Fisher information matrix, we show that there exists a threshold, which we named Carathéodory's limit, such that the optimal estimation is always feasible as far as the number of selected sensors is no less than that limit. We also investigated on the difference between the total sample number and the total sensor number. Discussions on some necessary conditions of sensor density and sensor deployment patterns are included too. We argue that sensor selection methods have potentials to save significant amount of energy for a large class of embedded wireless sensor networks without sacrificing estimation accuracy.

I. Introduction

A. Motivation

In future, wireless sensor networks (WSNs) might significantly change the world [9]. Such a network could contains thousands of low-cost embedded sensor nodes, with each node equipped with a microprocessor, a wireless radio and several physical sensors. Cooperating with each other, these tiny sensor nodes could collect valuable monitoring data of the environment and enable numerous new applications [7]. A key challenge for WSN is to design sensor operations that would conserve precious onboard battery energy to transmit just enough data and yet satisfy estimation accuracy requirements. In fact, comparing to the energy cost for wireless communication, other energy costs are normally ignored [18].

Under this background, the optimal sensor selection problem (SSP) attracted more and more attention in recent years. Many algorithms have been developed in order to select a certain number of sensors and achieve the optimum criterion. They are usually formulated as the minimization of a cost function [10], [28], [5], [3], [27] or the maximization of a utility function [2].

Despite the considerable number of publications on the topic, an important question that has not been studied in depth is that *if sensor selection is even feasible and worthwhile*. The fact that an algorithm could select, say, twenty sensors out of one hundred does not automatically justify the effort. What if we "save" 80% energy but collect garbage data? That is in fact a 20% of waste. In addition, if only a small set of networks are feasible for sensor selection, the impact of the method is also limited. In this work, we discuss some fundamental characteristics of sensor selection methods. The conclusion is that sensor selection is feasible for many densely deployed sensor networks. Based on the results, readers can estimate the potential energy savings due to sensor selection methods for a specific project.

One important difference between WSN and the existing Internet technology is that the information transmitted via WSN is usually the measurement on the physical world and therefore subject to certain constraints due to physical laws. Naturally, the networking problems are then strongly coupled with the physical systems. Studying such relations between networks and the physical world is the central mission for the research of cyber-physical systems [22]. The methods in this work are model-based approaches.

B. Counter Examples

To show that even an "optimal" SSP method could be improper, we can imagine some counter examples, as illustrated in Fig. 1. We want to get the best estimation from the sensor network with the least energy costs. Assuming there are n sensors in total. If *case a* is always true, then why bother to develop a generic SSP method to choose k_S sensors out of n? A brute force method that tests sensors one by one from the nearest might be good enough. Meanwhile, if *case b* holds, then, as we explained, SSP methods can do nothing but waste energy and collect useless data.

Sensor selection is more meaningful for *case c* and *case d*. If *case c* is true, then, naturally, we need a solid method to estimate the necessary number of selected sensor based on the requirement of the estimation accuracy. However, our analysis reveals an interesting phenomenon: *case d* is closer to the reality, even though it might be counter intuitive at the first look. In fact, the estimation error can be minimized after the number of selected sensors is equal or bigger than a threshold C_l, which we named *Carathéodory's limit*, and C_l is usually a number much smaller than n. Notice that the

978-1-4244-4524-0/09/$25.00 ©2009 AACC

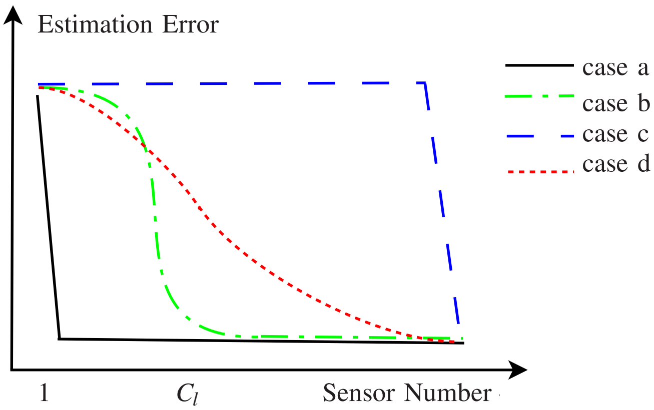

Fig. 1. Possible error curves for sensor election problems.

minimum error may be achievable with even less than C_l sensors, but there is no guarantee.

In this work, we will try to systematically answer the above questions in a framework based on the theory of optimal experimental design. One key idea is to incorporate the mighty Carathéodory's theorem, which has not been discussed with SSP before, into our analysis. Another simple yet important idea from us is to use *general sampling*, which will be addressed later, as a bridge toward Carathéodory's theorem.

For presentation purposes, we firstly introduce background knowledge in Sec. II-A, then review the related work in Sec. II-B. The problem is formulated in Sec. II-C, and discussed in the following sections. To verify the theory, we actually developed two sensor selection methods. Finally, Sec. III concludes the article and list some future work. Please check the notation list in the Appendix frequently.

II. Problem Formulation and Analysis

A. Carathéodory's theorem and Fisher information matrix

Carathéodory's theorem [19] plays the key role in our analysis. The theorem tells us the required number of vectors in order to represent a point in a convex domain.

Let us introduce the concept of convex hull.

Definition 1 (Convex hull): The convex hull of a set $\mathbb{S}_C$, denoted $\text{conv}(\mathbb{S}_C)$, is the set of all convex combinations of points in $\mathbb{S}_C$:

$$\text{conv}(\mathbb{S}_C) = \{\sum_{i=1}^{n} c_i \mathbf{x}_i | \mathbf{x}_i \in \mathbb{S}_C, c_i \geq 0, \sum_{i=1}^{n} c_i = 1\},$$

where $\mathbf{x_i}$ is a column vector.

Following our conventions on notations, Carathéodory's theorem can be presented as follows:

Theorem 2 (Carathéodory's theorem, based on [19], p.72): Let $\mathbb{S}_C$ be a subset of $\mathbb{R}^k$. Every element $\mathbf{x}$ in $\text{conv}(\mathbb{S}_C)$ can be expressed as a convex combination of no more than $k+1$ elements of $\mathbb{S}_C$. If $\mathbf{x}$ is on the boundary of $\text{conv}(\mathbb{S}_C)$, $k+1$ can be replaced by k.

Now, we formally introduce Fisher information matrix (FIM). For a linear observer[1], the ideal n sensor measurements, $\mathbf{s}$, are defined as

$$\mathbf{s} = \mathbf{A}^T \mathbf{q}^*,$$

where $\mathbf{q}^*$ is the vector of m unknown parameters under estimation. Thus, $\mathbf{q}^* \in \mathbb{R}^m$, $\mathbf{s} \in \mathbb{R}^n$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. In practice, the sensor measurement is always corrupted by noise $\bar{\mathbf{v}}$,

$$\begin{aligned} \mathbf{y} &= \mathbf{s} + \bar{\mathbf{v}}, \\ \bar{v}_i &= \frac{1}{n_i} \sum_{k=1}^{n_i} v_i[k], \end{aligned}$$

where n_i is the number of samples that sensor i takes in the time slot t_S; $v_i[k]$ is the *sensor noise* for the k-th sample and $\bar{v}_i$ is the associated *measurement noise*.

The statistical property of the sensor noise is determined only by the hardware characteristics of the onboard (physical) sensors. At the moment, we consider $\mathbf{v}$ as the independent white Gaussian noise such that $v_i \sim \mathcal{N}(0, \sigma_i^2)$, where σ_i is the standard deviation of the sensor noise. Other types of noises are out of the scope of this article but being discussed in another our work [20], [21]. An effective method to reject the assumed sensor noise is to take n_i *samples*, average them and result in *one measurement*. Then, the measurement is transmitted to the based station, which computes one *estimate* on the unknown parameters according to the received measurements from the network. It is easy to see that the measurement noise $\bar{\mathbf{v}}$ is reduced such that

$$\bar{v}_i \sim \mathcal{N}(0, \sigma_i^2 / n_i).$$

If $n_i \in \{0, 1\}$, we call it the *single sampling* scheme, which is the *only* scenario being discussed in the current literature, to our best knowledge. Obviously, since the number of total samples, i.e., the number of samples taken by the whole network, is always the same as the number of selected sensors in this scheme, it is impossible to distinguish the different characteristics between samples and sensors. Now, we consider a more generic *general sampling* scheme, where $n_i \in [0, c]$, $n_i \in \mathbb{Z}$, $c \in \mathbb{Z}$, and c is a constant positive integer. Then, given m parameters, i.e., $\mathbf{q} \in \mathbb{R}^m$, and n sensors, such that $\mathbf{y} \in \mathbb{R}^n$, FIM can be defined in Def. 3.[2]

Definition 3 (Fisher information matrix):

$$\begin{aligned} \mathbf{M} &= \mathbf{A}\Sigma^{-1}\mathbf{A}^T, \qquad (1) \\ \Sigma^{-1} &= \begin{bmatrix} n_1\sigma_1^{-2} & 0 & & \\ 0 & n_2\sigma_2^{-2} & & \\ & & \ddots & \\ & & & n_n\sigma_n^{-2} \end{bmatrix}. \end{aligned}$$

Do not get confused with the notation of sum, $\sum$, and the sigma matrix Σ.

An important application of FIM is to estimate the covariance matrix of the estimate, $\text{cov}(\hat{\mathbf{q}})$, where $\hat{\mathbf{q}}$ is the estimate on the true parameter $\mathbf{q}^*$. In plain words, the covariance matrix is a confidence ellipsoid that encloses the most likely values of the true parameter. We always want to minimize the ellipsoid since the smaller ellipsoid indicates a better prediction on the true parameter. The theorem of Cramér-Rao inequality [20], [15] told us that the minimum estimation

[1]For non-linear observers, we can piecewisely linearize it. Refer to our target tracking example in Sec. II-D.

[2]In some literatures [19], the $\mathbf{M}$ matrix in Def. 3 is called information matrix. We do not distinguish them in this work. Refer to [21], [17], [24], [19] for details.

error we can expect is bounded by $\mathbf{M}^{-1}$. In other words, if we maximize the volume of $\mathbf{M}$ (or minimize that of $\mathbf{M}^{-1}$), we will get the best unbiased estimation. If the maximum likelihood estimator, $\mathbf{q}_{ML}$, exists, $\mathrm{cov}(\mathbf{q}_{ML}) = \mathbf{M}^{-1}$ according to the Cramér-Rao inequality [8]. There are rich literatures [1] on how to precisely measure the volume of a matrix. In this work, we took the widely used D-optimality criterion [19], [1], Ψ, which is defined as

$$\Psi(\mathbf{M}) = -\ln\det(\mathbf{M}).$$

Such a criterion has been used by other works in sensor selection [11], [6], [23], [12].

B. Literature Review

The SSP has been discussed from different aspects before and the detailed formulations used in several prior works are usually diverse in nature [10], [28], [5], [3], [2], [27]. However, the main idea behind all these works is the same and can be summarized as balancing the trade off between the precise estimation and the energy costs on communications.

Many sensor-network-based systems are used to observe systems that are governed by certain physical laws. Naturally, the model of those systems could provide valuable information for us to simplify the problem [22]. In this work, we focus on model-based sensor selection methods.

A model-based geometric method is presented in [10]. Based on geometrical analysis of camera-like sensors, it is concluded that the sensor selection problem can be solved in polynomial time. The authors of this work also observe that four sensors can provide the position estimation as good as that from all sensors. The sensor selection problem can be studied from the perspective of information compression as well since the sensor data with high uncertainty or redundancy should not be transmitted via the communication channel. Uncertainty is measured by entropy whereas redundancy can be measured by correlations. For example, [28] presents a sensor selection method based on entropy filtering and Bayes' theorem. This work uses a grid-based method where the area of interest is segmented into many small cells, and the entropy of the target's location is computed based on a probability mass function. Since smaller entropy indicates less information uncertainty, [28] proposes an algorithm to select proper sensors by minimizing the entropy. Thus, only the "good" information with more certainty is sent to the base station. The probability mass function can be recursively updated according to the dynamics of the target using a Bayes' theorem based updating method. More details about this method are presented in several other works [14], [30], [29]. Based on the same framework, [26] proposes a faster heuristic entropy-based sensor selection method. In [25], a tempo-spatial correlation model is assumed. The variance of the estimation noise is minimized by selecting proper sensors.

In addition, sensor selection can be discussed within the framework of optimal estimation, where the estimation error is normally used as the cost function of the optimization. In previous works [5], [4], [3], [11], [12], and [23], the SSP is formulated as a constrained 0-1 integer programming problem, where the estimation error is minimized.

Because this approach is the most related, we include more details on the formulation. Along this approach, the decision on whether to select sensor i is indicated by a binary integer $n_i \in \{0,1\}$, with 1 as being selected and 0 as not.[3]

These type of SSP can be formulated as follows

Definition 4 (Binary SSP):

$$\begin{aligned} \hat{\mathbf{n}} &= \arg\min \Psi(\mathbf{M}(\mathbf{n})), \\ \mathbf{M} &= \sum_{i=1}^{n} n_i \mathbf{M}_i, \\ \text{subject to: } n_i &\in \{0,1\}, \\ \sum n_i &= k_S, \end{aligned}$$

where the k_S is the number of selected sensors "given by users." As what we explained after Fig. 1, finding the proper k_S is in fact not trivial. We may actually waste energy if the selection on k_S is not proper. We will resume the discussion shortly. $\mathbf{M}_i$ is the FIM that is associated with sensor i,

$$\begin{aligned} y_i &= \mathbf{a}_i^T \mathbf{q}^*, \\ \mathbf{M}_i &= \sigma_i^{-2} \mathbf{a}_i \mathbf{a}_i^T, \\ \mathbf{A} &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}. \end{aligned} \quad (2)$$

C. Formulations

As aforementioned, our approach makes better use of on-board processing power of the modern "smart sensors." For typical sensor nodes, the energy costs for such sampling and filtering operations are ignorable comparing to the energy consumption for wireless communications [16].

To simplify the notation, we use normalized sampling rate, $\mathbf{p}$, through the rest of our work, and $p_i = \bar{p}_i / \sum \bar{p}_i$, where $\bar{p}_i$ is the (physical) sampling rate for sensor i, defined as $\bar{p}_i = n_i / t_S$. It is immediate that $\sum p_i = 1$. Now, we have $\mathbf{M} = n_S \sum p_i \mathbf{M}_i$, where $\mathbf{M}_i$ is defined as in Eq. 2 and $n_S = \sum n_i$. Introducing $\bar{\mathbf{M}}$ as

$$\bar{\mathbf{M}} = \sum p_i \mathbf{M}_i,$$

then it is obvious that $\arg\min \Psi(\bar{\mathbf{M}}(\mathbf{p})) = \arg\min \Psi(\mathbf{M}(\mathbf{p}))$. In fact, $\Psi(\mathbf{M}) = \Psi(\bar{\mathbf{M}}) - m \ln n_S = \Psi(\bar{\mathbf{M}}) - c_1$, where c_1 is a constant for a fixed n_S.

Definition 5 (Sampling rate optimization problem (SROP)):

$$\begin{aligned} \hat{\mathbf{p}} &= \arg\min \Psi(\bar{\mathbf{M}}(\mathbf{p})), \\ \bar{\mathbf{M}} &= \sum_{i=1}^{n} p_i \mathbf{M}_i, \\ \text{subject to: } \mathbf{p} &\succeq 0, \\ \sum \mathbf{p} &= 1. \end{aligned}$$

If we multiply $\mathbf{p}$ in Def. 5 by k_S, it is obvious that the binary SSP is the SROP with more constraints. Therefore, the precision of binary SSP can not be better than that of SROP. Although SROP itself does not provide a solution for SSP, it establishes a bound of the estimation error. In addition, through the analysis of the SROP, we know a formulation that we named Relaxed Convex SSP (RCSSP) is solvable under reasonable conditions. The formulation on the problem is in Def. 7 and the analysis on the condition is presented in Sec. II-D. Now we have to introduce a new math symbol $\succeq^k$.

Definition 6 (Math symbol $\succeq^k$): If every entry of vector $\mathbf{v}$ is non-negative and at least k entries of $\mathbf{v}$ equal to 0, where k

[3]Notice that the notation $c \in \{a,b\}$ means that c has to be either a or b, while the notation $c \in [a,b]$ means that $c \in \mathbb{R}, a \le c \le b$.

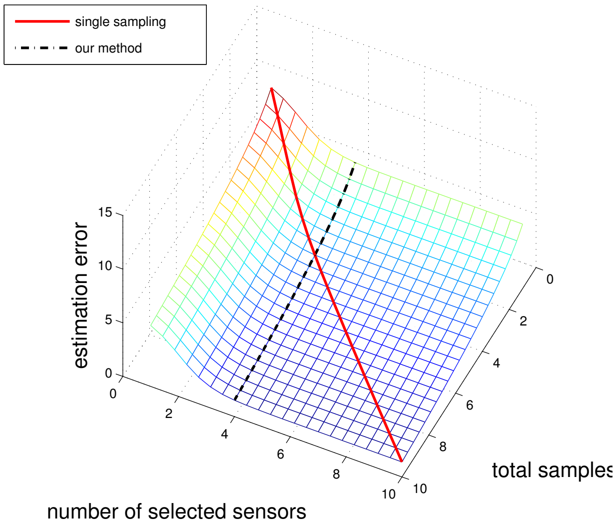

Fig. 2. Illustration on Carathéodory's limit.

is a non-negative integer number and $m \geq k$, then we denote it as $\mathbf{v} \succeq^k 0$.
A related symbol is $\succeq$, which is equal to $\succeq^0$. It is immediate that if $\mathbf{p} \succeq^{n+1} 0$, then $\mathbf{p} \succeq^n 0$.

Definition 7 (RCSSP):

$$\begin{aligned} \hat{\mathbf{p}} &= \arg\min \Psi(\bar{\mathbf{M}}(\mathbf{p})), \\ \bar{\mathbf{M}} &= \sum_{i-1}^{n} p_i \mathbf{M}_i, \\ \text{subject to}: \ \mathbf{p} &\succeq^{n-C_l-1} 0, \\ \sum \mathbf{p} &= 1, \end{aligned}$$

where $\mathbf{p} \in \mathbb{R}^n$, $C_l = m(m+1)/2$, and $\bar{\mathbf{M}}, \mathbf{M}_i \in \mathbb{R}^{m \times m}$.

D. Carathéodory's Limit

Notice that we only discuss Carathéodory's limit, C_l, under the context of sensor selection, where $C_l = m(m+1)/2$. The concept of Carathéodory's limit is more than just a number. It indeed represents a relationship among the number of "total samples," the "number of selected sensors," and the worst-case "estimation error." The concept is illustrated in Fig. 2. The term "our method" refers to the sensor selection methods that choose sensors no more than C_l, while the existing 0-1 integer programming approaches are depicted by the "single sampling" curve. As the figure shows, the worst-case estimation error reduces to the minimum when the number of selected sensors is about C_l and the error does not reduce further since that. On the other hand, the estimation error gradually decreases along the axis of "total samples." The contribution of the concept of Carathéodory's limit is due to the fact this structure in the solution domain of the SSP has not been reported.

The significance of the Carathéodory's limit is described as follows:

- It justifies the sensor selection methods, because the limit exits for generic WSN-based observation systems.
- The network designs and the sensor selection algorithm designs are connected via the limit. Therefore, systematic co-designs are possible. While the existing methods select sensors from a given network, we are more interested in designing the network and the sensor selection algorithm together, such that the worst-case performances, in terms of observation precision and energy costs, can be guaranteed.
- The limit establishes a baseline to compare different sensor selection methods. According to the structure of the solution domain, we conclude that it is fair to compare networks with the same number of total samples.

Introducing another notation "stack," which converts the lower diagonal part of a matrix to a column vector. For example

$$\begin{aligned} \bar{\mathbf{M}} &= \begin{bmatrix} \bar{\mathbf{M}}_{(1,1)} & \bar{\mathbf{M}}_{(1,2)} \\ \bar{\mathbf{M}}_{(2,1)} & \bar{\mathbf{M}}_{(2,2)} \end{bmatrix} \\ \text{stack}(\bar{\mathbf{M}}) &= \begin{bmatrix} \bar{\mathbf{M}}_{(1,1)} \\ \bar{\mathbf{M}}_{(2,1)} \\ \bar{\mathbf{M}}_{(2,2)} \end{bmatrix} \end{aligned}$$

Since the FIM is symmetric (cf. Eq.2) , $\text{stack}(\bar{\mathbf{M}})$ contains all information of $\bar{\mathbf{M}}$.

Theorem 8: If $\text{rank}(\mathbf{A}) = m$, the RCSSP in Def. 7 is solvable.

Proof If $\text{rank}(\mathbf{A}) = m$, there exists sampling rate vectors such that $\bar{\mathbf{M}}$ is full rank. That is, $\min \Psi(\bar{\mathbf{M}}(\mathbf{p}))$ exists for a set of $\mathbf{p}$. In fact, $p_i = 1/n$ is one example. Define $\bar{\mathbf{M}}^*$ as $\Psi(\bar{\mathbf{M}}^*) = \min \Psi(\bar{\mathbf{M}}(\mathbf{p}))$. It is obvious that $\text{stack}(\bar{\mathbf{M}}^*)$ is on a convex hull. Therefore, based on Def. 7,

$$\text{stack}(\bar{\mathbf{M}}^*) \in \text{conv}(\text{stack}(\mathbf{M}_1), \text{stack}(\mathbf{M}_2), \cdots, \text{stack}(\mathbf{M}_n)),$$

and $\text{stack}(\mathbf{M}_i)) \in \mathbb{R}^{C_l}$, where $i \in \{1, \cdots, n\}$. According to Carathéodory's theory, at most $C_l + 1$ elements are required to represent $\text{stack}(\bar{\mathbf{M}}^*)$. Therefore, at least $n - (C_l + 1)$ entries in $\mathbf{p}$ can be zeros. ■

Problems related to Theorem 8 have been discussed in [24], [19], [13]. Different from the existing works, our formation is specifically developed for SSP, so that we only require $\mathbf{A}$ to be full rank. For example, the interpretation of the requirement for target tracking applications is that the sensors should not be placed along a line within the domain where the target is moving. As the result, we conclude that the requirement is not demanding, therefore, many networks are sensor-selection feasible.

With longer derivations, we can prove that, instead of $C_l + 1$ sensors, only C_l sensors are required [21], but such a difference is not important to the conclusions of this article.

Because C_l is the worst-case bound, it is possible to develop algorithms that selects less than C_l sensors. In another piece of our work [21], we proposed a two-step scheme. Firstly, we use a light-weight heuristic method named hCOSS to select sensors, which usually picks C_l sensors or less. Secondly, in that cases when more than $C_l + 1$ sensors are selected, we can use an algorithm to eliminate unnecessary sensors and guarantee that no more than $C_l + 1$ sensors are selected. The elimination process is the key idea of our eCOSS algorithm.

An example of such scenario is shown in Fig. 3, where light sensors are used to locate the position of a lamp. We piecewisely linearize the non-linear sensor model and apply

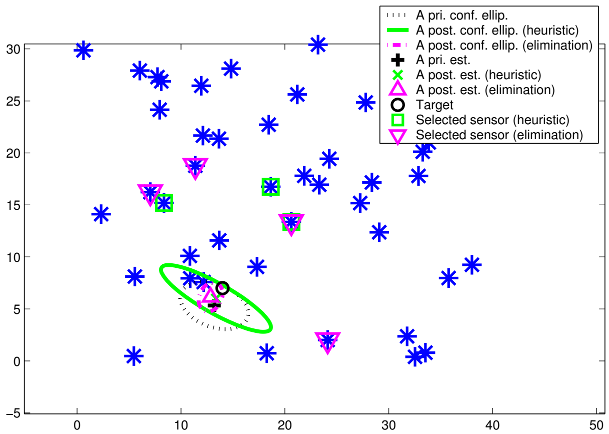

Fig. 3. Sensor selection by the hCOSS and eCOSS methods.

the least squares (LS) method to estimate the position of the lamp, which is referred as the *a priori* position, based on sensor data from the whole network. Then we select sensors based on the *a priori* position. We use the hCOSS and eCOSS methods to select no more than $C_l + 1$ sensors and estimate the position based on the data from the selected sensors. The position is referred as the *a posteriori* position. On a 3GHz PC, the hCOSS method can select sensors out of 60 candidates in about 8 ms, while the eCOSS method requires about 15 ms to 25 ms.

In an independent work [10], a non-real-time geometrical SSP method is proposed, and it is observed and proved that for 2D target tracking applications, "the estimates that are obtained by four sensors are as good as the estimates obtained from all sensors." Based on FIM, our method confirmed that conclusion and our method is applicable to general parameter estimation problems with arbitrary high dimensions. Therefore, our method can be used to observer more parameters.

E. Sensor-selection-feasible networks

Thanks to the aforementioned analysis, we are ready to answer some interesting design questions, such as the follows.

a) Necessary density condition: It is easy to imagine some scenarios where no sensor selection method could help. Such as, if there is only one sensor at all. Certainly, there is a requirement on the minimal sensor density, ρ_L, in order to justify the SSP methods. If the sensor's detection range is r, then ρ_L should satisfy

$$\rho_L > C_l / \pi r^2,$$

in order for a sensor selection algorithm to save energy without scarifying the estimation precision. If the sensor density falls below ρ_L, there is no guarantee on the estimation accuracy.

b) Monotone: Back to the questions following Fig. 1, is *case d* monotonic? Remind

$$\Psi(\mathbf{M}) = \Psi(n_S \sum_i \mathbf{M}_i) = \Psi(\bar{\mathbf{M}}) - m \ln n_S. \quad (3)$$

Therefore, for a deployed network with the sensor density no less than ρ_L, we can increase the total samples n_S to achieve *arbitrary accuracy*, since $\Psi(\mathbf{M})$ is infinitely small if $n_S \to +\infty$. Therefore, for the single sampling scheme, the error curve is monotonic and similar to *case d*, which is similar to the figures shown in an independent work [11]. In Fig. 2, *case d* is the "single sampling" curve.

c) Samples vs. sensors: What is the difference between samples and sensors? When will *case c* and *case d* hold? For single-sampling schemes, *case d* dominates. However, it is unfair to say that "the estimation error gradually is reduced as more sensors are selected," unless the constraint on single-sampling is properly stated. This property is illustrated in Fig. 2. In general, without such a constraint, the estimation error drops quickly along the increase of the selected sensors and stays at the minimum value after the sensor number reaches C_l, which is usually a quite small number. This scenario is depicted in *case c*, which is the "our method" curve in Fig. 2.

For a fixed n_S, the estimation of general sampling scheme is not worse than that of the single sampling method. However, if the single sampling scheme takes a larger n_S than the general sampling method, the former one has chances to offer more accurate estimation. Therefore we conclude that it is *unfair* to compare different sensor selection methods with different total sample numbers.

d) Sensor deployment pattern: Is there an ideal sensor deployment pattern for sensor selection purposes? The limit of $\text{rank}(\mathbf{A}) = m$ indicates that the sensor selection methods are not very "picky" on the sensors' locations. For example, for 2D tracking problems, where $\mathbf{q}^*$ is the position of the target, $\mathbf{a}_i$ are linear dependent only when the sensors and the target are along one line. But this pattern is intuitively improper. The conclusion is that many networks are sensor-selection feasible.

III. Conclusion and Future Work

Based on FIM and Carathéodory's theory, we discovered that the optimal estimate on unknown parameters is achievable by a relatively small number of sensors, which is named as Carathéodory's limit. The different impacts of sensors and samples are discussed in depth. The analysis also provides guidelines on sensor deployment. To verify the analysis, we developed two sensor selection methods that use the general sampling scheme. We conclude that sensor selection methods can significantly save valuable onboard energy for many CPSs that observe environments using WSNs.

As a future work, we plan to analyze our algorithms from the computer networking perspective in order to complement our algorithms with more realistic networks.

Appendix: Concepts and Notations

- FIM: Fisher information matrix.
- WSN: wireless sensor network.
- CPS: cyber-physical system.
- Measurement and sample: a sensor can take multiple samples, processes them and results in one measurement.
- Notation on scalars, matrices and vectors: a vector is indicated by a bold lower letter, such as $\mathbf{n}$, while a matrix is a capital bold letter, e.g., $\bar{\mathbf{M}}$. The scalar n should be not confused with vector $\mathbf{n}$. The subscripts in brackets indicate scalar entries of a matrix or vector. The

subscripts without brackets are instances of variables. For example, $\mathbf{M}_i(1,1)$ is the $(1,1)$th entry of the matrix $\mathbf{M}_i$.

- Index: the integers in [ ] are indices of discrete time or iteration number, such as $\bar{\mathbf{M}}(\mathbf{p}[k])$.
- conv( ): convex hall.
- cov( ): covariance.
- n : the total number of sensors.
- t_S: the sampling time. In t_S, sensor i collects n_i samples.
- n_S: the total number of samples of the whole network in t_S time slot.
- k_S: number of sensors to be selected by users.
- $\mathbf{n}$: the sample vector. The number n_i is the number of samples that sensor i collects in each t_S time slot.
- m: the number of unknown parameters.
- $\mathbf{M}$, $\bar{\mathbf{M}}$, and $\mathbf{M}_i$: FIM.
- c_1: a constant.
- C_l: Carathéodory's limit, which is $m(m+1)/2$.
- y_i: the ideal measurement value of sensor i.
- s_i: the measurement from sensor i, which is y_i plus the sensor noise.
- $\mathcal{N}(\mu,\sigma^2)$: Gaussian (normal) distribution with the expectation of μ and the standard deviation of σ.
- $\bar{\mathbf{p}}$: sampling rate of sensors.
- $\mathbf{p}$: normalized sampling rate of sensors.
- $\mathbf{q}$, $\mathbf{q}^*$, $\mathbf{q}_{ML}$: $\mathbf{q}$ is the position of the target. Specifically, $\mathbf{q}^*$ is the true position of the target and $\mathbf{q}_{ML}$ is the maximum likelihood estimate on $\mathbf{q}^*$.
- $\mathbf{a} \succeq b$: each entry of vector $\mathbf{a}$ is no less than the scalar b. $a_i \succeq b$.
- $\mathbf{v}$, $\bar{\mathbf{v}}$: measurement noise vector.

References

[1] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*. Claredon Press, Oxford, 1992.

[2] F. Bian, D. Kempe, and R. Govindan, "Utility-based sensor selection," in *Proceedings of International Conference on Information Processing in Sensor Networks (IPSN)*, Nashville, TN, Apr. 2006. [Online]. Available: http://citeseer.ist.psu.edu/747271.html

[3] A. L. Buczak, H. H. Wang, H. Darabi, and M. A. Jafari, "Genetic algorithm convergence study for sensor network optimization," *Information Sciences*, vol. 133, pp. 267–282, 2001.

[4] A. S. Chhetri, D. Morrell, and A. Papandreou-Suppappola, "Energy efficient target tracking in a sensor network using non-myopic sensor scheduling," in *Proceedings of IEEE International Conference on Information Fusion*, Philadelphia, PA, 2005.

[5] ——, "Sensor scheduling using a 0-1 mixed integer programming framework," in *Proceedings of IEEE Workshop on Sensor Array and Multichannel Processing*, Waltham, MA, July 2006, pp. 471–475.

[6] A. S. Chhetri, "Sensor scheduling and efficient algorithm implementation for target tracking," Ph.D. dissertation, Arizona State University, 2006.

[7] M. A. Demetriou and I. I. Hussein, "Estimation of spatially distributed processes using mobile spatially distributed sensor network," *SIAM J. Control and Optimization*, vol. 48, no. 1, pp. 266–291, 2009.

[8] A. F. Emery and A. V. Nenarokomovz, "Optimal experiment design," *Measurement Science & Technology*, vol. 9, no. 6, pp. 864–876, 1998. [Online]. Available: http://www.iop.org/EJ/abstract/0957-0233/9/6/003

[9] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lack City, Utah, USA, May 2001, pp. 2033–2036.

[10] V. Isler and R. Bajcsy, "The sensor selection problem for bounded uncertainty sensing models," *IEEE Transactions on Automation Science and Engineering*, vol. 3, no. 4, pp. 372–381, Oct. 2006. [Online]. Available: http://www.cs.rpi.edu/~isler/new/pub/pubs/isler05tase.pdf

[11] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, 2008. [Online]. Available: http://www.stanford.edu/~boyd/papers/sensor_selection.html

[12] L. Kaplan, "Local node selection for localization in a distributed sensor network," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 136 – 146, 2006.

[13] S. Karlin and W. Studden, *Tchebycheff systems:with applications in analysis and statistics*. New York: Wiley, 1966.

[14] J. Liu, J. Reich, and F. Zhao, "Collaborative in-network processing for target tracking," *EURASIP Journal on Applied Signal Processing*, vol. 4, pp. 378–391, Mar. 2003. [Online]. Available: http://citeseer.ist.psu.edu/liu02collaborative.html

[15] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.

[16] Moteiv Corporation, "Tmote Sky datasheet," Nov. 2006. [Online]. Available: http://www.moteiv.com/products/tmotesky.php

[17] A. Pazman, *Foundations of Optimum Experimental Design*. Veda, Bratislava, Czechoslovakia: D. Reidel Publishing Company, 1986.

[18] J. Polastre, R. Szewczyk, and D. Culler, "Telos: Enabling ultra-low power wireless research," in *Proceedings of International Symposium on Information Processing in Sensor Networks (IPSN)*, Apr. 2005, pp. 364–369. [Online]. Available: http://www.moteiv.com/products/docs/an002-telos.pdf

[19] S. D. Silvey, *Optimal Design: An introduction to the Theory for Parameter Estimation*. New York: Chapman and Hall, 1980.

[20] Z. Song, "Optimal observation problems involving wireless sensor networks," Ph.D. dissertation, Utah State University, 2007. [Online]. Available: http://sites.google.com/site/zsonghomepage/publication/zsongDiss.pdf

[21] Z. Song, Y. Chen, C. R. Sastry, and N. C. Tas, *Optimal Observation in Cyber-physical Systems: A Fisher-information-matrix-based Approach*. Lodon, UK: Springer-Verlag, to be published in 2009.

[22] C. Tricaud and Y. Chen, "Optimal mobile actuator/sensor network motion strategy for parameter estimation in a class of cyber physical systems," in *Accepted by IEEE American Control Conference (ACC)*, Jun. 2009.

[23] D. Uciński and M. Patan, "D-optimal design of a monitoring network for parameter estimation of distributed systems," *Journal of Global Optimization*, 2006.

[24] D. Uciński, *Optimal Measurement Methods for Distributed Parameter System Identification*. Boca Raton, FL: CRC Press, 2005.

[25] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: Theory and applications for wireless sensor networks," *Computer Networks Journal (Elsevier)*, vol. 45, no. 3, pp. 245 –259, June 2004.

[26] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proceedings of the Third International Symposium on Information Processing in Sensor Networks (IPSN)*. Berkeley, CA: ACM Press, 2004, pp. 36–45. [Online]. Available: http://citeseer.ist.psu.edu/wang04entropybased.html

[27] H. Wang, K. Yao, and D. Estrin, "Information-theoretic approaches for sensor selection and placement in sensor networks for target localization and tracking : Communication and signal processing in sensor networks," *Journal of Communication and Networks*, vol. 7, no. 4, pp. 438–449, 2005.

[28] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration for tracking applications," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 61–72, Mar. 2002.

[29] F. Zhao and L. Guibas, *Wireless Sensor Networks: An Information Processing Approach*. Morgan Kaufmann Publishers, 2004.

[30] F. Zhao, J. Liu, J. Liu, L. Guibas, and J. Reich, "Collaborative signal and information processing: An information-directed approach," *IEEE Proceedings*, vol. 91, no. 8, pp. 1199–1209, Aug. 2003.