Spectral Compressive Sensing with Model Selection
Zhenqi Lu, Rendong Ying, Sumxin Jiang, Zenghui Zhang, Peilin Liu, Wenxian Yu

Abstract

The performance of existing approaches to the recovery of frequency-sparse signals from compressed measurements is limited by the coherence of required sparsity dictionaries and the discretization of frequency parameter space. In this paper, we adopt a parametric joint recovery-estimation method based on model selection in spectral compressive sensing. Numerical experiments show that our approach outperforms most state-of-the-art spectral CS recovery approaches in fidelity, tolerance to noise and computation efficiency.

Index Terms

Compressive sensing, frequency-sparse signal, model selection, parametric estimation, maximum likelihood estimator

I. INTRODUCTION

One of the recent research interests of compressive sensing (CS) has focused on the recovery of signals that are spectrally sparse from a reduced number of measurements [1]–[6]. A great many applications, including spectrum sensing [7] and wideband communication [3], [8], feature smooth or modulated signals that can be modelled as a superposition of a small number of sinusoids. Recovery of such frequency-sparse signals brings about a novel issue in the formulation of CS recovery problem: signal representations in frequency domain have a continuous parameter space, while recent CS researches [9]–[11] are rooted on signal decomposition in a discretized dictionary.

An intuitive solution to this problem is a denser sampling of the parameter space, which improves the compressibility of signal representations. But increasing the resolution of parameter sampling worsens the coherence between dictionary elements, which results in loss of sparsity and uniqueness of signal representations. Such ambiguity prevents certain algorithms [9], [10] from achieving the sparse representation successfully. Initial contributions to spectral CS recovery are concentrated on the recovery algorithm, the optimization problem formulation and the sparsity prior to combat the intricacy in signal representations [5], [12]–[16].

Estimation of the frequencies, amplitudes and phase shifts of sinusoids embedded in noise is a fundamental problem in time series analysis and other more general statistical signal processing problems. Classical treatments of this problem are restricted to the Fourier frequencies. This implicit discretization barrier is overcome due to the introduction of the minimum description length (MDL) principle for model selection [17]–[19]. In this paper, we improve over existing approaches by applying model selection to spectral CS. We take into consideration parametric joint recovery-estimation methods, which determine parameters by minimizing a log-likelihood function of compressed measurements.

The novelty of our approach is that it performs estimation from compressed measurements rather than from signal samples. Since the log-likelihood minimization is in practice equivalent to an $\ell_2$-norm minimization, the estimation performance is guaranteed as long as the sensing matrix satisfies restricted isometry property (RIP), and thus distance-preserving [10]. It can be shown that random marices from Gaussian, Rademacher, or more generally a sub-Gaussian distribution have the RIP with high probability under certain conditions [20]. We solve the optimization problem through an iterative greedy approach to outperform state-of-the-art spectral CS approaches. Experimental results show improved reconstruction fidelity against existing approaches, from both noiseless and noisy measurements. In addition, our approach is essentially greedy, and is thus more computationally efficient than optimization based approaches. Furthermore, compared to traditional model selection estimators [17]–[19], our approach reduces the number of samples needed and consequently the computational load of estimation.

This work is partially supported by the National Natural Science Foundation of China under grant number 61171171 and 61102169.

Copyright 2014 IEEE. Published in the IEEE 2014 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), scheduled for 4-9 May 2014 in Florence, Italy. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works, must be obtained from the IEEE. Contact: Manager, Copyrights and Permissions / IEEE Service Center / 445 Hoes Lane / P.O. Box 1331 / Piscataway, NJ 08855-1331, USA. Telephone: + Intl. 908-562-3966.
II. PROBLEM FORMULATION AND RELATED PRIOR WORK

The problem of recovering frequency-sparse signals from compressed measurements is formulated as follows: Let $s(t)$ denote an unidimensional real-valued frequency-sparse signal composed of $K$ sinusoids with unknown frequencies $\omega_j$, amplitudes $a_j$, and phase shifts $\phi_j$, and $x(t)$ be $s(t)$ corrupted by additive noise $\xi(t)$ with unknown noise level.

\[ x(t) = s(t) + \xi(t) = \sum_{j=1}^{K} a_j \sin(\omega_j t + \phi_j) + \xi(t). \]  

Let $s = \{s_t\}_{t=1}^{N}$ and $x = \{x_t\}_{t=1}^{N}$ be the observed samples of $s(t)$ and $x(t)$ at discrete times $\{t_j = j\}_{j=1}^{N}$, and $m = \Phi x \in \mathbb{R}^M$ be the compressed measurements of $x$, where $\Phi \in \mathbb{R}^{M \times N}$ is the sensing matrix.

Given the compressed measurements $m$, the problem is to recover the $K$ sinusoids in (1). Many different approaches have been suggested in the literature for spectral CS recovery. Under certain conditions for matrix $\Phi$, one can recover signal $s$ through an $\ell_1$-norm optimization problem [9], [10], denoted as $\ell_1$-synthesis,

\[ \hat{s} = F \hat{\eta}, \hat{\eta} = \arg \min_{\eta \in \mathbb{R}^N} \|\eta\|_1 \text{ s.t. } \|m - \Phi F \eta\|_2 \leq \epsilon, \]

where $\epsilon$ is an appropriately chosen bound on the noise level, $F$ is the orthonormal DFT basis, and $\eta$ is the DFT coefficients of signal samples $s$. The optimal recovery of signal $s$ by optimizing (2) is feasible provided that the decomposition of $s$ in the DFT basis $F$ is $K$-sparse, i.e. $\|\eta\|_0 = K$ [9], [10]. Unfortunately, not only the DFT coefficients of frequency-sparse signals are not sparse, but even worse, they are just barely compressible. One way to remedy this problem would be to employ a redundant DFT frame

\[ \Psi(c) := [e(0) \ e(\Delta) \cdots e(2\pi - \Delta)], \Delta := 2\pi/cN, \]

\[ e(\omega) := \frac{1}{\sqrt{N}} \left[ 1 e^{j\omega} e^{j2\omega} \cdots e^{j\omega(N-1)} \right]^T, \]

as a substitute for $F$. But the redundant DFT frame violates the incoherence requirement for the dictionary [5].

It has recently been shown that the incoherence condition of dictionary $D$ is not necessary concerning the recovery of signal $x$, provided that the frame coefficients $D^H x$ are sufficiently sparse [12], where $(\cdot)^H$ designates the Hermitian operation. Under this circumstance, $\ell_1$-analysis yields good recovery result for signal $x$. However, the redundant DFT frame coefficients of frequency-sparse signals $\Psi(c)^H s$ do not have the sparsity property.

An alternative approach is to benefit from structured sparsity by using a coherence inhibition signal model [5]. The resulting Structured Iterative Hard Thresholding (SIHT) algorithm is able to recover frequency-sparse signals by selecting elements with low coherence in an redundant DFT frame. Other algorithms with similar flavor to SIHT include Band-excluded Orthogonal Matching Pursuit (BOMP), which takes advantage of band-exclusion [15]. However, the reconstruction fidelity of SIHT and BOMP is substantially limited due to the simplicity in the formulation of algorithms.

One way to remedy the discretization of parameter space is the polar interpolation approach, and the corresponding algorithm is named Continuous Basis Pursuit (CBP) [16]. Like other optimization based algorithms, CBP suffers from its high computational complexity. A novel algorithm, Band-excluded Interpolating Subspace Pursuit (BISP), combining the merits of band-exclusion and polar interpolation, has been proposed more recently [14], [21].

By incorporating polar interpolation with greedy algorithm, BISP improves the convergence rate of CBP while only inducing an amenable reduction in performance.

Recent advances in convex geometry has proved that frequency-sparse signals can be recovered from random subsamples via atomic norm minimization [13], which can be implemented as a semidefinite program (SDP) [22]. Though atomic norm has several appealing properties, SDP is in practice computationally expensive. Moreover, the formulation of SDP is limited to random subsampling matrix, and no discussion for arbitrary measurement settings is provided.

III. MODEL SELECTION FOR SPECTRAL COMPRESSIVE SENSING

In this paper, we adopt a parametric joint recovery-estimation method, which estimates the unknown frequencies, amplitudes, and phase shifts. Under the assumption of white Gaussian noise, with the number of sinusoids $K$ a


priori known, a common method to estimate the $3K$ parameters is by maximizing the likelihood function $L$ of observed data $x$ [17]

$$L(\theta_K, x) = \prod_{t=1}^{N} e^{-|x_t - \sum_{j=1}^{K} a_j \sin(\omega_j t + \phi_j)|^2},$$

(4)

where $\theta_K = \{a_j, \omega_j, \phi_j\}_{j=1}^{K}$ contains the $3K$ parameters of the $K$ sinusoids. In spectral CS recovery problem, the signal samples $x$ is not available, and as a substitute the compressed measurements $m$ is observed. Thus the estimation method is reformulated as the minimization of the log-likelihood function of $m$

$$\hat{\theta}_K = \arg\min_{\theta_K} -\ln L(\theta_K, m),$$

(5)

which is equivalent to an $\ell_2$-norm minimization problem

$$\hat{\theta}_K = \arg\min_{\theta_K} \left\|m - \sum_{j=1}^{K} \Phi s_j \right\|_2^2,$$

(6)

where $s_j = \{a_j \sin(\omega_j t + \phi_j)\}_{t=1}^{N}$ is the sample vector of $K$ sinusoids with estimated $3K$ parameters. Obviously, (6) can be reformulated as the following $\ell_2$-norm minimization problem

$$\hat{\theta}_K = \arg\min_{\theta_K} \left\|m - \sum_{j=1}^{K} \Phi \left(a_{1,j} \sin \omega_j + a_{2,j} \cos \omega_j\right) \right\|_2^2,$$

(7)

where $\hat{\theta}_K = \{\omega_j, a_{1,j}, a_{2,j}\}_{j=1}^{K}$ contains the reformulated $3K$ parameters with $a_{1,j} = a_j \cos \phi_j$, $a_{2,j} = a_j \sin \phi_j$ designating the amplitudes of sine and cosine sinusoids, and $\sin \omega_j = \{\sin n \omega_j\}_{n=1}^{N}$, $\cos \omega_j = \{\cos n \omega_j\}_{n=1}^{N}$ are the samples of sinusoids with frequency $\omega_j$.

In recent parametric estimation works, the best $K$ matching sinusoids are iteratively [17]. In this paper, we reformulate previous methods as a spectral CS recovery method as shown in Algorithm 1, which deviates from previous approaches in that the input is compressed measurements rather than signal samples. In each iteration, the estimated compressed measurements of other $K-1$ sinusoids are trimmed from input compressed measurements $m$, and then the parameters of the best matching sinusoid to the residual measurements $r$ are estimated through function $R(\Phi, r)$.

The function $\{\hat{\omega}, \hat{a}, \hat{\phi}\} = R(\Phi, r)$, as shown in Algorithm 2, estimates the parameters $\hat{\theta} = \{\hat{\omega}, \hat{a}, \hat{\phi}\}$ of the best matching sinusoid from residual CS measurement vector $r$. The algorithm solves the parametric estimation by minimizing the following log-likelihood function

$$\hat{\theta} = \arg\min_{\theta} -\ln L(\theta, r),$$

(8)

which is equivalent to an $\ell_2$-norm minimization problem

$$\hat{\theta} = \arg\min_{\theta} \left\|r - \Phi \left(a_1 \sin \omega + a_2 \cos \omega\right) \right\|_2^2,$$

(9)

where $\theta = \{\omega, a_1, a_2\}$ contains the reformulated sinusoid parameters. (9) can be rewritten in matrix form

$$\hat{\theta} = \arg\min_{\theta} \left\|r - A_\omega a \right\|_2^2,$$

(10)

where $A_\omega = [\Phi \sin \omega, \Phi \cos \omega]$ is composed of the compressed measurements of sinusoid samples at frequency $\omega$, and $a = [a_1, a_2]^T$ contains the amplitudes of sine and cosine sinusoids, respectively. Given frequency $\omega$ a priori known, (10) is converted to the estimation of amplitudes $a$ by solving an $\ell_2$-norm minimization problem

$$\hat{a} = \arg\min_{a \in \mathbb{R}^2} \left\|r - A_\omega a \right\|_2^2,$$

(11)

to which the solution is given in a simple form $\hat{a} = A_\omega^\dagger r$, where $A_\omega^\dagger = (A_\omega^H A_\omega)^{-1} A_\omega^H$ denotes the pseudoinverse of matrix $A_\omega$. Applying (11) on (10) yields the following two-step optimization problem, which estimates frequency $\omega$ and amplitudes $a$ consecutively.

$$\hat{\omega} = \arg\min_{\omega \in [0, \pi]} \left\|r - A_\omega A_\omega^\dagger r \right\|_2^2, \hat{a} = A_\omega^\dagger r.$$
Algorithm 1: Model Selection for Spectral Compressive Sensing

Input: CS matrix $\Phi \in \mathbb{R}^{M \times N}$, CS measurement vector $m \in \mathbb{R}^{M}$, sparsity $K$
Output: Reconstructed frequency-sparse signal $\hat{s}$

Initialize: $\hat{s}(j) = 0$, $j = 1, \ldots, K$

while halting criterion false do
  for $i = 1$ to $K$ do
    \{form residual measurement\}
    $r \leftarrow m - \sum_{j \neq i}^{K} \Phi \hat{s}(j)$
    \{estimate sinusoid parameters\}
    $\{\hat{\omega}_i, \hat{a}_i, \hat{\phi}_i\} \leftarrow R(\Phi, r)$
    \{form sinusoid estimate\}
    $\hat{s}(i) \leftarrow \left\{ \hat{a}_i \sin\left( \hat{\omega}_i t + \hat{\phi}_i \right) \right\}_{t=1}^{N}$
  end for
end while

return $\hat{s} \leftarrow \sum_{j=1}^{K} \hat{s}(j)$

Algorithm 2: Sinusoid Parametric Estimation $R(\Phi, r)$

Input: CS matrix $\Phi \in \mathbb{R}^{M \times N}$, residual CS measurement vector $r \in \mathbb{R}^{M}$
Output: Sinusoid parameter estimates $\hat{\omega}, \hat{a}, \hat{\phi}$

Initialize: $\alpha = 0$, $\beta = \pi$, $S = \infty$

while halting criterion false do
  \{frequency estimate\}
  $\{\omega_i\}_{i=0}^{N} \leftarrow \{\alpha + i (\beta - \alpha) / N\}_{i=0}^{N}$
  for $i = 0$ to $N$ do
    \{calculate compressed measurement\}
    $A_{\omega_i} \leftarrow [\Phi \sin(\omega_i), \Phi \cos(\omega_i)]$
    \{calculate square error\}
    $S_{\omega_i} \leftarrow \| r - A_{\omega_i} A_{\omega_i}^\dagger r \|_2^2$
    if $S_{\omega_i} < S$ then
      $S \leftarrow S_{\omega_i}, j \leftarrow i$
      \{update parameter estimate\}
      $\hat{a} = [\hat{a}_1, \hat{a}_2]^T \leftarrow A_{\omega_i}^\dagger r, \hat{\omega} \leftarrow \omega_i$
    end if
  end for
  \{frequency range refinement\}
  $\alpha \leftarrow \max \{\omega_{j-1}, \alpha\}, \beta \leftarrow \min \{\omega_{j+1}, \beta\}$
end while

return $\hat{\omega}, \hat{a} \leftarrow (\hat{a}_1^2 + \hat{a}_2^2)^{\frac{1}{2}}, \hat{\phi} \leftarrow \arctan(\hat{a}_2/\hat{a}_1)$

In Algorithm 2, the estimation of frequency $\omega$ and amplitudes $a$ are carried out in an iterative form to gradually converge to the numerical solution to (9). In each iteration, the minimum $\ell_2$-norm error is calculated for each frequency point sampled from predetermined frequency range with equal interval, and then, the frequency range is shrunk to the neighborhood of the frequency point with least $\ell_2$-norm error, through which the estimation precision is improved.

IV. Numerical Experiments

We compared the reconstruction performance of our algorithm, denoted as MDS, to state-of-the-art methods including $\ell_1$-analysis, $\ell_1$-synthesis, SIHT, BISP, BOMP, SDP and CBP, from both noisy and noise-free measure-
We chose the normalized $\ell_2$-norm error as major performance measure, which is defined as $\|x - \hat{x}\|_2 / ||x||_2$, between the original signal $x$ and the recovered signal $\hat{x}$. To evaluate and compare to other algorithms the frequency estimation performance of MDS, we generated frequency-sparse signals of length $N = 128$ composed of $K = 3$ real-valued sinusoids with frequencies selected uniformly at random, unit amplitudes, and zero phase shifts. The frequencies were well-separated so that no two tones were closer than $\pi/N$ to keep compatible with algorithms based on coherence inhibition signal model, including SIHT, BOMP, and BISP. Such separation is reasonable due to recent theoretical advances in relation between signal recoverability and minimum separation between spectral spikes \cite{13}, \cite{23}. Additionally, we generated frequency-sparse signals with same frequencies but amplitudes and phase shifts selected uniformly at random to evaluate the parametric estimation performance of MDS. The two parameter settings are denoted in our experiment as MDS-FREQ and MDS-SINU, respectively. We performed Monte Carlo experiments and averaged over 600 trials. The sensing matrix $\Phi$ was chosen as the Gaussian random matrix\footnote{For the SDP algorithm we used a random subsampling matrix, as the algorithm is only defined for such a sensing matrix.} and the redundant DFT frame was with $c = 5$.

In the first experiment, we evaluated the reconstruction performance from noiseless measurements with $M$ varying from 15 to 65. We set $\epsilon = 10^{-10}$ for relevant algorithms. The results of numerical experiment are shown in Figure \ref{fig:noiseless}. In the noiseless case, SDP obtains the best result. When the number of measurements is sufficiently large, the frequency estimation performance of MDS outperforms CBP, whereas for $M$ smaller than 40 it is worse than CBP and $\ell_1$-synthesis, while still better than other algorithms. The parametric estimation performance of MDS is similar to the frequency estimation. Among other algorithms, CBP outperforms $\ell_1$-synthesis and remains static precision level for a wide range of $M$. Though the redundant DFT coefficients recovered by $\ell_1$-synthesis is actually not sparse and exhibits severe frequency mismatch phenomenon, the signal is in practice reconstructed accurately. The performance of $\ell_1$-analysis, BOMP, and SIHT is the worst among the algorithms tested.

In the second experiment, we included additive Gaussian noise in the signal model. We fixed $M = 64$ and explored

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{noiseless.png}
\caption{Signal reconstruction performance in noiseless case.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{noisy.png}
\caption{Signal reconstruction performance in noisy case.}
\end{figure}

\footnote{The authors would like to thank Marco F. Duarte, Karsten Fyhn, Boaz Nadler and Gongguo Tang (listed in alphabetical order) for providing the implementation of their algorithms.}
### Table I: Average computation times in seconds.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Noiseless</th>
<th>Noisy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell_1$-analysis</td>
<td>232.5267</td>
<td>329.2279</td>
</tr>
<tr>
<td>$\ell_1$-synthesis</td>
<td>9.6257</td>
<td>12.8537</td>
</tr>
<tr>
<td>SIHT</td>
<td>1.0950</td>
<td>1.1129</td>
</tr>
<tr>
<td>SDP</td>
<td>39.2334</td>
<td>59.1094</td>
</tr>
<tr>
<td>BOMP</td>
<td>0.0398</td>
<td>0.0396</td>
</tr>
<tr>
<td>CBP</td>
<td>133.7628</td>
<td>115.8614</td>
</tr>
<tr>
<td>BISP</td>
<td>17.7901</td>
<td>15.7219</td>
</tr>
<tr>
<td>MDS-FREQ</td>
<td>0.9900</td>
<td>0.9499</td>
</tr>
<tr>
<td>MDS-SINU</td>
<td>1.1507</td>
<td>1.0233</td>
</tr>
</tbody>
</table>

In the noisy case, the frequency estimation performance of our algorithm outperforms other algorithms, and when random amplitudes and phase shifts are involved, only negligible deteriorations are induced. This is because model selection relies less on signal sparsity, and more on the matching to superposition of sinusoids. Among other algorithms, CBP and SDP obtain the best result, and $\ell_1$-synthesis exhibits high fidelity when SNR level is sufficiently high. Despite its satisfactory performance in low SNR level, the performance of BISP is mediocre when noise level is low.

The computation time is of equal importance, and the average computation times are listed in Table I. The table shows that the excellent performance of frequency estimation and parameter estimation of MDS is enhanced by its high computational efficiency. Moreover, it is observed that the distinguished performance of CBP and SDP are restrained by their high computational expense. In addition, MDS only requires the sensing matrix to have the RIP.

This flexibility on measurement scheme increases its performance advantage over SDP.

#### References


---

3We set $M = 64$ for the noiseless case and SNR = 30 for the noisy case.