CONDITIONAL PRIOR BASED LMMSE ESTIMATION OF SPARSE SIGNALS

Martin Sundin, Magnus Jansson, Saikat Chatterjee

ACCESS Linnaeus Center, Electrical Engineering
KTH Royal Institute of Technology, Stockholm, Sweden

masundi@kth.se, magnus.jansson@ee.kth.se, sach@kth.se

ABSTRACT

We derive a linear minimum mean square error estimator for sparse vector estimation from an underdetermined set of linear equations. The derivation of the estimator uses a prior distribution conditioned on the support set of the underlying sparse vector. The estimator is used in the architecture of the standard orthogonal matching pursuit algorithm to achieve a better performance.

Index Terms— Bayesian estimation, minimum mean square error estimation, sparse estimation, compressed sensing.

1. INTRODUCTION

In recent years the problem of estimating a sparse vector from an underdetermined set of linear measurements has recieved considerable interest. In this paper we consider the application area of *Compressed Sensing* (CS) [1]. Let us state the standard CS problem where we estimate a sparse signal $\mathbf{x} \in \mathbb{R}^n$ from the linear measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a matrix representing the sampling system, $\mathbf{y} \in \mathbb{R}^m$ represents a vector of measurements and $\mathbf{w} \in \mathbb{R}^m$ is additive noise representing measurement errors. A K-sparse signal vector consists of at most K non-zero scalar components. We consider the setup $K < m < n$ (underdetermined system of linear equations), the task is to reconstruct $\mathbf{x}$ from $\mathbf{y}$ as $\hat{\mathbf{x}}$. For a sparse signal vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^\top$, the support set $\mathcal{I} \subset \{1, 2, \ldots, n\}$ is defined as $\mathcal{I} = supp(\mathbf{x}) = \{i : x_i \neq 0\}$. For a K-sparse vector $\mathbf{x} \in \mathbb{R}^n$, $|\mathcal{I}| = \|\mathbf{x}\|_0 \leq K$, where $|\mathcal{I}|$ denotes the cardinality of $\mathcal{I}$. In this paper, we assume that $|\mathcal{I}| = K$. For CS reconstruction, three main approaches are used in the literature: (1) convex relaxations [1, 2, 3, 4, 5], (2) Bayesian methods [6, 7, 8, 9] and (3) greedy searches [10, 11, 12]. Among these three approaches, greedy search solutions are computationally simpler. In this paper, we endeavor to improve greedy search methods using Bayesian estimation principles.

In greedy searches, the main tasks are finding the support-set $\mathcal{I}$ of the underlying sparse vector $\mathbf{x}$ and then estimating the non-zero values. For example, the orthogonal matching pursuit (OMP) [10] algorithm detects the non-zero coordinates sequentially. OMP uses matched filter based detection and least-squares based estimation. The standard OMP does not use a Bayesian estimation framework.

Let us consider a scenario where the statistics $p(\mathbf{x})$ of the unknown sparse vector $\mathbf{x}$ is known a piori. In that case, improvement of greedy search algorithms using the Bayesian framework remains as a challenging task. In this pursuit, [13] incorporates the minimum mean square (MMSE) estimate of the non-zero coefficients where it is assumed that the non-zero coefficients are iid Gaussian. The technique of [13] only improves signal value estimation, but not the non-zero coefficient detection. It is interesting to use the MMSE and linear MMSE (LMMSE) estimator for CS reconstruction. The standard MMSE estimator can be found by minimizing the Mean Square Error (MSE) with respect to the posterior distribution $p(\mathbf{x}|\mathbf{y})$. The MMSE estimator was pursued in [7, 8] and we note that the MMSE estimate is computationally intractable in general due to an exhaustive search requirement over all possible support sets. To achieve a practical solution, approximations are important. In this regard, we note that [7] provides an approximation of the standard MMSE estimator by a truncated version. We mention that a convex method for estimating sparse vectors from noisy linear measurements may use ℓ_1-norm penalized least squares, such as in standard LASSO algorithm [2, 6, 5]. The ℓ_1-penalized estimator can be considered as the Maximum A Posteriori estimator when the source has a Laplacian prior distribution.

In this paper, we develop a LMMSE estimator where we minimize the MSE of $\hat{x}_i$ with respect to the conditional prior distribution $p(\mathbf{x}|i \in \mathcal{I})$. This is different from the standard LMMSE estimator which minimizes the MSE with respect to the prior distribution $p(\mathbf{x})$. We use the conditional prior LMMSE estimator to modify standard greedy algorithms, for example the OMP and Projection based OMP (POMP) [14] algorithms. The new algorithms are referred to as Conditional Prior based OMP (CpOMP) and Conditional Prior based POMP (CpPOMP). Through simulations, we evaluate the performance of the CpOMP and CpPOMP algorithms showing that they provide better performance than standard OMP and POMP in terms of lower MSE and support cardinality error.

Notation: In this paper we write $\mathbf{I}_p$ for the $p \times p$ identity matrix. We let $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$ have column vectors of unit length and use $\mathbf{A}_\mathcal{I} \in \mathbb{R}^{m\times|\mathcal{I}|}$ to denote the matrix consisting of the columns of $\mathbf{A}$ indexed by $i \in \mathcal{I}$. We use $(.)^\top$ and $(.)^{-1}$ for the matrix transpose and inverse repectively. The expectation value of a random variable is denoted by $E[.]$ and $\mathbf{z}|H$ is used to for the random variable $\mathbf{z}$ conditioned on H, i.e. $\mathbf{z}|H \sim p(\mathbf{z}|H)$. We use $\emptyset$ to denote the empty set.

2. LMMSE ESTIMATION OF SPARSE VECTORS

The LMMSE estimator or Wiener filter is one of the fundamental tools in signal processing. It is used in a multitude of applications and is described in several textbooks [15, 16]. The LMMSE estimator of a random variable x from observations of a random variable $\mathbf{y}$, $\hat{x} = \mathbf{b}^\top\mathbf{y} + c$, is derived by minimizing the MSE

$$\text{MSE} = E[(x - \hat{x})^2]$$

This work was partially supported by the Swedish Research Council under contract 621-2011-5847.

with respect to $\mathbf{b}$ and c, where the expectation value is taken with respect to the (joint) probability distribution $p(x, \mathbf{y})$ of x and $\mathbf{y}$. The LMMSE estimator is found to be

$$\hat{x} = E[x] + \mathbf{C}(x, \mathbf{y})\mathbf{C}(\mathbf{y})^{-1}(\mathbf{y} - E[\mathbf{y}]) \tag{2}$$

where $\mathbf{C}(x, \mathbf{y}) = \text{Cov}(x, \mathbf{y}) = E[x\mathbf{y}^\top] - E[x]E[\mathbf{y}^\top]$ and $\mathbf{C}(\mathbf{y}) = \text{Cov}(\mathbf{y}, \mathbf{y})$.

For random sparse vectors the standard MMSE estimator of $\mathbf{x}$ given $\mathbf{y}$ in (1) is given by [7, 8]

$$\hat{\mathbf{x}} = E[\mathbf{x}|\mathbf{y}] = \sum_{|\mathcal{I}|=K} p(\mathcal{I}|\mathbf{y})\hat{\mathbf{x}}|\mathcal{I}$$

where $\hat{\mathbf{x}}|\mathcal{I}$ is the MMSE estimator of $\mathbf{x}$ given that $supp(\mathbf{x}) = \mathcal{I}$. The probability $p(\mathcal{I}|\mathbf{y})$ can be computed using Bayes theorem

$$p(\mathcal{I}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathcal{I})p(\mathcal{I})}{p(\mathbf{y})}$$

where $p(\mathcal{I}) = \text{Pr}(supp(\mathbf{x}) = \mathcal{I})$. When $\mathbf{x}_\mathcal{I} \sim N(\mathbf{0}, \sigma_x^2\mathbf{I}_K)$ and $\mathbf{w} \sim N(\mathbf{0}, \sigma_w^2\mathbf{I}_m)$ are uncorellated, we find that

$$p(\mathbf{y}|\mathcal{I}) \propto det(\mathbf{C}(\mathbf{y}|\mathcal{I}))^{-1/2} e^{-\frac{1}{2}\mathbf{y}^\top\mathbf{C}(\mathbf{y}|\mathcal{I})^{-1}\mathbf{y}}$$

where $\mathbf{C}(\mathbf{y}|\mathcal{I}) = \sigma_x^2\mathbf{A}_\mathcal{I}\mathbf{A}_\mathcal{I}^\top + \sigma_w^2\mathbf{I}_m$. We see that the MMSE estimator is computationally intractable in general since the probability $p(\mathbf{y}|\mathcal{I})$ needs to be evaluated for all support sets.

3. CONDITIONAL PRIOR AND LMMSE

The main difficulty of sparse estimation is to determine whether a component is in the support set or not. In this section, we develop an LMMSE estimator under a conditional prior distribution $p(\mathbf{x}|H_i)$, where H_i is a hypothesis about the support set. After estimation we then choose the hypothesis which best describes the measurements. An exhaustive search corresponds to setting $H_i = \{\mathcal{I}_i = \mathcal{I}\}$ for $i = 1, 2, \ldots, \binom{n}{K}$. Since an exhaustive search is computationally intractable for large systems we instead consider the least restrictive hypothesis possible, namely $H_i = \{i \in \mathcal{I}\}$ for $i = 1, 2, \ldots, n$. Under the hypothesis $i \notin \mathcal{I}$, the MMSE estimator of x_i is the trivial estimator

$$\hat{x}_i|(i \notin \mathcal{I}) = 0$$

To find the LMMSE estimator for $\hat{x}_i|(i \in \mathcal{I})$, we need to take the distributions $p(\mathcal{I})$ and $p(\mathbf{x}|\mathcal{I})$ into account.

3.1. Noise free and uncorrelated signals

For simplicity we first consider the noise-free case with uniformly distributed support sets, i.e. $p(\mathcal{I}) = \binom{n}{K}^{-1}$, $\mathbf{w} = \mathbf{0}$. We let the components of $\mathbf{x}_\mathcal{I}|\mathcal{I}$ be uncorrelated with zero mean and variance σ_x^2. Under the hypothesis $i \in \mathcal{I}$ the MSE of the linear estimator $\hat{x}_i|(i \in \mathcal{I}) = \mathbf{b}^\top\mathbf{y}$ becomes

$$\begin{aligned}\text{MSE}|(i \in \mathcal{I}) &= E\left[(x_i - \mathbf{b}^\top\mathbf{y})^2 | i \in \mathcal{I}\right] \\ &= E_\mathcal{I}\left[(1 - \mathbf{b}^\top\mathbf{a}_i)^2\sigma_x^2 + \sum_{j\in\mathcal{I}\setminus\{i\}} (\mathbf{b}^\top\mathbf{a}_j)^2\sigma_x^2 \middle| i \in \mathcal{I}\right] \\ &= (1 - \mathbf{b}^\top\mathbf{a}_i)^2\sigma_x^2 + \sum_{j\neq i}(\mathbf{b}^\top\mathbf{a}_j)^2\sigma_x^2\text{Pr}(j \in \mathcal{I}|i \in \mathcal{I})\end{aligned} \tag{3}$$

where we used that $E[.] = E_\mathcal{I}[E_{\mathbf{x}|\mathcal{I}}[.]]$. We find that

$$\text{Pr}(j \in \mathcal{I}|i \in \mathcal{I}) = \frac{\binom{n-2}{K-2}}{\binom{n-1}{K-1}} = \frac{K-1}{n-1} = \rho_1$$

for $j \neq i$. Now we state the following proposition.

Proposition 1. *Minimizing* (3) *with the condition* $K > 1$ *gives the LMMSE estimator*

$$\hat{x}_i|(i \in \mathcal{I}) = \frac{\mathbf{a}_i^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{y}}{\rho_1 + (1-\rho_1)\mathbf{a}_i^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{a}_i}. \tag{4}$$

Proof. The Mean Square Error $\text{MSE}|(i \in \mathcal{I})$ can be written as

$$\begin{aligned}\text{MSE}|(i \in \mathcal{I}) &= \sigma_x^2 - 2\sigma_x^2\mathbf{b}^\top\mathbf{a}_i \\ &+ \sigma_x^2\mathbf{b}^\top\left((1-\rho_1)\mathbf{a}_i\mathbf{a}_i^\top + \rho_1\mathbf{A}\mathbf{A}^\top\right)^{-1}\mathbf{b}\end{aligned}$$

Minimizing with respect to $\mathbf{b}$ we find that

$$\mathbf{b} = \left((1-\rho_1)\mathbf{a}_i\mathbf{a}_i^\top + \rho_1\mathbf{A}\mathbf{A}^\top\right)^{-1}\mathbf{a}_i$$

Using the Sheerman-Morrison-Woodbury formula [17] gives us that

$$\begin{aligned}\hat{x}_i|(i \in \mathcal{I}) =& \mathbf{a}_i^\top\left((1-\rho_1)\mathbf{a}_i\mathbf{a}_i^\top + \rho_1\mathbf{A}\mathbf{A}^\top\right)^{-1}\mathbf{y} \\ =& \mathbf{a}_i^\top(\rho_1\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{y} \\ &- (1-\rho_1)\frac{\mathbf{a}_i^\top(\rho_1\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{a}_i\mathbf{a}_i^\top(\rho_1\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{y}}{1 + (1-\rho_1)\mathbf{a}_i(\rho_1\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{a}_i} \\ =& \frac{\mathbf{a}_i^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{y}}{\rho_1 + (1-\rho_1)\mathbf{a}_i(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{a}_i}\end{aligned}$$

This proves the proposition. □

We see that the estimator (4) converges to the LMMSE estimator with unconditional prior in the limit $\rho_1 \to 1$, i.e. $\hat{x}_i \to \mathbf{e}_i^\top\mathbf{A}^\dagger\mathbf{y}$ where $(.)^\dagger$ denotes the Moore-Penrose pseudoinverse. In the limit $\rho_1 \to 0$, the estimator converges to $\hat{x}_i = \mathbf{e}_i^\top\mathbf{A}^\dagger\mathbf{y}/\mathbf{a}_i^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{a}_i$. This estimator was earlier derived in [18], using an alternative approach, where it was used only in the detection step of OMP. For $K = 1$, the minima of (3) is non-unique. The solution minimizing $||\mathbf{b}||_2$ is given by

$$\hat{x}_i|(\{i\} = \mathcal{I}) = \mathbf{a}_i^\top\mathbf{y}$$

In each iteration of OMP, one new element is included in the estimated support set. Given a partially estimated support set $\mathcal{I}_s$, with $|\mathcal{I}_s| = s \leq K$, OMP assumes that $\mathcal{I}_s \subset \mathcal{I}$ and estimates the components of $\mathbf{x}_{\mathcal{I}_s}$. The LMMSE estimator of $\mathbf{x}_{\mathcal{I}_s}$ under the hypothesis $\mathcal{I}_s \subset \mathcal{I}$ can be derived in the same way as the estimator (4) was derived. The MSE of $\hat{\mathbf{x}}_{\mathcal{I}_s}|(\mathcal{I}_s \subset \mathcal{I}) = \mathbf{B}_s^\top\mathbf{y}$ is given by

$$\begin{aligned}\text{MSE}|(\mathcal{I}_s \subset \mathcal{I}) =& E\left[\left|\left|\mathbf{x}_{\mathcal{I}_s} - \mathbf{B}_s^\top\mathbf{y}\right|\right|_2^2 | \mathcal{I}_s \subset \mathcal{I}\right] \\ =& \sigma_x^2||\mathbf{I}_s - \mathbf{B}_s^\top\mathbf{A}_{\mathcal{I}_s}||_F^2 \\ &+ \sigma_x^2\sum_{j\notin\mathcal{I}_s}||\mathbf{B}_s^\top\mathbf{a}_j||_2^2\text{Pr}(j \in \mathcal{I}|\mathcal{I}_s \subset \mathcal{I})\end{aligned} \tag{5}$$

where

$$\text{Pr}(j \in \mathcal{I}|\mathcal{I}_s \subset \mathcal{I}) = \frac{\binom{n-s-1}{K-s-1}}{\binom{n-s}{K-s}} = \frac{K-s}{n-s} = \rho_s$$

for $j \notin \mathcal{I}_s$. Minimizing (5) with respect to $\mathbf{B}_s$ we find the LMMSE estimator

$$\hat{\mathbf{x}}_{\mathcal{I}_s}|(\mathcal{I}_s \subset \mathcal{I}) = \mathbf{A}_{\mathcal{I}_s}^\top \left((1-\rho_s)\mathbf{A}_{\mathcal{I}_s}\mathbf{A}_{\mathcal{I}_s}^\top + \rho_s \mathbf{A}\mathbf{A}^\top\right)^{-1} \mathbf{y} \quad (6)$$

The estimator (6) can be simplified using matrix inverse identities [17] to lower the cost of computing the estimate. We see that $\rho_K = 0$, so

$$\hat{\mathbf{x}}_{\mathcal{I}_K}|(\mathcal{I}_K = \mathcal{I}) = \mathbf{A}_{\mathcal{I}_K}^\dagger \mathbf{y}$$

Thus for this scenario, the Bayesian estimator becomes the usual least square estimator in the final iteration.

3.2. Noisy and correlated signals

When the components of $\mathbf{x}_\mathcal{I}|\mathcal{I}$ are zero mean and correlated, covariance matrices

$$\mathbf{C}(\mathbf{u},\mathbf{v}|\mathcal{I}_s) = E[\mathbf{u}\mathbf{v}^\top|\mathcal{I}_s \subset \mathcal{I}] = \sum_{\mathcal{I},|\mathcal{I}|=K} p(\mathcal{I}|\mathcal{I}_s \subset \mathcal{I}) E[\mathbf{u}\mathbf{v}^\top|supp(\mathbf{x}) = \mathcal{I}]$$

are non-diagonal in general. For the linear estimator $\hat{\mathbf{x}}_{\mathcal{I}_s}|(\mathcal{I}_s \subset \mathcal{I}) = \mathbf{B}^\top\mathbf{y}$, the MSE becomes

$$\text{MSE}|(\mathcal{I}_s \subset \mathcal{I}) = E\left[||\mathbf{x}_{\mathcal{I}_s} - \mathbf{B}^\top\mathbf{y}||_2^2|\mathcal{I}_s \subset \mathcal{I}\right] \quad (7)$$
$$= tr(\mathbf{C}(\mathbf{x}_{\mathcal{I}_s}|\mathcal{I}_s)) + tr(\mathbf{B}^\top\mathbf{C}(\mathbf{y}|\mathcal{I}_s)\mathbf{B}) - 2tr(\mathbf{B}^\top\mathbf{C}(\mathbf{y},\mathbf{x}_{\mathcal{I}_s}|\mathcal{I}_s))$$

where $\mathbf{C}(\mathbf{y}|\mathcal{I}_s) = \mathbf{A}\mathbf{C}(\mathbf{x}|\mathcal{I}_s)\mathbf{A}^\top + \mathbf{C}(\mathbf{w})$ is the conditional covariance of $\mathbf{y}$, $\mathbf{C}(\mathbf{y},\mathbf{x}_{\mathcal{I}_s}|\mathcal{I}_s) = \mathbf{A}\mathbf{C}(\mathbf{x},\mathbf{x}_{\mathcal{I}_s}|\mathcal{I}_s)$ and we used the notation $\mathbf{C}(\mathbf{u}|\mathcal{I}_s) = \mathbf{C}(\mathbf{u},\mathbf{u}|\mathcal{I}_s)$. Minimizing (7) with respect to $\mathbf{B}$ we find the LMMSE estimator

$$\hat{\mathbf{x}}_{\mathcal{I}_s}|(\mathcal{I}_s \subset \mathcal{I}) = \mathbf{C}(\mathbf{x}_{\mathcal{I}_s},\mathbf{y}|\mathcal{I}_s)\mathbf{C}(\mathbf{y}|\mathcal{I}_s)^{-1}\mathbf{y}. \quad (8)$$

Expressing the estimator as (8) makes the similarity with (2) apparent.

Remark: Although we have chosen to concentrate on the underdetermined setup ($m < n$), the results of this paper can also be applied to sparse problems with an overdetermined setup ($m \geq n$).

4. CONDITIONAL PRIOR BASED GREEDY ALGORITHMS

Based on the structure of the OMP algorithm we develop the Conditional Prior based OMP (CpOMP) algorithm by using the conditional prior LMMSE estimator (4) in the detection step and (6) in the estimation step. In the implementation we assume that $\mathbf{x}$ and $\mathbf{w}$ are zero mean and uncorrelated, the non-zero components of $\mathbf{x}$ are iid and that the noise is white. The CpOMP algorithm is summarized in Algorithm 1.

Another modification of the OMP algorithm is the Projection-Based OMP (POMP) described in [14]. POMP improves the performance of OMP at the cost of higher complexity. In each iteration POMP includes the L components with largest amplitude of $\mathbf{A}^\top\mathbf{r}$ in an intermediate support set, i.e.

$$\mathcal{J} = \hat{\mathcal{I}} \cup \{L \text{ largest components of } |\mathbf{A}^\top\mathbf{r}|\}$$

where the integer L is a user defined parameter. The vector $\mathbf{x}_\mathcal{J}$ is then estimated using least square minimization. In the final stage only the element of largest amplitude of $\mathcal{J}\backslash\hat{\mathcal{I}}$ is included in the estimated support set $\hat{\mathcal{I}}$, for more details see [14]. POMP can be adapted to random sparse signals using the Conditional Prior based framework, i.e. we replace the detection and estimation steps of POMP by the steps 3. and 5. of the CpOMP algorithm with $\rho_s = \max(0, K-s)/(n-s)$. We refer to the Conditional Prior based version of POMP as CpPOMP.

Algorithm 1 CpOMP: Conditional Prior based OMP

Input: $\mathbf{y} \in \mathbb{R}^m, \mathbf{A} \in \mathbb{R}^{m\times n}, K, \gamma = SNR^{-1} = \sigma_w^2/\sigma_x^2$

Initialization:

1: $\mathbf{r} = \mathbf{y}, \hat{\mathcal{I}} = \emptyset, \hat{\mathbf{x}} = \mathbf{0}, \mathbf{D} = \left(\mathbf{A}\mathbf{A}^\top + \gamma\mathbf{I}_m\right)^{-1}$.

Iteration:

1: **repeat**
2: $\rho = (K - |\hat{\mathcal{I}}| - 1)/(n - |\hat{\mathcal{I}}| - 1)$
3: $i = \arg\max_{j\notin\hat{\mathcal{I}}} \left|\mathbf{a}_j^\top\mathbf{D}\mathbf{r}\right| / (\rho + (1-\rho)\mathbf{a}_j^\top\mathbf{D}\mathbf{a}_j)$
4: $\hat{\mathcal{I}} \leftarrow \hat{\mathcal{I}} \cup \{i\}$
5: $\hat{\mathbf{x}}_{\hat{\mathcal{I}}} = \mathbf{A}_{\hat{\mathcal{I}}}^\top\left((1-\rho)\mathbf{A}_{\hat{\mathcal{I}}}\mathbf{A}_{\hat{\mathcal{I}}}^\top + \rho\mathbf{A}\mathbf{A}^\top + \gamma\mathbf{I}_m\right)^{-1}\mathbf{y}$
6: $\mathbf{r} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$
7: **until** $|\hat{\mathcal{I}}| = K$ *or* $||\mathbf{r}||_2 \leq \epsilon$.

Output: $\hat{\mathcal{I}}, \hat{\mathbf{x}}$

5. COMPUTATIONAL COMPLEXITY

The naive implementation of CpOMP has a higher computational complexity, i.e. runtime, than the naive implementation of OMP. Assuming that the algorithms terminates when $|\mathcal{I}| = K$, OMP requires the computation of $\mathcal{O}(nK)$ correlations requiring $\mathcal{O}(nmK)$ operations and solving K least squares problems requiring $\mathcal{O}(mK^3)$ operations in total. So the complexity of OMP is $\mathcal{O}(mK \cdot \max(K^2, n))$.

CpOMP requires computing the matrix $\mathbf{A}^\top\mathbf{D}$ requiring $\mathcal{O}(n^2m)$ operations, correlations requiring $\mathcal{O}(nmK)$ operations and computing K estimates $\hat{\mathbf{x}}_{\hat{\mathcal{I}}}$ requiring $\mathcal{O}(Km^3)$ operations in total. The complexity of CpOMP is thus $\mathcal{O}(m \cdot \max(n^2, Km^2))$.

We should mention that faster implementations of OMP and other pursuit algorithms exists, see e.g. [19] and references therein.

6. NUMERICAL EVALUATION

To compare the algorithms we numerically evaluated their performance using Monte-Carlo simulations. The simulation was performed as follows.

1. Generate a $m \times n$ matrix $\mathbf{A}$ by drawing its components from $N(0, 1)$ and normalize the column vectors.
2. Draw a subset $\mathcal{I} \subset \{1, 2, \ldots, n\}$ with K elements uniformly at random.
3. Generate the signal vector $\mathbf{x} \in \mathbb{R}^n$ by setting $\mathbf{x}_{\mathcal{I}^c} = \mathbf{0}$ and draw $\mathbf{x}_\mathcal{I}$ from $N(\mathbf{0}, \sigma_x^2\mathbf{I}_K)$. Draw $\mathbf{w}$ from $N(\mathbf{0}, \sigma_w^2\mathbf{I}_m)$.
4. Calculate the measurement $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$.

From $\mathbf{y}$ we then estimate $\mathbf{x}$ using OMP, POMP, Subspace Pursuit (SP) [11], Gaussian based Matching Pursuit (GMP) [13] CpOMP and CpPOMP. We compared the performance by empirically evaluating two performance measures: the Normalized Mean Square Error (NMSE)

$$\text{NMSE} = \frac{E[||\hat{\mathbf{x}} - \mathbf{x}||_2^2]}{E[||\mathbf{x}||_2^2]}$$

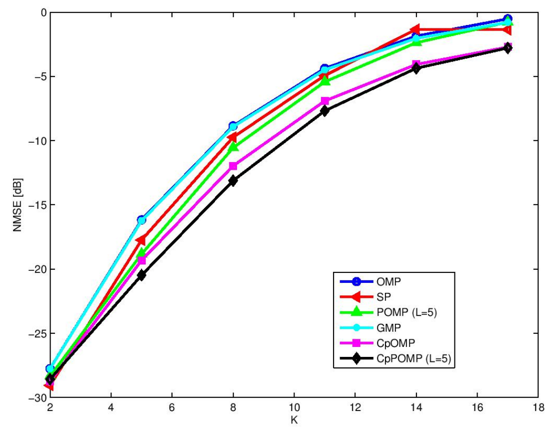

Fig. 1. NMSE [dB] for SNR = 20 dB.

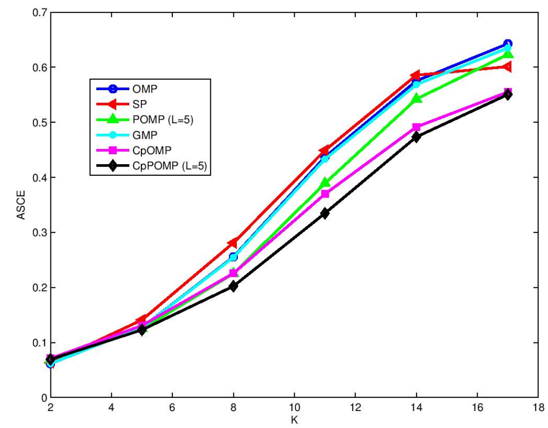

Fig. 2. ASCE for SNR = 20 dB.

and the Average Support Cardinality Error (ASCE)

$$\text{ASCE} = 1 - \frac{E[|\hat{\mathcal{I}} \cap \mathcal{I}|]}{|\mathcal{I}|}.$$

Each of the performance measures show different aspects of the sparse estimation problem. Which measure to be preferred should ultimately depend on the scenario or application where the methods are applied. In the simulations we kept the Signal-to-Noise-Ratio (SNR)

$$\text{SNR} = \frac{E[||\mathbf{Ax}||_2^2]}{E[||\mathbf{w}||_2^2]} = \frac{K\sigma_x^2}{m\sigma_w^2}$$

fixed. We generated $T = 50$ random 30×100 matrices. For each matrix and each value of K we estimated $N = 500$ random signals from measurements. For varying SNR and $\alpha = m/n$ we fixed the sparsity to $K = 11$.

Figure 1 and 2 shows the NMSE and ASCE for different K and SNR $= 20$ dB. We observe that CpOMP and CpPOMP improves the NMSE by 2.5 dB for $K \geq 6$ and the ASCE by 5% for $K \geq 10$ compared to their non-Bayesian counterparts OMP and POMP. In total CpPOMP achieved an improvement of 4 dB in NMSE and 8 $\%$ in ASCE compared to the standard OMP.

The performance for noise free measurements is shown in figure 3 and 4. CpOMP and CpPOMP improves the NMSE with more than 3.2 dB and the ASCE by more than 3% for $K \leq 10$ over OMP and POMP. The total improvement of CpPOMP over OMP was 4.6 dB in NMSE for and 9% in ASCE for $K \geq 10$.

Figure 5 and 6 shows the NMSE for varying SNR and $\alpha = \frac{m}{n}$ for $K = 11$. We observe that the performance gain of CpOMP and CpPOMP over OMP and POMP increases with SNR. The NMSE is lower by 3 dB for high SNR. In figure 6 we see that CpOMP and CpPOMP have lower NMSE than OMP and POMP for $\alpha \leq 0.4$, i.e. the proposed bayesian methods only perform better than the deterministic methods when the fraction of measurements is less than 40%.

7. CONCLUSION

In this paper we derived a LMMSE estimator of random sparse vectors by minimizing the MSE with respect to a conditional prior distribution. We used the LMMSE estimator to modify the OMP and POMP algorithms. Through numerical simulations it was shown that the modified algorithms improved the empirical performance of the algorithms.

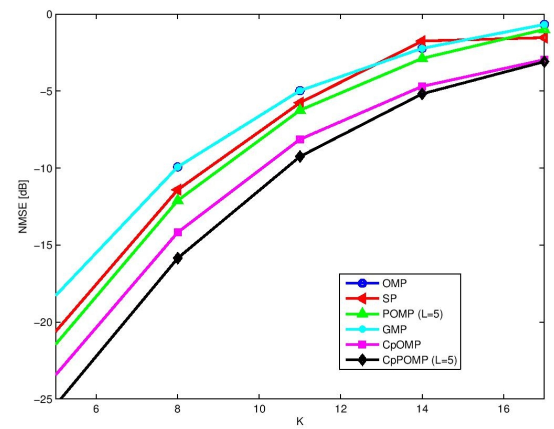

Fig. 3. NMSE [dB] for noise-free measurements.

8. ACKNOWLEDGEMENTS

The authors want to thank the anonymous reviewers for useful comments and remarks.

9. REFERENCES

[1] D.L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[2] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.

[3] E.J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete fre-

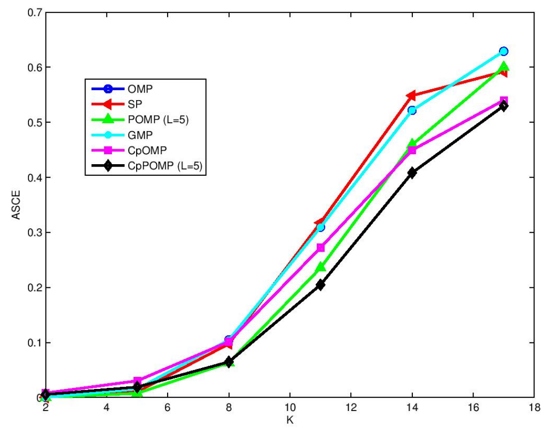

Fig. 4. ASCE for noise-free measurements.

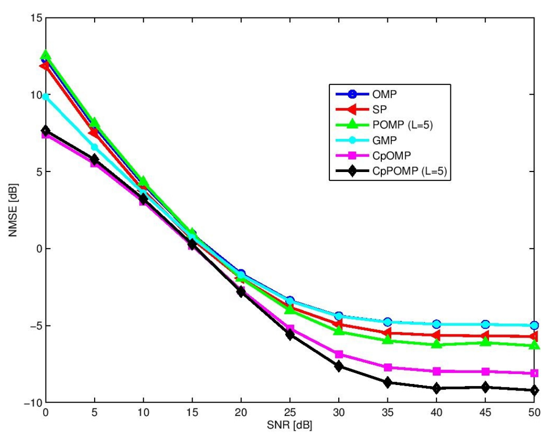

Fig. 5. NMSE [dB] versus SNR for $K = 11$.

quency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] E.J. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[5] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[6] Shihao Ji, Ya Xue, and L. Carin, "Bayesian compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[7] Erik G. Larsson and Yngve Selen, "Linear regression with a sparse parameter vector," *Signal Processing, IEEE Transactions on*, vol. 55, no. 2, pp. 451 –460, Feb. 2007.

[8] J.S. Turek, I. Yavneh, and M. Elad, "On mmse and map denoising under sparse representation modeling over a unitary dictionary," *Signal Processing, IEEE Transactions on*, vol. 59, no. 8, pp. 3526 –3535, Aug. 2011.

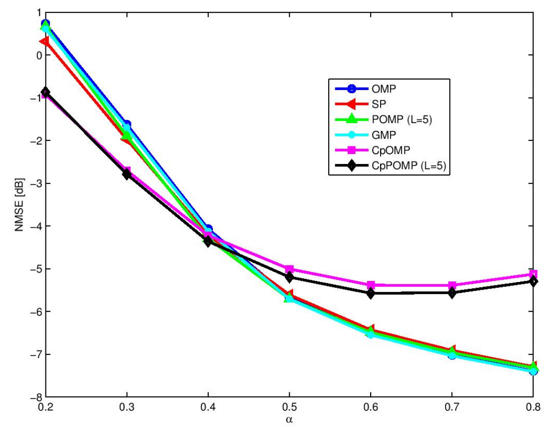

Fig. 6. NMSE [dB] versus $\alpha = m/n$ for $K = 11$.

[9] Cedric Herzet and Angélique Drémeau, "Bayesian Pursuit Algorithms," *Submitted to IEEE Trans. On Signal Processing*, 2012.

[10] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theor.*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[11] Wei Dai and Olgica Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theor.*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[12] Deanna Needell and Joel A. Tropp, "Cosamp: iterative signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, no. 12, pp. 93–100, Dec. 2010.

[13] S. Chatterjee, D. Sundman, and M. Skoglund, "Robust matching pursuit for recovery of gaussian sparse signal," in *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE*, Jan. 2011, pp. 420–424.

[14] S. Chatterjee, K. V S Hari, P. Händel, and M. Skoglund, "Projection-based atom selection in orthogonal matching pursuit for compressive sensing," in *Communications (NCC), National Conference on*, Feb. 2012, pp. 1–5.

[15] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Number v. 1 in Prentice Hall Signal Processing Series. Prentice Hall, 1993.

[16] M.H. Hayes, *Statistical digital signal processing and modeling*, John Wiley & Sons, 1996.

[17] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 1996.

[18] M. Sundin, D. Sundman, and M. Jansson, "Beamformers for sparse recovery," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, 2013.

[19] Ron Rubinstein, Michael Zibulevsky, and Michael Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *Technical Report - CS Technion*, April 2008.