LARGE SCALE 2D SPECTRAL COMPRESSED SENSING IN CONTINUOUS DOMAIN

Jian-Feng Cai[1], Weiyu Xu[2], and Yang Yang[3]

[1]Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China

[2]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA

[3]Department of Mathematics, University of Iowa, Iowa City, IA 52242 USA

ABSTRACT

We consider the problem of spectral compressed sensing in continuous domain, which aims to recover a 2-dimensional spectrally sparse signal from partially observed time samples. The signal is assumed to be a superposition of s complex sinusoids. We propose a semidefinite program for the 2D signal recovery problem. Our model is able to handle large scale 2D signals of size 500×500, whereas traditional approaches only handle signals of size around 20×20.

Index Terms— Compressed sensing, sparse recovery, Toeplitz matrices, matrix completion

1. INTRODUCTION

Spectral compressed sensing aims to recover a spectrally sparse signal from sub-Nyquist sampling of time samples. Consider a frequency-sparse signal $X^\star$ which consists of a superposition of s complex sinusoids with 2-dimensional frequencies $\boldsymbol{f}_p = (f_{p1}, f_{p2})$, where $f_{pi} \in [0, 1), i = 1, 2$. Suppose we observe m regularly-spaced time samples at times (j, k) which form an index set $T \subset \{0, \ldots, n - 1\} \times \{0, \ldots, n - 1\}$. We would like to recover the signal of the following form:[1]

$$X^\star(j, k) = \sum_{p=1}^{s} c_p e^{i2\pi(f_{p1}j + f_{p2}k)},$$
$$(j, k) \in T,$$
(1.1)

where c_p is the coefficient of each complex sinusoid.

In the foundational works of Candès *et al.* [1] and Donoho [2], successful signal recoveries from very few samples were guaranteed with high probabilities under the assumption that the unknown frequencies lie on a grid. However, in real-world applications, frequencies can possibly lie anywhere on the continuous domain [3, 4]. If they are still assumed to lie

on a certain grid, the recovery might suffer from the "basis mismatch" issue and the recovery might be far from the original true signal [5]. Other approaches for frequency-sparse signal recovery have also been proposed, for example, finite rate of innovation (FROI) sampling [6, 7]. Recently, people have been working on the recovery based on continuous frequency domain. Tang *et al.* [8] showed that $\mathcal{O}(s \log s \log n)$ random samples guarantee exact frequency recovery of a 1D spectrally sparse signal

$$\boldsymbol{x}^\star(j) = \sum_{p=1}^{s} c_p e^{i2\pi f_p j}, \; j = 0, \ldots, n - 1,$$
(1.2)

with high probability over continuous frequency domain, given the condition that the frequencies are well separated. An atomic norm minimization approach was proposed therein, which was reformulated as an equivalent solvable semidefinite program (SDP). Despite the 1D case, there are more applications involving the recovery of d-dimensional signals ($d \geq 2$). For example, super-resolution imaging [9] and nuclear magnetic resonance (NMR) spectroscopy [10]. Chi and Chen [11, 12] extended Tang's approach to 2D frequency models. However, the equivalence between the atomic norm minimization and the proposed SDP was not guaranteed. Xu *et al.* [13] proposed equivalent SDP formulations of atomic norm minimization to recover signals with 2D off-the-grid frequencies using theories from trigonometric polynomials. Moreover, the SDP formulations were generalized to higher dimensions as well.

In Chi's and Xu's approaches, the $n \times n$ signal matrix X appeared as an $n^2 \times 1$ vector in the positive semidefinite constraints. The resulting SDP was on a space of dimension at least $\mathcal{O}(n^2) \times \mathcal{O}(n^2)$. For example, to recover a 100×100 signal X, we would have to solve an SDP of size at least 10001×10001, which is computationally prohibitive.

In this paper, we propose a model to recover 2D spectrally sparse signals, which reduces the size of the SDP to $\mathcal{O}(n) \times \mathcal{O}(n)$, while at the same time requires a small number of observed samples to guarantee exact recovery with high probability, which means the model is able to handle large scale 2D signals compared to existing approaches.

The remainder of this paper is organized as follows. In

[1]Here we consider square matrices for simplicity, non-square cases can be readily generalized.

Section 2, we introduce our proposed SDP formulation for large scale 2D signal recovery. In Section 3, we give numerical experiments to validate our model.

2. SEMIDEFINITE FORMULATION FOR 2D SIGNAL RECOVERY

Notice that the 2D signal (1.1) has the matrix form

$$X^\star = V_1 C V_2^H$$
$$= \sum_{p=1}^{s} c_p \boldsymbol{v_{p1}} \boldsymbol{v_{p2}}^H \qquad (2.1)$$

where C is a diagonal matrix and V_1 and V_2 are Vandermonde matrices of the following form[2]:

$$C = \begin{bmatrix} c_1 & & \\ & \ddots & \\ & & c_s \end{bmatrix},$$

$$V_1 = [\boldsymbol{v_{11}}, \cdots, \boldsymbol{v_{s1}}]$$
$$= \begin{bmatrix} e^{i2\pi f_{11}0} & \cdots & e^{i2\pi f_{s1}0} \\ \vdots & \ddots & \vdots \\ e^{i2\pi f_{11}(n-1)} & \cdots & e^{i2\pi f_{s1}(n-1)} \end{bmatrix}, \qquad (2.2)$$

$$V_2 = [\boldsymbol{v_{12}}, \cdots, \boldsymbol{v_{s2}}]$$
$$= \begin{bmatrix} e^{-i2\pi f_{12}0} & \cdots & e^{-i2\pi f_{s2}0} \\ \vdots & \ddots & \vdots \\ e^{-i2\pi f_{12}(n-1)} & \cdots & e^{-i2\pi f_{s2}(n-1)} \end{bmatrix}.$$

We assume that $s \ll n$. The ith ($i = 1, 2$) frequency components f_{pi} are distinct across p ($p = 1, \ldots s$), so that rank $(X^\star) = s$ and can be considered low-rank. The recovery problem can thus be formulated as the following rank minimization problem:

$$\min_{X, V_1, C, V_2} \quad \text{rank}(X)$$
$$\text{s.t.} \quad X = V_1 C V_2^H \qquad (2.3)$$
$$\text{where } V_1, V_2 \text{ are full rank Vandermonde matrices,}$$
$$C \text{ diagonal as in (2.2)},$$
$$X(j, k) = X^\star(j, k), (j, k) \in T,$$

where T is the set of indices of the observed entries and $|T| = m$.

---

[2]In (2.2), the negative exponents in entries of V_2 arise from the conjugate transpose of $\boldsymbol{v_{p2}}$ in (2.1) in order to be consistent with the original form of the signal (1.1).

We propose the following semidefinite program for Problem (2.3)

$$\min_{T_1, T_2, X} \quad \frac{1}{2}\text{trace}(T_1) + \frac{1}{2}\text{trace}(T_2)$$
$$\text{s.t.} \quad \begin{bmatrix} T_1 & X \\ X^H & T_2 \end{bmatrix} \succeq 0, \qquad (2.4)$$
$$T_1, T_2 \text{ are Toeplitz,}$$
$$X(j, k) = X^\star(j, k), (j, k) \in T,$$

The advantage of this SDP over the ones proposed in [12, 13] is that the size of the semidefinite constraint is $2n \times 2n$ instead of $(n^2 + 1) \times (n^2 + 1)$, which significantly reduces the memory needed for computation.

2.1. Why it works

Our SDP model (2.4) is inspired from the SDP model in [14]

$$\min_{W_1, W_2, X} \quad \frac{1}{2}\text{trace}(W_1) + \frac{1}{2}\text{trace}(W_2)$$
$$\text{s.t.} \quad \begin{bmatrix} W_1 & X \\ X^H & W_2 \end{bmatrix} \succeq 0, \qquad (2.5)$$
$$X(j, k) = X^\star(j, k), (j, k) \in T,$$

which is equivalent to the problem of minimizing the *nuclear norm* of a matrix X:

$$\min_X \quad ||X||_*$$
$$\text{s.t.} \quad X(j, k) = X^\star(j, k), (j, k) \in T. \qquad (2.6)$$

It has been proved to be an effective approach in the *low-rank matrix completion problem* [15].

Our model (2.4) for 2D signal recovery can be viewed as an extension of the low-rank matrix completion problem in the following sense:

First, model (2.4) will yield a low-rank solution. Notice that it minimizes the trace of a positive semidefinite matrix $\begin{bmatrix} T_1 & X \\ X^H & T_2 \end{bmatrix}$, which is equivalent to minimizing the sum of all the (nonnegative) eigenvalues. It can be viewed as a relaxation of minimizing the number of nonzero eigenvalues, therefore minimizing $\text{rank}\left(\begin{bmatrix} T_1 & X \\ X^H & T_2 \end{bmatrix}\right)$ and hence $\text{rank}(X)$.

Second, we would like to incorporate the additional structure of the true signal matrix $X^\star = V_1 C V_2^H$ into the recovered solution X_{rec}, therefore imposing the Toeplitz constraints on matrices T_1 and T_2 so that the problem becomes our model (2.4). This idea works due to the Caratheodory-Toeplitz Lemma [8], which states that for Toeplitz positive semidefinite matrices T_1 and T_2, we have the following decompositions:

$$T_1 = U_1 D_1 U_1^H, \quad T_2 = U_2 D_2 U_2^H \qquad (2.7)$$

where D_1, D_2 are $r \times r$ diagonal matrices and U_1, U_2 are full rank $n \times r$ Vandermonde matrices. Following the similar argument in [8], the Vandermonde decomposition (2.7) together with the positive semidefinite constraint in (2.4) imply that X must have the following form

$$X = U_1 A U_2^H \tag{2.8}$$

Notice that U_l in (2.8) have exactly the same form as V_l in (2.2) ($l = 1, 2$) (they are all Vandermonde matrices). The only difference between the true signal $X^\star = V_1 C V_2^H$ and the recovered signal $X_{\text{rec}} = U_1 A U_2^H$ is that the matrix C is required to be diagonal in the true signal, whereas in the recovered signal, there is no such restriction on the matrix A, which means that the recovered signal is a structural approximation of the true signal. Therefore, by solving Problem (2.4), we expect to get a low-rank solution X_{rec} that resembles the form of the true solution $X^\star$.

In a nutshell, (2.4) can be viewed as a low-rank matrix completion problem where the matrix has the additional structure of Vandermonde factors. We will demonstrate through numerical experiments in Section 3 that the solution X_{rec} is exactly $X^\star$ with high probability given only a small number of randomly observed time samples.

3. NUMERICAL EXPERIMENTS

We solve (2.4) by using the alternating direction method of multipliers (ADMM). To synthesize the $n \times n$ true signal matrix $X^\star$, the s frequency pairs are randomly drawn from $[0,1]^2$ with a minimum separation condition [8] $\Delta_{\min} \triangleq \min_{p \neq q}\{|f_{p1} - f_{q1}|, |f_{p2} - f_{q2}|\} \geq 1/n$, where $p, q = 1, \ldots, s$. The amplitudes $|c_p|$ are drawn randomly from the distribution $0.5 + w^2$ with w a zero mean unit variance Gaussian random variable. The phases $e^{i\phi_p}$ are drawn uniformly at random in $[0, 2\pi)$. A total of m observed entries are randomly chosen from $X^\star$. Different values of parameters m and s are used in the implementations of ADMM below to investigate the dependence of m on s. The recovery is considered successful if the recovered signal X_{rec} and the true synthetic signal $X^\star$ satisfy $\frac{\|X_{\text{rec}} - X^\star\|_F}{\|X^\star\|_F} \leq 10^{-3}$.

3.1. ADMM Details

To apply ADMM, we first reformulate (2.4) as

$$\min_{M,N} \quad \text{trace}(M)$$
$$\text{s.t.} \quad M \succeq 0,$$
$$N = \begin{bmatrix} T_1 & X \\ X^H & T_2 \end{bmatrix}, \tag{3.1}$$
$$T_1, T_2 \text{ are Toeplitz}$$
$$X(j,k) = X^\star(j,k), (j,k) \in T,$$
$$M - N = 0.$$

This new formulation splits the semidefinite constraint and the Toeplitz constraint so that they contain M and N separately.

ADMM consists of the following three updates:

$$M^{k+1} = \arg\min_M \quad \text{trace}(M) + \frac{\rho}{2}\|M - N^k + U^k\|_F^2 \tag{3.2}$$
$$\text{s.t.} \quad M \succeq 0,$$

$$N^{k+1} = \arg\min_N \quad \frac{\rho}{2}\|M^{k+1} - N + U^k\|_F^2 \tag{3.3}$$
$$\text{s.t.} \quad N = \begin{bmatrix} T_1 & X \\ X^H & T_2 \end{bmatrix},$$
$$T_1, T_2 \text{ are Toeplitz,}$$
$$X(j,k) = X^\star(j,k), (j,k) \in T,$$

$$U^{k+1} = U^k + M^{k+1} - N^{k+1}, \tag{3.4}$$

where ρ is a user-specified parameter in ADMM [16] . We pick $\rho = 0.1$ in the experiments.

M-update can be simplified to

$$M^{k+1} = V D_+ V^H, \tag{3.5}$$

where D_+ is D with all negative diagonal entries replaced with 0. And $V D V^H$ is the eigendecomposition of $N^k - U^k - \frac{1}{\rho}I$.

N-update contains 4 blocks that can be simplified to

$$T_1^{k+1}(i+l, i) = \sum_{i=\max\{1, 1-l\}}^{\min\{n-l, n\}} [M_1^{k+1} + U_1^k](i+l, i)/(n-l), \tag{3.6}$$

$$T_2^{k+1}(i+l, i) = \sum_{i=\max\{1, 1-l\}}^{\min\{n-l, n\}} [M_2^{k+1} + U_2^k](i+l, i)/(n-l), \tag{3.7}$$

for each $l = -(n-1), \ldots, 0, \ldots, n-1$
and $\forall i = \max\{1, 1-l\}, \ldots, \min\{n-l, n\}$.

$$X^{k+1}(j,k) = \begin{cases} X^\star(j,k), & (j,k) \in T, \\ \left[M_3^{k+1} + U_3^k\right](j,k), & \text{otherwise.} \end{cases} \tag{3.8}$$

ADMM terminates when both the primal residual $r^k = M^k - N^k$ and dual residual $s^k = \rho(N^{k-1} - N^k)$ satisfy the following conditions, as detailed in [16].

$$\|r^k\|_F \leq 2n\epsilon^{\text{abs}} + \epsilon^{\text{rel}} \max\{\|M^k\|_F, \|N^k\|_F\}, \tag{3.9}$$
$$\|s^k\|_F \leq 2n\epsilon^{\text{abs}} + \epsilon^{\text{rel}}\|\rho U^k\|_F, \tag{3.10}$$

where ϵ^{abs} and ϵ^{rel} are user-specified parameters and here we set both of them to be 10^{-5}. Once the algorithm terminates, the upper right block of N is the output X_{rec}. The entire ADMM algorithm can be summarized as:

Algorithm 1 ADMM for (2.4)

1: **Input** $X^{\star}$, the observed indices set T
2: **while** (3.9) and (3.10) are not satisfied simultaneously
3:      **do** (3.5)(3.6)(3.7)(3.8)(3.4)
4: **Output** X_{rec}.

3.2. Phase Transition

We investigate the dependence of m on s. Here we fix $n = 50$ and implement ADMM with different m and s values. For each (m, s) pair, the simulation is repeated 20 times. The gray level indicates the number of successful recoveries among all 20 repetitions. In Figure 1, we compare between the phase transition plots of the proposed model (2.4) and the nuclear norm minimization in [14] (i.e., after removing the Toeplitz constraint in (2.4)).

Fig. 1. m vs s phase transition plots of model (2.4) (left) and the nuclear norm minimization (right)

We can see that the Toeplitz constraint in our model improves the phase transition because it captures the structure of the signal (i.e., the Vandermonde factors). Also, Figure 1 (left) indicates that the dependence of m on s is linear.

3.3. Comparison with Existing Approach

Here we compare our approach with Chi's approach (i.e., solving SDP (16) in [12]). We implemented all the experiments on a PC with a 8-core Intel i7-4790 3.60 GHz CPU and 16 GB memory. First, we synthesized 2D signal matrices with different sizes n from $s = 5$ frequency components and randomly drew m time samples. Then we implemented Chi's approach via CVX, and our proposed SDP (2.4) via both CVX and ADMM. The running times are listed in Table 1. We can see that as the size of the SDP slowly increases, the time required to solve the SDP in [12] grows quickly, whereas the time required for our proposed SDP remains small. Actually, as n goes beyond 23, the SDP in [12] crashes since the size of the SDP is too large to compute. On the contrary, our proposed SDP is still capable of handling large scale problems in a reasonable amount of time, which will be demonstrated in Section 3.4.

n	m	[12] via CVX	(2.4) via CVX	(2.4) via ADMM
15	80	38.52	0.55	0.050
16	90	60.57	0.26	0.059
17	100	103.47	0.29	0.060
18	110	161.27	0.49	0.063
19	120	239.09	0.39	0.064
20	130	363.61	0.43	0.074
21	140	534.03	0.68	0.072
22	150	850.19	0.54	0.076
23	160	crashed	0.60	0.087

Table 1. Comparison of running times (in seconds) between existing approach [12] and the proposed SDP formulation (2.4) for different sizes n and observations m

3.4. A Large Scale Example

We implemented ADMM on SDP (2.4) to recover a 500×500 signal generated from $s = 10$ random frequencies on $[0, 1) \times [0, 1)$ with $m = 5000$ (2%) randomly observed samples. The ADMM implementation details are demonstrated in Figure 2. The relative error $\frac{\|X^{\star} - X_{\text{rec}}\|_F}{\|X^{\star}\|_F} = 1.4633 \times 10^{-4}$, indicating successful recovery. The Fast ADMM by Goldstein *et al.*[17] was applied to accelerate convergence. The running time is around 24 min on the same PC.

Fig. 2. Objective values vs number of iterations(left) Primal and dual residuals vs number of iterations (right)

The size of SDP (2.4) for this problem is 1000×1000. If the approaches in [12, 13] were adopted, it would become $250,001 \times 250,001$, which would be impossible to compute via existing solvers. Therefore, our approach achieved better performance in terms of memory savings.

4. CONCLUSION

In this paper, we proposed a model for the recovery of large scale 2D spectrally sparse signal in continuous domain. This model is able to handle 2D signal matrices with much larger size (500×500) compared with existing approaches (around 20×20). We demonstrated through numerical experiment that the proposed SDP formulation indeed outperforms existing approaches in terms of running time and memory.

5. REFERENCES

[1] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[2] David L Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] Chaitanya Ekanadham, Daniel Tranchina, and Eero P Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE transactions on signal processing*, vol. 59, no. 10, pp. 4735–4744, 2011.

[4] Christopher E Parrish and Robert D Nowak, "Improved approach to lidar airport obstruction surveying using full-waveform data," *Journal of Surveying Engineering*, vol. 135, no. 2, pp. 72–82, 2009.

[5] Yuejie Chi, Louis L Scharf, Ali Pezeshki, and A Robert Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[6] Martin Vetterli, Pina Marziliano, and Thierry Blu, "Sampling signals with finite rate of innovation," *IEEE transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[7] Thierry Blu, Pier-Luigi Dragotti, Martin Vetterli, Pina Marziliano, and Lionel Coulot, "Sparse sampling of signal innovations," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.

[8] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[9] Michael J Rust, Mark Bates, and Xiaowei Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," *Nature methods*, vol. 3, no. 10, pp. 793–796, 2006.

[10] Jiaxi Ying, Hengfa Lu, Qingtao Wei, Jian-Feng Cai, Di Guo, Jihui Wu, Zhong Chen, and Xiaobo Qu, "Hankel matrix nuclear norm regularized tensor completion for n-dimensional exponential signals," *arXiv preprint arXiv:1604.02100*, 2016.

[11] Yuejie Chi and Yuxin Chen, "Compressive recovery of 2-d off-grid frequencies," in *2013 Asilomar Conference on Signals, Systems and Computers*. IEEE, 2013, pp. 687–691.

[12] Yuejie Chi and Yuxin Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1030–1042, 2015.

[13] Weiyu Xu, Jian-Feng Cai, Kumar Vijay Mishra, Myung Cho, and Anton Kruger, "Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional (d ≥ 2) off-the-grid frequencies," in *Information Theory and Applications Workshop (ITA), 2014*. IEEE, 2014, pp. 1–4.

[14] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[15] Emmanuel J Candès and Benjamin Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[16] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[17] Tom Goldstein, Brendan O'Donoghue, Simon Setzer, and Richard Baraniuk, "Fast alternating direction optimization methods," *SIAM Journal on Imaging Sciences*, vol. 7, no. 3, pp. 1588–1623, 2014.