A Fast Algorithm for Reconstruction of Spectrally Sparse Signals in Super-Resolution

Jian-Feng Cai[a], Suhui Liu[a], and Weiyu Xu[b]

[a]Department of Mathematics, University of Iowa, Iowa City, IA 52242;
[b]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242.

ABSTRACT

We propose a fast algorithm to reconstruct spectrally sparse signals from a small number of randomly observed time domain samples. Different from conventional compressed sensing where frequencies are discretized, we consider the super-resolution case where the frequencies can be any values in the normalized continuous frequency domain $[0, 1)$. We first convert our signal recovery problem into a low rank Hankel matrix completion problem, for which we then propose an efficient feasible point algorithm named projected Wirtinger gradient algorithm(PWGA). The algorithm can be further accelerated by a scheme inspired by the fast iterative shrinkage-thresholding algorithm (FISTA). Numerical experiments are provided to illustrate the effectiveness of our proposed algorithm. Different from earlier approaches, our algorithm can solve problems of large scale efficiently.

Keywords: Spectral compressed sensing, low rank, Hankel matrix, super resolution

1. INTRODUCTION

The task of reconstructing signals from their sampling measurements arises frequently from signal processing, and it has numerous practical applications in radar, sonar, array processing, wireless communication, seismology, fluorescence microscopy, etc. In this paper, we are interested in recovering *spectrally sparse signals* from their small number of discrete time samples. Here we say a signal is spectrally sparse if it is a weighted sum of 1-dimensional(1-D) complex sinusoids at, say, R distinct *continuous* frequencies in the unit interval $[0, 1)$. Such a signal model covers signals in various applications, for example, in acceleration of medical imaging,[1] analog-to-digital conversion,[2] and inverse scattering in seismic imaging.[3]

Early conventional approaches, such as Prony's method,[4] ESPRIT,[5] and the matrix pencil method,[6] use sampling rates satisfying the Nyquist-Shannon sampling theorem. The new technique compressed sensing (CS)[7, 8] can significantly reduce the sampling rates that the Nyquist-Shannon sampling theorem requires, provided that the signal of interest is spectrally sparse. However, CS generally assumes the frequencies take values on a finite discrete subset of the frequency domain $[0, 1)$. When the discretization is not fine enough, this will cause *basis mismatch*[9] in signal recovery, where non-negligible signal recovery errors occur from the impact of discretization errors on CS signal recovery procedures. One may make the grid discretization very fine and reduce signal recovery error,[10] which, however, will leads to an undesirably large dictionary for signal recovery.

Recently there have been growing interests in designing new algorithms which can recover the continuous-valued parameters *precisely* even from a small number of discrete nonuniform time samples. In,[11] the authors proposed to use total variation minimization to find continuous-valued shifted delta functions from equi-spaced low-frequency samples. In[12] and,[13] the atomic norm[14] and the nuclear norm of Hankel matrices are minimized respectively to recover spectrally sparse signals with continuous-valued frequencies from nonuniform samples. Though robust signal recovery is guaranteed theoretically, these convex optimization based methods in[11–13] are not computationally efficient. They are all implemented by semi-definite programming (SDP) whose variables

are matrices containing $O(N^2)$ entries with N the dimension of the signal. To solve the resulting SDP's, off-the-shelf algorithms such as SDPT3[15] use interior point methods which requires computing a Hessian matrix of size $O(N^4)$ in its Newton step. First-order methods, such as alternating direction method of multipliers (ADMM) and proximal point algorithm (PPA), need a dual matrix that is unstructured,[16] and, consequently, these algorithms require memory of size $O(N^2)$. Therefore, these convex optimization approaches are not suitable for recovering signals of large dimensions.

To efficiently recover large scaled signals, this paper proposes a fast algorithm, which first converts the signal recovery problem to a low-rank Hankel matrix completion problem and then employ a projected Wirtinger gradient descent (PWGD) algorithm. Since our proposed algorithm is a feasible point algorithm, it uses only $O(NR)$ memory. Since the number of sinusoids, R, is usually much smaller than N, the proposed algorithm provides efficient large scale signal recovery. To speed up our proposed algorithm, an acceleration technique scheme similar to FISTA[17] is given. The practical applicability of our algorithm is validated by numerical experiments, which show our algorithms can recover large scaled signals as a superposition of multiple sinusoids.

The paper is organized as follows. In Section 2, we describe our signal model, give essential concepts about Hankel matrix, and formulate the signal recovery problem. Our iterative algorithm is presented in Section 3. In Section 4, some numerical experiments are provided to demonstrate the performance of our algorithm. We also provide an acceleration for the proposed algorithm. Finally, we conclude our paper with a discussion of future work.

2. PROBLEM FORMULATION

In this section, we give some preliminaries on our signal model and the formulation of the signal reconstruction problem considered in this paper.

2.1 Signal model

The signal of our interest $x^{(\text{true})}(t)$, $t \in \mathbb{R}$, is assumed as a linear combination of complex sinusoids at R distinct frequencies $f_k^{(\text{true})} \in [0, 1)$ for $1 \leq k \leq R$, i.e.,

$$x^{(\text{true})}(t) = \sum_{k=1}^{R} d_k^{(\text{true})} e^{2\pi \iota f_k^{(\text{true})} t}, \quad t \geq 0,$$

where $\iota = \sqrt{-1}$. Here the frequencies $f_k^{(\text{true})}$'s are normalized to be in $[0, 1)$ so that the signal can be uniquely determined by its time domain samples at integer points, and the associated coefficients $d_k^{(\text{true})}$'s are the complex amplitudes. This model covers a wide range of signals in wireless communication, biology, automation, imaging science, seismology, etc.

To reconstruct the signal $x^{(\text{true})}(t)$, early methods (e.g. Prony's method, the matrix pencil method, MUSIC) need time domain samples on uniformly sampled integer time points. More specifically, they use the following M samples in the time domain $x^{(\text{true})}(t)$, $t = 0, 1, \ldots, M-1$; and then, in order to get the frequencies of $x^{(\text{true})}(t)$, these early methods used linear algebra techniques involving linear structured matrices such as Hankel and Toeplitz matrices. However, due to physical measurement limitations, it is usually hard to get all the M samples of $x^{(\text{true})}(t)$, $t = 0, 1, \ldots, M-1$, especially for signals with very high frequencies (before normalization).[2] So in this paper, we will consider non-uniform sampling in the time domain. We denote the underlying uniformly-sampled true signal as

$$\boldsymbol{x}^{(\text{true})} = [x^{(\text{true})}(0), \ x^{(\text{true})}(1), \ldots, \ x^{(\text{true})}(2N-2)]^T \in \mathbb{C}^{2N-1},$$

where N is a large integer. However, we consider the case where only M $(M < 2N-1)$ entries of $\boldsymbol{x}^{(\text{true})}$ are observed. More precisely, let $\Theta \subseteq \{0, 1, \ldots, 2N-2\}$ be the set of indices of observed entries of $\boldsymbol{x}^{(\text{true})}$, and we will reconstruct the true vector $\boldsymbol{x}^{(\text{true})}$ from

$$\boldsymbol{y} = \boldsymbol{x}_{\Theta}^{(\text{true})} := \{x_t^{(\text{true})} \mid t \in \Theta\}. \tag{1}$$

In this way, the sampling rate is significantly reduced. The same signal model is also considered in.[7, 12, 13]

2.2 Hankel Matrix Completion

Our proposed algorithm first converts the reconstruction of $\boldsymbol{x}^{(\text{true})}$ from (1) to a Hankel matrix completion problem. This strategy has also been used in.[13]

Let $\mathcal{H}$ be a linear operator that maps a vector in $\mathbb{C}^{2N-1}$ to a $N \times N$ Hankel matrix as follows

$$\mathcal{H} : \ \boldsymbol{x} \in \mathbb{C}^{2N-1} \longrightarrow \mathcal{H}\boldsymbol{x} \in \mathbb{C}^{N \times N}, \qquad [\mathcal{H}\boldsymbol{x}]_{jk} = x_{j+k}, \quad 0 \le j, k \le N-1.$$

Define $\boldsymbol{H}^{(\text{true})} = \mathcal{H}\boldsymbol{x}^{(\text{true})}$. It can be checked that the rank of $\boldsymbol{H}^{(\text{true})}$ is R, due to the following factorization

$$\mathcal{H}\boldsymbol{x}^{(\text{true})} = \begin{bmatrix} 1 & \cdots & 1 \\ e^{2\pi \imath f_1^{(\text{true})}} & \cdots & e^{2\pi \imath f_R^{(\text{true})}} \\ \vdots & \vdots & \vdots \\ e^{2\pi \imath (M-1) f_1^{(\text{true})}} & \cdots & e^{2\pi \imath (M-1) f_R^{(\text{true})}} \end{bmatrix} \begin{bmatrix} d_1^{(\text{true})} & & \\ & \ddots & \\ & & d_R^{(\text{true})} \end{bmatrix} \begin{bmatrix} 1 & e^{2\pi \imath f_1^{(\text{true})}} & \cdots & e^{2\pi \imath (M-1) f_1^{(\text{true})}} \\ \vdots & \vdots & & \vdots \\ 1 & e^{2\pi \imath f_R^{(\text{true})}} & \cdots & e^{2\pi \imath (M-1) f_R^{(\text{true})}} \end{bmatrix}$$

Then, instead of constructing the true signal $\boldsymbol{x}^{(\text{true})}$ directly, we reconstruct the rank-R Hankel matrix $\mathcal{H}\boldsymbol{x}^{(\text{true})}$. Since $\mathcal{H}$ is one-to-one from a vector in $\mathbb{C}^{2N-1}$ to an $N \times N$ Hankel matrix, one can easily convert the reconstructed Hankel matrix back to a signal.

Now the signal reconstruction problem is formulated as

$$\begin{aligned} \text{Find} \qquad & \text{matrix } \boldsymbol{X} \\ \text{subject to} \quad & \text{rank}(\boldsymbol{X}) \le R, \\ & X_{jk} = H_{jk}^{(\text{true})}, \ (j,k) \in \Omega, \\ & \boldsymbol{X} \text{ is Hankel}, \end{aligned} \tag{2}$$

where

$$\Omega = \{(j,k) \mid j+k \in \Theta, \ 0 \le j, k \le N-1\}$$

is the positions of known entries in $\boldsymbol{H}^{(\text{true})}$. Since $\mathcal{H}$ is one-to-one from $\mathbb{C}^{2N-1}$ to the set of all $N \times N$ Hankel matrix, reconstructing $\boldsymbol{x}^{(\text{true})}$ is equivalent to reconstructing $\boldsymbol{H}^{(\text{true})}$. The problem (2) is referred to low-rank Hankel matrix completion.

Following generic low-rank matrix completion,[18] (2) is converted in[13] to a rank minimization problem and further relaxed to

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_* \quad \text{s.t.} \quad X_{jk} = H_{jk}^{(\text{true})}, \ (j,k) \in \Omega, \ \text{and } \boldsymbol{X} \text{ is Hankel}. \tag{3}$$

Here $\|\cdot\|_*$ is the sum of all the singular values, namely the nuclear norm. It was shown that, if Θ is uniformly randomly drawn from all subsets of $\{0, 1, \ldots, 2N-2\}$ with cardinality $M \ge O(R \log^4 N)$, and certain separation conditions between frequencies are satisfied, then the solution of (3) recover $\boldsymbol{H}^{(\text{true})}$ perfectly with dominant probability. Similar models are considered in.[19]

Though (3) is a convex optimization problem, there were no efficient ways to compute it for large problem dimensions. The unknown matrix is $N \times N$, compared with $O(N)$ unknowns in $\boldsymbol{x}^{(\text{true})}$. One may convert (3) to an SDP and then employ available packages such as SDPT3.[15] However, these packages use second-order methods, which require solving a huge linear system of order $O(N^2) \times O(N^2)$ at each step. Also, it is not straightforward[16] to adapt nuclear norm minimization algorithms (e.g.[20]) for generic low-rank matrix completion to solving (3), as the Hankel constraint invokes $O(N^2)$ linear equality constraints.

In this paper, instead of considering convex optimizations (3), we aim at attacking the original non-convex problem (2) directly. Non-convex algorithms has been proven to have the advantage of fast convergence in sparsity and low-rank reconstruction.[21, 22] We propose an efficient algorithm based on projected Wirtinger gradient descent for this particular spectrally sparse signal recovery problem.

3. ALGORITHMS

In this section, we present our projected Wirtinger gradient algorithm for solving (2). Let us define the set of all complex-valued matrices with rank no greater than R as

$$\mathscr{R}_{\mathbb{C}}^R = \{\boldsymbol{L} \in \mathbb{C}^{N \times N} | \mathrm{rank}(\boldsymbol{L}) \le R\}. \tag{4}$$

Similarly, define the set of all complex-valued Hankel matrices that are consistent with the observed data

$$\mathscr{H} = \{\mathcal{H}\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{C}^{2N-1}, \ \boldsymbol{x}_\Theta = \boldsymbol{x}_\Theta^{(\mathrm{true})}\}. \tag{5}$$

The set $\mathscr{R}_{\mathbb{C}}^R$ is a smooth manifold and $\mathscr{H}$ is an affine space. Then, our signal recovery problem and also (2) can be formulated as the following optimization problem

$$\min_{\boldsymbol{L} \in \mathscr{R}_{\mathbb{C}}^R, \boldsymbol{H} \in \mathscr{H}} \quad \frac{1}{2}\|\boldsymbol{L} - \boldsymbol{H}\|_F^2. \tag{6}$$

We will employ a projected gradient descent algorithm to solve (6). The objective $F(\boldsymbol{L}, \boldsymbol{H}) := \frac{1}{2}\|\boldsymbol{L} - \boldsymbol{H}\|_F^2$ is a real-valued function with complex variables, which is not differentiable in the ordinary complex calculus sense. Nevertheless, $F(\boldsymbol{L}, \boldsymbol{H})$ is differentiable with respect to the real and imaginary parts of its variables. Thus, our gradient flow is performed on the real and imaginary parts respectively. Thanks to the Wirtinger calculus,[23] our proposed algorithm is given as follows: at iteration t, we have

$$\begin{cases} \boldsymbol{L}_{t+1} \in \mathcal{P}_{\mathscr{R}_{\mathbb{C}}^R}(\boldsymbol{L}_t - \delta_1(\boldsymbol{L}_t - \boldsymbol{H}_t)), \\ \boldsymbol{H}_{t+1} \in \mathcal{P}_{\mathscr{H}}(\boldsymbol{H}_t - \delta_2(\boldsymbol{H}_t - \boldsymbol{L}_{t+1})), \end{cases} \tag{7}$$

where $\delta_1 > 0$ and $\delta_2 > 0$ are step sizes, and $\mathcal{P}_{\mathscr{R}_{\mathbb{C}}^R}$ and $\mathcal{P}_{\mathscr{H}}$ are projections onto $\mathscr{R}_{\mathbb{C}}^R$ and $\mathscr{H}$ respectively. We call (7) projected Wirtinger gradient descent (PWGD).

It remains to find out $\mathcal{P}_{\mathscr{R}_{\mathbb{C}}^R}$ and $\mathcal{P}_{\mathscr{H}}$ respectively. Since $\mathcal{P}_{\mathscr{R}_{\mathbb{C}}^R}(\boldsymbol{X})$ is the best rank-R approximation to $\boldsymbol{X}$, according to Eckhart-Young Theorem,[24]

$$\mathcal{P}_{\mathscr{R}_{\mathbb{C}}^R}(\boldsymbol{X}) = \boldsymbol{U}_R \boldsymbol{\Sigma}_R \boldsymbol{V}_R^*,$$

where the columns of $\boldsymbol{U}_R$ and $\boldsymbol{V}_R$ are the first R left and right singular vectors of $\boldsymbol{X}$ respectively and $\boldsymbol{\Sigma}_R$ is a diagonal matrix with diagonals corresponding singular values. By direct calculation, the closed form of $\mathcal{P}_{\mathscr{H}}$ is given by

$$\mathcal{P}_{\mathscr{H}}(\boldsymbol{X}) = \mathcal{H}\boldsymbol{z}, \quad \text{where} \quad z_j = \begin{cases} x_j^{(\mathrm{true})}, & \text{if } j \in \Theta, \\ \mathrm{mean}\{X_{kl} \mid k + l = j\}, & \text{otherwise.} \end{cases} \tag{8}$$

The proposed PWGD algorithm (7) is a feasible point algorithm. The iterates $\boldsymbol{L}_t$ and $\boldsymbol{H}_t$ are always in their feasible sets $\mathscr{R}_{\mathbb{C}}^R$ and $\mathscr{H}$ respectively. This property can significantly reduce the computational cost and storage, when R is small compared to N. Since $\boldsymbol{L}_t \in \mathscr{R}_{\mathbb{C}}^R$, it is stored in a factorization form and only $O(NR)$ memory is necessary. Also, the Hankel matrix $\boldsymbol{H}_t$ can be represented by its parameters, which is of size only $O(N)$. Furthermore, in Step 1 of (7), it needs to compute only the first R singular values and their corresponding singular vectors of $\boldsymbol{L}_t - \delta_1(\boldsymbol{L}_t - \boldsymbol{H}_t)$ in the computation of the projection. This can be done by, e.g., Krylov subspace methods, which invokes only the matrix-vector product of $\boldsymbol{L}_t - \delta_1(\boldsymbol{L}_t - \boldsymbol{H}_t)$. For the matrix-vector product of $\boldsymbol{L}_t$, since $\boldsymbol{L}_t$ is rank R and in a factorization form, it can be done in $O(NR)$ operations. The matrix-vector product of the Hankel matrix $\boldsymbol{H}_t$ is implemented by fast Fourier transform,[24] which needs only $O(N \log N)$ operations. Step 2 of (7) needs averages of $\boldsymbol{L}_{t+1}$ along anti-diagonals.

4. NUMERICAL EXPERIMENTS

In this section, we use numerical experiments to demonstrate the effectiveness and efficiency of our proposed algorithms.

4.1 Phase Transition

We first illustrate that our proposed algorithm is able to recovery spectrally sparse signals from their very limited time domain samples. We fix the dimension of the signal to be 127 (i.e. $N = 64$), and we vary the sparsity R and the number of samples M. For each (R, M) pair, 100 Monte Carlo trials were conducted. For each trial, the true signal is synthesized by randomly generating the true frequencies $f_k^{(\text{true})}$'s and magnitudes $d_k^{(\text{true})}$'s, which are independently uniformly distributed on $[0, 1)$ and the unit circle respectively. We then get M samples uniformly at random. The PWGD is executed by setting the parameters $\delta_1 = \delta_2 = 0.9999$, and we stop the algorithm when $\|\boldsymbol{H}_{t+1} - \boldsymbol{H}_t\|_F / \|\boldsymbol{H}_t\|_F \le 10^{-4}$. Signal recovery for each trial is considered successful if the relative error satisfies $\|\hat{\boldsymbol{x}} - \boldsymbol{x}^{(\text{true})}\|_2 / \|\boldsymbol{x}^{(\text{true})}\|_2 \le 5 \times 10^{-3}$, where $\hat{\boldsymbol{x}}$ denotes the solution returned by the PWGD. Figure 1 illustrates the results of the Morte Carlo experiments. Here the horizontal axis corresponds to the number M of the samples (i.e. the size of the observation location set Θ), while the vertical axis corresponds to the sparsity level R. The empirical success rate is reflected by the color of each cell. It can be seen from the figure that our proposed algorithm have a high rate of successful recovery if M exceeds than certain thresholds for a given R.
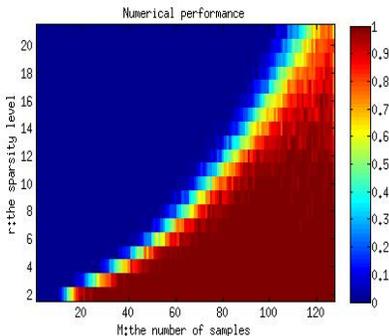

Figure 1: Phase transition plots where frequency locations are randomly generated. The empirical success rate is calculated by averaging 100 Monte Carlo trials.

4.2 Signals of large dimension

Next we demonstrate that our proposed algorithm is able to recover signals of large scale, and compare it with the Enhance Matrix Completion (EMaC) in.[13] As we have argued, different from existing convex optimization based methods such as EMaC, our proposed algorithm is able to work with high-dimensional spectrally sparse signals. In Table 1, the elapsed time for signals of different dimensions are listed. For our algorithm, we use the same settings as in the previous section. For EMaC algorithm, we used CVX software to solve it. From the table, we can see that PWGD can greatly speed up the the signal recovery for moderate dimensions and also work well for signals of high dimensions.

	PWGD	EMaC
the signal with $N = 51, R = 1, M = 10$	0.34	46.8
the signal with $N = 51, R = 3, M = 20$	0.46	58.0
the signal with $N = 101, R = 5, M = 40$	0.95	out of memory
the signal with $N = 501, R = 5, M = 100$	12.7	out of memory
the signal with $N = 2501, R = 13, M = 500$	133	out of memory
the signal with $N = 2501, R = 25, M = 1000$	91.4	out of memory
the signal with $N = 5001, R = 20, M = 1000$	645	out of memory
the signal with $N = 5001, R = 31, M = 2000$	403	out of memory

Table 1: Elapsed time in seconds for signals of different dimensions.

4.3 Acceleration by a FISTA-like Scheme

Finally, we propose a scheme to accelerate the convergence of the PWGD algorithm (7). Our scheme borrows from the fast iterative shrinkage-thresholding algorithm (FISTA),[17] which has been proven to be efficient in minimizing the sum of two convex functions with one having a Lipschitz continuous gradient. The basic idea is to use a specific linear combination of two successive iterates. Although our problem is non-convex, we still employ the linear combination scheme in FISTA for our model.

Our PWGD with FISTA scheme, called PWGD-FISTA, is constructed as follows: Given $k_0 = 1$, we generate $\{\boldsymbol{L}_t, \boldsymbol{H}_t\}$ by

$$
\begin{cases}
\boldsymbol{L}_{t+1} \in \mathcal{P}_{\mathscr{R}_{\mathbb{C}}^R}(\boldsymbol{L}_t - \delta_1(\boldsymbol{L}_t - \tilde{\boldsymbol{H}}_t)), \\
\boldsymbol{H}_{t+1} \in \mathcal{P}_{\mathscr{H}}(\boldsymbol{H}_t - \delta_2(\tilde{\boldsymbol{H}}_t - \boldsymbol{L}_{t+1})), \\
k_{t+1} = \frac{\sqrt{1+4k_t^2}+1}{2}, \\
\tilde{\boldsymbol{H}}_{t+1} = \boldsymbol{H}_{t+1} + \frac{k_t-1}{k_{t+1}}(\boldsymbol{H}_{t+1} - \boldsymbol{H}_t)
\end{cases}
\tag{9}
$$

Since $\mathscr{H}$ is an affine subspace, the linear combination in the last line of (9) does not change the feasibility of $\tilde{\boldsymbol{H}}_{t+1}$, i.e., $\tilde{\boldsymbol{H}}_{t+1} \in \mathscr{H}$. This guarantees that the computational complexity and required storage of Step 1 and Step 2 in the PWGD-FISTA algorithm are the same as that in the PWGD algorithm (7). Also, the computational effort in Step 3 and Step 4 of (9) is negligible compared with that in Step 1 and Step 2. Therefore, the PWGD-FISTA algorithm preserves the computational simplicity of the PWGD algorithm. Figure 2 depicts the convergence curves of PWGD and PWGD-FISTA. We see clearly that the PWGD-FISTA Algorithm converges faster than the PWGD algorithm. Roughly, the PWGD-FISTA needs only 2/3 number of iterations that PWGD requires to get solutions of the same accuracy.
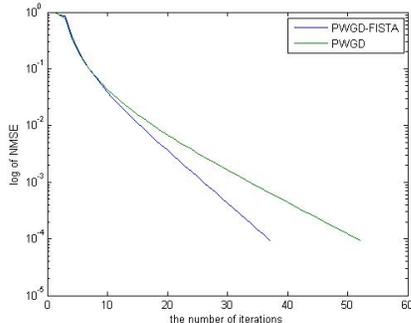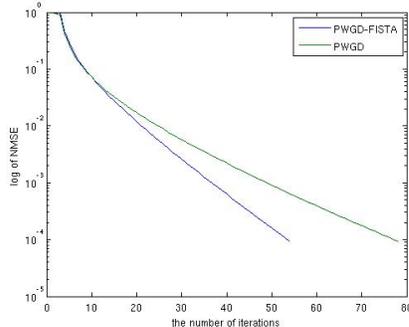
(a) $N = 501$, $R = 8$, $M = 200$ (b) $N = 5001$, $R = 20$, $M = 1000$

Figure 2: Convergence rate comparison of PWGD and PWGD-FISTA .

5. CONCLUSION

In this paper, a fast iterative algorithm is proposed for recovering spectrally sparse signals whose frequencies can be any values in the continuous domain $[0, 1)$ from a small amount of time domain samples. Different from existing algorithms, our proposed algorithm is able to deal with signals of large dimension. Inspired by the scheme in FISTA, we also provided an acceleration of the proposed algorithm. In the future, we will extend our algorithms to signal recovery from noisy samples and signals with multivariate frequencies.

REFERENCES

[1] Lustig, M., Donoho, D., and Pauly, J. M., "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine* **58**(6), 1182–1195 (2007).

[2] Tropp, J. A., Laska, J. N., Duarte, M. F., Romberg, J. K., and Baraniuk, R. G., "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *Information Theory, IEEE Transactions on* **56**(1), 520–544 (2010).

[3] Borcea, L., Papanicolaou, G., Tsogka, C., and Berryman, J., "Imaging and time reversal in random media," *Inverse Problems* **18**(5), 1247–1279 (2002).

[4] Scharf, L. L., [*Statistical signal processing*], vol. 98, Addison-Wesley Reading, MA (1991).

[5] Roy, R. and Kailath, T., "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing* **37**(7), 984–995 (1989).

[6] Hua, Y. and Sarkar, T. K., "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *Acoustics, Speech and Signal Processing, IEEE Transactions on* **38**(5), 814–824 (1990).

[7] Candès, E. J., Romberg, J., and Tao, T., "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory* **52**(2), 489–509 (2006).

[8] Donoho, D. L., "Compressed sensing," *IEEE Trans. Inform. Theory* **52**(4), 1289–1306 (2006).

[9] Chi, Y., Scharf, L. L., Pezeshki, A., and Calderbank, A. R., "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.* **59**(5), 2182–2195 (2011).

[10] Tang, G., Bhaskar, B. N., Shah, P., and Recht, B., "Sparse recovery over continuous dictionaries — just discretize," *Proceedings of 2013 Asilomar Conference on Signals, Systems and Computers*, 1043–1047 (2013).

[11] Candès, E. J. and Fernandez-Granda, C., "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics* **67**(6), 906–956 (2014).

[12] Tang, G., Bhaskar, B. N., Shah, P., and Recht, B., "Compressive sensing off the grid," *IEEE Transactions on Information Theory* **59**(11), 7465–7490 (2013).

[13] Chen, Y. and Chi, Y., "Robust spectral compressed sensing via structured matrix completion," *IEEE Transactions on Information Theory* **60**(10), 6576 – 6601 (2014).

[14] Chandrasekaran, V., Recht, B., Parrilo, P. A., and Willsky, A. S., "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics* **12**(6), 805–849 (2012).

[15] Toh, K., Todd, M., and Tutuncu, R., "SDPT3 – a Matlab software package for semidefinite programming," *Optimization Methods and Software* **11**(12), 545–581 (1999).

[16] Fazel, M., Pong, T. K., Sun, D., and Tseng, P., "Hankel matrix rank minimization with applications to system identification and realization," *SIAM J. Matrix Anal. Appl.* **34**(3), 946–977 (2013).

[17] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences* **2**(1), 183–202 (2009).

[18] Candes, E. and Recht, B., "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics* **9**(6), 717–772 (2009).

[19] Cai, J.-F., Qu, X., Xu, W., and Ye, G.-B., "Robust recovery of complex exponential signals from random gaussian projections via low rank hankel matrix reconstruction," *arXiv preprint arXiv:1503.02893* (2015).

[20] Cai, J.-F., Candès, E. J., and Shen, Z., "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimiz.* **20**(4), 1956–1982 (2010).

[21] Blumensath, T. and Davies, M. E., "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis* **27**(3), 265–274 (2009).

[22] Jain, P., Meka, R., and Dhillon, I. S., "Guaranteed rank minimization via singular value projection," in [*Advances in Neural Information Processing Systems*], 937–945 (2010).

[23] Fischer, R. F., [*Precoding and signal shaping for digital transmission*], John Wiley & Sons (2005).

[24] Golub, G. H. and Van Loan, C. F., [*Matrix computations*], Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, third ed. (1996).