MAXIMUM LIKELIHOOD ESTIMATION UNDER PARTIAL SPARSITY CONSTRAINTS

Tirza Routtenberg[a], *Yonina C. Eldar*[b], *and Lang Tong*[a]

[a]Cornell University, Ithaca, NY 14850, New York
[b]Technion - Israel Institute of Technology, Haifa 32000, Israel

ABSTRACT

We consider the problem of estimating two deterministic vectors in a linear Gaussian model where one of the unknown vectors is subject to a sparsity constraint. We derive the maximum likelihood estimator for this problem and develop the Projected Orthogonal Matching Pursuit (POMP) algorithm for its practical implementation. The corresponding constrained Cramér-Rao bound (CCRB) on the mean-square-error is developed under the sparsity constraint. We then show that estimation in linear dynamical systems with a sparse control can be formulated as a special case of this problem.

***Index Terms*—** Sparsity, compressed sensing, maximum likelihood estimation, constrained Cramér-Rao

1. INTRODUCTION

In recent years, compressed sensing (CS) has attracted considerable attention in many areas by suggesting the possibility to surpass the traditional limits of sampling theory. CS builds upon the fundamental fact that many signals can be represented by using only a few non-zero coefficients in a suitable basis or dictionary. Nonlinear optimization can then enable the recovery of such signals from much fewer measurements than what the dimension of the unknown signal suggests [1], [2].

The estimation of a sparse vector in the presence of an unknown non-sparse vector is an important problem that arises in many applications, such as the recovery of sparsely corrupted signals [3], [4]. In [3], an ad hoc recovery algorithm was explored for the related recovery problem of sparse signal corrupted by sparse noise. Another familiar case is the problem of recovering a low rank matrix from the sum of the matrix and a sparse matrix representing errors. It was proven in [5] that under certain assumptions, both the low-rank and the sparse components can be recovered by solving a convex optimization problem. Recently, there are several works on CS in dynamic models, such as the robust smoothing algorithms developed in [6] for dynamical processes contaminated with sparse outliers.

In this paper, we consider maximum likelihood (ML) estimation of a combination of a sparse and non-sparse vector in linear models with Gaussian noise. We show that ML estimation is attained by a two-stage procedure: sparse recovery by any classical sparse method applied on the projected measurements followed by non-sparse estimation on the residual. In this work, we develop the Projected Orthogonal Matching Pursuit (POMP) algorithm in which the sparse recovery stage is performed via a variant of the Orthogonal Matching Pursuit (OMP) algorithm. An alternative suboptimal method and oracle estimator are also proposed whose performance can be used as lower and upper bounds on the mean-square-error (MSE), respectively. Another contribution of this paper is the derivation of the constrained Cramér-Rao bound (CCRB) [7] for this problem, which is an extension of the existing sparse CCRB in [8].

Finally, we show that the batch estimation of states and sparse control in dynamic systems can be formulated as the estimation of sparse and non-sparse vectors under linear models. The intuition behind the sparsity constraint is that control systems sometimes operate with an on-off control.

In the sequel, we denote vectors by boldface lowercase letters and matrices by boldface uppercase letters. The $K \times K$ identity matrix is denoted by $\mathbf{I}_K$. Let $\Lambda \subset \{1, \ldots, n\}$ be a subset of indices and $\Lambda^c = \{1, \ldots, n\} \setminus \Lambda$. By $\mathbf{x}_\Lambda$ we mean the length n vector obtained by setting the entries of $\mathbf{x}$ indexed by Λ^c to zero. $||\cdot||_0$ is the ℓ_0 seminorm, which is equal to the vector's number of nonzero entries. $\mathbf{A}_\Lambda$ denotes the submatrix of $\mathbf{A}$ made of the columns indexed by Λ and $\mathbf{A}^\dagger$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{A}$.

2. ESTIMATION METHODS

Consider the measurement model

$$\mathbf{y} = \mathbf{L}\mathbf{x} + \mathbf{S}\mathbf{u} + \mathbf{w}, \quad ||\mathbf{u}||_0 = s, \tag{1}$$

where $\mathbf{L} \in \mathbf{R}^{m\times l}$, $\mathbf{S} \in \mathbf{R}^{m\times n}$, and $\mathbf{w}$ is a zero mean Gaussian vector with covariance matrix $\sigma^2\mathbf{I}_m$. The unknown sparse and non-sparse deterministic vectors are $\mathbf{u} \in \mathbf{R}^n$ and $\mathbf{x} \in \mathbf{R}^l$, respectively, where it is assumed that $l \ll m < n$ and $||\mathbf{u}||_0 = s$. Our goal is to derive the ML estimator for both $\mathbf{u}$ and $\mathbf{x}$ from the underdetermined measurements $\mathbf{y}$. The difference between this problem and classical CS is the additional parameter of interest, $\mathbf{x}$, which is a non-sparse vector.

A related sparse recovery problem is discussed in [3], in which the model in (1) is assumed without the random noise $\mathbf{w}$. In addition, the non-sparse vector $\mathbf{x}$ is assumed to be a noise signal, i.e. a nuisance parameter. The performance of ℓ_1-recovery techniques are analyzed by using uncertainty relations and practicable recovery algorithm is provided as an ad-hoc solution. This algorithm is similar to the proposed method, where we investigate the non-Bayesian perspective of this problem and shown that this method is in fact the ML estimator.

2.1. ML estimation

We now derive the ML estimator (or equivalently, least squares) for (1) and show that it is a two-stage procedure: sparse recovery of $\mathbf{u}$ from the projected measurements and then non-sparse ML estimation of $\mathbf{x}$ from the residual.

This research was partially supported by the joint Technion-Cornell Innovation Institute under Grant NYC-3021. The work of L. Tong was partially supported by the National Science Foundation under Grant no. CNS-1135844 and the work of Y. C. Eldar was supported by the Israel Science Foundation under Grant no. 170/10

Incorporating the sparsity constraint, the constrained ML estimate of $\mathbf{x}$ and $\mathbf{u}$ from (1) is given by

$$\min_{\mathbf{x},\mathbf{u}} ||\mathbf{y}-\mathbf{L}\mathbf{x}+\mathbf{S}\mathbf{u}||_2^2, \quad s.t. \; ||\mathbf{u}||_0 \le s. \tag{2}$$

For a given s-sized index set of $\mathbf{u}$, Λ, the ML problem in (2) is equivalent to

$$\min_{\tilde{\mathbf{v}}_\Lambda} ||\mathbf{y}-\tilde{\mathbf{G}}_\Lambda\tilde{\mathbf{v}}_\Lambda||_2^2, \tag{3}$$

where $\tilde{\mathbf{G}}_\Lambda \triangleq [\mathbf{L} \vdots \mathbf{S}_\Lambda]$ and $\tilde{\mathbf{v}}_\Lambda \triangleq \left[\mathbf{x}^T, \mathbf{u}_\Lambda^T\right]^T \in \mathbb{R}^{l+s}$. Under the assumption that $\tilde{\mathbf{G}}_\Lambda$ is a full-rank matrix for any Λ, the solution to the minimization problem in (3) is given by

$$\hat{\tilde{\mathbf{v}}}_\Lambda = (\tilde{\mathbf{G}}_\Lambda^T\tilde{\mathbf{G}}_\Lambda)^{-1}\tilde{\mathbf{G}}_\Lambda^T\mathbf{y}. \tag{4}$$

By using the blockwise inversion of matrices,

$$(\tilde{\mathbf{G}}_\Lambda^T\tilde{\mathbf{G}}_\Lambda)^{-1} = \begin{bmatrix} \mathbf{L}^T\mathbf{L} & \mathbf{L}^T\mathbf{S}_\Lambda \\ \mathbf{S}_\Lambda^T\mathbf{L} & \mathbf{S}_\Lambda^T\mathbf{S}_\Lambda \end{bmatrix}^{-1} =$$
$$\begin{bmatrix} \left(\mathbf{L}^T\mathbf{L}\right)^{-1} + \mathbf{L}^\dagger\mathbf{S}_\Lambda\mathbf{K}^{-1}\mathbf{S}_\Lambda^T(\mathbf{L}^\dagger)^T & -\mathbf{L}^\dagger\mathbf{S}_\Lambda\mathbf{K}^{-1} \\ -\mathbf{K}^{-1}\mathbf{S}_\Lambda^T(\mathbf{L}^\dagger)^T & \mathbf{K}^{-1} \end{bmatrix}, \tag{5}$$

where $\mathbf{K} = \mathbf{S}_\Lambda^T\mathbf{P}_\mathbf{L}^\perp\mathbf{S}_\Lambda$, $\mathbf{L}^\dagger = \left(\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{L}^T$, and $\mathbf{P}_\mathbf{L}^\perp = \mathbf{I}_m - \mathbf{L}\mathbf{L}^\dagger$ is the orthogonal projection onto $\mathcal{R}(\mathbf{L})^\perp$. Substituting (5) in (4), we obtain the ML estimator of $\tilde{\mathbf{v}}_\Lambda$:

$$\hat{\tilde{\mathbf{v}}}_\Lambda = \begin{bmatrix} \mathbf{L}^\dagger\left(\mathbf{I}_m - \mathbf{S}_\Lambda\mathbf{K}^{-1}\mathbf{S}_\Lambda^T\mathbf{P}_\mathbf{L}^\perp\right) \\ \mathbf{K}^{-1}\mathbf{S}_\Lambda^T\mathbf{P}_\mathbf{L}^\perp \end{bmatrix}\mathbf{y}. \tag{6}$$

Therefore, the ML estimators of $\mathbf{u}$ and $\mathbf{x}$ are given by

$$\hat{\mathbf{u}}_{\hat{\Lambda}} = (\mathbf{S}_{\hat{\Lambda}}^T\mathbf{P}_\mathbf{L}^\perp\mathbf{S}_{\hat{\Lambda}})^{-1}\mathbf{S}_{\hat{\Lambda}}^T\mathbf{P}_\mathbf{L}^\perp\mathbf{y}, \tag{7}$$

and

$$\hat{\mathbf{x}} = \mathbf{L}^\dagger\left(\mathbf{y} - \mathbf{S}\hat{\mathbf{u}}\right), \tag{8}$$

respectively, where $\hat{\Lambda}$ is the ML estimated subset of nonzero indices of $\mathbf{u}$. By substituting (6) into (2), the estimator of Λ is

$$\begin{aligned} \hat{\Lambda} &= \arg\min_\Lambda ||\mathbf{y}-\mathbf{L}\hat{\mathbf{x}}-\mathbf{S}_\Lambda\hat{\mathbf{u}}_\Lambda||_2^2 \\ &= \arg\min_\Lambda ||\mathbf{P}_\mathbf{L}^\perp\mathbf{S}_\Lambda\hat{\mathbf{u}}_\Lambda||_2^2. \end{aligned} \tag{9}$$

It can be verified that the estimators in (7) and (9) produce the estimate of the sparse vector $\mathbf{u}$ from the projected observations,

$$\mathbf{P}_\mathbf{L}^\perp\mathbf{y} = \mathbf{P}_\mathbf{L}^\perp\mathbf{S}\mathbf{u}. \tag{10}$$

Therefore, ML estimation is composed of a two-stage procedure. First, $\mathbf{y}$ is projected onto $\mathcal{R}(\mathbf{L})^\perp$ by using (10). The sparse vector $\mathbf{u}$ can then be recovered from $\mathbf{P}_\mathbf{L}^\perp\mathbf{y}$ by using any standard sparse recovery method on the projected measurements in (10). Finally, $\mathbf{x}$ is estimated by using the residual $\mathbf{y} - \mathbf{S}\hat{\mathbf{u}}$, as described in (8).

2.2. The POMP method

Sparse recovery of $\mathbf{u}$ from the noisy measurements $\mathbf{P}_\mathbf{L}^\perp\mathbf{y}$ in (10) can be achieved by any existing sparse recovery method. The two main strategies for this NP-hard estimation are ℓ_1 relaxation and greedy methods, such as the OMP technique. The basic principle behind greedy algorithms is to iteratively find the support set of the sparse vector and reconstruct the signal using the restricted support ML estimate. The OMP method [9] proceeds by finding the column of the CS matrix that correlates most to the signal residual, which is obtained by subtracting the contribution of a partial estimate of the signal from the measurements. In this paper, the implementation of the two-stage ML estimation is exemplified by using OMP.

Our proposed POMP method is a modified version of the regular OMP by taking into account both the additional non-sparse vector $\mathbf{x}$ and the sparsity of $\mathbf{u}$. This method consists of two stages: first, it iteratively finds the support set of $\mathbf{u}$ by using the projected measurements $\mathbf{P}_\mathbf{L}^\perp\mathbf{y}$. At each iteration, the algorithm proceeds by finding the column of the projected CS matrix, $\mathbf{P}_\mathbf{L}^\perp\mathbf{S}$, that correlates most to the current projected signal residual, $\mathbf{P}_\mathbf{L}^\perp\mathbf{r}^{(i)}$, where i is the iteration index and $\mathbf{r}^{(i)}$ is obtained by subtracting the contribution of a partial estimate of the signal from the projected measurements. Second, the non-sparse vector $\mathbf{x}$ is estimated by substituting the final estimator $\mathbf{u}^{(i)}$ in (8). The resulting POMP algorithm is described in Table 1.

Table 1. The POMP algorithm

Initialization: Fix $i = 0$ and set the temporary solution, residual, and support to:

$$\hat{\mathbf{u}}^{(i)} = \mathbf{0}, \quad \mathbf{r}^{(i)} = \mathbf{P}_\mathbf{L}^\perp\mathbf{y}, \quad \mathcal{S}^{(i)} = \emptyset.$$

Main iteration: Increment i and apply

1. **Update support:** Find j_0 such that
$$j_0 \quad = \quad \arg\max_j \frac{\left(\mathbf{s}_j^T\mathbf{P}_\mathbf{L}^\perp\mathbf{r}^{(i-1)}\right)^2}{\mathbf{s}_j^T\mathbf{P}_\mathbf{L}^\perp\mathbf{s}_j}.$$
2. **Update solution:** Compute
$$\hat{\mathbf{u}}^{(i)} = \min_\mathbf{u} ||\mathbf{P}_\mathbf{L}^\perp\mathbf{y} - \mathbf{P}_\mathbf{L}^\perp\mathbf{S}\mathbf{u}||_2^2, \quad s.t. \; \mathbf{u} \in \mathcal{S}^{(i)}.$$
3. **Update residual:** Compute
$$\mathbf{r}^{(i)} = \mathbf{P}_\mathbf{L}^\perp\mathbf{y} - \mathbf{P}_\mathbf{L}^\perp\mathbf{S}\hat{\mathbf{u}}^{(i)}$$
4. **Stopping rule:** If $||\mathbf{r}^{(i)}||_2^2 < T_1$ stop.

Output: $\hat{\mathbf{u}} = \hat{\mathbf{u}}^{(i)}$ and $\hat{\mathbf{x}} = \mathbf{L}^\dagger\left(\mathbf{y} - \mathbf{S}\hat{\mathbf{u}}\right)$.

2.3. Comparison with other methods

In our simulations, we compare our approach with a suboptimal method and oracle estimation. The MSE of these methods can be used as an upper and lower bound on the MSE of the proposed POMP algorithm.

- Suboptimal sparse recovery: Instead of solving (2), we can solve
$$\min_\mathbf{v} ||\mathbf{y} - \mathbf{G}\mathbf{v}||_2^2, \quad s.t. \; ||\mathbf{v}||_0 \le l+s, \tag{11}$$
by using any standard sparse recovery method, where $\mathbf{G} = [\mathbf{L} \vdots \mathbf{S}]$ and $\mathbf{v} = \left[\mathbf{x}^T, \mathbf{u}^T\right]^T$. The estimate $\hat{\mathbf{x}}$ is obtained by selecting the first l elements of $\hat{\mathbf{v}}$, and an estimate of $\mathbf{u}$ is obtained by selecting the last n elements of $\hat{\mathbf{v}}$. This approach has higher computational complexity and does not utilize the prior information that the first l elements of $\mathbf{v}$ are certainly in the support of $\mathbf{v}$. Thus, the MSE of the solution of (11) can be used as an upper bound on the MSE of the solution of (2).

- Oracle ML estimation: The oracle ML estimator is based on prior knowledge of the support Λ of the sparse signal $\mathbf{u}$. With this added information, both $\mathbf{x}$ and $\mathbf{u}$ can be reconstructed by direct ML estimation via

$$\min_{\mathbf{x},\mathbf{u}} ||\mathbf{y} - \mathbf{L}\mathbf{x} - \mathbf{S}\mathbf{u}||_2^2, \quad s.t. \text{ supp}(\mathbf{u}) = \Lambda. \tag{12}$$

The solution is given by

$$\begin{aligned} \hat{\mathbf{u}}_\Lambda &= (\mathbf{S}_\Lambda^T \mathbf{P}_\mathbf{L}^\perp \mathbf{S}_\Lambda)^{-1} \mathbf{S}_\Lambda^T \mathbf{P}_\mathbf{L}^\perp \mathbf{y} && (13)\\ \hat{\mathbf{x}} &= \mathbf{L}^\dagger (\mathbf{y} - \mathbf{S}_\Lambda \hat{\mathbf{u}}_\Lambda). && (14) \end{aligned}$$

Clearly, the MSE of the ML oracle estimator is always lower than the MSE of the proposed method.

3. THE CCRB

We next derive the CCRB for the model in (1). The derivations are based on the general CCRB [7] and the more specific CCRB for sparse estimation [8].

The model in (1) can be rewritten as

$$\mathbf{y} = \mathbf{G}\mathbf{v} + \mathbf{w}, \tag{15}$$

where $\mathbf{G} = [\mathbf{L} \,\vdots\, \mathbf{S}]$ and $\mathbf{v} = \left[\mathbf{x}^T, \mathbf{u}^T\right]^T$. The CCRB for estimating $\mathbf{v}$ is given by [7]

$$\text{Cov}(\hat{\mathbf{v}}) \succeq \mathbf{U}(\mathbf{U}^T \mathbf{J} \mathbf{U})^\dagger \mathbf{U}^T, \tag{16}$$

where $\mathbf{J}$ is the Fisher information matrix (FIM) and matrix $\mathbf{U}$ spans the constraint. The CCRB in (16) is a lower bound on the MSE matrix of any constrained locally-unbiased estimator of $\mathbf{v}$ in the sense that the estimator's bias, $\mathbf{b} \stackrel{\triangle}{=} \text{E}[\hat{\mathbf{v}}] - \mathbf{v}$, satisfies

$$\frac{\partial \mathbf{b}}{\partial \mathbf{v}} \mathbf{U} = \mathbf{0}. \tag{17}$$

It can be verified that the FIM for this problem is given by

$$\mathbf{J} = \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{L}^T\mathbf{L} & \mathbf{L}^T\mathbf{S} \\ \mathbf{S}^T\mathbf{L} & \mathbf{S}^T\mathbf{S} \end{bmatrix}. \tag{18}$$

In addition, similar to the derivations in [8], a possible choice for the matrix $\mathbf{U}$ corresponding to the sparsity constraints is

$$\mathbf{U} = \begin{bmatrix} \mathbf{I}_l & \mathbf{0}_{l\times s} \\ \mathbf{0}_{n\times l} & \mathbf{U}_s \end{bmatrix}, \tag{19}$$

where $\mathbf{U}_s = [\mathbf{e}_{i1+l}, \ldots, \mathbf{e}_{is+l}] \in \mathbf{R}^{n\times s}$, in which $\mathbf{e}_j$ is the jth column of the identity matrix $\mathbf{I}_n$. By substituting (18) and (19) in (16) and using $\mathbf{S}\mathbf{U}_s = \mathbf{S}_\Lambda$, we obtain the CCRB for our problem:

$$\text{Cov}(\hat{\mathbf{v}}) \succeq \sigma^2 \mathbf{U} \left(\begin{bmatrix} \mathbf{L}^T\mathbf{L} & \mathbf{L}^T\mathbf{S}_\Lambda \\ \mathbf{S}_\Lambda^T\mathbf{L} & \mathbf{S}_\Lambda^T\mathbf{S}_\Lambda \end{bmatrix} \right)^{-1} \mathbf{U}^T. \tag{20}$$

In particular, (20) implies that the bounds on the MSEs of $\mathbf{u}$ and $\mathbf{x}$ are

$$\begin{aligned} \text{E}\left[||\hat{\mathbf{u}} - \mathbf{u}||_2^2\right] &\geq \sigma^2 \text{trace}\left(\mathbf{U}_s (\mathbf{S}_\Lambda^T \mathbf{P}_\mathbf{L}^\perp \mathbf{S}_\Lambda)^{-1} \mathbf{U}_s^T\right) \\ &= \sigma^2 \text{trace}\left((\mathbf{S}_\Lambda^T \mathbf{P}_\mathbf{L}^\perp \mathbf{S}_\Lambda)^{-1}\right) \qquad (21) \end{aligned}$$

and

$$\begin{aligned} &\text{E}\left[||\hat{\mathbf{x}} - \mathbf{x}||_2^2\right] \geq \\ &\sigma^2 \text{trace}\left((\mathbf{L}^T\mathbf{L})^{-1} + \mathbf{L}^\dagger \mathbf{S}_\Lambda (\mathbf{S}_\Lambda^T \mathbf{P}_\mathbf{L}^\perp \mathbf{S}_\Lambda)^{-1} \mathbf{S}_\Lambda^T (\mathbf{L}^\dagger)^T\right). \qquad (22) \end{aligned}$$

It can be shown that the bound in (20) is the MSE of the oracle estimator in (13)-(14).

The CCRB in (20) is a lower bound on the MSE matrix of any estimator which satisfies the condition in (17) with the matrix $\mathbf{U}$ from (19). This condition implies that $\hat{\mathbf{x}}$ should be a locally unbiased estimator in the classical sense and $\hat{\mathbf{u}}$ should be a locally unbiased estimator in the sense of [8] and [10]. For known non-sparse vector $\mathbf{x}$, the CCRB in (20) reduces to the bound in [8], as we expect.

4. APPLICATION: STATE AND SPARSE INPUT ESTIMATION FOR DYNAMICAL SYSTEMS

In this section, we consider a special case of the model (1) in the setting of dynamical systems with sparse control where our goal is to estimate both the system state and the control input. The intuition behind the sparsity constraint is that control systems sometimes operate with an on-off control. By utilizing the sparsity structure of the control input, recovery of the initial state and control can be performed from compressed measurements with lower dimension than the dimension of the unknown vectors, $\mathbf{x}$ and $\mathbf{u}$.

Consider the following MIMO state space discrete-time model

$$\Phi(\mathbf{A}, \mathbf{B}, \mathbf{C}) : \left\{ \begin{array}{c} \mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t \\ \mathbf{z}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t \end{array} \right., \quad t = 0, 1, \ldots, \tag{23}$$

where the state vector is $\mathbf{x}_t \in \mathbb{R}^l$, the input control vector is $\mathbf{u}_t \in \mathbb{R}^{K_u}$, $\mathbf{z}_t \in \mathbf{R}^{K_z}$ are the observations, and $\mathbf{x}_0 \in \mathbf{R}^l$ is the system's initial condition. The state transfer matrix, control matrix, and observation matrix are $\mathbf{A} \in \mathbf{R}^{l\times l}$, $\mathbf{B} \in \mathbf{R}^{l\times K_u}$, $\mathbf{C} \in \mathbf{R}^{K_z\times l}$, and $\mathbf{D} \in \mathbf{R}^{K_z\times K_u}$, respectively, and they are assumed to be known. Our goal is to estimate $\mathbf{x}_0$ and $\mathbf{u}$ from compressed measurements of $\mathbf{z}$, where we assume that $\mathbf{u}$ is sparse.

The recursive model in (23) can be reformulated in a "batch" form as follows:

$$\mathbf{z} \;=\; \mathcal{O}_N \mathbf{x}_0 + \mathbf{\Gamma}_N \mathbf{u}, \tag{24}$$

where

$$\mathcal{O}_N = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{N-1} \end{bmatrix},$$

$$\mathbf{\Gamma}_N = \begin{bmatrix} \mathbf{D} & \mathbf{0} & & \ldots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \ldots & \mathbf{0} \\ \vdots & & & \ldots & \mathbf{0} \\ \mathbf{CA}^{N-2}\mathbf{B} & \mathbf{CA}^{N-3}\mathbf{B} & \ldots & \mathbf{CB} & \mathbf{D} \end{bmatrix},$$

$$\mathbf{z} = [\mathbf{z}_0^T, \ldots, \mathbf{z}_{N-1}^T]^T \in \mathbf{R}^{NK_z},$$

$$\mathbf{u} = [\mathbf{u}_0^T, \ldots, \mathbf{u}_{N-1}^T]^T \in \mathbf{R}^n.$$

It can be seen that $\mathbf{\Gamma}_N \in \mathbf{R}^{NK_z\times n}$ is a block-Toeplitz lower-triangular matrix. In addition, the system matrices $\mathbf{A}$ and $\mathbf{C}$ are assumed to be full-rank matrices thus, $\mathcal{O}_N \in \mathbf{R}^{NK_z\times l}$ is a full-rank matrix. The vector $\mathbf{u}$ is assumed to be a sparse vector with, at most, s nonzero values, i.e. $||\mathbf{u}||_0 \leq s$.

The compressed signal of the measurements vector $\mathbf{z}$ from (24) is given by

$$\mathbf{y} \quad = \quad \mathbf{\Psi z} = \mathbf{L x}_0 + \mathbf{S u}, \tag{25}$$

where the CS matrix is $\mathbf{\Psi} \in \mathbf{R}^{m \times NK_z}$, $\mathbf{L} = \mathbf{\Psi}\mathcal{O}_N \in \mathbf{R}^{m \times l}$, $\mathbf{S} = \mathbf{\Psi}\mathbf{\Gamma}_N \in \mathbf{R}^{m \times n}$, and $n = NK_u$. The model in (25), with the addition of Gaussian noise, is identical to (1) with $\mathbf{x}_0 = \mathbf{x}$.

From a dynamic system perspective, the problem considered here is slightly different than the standard Kalman filtering-based state estimation, since the estimation problem includes not only estimating the states but also the sparse control.

5. SIMULATIONS

We consider the state space model from (24) with the initial state $\mathbf{x}_0 = [78, 70]^T$, $N = 100$ samples, $K_u = 1$, and the control vector satisfies $||\mathbf{u}||_0 = 3$. The index set Λ was randomly chosen uniformly and the values were $\mathbf{u}_\Lambda = [12, 10, 5]^T$. The system matrices were

$$\mathbf{A} = \begin{bmatrix} 0.98 & 0 \\ 0 & 0.9 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \ \mathbf{C} = \mathbf{I}_2, \ \mathbf{D} = \begin{bmatrix} 0.75 \\ 0.75 \end{bmatrix}.$$

The CS matrix, $\mathbf{\Psi}$, was a random Gaussian matrix and only 70% of the measurments were used. The performance was evaluated using 5000 Monte-Carlo simulations.

The root MSE (RMSE) of the estimators of $\mathbf{x}_0$ and $\mathbf{u}$ by using the POMP algorithm, suboptimal OMP-type algorithm, and oracle estimate compared to the proposed CCRB in (21)-(22) are presented in Fig. 1 versus signal-to-noise ratio (SNR), where the SNR is defined as SNR $= \frac{||\mathbf{L}||_2^2 + ||\mathbf{S}||_2^2}{\sigma^2}$. It can be seen that the proposed algorithm achieves good results and asymptotically attains the corresponding CCRB.

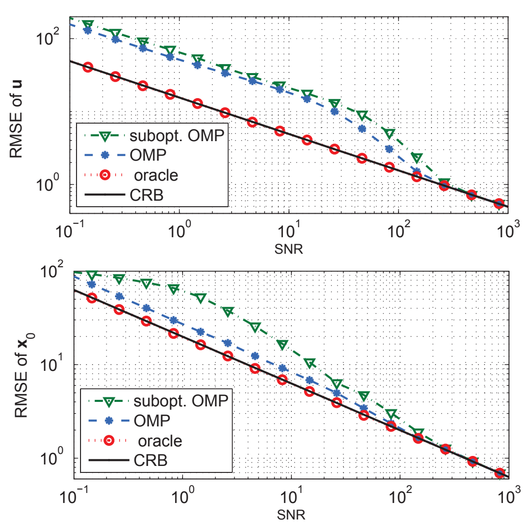

Fig. 1. The RMSE of the estimators of $\mathbf{u}$ (upper) and $\mathbf{x}_0$ (lower) vs. SNR for $s = 3$, $N = 100$.

Fig. 2 shows the probability of correct support detection for known sparsity order for both the proposed POMP and the suboptimal algorithms. This probability is given by

$$P_d = \frac{\#(\text{supp}(\mathbf{u}) \bigcap \text{supp}(\hat{\mathbf{u}}))}{\#(\text{supp}(\mathbf{u}))},$$

where for the suboptimal method also the probability of $\mathbf{x}_0$-support detection is presented. It can be seen that the probability of detection is higher for the POMP method at any SNR and that at high SNR this probability approaches 1 for both methods.

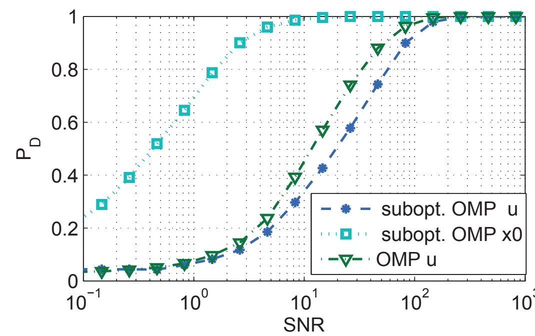

Fig. 2. Probability of support detection vs. SNR for $s = 3$ and $N = 100$.

6. CONCLUSION

We presented the problem of joint sparse and non-sparse estimation in linear models with Gaussian noise. The corresponding ML estimator has been shown to satisfy a separation principle, i.e. it can be separated into two stages: sparse recovery followed by non-sparse estimation. The ML estimation is implemented via the proposed POMP method. In addition, the CCRB that we developed for this problem is shown to equal the MSE of the oracle estimator. Also, the batch estimation of states and the input in a dynamic system with a sparse input is shown to be equivalent to this estimation problem. Finally, the performance of the proposed methods and new CCRB are presented for the problem of control and states estimation in dynamical systems. In case of high SNR, the MSE of the proposed POMP algorithm is shown to attain the CCRB.

7. REFERENCES

[1] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," *Compressed Sensing: Theory and Applications*, Edited by Y. C. Eldar and G. Kutyniok, Cambridge University Press, 2012.

[2] M. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, pp. 4053–4085, Sept. 2011.

[3] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inform. Theory*, vol. 58, pp. 3115–3130, May 2012.

[4] J. Laska, M. Davenport, and R. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the Pursuit of Justice," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, pp. 1556–1560, Nov. 2009.

[5] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?,” *J. ACM*, vol. 58, no. 3, p. 11, 2011.

[6] S. Farahmand, G. Giannakis, and D. Angelosante, “Doubly robust smoothing of dynamical processes via outlier sparsity constraints,” *IEEE Trans. Signal Process.*, vol. 59, pp. 4529–4543, Oct. 2011.

[7] J. Gorman and A. Hero, “Lower bounds for parametric estimation with constraints,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 1285–1301, Nov. 1990.

[8] Z. Ben-Haim and Y. C. Eldar, “The Cramér-Rao bound for estimating a sparse parameter vector,” *IEEE Trans. Signal Process*, vol. 58, pp. 3384 –3389, June 2010.

[9] S. Mallat and Z. Zhang, “Matching pursuit with time-frequency dictionaries,” *IEEE Trans. Signal Process*, vol. 41, pp. 3397–3415, Dec. 1993.

[10] T. Routtenberg and J. Tabrikian, “Performance bounds for constrained parameter estimation,” in *IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 513–516, 2012.