Sensor Selection via Convex Optimization

Siddharth Joshi and Stephen Boyd, *Fellow, IEEE*

Abstract—We consider the problem of choosing a set of k sensor measurements, from a set of m possible or potential sensor measurements, that minimizes the error in estimating some parameters. Solving this problem by evaluating the performance for each of the $\binom{m}{k}$ possible choices of sensor measurements is not practical unless m and k are small. In this paper, we describe a heuristic, based on convex optimization, for approximately solving this problem. Our heuristic gives a subset selection as well as a bound on the best performance that can be achieved by any selection of k sensor measurements. There is no guarantee that the gap between the performance of the chosen subset and the performance bound is always small; but numerical experiments suggest that the gap is small in many cases. Our heuristic method requires on the order of m^3 operations; for $m = 1000$ possible sensors, we can carry out sensor selection in a few seconds on a 2-GHz personal computer.

Index Terms—Convex optimization, experiment design, sensor selection.

I. INTRODUCTION

WE study the problem of selecting k sensors, from among m potential sensors. Each sensor gives a linear function of a parameter vector x, plus an additive noise; we assume these measurement noises are independent identically distributed zero-mean Gaussian random variables. The sensor selection, i.e., the choice of the subset of k sensors to use, affects the estimation error covariance matrix. Our goal is to choose the sensor selection to minimize the determinant of the estimation error covariance matrix, which is equivalent to minimizing the volume of the associated confidence ellipsoid. One simple method for solving the sensor selection problem is to evaluate the performance for all $\binom{m}{k}$ choices for the sensor selection, but evidently this is not practical unless m or k is very small. For example, with $m = 100$ potential sensors, from which we are to choose $k = 25$, there are on the order of 10^{23} possible choices, so direct enumeration is clearly not possible.

In this paper we describe a new method for approximately solving the sensor selection problem. Our method is based on convex optimization, and is therefore tractable, with computational complexity growing as m^3. For $m = 1000$, the method can be carried out in a few seconds, on a 2-GHz personal computer; for $m = 100$, the method can be carried out in milliseconds. The method provides both a suboptimal choice of sensors,

and a bound on the performance that can be achieved over all possible choices. Thus, we get a suboptimal design, and a bound on how suboptimal it is. Numerous numerical experiments suggest that the gap between these two is often small. Our basic method can be followed by any local optimization method. We have found, for example, that a greedy algorithm that considers all possible swaps between the set of selected and unselected sensors, accepting any swaps that improve the objective, can give a modest improvement in the quality of the sensor selection. When this local search terminates it gives a *2-opt* sensor selection, i.e., one for which no swap of a selected and an unselected sensor has better objective value.

1) Prior and Related Work: The sensor selection problem arises in various applications, including robotics [2], sensor placement for structures [3], [4], target tracking [5], [6], chemical plant control [7], and wireless networks [8]. Sensor selection in the context of dynamical systems is studied in, e.g., [9]–[11]. Sensor selection, with a rather different setup from ours, has been studied in sensor network management [12], hypothesis testing in a sensor network [13], and discrete-event systems [14]. The sensor selection problem formulation we use in this paper can be found in, e.g., [15]. The sensor selection problem (and various extensions described in Section V) can be formulated in an information theoretic framework [16]–[19], and in a Bayesian framework [20], [21]. (We will comment on this in more detail later.)

The complexity of a sensor selection problem (though not the one we consider) is considered in [22], where the authors show that it is NP-hard. (As far as we know, NP-hardness of the sensor selection problem we consider has not been established.) The sensor selection problem can be exactly solved using global optimization techniques, such as branch and bound [23], [24]. These methods can, and often do, run for very long times, even with modest values of k and m.

Several heuristics have been proposed to approximately solve the sensor selection problem. These include genetic algorithms [15], and application specific local search methods. Local optimization techniques, similar to the one we describe, are summarized in [25] and [26]. While these heuristics can produce good suboptimal sensor selections, they do not yield any guarantees or bounds on the performance that is achievable. In any case, any local optimization method, including the ones described in these papers, and generic methods such as randomized rounding [27], can be incorporated into our method.

The sensor selection problem is closely related to the D-optimal experiment design problem [28], [29]. Here, too, we are to choose a subset of possible measurements from a palette of choices. In D-optimal experiment design, however, we consider the case when k and m both grow, with a constant ratio; the question is not which sensors to use, but how frequently to use

each one. The standard convex relaxation for the D-optimal experiment design problem (see, e.g., [30, Sec. 7.5]) leads to a convex problem that is similar to ours, but different; we will discuss the differences in more detail below.

Finally, we note that the idea of using convex relaxation as the basis for a heuristic for solving a combinatorial problem is quite old, and has been observed to give very good results in many applications. Recent problems that are solved using this general technique are compressed sensing [31], sparse regressor selection [32], sparse signal detection [33], sparse decoding [34], and many others. Other applications that use convex relaxations include portfolio optimization with transaction costs [35], controller design [36], and circuit design [37].

2) Outline: The rest of this paper is organized as follows. In Section II, we formally describe the sensor selection problem. In Section III, we describe the basic convex relaxation, an approximate relaxation that can be solved even more efficiently, and a local optimization method to improve the basic sensor selection. We illustrate the method, with and without local optimization, with a numerical example, in Section IV. In Section V, we describe a number of variations and extensions on the sensor selection problem, than can be incorporated in the convex optimization framework, including different objective functions, map estimation, constraints on sensors, and a robust version of the sensor selection problem.

II. SENSOR SELECTION

A. Parameter Estimation

Suppose we are to estimate a vector $x \in \mathbf{R}^n$ from m linear measurements, corrupted by additive noise,

$$y_i = a_i^T x + v_i, \quad i = 1, \ldots, m \tag{1}$$

where $x \in \mathbf{R}^n$ is a vector of parameters to estimate, and $v_1, \ldots, v_m$ are independent identically distributed $\mathcal{N}(0, \sigma^2)$ random variables. We assume that $a_1, \ldots, a_m$, which characterize the measurements, span $\mathbf{R}^n$. The maximum-likelihood estimate of x is

$$\hat{x} = \left(\sum_{i=1}^m a_i a_i^T \right)^{-1} \sum_{i=1}^m y_i a_i. \tag{2}$$

The estimation error $x - \hat{x}$ has zero mean and covariance

$$\Sigma = \sigma^2 \left(\sum_{i=1}^m a_i a_i^T \right)^{-1}.$$

The η-*confidence ellipsoid* for $x - \hat{x}$, which is the minimum volume ellipsoid that contains $x - \hat{x}$ with probability η, is given by

$$\mathcal{E}_\alpha = \{ z \mid z^T \Sigma^{-1} z \leq \alpha \} \tag{3}$$

where $\alpha = F_{\chi_n^2}^{-1}(\eta)$. ($F_{\chi_n^2}$ is the cumulative distribution function of a χ-squared random variable with n degrees of freedom.)

A scalar measure of the quality of estimation is the volume of the η-confidence ellipsoid

$$\mathbf{vol}(\mathcal{E}_\alpha) = \frac{(\alpha\pi)^{n/2}}{\Gamma\left(\frac{n}{2+1}\right)} \det \Sigma^{1/2} \tag{4}$$

where Γ is the Gamma function. Another scalar measure of uncertainty, that has the same units as the entries in the parameter x, is the *mean radius*, defined as the geometric mean of the lengths of the semi-axes of the η-confidence ellipsoid

$$\rho(\mathcal{E}_\alpha) = \sqrt{\alpha}(\det \Sigma)^{1/2n}. \tag{5}$$

We will be interested in volume ratios, so it is convenient to work with the log of the volume

$$\log \mathbf{vol}(\mathcal{E}_\alpha) = \beta - \left(\frac{1}{2}\right) \log \det \left(\sum_{i=1}^m a_i a_i^T \right) \tag{6}$$

where β is a constant that depends only on σ, n, and η. The log volume of the confidence ellipsoid, given in (6), gives a quantitative measure of how informative the collection of m measurements is.

B. Sensor Selection Problem

Now we can describe the *sensor selection problem*. We consider a set of m potential measurements, characterized by $a_1, \ldots, a_m \in \mathbf{R}^n$; we are to choose a subset of $k \ (\geq n)$ of them that minimizes the log volume (or mean radius) of the resulting confidence ellipsoid. This can be expressed as the optimization problem

$$\begin{aligned} \text{maximize} \quad & \log \det \left(\sum_{i \in \mathcal{S}} a_i a_i^T \right) \\ \text{subject to} \quad & |\mathcal{S}| = k \end{aligned} \tag{7}$$

where $S \subseteq \{1, \ldots, m\}$ is the optimization variable, and $|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$. (We interpret $\log \det \left(\sum_{i \in \mathcal{S}} a_i a_i^T \right)$ as $-\infty$ if $\sum_{i \in \mathcal{S}} a_i a_i^T$ is singular.) We let $p^\star$ denote the optimal value of the sensor selection problem.

We can rewrite the problem (7) as

$$\begin{aligned} \text{maximize} \quad & \log \det \left(\sum_{i=1}^m z_i a_i a_i^T \right) \\ \text{subject to} \quad & \mathbf{1}^T z = k \\ & z_i \in \{0, 1\}, \quad i = 1, \ldots, m \end{aligned} \tag{8}$$

with variable $z \in \mathbf{R}^m$. (The vector $\mathbf{1}$ is the vector with all entries one.) Here z_i encodes whether the ith measurement (or sensor) is to be used. This problem is a Boolean-convex problem, since the objective is a concave function of z for $z_i \geq 0$ (see, e.g., [30, Sec. 3.1.5]), the sum constraint is linear, and the last m constraints restrict z to be Boolean (i.e., 0–1).

III. CONVEX RELAXATION

A. The Relaxed Sensor Selection Problem

By replacing the nonconvex constraints $z_i \in \{0,1\}$ with the convex constraints $z_i \in [0,1]$, we obtain the *convex relaxation* of the sensor selection problem (7):

$$\begin{aligned} \text{maximize} \quad & \log \det \left(\sum_{i=1}^{m} z_i a_i a_i^T \right) \\ \text{subject to} \quad & \mathbf{1}^T z = k \\ & 0 \le z_i \le 1, \quad i = 1, \ldots, m \end{aligned} \tag{9}$$

where $z \in \mathbf{R}^m$ is the variable. This problem, unlike the original sensor selection problem (7), is convex, since the objective (to be maximized) is concave, and the equality and inequality constraints on z are linear. It can be solved efficiently, for example, using interior-point methods [30]. These methods typically require a few tens of iterations; each iteration can be carried out (as we will see below) with a complexity of $O(m^3)$ operations, so the overall complexity is $O(m^3)$ operations. We will let $z^\star$ denote a solution of the relaxed problem (9).

The relaxed sensor selection problem (9) is not equivalent to the original sensor selection problem (7); in particular, $z_i^\star$ can be fractional. We can say, however, that the optimal objective value of the relaxed sensor selection problem (9), which we denote U, is an *upper bound* on $p^\star$, the optimal objective value of the sensor selection problem (8). To see this, we note that the feasible set for the relaxed problem contains the feasible set for the original problem; therefore, its optimal value cannot be smaller than that of the original problem.

We can also use the solution $z^\star$ of the relaxed problem (9) to generate a suboptimal subset selection $\hat{S}$. There are many ways to do this; but we describe here the simplest possible method. Let $z_{i_1}^\star, \ldots, z_{i_m}^\star$ denote the elements of $z^\star$ rearranged in descending order. (Ties can be broken arbitrarily.) Our selection is then

$$\hat{S} = \{i_1, \ldots, i_k\}$$

i.e., the indexes corresponding to the k largest elements of $z^\star$. We let $\hat{z}$ be the corresponding 0–1 vector. The point $\hat{z}$ is feasible for the sensor selection problem (8); the associated objective value

$$L = \log \det \left(\sum_{i=1}^{m} \hat{z}_i a_i a_i^T \right)$$

is then a *lower bound* on $p^\star$, the optimal value of the sensor selection problem (8).

The difference between the upper and lower bounds on $p^\star$,

$$\begin{aligned} \delta &= U - L \\ &= \log \det \left(\sum_{i=1}^{m} z_i^\star a_i a_i^T \right) - \log \det \left(\sum_{i=1}^{m} \hat{z}_i a_i a_i^T \right) \end{aligned}$$

is called the *gap*. The gap is always nonnegative; if it is zero, then $\hat{z}$ is actually optimal for the sensor selection problem (8); more generally, we can say that the subset selection $\hat{z}$ is no more than δ-suboptimal.

We can relate the gap δ, which is a difference of log-determinants, to geometric properties of confidence ellipsoids. A gap of δ corresponds to a ratio of $\exp(\delta/2)$ in confidence ellipsoid volume. In terms of the mean radius ρ, a gap of δ corresponds to a ratio $\exp(\delta/2n)$.

Not much can be said about the gap, in general; for example, there are no generic useful bounds on how large it can be. The gap is, however, very useful when evaluated for a given problem instance.

B. Relation to D-Optimal Experiment Design

Our sensor selection problem, and relaxed sensor selection problem, are closely related to D-optimal experiment design. In D-optimal experiment design, we have a set of potential measurements or sensors. In this case, however, we can use any one sensor multiple times; the problem is to choose which sensors to use, and for each one, how many times to use it, while keeping the total number of uses less than or equal to k. In contrast, in our sensor selection problem, we can use each potential sensor at most once. One method for approximately solving the D-optimal experiment design problem is to form a convex relaxation, that is the very similar to ours; however, the upper bound constraints $z_i \le 1$ are not present, and the relaxed variables are normalized to have sum one (and not k); see, e.g., [30, Sec. 7.5]. The variables in the relaxed D-optimal experiment design problem also have a different interpretation: z_i is the frequency with which sensor i is to be used, when a large number of measurements is made.

C. The Dual Problem

In this section we describe a dual for the relaxed sensor selection problem, which has an interesting interpretation in terms of covering ellipsoids. The dual of the relaxed sensor selection problem is

$$\begin{aligned} \text{minimize} \quad & \log \det \Lambda^{-1} - n + k\nu + \mathbf{1}^T \mu \\ \text{subject to} \quad & \nu + \mu_i \ge a_i^T \Lambda a_i, \quad i = 1, \ldots, m \\ & \mu_i \ge 0, \quad i = 1, \ldots, m, \end{aligned} \tag{10}$$

with variables $\Lambda \in \mathbf{S}^n$, $\mu \in \mathbf{R}^n$, and $\nu \in \mathbf{R}$. (The set of $n \times n$ symmetric matrices is denoted by $\mathbf{S}^n$.) See the Appendix for the derivation.

This dual problem can be interpreted as the problem of finding the minimum volume covering ellipsoid with outlier detection; see [38], [39]. (This should not be surprising because the dual of the D-optimal experiment design problem is the minimum volume covering ellipsoid problem [30, 7.5.3].)

If μ_i are set to 0, the optimal solution ν^* and Λ^* determine the minimum volume ellipsoid, given by $\{x \mid x^T \Lambda^* x \le \nu^*\}$, that contains the m points $a_1, \ldots, a_m$. When the variable μ_i is positive, a_i is allowed to be outside this ellipsoid, i.e., it is an outlier. We now show that, at optimality, at most k of the μ_i are nonzero. This can be inferred in many ways. Let $\Lambda^\star$, $\mu^\star$, and $\nu^\star$ be an optimal solution of the problem (10). The dual variables μ_i are associated with the inequalities $z_i \le 1$ and by complementary slackness we have

$$\mu_i^\star (1 - z_1^\star) = 0, \quad i = 1, \ldots, m.$$

Since all the $z_i^\star$ are positive and sum to k, at most k of the $z_i^\star$ are 1 and thus at most k of the $\mu_i^\star$ are nonzero. Therefore, the solution of the problem (10) determines a covering ellipsoid for the m points $a_1, \ldots, a_m$, with at most k outliers.

D. Approximate Relaxed Sensor Selection

It is not necessary to solve the relaxed sensor selection problem (9) to high accuracy, since we use it only to get the upper bound U, and to find the indexes associated with the largest k values of its solution. In this section we describe a simple method for solving it approximately but very efficiently, while retaining a provable upper bound on $p^\star$. This can be done by solving a smooth convex problem, which is closely related to the subproblems solved in an interior-point method for solving the relaxed problem.

The *approximate relaxed sensor selection problem* is

$$
\begin{aligned}
\text{maximize} \quad & \psi(z) = \log \det \left(\sum_{i=1}^{m} z_i a_i a_i^T \right) \\
& + \kappa \sum_{i=1}^{m} \left(\log(z_i) + \log(1 - z_i) \right) \\
\text{subject to} \quad & \mathbf{1}^T z = k
\end{aligned}
\tag{11}
$$

with variable $z \in \mathbf{R}^m$. Here κ is a positive parameter that controls the quality of approximation. In the approximate relaxed sensor selection problem, we have implicit constraints that $z_i \in (0, 1)$. The function ψ is concave and smooth, so the problem (11) can be efficiently solved by Newton's method, which we describe in detail below. Let z^* denote the solution of the approximate relaxed sensor selection problem (11).

A standard result in interior-point methods [30, Sec. 11.2] is that z^* is at most $2m\kappa$ suboptimal for the relaxed sensor selection problem (9):

$$
\log \det \left(\sum_{i=1}^{m} z_i^* a_i a_i^T \right) \geq \log \det \left(\sum_{i=1}^{m} z_i^\star a_i a_i^T \right) - 2m\kappa
$$
$$
= U - 2m\kappa.
$$

In particular, we can use

$$
\tilde{U} = \log \det \left(\sum_{i=1}^{m} z_i^* a_i a_i^T \right) + 2m\kappa
\tag{12}
$$

as an upper bound on $p^\star$.

We can use this bound to choose κ so that, in terms of ρ, the increase in gap contributed by the term $2m\kappa$, which is a factor of $\exp(2m\kappa/2n)$, is small, say, 1%. This corresponds to $\kappa \approx 0.01 \, n/m$.

Newton's Method:

We now briefly describe Newton's method for solving (11); for full details, see, e.g., [30, Sec. 10.2]. As an initial (feasible) point, we take $z = (k/m)\mathbf{1}$. At each step, we compute the Newton search step δZ_{nt}, which can be expressed as

$$
\Delta Z_{nt} = -(\nabla^2 \psi)^{-1} \nabla \psi + \left(\frac{\mathbf{1}^t (\nabla^2 \psi)^{-1} \nabla \psi}{\mathbf{1}^t (\nabla^2 \psi)^{-1} \mathbf{1}} \right) (\nabla^2 \psi)^{-1} \mathbf{1}.
\tag{13}
$$

We then use a backtracking line search to compute a step size $s \in (0, 1]$, and update z by replacing it with $z + s\Delta z_{nt}$. We stop when the Newton decrement $(-\nabla \psi^t \Delta z_{nt})^{1/2}$ is small. The total number of steps required is typically ten or fewer.

For completeness we give expressions for the derivatives of ψ. Its gradient is given by

$$
(\nabla \psi)_i = a_i^T W a_i + \frac{\kappa}{z_i} - \frac{\kappa}{(1 - z_i)}, \quad i = 1, \ldots, m,
$$

where

$$
W = \left(\sum_{i=1}^{m} z_i a_i a_i^T \right)^{-1}.
$$

The Hessian $\nabla^2 \psi$ is given by

$$
\nabla^2 \psi = -(A W A^T) \circ (A W A^T)
$$
$$
- \kappa \, \mathbf{diag} \left(\frac{1}{z_1^2} + \frac{1}{(1 - z_1)^2}, \ldots, \frac{1}{z_m^2} + \frac{1}{(1 - z_m)^2} \right)
$$

where $\circ$ denotes the Hadamard (elementwise) product and A is the *measurement matrix*

$$
A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}.
\tag{14}
$$

We can give a complexity analysis for computing the Newton step Δz_{nt} using (13). We first form $\sum_{i=1}^{m} z_i a_i a_i^T$, which costs $O(mn^2)$ operations, and compute its Cholesky factor, which costs $O(n^3)$. We then form $A W A^T$ and $\nabla^2 \psi$, which costs $O(m^2 n)$. We compute its Cholesky factorization, which costs $O(m^3)$ (which dominates all other costs so far). Once we have computed the Cholesky factorization of $\nabla^2 \psi$, we can compute Δz_{nt} at cost $O(m^2)$. Thus, the overall cost is $O(m^3)$. Moreover, the hidden constant is quite modest, since the cost is dominated by the Cholesky factorization of an $m \times m$ matrix, which can be carried out in $(1/3)m^3$ operations.

E. Local Optimization

The construction of a feasible selection $\hat{z}$ from the solution of the (approximate) relaxed problem (11)z^* can be (possibly) improved by a local optimization method. One simple method to carry this out is to start from $\hat{z}$, and check sensor selections that can be derived from $\hat{z}$ by swapping one of the k chosen sensors (i.e., $i_1, \ldots, i_k$) with one of the $m - k$ sensors not chosen. For similar methods, see, e.g., Fedorov's exchange algorithm [29], [40] or Wynn's algorithm [41].

We can determine whether a sensor swap increases the objective value more efficiently than by computing the new objective value from scratch. Suppose we are to evaluate the change in objective value when sensor j is removed from our selection, and sensor l (which was not originally chosen) replaces it. We let $\hat{\Sigma}$ denote the error covariance with the original subset selection

$$
\hat{\Sigma} = \left(\sum_{i=1}^{m} \hat{z}_i a_i a_i^T \right)^{-1}
$$

and we let $\tilde{\Sigma}$ denote the error covariance when sensor j is swapped with sensor l

$$\tilde{\Sigma} = \left(\sum_{i=1}^m \hat{z}_i a_i a_i^T - a_j a_j^T + a_l a_l^T \right)^{-1}.$$

Using the low-rank update formula for the determinant of a matrix, we have

$$\begin{aligned} \det \tilde{\Sigma}^{-1} &= \det \left(\hat{\Sigma}^{-1} - a_j a_j^T + a_l a_l \right) \\ &= \det \hat{\Sigma}^{-1} \det \left(I + \begin{bmatrix} a_j^T \\ a_l^T \end{bmatrix} \hat{\Sigma} \begin{bmatrix} -a_j & a_l \end{bmatrix} \right). \end{aligned}$$

We can determine whether swapping j and l increases the objective, i.e., whether $\det \hat{\Sigma}^{-1} > \det \hat{\Sigma}^{-1}$, by evaluating the determinant of the 2×2 matrix

$$S = I + \begin{bmatrix} a_j^T \\ a_l^T \end{bmatrix} \hat{\Sigma} [-a_j \ a_l].$$

The computation effort required to calculate this matrix is $O(n^2)$. (In contrast, computing $\det \tilde{\Sigma}$ from scratch requires $O(n^3)$, so the savings here is $O(n)$.) A small gain in efficiency can be obtained by recognizing

$$\begin{aligned} \det S &= 1 + a_l^T \hat{\Sigma} a_l - a_j^T \hat{\Sigma} a_j + (a_l^T \hat{\Sigma} a_j)^2 \\ &\quad - (a_l^T \hat{\Sigma} a_l)(a_j^T \hat{\Sigma} a_j) \\ &\leq 1 + a_l^T \hat{\Sigma} a_l - a_j^T \hat{\Sigma} a_j \end{aligned}$$

remembering the previously calculated products of the form $a_i^T \hat{\Sigma} a_i$, and checking $a_l^T \hat{\Sigma} a_l > a_j^T \hat{\Sigma} a_j$ before calculating $\det S$.

Now we continue our description of the local optimization method. Given the current sensor selection, we attempt a search over all possible $k(m-k)$ swaps. If we find that no swap increases the objective value, the algorithm terminates. The solution so obtained is called *2-opt*, because exchanging any one selected sensor with any unselected one will not improve the solution.

If, however, we encounter a swap that increases the objective value, we (greedily) update $\hat{z}$ to correspond to the new sensor selection, replacing $\hat{\Sigma}$ by $\tilde{\Sigma}$. The matrix $\tilde{\Sigma}$ can be evaluated efficiently using the matrix inversion lemma (also known as Woodbury formula):

$$\tilde{\Sigma} = \hat{\Sigma} - \hat{\Sigma} \begin{bmatrix} -a_j & a_l \end{bmatrix} \left(I + \begin{bmatrix} a_j^T \\ a_l^T \end{bmatrix} \hat{\Sigma} \begin{bmatrix} -a_j & a_l \end{bmatrix} \right)^{-1} \begin{bmatrix} a_j^T \\ a_l^T \end{bmatrix} \hat{\Sigma}.$$

The computation effort required to calculate $\tilde{\Sigma}$ given $\hat{\Sigma}$ is $O(n^2)$. With the new sensor selection $\hat{z}$ we restart the search for an improving swap.

The local optimization algorithm must terminate because there is only a finite number of sensor selections that are better than the original one. The total number of local optimization steps can be very large (in theory); so we can simply limit the number of steps taken, say to N_{loc}. (We should mention that we have never observed an example that requires a very large number of local optimization steps.) If N_{loc} is chosen to grow no faster than m^3/n^2, then the total computational effort of the

local optimization method will be $O(m^3)$, the same as solving the relaxed sensor problem.

More Sophisticated Local Optimization:

The local optimization method described above does not use the solution of the (approximate) relaxed sensor selection problem z^*; Instead it proceeds directly from the rounded estimate $\hat{Z}$. More sophisticated rounding methods can use the approximate relaxed point z^*. For example, in a randomized rounding scheme, z_i^* is interpreted as the probability of selecting sensor i. In the local optimization method, we can use z^* to order the sensors which are checked for possible swapping. (In the local optimization described above, the sensors are checked according to their index.) More Specifically, we choose unselected sensors in descending order of the z_i^* values, and we pick the selected sensors in ascending order of the z_i^* values. The intuition behind this scheme is that a sensor with higher z_i^* is more likely to be in the globally optimal sensor selection. To determine the ordering we need to sort the sensors according to the values only one time, and then maintain the ordering when a swap is taken. The initial sorting requires a computation effort of $o(m \log M)$, which for practical values of m and k is dominated by the computational effort needed to check the $k(m-k)$ swaps. We can also restrict the swaps to be among those sensors for which z_i^* is in the interval $[0.1, 0.9]$ (or some interval, possibly symmetric, around $1/2$). This drastically reduces the number of swaps to be checked (and number of sensors to be sorted), and therefore speeds up the local optimization.

IV. EXAMPLE

In this section, we illustrate the sensor selection method with a numerical example. We consider an example instance with $m = 100$ potential sensors and $n = 20$ parameters to estimate. The measurement vectors $a_1, \ldots, a_m$ are chosen randomly, and independently, from an $\mathcal{N}(0, I/\sqrt{n})$ distribution. We solve the relaxed problem (11), with $\kappa = 10^{-3}$, and find suboptimal subset selections, with and without local optimization, for $k = 20, 21, \ldots, 40$.

To solve each approximate relaxed problem requires 11 Newton steps, which would take a few milliseconds in a C implementation, run on a typical 2-GHz personal computer. For each problem instance, the (basic) local search checks 4000–12 000 sensor swaps, and around 3–20 swaps are taken before a 2-opt solution is found. We also the run the restricted version on the local search, which only considers sensors with z_i^* value in the interval $[0.1, 0.9]$. This local search produces an equally good final sensor selection, while checking a factor 10–15 times fewer swaps than the basic method. (In any case, the basic local search only takes milliseconds to complete, on a typical personal computer, for a problem instance of this size.)

To show the quality of the sensor subsets chosen, we evaluate the upper bound $\tilde{U}$ [given by (12)], the lower bound L using the simple selection rule, and the (possibly) better lower bounds L_{loc} and $L_{\mathrm{loc(r)}}$ obtained after local optimization and restricted local optimization, respectively, for each value of k. The top half of Fig. 1 shows $\tilde{U}$, L, L_{loc}, and $L_{\mathrm{loc(r)}}$, and the bottom half
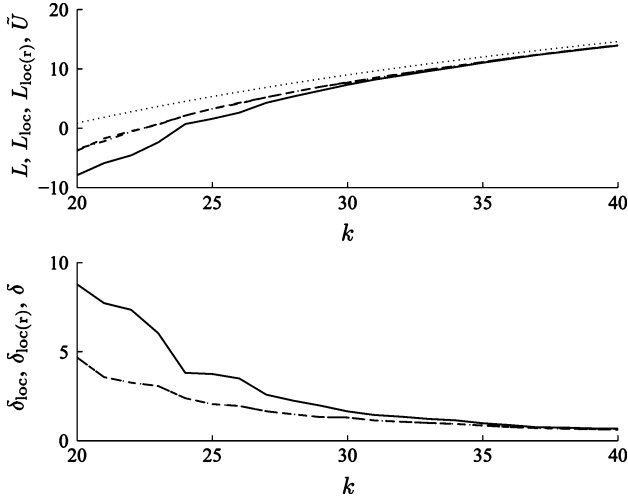
Fig. 1. Top: Upper bound $\tilde{U}$ (top curve); lower bounds L_{loc} and $L_{\text{loc(r)}}$ (middle curves); lower bound L (bottom curve). Bottom: Gap δ (top curve); δ_{loc} and $\delta_{\text{loc(r)}}$ (bottom curves).
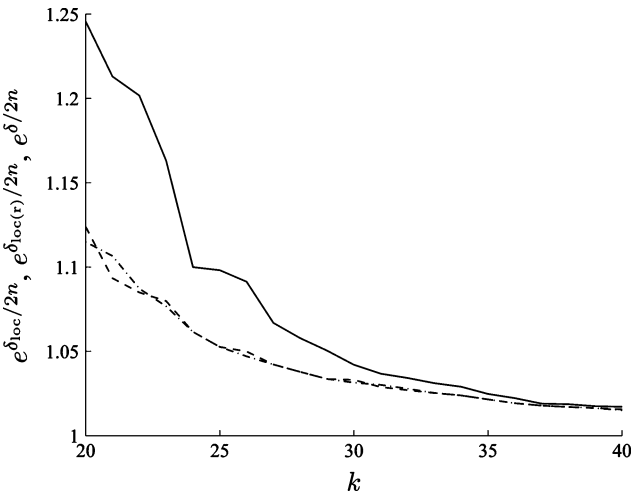

Fig. 2. Gaps expressed as ratios of mean radii: $\exp(\delta/2n)$ (top curve); $\exp(\delta_{\text{loc}}/2n)$ and $\exp(\delta_{\text{loc(r)}}/2n)$ (bottom curves).

shows the gaps $\delta = \tilde{U} - L$, $\delta_{\text{loc}} = \tilde{U} - L_{\text{loc}}$, and $\delta_{\text{loc(r)}} = \tilde{U} - L_{\text{loc(r)}}$. We also express the gaps as the ratio of mean radii $\exp(\delta/2n)$, $\exp(\delta_{\text{loc}}/2n)$, and $\exp(\delta_{\text{loc(r)}}/2n)$, in Fig. 2.

These plots show that very good sensor selections are obtained. For example, with $k = 25$, the relaxation followed by 2-opt local optimization produces a design which is at most 5.3% suboptimal, as measured by mean radius of the confidence ellipsoid. (This is only a bound; it is likely that the sensor selection found is closer to optimal than 5.3%.) We can see that restricted local optimization performs as well as basic local optimization; the two curves are barely distinguishable. (In the figures, the values corresponding to the basic local optimization are shown by the dashed curve, and to the restricted local optimization are shown by the dash-dotted curve.) To find the globally optimally best sensor selection by direct enumeration would require evaluating the objective $\binom{100}{25}$ times, which is on the order of 10^{23} times, and clearly not practical.

V. EXTENSIONS

A. Other Measures of Estimation Quality

So far we have used the volume of the confidence ellipsoid as our measure of the quality of estimation obtained by a sensor subset selection. Several other measures can be used instead of this one.

Mean Squared Error:

The mean squared error in estimating the parameter x is

$$\mathbf{E}(\|X - \hat{X}\|^2) = \mathbf{tr}(\Sigma).$$

The associated sensor selection problem can be expressed as the optimization problem

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{tr}\left(\sum_{i=1}^{m} z_i a_i a_i^T\right)^{-1} \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& z_i \in \{0,1\}, \quad i = 1, \ldots, m
\end{aligned}
$$

with variable $z \in \mathbf{R}^m$. The optimization problem obtained by relaxing the 0–1 constraint is convex; see [30, Sec. 7.5]. (In the context of experiment design, this measure leads to so-called A-optimal experiment design.)

Worst Case Error Variance:

The variance of the estimation error in the direction $q \in \mathbf{R}^n$, With $\|q\| = 1$, is

$$\mathbf{E}(q^T x)^2 = q^T \Sigma q.$$

The worst case variance of the estimation error, over all directions, is

$$\max_{\|q\|=1} q^T \Sigma q = \lambda_{\max}(\Sigma)$$

the maximum eigenvalue of Σ. The associated sensor selection problem can be expressed as the optimization problem

$$
\begin{aligned}
\text{maximize} \quad & \lambda_{\min}\left(\sum_{i=1}^{m} z_i a_i a_i^T\right) \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& z_i \in \{0,1\}, \quad i = 1, \ldots, m
\end{aligned}
$$

with variable $z \in \mathbf{R}^m$. Relaxing the 0–1 constraint we obtain a convex problem. (In the context of experiment design, this measure leads to so-called e-optimal experiment design.)

1) Worst Case Coordinate Error Variance: The variance of the ith coordinate of the estimation error, $x_i - \hat{x}_i$, is Σ_{ii}. The worst case coordinate error variance is the largest diagonal entry of the covariance matrix. Choosing the sensor subset to minimize this measure can be expressed as the problem

$$
\begin{aligned}
\text{minimize} \quad & \max_{j=1,\ldots,n} \left(\left(\sum_{i=1}^{m} z_i a_i a_i^T\right)^{-1}\right)_{jj} \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& z_i \in \{0,1\}, \quad i = 1, \ldots, m \quad (15)
\end{aligned}
$$

with variable z. Relaxing the 0–1 constraint we obtain a convex problem.

In fact, the problem can be transformed to a semidefinite program (SDP), and therefore efficiently solved. Writing the problem (15) in epigraph form and relaxing the 0–1 constraint, we obtain

$$
\begin{aligned}
\text{minimize} \quad & t \\
\text{subject to} \quad & t \geq e_j^T \left(\left(\sum_{i=1}^m z_i a_i a_i^T \right)^{-1} \right) e_j, \quad j = 1, \ldots, m \\
& \mathbf{1}^T z = k \\
& 0 \leq z_i \leq 1, \quad i = 1, \ldots, m
\end{aligned}
$$

with variables $z \in \mathbf{R}^m$ and $t \in \mathbf{R}$. The vector e_i is the vector with 1 in the ith entry and 0 in the rest of the entries. This problem is equivalent to

$$
\begin{aligned}
\text{minimize} \quad & t \\
\text{subject to} \quad & \begin{bmatrix} t & e_j^T \\ e_j & \sum_{i=1}^m z_i a_i a_i^T \end{bmatrix} \succeq 0, \quad j = 1, \ldots, m \\
& \mathbf{1}^T z = k \\
& 0 \leq z_i \leq 1, \quad i = 1, \ldots, m
\end{aligned}
$$

with variables z and t. (The symbol $\succeq$ represents inequality with respect to the positive semidefinite matrix cone.) This is a semidefinite program.

B. Sensor Selection Constraints

Many constraints on the selection of the sensors can be represented as linear equalities or inequalities on the variable z, and so are easily incorporated into the convex relaxation. We describe some typical cases below.

1) Logical Constraints:
- *"Only when" constraints.* The constraint that sensor i can be chosen only when sensor j is also chosen can be expressed as $z_i \leq z_j$.
- *"Not both" constraints.* The constraint that sensor i and sensor j cannot both be chosen can be expressed as $z_i + z_j \leq 1$.
- *"At least one of'" constraints.* To require that one of sensor i or sensor j be chosen, we impose the constraint $z_i + z_j \geq 1$.

These are easily extended to more complex situations. For example, to require that exactly two of the four sensors i, j, k, and l be chosen, we impose the linear equality constraint $z_i + z_j + z_k + z_l = 2$.

2) Budget Constraints: In addition to limiting the number of sensors chosen to k, we can impose other resource limitations on the sensor selection. Suppose that c_i is some cost (say, in dollars, power, or weight) associated with choosing sensor i. We can impose a budget constraint on the selection, i.e., a maximum allowed cost for the selection, as $c^T z \leq B$, where B is the budget.

C. Vector Measurements

In our setup so far, each sensor gives a scalar measurement. Now suppose the measurements y_i are vectors, i.e., sensor i gives not one, but several scalar measurements of the parameters. The m potential measurements are

$$
y_i = A_i^T x + v_i, \quad i = 1, \ldots, m
$$

where $y_i \in \mathbf{R}^{m_i}$, $x \in \mathbf{R}^n$. The measurement noises $v_1, \ldots, v_m$ are independent random variables, with $\mathcal{N}(0, \sigma^2 I)$ distribution. The sensor selection problem can be expressed as

$$
\begin{aligned}
\text{maximize} \quad & \log \det \left(\sum_{i=1}^m z_i A_i A_i^T \right) \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& z_i \in \{0, 1\}, \quad i = 1, \ldots, m
\end{aligned}
$$

with variable $z \in \mathbf{R}^m$. Relaxing the 0–1 constraint we obtain a convex problem. (The same problem can be also be obtained by associating each component of y_i with a separate measurement, and adding constraints that require that if any scalar measurement from y_i is used, all must be.)

D. MAP Estimation

We have so far worked with maximum likelihood estimation. We can easily extend the method to the Bayesian framework. Suppose the prior density of x is $\mathcal{N}(0, \Sigma_x)$. The maximum *a posteriori* probability (MAP) estimate of x, with selected sensors characterized by z, is

$$
\hat{x}_{map} = \left(\sigma^{-2} \sum_{i=1}^m z_i a_i a_i^T + \Sigma_x^{-1} \right)^{-1} \sum_{i=1}^m y_i a_i.
$$

The estimation error $x - \hat{x}_{map}$ has zero mean and covariance

$$
\Sigma_{map}(z) = \left(\sigma^{-2} \sum_{i=1}^m z_i a_i a_i^T + \Sigma_x^{-1} \right)^{-1}. \tag{16}
$$

The problem of choosing k sensors to minimize the volume of the resulting η-confidence ellipsoid reduces to

$$
\begin{aligned}
\text{maximize} \quad & \log \det \left(\sigma^{-2} \sum_{i=1}^m z_i a_i a_i^T + \Sigma_x^{-1} \right) \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& z_i \in \{0, 1\}, \quad i = 1, \ldots, m
\end{aligned} \tag{17}
$$

with variable $z \in \mathbf{R}^m$. Relaxing the 0–1 constraint results in a convex optimization problem.

Since the $\log \det$ of the covariance matrix of a Gaussian random variable is the entropy of the random variable (differing by a constant), the problem (17) can be obtained via an information theoretic approach. Let $\tilde{y}(z)$ be the sensor measurement vector when sensors characterized by z are chosen. The problem of choosing k sensors to minimize the entropy of the random variable $x|\tilde{y}(z)$, or to maximize the mutual information between x and the resulting measurement vector $\tilde{y}(z)$, is the problem (17).

E. Estimating a Linear Function of the Parameters

Suppose the goal is to estimate $\tilde{x} = Cx \in \mathbf{R}^{\tilde{n}}$, a linear function of the parameter x, where $C \in \mathbf{R}^{\tilde{n} \times n}$ has rank $\tilde{n}$. The prior density of x is $\mathcal{N}(0, \Sigma_x)$, so the prior density of $\tilde{x}$ is $\mathcal{N}(0, \Sigma_{\tilde{x}})$, where $\Sigma_{\tilde{x}} = C\Sigma_x C^T$. The covariance of the error of the MAP estimate of $\tilde{x}$ is $C\Sigma_{map}(z)C^T$, where $\Sigma_{map}(z)$ is given by (16). The problem of choosing k sensors to minimize the volume of the resulting confidence ellipsoid is

$$
\text{minimize} \quad \log \, \det \left(C \left(\sigma^{-2} \sum_{i=1}^{m} z_i a_i a_i^T + \Sigma_{\tilde{x}}^{-1} \right)^{-1} C^T \right)
$$
$$
\text{subject to} \quad \mathbf{1}^T z = k
$$
$$
z_i \in \{0, 1\}, \quad i = 1, \ldots, m, \tag{18}
$$

with variable $z \in \mathbf{R}^m$.

Relaxing the constraints $z_i \in \{0, 1\}$ to $0 \le z_i \le 1$ yields a convex problem. This relaxed problem can be solved directly by the Newton's method described in Section III-D, for which we need the gradient and Hessian of the objective function. The objective function is

$$
\phi(z) = \log \, \det \left(CXC^T \right)
$$

where $X \in \mathbf{S}^n$ is

$$
X = \left(\sigma^{-2} \sum_{i=1}^{m} z_i a_i a_i^T + \Sigma_{\tilde{x}}^{-1} \right)^{-1}.
$$

(To simplify notation we do not write X explicitly as a function of z.) The gradient of the function ϕ is

$$
(\nabla \phi)_i = -\sigma^{-2} a_i^T X C^T \left(CXC^T \right)^{-1} CXa_i, \quad i = 1, \ldots, m.
$$

The Hessian of the function ϕ is

$$
(\nabla^2 \phi)_{ij} = 2\sigma^{-4} \left(a_i^T X a_j \right) \left(a_i^T X C^T \left(CXC^T \right)^{-1} CXa_j \right)
$$
$$
- \sigma^{-4} \left(a_i^T X C^T \left(CXC^T \right)^{-1} CXa_j \right)^2,
$$
$$
i = 1, \ldots, m, \quad j = 1, \ldots, m
$$

which can be written compactly as

$$
\nabla^2 \phi = \sigma^{-4} (2P - Q) \circ Q,
$$

where

$$
P = AXA^T,
$$
$$
Q = AXC^T \left(CXC^T \right)^{-1} CXA^T
$$

and A is given by (14).

The problem (18) can also be solved by transforming it to a standard one. We introduce a new variable, a lower triangular matrix $L \in \mathbf{R}^{\tilde{n} \times \tilde{n}}$, and write the relaxed version of the problem (18) as

$$
\text{minimize} \quad \log \, \det (LL^T)^{-1}
$$
$$
\text{subject to} \quad \mathbf{1}^T z = k
$$
$$
0 \le z_i \le 1, \quad i = 1, \ldots, m
$$
$$
(LL^T)^{-1} \succeq C \left(\sigma^{-2} \sum_{i=1}^{m} z_i a_i a_i^T + \Sigma_{\tilde{x}}^{-1} \right)^{-1} C^T
$$
$$
L \text{ lower triangular} \tag{19}
$$

with variables $z \in \mathbf{R}^m$ and $L \in \mathbf{R}^{\tilde{n} \times \tilde{n}}$. The objective function is

$$
\log \, \det (LL^T)^{-1} = -2 \log \, \det (L) = -2 \sum_{i=1}^{\tilde{n}} \log L_{ii}
$$

since L is lower triangular. The constraint that the lower triangular matrix L is invertible is implicit, since the objective function requires that $L_{ii} > 0$.

The matrix inequality

$$
(L^T)^{-1} L^{-1} \succeq C \left(\sigma^{-2} \sum_{i=1}^{m} z_i a_i a_i^T \Sigma_{\tilde{x}}^{-1} \right)^{-1} C^T
$$

can be written as

$$
I \succeq L^T C \left(\sigma^{-2} \sum_{i=1}^{m} z_i a_i a_i^T + \Sigma_{\tilde{x}}^{-1} \right)^{-1} C^T L
$$

which is equivalent to

$$
\begin{bmatrix} I & L^T C \\ C^T L & \sigma^{-2} \sum_{i=1}^{m} z_i a_i a_i^T + \Sigma_{\tilde{x}}^{-1} \end{bmatrix} \succeq 0.
$$

(Here we use $\sum_{i=1}^{m} z_i a_i a_i^T + \Sigma_{\tilde{x}}^{-1} \succ 0$.)

The problem (19) is therefore equivalent to

$$
\text{maximize} \quad \sum_{i=1}^{\tilde{n}} \log L_{ii}
$$
$$
\text{subject to} \quad \mathbf{1}^T z = k
$$
$$
0 \le z_i \le 1, \quad i = 1, \ldots, m
$$
$$
\begin{bmatrix} I & L^T C \\ C^T L & \sigma^{-2} \sum_{i=1}^{m} z_i a_i a_i^T + \Sigma_{\tilde{x}}^{-1} \end{bmatrix} \succeq 0
$$
$$
L \text{ lower triangular} \tag{20}
$$

with variables $z \in \mathbf{R}^m$ and $L \in \mathbf{R}^{\tilde{n} \times \tilde{n}}$.

A similar approach can handle the problem of estimating a linear function of the variable x in the maximum likelihood framework, but this requires additional technical conditions.

F. Robust Sensor Selection

In this section we consider the sensor problem with some uncertainty in the measurement vectors. The uncertainty is characterized by a given set $\mathcal{A}$ in which the measurement matrix A, given by (14), can take any value. In terms of A, the objective of the sensor selection problem (8) can be written as

$$
\log \, \det (A^T \, \mathbf{diag} \, (z) A)
$$

where $\mathbf{diag} \, (z) \in \mathbf{R}^{m \times m}$ is the diagonal matrix with entries $z_1, \ldots, z_m$.

In the *robust sensor selection problem* we choose k sensors to minimize the worst case mean radius of the resulting confidence ellipsoid, which can be written as

$$
\text{maximize} \quad \inf_{A \in \mathcal{A}} \log \, \det (A^T \, \mathbf{diag} \, (z) A)
$$
$$
\text{subject to} \quad \mathbf{1}^T z = k
$$
$$
z_i \in \{0, 1\}, \quad i = 1, \ldots, m \tag{21}
$$

with variables $z \in \mathbf{R}^m$. The problem data is the set x$\mathcal{A}$.

The objective function of the robust optimization is problem (21), for $z_i \geq 0$, is the infimum of a family of concave functions, and therefore concave. Thus, the problem (21), after relaxing the 0–1 constraints, is a convex optimization problem.

The relaxed robust sensor selection problem can be written as

$$
\begin{aligned}
\text{maximize} \quad & \log \det (U) \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& 0 \leq z_i \leq 1, \quad i = 1, \dots, m \\
& U \preceq A^T \operatorname{diag}(z) A, \quad \text{for all } A \in \mathcal{A},
\end{aligned}
$$

with variables $z \in \mathbf{R}^m$ and the symmetric (positive definite) matrix $U \in \mathbf{R}^{m \times m}$. If the set $\mathcal{A}$ is finite, this is a standard convex optimization problem. If the set $\mathcal{A}$ is not finite, which usually is the case, the problem is a semi-infinite convex optimization problem, which can be solved using various general techniques, such as sampling; see, e.g., [42] and [43]. In some cases (as we will see below), the semi-infinite problem can be simplified and solved.

We now consider the specific uncertainty model

$$
A \in \mathcal{A} = \left\{ \bar{A} + \Delta \mid \|\|\Delta\|_2 \leq \epsilon \right\} \tag{22}
$$

the semi-infinite constraint can represented as a (simple) linear matrix inequality (LMI) constraint, thereby simplifying the robust sensor selection problem to a standard SDP. The constraint $\|\Delta\|_2 \leq \epsilon$ can be written as $\Delta^T \Delta \preceq \epsilon^2 I$. The semi-infinite constraint in terms of $\bar{A}$ and Δ is

$$
\begin{aligned}
U \preceq \bar{A}^T \operatorname{diag}(z) \bar{A} + \bar{A}^T \operatorname{diag}(z) \Delta + \Delta^T \operatorname{diag}(z) \bar{A} \\
+ \Delta^T \operatorname{diag}(z) \Delta, \quad \text{for all } \Delta^T \Delta \preceq \epsilon^2 I.
\end{aligned}
$$

Theorem 3.3 in [44, Sec. 3] implies that the above semi-infinite quadratic matrix inequality holds if and only if the matrix inequality

$$
\begin{bmatrix}
\bar{A}^T \operatorname{diag}(z) \bar{A} - U - tI & \bar{A}^T \operatorname{diag}(z) \\
\operatorname{diag}(z) \bar{A} & \operatorname{diag}(z) + \left(\frac{t}{\epsilon^2}\right) I
\end{bmatrix} \succeq 0
$$

is feasible for some $t \geq 0$. The matrix inequality is linear in z, U, and t. The relaxed robust sensor selection for the uncertainty model (22) is [see the equation at the bottom of the page], with variables z, U, and t.

G. Example

In this section, we consider an example that combines three of the extensions. We consider a discrete-time linear dynamical system

$$
x(t+1) = Fx(t), \qquad t = 1, \dots, m+1 \tag{23}
$$

where $x(t) \in \mathbf{R}^n$ is the state at time t, and $F \in \mathbf{R}^{n \times n}$ is the dynamics matrix, which we assume is invertible. We have linear noise corrupted measurements

$$
y(t) = Hx(t) + v(t), \qquad t = 1, \dots, m \tag{24}
$$

where $y(t) \in \mathbf{R}^p$ is the measurement at time t, $v(t) \in \mathbf{R}^p$ is the measurement noise at time t, and $H \in \mathbf{R}^{p \times n}$ is the measurement matrix. We assume the noise vectors $v_1, \dots, v_m$ are independent identically distributed $\mathcal{N}(0, I)$ random variables. The initial state $x(1)$ has a prior probability density $\mathcal{N}(0, \Sigma_1)$, and is independent of the noise vectors.

We consider the problem of choosing a set of k (vector) measurements out of the m (vector) measurements of the state, in order to minimize the mean squared error in estimating $x(m+1)$. This corresponds to choosing a set of k *times* (out of the m possible times) at which to obtain the measurements.

We can express the measurements as

$$
\begin{bmatrix} y(1) \\ \vdots \\ y(m) \end{bmatrix} = \begin{bmatrix} HF^{-m} \\ \vdots \\ HF^{-1} \end{bmatrix} x(m+1) + \begin{bmatrix} v(1) \\ \vdots \\ v(m) \end{bmatrix}.
$$

The prior density of $x(m+1)$ is $\mathcal{N}(0, \Sigma_{m+1})$, where $\Sigma_{m+1} = F^m \Sigma_1 F^{mT}$. The MAP estimation error for $x(m+1)$ is zero mean, with covariance

$$
\Sigma(z) = \left(\sum_{i=1}^m z_i (HF^{-m+i-1})^T HF^{-m+i-1} + \Sigma_{m+1}^{-1} \right)^{-1}
$$

where $z \in \mathbf{R}^m$ characterizes the selected measurement times. The problem of choosing k times at which to take state measurements, in order to minimize the resulting mean square estimation error of $x(m+1)$, can be formulated as

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{tr}(\Sigma(z)) \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& z \in \{0, 1\}, \quad i = 1, \dots, m \tag{25}
\end{aligned}
$$

$$
\begin{aligned}
\text{maximize} \quad & \log \det (U) \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& 0 \leq z_i \leq 1, \quad i = 1, \dots, m \\
& \begin{bmatrix} \bar{A}^T \operatorname{diag}(z) \bar{A} - U - tI & \bar{A}^T \operatorname{diag}(z) \\ \operatorname{diag}(z) \bar{A} & \operatorname{diag}(z) + \left(\frac{t}{\epsilon^2}\right) I \end{bmatrix} \succeq 0 \\
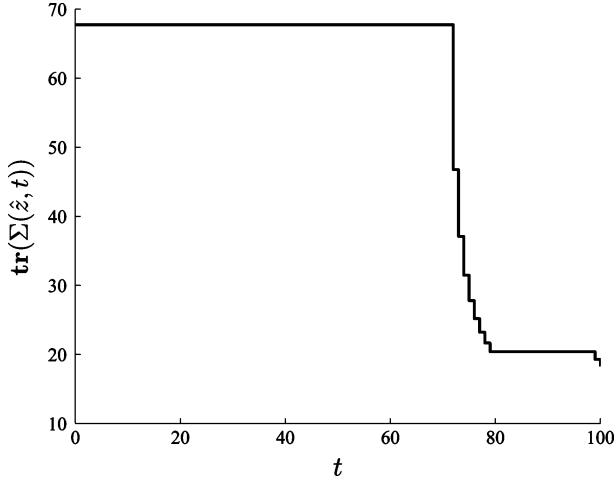& t \geq 0,
\end{aligned}
$$

Fig. 3. Mean-square error versus time for the sample time selection characterized by $\hat{z}$.

where z is variable. Relaxing the 0–1 constraints we obtain a convex optimization problem, which can be transformed to the semidefinite program

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{tr}(Y) \\
\text{subject to} \quad & \mathbf{1}^T z = k \\
& 0 \leq z_i \leq 1, \quad i = 1, \ldots, m \\
& \begin{bmatrix} Y & I \\ I & \Sigma(z)^{-1} \end{bmatrix} \succeq 0
\end{aligned} \tag{26}
$$

with variables $z \in \mathbf{R}^m$ and $Y \in \mathbf{S}^n$.

We now consider a numerical instance of the problem. We take state dimension $n = 5$, measurement dimension $p = 2$, over time interval $m = 100$, out of which we are to choose $k = 10$ times. We take the covariance of $x(1)$ to be $\Sigma_1 = I$. The dynamics matrix F has eigenvalues

$$0.99 \pm 0.05i, \quad 1.01 \pm 0.03i, \quad 0.98$$

i.e., a slowing growing, and a slowing decaying, oscillatory modes. The entries of the matrix H are chosen independently from the uniform distribution on $[-0.1, 0.1]$.

We solve the semidefinite program (26) using CVX[45] to obtain the solution of the relaxed problem $z^\star$, and select the k times with the largest values of $z_i^\star$. The chosen times are

$$72, \ldots, 79, \quad 99, \quad 100.$$

The objective value (mean square error) for this choice of 10 sample times is very close to the lower bound, given by the optimal value of the problem (25), so our choice is near globally optimal (and in particular, there is no need for local optimization).

In Fig. 3 we plot the mean square estimation error of $x(m+1)$, given the chosen measurements up to time t, which is given by $\mathbf{tr}(\Sigma(\hat{z}, t))$, where

$$
\Sigma(\hat{z}, t) = \left(\sum_{i=1}^{t} \hat{z}_i (HF^{-m+i-1})^T HF^{-m+i-1} + \Sigma_{m+1}^{-1} \right)^{-1},
$$
$$t = 0, \ldots, m$$

where $\hat{z}$ is the (0–1) sample time selection. This mean square error drops after each measurement is taken. We can see that the largest drop in mean square estimation error occurs during the burst of measurements taken over the interval $t = 72$ to $t = 79$; a small further improvement occurs in the last two time steps.

VI. Conclusion

The problem of choosing k sensors or measurements, from among a set of candidate measurements, in order to obtain the best resulting estimate of some parameters, is in general a difficult combinatorial problem. We have shown, however, that convex relaxation, followed by a local optimization method, can often work very well. In particular, this method produces not only a suboptimal choice of measurements, but also, a bound on how well the globally optimal choice does. The performance achieved by the suboptimal choice is often very close to the global bound, which certifies that the choice is nearly optimal. Our method does not give a prior guarantee on this gap; but each time the method is used, on a particular problem instance, we get a specific bound.

Appendix
Derivation of the Dual Problem

In this section we derive the dual of the relaxed sensor selection problem (9). We introduce a new variable $X = \sum_{i=1}^{m} z_i a_i a_i^T$ and write the relaxed sensor selection problem (9) as

$$
\begin{aligned}
\text{minimize} \quad & \log \det X^{-1} \\
\text{subject to} \quad & X = \sum_{i=1}^{m} z_i a_i a_i^T \\
& \mathbf{1}^T z = k \\
& 0 \leq z_i \leq 1, \quad i = 1, \ldots, m
\end{aligned} \tag{27}
$$

with variables $z \in \mathbf{R}^m$ and $X \in \mathbf{S}^n$ (set of symmetric $n \times n$ matrices). To form the Lagrangian of the problem (27) we introduced Lagrange multipliers λ_i for $z_i \geq 0$, μ_i for $z_i \leq 1$, ν for $\mathbf{1}^T z = k$, and Λ for $X = \sum_{i=1}^{m} z_i a_i a_i^T$. The Lagrangian L is

$$
\begin{aligned}
L(z, X, \Lambda, \lambda, \mu, \nu) = {} & \log \det X^{-1} + \mathbf{tr}(X\Lambda) \\
& - \sum_{i=1}^{m} z_i a_i^T \Lambda a_i + \nu(\mathbf{1}^T z - k) \\
& - \lambda^T z - \mu^T (\mathbf{1} - z)
\end{aligned}
$$

where $\Lambda \in \mathbf{S}^n$, $\lambda \in \mathbf{R}^m$, $\mu \in \mathbf{R}^m$, and $\nu \in \mathbf{R}$. Rearranging the terms, we get

$$
\begin{aligned}
L(z, X, \Lambda, \lambda, \mu, \nu) = {} & \log \det X^{-1} + \mathbf{tr}(X\Lambda) - k\nu - \mathbf{1}^T \mu \\
& + \sum_{i=1}^{m} z_i \left(\nu - a_i^T \Lambda a_i - \lambda_i + \mu_i \right).
\end{aligned}
$$

The Lagrange dual function g is given by

$$
g(\Lambda, \lambda, \mu, \nu) = \inf_{z, X} L(z, X, \Lambda, \lambda, \mu, \nu).
$$

$$g(\Lambda, \lambda, \mu, \nu) = \begin{cases} \log \det \Lambda + n - k\nu - \mathbf{1}^T \mu & \text{if } \nu + \mu_i - a_i^T \Lambda a_i = \lambda_i, \ i = 1, \dots, m \\ -\infty & \text{otherwise} \end{cases}.$$

Minimum of L over z_i is bounded only if

$$\nu - a_i^T \Lambda a_i - \lambda_i + \mu_i = 0.$$

Minimizing L over X yields $X^{-1} = \Lambda$. The Lagrange dual function g is [see the equation at the top of the page]. The dual problem is

$$\begin{aligned} \text{maximize} \quad & \log \det \Lambda + n - k\nu - \mathbf{1}^T \mu \\ \text{subject to} \quad & \nu + \mu_i - a_i^T \Lambda a_i = \lambda_i, \quad i = 1, \dots, m \\ & \lambda_i \geq 0, \quad i = 1, \dots, m \\ & \mu_i \geq 0, \quad i = 1, \dots, m \end{aligned} \tag{28}$$

with variables Λ, λ, μ, and ν. (The constraint $\Lambda \succ 0$, i.e., Λ positive definite, is implicit.) The variable λ can be eliminated, and we write the dual problem as

$$\begin{aligned} \text{minimize} \quad & \log \det \Lambda^{-1} - n + k\nu + \mathbf{1}^T \mu \\ \text{subject to} \quad & \nu + \mu_i \geq a_i^T \Lambda a_i, \quad i = 1, \dots, m \\ & \mu_i \geq 0, \quad i = 1, \dots, m \end{aligned} \tag{29}$$

with variables $\Lambda \in \mathbf{S}^n$, $\mu \in \mathbf{R}^n$, and $\nu \in \mathbf{R}$.

REFERENCES

[1] C. Guestrin, A. Krause, and A. Singh, "Near-optimal sensor placements in Gaussian processes," School of Computer Science, Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep., 2005.

[2] G. Hovland and B. McCarragher, "Dynamic sensor selection for robotic systems," in *Proc. IEEE Int. Conf. Robotics Automation*, 1997, vol. 1, pp. 272–277.

[3] D. Kammer, "Sensor placement for on-orbit modal identification and correlation of large space structures," *J. Guid., Control, Dynam.*, vol. 14, pp. 251–259, 1991.

[4] K. Kincaid and S. Padula, "D-optimal designs for sensor and actuator locations," *Compute. Oper. Res.*, vol. 29, no. 6, pp. 701–713, 2002.

[5] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proc. 3rd Int. Symp. Information Processing in Sensor Networks*, Berkeley, CA, 2004, pp. 36–45.

[6] V. Isler and R. Bajcsy, "The sensor selection problem for bounded uncertainty sensing models," in *Proc. 4th Int. Symp. Information Processing in Sensor Networks*, Los Angeles, CA, 2005.

[7] K. Kookos and J. Perkins, "A systematic method for optimum sensor selection in inferential control systems," *Ind. Eng. Chem. Res.*, vol. 38, no. 11, pp. 4299–4308, 1999.

[8] F. Zhao and L. Guibas, *Wireless Sensor Networks: An Information Processing Approach*. San Mateo, CA: Morgan Kaufmann, 2004.

[9] V. Gupta, T. Chung, B. Hassibi, and R. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, 2006.

[10] M. Kalandros and L. Pao, "Controlling target estimate covariance in centralized multi-sensor systems," in *Proc. Amer. Control Conf.*, 1998, vol. 5.

[11] Y. Oshman, "Optimal sensor selection strategy for discrete-time state estimators," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 2, pp. 307–314, Apr. 1994.

[12] H. Rowaihy, S. Eswaran, M. Johnson, D. Verma, A. Bar-Noy, T. Brown, and T. L. Porta, "A survey of sensor selection schemes in wireless sensor networks," in *Proc. SPIE*, 2007, vol. 6562.

[13] R. Debouk, S. Lafortune, and D. Teneketzis, "On an optimization problem in sensor selection," *Discrete Event Dynam. Syst.*, vol. 12, no. 4, pp. 417–445, 2002.

[14] S. Jiang, R. Kumar, and H. Garcia, "Optimal sensor selection for discrete-event systems with partial observation," *IEEE Trans. Autom. Control*, vol. 48, no. 3, pp. 369–381, Mar. 2003.

[15] L. Yao, W. Sethares, and D. Kammer, "Sensor placement for on-orbit modal identification via a genetic algorithm," *Amer. Inst. Aeronaut. Astronaut. J.*, vol. 31, no. 10, pp. 1922–1928, 1993.

[16] M. Chu, H. Haussecker, and F. Zhao, "Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks," *Int. J. High Perform. Comput. Appl.*, vol. 16, no. 3, p. 293, 2002.

[17] J. Manyika and H. Durrant-Whyte, *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach*. Upper Saddle River, NJ: Prentice-Hall PTR, 1995.

[18] E. Ertin, J. Fisher, and L. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *Proc. IPSN*, 2003, vol. 3.

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

[20] S. Crary and Y. Jeong, "Bayesian optimal design of experiments for sensor calibration," in *Proc. 8th Int. Conf. Solid-State Sensors Actuators and Eurosensors IX. Transducers*, 1995, vol. 2.

[21] C. Giraud and B. Jouvencel, "Sensor selection: A geometrical approach," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 1995, vol. 2.

[22] F. Bian, D. Kempe, and R. Govindan, "Utility based sensor selection," in *Proc. 5th Int. Conf. Information Processing Sensor Networks*, Nashville, TN, 2006, pp. 11–18.

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

[24] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *Oper. Res.*, vol. 14, pp. 699–719, 1966.

[25] N. Nguyen and A. Miller, "A review of some exchange algorithms for constructing discrete D-optimal designs," *Comput. Statist. Data Anal.*, vol. 14, pp. 489–498, 1992.

[26] R. John and N. Draper, "D-Optimality for regression designs: A review," *Technometrics*, vol. 17, no. 1, pp. 15–23, 1975.

[27] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[28] F. Pukelsheim, *Optimal Design of Experiments*. Philadelphia, PA: SIAM, 2006.

[29] V. Fedorov, *Theory of Optimal Experiments*. New York: Academic, 1972.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[31] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[32] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[33] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[34] J. Feldman, D. Karger, and M. Wainwright, "LP decoding," in *Proc. 41st Allerton Conf. Communications, Control, Computing*, Oct. 2003, pp. 1–3.

[35] M. Lobo, M. Fazel, and S. Boyd, "Portfolio optimization with linear and fixed transaction costs," *Ann. Oper. Res.*, vol. 152, no. 1, pp. 341–365, 2007.

[36] A. Hassibi, J. How, and S. Boyd, "Low-authority controller design via convex optimization," in *Proc. 37th IEEE Confe. Decision Control*, 1998, vol. 1.

[37] L. Vandenberghe, S. Boyd, and A. E. Gamal, "Optimal wire and transistor sizing for circuits with non-tree topology," in *Proc. 1997 IEEE/ACM Int. Conf. Computer Aided Design*, 1997, pp. 252–259.

[38] P. Sun and R. Freund, "Computation of minimum-volume covering ellipsoids," *Oper. Res.*, vol. 52, no. 5, pp. 690–706, 2004.

[39] A. Dolia, S. Page, N. White, and C. Harris, "D-optimality for minimum volume ellipsoid with outliers," in *Proc. 7th Int. Conf. Signal/Image Processing Pattern Recognition*, 2004, pp. 73–76.

[40] A. Miller and J. Nguyen, "Algorithm AS 295: A Fedorov exchange algorithm for D-optimal design," *Appl. Statist.*, vol. 43, no. 4, pp. 669–677, 1994.

[41] H. Wynn, "Results in the theory and construction of D-optimum experimental designs," *J. Roy. Statist. Soc. Series B (Methodological)*, vol. 34, no. 2, pp. 133–147, 1972.

[42] R. Hettich and K. Kortanek, "Semi-infinite programming: Theory, methods, and applications," *SIAM Rev.*, vol. 35, no. 5, pp. 380–429, 1993.

[43] A. Mutapcic and S. Boyd, Cutting-Set Methods for Robust Convex Optimization With Worst-Case Oracles 2007 [Online]. Available: www.stanford.edu/~boyd/papers/prac robust.html

[44] Z. Luo, J. Sturm, and S. Zhang, "Multivariate nonnegative quadratic mappings," *SIAM J. Optim.*, vol. 14, no. 4, pp. 1140–1162, 2004.

[45] M. Grant, S. Boyd, and Y. Ye, CVX Version 1.1. Matlab Software for Disciplined Convex Programming 2007 [Online]. Available: www.stanford.edu/~boyd/cvx/.

Siddharth Joshi received the B. Tech. (with Honors) degree in electrical engineering from the Indian Institute of Technology, Kharagpur, India, and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 2002 and 2004, respectively. He is currently working towards the Ph.D. degree in electrical engineering at Stanford University.

His current interests include application of convex optimization to various engineering applications.

Mr. Joshi is a recipient of the Stanford Graduate Fellowship.

Stephen Boyd (S'82–M'85–SM'97–F'99) received the A.B. degree in mathematics from Harvard University, Cambridge, MA, in 1980 and the Ph.D. in electrical engineering and computer science from the University of California, Berkeley, in 1985.

He is the Samsung Professor of engineering and Professor of electrical engineering in the Information Systems Laboratory at Stanford University. His current research focus is on convex optimization applications in control, signal processing, and circuit design.