Compressive Parameter Estimation with Earth Mover's Distance via K-Median Clustering

Dian Mo and Marco F. Duarte

Department of Electrical and Computer Engineering
University of Massachusetts, Amherst, MA 01003

ABSTRACT

In recent years, sparsity and compressive sensing have attracted significant attention in parameter estimation tasks, including frequency estimation, delay estimation, and localization. Parametric dictionaries collect observations for a sampling of the parameter space and can yield sparse representations for the signals of interest when the sampling is sufficiently dense. While this dense sampling can lead to high coherence in the dictionary, it is possible to leverage structured sparsity models to prevent highly coherent dictionary elements from appearing simultaneously in a signal representation, alleviating these coherence issues. However, the resulting approaches depend heavily on a careful setting of the maximum allowable coherence; furthermore, their guarantees apply to the coefficient vector recovery and do not translate in general to the parameter estimation task. We propose a new algorithm based on optimal sparse approximation measured by earth mover's distance (EMD). We show that EMD provides a better-suited metric for the performance of parametric dictionary-based parameter estimation. We leverage K-median clustering algorithms to solve the EMD-optimal sparse approximation problem, and show that the resulting compressive parameter estimation algorithms provide satisfactory performance without requiring control of dictionary coherence.

Keywords: compressive sensing, parameter estimation, earth mover's distance, parametric dictionaries, K-median clustering

1. INTRODUCTION

Compressive sensing (CS) has emerged as a framework for integrated sensing and compression of signals that are known to be sparse or compressible in some transform.[1–3] CS has attracted significant attention in recent years; its application has been extended from signal recovery to parameter estimation through the design of parametric dictionaries (PDs) that yield sparse representations for the signal of interest. The PD consists of a set of signal observations corresponding to a discrete set of parameter values. Intuitively, a PD corresponds to a discrete sampling of the continuous parameter space (such as the possible values of the frequencies, delays, and locations, respectively). Making this connection between parameter estimation and sparsity allows for compressive parameter estimation algorithms that rely on the rich standard (sparsity-based) CS framework. The resulting dictionary coefficient vector obtained from CS signal recovery is interpreted by matching parameter values to the nonzero entries of the recovered vector. This sparsity-based approach has been previously formulated for landmark parameter estimation problems, including localization and bearing estimation,[4–15] frequency estimation (FE),[16–21] and time delay estimation (TDE).[22–24]

Unfortunately, compressive parameter estimation with PDs can only be perfect in the contrived case when the observed parameter values are contained in the set of parameters sampled by the PD; fortunately, it may give low estimation error if the true parameters are very close to some sampled parameter.[15] Thus, a dense sampling of the parameter space may be able to improve the average parameter estimation error; however, the resulting PD will have a significantly larger coherence (i.e., the largest inner product between its normalized columns will become closer to one), which is known to hamper the performance of sparse approximation algorithms.[25] Previous approaches address this coherence problem by leveraging structured sparsity models[26] in order to inhibit highly coherent dictionary elements from appearing simultaneously in the recovered signal's representation.[15, 20–22, 27] However, the performance of the resulting algorithms will be dependent on a careful setting of the maximum allowable value of the coherence, which must compensate the performance of the algorithms with the spacing of the parameters that can be observed simultaneously and the resolution of the PD.

A second issue that arises from the use of PDs is that almost all proposed CS recovery algorithms guarantee stable recovery of the coefficient vector as measured by the ℓ_2 norm, i.e., the estimated coefficient vector is close to the true vector in Euclidean distance. Most of these algorithms link the proof of such guarantees to the core thresholding operation, which sets all components of an input vector to zero except for those with the largest magnitudes and returns the optimal sparse approximation to the input vector, again in terms of the ℓ_2 distance. Such guarantees have very limited impact for PD-based compressive parameter estimation, since these guarantees can only be linked to accurate estimation of the support (i.e., the indices of the nonzero entries in the vector, which are translated into parameter estimates) only in the most demanding case of perfect signal recovery. Consider the simple example where a canonical basis vector $c = e_i \in \mathbb{R}^N$ of sparsity 1 (its only nonzero entry is the ith entry) encodes the Fourier Transform coefficients of a complex exponential of frequency f_i. When perfect recovery is obtained, it is easy to see that the frequency estimate is accurate. However, if perfect recovery is not possible, two candidate 1-sparse recovered vectors $\widehat{c} = e_{i+1}$ and $\widehat{c} = e_{i+2}$ provide the same ℓ_2 recovery error, as their supports are disjoint with that of c; nonetheless, the frequency estimate $\widehat{f} = f_{i+1}$ from the first vector will have smaller error than that from the second vector $\widehat{f} = f_{i+2}$, as the elements of the Fourier transform are ordered so that $f_i < f_{i+1} < f_{i+2}$. This example motivates the need for new performance metrics that can capture the difference between these two candidates and prefer the former over the latter.

Several distance metrics that measure coefficient vector error in terms of similarity between sparse supports rather than ℓ_p distance are available, and can be applied to the PD setting. The Hamming distance measures the number of coefficients that are either both zero or both nonzero in the coefficient vector c and its estimate $\widehat{c}$, and certain CS recovery approaches consider this criterion.[28–31] Unfortunately, Hamming distances only control the number of errors committed in parameter estimation, but not in the magnitude of the errors that occur. As an alternative, earth mover's distance (EMD)[32] quantifies the magnitudes of the errors by minimizing the amount and distance of "mass" that would need to flow between the entries of the coefficient vector $\widehat{c}$ to become equal to c. Using this notion of distance in compressive parameter estimation leverages the fact that the elements of the dictionary (and, by extension, the entries of the coefficient vector) are sorted by the corresponding parameter values, so that the EMD between c and $\widehat{c}$ is indicative of the parameter estimation error $\theta - \widehat{\theta}$. Very recently, the EMD has been integrated within CS to provide recovery algorithms for sparse and nearly sparse signals, where the accuracy is measured in terms of the EMD.[33]

In this paper, we propose a new framework for compressive parameter estimation that leverages the EMD to measure estimation error. Our framework replaces the use of the ℓ_2 norm in the sparse approximation algorithms by EMD. The prevalence of thresholding in the standard algorithms is explained by the fact that thresholding finds the best sufficiently sparse approximation to the input vector in the ℓ_2 sense. Therefore, we propose to switch K-thresholding into a search for the best K-sparse approximation to the input in the EMD sense, which is well known to be solved by the K-median clustering procedure.[33] The proposed parameter estimation algorithms can natively avoid highly coherent dictionary elements without prior knowledge about the structure of the PD, eliminating the need for the selection of a coherence inhibition parameter. Additionally, we show that the use of manifold interpolation approaches[21, 22] significantly improves the performance of the proposed methods.

This paper is organized as follows. Section 2 provides background on compressive parameter estimation. Section 3 motivates the use of EMD for this problem when PDs are involved, and Section 4 introduces our proposed formulation for EMD-based compressive parameter estimation. Section 5 provides numerical simulation results that verify the advantages of the proposed approach, and Section 6 offers conclusions and directions for future work.

2. COMPRESSIVE PARAMETER ESTIMATION

CS describes a reduced-rate acquisition framework in which a discrete signal $x \in \mathbb{C}^N$ is compressed using a linear dimensionality-reducing operator $\Phi \in \mathbb{R}^{M \times N}$ to obtain a measurement $y = \Phi x \in \mathbb{C}^M$, where $M \ll N$. CS shows that it is possible to recover an accurate estimate $\widehat{x}$ of the signal x from measurement y when x is known to be sparse in a transform domain, i.e., the coefficient vector $c = \Psi^* x$ of x via the transform matrix Ψ has only a small number $K \le M$ of nonzero components.

A set of parametric signals is defined via a mapping $\mathcal{M} : \Theta \to X$ from a parameter space $\Theta \subseteq \mathbb{C}^K$ to a signal space $X \subseteq \mathbb{C}^N$ that connects each parameter value $\theta \in \Theta$ with a signal $x = \mathcal{M}(\theta) \in X$. The parameter estimation problem from the noisy measurement $y = x + n$, where n denotes additive white Gaussian noise (AWGN), is to find the closed signal $\widehat{x} \in \mathcal{M}$ to y and then invert the mapping to estimate the corresponding parameter $\widehat{\theta}$. The CS theory suggests that a random projection operator $\Phi \in \mathbb{R}^{M \times N}$ with $M \ll N$ approximately preserves the distance between any signal contained in the set $\{\mathcal{M}(\theta) : \theta \in \Theta\}$ and a fixed arbitrary signal.[34] In other words, compressive parameter estimation can be performed from the noisy compressed measurement $y = \Phi x + n$ without having to recover the full signal $\widehat{x}$ from y to then estimate the parameter $\widehat{\theta}$ from $\widehat{x}$. However, such direct approaches assume that the mapping $\mathcal{M}$ is known and available, usually in the form of a nonlinear low-dimensional manifold model, which can be arbitrarily complex to leverage.[34]

As an alternative to manifold models, parametric dictionaries (PDs) have adapted the sparse signal model to perform parameter estimation directly from CS measurements without having to recover the full signal.[4–24] In many practical applications, high-complexity parametric signals can be expressed or approximated by the linear combination of low-complexity parametric signals, i.e., they correspond to the linear combination of some number K of parametric signals with distinct parameters $x = \sum_{k=1}^{K} c_k \mathcal{M}(\theta_k)$. One can build a PD as a collection of samples from the set of parametric signals $\Psi = [\mathcal{M}(\omega_1), \mathcal{M}(\omega_2), \ldots, \mathcal{M}(\omega_L)]$ that correspond to a sampling of the parameter space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_L\} \subseteq \Theta$. If the sampling set is large and dense enough so that unknown parameters are all contained in the sampling set $\{\theta_1, \theta_2, \ldots, \theta_K\} \subseteq \Omega$, then the signal can be written as the product of the PD and a coefficient vector $x = \Psi c$, where the coefficient vector has at most K non-zero components. By introducing a PD, sparse approximation is retrofitted into parameter estimation, which allows very simple integration into the CS framework. In this case, finding the unknown parameters reduces to finding at most K dictionary elements whose linear combination correspond to a CS measurement vector close enough to y, that is, to obtain an estimate of the coefficient vector $\widehat{c}$ from the measurements $y = \Phi x + n = \Phi \Psi c + n$.

A limitation that is implicit from the definition of the PD is that parameter estimation can be perfectly solved only if all the values for the observed parameters are among the sampled parameters $\Omega \subseteq \Theta$. Fortunately, when this perfect inclusion case is not met for some observed parameter θ_k, having a densely sampling of the parameter space in the PD increases the chance that the observation for some sampled parameter $\mathcal{M}(\omega_l)$ is sufficiently close to the observed vector, i.e., $\|\mathcal{M}(\omega_l) - \mathcal{M}(\theta_k)\|$ is very small, in which case it is possible to control the accuracy of parameter estimation. However, increasing the parameter space sampling resolution inherently increases the similarity between dictionary elements for adjacent parameter values, as measured by the coherence:

$$\mu(\Psi) := \max_{1 \le i \ne j \le L} \frac{|\langle \mathcal{M}(\omega_i), \mathcal{M}(\omega_j) \rangle|}{\|\mathcal{M}(\omega_i)\|_2 \|\mathcal{M}(\omega_j)\|_2}.$$

The higher similarity that adjacent dictionary elements have, the closer that $\mu(\Psi)$ is to one, and the more difficult that it is to distinguish between them, severely hampering the performance of parameter estimation with densely sampled PDs.[25]

Structured sparsity has emerged as an alternative signal model formulation to address the coherence issue in PDs.[26] In contrast with classical sparsity that allows any sparse vector to appear as the signal representation, structured sparsity only allows for sparse vectors that exhibit particular additional structure. Structured sparsity is introduced in CS recovery by replacing the thresholding operation in standard sparse recovery algorithms by a structured sparse approximation algorithm that provides the the best approximation to the input within the set of allowable (structured) sparse signals. To alleviate coherence issues in PD-based parameter estimation, one can design a coherence-inhibiting structured sparsity model in which the K nonzero entries in a vector must correspond to dictionary elements that have low coherence (i.e., normalized inner product), in order to inhibit highly coherent dictionary elements from appearing simultaneously in the signal representation.[15, 20–22] This structured sparsity model is defined by a maximum allowed coherence level ν that defines the restriction on the choice of dictionary elements that can appear simultaneously. Such coherence-inhibiting framework has resulted in a variety of PD-tailored algorithms, including structured iterative hard thresholding (SIHT),[15, 20] band-excluded orthogonal matching pursuit (BOMP),[27] and band-excluded interpolating subspace pursuit (BISP).[21]

Although it is clear that an appropriate choice of the maximum coherence parameter ν can improve the performance of PD-based parameter estimation, there has not been much research on the sensitivity of the

aforementioned algorithms to the choice of value of ν. Intuitively, setting the parameter too low results in performance similar to that of standard CS algorithms, which is poor. Alternatively, setting the parameter too high results in strict requirements on the minimum separation of the parameter values of the observed signal, resulting in suboptimal performance for observations that do not meet this requirement. Our simulations in Section 5 illustrate the sensitivity of the choice of the value of ν in the performance of PD-based compressive parameter estimation.

3. EARTH MOVER'S DISTANCE AND PARAMETRIC DICTIONARIES

The approaches to parameter estimation we have described so far guarantee stable recovery of the coefficient vector c under the ℓ_2 norm, which underlies the sparse approximation algorithms used by them. More precisely, both the thresholding operation in classical CS and and the coherence-inhibiting approximation in structured CS involve approximation steps that return the approximation of the coefficient vector with the lowest Euclidean-norm error $\|c - \widehat{c}\|_2$ among the class of signals of interest (sparse or structured sparse, respectively). However, such guarantees cannot be translated to the error of parameter estimation $\|\theta - \widehat{\theta}\|$. This is due to the fact that, except for the stringent case of perfect recovery where $\widehat{c} = c$, an ℓ_2-norm guarantee cannot translate into a guarantee on the selection of the support (nonzero entries) of c, which are interpreted to obtain parameter estimates. Thus, whenever perfect recovery is not possible, one cannot obtain performance bounds on PD-based parameter estimation. While alternative approaches can provide guarantees on perfect recovery of the signal support, they do not provide control on the types of errors that may arise when perfect support estimation is not achievable.[28–31]

Interestingly, the earth mover's distance (EMD) provides a notion of error (i.e., distance between two vectors) that is significantly better suited for application to PD coefficient vectors. The EMD between two vectors p, q relies on the notion of mass being assigned to each entry of the two vectors involved, with the goal being to transfer mass between the entries of the first vector in order to match mass of the entries of the second vector. The EMD metric captures the difference between two vectors by finding the flow with the smallest amount of work (measured as the product of the mass to be moved and the transport distance the mass should move) among all flows that when applied to the first vector yield the second one; the distance itself is the value of such minimum work.[32] For two vectors $p, q \in \mathbb{R}^L$, if f_{ij} denotes the flow mass between the ith component of p and the jth component of q, then the EMD is defined as

$$\text{EMD}(p, q) := \min_f \sum_{i,j=1}^{L} f_{ij}|i - j| \text{ such that } \sum_j f_{ij} = |p_i| \ \forall \ i = 1, \ldots, N, \sum_i f_{ij} = |q_j| \ \forall \ j = 1, \ldots, N. \quad (1)$$

In the case where $\sum_i |p_i| \neq \sum_i |q_i|$, one can easily add a new flow source or sink that provides or receives the difference with a larger transport cost, i.e., $L + 1$.

Consider now the specific case in which the vectors c and $\widehat{c}$ represent the true and estimated K-sparse PD coefficient vectors, respectively, with respective supports $\Gamma = \{\gamma_1, \ldots, \gamma_K\}$ and $\widehat{\Gamma} = \{\widehat{\gamma}_1, \ldots, \widehat{\gamma}_K\}$. These PD coefficients correspond to parameter values $\theta = [\gamma_1 \Delta \ \ldots \ \gamma_K \Delta]$ and estimates $\widehat{\theta} = [\widehat{\gamma}_1 \Delta \ \ldots \ \widehat{\gamma}_K \Delta]$. In the case where the estimate is imperfect, the supports of c and $\widehat{c}$ do not exactly match each other and the standard CS error $\|c - \widehat{c}\|_2$ is potentially independent of the estimation error $\|\theta - \widehat{\theta}\|$. In contrast, in this case one can write the EMD metric as

$$\text{EMD}(c, \widehat{c}) = \min_f \sum_{i,j=1}^{K} f_{ij}|\gamma_i - \widehat{\gamma}_j| \text{ such that } \sum_j f_{ij} = |c[\gamma_i]|, \ \sum_i f_{ij} = |\widehat{c}[\widehat{\gamma}_j]| \ \forall \ i, j = 1, \ldots, K. \quad (2)$$

Assuming that the estimates of the parameters are sufficiently accurate so that $c[\gamma_i] \approx c[\widehat{\gamma}_i]$ for $i = 1, \ldots, K$, the minimizing flow would yield (perhaps approximately) $f_{i,i} = |c[\gamma_i]|$, with negligibly small values $f_{i,j}$ for $i \neq j$. In such case, it is easy to see that the EMD in (2) is approximately

$$\text{EMD}(c, \widehat{c}) \approx \sum_{i=1}^{K} |c[\gamma_i]||\gamma_i - \widehat{\gamma}_i|,$$

4

where each γ_i is matched to a $\widehat{\gamma}_i$. This EMD value which yields an amplitude-weighted metric for parameter estimation. In particular, when $c[\gamma_i]$ are approximately equal to a constant C, we can further write that

$$\text{EMD}(c, \widehat{c}) \approx C \sum_{i=1}^{K} |\gamma_i - \widehat{\gamma}_i| = \frac{C}{\Delta} \sum_{i=1}^{K} |\theta_i - \widehat{\theta}_i| = \frac{C}{\Delta} \|\theta - \widehat{\theta}\|_1. \tag{3}$$

This approximation motivates the use of EMD as a metric on the PD coefficients, in order to be able to control the performance of sparsity-based or compressive parameter estimation.

4. PARAMETER ESTIMATION WITH EARTH MOVER'S DISTANCE

The aforementioned properties of the EMD motivate its integration into the CS framework.[33] We develop here a greedy approach to EMD-based sparse recovery with PDs that follows the principles of structured sparse signal recovery algorithms.[26]

4.1 Optimal Sparse Approximation with EMD Metric

A key component to the integration of EMD and CS is the formulation of an optimal sparse approximation procedure that relies on the EMD metric. First, consider the problem of finding a K-sparse vector $\widehat{c}$ with fixed support Γ that has smallest EMD to an arbitrary vector c. In this case, the minimum-cost flow in the EMD definition (1) is achieved when the flow is active only between each entry of the vector c and the nearest non-zero entry of the vector $\widehat{c}$. Since the entries of $\widehat{c}$ are arbitrary, we minimize this EMD by moving the mass from each entry of the vector c into the corresponding nonzero entry in $\widehat{c}$ that minimizes the transport distance. That is, we choose each nonzero value $\widehat{c}[\gamma_i]$ so that it receives through the flow all the mass from all entries of c that are closest to it. This choice of $\widehat{c}$ will yield

$$\text{EMD}(c, \widehat{c}) = \sum_{i=1}^{L} |c_i| \min_{j \in \Gamma} |i - j|. \tag{4}$$

In words, the search for the EMD-closest sparse vector $\widehat{c}$ with fixed support finds a partition of the components of c into K different groups, picking the nonzero components of the vector $\widehat{c}$ so that the flow work cost function is minimized. However, to achieve optimal sparse approximation using EMD, we must address the additional problem of determining the optimal selection of the support of $\widehat{c}$ (i.e., the indices of its K nonzero entries) that minimizes the EMD in (4).

A well-known method to solve this search problem is given by K-median clustering.[35] Consider a setting where L points $p_1, \ldots, p_L$ have mutual distances $d(p_i, p_j)$ and weights w_i, $i, j = 1, \ldots, L$. The goal of K-median clustering is to find a set $\Lambda \subseteq \{p_1, \ldots, p_L\}$ of K centroids that minimize the objective function

$$\sum_{i=1}^{L} w_i \min_{j \in \Lambda} d(p_i, p_j).$$

By setting $p_i = i$, $d(p_i, p_j) = |j - i|$ and $w_i = |c_i|$ for $i, j = 1, \ldots, L$, it is easy to see that the resulting K-median clustering setup solves for the minimum EMD cost in (4) among all sparse signals $\widehat{c}$. In fact, the set of centroids Λ resulting from the K-median clustering of the points $p_1, \ldots, p_L$ provides the support Ω for the optimal K-sparse approximation $\widehat{c}$; the magnitude of those entries are then obtained trivially.

A common K-median clustering approach proceeds iteratively, starting from initial centroids that are randomly chosen, and assigning each point to the cluster with the nearest centroid; the cluster centers are then updated to be the weighted median of the cluster's points.[35] This algorithm can be adapted for optimal sparse approximation in the EMD sense as shown in Algorithm 1. This EMD-based approach, in contrast with structured sparsity-based approaches, does not require the tuning of the maximum allowed coherence level ν and is more robust to variations in the level of coherence of the PD elements, as highly coherent elements are naturally assigned into the same cluster by the K-median clustering step; numerical evidence of this property will be shown in Section 5.

Algorithm 1 EMD sparse approximation support finder $\Sigma = \mathbb{C}(c, K)$

Input: coefficient vector c, sparsity K
Output: approximation vector support Σ
 1: Initialize: choose Σ as a random K-element subset of $\{1, \ldots, L\}$.
 2: **repeat**
 3:   $g_i = \arg\min_{j=1,\ldots,K} |i - s_j|$ for each $i = 1, \ldots, L$
 4:   $\sigma_j = \text{median}\{i \cdot |c_i| : g_i = j\}$ for each $j = 1, \ldots, K$
 5:   $\Sigma = \{\sigma_1, \ldots, \sigma_K\}$
 6: **until** S does not change

Algorithm 2 Clustering Subspace Pursuit (CSP)

Input: measurement vector y, measurement matrix Φ, sparsity K, set of PD-sampled parameter values Ω
Output: estimated signal x, estimated parameter values θ
 1: Initialize $x = 0$, $\Sigma = \emptyset$; generate PD Ψ from Ω.
 2: **repeat**
 3:   $y_r = y - \Phi x$ {Compute residual}
 4:   $c_r = (\Phi\Psi)^T y_r$ {Obtain proxy estimate of PD coefficients}
 5:   $\Sigma = \Sigma \cup \mathbb{C}(c_r, K)$ {Augment support estimate using EMD-optimal sparse approximation of proxy}
 6:   $c = (\Phi\Psi_\Sigma)^+ y$ {Update PD coefficient proxy to be $2K$-sparse}
 7:   $\Sigma = \mathbb{C}(c, K)$ {Obtain EMD-optimal sparse approximation of proxy}
 8:   $x = \Psi_\Sigma c$ {Assemble signal estimate}
 9:   $\theta = \Omega_\Sigma$ {Assemble parameter estimates}
10: **until** a convergence criterion is met

4.2 Clustering Subspace Pursuit

Since most existing sparse recovery algorithms rely on a thresholding operation to obtain optimal sparse approximations in the ℓ_2 metric, it is particularly easy to modify the existing framework to achieve sparse recovery with an EMD metric. We propose a new PD-based parameter estimation algorithm, called Clustering Subspace Pursuit (CSP), and shown in Algorithm 2. CSP merges the Subspace Pursuit algorithm[36] with EMD-based sparse approximation (Algorithm 1). CSP replaces the thresholding steps from SP that find candidate support sets Σ from a residual (or proxy) coefficient vector c_r by K-median clustering steps. CSP avoids highly coherent dictionary elements from appearing simultaneously in the signal representation, while removing the requirement for a coherence-inhibiting parameter ν that is required when structured sparsity is used. Furthermore, CSP has the potential to provide EMD-based error guarantees for parameter estimation.

5. NUMERICAL SIMULATIONS

To evaluate the performance of our proposed approach, we consider the time delay estimation problem where the signal is measured using CS. The continuous signal model is a chirp waveform with time delay s defined as

$$g(t, s) := p(t - s) \exp\left(j2\pi\left(f_0 + f_\Delta \frac{t - s}{2T}\right)(t - s)\right)$$

where $f_0 = 1\text{MHz}$ is the chirp center frequency, $f_\Delta = 5\text{MHz}$ is the frequency sweep range, and $p(t)$ is a raised cosine pulse that windows the chirp signal in time:

$$p(t) = \begin{cases} 1 + \cos(2\pi t/T), & t \in (0, T), \\ 0 & \text{otherwise.} \end{cases}$$

Here, $T = 1\mu s$ is the duration of the chirp signal. We sample the chirp signal at a frequency $f_s = 50\text{MHz}$ and collect $N = 500$ samples to generate the discrete signal x_s, whose samples can be written as

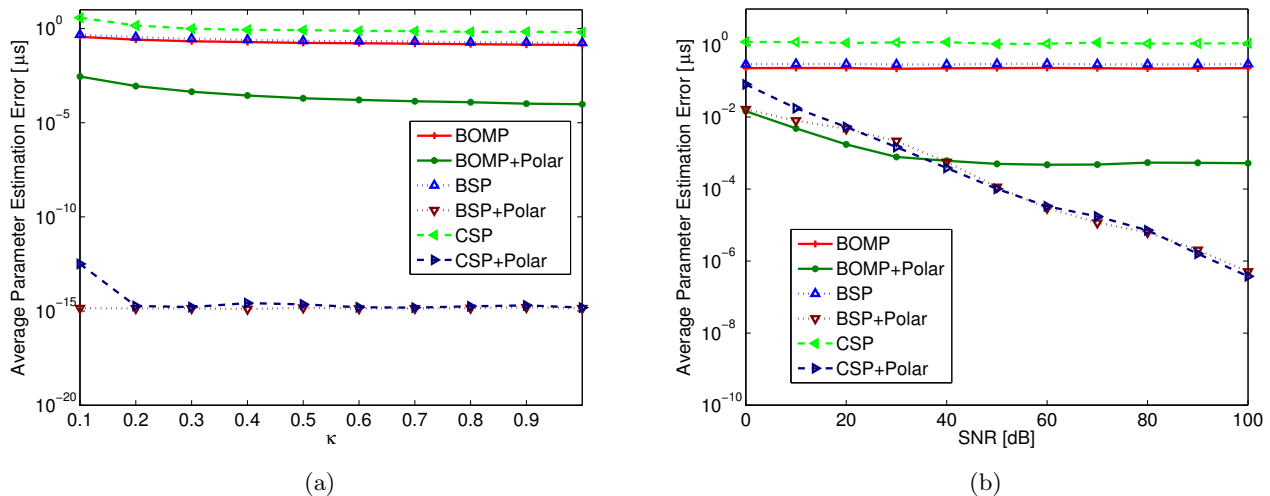$$g_s[n] = \frac{1}{\sqrt{1.5 T f_s}} g\left(\frac{n - 1}{f_s}, s\right), \quad n = 1, \ldots, N,$$

6

Figure 1: Average delay estimation error as a function of (a) the CS sub-sampling ratio for noiseless measurements, and (b) the SNR level for noisy measurements, where the chirp duration is 1μs.

where the coefficient $\frac{1}{\sqrt{1.5Tf_s}}$ normalizes the chirp. The observed signal can then be written as $x = \sum_{i=1}^{K} a_k g_{s_k}$, where the parameters s_k are selected at arbitrary resolution from the range $[0, 10\mu s]$, with a minimum separation between parameters of $T_s = 0.02\mu$s, and a_k is the magnitude coefficient for the kth component. We generate a PD for this problem by sampling the the time delay (i.e., parameter) space with a spacing of $T_s = 0.02\mu$s (matching the sampling period); this PD can be written as $\Psi = [g_0, g_{T_s}, g_{2T_s}, \ldots, g_{(N-1)T_s}]$, and its entries are given by $\Psi[i,j] = g_{(j-1)T_s}[i]$. The signal x is then sensed using a CS-based random demodulator[37] simulated by an $M \times N$ CS matrix Φ for a variety of values of M.

Our experiments compare the time delay estimation performance of CSP to that of two existing baseline algorithms: band-excluding subspace pursuit (BSP) and band-excluded orthogonal matching pursuit (BOMP).[27] Furthermore, we integrate polar interpolation within these algorithms to accommodate arbitrary values for the delay outside of the sampled set;[21, 38] note in particular that the BSP+Polar algorithm is equivalent to BISP.[21] We set the maximum allowed coherence level to $\nu = 0.001$ for all structured sparsity (band-excluded) algorithms.

Our first experiments consider compressive time delay estimation from noiseless measurements as a function of the CS subsampling rate $\kappa = M/N$, as well as for noisy measurements under AWGN with fixed subsampling rate $\kappa = 0.4$ as the function of the SNR level. Figure 1 shows the performance of the algorithms in these setups averaged over 1000 randomized realizations. While the performance of CSP does not match that of the band-excluding algorithms when no interpolation is used, there is a significant improvement in estimation performance when polar interpolation is added to the algorithms. This is indicative of small biases (on the order of the value of T_s) that are observed in the output from CSP and easily corrected by interpolation. When polar interpolation is added, BOMP estimates only one time delay at a time; the interference from the remaining copies of the delayed signal can cause noticeable errors in the interpolation stage.

Our next experiment evaluates the role that the PD coherence level has on the performance of the algorithms. To vary this coherence level, we vary the duration of chirp signal within the range $[1\mu s, 5\mu s]$, which is expected to increase the number of PD element pairs that are coherent (i.e., change the necessary band exclusion). We expect that the band-exclusion algorithms will be sensitive to the fixed choice of the maximum allowed coherence level ν. For the duration $T = 2\mu$s, Figure 2 replicates the setup of Figure 1 and shows decreased performance for all algorithms except for CSP and CSP+Polar. Clearly, the drop in performance in the structured sparsity-based algorithms is due to a suboptimal choice of the parameter ν for the PDs that feature increased coherence. This gap in performance is expected to become more significant as the spacing between the delays (or parameters in general) becomes smaller and leads to the selection of elements with coherence above the allowed value ν.
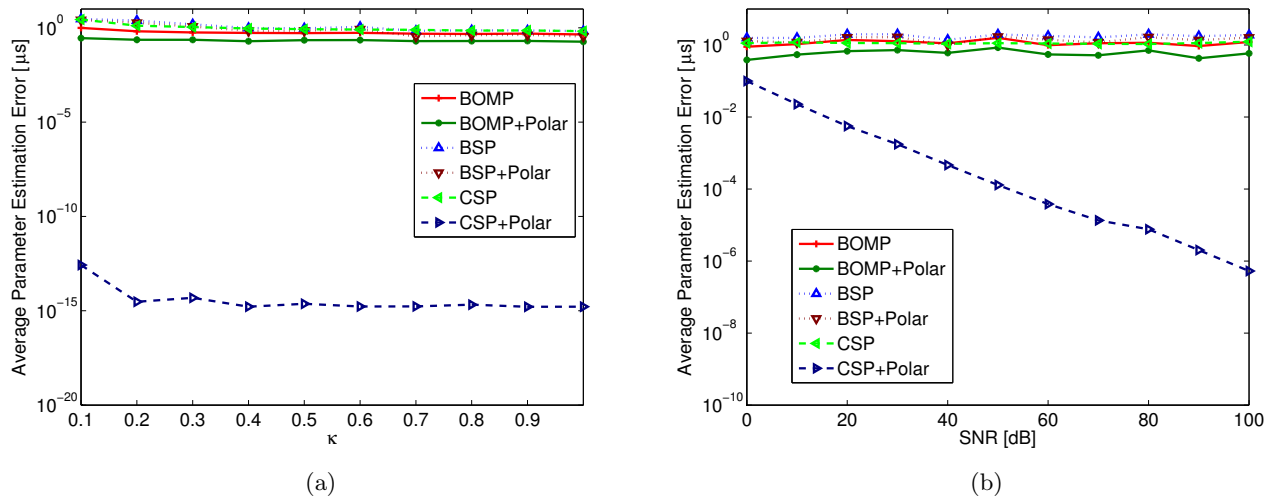
Figure 2: Average delay estimation error as a function of (a) the CS sub-sampling ratio for noiseless measurements, and (b) the SNR level for noisy measurements, where the chirp duration is $2\mu s$.

Our last experiment focuses on this sensitivity on the choice of the maximum allowed coherence parameter ν for BSP in contrast with CSP. We test the performance of these algorithms as a function of the CS subsampling ratio κ, with and without polar interpolation, for a variety of chirp lengths. Figure 3 shows the average performance over 100 randomized realizations per setup, and shows both a range of degradation levels in BSP performance and a sharp degradation in BSP+Polar performance as the PD coherence varies. This is in contrast to CSP and CSP+Polar, whose performances are essentially stable over the choice of PD.

6. CONCLUSION

The use of PDs for sparsity-based and compressive parameter estimation introduce a tradeoff between accuracy of the sparse signal model, which requires dense parameter space sampling in the PD, and poor estimation performance caused by the high coherence between PD elements, which is known to affect sparse approximation algorithms. While structured sparsity can address some of these issues, our experiments show a close dependence between the application of interest and the coherence-controlling parameter ν, which may not be easy to identify in the most challenging applications. Motivated by the properties of PDs and their interpretation for parameter estimation, we have introduced the use of EMD-based sparse approximation within PD-based parameter estimation algorithms. We highlighted the connection between EMD sparse approximation and K-median clustering, which in combination with existing greedy sparse recovery algorithms provide us with an intuitive computational framework for PD-based parameter estimation. The proposed CSP algorithms matches or improves the performance of existing structured sparsity-based algorithms for parameter estimation without requiring careful control of the coherence of the PD, which is likely key to guarantee acceptable estimation performance. In future work, we expect to explore the potential for algorithms based on K-median clustering to provide recovery guarantees that rely on the EMD metric, which will immediately translate into parameter estimation performance guarantees. Future work will also address the sensitivity of the algorithm to the presence of noise and characterize the estimation bias that is observed by CSP and corrected by interpolation. More sophisticated algorithms should be formulated to involve the thresholding level, which may alleviate the influence of noise, and extend to higher dimensional parameter space.

REFERENCES

[1] Donoho, D. L., "Compressed sensing," *IEEE Trans. Info. Theory* **52**, 1289–1306 (Sep. 2006).

[2] Candès, E. J., "Compressive sampling," in [*Proc. Int. Congress of Mathematicians*], **3**, 1433–1452 (Aug. 2006).

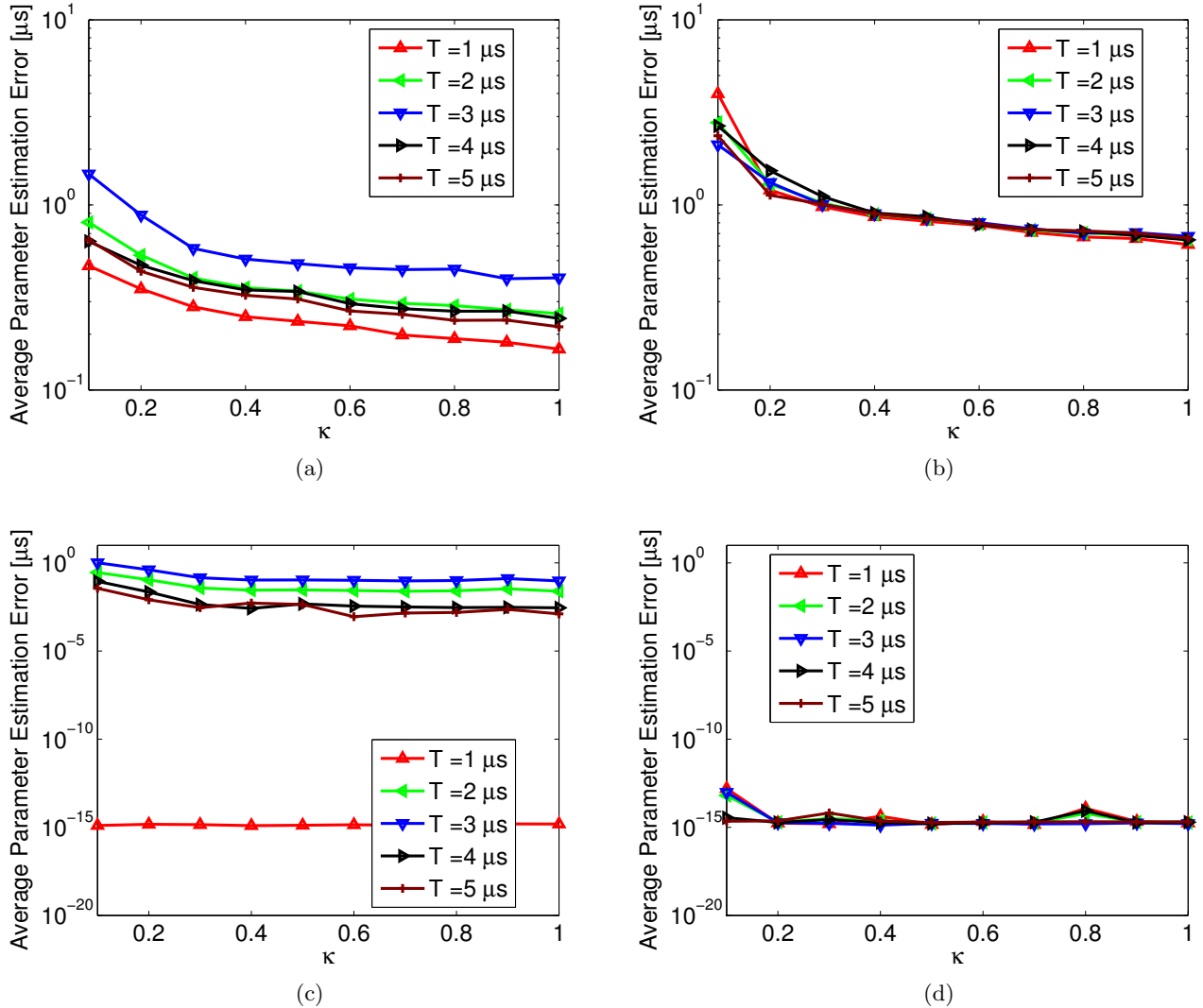[3] Baraniuk, R. G., "Compressive sensing," *IEEE Signal Proc. Mag.* **24**(4), 118–120, 124 (July 2007).

8

Figure 3: Average delay estimation error for (a) BSP, (b) CSP, (c) BSP+Polar, and (d) CSP+Polar algorithms as a function of the sub-sampling ratio κ for several chirp lengths T.

[4] Gorodnitsky, I. F. and Rao, B. D., "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Proc.* **45**, 600–616 (Mar. 1997).

[5] Malioutov, D., Çetin, M., and Willsky, A. S., "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Proc.* **53**, 3010–3022 (Aug. 2005).

[6] Karabulut, G. Z., Kurt, T., and Yongaçoglu, A., "Estimation of direction of arrival by matching pursuit," *EURASIP J. Wireless Comm.* **2**, 197–205 (2005).

[7] Varshney, K. R., Çetin, M., Fisher, J. W., and Willsky, A. S., "Sparse representation in structured dictionaries with application to synthetic aperture radar," *IEEE Trans. Signal Proc.* **56**, 3548–3561 (Aug. 2008).

[8] Cevher, V., Gurbuz, A. C., McClellan, J. H., and Chellappa, R., "Compressive wireless arrays for bearing estimation," in [*IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*], 2497–2500 (Apr. 2008).

[9] Gurbuz, A. C., Cevher, V., and McClellan, J. H., "A compressive beamformer," in [*IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*], 2617–2620 (Apr. 2008).

[10] Stojanovic, I., Karl, W. C., and Çetin, M., "Compressed sensing of mono-static and multi-static SAR," in [*Algorithms for Synthetic Aperture Radar Imagery XVI*], *Proc. SPIE* **7337** (Apr. 2009).

[11] Christensen, M. G. and Nielsen, J. K., "Joint direction-of-arrival and order estimation in compressed sensing using angles between subspaces," in [*IEEE Statistical Signal Proc. Workshop (SSP)*], 449–452 (Jun. 2011).

[12] Herman, M. and Strohmer, T., "High resolution radar via compressed sensing," *IEEE Trans. Signal Proc.* **57**, 2275–2284 (June 2009).

[13] Kyriakides, I., "Adaptive compressive sensing and processing of delay-doppler radar waveforms," *IEEE Trans. Signal Proc.* **60**, 730–739 (Feb. 2012).

[14] Sen, S., Tang, G., and Nehorai, A., "Multiobject optimization of OFDM radar waveform for target detection," *IEEE Trans. Signal Proc.* **59**, 639–652 (Feb. 2011).

[15] Duarte, M. F., "Localization and bearing estimation via structured sparsity models," in [*IEEE Statistical Signal Proc. Workshop (SSP)*], (Aug. 2012).

[16] Chen, S. and Donoho, D., "Application of basis pursuit in spectrum estimation," in [*IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP)*], **3**, 1865–1868 (May 1998).

[17] Fuchs, J.-J., "Minimal ℓ_1-norm reconstruction function for oversampled signals: application to time-delay estimation," *IEEE Trans. Info. Theory* **46**, 1666–1673 (July 2000).

[18] Bourguignon, S., Carfantan, H., and Idier, J., "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Selected Topics in Signal Proc.* **1**, 575–585 (Dec. 2007).

[19] Austin, C. D., Moses, R. L., Ash, J. N., and Ertin, E., "On the relation between sparse reconstruction and parameter estimation with model order selection," *IEEE J. Selected Topics in Signal Proc.* **4**, 560–570 (June 2010).

[20] Duarte, M. F. and Baraniuk, R. G., "Spectral compressive sensing," *Appl. Comput. Harmon. Analysis* **35**, 111–129 (July 2013).

[21] Fyhn, K., Dadkhahi, H., and Duarte, M. F., "Spectral compressive sensing via polar interpolation," in [*IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*], 6225–6229 (May 2013).

[22] Fyhn, K., Jensen, S. H., and Duarte, M. F., "Compressive parameter estimation for sparse translation-invariant signals using polar interpolation," (May 2013). Preprint. Available at `http://arxiv.org/pdf/1305.3483`.

[23] Eftekhari, A., Romberg, J. K., and Wakin, M. B., "Matched filtering from limited frequency samples," *IEEE Trans. Info. Theory* **59**, 3475–3496 (June 2013).

[24] Jamali-Rad, H. and Leus, G., "Sparsity-aware TDOA localization of multiple sources," in [*IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*], 4021–4025 (May 2013).

[25] Tropp, J. A., "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theory* **50**, 2231–2242 (Oct. 2004).

[26] Baraniuk, R. G., Cevher, V., Duarte, M. F., and Hegde, C., "Model-based compressive sensing," *IEEE Trans. Info. Theory* **56**, 1982–2001 (Apr. 2010).

[27] Fannjiang, A. C. and Liao, W., "Coherence-pattern guided compressive sensing with unresolved grids," *SIAM J. Imaging Science* **5**(1), 179–202 (2012).

[28] Fletcher, A. K., Rangan, S., and Goyal, V. K., "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Info. Theory* **55**, 5758–5772 (Dec. 2009).

[29] Karbasi, A., Hormati, A., Mohajer, S., and Vetterli, M., "Support recovery in compressed sensing: An estimation theoretic approach," in [*IEEE Int. Symp. Info. Theory (ISIT)*], 679–683 (June 2009).

[30] Reeves, G. and Gastpar, M., "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Info. Theory* **58**, 3065–3092 (May 2012).

[31] Bajwa, W., Calderbank, R., and Jafarpour, S., "Why Gabor frames? Two fundamental measures of coherence and their role in model selection," *J. Communication Networks* **12**, 289–307 (Aug. 2010).

[32] Rubner, Y., Tomasi, C., and Guibas, L., "A metric for distributions with applications to image databases," in [*Int. Conf. Computer Vision (ICCV)*], 59–66 (Jan. 1998).

[33] Indyk, P. and Price, E., "K-median clustering, model-based compressive sensing, and sparse recovery for earth mover distance," in [*Symp. Theory of Computing (STOC)*], 627–636 (June 2011).

[34] Wakin, M. B. and Baraniuk, R. G., "Random projections of smooth manifolds," *Found. Comput. Math.* **9**, 51–77 (Feb. 2009).

[35] Bradley, P. S., Mangasarian, O. L., and Street, W. N., "Clustering via concave minimization," in [*Advances in Neural Information Processing Systems (NIPS)*], 368–374 (Dec. 1997).

[36] Dai, W. and Milenkovic, O., "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Info. Theory* **55**, 2230–2249 (May 2009).

[37] Tropp, J. A., Laska, J. N., Duarte, M. F., Romberg, J. K., and Baraniuk, R. G., "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Info. Theory* **56**, 520–544 (Jan. 2010).

[38] Ekanadham, C., Tranchina, D., and Simoncelli, E. P., "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Proc.* **59**, 4735–4744 (Oct. 2011).