A Convex Optimization Approach to Worst-Case Optimal Sensor Selection

Yin Wang   Mario Sznaier   Fabrizio Dabbene

Abstract— This paper considers the problem of optimal sensor selection in a worst-case setup. Our objective is to estimate a given quantity based on noisy measurements and using no more than n sensors out of a total of N available, possibly subject to additional selection constraints. Contrary to most of the literature, we consider the case where the only information available about the noise is a deterministic set-membership description and the goal is to minimize the worst-case estimation error. While in principle this is a hard, combinatorial optimization problem, we show that tractable convex relaxations can be obtained by using recent results on polynomial optimization.

I. INTRODUCTION AND MOTIVATION

Sensor selection problems, that is selecting n out of N possible sensors in order to optimize a given criteria, are ubiquitous in areas as diverse as robotics, chemical process control, structural monitoring and target localization and tracking. A large portion of the existing literature adopts a stochastic approach, where the goal is to select a suite of sensors that minimize the covariance of the estimation error [1], or the Shannon entropy [2]. Alternative formulations focus on guaranteeing coverage of a given region or covering the maximum number of targets with the minimum number of sensors. A survey of these formulations can be found in [3].

Since the sensor selection problem is known to be generically NP-hard, several relaxations have been proposed [1], [4]–[9]. While these approaches provide efficient methods for sensor selection in the case of measurements corrupted by Gaussian noise, many practical situations do not fit this framework. Indeed, often no information is available about the distribution of the noise, beyond its support set. In these cases, a set-membership approach can be adopted, seeking to minimize either the average or worst-case estimation error over the set of estimates consistent with the available

measurements and noise description [10]. The latter is precisely the framework adopted in the present paper. Motivated by [10]–[14] we seek to select n sensors out of a total of N available in such a way that the resulting estimator is globally worst-case optimal (in the sense that it minimizes the maximum estimation error with respect to all admissible noise sequences and initial conditions) over all feasible selections. The main result of the paper is an efficient convex optimization approach to solving this problem, based on recent results on semi-algebraic and polynomial $0 - 1$ optimization. These results provide a hierarchy of increasingly tight convex relaxations that converge to the optimal solution, along with conditions that certify that this optimum has indeed been achieved. Further, in the case of ℓ^2 bounded noise, explicit bounds on the order of the convex relaxation needed to guarantee optimality are available. These results are illustrated with some academic examples comparing the performance of the resulting estimators against those obtained using the techniques in [1], [13].

II. PRELIMINARIES

A. Notation and Definitions

$\|\mathbf{x}\|_p$	ℓ^p-norm of the vector $\mathbf{x} \in \mathbf{R}^n$
	$\|\mathbf{x}\|_p^p \doteq \sum_i x_i^p,\ p < \infty$
	$\|\mathbf{x}\|_\infty \doteq \sup_i \|x_i\|$
$\|\mathbf{S}\|_{\ell^p \to \ell^q}$	ℓ^p to ℓ^q induced norm of S
	$\|\mathbf{S}\|_{\ell^p \to \ell^q} \doteq \sup_{\|\mathbf{x}\|_p=1} \|\mathbf{S}x\|_q$
$\mathbf{M} \succeq \mathbf{L}$	$\mathbf{M} - \mathbf{L}$ is positive semidefinite

B. Background on Information Based Complexity and Worst-Case Optimal Estimation

Next, we recall some key results from Information Based Complexity (IBC) [11], [13] required to establish worst-case optimality of the proposed method.

Consider three Banach spaces, X, Y, Z over the real field and two linear operators $S_y : X \to Y$ and $S_z : X \to Z$, where S_y is of the form $\mathbf{y} = \mathbf{C}\mathbf{x} + \eta$ with $\mathbf{C}$ full column rank. The only information available about the noise η is that it belongs to a bounded set $\mathcal{N}$. The

goal of IBC based worst-case estimation is to develop optimal (in a sense precisely defined below) estimators of $\mathbf{z}$ from the noisy measurements $\mathbf{y}$. Define the set

$$\mathcal{T}(\mathbf{y}) \doteq \{\mathbf{x} \in X : \mathbf{y} = \mathbf{C}\mathbf{x} + \eta \text{ for some } \eta \in \mathcal{N}\} \quad (1)$$

containing all possible elements in X that could have generated the observed data. Then for each $\mathbf{y}$, the local error $e(\mathcal{A}, \mathbf{y})$ of a given estimation algorithm $\mathcal{A}$, defined as $\hat{\mathbf{z}} = \mathcal{A}(\mathbf{y})$, is given by

$$e(\mathcal{A}, \mathbf{y}) \doteq \sup_{\mathbf{x} \in \mathcal{T}(\mathbf{y})} \|S_z \mathbf{x} - \mathcal{A}(\mathbf{y})\|. \quad (2)$$

The global error $e(\mathcal{A})$ of an algorithm $\mathcal{A}$ is defined as

$$e(\mathcal{A}) \doteq \sup_{\mathbf{y}} e(\mathcal{A}, \mathbf{y}). \quad (3)$$

An algorithm $\mathcal{A}_o$ is called globally optimal if

$$e(\mathcal{A}_o) = e_o \doteq \inf_{\mathcal{A}} e(\mathcal{A}). \quad (4)$$

Intuitively, an algorithm is globally optimal if it minimizes the worst-case (over all possible measurements) estimation error. Globally optimal algorithms have been extensively studied [11]–[14]. In particular, the main results in [13] show that in the case of spaces equipped with either ℓ^∞ or ℓ^2 norms, the optimal worst-case estimator is linear, of the form:

$$\hat{\mathbf{z}} = S_z \mathbf{C}^\dagger \mathbf{y} \quad (5)$$

where $\mathbf{C}^\dagger$ is a suitable constructed left inverse of $\mathbf{C}$[1]. Further, in the case where the noise set $\mathcal{N}$ is an ϵ ball in the ℓ^p norm and error is measured in the ℓ^q topology, then the optimal worst-case error is given by:

$$e_o = \|S_z \mathbf{C}^\dagger\|_{\ell^p \to \ell^q} \epsilon. \quad (6)$$

C. The General Problem of Moments

Given a compact semi-algebraic set $K \subset \mathbf{R}^n$ defined by d multivariate polynomial inequalities of the form $g_k(x) \geq 0$, $k = 1, \cdots, d$, and a multi-sequence of scalars $\{m_{\boldsymbol{\alpha}}\}$, indexed by a multi-index $\boldsymbol{\alpha} \in \mathbf{N}^n$, the *K-moment problem* is to determine whether there exist a probability measure μ supported on K that has $\{m_{\boldsymbol{\alpha}}\}$ as its $\boldsymbol{\alpha}^{th}$ moments. That is:

$$m_{\boldsymbol{\alpha}} = \mathbf{E}_\mu(\mathbf{x}^{\boldsymbol{\alpha}}) \doteq \int_{\mathbf{K}} \mathbf{x}^{\boldsymbol{\alpha}} \mu(dx) \quad (7)$$

where $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$. As shown in [15], existence of μ is equivalent to $\mathbf{M}(m_{\boldsymbol{\alpha}}) \succeq 0$ and $\mathbf{L}(g_k m_{\boldsymbol{\alpha}}) \succeq 0$. In the sequel, we will consider truncated version of

these matrices, containing moments of order up to $2T$. The truncated moment matrix is given by

$$\mathbf{M}_T(\mathbf{m})(i, j) = m_{\boldsymbol{\alpha}^{(i)} + \boldsymbol{\alpha}^{(j)}} \text{ for all } i, j \leq S_T \quad (8)$$

where $S_T \doteq \binom{T+n}{n}$ denotes the number of moments in $\mathbf{R}^n$ up to order T. Similarly, if $g_k(x) = \sum_{\boldsymbol{\beta}} g_{k,\boldsymbol{\beta}^{(l)}} x^{\boldsymbol{\beta}^{(l)}}$ is one of the polynomials defining the set K with degree δ_k, the corresponding truncated matrix is given by:

$$\mathbf{L}_T(g_k \mathbf{m})(i, j) = \sum_{\boldsymbol{\beta}} g_{k,\boldsymbol{\beta}^{(l)}} m_{\boldsymbol{\beta}^{(l)} + \boldsymbol{\alpha}^{(i)} + \boldsymbol{\alpha}^{(j)}}$$
$$\text{for all } i, j \leq S_{T - \frac{\delta_k}{2}} \quad (9)$$

In the context of the problems addressed in this paper, the results above allow for developing convex finite dimensional relaxations of polynomial optimization problems over semi-algebraic sets. As shown in [15], polynomial optimization problems of the form:

$$p_K^* := \min_x \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} x^{\boldsymbol{\alpha}} \text{ subject to } x \in K \quad (10)$$

can be reduced to a monotonically convergent sequence of Linear Matrix Inequalities (LMI) optimization problems of the form:

$$p_T^* = \min_{\mathbf{m}} \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} m_{\boldsymbol{\alpha}} \text{ subject to:}$$
$$\mathbf{M}_T(\mathbf{m}) \succeq 0, \quad (11)$$
$$\mathbf{L}_T(g_k \mathbf{m}) \succeq 0, k = 1, \cdots, d,$$

with $p_T^* \uparrow p_K^*$ as $T \to \infty$.

Remark 1: If for some T the sequence of relaxations satisfies rank $[\mathbf{M}_T(\mathbf{m})] = $ rank $[\mathbf{M}_{T-1}(\mathbf{m})]$ (the so called flat extension property), then the Tth order relaxation is indeed exact [16].

D. Moments-based Nonlinear 0-1 Programs

In this paper, we will reduce the sensor selection optimization program to a polynomial optimization problem in 0-1 variables of the form:

$$p^* := \min_{x \in \{0,1\}^N} p(x) \text{ subject to } x \in K \quad (12)$$

where N denotes the total number of sensors available. In this case, as shown in [17], the sequence of relaxations (11) of problem (12) is guaranteed to find the optimal solution for some finite order matrices of order $T \leq N$.

III. THE CONSTRAINED SENSOR SELECTION PROBLEM

As briefly outlined in section II-B, the results in [13] show that, in the absence of constraints, worst-case optimal estimators either entail using all available

---

[1]In the case of ℓ^2 bounded noise, $\mathbf{C}^\dagger$ is the Moore–Penrose pseudo-inverse, leading to the usual least square estimator.

measurements (when using the ℓ^2 metric), or a subset of these with cardinality n_x (the dimension of X), when considering the ℓ^∞ topology. In this case, the specific optimal subset can be found by solving $2n_z$ linear programming problems, where n_z denotes the dimension of the quantity to be estimated. However, in many cases of practical interest, sensor choices are subject to additional selection rules. For example, transmission medium access constraints or power considerations may preclude certain sensors from being simultaneously used or limit the number of sensors that can chosen in some given subsets. In this context, the goal of this paper can be simply stated as deciding a strategy for selecting n sensors such that the estimator $\hat{z} = S_z \mathbf{C}_{\text{selected}}^\dagger \mathbf{y}_{\text{selected}}$ is globally (worst-case) optimal, under given selection constraints. Formally, the problem can be stated as:

$$\min_{\mathbf{C}_s \text{ admissible}} \|S_z \mathbf{C}_s^\dagger\| \tag{13}$$

where $\mathbf{C}_s$, $s = 1, 2, \cdots, n_s \le \binom{N}{n}$ and $\|.\|$ denote all possible admissible choices of n sensors from the N available, and a suitable norm, respectively. In principle, this leads to a hard, combinatorial optimization problem. However, as we show in the sequel, it can be reduced to a 0-1 polynomial optimization, which in turn can be relaxed to a convex optimization by appealing to moments concepts.

For simplicity, in the sequel we will make the assumption that z, the quantity to be estimated, is a scalar. Note that in the ℓ^∞ case this assumption can be made without loss of generality, since vector optimal estimators simply consist of the optimal scalar estimator $\hat{z}_i$ of each component of $\mathbf{z}$ [13]. In the case of ℓ^2 norms, this assumption is equivalent to designing independent estimators (and hence potentially using different set of sensors) for each component, rather than a single (coupled) estimator.

A. ℓ^2 bounded noise

Consider first the case were the measurement noise is characterized in terms of the ℓ^2 norm. In this case, for a given choice s of sensors, $\mathbf{C}^\dagger$ in (6) is given by $\mathbf{C}^\dagger \doteq \left(\mathbf{C}_s^T \mathbf{C}_s\right)^{-1} \mathbf{C}_s^T$, where $\mathbf{C}_s$ denotes the submatrix of $\mathbf{C}$ corresponding to the chosen sensors. Thus, by adding binary variables $v_i \in \{0, 1\}$, $i = 1, 2, \cdots, N$ and forming the sensor matrix $\mathbf{C}_v \doteq [v_1 \mathbf{c}_1^T \ v_2 \mathbf{c}_2^T \cdots v_N \mathbf{c}_N^T]^T$, where $\mathbf{c}_i$ denote the rows of $\mathbf{C}$, problem (13) can be rewritten in the following form:

$$\min_{v \in \mathcal{S}_v} \ \|\mathbf{S}_z(\mathbf{C}_v^T \mathbf{C}_v)^{-1} \mathbf{C}_v^T\|_2 \text{ subject to:}$$
$$\mathbf{1}^T \mathbf{v} = n \tag{14}$$
$$v_i \in \{0, 1\}, \quad i = 1, \cdots, N$$

where $\mathcal{S}_v$ is the set of admissible selections and $\mathbf{1}$ denotes a vector with all entries one. As we show next, the problem above is equivalent to a polynomial minimization over a semi-algebraic set. To this effect, note that the objective in (14) can be rewritten as:

$$\|\mathbf{S}_z(\mathbf{C}_v^T \mathbf{C}_v)^{-1} \mathbf{C}_v^T\|_2^2 = \|\mathbf{S}_z \frac{adj(\mathbf{C}_v^T \mathbf{C}_v)}{det(\mathbf{C}_v^T \mathbf{C}_v)} \mathbf{C}_v^T\|_2^2$$
$$= \frac{\|\mathbf{P}_v\|_2^2}{d^2(\mathbf{v})} = \frac{p^2(\mathbf{v})}{d^2(\mathbf{v})} \tag{15}$$

where $adj(\mathbf{A})$ and $det(\mathbf{A})$ denote the adjugate matrix and determinant of a given matrix $\mathbf{A}$, respectively. Here $\mathbf{P}_v = [p_1(\mathbf{v}), p_2(\mathbf{v}), \cdots, p_N(\mathbf{v})]$ is a vector whose entries are polynomials in the variables $v_1, v_2, \cdots, v_N$. Since $d(\mathbf{v}) = det(\mathbf{C}_v^T \mathbf{C}_v)$ is also polynomial, it follows that, by introducing an additional variable β, problem (14) can be recast as the following (constrained) polynomial optimization problem:

$$\min \beta \text{ subject to:}$$
$$p^2(\mathbf{v}) - \beta d^2(\mathbf{v}) = 0$$
$$\mathbf{1}^T \mathbf{v} = n \tag{16}$$
$$v_i \in \mathcal{S}_v \text{ and } v_i^2 - v_i = 0, \quad i = 1, \cdots, N$$

where the last equality enforces the condition $v_i \in \{0, 1\}$. In the case where $\mathcal{S}_v$ is semi-algebraic (e.g. defined by a set of polynomial constraints) then problem (16) can be solved using the moments-based method outlined in section II-C. In particular, simple constraints of the form:
(i) no more than n_i sensors can be simultaneously chosen from a given set $\mathcal{S}_i$.
(ii) sensors v_i and v_j cannot be simultaneously chosen.
can be directly incorporated when forming the moments matrix $\mathbf{M}(\mathbf{m})$, eliminating the need for considering the moments localizing matrix $\mathbf{L}(\mathbf{m})$ in the optimization problem (11). In the first case, this is accomplished by setting all entries of $\mathbf{M}$ of the form , $m_{\boldsymbol{\alpha}} = \mathbf{E}_\mu(v_1^{\alpha_1} v_2^{\alpha_2} \cdots v_n^{\alpha_n}) = 0$ whenever $\sum_{i \in \mathcal{S}_i} \alpha_i > n_i$. Similarly, pairwise sensor exclusion constraints reduce to setting $m_{\boldsymbol{\alpha}} = 0$ whenever $\alpha_i + \alpha_j > 1$.

Note that since (16) involves polynomials of degree $2n + 1$ in $N + 1$ variables, this method requires considering moment matrices of size at least $n_m \times n_m$ where $n_m = \binom{n+N+2}{N+1}$. Thus, a more efficient implementation can be obtained by simply performing a line search on β, leading to the following algorithm:

3

Algorithm I: ℓ^2 norm based sensor selection

1.- Initialize β_0 equal to a positive scalar.
2.- Repeat

    (a) Solve

$$\mathbf{v}_k = \underset{\mathbf{v}}{\operatorname{argmin}} p^2(\mathbf{v}) - \beta_k d^2(\mathbf{v})$$
$$s.t.$$
$$\mathbf{1}^T\mathbf{v} = n$$
$$v_i \in \mathcal{S}_v \text{ and } v_i^2 = v_i, \quad i = 1, \cdots, N$$

    (b) Update $\beta_{k+1} = \frac{p^2(\mathbf{v}_k)}{d^2(\mathbf{v}_k)}$.

    Until β_{k+1} converges.

Remark 2: Note that each iteration of Algorithm I involves only polynomials of degree 2n in the variables $v_i \in \{0,1\}$. Thus, as outlined in Section II-D only moments of order up to $2n$ need to be considered.

Theorem 1: The sequence β_k in Algorithm I converges to the solution of Problem (16).

    Proof: To establish this result, we will show first that the sequence β_k is non-increasing. To this effect, let $J(\mathbf{v}, \beta) \doteq p^2(\mathbf{v}) - \beta d^2(\mathbf{v})$. Since $\mathbf{v}_k$ is a feasible solution for the optimization at step $k+1$, it follows that $J(\mathbf{v}_{k+1}, \beta_{k+1}) \leq J(\mathbf{v}_k, \beta_{k+1}) = 0$ (by definition). Thus $p^2(\mathbf{v}_{k+1}) - \beta_{k+1} d^2(\mathbf{v}_{k+1}) \leq 0$, or, equivalently, $\beta_{k+2} \doteq \frac{p^2(\mathbf{v}_{k+1})}{d^2(\mathbf{v}_{k+1})} \leq \beta_{k+1}$. Since β_k is a non-increasing sequence, bounded below by zero, it follows that it converges from above to some limit $\tilde{\beta}$. Similarly, since $\mathbf{v}_k$ is bounded, it has a convergent subsequence $\mathbf{v}_{k_i} \to \tilde{\mathbf{v}}$. By construction $p^2(\tilde{\mathbf{v}}^*) - \tilde{\beta} d^2(\tilde{\mathbf{v}}^*) = 0$. Hence, for any $\epsilon > 0$, there exists some K such that:

$$p^2(\mathbf{v}_{k+1}) - \beta_k d^2(\mathbf{v}_{k+1}) \geq -\epsilon \,\forall\, k \geq K \qquad (17)$$

Denote by β^* and $\mathbf{v}^*$ the solution to (16). Then

$$p^2(\mathbf{v}_{k+1}) - \beta_k d^2(\mathbf{v}_{k+1}) \leq p^2(\mathbf{v}^*) - \beta_k d^2(\mathbf{v}^*)$$
$$\leq p^2(\mathbf{v}^*) - \tilde{\beta} d^2(\mathbf{v}^*) = -d^2(\mathbf{v}^*)(\tilde{\beta} - \beta^*) \qquad (18)$$

If $\tilde{\beta} > \beta^*$, then choosing $\epsilon < d^2(\mathbf{v}^*)(\tilde{\beta} - \beta^*)$ leads to a contradiction between (17) and (18). $\qquad\square$

B. ℓ^∞ bounded noise

In this case worst-case optimal estimators have the form $\hat{z} = \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{y}}$, where $\tilde{\mathbf{C}}$ is a full rank matrix formed from the rows of $\mathbf{C}$ and $\tilde{\mathbf{y}}$ denotes the subset of the available measurements $\mathbf{y}$ corresponding to the chosen rows (see [13] for details). The corresponding worst-case estimation error is given by

$$e_o \doteq \|\mathbf{S}_z \tilde{\mathbf{C}}^{-1}\|_{\ell^\infty \to \ell^\infty} \qquad (19)$$

Proceeding as in Section III-A, by introducing binary variables v_i, the optimal cost (19) can be rewritten as

$$e_o \doteq \|\mathbf{S}_z(\mathbf{C}_v^T\mathbf{C}_v)^{-1}\mathbf{C}_v^T\|_{\ell^\infty \to \ell^\infty}$$
$$\text{subject to: } \mathbf{1}^T\mathbf{v} = n \qquad (20)$$

where the constraint enforces the condition that exactly n rows of $\mathbf{C}_v$ must be selected. Thus, in this case the constrained sensor selection problem reduces to the following semi-algebraic optimization.

$$\min \quad \frac{\sum_i |p_i(\mathbf{v})|}{d(\mathbf{v})}$$
$$s.t.$$
$$\mathbf{1}^T\mathbf{v} = n$$
$$v_i \in \{0,1\}, \quad i = 1, \cdots, N$$
$$v_i \in \mathcal{S}_v$$

In turn, introducing the additional variable β leads to:

$$\min_{\{\beta, \mathbf{x}_+, \mathbf{x}_-, \mathbf{v}\}} \quad \beta$$
$$s.t.$$
$$\mathbf{1}^T\mathbf{v} = n$$
$$\mathbf{x}_+ \geq 0$$
$$\mathbf{x}_- \geq 0$$
$$x_+(i) - x_-(i) = p_i(\mathbf{v})$$
$$\sum_i [x_+(i) + x_-(i)] = \beta d(\mathbf{v})$$
$$v_i \in \{0,1\}, \quad i = 1, \cdots, N$$
$$v_i \in \mathcal{S}_v$$

$$(21)$$

which can be reduced to a sequence of LMI optimization problems by using a moments–based relaxation. It is worth emphasizing that in this case, due to the additional real variables $\mathbf{x}_+, \mathbf{x}_-$ and β, the results in section II-D no longer apply. Nevertheless, the results in II-C show that the sequence of LMI problems converge (monotonically) to the optimal solution and provide optimality certificates for finite order relaxations (the flat extension property).

Bound on the number of times that sensor 1 can be active	3	2	1
proposed method (time instant, sensor)	$\{(1,1)\ (2,1)$ $(3,1)\}$	$\{\ (1,1)\ (2,1)$ $(3,2\}$	$\{(1,1)\ (2,2),$ $(3,3)\}$
method in [13]	All	NA	NA
max-det criterion used in [1]	$\{(1,1)\ (2,1)$ $(3,1)\ \}$	NA	NA
random selection	$\{(1,1)\ (2,2)$ $(3,1)\ \}$	$\{(1,2)\ (2,1)$ $(3,1)\ \}$	$\{(1,1)\ (2,2)$ $(3,2)\ \}$

TABLE I

COMPARISON OF SENSOR CHOICES FOR DIFFERENT ALGORITHMS

4
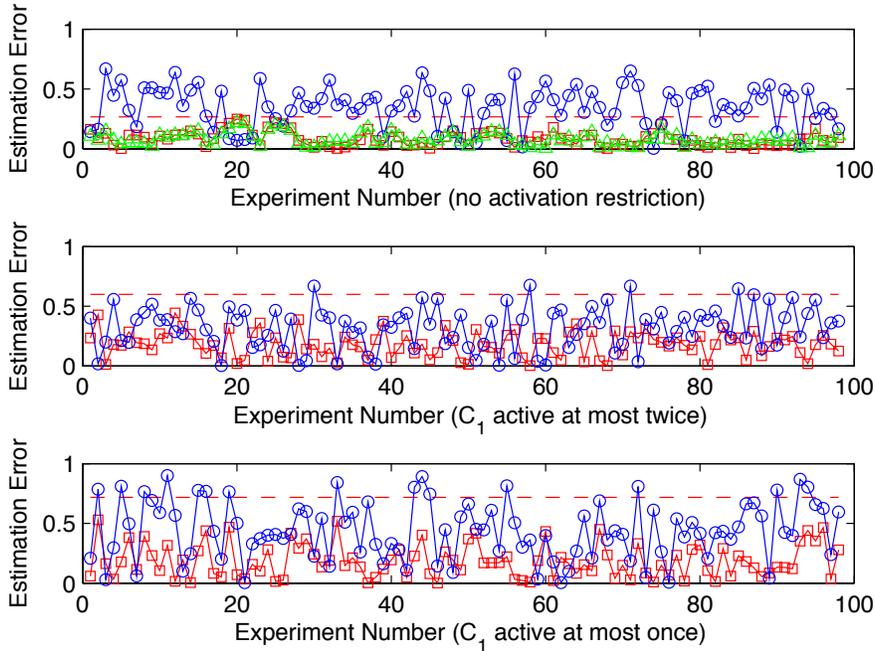
Fig. 1. Comparison of the performance of the sensor selections strategies listed in Table 1, corresponding to different activation constraints. In all cases the dashed line indicates the (optimal) worst-case bound. Top: No constraints; red squares (proposed method): Mean=-5.00e-3, Std=1.02e-1; blue circles (random selection): Mean=6.18e-2, Std=3.86e-1; Green triangles [13]: Mean=-8.38e-3, Std=9.43e-2. Center: C_1 active at most twice; Red squares (proposed method): Mean=1.34e-2, Std=2.10e-1; Blue circles (random selection): Mean=2.60e-2, Std=3.68e-1. Bottom: C_1 active at most once; Red squares line (proposed method): Mean=4.43e-2, Std=2.28e-1; Blue circles line (random selection): Mean=1.12e-1, Std=4.93e-1.

IV. EXPERIMENTS

In this section we illustrate our results with two academic examples.

A. Example1

Consider the linear system:

$$x(k+1) = \begin{bmatrix} -0.7321 & 0.7321 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x(k)$$

$$y(k) = Cx(k) + \eta(k)$$

$$C = \begin{bmatrix} -4.1466 & -2.7676 & -3.8185 \\ 0.5857 & -0.7998 & -0.1319 \\ 0.1676 & 0.9364 & -0.3085 \end{bmatrix}$$

where $\|\eta\|_2 \leq 1$. Assume that the sensor $C(1,:)$, denoted as C_1, has high energy requirements and therefore it is of interest to limit its use. The goal is to estimate, at each instant $x(k)$ using the measurements $y(k+2)$, $y(k+1)$, $y(k)$. Table I summarizes the sensor choices obtained using the proposed algorithm, the one in [13] and the one in [1]. Figure 1 shows the results of 100

experiments with random initial conditions corresponding to these sensor selections. As illustrated there, as the constraints on usage of sensor 1 tighten, the estimation error increases. In all cases, the estimation error for the proposed method remains below the worst case bound (dashed line) while the error corresponding to a random sensor selection exceeds it.

B. Example2

Consider now the case of up to 9 measurements of a vector $\mathbf{x} \in R^3$ given by $y = Cx + \eta$, where $\|\eta\|_2 \leq 1$ and

$$C = \begin{bmatrix} -4.0988 & 2.2718 & 2.1652 \\ -4.1893 & 2.3651 & -4.1625 \\ 3.9169 & 3.8809 & 0.6934 \\ 3.4704 & -3.7338 & -2.9529 \\ -4.5053 & 1.9238 & -2.0671 \\ 2.4785 & 0.3474 & -1.7208 \\ 4.2102 & 3.2847 & 0.2299 \\ -3.7732 & -4.0766 & 1.7296 \\ 2.1043 & 4.8087 & 4.6235 \end{bmatrix}$$
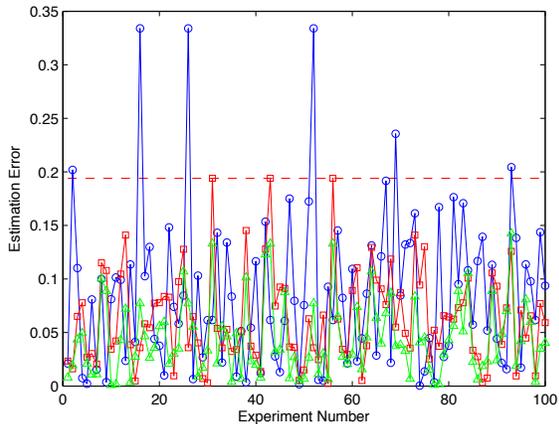
Fig. 2. Estimation error for different sensor selections in Example 2. Red squares (proposed method): Mean=-6.40e-3, Std=7.67e-2; Blue circles (max-det selection): Mean=1.07e-2, Std=1.12e-1; Green triangles (All sensors): Mean=-5.56e-3, Std=5.63e-2.

The goal is to select 3 sensors out of the available 9 to estimate $z \doteq x_1 + x_2 + x_3$. In this case the optimal choice according to the max-det criterion [1] is sensors #2, #4 and #8, while the proposed method selects #2, #3 and #9. Figure 2 shows the estimation error corresponding to these choices for 100 random experiments. Note that in some cases the max-det selection error (blue circles) indeed exceeds the optimal worst-case bound achieved by the proposed method (red squares). As before, for comparison purposes we also show (green triangles) the error achieved by a worst-case optimal estimator that uses all the sensors available, illustrating the additional cost incurred when using only a subset of these.

V. CONCLUSIONS

This paper considers a constrained sensor selection problem in a deterministic, set-membership setting. Such situations arise in cases where no information is available about the distribution of the measurement noise, other than its support set. Our main result shows that in this case worst-case optimal selections can be found by solving a sequence of convex optimization problems. Moreover, in the case of ℓ^2 bounded noise, a priori bounds on the size of the relaxation can be obtained by exploiting recent results on 0-1 polynomial optimization. These results were illustrated using simple examples comparing the performance of the resulting suit of sensors against those obtained using currently existing methods. Future research seeks to extend the results presented here to a probabilistic setup, where risk-adjusted rather than worst-case optimality is considered.

REFERENCES

[1] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 451–462, 2009.

[2] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 22, p. 1650, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007.

[3] H. Rowaihy, S. Eswaran, M. Johnson, D. Verma, A. Bar-Noy, T. Brown, and T. La Porta, "A survey of sensor selection schemes in wireless sensor networks," in *Proc. of SPIE*, vol. 6562, pp. 65621A1–13, 2007.

[4] F. Altenbach, S. Corroy, G. Böcherer, and R. Mathar, "Strategies for distributed sensor selection using convex optimization," *arXiv preprint arXiv:1208.1111*, 2012.

[5] Y. Mo, R. Ambrosino, and B. Sinopoli, "A convex optimization approach of multi-step sensor selection under correlated noise," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pp. 186–193, IEEE, 2009.

[6] V. Kekatos and G. Giannakis, "Selecting reliable sensors via convex optimization," in *Proc. IEEE Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2010.

[7] A. Carmi, "Sensor scheduling via compressed sensing," in *Information Fusion (FUSION), 2010 13th Conference on*, pp. 1–8, IEEE, 2010.

[8] 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.

[9] A. Krause and C. Guestrin, "Near-optimal nonmyopic value of information in graphical models," *arXiv preprint arXiv:1207.1394*, 2012.

[10] J. Traub, G. Wasilkowski, and H. Wozniakowski, *Information-Based Complexity*. Academic Press, New York, 1988.

[11] C. A. Micchelli and T. J. Rivlin, *Optimal Estimation in Approximation Theory*, ch. A survey of optimal recovery, pp. 1–54. Plenum Press, 1977.

[12] M. Milanese and R. Tempo, "Optimal algorithms theory for robust estimation and prediction," *IEEE Transactions on Automatic Control*, vol. 30, pp. 730–738, 1985.

[13] R. Tempo, "Robust estimation and filtering in the presence of bounded noise," *Automatic Control, IEEE Transactions on*, vol. 33, pp. 864 –867, sep 1988.

[14] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: an overview.," *Automatica*, vol. 27, pp. 997–1009, 1991.

[15] J. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. Optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[16] R. E. Curto and L. A. Fialkow, "Solution of the truncated complex moment problem for flat data," *Memoirs of the American Mathematical Society*, vol. 568, no. 119, 1996.

[17] J. B. Lasserre, "An explicit exact sdp relaxation for nonlinear 0-1 programs," in *Proceedings of the 8th International IPCO Conference on Integer Programming and Combinatorial Optimization*, (London, UK, UK), pp. 293–303, Springer-Verlag, 2001.