Upper Bounds for the Sensor Subset Selection Problem

Farhad Ghassemi and Vikram Krishnamurthy
Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, B.C., Canada
farhad@ece.ubc.ca and vikramk@ece.ubc.ca

Abstract – In this paper, we study the sensor subset selection problem with the determinant of the (Bayesian) Fisher information matrix (FIM) as the metric of estimation accuracy. As a combinatorial optimization problem, we analyze two well-known upper bounds for this problem: (i) the Lagrangian bound and (ii) the continuous bound. We show that the determinant of the FIM is a supermodular function from which it follows that the Lagrangian bound can be computed in polynomial time. We note that the continuous relaxation of the sensor subset selection problem can be transformed to a convex optimization problem from which it follows that the continuous bound is also computable in polynomial time. We also point to the benefit of using the natural selection process to solve the continuous relaxation of a variation of the sensor subset selection problem where sensors are allowed to make more than one measurement.

Keywords: Sensor management, non-linear parameter estimation, resource allocation, knapsack problem, supermodularity, natural selection.

1 Introduction

For a network of fixed sensors estimating the value of an unknown time-invariant parameter, the sensor subset selection problem is how to select M out of the total of N sensors such that a metric of estimation accuracy is maximized. Due to recent widespread applications of sensor networks in pervasive surveillance and environment monitoring, this problem has gained considerable attention [1–4]. The sensor subset selection problem is inherently a combinatorial optimization problem, i.e. a 0-1 integer program. This problem can be considered as an instance of resource allocation problems [5] or knapsack problems [6] with special characteristics of its own.

Carrying out an exhaustive search to solve the sensor subset selection problem is prohibitively expensive because \( \binom{N}{M} \) subsets must be evaluated for this purpose. Instead, we study the Lagrangian and continuous relaxations of the sensor subset selection problem which can be used in branch-and-bound algorithms [6] to avoid an exhaustive search.

We formulate the sensor subset selection problem with the determinant of FIM [7] as the metric of estimation accuracy. We first show that the determinant of the FIM can be represented by a homogenous polynomial as a function of decision variables, i.e. the activation command for sensors (see Proposition 1). We then prove that the determinant of the FIM is a supermodular function (see Proposition 2) and, as such, the formulated sensor subset selection problem is a supermodular knapsack problem. Gallo and Simeone [8] have shown that the Lagrangian relaxation of a supermodular knapsack problem can be solved in polynomial time by an outer approximation algorithm.

However, we also show that the continuous bound is tighter than the Lagrangian bound while it can be still found in polynomial time by transforming the continuous relaxation problem to a convex optimization problem. While, in general, standard algorithms for constrained convex optimization [9] can be exploited, we devise a very efficient algorithm for solving the continuous relaxation of a variation of the sensor subset selection problem. In this variation, called the measurement allocation problem, sensors are allowed to make more than one measurement and the cost of making each measurement is equal for all sensors. Knowing that the determinant can be represented by a homogenous polynomial, we then show that what is known as the natural selection process in population genetics [10] shares a unique stable equilibrium with the optimal allocation vector in the measurement allocation problem. Hence, by updating an initial allocation vector via the natural selection process, a sequence of allocation vectors are generated that eventually converges to the optimal allocation vector.

This paper is organized as follows. Section 2 formulates the sensor subset selection problem and proves
several important properties of its objective function. Sections 3 and 4 discuss the characteristics of the Lagrangian bound and the characteristics of the continuous bound, respectively. Section 5 gives some concluding remarks.

2 Formulation of the Sensor Subset Selection Problem

In this section, we first formally define the sensor subset selection problem with the determinant of the FIM as the objective function. Then, we derive several important properties of this objective function that will be used subsequently.

We start by denoting the set of all sensors by \( \mathcal{N} = \{1, 2, \ldots, N\} \) and the parameter vector which is unknown to sensors and must be estimated by \( \mathbf{p} \in \mathbb{R}^L \). The measurement model of sensors is characterized by

\[
\mathbf{z}_i = \mathbf{h}_i(\mathbf{p}) + \mathbf{v}_i \tag{1}
\]

for \( i = 1, 2, \ldots, N \) where \( \mathbf{z}_i \in \mathbb{R}^L \) is the measurement vector of the \( i \)-th sensor, \( \mathbf{h}_i \) is a vector-valued differentiable function of \( \mathbf{p} \), and \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N \) are independent zero-mean Gaussian random error vectors. The covariance matrix of \( \mathbf{v}_i \) is denoted by \( \mathbf{R}_i \). Later in this section, we comment on the implications of using a more general measurement model. Typically, the size of measurement vectors \( \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N \) is one, i.e. \( D = 1 \). However, as we see in the numerical example of Section 4, it can be also higher.

We use vector \( \mathbf{x} \in \{0, 1\}^N \) to denote the sensors that are selected. If the \( i \)-th element of \( \mathbf{x} \) is 1, sensor \( i \) is selected and if its \( i \)-th element is 0, sensor \( i \) is not selected. The measurements that are collected from the selected sensors are denoted by \( \mathcal{Z}(\mathbf{x}) \). The vectors corresponding to sets \( \mathcal{N} \), \( \{i\} \), and \( \emptyset \) are denoted by \( \mathbf{e} \), \( \mathbf{e}_i \) and \( \mathbf{0} \), respectively.

**Definition 1** The sensor subset selection problem is defined as solving one of the following:

1. **If \( \mathbf{p} \) is regarded as a non-random parameter [7]**,

\[
\max_{\mathbf{x} \in \{0, 1\}^N} \det(\mathbf{J}(\mathbf{x})) \quad \text{such that} \quad \mathbf{c}^T \mathbf{x} \leq M, \tag{2}
\]

where \( \mathbf{c} \in \mathbb{R}^N_+ \), \( M \in \mathbb{R}_+ \) and

\[
\mathbf{J}(\mathbf{x}) = \mathbb{E}_{\mathcal{Z}(\mathbf{x})}[\mathbf{p}] \left\{ \frac{\partial^T}{\partial \mathbf{p}} \log(p(\mathcal{Z}(\mathbf{x})|\mathbf{p})) + \frac{\partial}{\partial \mathbf{p}} \log(p(\mathcal{Z}(\mathbf{x})|\mathbf{p})) \right\} \tag{3}
\]

for \( \mathbf{x} \neq \mathbf{0} \) and \( \mathbf{J}(\mathbf{0}) = 0 \). Furthermore, it is assumed that \( \sum_{i=1}^N c_i > M > 0 \).

2. **If \( \mathbf{p} \) is regarded as a random parameter, i.e. a prior density \( p(\mathbf{p}) \) is given for \( \mathbf{p} \),**

\[
\max_{\mathbf{x} \in \{0, 1\}^N} \det(\mathbf{J}_B(\mathbf{x}))
\]

such that \( \mathbf{c}^T \mathbf{x} \leq M, \tag{4}
\]

where \( \mathbf{c} \in \mathbb{R}^N_+ \), \( M \in \mathbb{R}_+ \) such that \( \sum_{i=1}^N c_i > M > 0 \). Furthermore,

\[
\mathbf{J}_B(\mathbf{x}) = \mathbb{E}_p(\mathbf{p}) \{ \mathbf{J}(\mathbf{x}) \} \quad \text{where} \quad \mathbb{E}_p(\mathbf{p}) = \mathbb{E}_p(\mathbf{p}) \left\{ \frac{\partial^T}{\partial \mathbf{p}} \log(p(\mathbf{p})) + \frac{\partial}{\partial \mathbf{p}} \log(p(\mathbf{p})) \right\}. \tag{5}
\]

For the rest of this paper, we assume that the sensor subset selection problem is given by (2). However, most results either directly hold or can be worked out similarly for (4). We make a remark in this regard later.

The following proposition states how the FIM and its determinant can be expanded.

**Proposition 1** If the measurement model of the sensors is governed by (1), it holds that

1. \( \mathbf{J}(\mathbf{x}) = \sum_{i=1}^N x_i \mathbf{S}_i \tag{7} \)

where

\[
\mathbf{S}_i = \frac{\partial^T \mathbf{h}_i}{\partial \mathbf{p}} \mathbf{R}_i^{-1} \frac{\partial \mathbf{h}_i}{\partial \mathbf{p}}. \tag{8}
\]

2. \( \det(\mathbf{J}(\mathbf{x})) = \sum_{a_1=1}^N \sum_{a_2=1}^N \cdots \sum_{a_L=1}^N w_{a_1, a_2, \ldots, a_L} x_{a_1} x_{a_2} \cdots x_{a_L} \tag{9} \)

where

\[
w_{a_1, a_2, \ldots, a_L} = \det \left( \begin{bmatrix} (\mathbf{S}_{a_1})_{1,1} & (\mathbf{S}_{a_2})_{2,2} & \cdots & (\mathbf{S}_{a_L})_{L,L} \end{bmatrix} \right). \tag{10}
\]

We omit the proof of Proposition 1 here. We note that the second part of Proposition 1 states that the determinant of the FIM can be expressed as a homogeneous polynomial. The coefficient of each term of the polynomial, i.e. \( w_{a_1, a_2, \ldots, a_L} \), is the determinant of a matrix that is constructed by having the first column of \( \mathbf{S}_{a_1} \) as its first column, the second column of \( \mathbf{S}_{a_2} \) as its second column and so on.

Motivated by Proposition 1, we define the following function.
**Definition 2** Let $S_1, S_2, \cdots, S_N$ be some positive semi-definite matrices. We define function $v$ such that for any $x \in \mathbb{R}^N$, $v(x) = \det(\sum_{i=1}^N x_i S_i)$.

As we assume for the rest of the paper, we note that if $S_1, S_2, \cdots, S_N$ are given by (8), function $v$ then extends the definition of the determinant of the FIM to the set of real vectors. We next prove that function $v$ is supermodular over the set of non-negative real vectors.

**Proposition 2** If $x, x' \in \mathbb{R}_+^N$, it then holds that

$$v(x) + v(x') \leq v(x \vee x') + v(x \wedge x')$$  \hspace{1cm} (11)

where $x \vee x'$ denotes the componentwise maximum and $x \wedge x'$ denotes the componentwise minimum of $x$ and $x'$.

**Proof** We first note that function $v$ is non-negative and non-decreasing over $\mathbb{R}_+^N$. Inequality (11) can then be easily proved if $\sum_{i=1}^N x_i S_i$ and/or $\sum_{i=1}^N x'_i S_i$ are singular. We, therefore, consider the region of $\mathbb{R}^N_+$ where $\sum_{i=1}^N x_i S_i$ is non-singular for a given $x$. We denote this region by $\mathcal{X}$. To prove the supermodularity of function $v$ over $\mathcal{X}$, it suffices to show that $\frac{\partial^2 v(x)}{\partial x_j \partial x_i} \geq 0$ for any $x \in \mathcal{X}$ [11]. We note that

$$\frac{\partial v(x)}{\partial x_i} = \det \left( \sum_{k=1}^N x_k S_k \right) \text{trace} \left( \left( \sum_{k=1}^N x_k S_k \right)^{-1} S_i \right)$$  \hspace{1cm} (12)

and

$$\frac{\partial^2 v(x)}{\partial x_j \partial x_i} = \det \left( \sum_{k=1}^N x_k S_k \right) \left( \text{trace} \left( \left( \sum_{k=1}^N x_k S_k \right)^{-1} S_i \right) \text{trace} \left( \left( \sum_{k=1}^N x_k S_k \right)^{-1} S_j \right) - \text{trace} \left( \left( \sum_{k=1}^N x_k S_k \right)^{-2} S_i S_j \right) \right) \geq 0$$  \hspace{1cm} (13)

where we have used the fact that for any two positive semi-definite matrices $A$ and $B$, $\text{trace}(A)\text{trace}(B) \geq \text{trace}(AB)$ [12].

**Remark** We note that from (5), we can also write

$$J_B(x) = Q + E_{p(p)} \left\{ \sum_{i=1}^N x_i S_i \right\}$$

$$= Q + \sum_{i=1}^N x_i E_{p(p)} \{ S_i \}.$$  \hspace{1cm} (14)

where $S_1, S_2, \cdots, S_N$ are given by (8). Since all matrices in the above summation are positive semi-definite, the proof of Proposition 2 can be repeated to show that the determinant of $J_B(x)$ is also a supermodular function. However, the determinant of $J_B(x)$ cannot be represented as a homogenous polynomial because of the presence of matrix $Q$ which causes some terms in the polynomial to have a degree lower than $L$. However, if the inequality constraint of the problem (4) is replaced with an equality constraint, i.e. $\sum_{i=1}^N \frac{c_i}{M} x_i = 1$, one can always multiply the terms of the polynomial resulted from the expansion of the determinant with $\sum_{i=1}^N \frac{c_i}{M} x_i$ as many times as needed without affecting the value of the polynomial in order to obtain a homogenous representation.

### 3 Characteristics of the Lagrangian Relaxation

In this section, we study the Lagrangian relaxation of the problem (2). The Lagrangian relaxation yields a combinatorial optimization problem where the equality constraint of the problem (2) is moved into the objective function. The fact that the determinant is a supermodular function gives special characteristics to the Lagrangian relaxation of the problem (2). These characteristics were first noted by Gallo and Simeone in [8] for the more general case of the supermodular knapsack problem. The following proposition summarizes those results.

**Proposition 3** [8] Consider the following combinatorial optimization problem

$$\max_{x \in \{0,1\}^N} q(x)$$

such that $c^T x \leq M$

where $c \in \mathbb{R}^N_+$, $M \in \mathbb{R}_+$, $\sum_{i=1}^N c_i > M > 0$ and function $q$ is a generic function defined over $\{0,1\}^N$ with the following properties:

(i) $q(0) = 0$,

(ii) $e = [1,1,\cdots,1]^T$ is a maximum point of $q$ over $\{0,1\}^N$,

(iii) $q$ is a supermodular function over $\{0,1\}^N$.

Then, the dual function of the problem (15) is a piecewise linear convex function with at most $N$ breakpoints where the largest breakpoint is given by $u = \max_i \frac{q(e)}{c_i} | i = 1, 2, \cdots, N |$.

The main point of Proposition 3 is that the number of breakpoints of the dual function of the problem (15) does not grow exponentially with $N$ and, in fact, it is at most $N$. The piecewise linearity and convexity of the dual function, otherwise, hold even if properties (i) to (iii) are not satisfied.
It is easy to see that the objective function of the problem (2) has all the properties listed in Proposition 3. Condition (i) follows immediately from the definition of the objective function of this problem. Condition (ii) follows from the fact that the objective function is non-decreasing and, finally, condition (iii) follows from Proposition 2.

### 3.1 Outer Approximation Algorithm for Solving the Lagrangian Relaxation

As a result of these observations, an outer approximation algorithm [6] with at most N iterations can be applied to find the Lagrangian bound for the problem (2). We introduce some notation in order to describe this algorithm. The Lagrangian of the problem (2) is given by

\[
L(x, u) = v(x) - u(c^Tx - M).
\]

where \( u \in \mathbb{R}_+ \) is the Lagrange multiplier. The dual function is an upper bound for the maximum of function \( v \) over the feasible region. The dual function is obtained by maximizing the Lagrangian with respect to \( x \), i.e.

\[
\phi(u) = \max_{x \in [0, 1]^N} L(x, u).
\]

The Lagrangian bound for the problem (2) is given by

\[
\phi^* = \min_{u \in \mathbb{R}_+} \phi(u).
\]

In order to find this bound, we note that, for any non-negative \( u \), a line that passes through \((u, \phi(u))\) and has a slope equal to the subgradient of function \( \phi \) at that point is a lower bound for function \( \phi \). A subgradient of function \( \phi \) at \( u \) is given by

\[
g(u) = -(c^Tx(u) - M)
\]

where \( x(u) \) is the vector that maximizes \( L(x, u) \). Similarly, for any non-negative \( u' \) and \( u'' \), where \( u' < u'' \), \( g(u') < 0 \) and \( g(u'') > 0 \), function

\[
\phi_o(u; u', u'') = \max_u \{\phi(u') + g(u')(u - u'), \phi(u'') + g(u'') (u - u'')\}
\]

is a lower bound (and an outer approximation) of function \( \phi \). It is easy to check that the minimizer of function \( \phi_o \) is given by

\[
u^* = \frac{\phi(u'') - g(u'')u'' - (\phi(u') - g(u')u')}{g(u') - g(u'')}
\]

Clearly, by replacing \( u' \) with \( u^* \) if \( g(u^*) \leq 0 \) and by replacing \( u'' \) with \( u^* \) if \( g(u^*) > 0 \), another outer approximation of function \( \phi \) can be obtained whose minimizer is closer to the minimizer of function \( \phi \). Since function \( \phi \) has a maximum of \( N \) breakpoints, after \( N \) iterations, an approximation is obtained for which

\[
\phi_o(u^*, u', u'') = \phi(u^*)
\]

We note that in order to compute \( g(u^*) \), \( x(u^*) \) must be known. This, in turn, requires solving the following problem

\[
\max_{x \in [0, 1]^N} v(x) - u^*c^Tx.
\]

However, since the objective function of the problem (22) is supermodular, a solution for this problem can be found in polynomial time. Consequently, the Lagrangian bound can be found in polynomial time.

### 4 Characteristics of the Continuous Relaxation

After studying the Lagrangian bound, we now consider the continuous relaxation of the problem (2) and study the characteristics of the continuous bound. Let \( \nu_c^* \) denote the continuous bound for the problem (2). This bound is computed via

\[
\max_{x \in [0, 1]^N} v(x)
\]

such that \( c^Tx \leq M \).

The following result establishes a relationship between the Lagrangian bound and the continuous bound.

**Proposition 4** For the problem (2), it holds that \( \nu_c^* \leq \phi^* \), i.e. the continuous bound is tighter than the Lagrangian bound.

**Proof** We note that

\[
\nu_c^* = \max_{x \in [0, 1]^N} \{v(x) | c^Tx \leq M\}
\]

\[
\leq \min_{u \in \mathbb{R}_+} \max_{x \in [0, 1]^N} L(x, u)
\]

\[
= \min_{u \in \mathbb{R}_+} \max_{x \in [0, 1]^N} L(x, u)
\]

\[
= \phi^*.
\]

The second line in (24) is due to the fact that the dual function of the problem (23) is an upper bound for the objective function of this problem over the feasible region. The third line follows from the fact that the maximum of function \( L \) for a given \( u \) always occurs at one of the extreme points of \([0, 1]^N\). To prove this, we first note that \( \frac{\partial^2 v(x)}{\partial x_i^2} \geq 0 \) for \( i = 1, 2, \cdots, N \) which was obtained as part of the proof of Proposition 2. Hence, \( \frac{\partial v(x)}{\partial x_i} \) is a non-decreasing function. We next note that

\[
\frac{\partial L(x, u)}{\partial x_i} = \frac{\partial v(x)}{\partial x_i} - uc_i.
\]

Therefore, depending on the value of \( u \) and \( a_i \), three different patterns emerge for \( \frac{\partial L(x, u)}{\partial x_i} \) as \( x_i \) increases from
0 to 1. It can be increasing, decreasing, or decreasing first and then increasing from some point through this interval. However, it is never possible that \( \frac{\partial L(x, u)}{\partial x_i} \) increases first and then decreases. This, in turn, shows that every element of the maximizer of \( L(x, u) \) must be either 0 or 1.

In order to solve the problem (23), we note that it shares identical maximizer(s) with the following problems:

\[
\max_{x \in [0, 1]^N} \log(v(x)) \quad \text{such that} \quad c^T x \leq M \tag{23'}
\]

and

\[
\max_{x \in [0, 1]^N} v(x)^{\frac{1}{2}} \quad \text{such that} \quad c^T x \leq M. \tag{23''}
\]

This follows immediately from the fact that functions \( f(x) = x^2 \) and \( f(x) = \log(x) \) are monotone increasing over \( \mathbb{R}_+ \). However, while \( v(x) \) is not a concave function of \( x \) over \([0, 1]^N\), both \( \log(v(x)) \) and \( v(x)^{\frac{1}{2}} \) are. Therefore, the problems (23') and (23'') have a unique maximizer which can be found in polynomial time by using algorithms for constrained convex optimization [9]. Consequently, the problem (23) has a unique maximizer which can be found in polynomial time.

### 4.1 Natural Selection Process for Solving the Continuous Relaxation

We next introduce an algorithm for solving the continuous relaxation of the following variation of the problem (2):

**Definition 3** The measurement allocation problem is posed as solving:

\[
\max_{x \in \mathbb{Z}_+} \quad v(x) \quad \text{such that} \quad e^T x = M \tag{26}
\]

where \( M \in \mathbb{Z}_+ \), \( M < N \) and \( J(x) \) is given by (3).

In this variation of the problem (2), sensors are allowed to measure more than one measurement and the cost of making one measurement is identical for all sensors. By defining \( m = \frac{x}{M} \) and taking into account Proposition 1, the continuous relaxation of the problem (26) can be written as

\[
\max_{m \in \Delta^{N \times 1}} \sum_{1 \leq a_1, a_2, \ldots, a_L \leq N} w_{a_1 a_2 \ldots a_L} m_{a_1} m_{a_2} \cdots m_{a_L} \tag{27}
\]

where \( \Delta^{N \times 1} = \{ m \in \mathbb{R}_+^N | \sum_{i=1}^N m_i = 1 \} \) denotes the standard simplex in \( \mathbb{R}^N \).

The problem (27) involves the maximization of a homogenous polynomial over the standard simplex. A general homogenous polynomial may have many local maximizers over the standard simplex. In fact, finding the global maximizer of a homogenous polynomial over the standard simplex is known to be NP-Hard even for \( L = 2 \) as it can be associated with the maximum clique problem [13]. However, as we discussed earlier, the problem (27), in our particular case, has a unique maximizer. Hence, our task for finding its maximizer becomes much simpler.

We next show that the so-called natural selection process in population genetics [10] can be used to find the maximizer of the problem (27). The use of the natural selection process for finding the local maximizers of a quadratic homogenous polynomial, i.e. \( L = 2 \), over the standard simplex is noted in [14]. More generally, it can be shown that the set of local maximizers of a homogenous polynomial is equal to the set of the stable equilibria of the natural selection process (see Proposition 5 below). Thus, if a problem such as (27) has only a unique maximizer by updating an allocation vector \( m \) via the natural selection process, this maximizer can be found.

To prove the desired result, it is more convenient that we first symmetrize (hyper-)matrix \( W \). Let us denote by \( a(a_1, a_2, \ldots, a_L) \) all permutations of \( a_1, a_2, \ldots, a_L \).

**Definition 4** For (hyper-)matrix \( W \), its symmetrization is given by (hyper-)matrix \( U \) where

\[
u_{a_1 a_2 \ldots a_L} = \frac{1}{L!} \sum_{\beta \in a(a_1, a_2, \ldots, a_L)} w_{\beta} \tag{28}\]

for \( 1 \leq a_1, a_2, \ldots, a_L \leq N \).

For instance, if \( W \) is a \( N \times N \) matrix, then \( U \) is given by

\[
u_{ij} = \frac{w_{ij} + w_{ji}}{2} \tag{29}\]

and if \( W \) is a \( N \times N \times N \) hyper-matrix, then \( U \) is given by

\[
u_{ijk} = \frac{w_{ijk} + w_{ikj} + w_{jik} + w_{kij} + w_{kji} + w_{kij}}{6} \tag{30}\]

for \( 1 \leq i, j, k \leq N \).

It is easy to see that, in general, the value of a homogenous polynomial does not change if \( W \) is replaced with its symmetrized version \( U \). Furthermore, it can be shown that when the elements of \( W \) are obtained via (10), then the elements of \( U \) are non-negative. The proof for this assertion in the general case is rather cumbersome. Instead, we just verify it when \( L = 2 \) and then take another approach to ensure the non-negativity of the elements \( U \). For \( L = 2 \) and \( 1 \leq i, j \leq N \) where \( j \neq i \), we note that \( u_{ii} = w_{ii} = \text{det}(S_i) \geq 0 \). Furthermore,

\[
u_{ij} = \frac{1}{2}(w_{ij} + w_{ji}) = \text{det}(S_i + S_j) - \text{det}(S_i) - \text{det}(S_j) \geq 0 \tag{31}\]
where the latter holds because the determinant is super-additive over the set of positive semi-definite matrices. However, even if there are elements of $U$ that are negative, we note that the maximizer of problem (27) does not change if a constant is added to all elements of $U$. Hence, the non-negativity of the elements of $U$ can be ensured by adding an appropriate constant.

**Remark** Although never needed, some of the earlier results of this paper can be also established by using the same technique discussed above. In particular, we note that by replacing $w_{a_1a_2\ldots a_l}$ with $w_{a_1a_2\ldots a_l} + \gamma b_{a_1}b_{a_2}\cdots b_{a_l}$ in (9) for a given $\gamma$, the maximizer of the problem (2) does not change. Hence, it can be always ensured that the coefficients of the objective function of this problem are non-negative by adding an appropriate constant. The supermodularity of the objective function then follows from the non-negativity of the coefficients [6], without resorting to Proposition 2. Furthermore, in Proposition 4, it is no longer to show that the derivative of the objective function with respect to a decision variable is non-decreasing since this is readily known for a polynomial with non-negative coefficients.

The fact that (hyper-)matrix $U$ is symmetric and its elements are non-negative can be used to prove the following proposition.

**Proposition 5** If allocation vector $m$ is updated via the natural selection process, i.e.

$$m_i' = \frac{m_i v_i(m)}{v(m)}$$  \hspace{1cm} (32)

for $1 \leq i \leq N$ where

$$v_i(m) = \sum_{1 \leq a_1, a_2, \ldots, a_{l-1} \leq N} u_{a_1a_2\ldots a_{l-1}} m_{a_1} m_{a_2} \cdots m_{a_{l-1}}$$  \hspace{1cm} (33)

and

$$v(m) = \sum_{1 \leq a_1, a_2, \ldots, a_{L} \leq N} u_{a_1a_2\ldots a_{L}} m_{a_1} m_{a_2} \cdots m_{a_{L}}$$  \hspace{1cm} (34)

it then holds that $v(m') \geq v(m)$.

**Proof** Note that in (33), $v_i(m)$ is the partial weighted sum of $U$ when the first digit in the index is fixed to $i$. We first somehow extend this notation to address the partial weighted sums in other dimensions. Let $I$ be a fixed number between 1 and $L$. We then define

$$v_i^I(m) = \sum_{1 \leq a_1, a_2, \ldots, a_{l-1} \leq N} u_{a_1a_2\ldots a_{l-1}} m_{a_1} m_{a_2} \cdots m_{a_{l-1}}$$

position $I$

(35)

Therefore, $v_i^I(m)$ is the partial weighted sum of $U$ when the index is fixed to $i$ in dimension $I$. For instance, when $L = 2$, $v_i^1(m)$ means the partial weighted sum of the first row, $v_i^2(m)$ means the partial weighted sum of the first column and so on. Note that $v_i(m) = v_i^I(m)$.

Furthermore, since $U$ is symmetric, we have

$$v_i(m) = v_i^I(m)$$  \hspace{1cm} (36)

for $1 \leq I \leq L$ and $1 \leq i \leq N$. We next rewrite the main inequality of the proposition in a form that has already been proved in the literature. We note that

$$v(m') = \sum_{1 \leq a_1, a_2, \ldots, a_{l} \leq N} u_{a_1a_2\ldots a_{l}} m_{a_1} m_{a_2} \cdots m_{a_{l}}$$

$$= \sum_{1 \leq a_1, a_2, \ldots, a_{l} \leq N} u_{a_1a_2\ldots a_{l}} \frac{m_{a_1} v_{a_1}(m) m_{a_2} v_{a_2}(m) \cdots m_{a_{l}} v_{a_{l}}(m)}{v(m)}$$  \hspace{1cm} (37)

Therefore to prove that $v(m') \geq v(m)$, we must show

$$\sum_{1 \leq a_1, a_2, \ldots, a_{l} \leq N} u_{a_1a_2\ldots a_{l}} m_{a_1} m_{a_2} \cdots m_{a_{l}} v_{a_1}(m) v_{a_2}(m) \cdots v_{a_{l}}(m)$$

$$\geq (v(m))^{L+1}.$$  \hspace{1cm} (38)

Using (36), we can write the above as

$$\sum_{1 \leq a_1, a_2, \ldots, a_{l} \leq N} u_{a_1a_2\ldots a_{l}} m_{a_1} m_{a_2} \cdots m_{a_{l}} v_1^1(m) v_2^2(m) \cdots v_l^l(m)$$

$$\geq (v(m))^{L+1}.$$  \hspace{1cm} (39)

The above inequality has been proved in [15] for (hyper-)matrices with non-negative elements.

According to Proposition 5, when $L = 2$, the allocation vector is updated as follows:

$$m_1' = \frac{\text{sum of elements in the first row of } A}{\text{sum of all elements in } A}$$  \hspace{1cm} (40)

$$m_2' = \frac{\text{sum of elements in the second row of } A}{\text{sum of all elements in } A}$$  \hspace{1cm} (41)

and so on for $m_3'$ to $m_N'$ where $A = [w_{ij}m_{ij}]$.

To initialize the natural selection process, a reasonable choice is the centroid of $\Delta^{N-1}$, i.e. $[\frac{x_1}{L}, \ldots, \frac{x_l}{L}]^T$. This choice ensures that the algorithm is not initially biased in favor of any particular solution. The overall algorithm for solving the continuous relaxation of the measurement allocation problem is given below.

---

**Algorithm 1: The Continuous Bound for the Measurement Allocation Problem**

**Inputs:**
A. A set of sensors: \( N = \{1, 2, \ldots, N\} \).
B. (Hyper-)matrix of the sensor subset selection problem: \( W \).
C. Total number of sensors to be selected: \( M \).
D. Termination criterion: \( \epsilon \).

Output:

A. An estimate for the continuous bound of the measurement allocation problem: \( \hat{v}_c \).

Steps:

1. Let \( m = [\frac{1}{x}, \frac{1}{x}, \ldots, \frac{1}{x}]^T \).
2. Compute
   \[ m'_i = \frac{m_i v_i(m)}{v(m)} \]
   for \( 1 \leq i \leq N \) where \( v_i \) and \( v \) are given by (33) and (34), respectively.
3. If \( \frac{\|m' - m\|}{\|m\|} > \epsilon \), go to Step 2. Otherwise let \( \hat{v}_c = M^P v(m') \) and return.

Algorithm 1 typically converges in a few tens of iterations for \( \epsilon \) as low as \( 10^{-6} \) (see below for a numerical example). Each iteration of this algorithm is dominated by \( N^L \) operations because the computation of \( v_i \) in (33) requires \( N^{L-1} \) operations and \( v_i \) must be computed for \( N \) sensors. As a result, for unknown parameters with two or three components, i.e. \( L = 2 \) or 3, it only takes a few milliseconds on a modest personal computer to find the optimal allocation vector even when \( N = 100 \).

4.2 Numerical Example

To study the behavior of Algorithm 1 numerically, we study the problem (26) in the context of bearings-only localization where the position of a target must be estimated by measuring the direction-of-arrival (DOA) of a signal emitted from the target to the sensors [16, 17].

Figure 1 shows the geometry of bearings-only localization in three-dimensional space. The figure indicates that the true DOA to sensor \( i \) can be specified by two angles. If the position of the target is denoted by \( p = [x_t, y_t, z_t] \) and the position of sensor \( i \) is denoted by \( p_i = [x_i, y_i, z_i] \), these angles are related to the position of the target and the position of the sensor via

\[
\theta_i(p) = \text{atan2}(y_t - y_i, x_t - x_i) \]  \hspace{1cm} (42)
\[
\beta_i(p) = \text{atan2}(z_t - z_i, \sqrt{(y_t - y_i)^2 + (x_t - x_i)^2}) \]  \hspace{1cm} (43)

where \( \text{atan2}(y, x) \) is the four-quadrant arc tangent of \( y/x \) in the interval \([-\pi, \pi]\). As a result, we can write function \( h_i \) in the measurement model (1) as

\[
\hat{h}_i(p) = \begin{bmatrix} \theta_i(p) \\ \beta_i(p) \end{bmatrix} \]  \hspace{1cm} (44)

and note that

\[
\frac{\partial h_i}{\partial p} = \begin{bmatrix} -\frac{\sin(\theta_i)}{r_{xyi}} & \cos(\theta_i) \\ -\frac{\sin(\beta_i)\cos(\theta_i)}{r_i} & -\frac{\beta_i}{r_{xyi}} & 0 \end{bmatrix} \]  \hspace{1cm} (45)

where \( r_{xyi} = \sqrt{(y_t - y_i)^2 + (x_t - x_i)^2} \) and \( r_i = \sqrt{(y_t - y_i)^2 + (x_t - x_i)^2 + (z_t - z_i)^2} \). Consequently, \( S_i \) can be computed using (8) and \( W \) can be computed using (10).

For bearings-only localization in two-dimensional space, we note that just one angle is sufficient to specify the DOA. Hence, \( h_i(p) = \theta_i(p) \) and

\[
\frac{\partial \theta_i}{\partial p} = \begin{bmatrix} -\frac{\sin(\theta_i)}{r_{xyi}} & \cos(\theta_i) \end{bmatrix} \]  \hspace{1cm} (46)

We generated 500 networks in both two-dimensional space and three-dimensional space with 100 sensors in each network. The target was located at the origin of a Cartesian coordinate system and the sensors were spread around the target in a \( 200 \times 200 \) square or a \( 200 \times 200 \times 200 \) cube. The positions of the sensors were drawn from a uniform distribution. It was assumed that the standard deviation of the measurement error for each angle of the DOA is 10 degrees and that the measurement error for one angle is independent from the measurement error for the other angle. A histogram of the number of iterations needed for the convergence of Algorithm 1 is shown in Figure 2 when \( \epsilon \) was set to \( 10^{-6} \). On average, it took 61.2 iterations for the algorithm to converge when localization was carried out in two-dimensional space while it took 85.1 iterations for the algorithm to converge when localization was carried out in three-dimensional space.
5 Conclusions

In this paper, we analyzed the Lagrangian and continuous relaxations of the sensor subset selection problem with determinant-based functions of the FIM as the metrics of estimation accuracy. Based on these results, to obtain an upper bound for the sensor subset selection problem in polynomial time, one can solve the continuous relaxation of this problem with the log-determinant or the L-th root of the determinant as the objective function. On the other hand, when the constraint of making one measurement by each sensor is removed, an upper bound can be obtained in polynomial time by using the natural selection process with the determinant as the objective function.

References


