Line Spectral Estimation Based on Compressed Sensing with Deterministic Sub-Nyquist Sampling

Shan Huang, Hong Sun, Haijian Zhang, and Lei Yu

Abstract—As an alternative to the traditional sampling theory, compressed sensing allows acquiring much smaller amount of data, still estimating the spectra of frequency-sparse signals accurately. However, compressed sensing usually requires random sampling in data acquisition, which is difficult to implement in hardware. In this paper, we propose a deterministic and simple sampling scheme, that is, sampling at three sub-Nyquist rates which have coprime undersampled ratios. This sampling method turns out to be valid through numerical experiments. A complex-valued multitask algorithm based on variational Bayesian inference is proposed to estimate the spectra of frequency-sparse signals after sampling. Simulations show that this method is feasible and robust at quite low sampling rates.

Index Terms—Line spectral estimation, Compressed sensing, Deterministic sub-Nyquist sampling.

I. INTRODUCTION

LINE spectral estimation has numerous applications in sonar, radar, underwater surveillance, communications, geophysical exploration, speech analysis, nuclear physics and other fields. In general, the sampling rate of the signal is required to be higher than twice the highest frequency (i.e., Nyquist rate). In some applications, high-speed analog-to-digital converters that increase the sampling rate or density are very expensive. Emerging compressed sensing (CS) goes against the common knowledge in data acquisition. CS theory asserts that one can recover certain signals and images from far fewer samples or measurements than traditional methods use [1].

Many researchers have utilized CS to estimate the spectra of frequency-sparse signals [2][3][4]. A source localization method based on a sparse representation of sensor measurements with an overcomplete basis was proposed in [5]. The authors in [6] addressed the problem of estimating spectral lines from irregularly sampled data within the framework of sparse representations. The uniqueness conditions of the sparse solution with different patterns of samples were analyzed. In [7], the effect of "basis mismatch" caused by grid discretization was analyzed. To deal with basis mismatch, some articles used grid refinement to approximate the true grid [8][9][10]. The atomic norm-based methods make line spectral estimation cast into a convex semidefinite program optimization, which deals with continuous-valued frequencies and completely eliminates basis mismatch [11][12][13]. However, these methods usually require random sampling, which is difficult or complicated to implement.

In this paper, we focus on line spectral estimation with deterministic sub-Nyquist sampling. The union of three series of undersampled samples at coprime ratios is enough to estimate the spectra of frequency-sparse signals. Then an algorithm based on variational Bayesian inference is employed to connect the samples. This method may be realized through three undersampled channels, the hardware is convenient to implement. The paper is organized as follows: Section II gives the smapling strategy. Section III demonstrates our algorithm. Simulation results are shown in Section IV. The last section draws conclusions.

II. SAMPLING STRATEGY

Consider the line spectral estimation problem where the observed signal is a summation of K complex sinusoids:

$$y(m) = \sum_{k=1}^{K} c_k e^{j\omega_k m}, \tag{1}$$

where $j = \sqrt{-1}$, $\omega_k \in [0, 2\pi)$ and c_k denote the angular frequency and the complex amplitude of the k-th component, respectively. When $m = 1, 2, \cdots, M$, it implies normal sampling, which is studied in conventional methods such as MUSIC [14]. In the methods based on CS, m is selected at random from the index set $[N] \triangleq \{1, 2, \cdots, N\}$. However, this pattern of sampling often leads to complex hardware. For example, a new type of data acquisition system called a random demodulator is studied to ensure the randomness of sampling in [15].

The proposed deterministic scheme is to sample at three coprime undersampled ratios p, q, r, in other words, we need the samples with indices

$$\mathcal{I} = \{p, 2p, \cdots\} \cup \{q, 2q, \cdots\} \cup \{r, 2r, \cdots\}. \tag{2}$$

It is worth mentioning that sampling at two coprime undersampled ratios sometimes also yields correct results but three rates guarantee a high probability of success. The process diagram of sampling is shown in Fig. 1.

After sampling, the samples are in chronological order, we select consecutive M samples with indices $t_1, t_2, \cdots, t_M$ to constitute a column vector

$$\boldsymbol{y} = \begin{bmatrix} y(t_1) & y(t_2) & \cdots & y(t_M) \end{bmatrix}^T, \tag{3}$$

where $[*]^T$ denotes the transpose operation. Assume that the frequencies are aligned with a uniform grid, i.e.,

$$\omega_n = 2\pi n/N, \ n = 1, 2, \cdots, N. \tag{4}$$

The authors are with Signal Processing Laboratory, School of Electronic Information, Wuhan University, Wuhan 430072, China (e-mail: staronice@whu.edu.cn; hongsun@whu.edu.cn; haijian.zhang@whu.edu.cn; ly.wd@whu.edu.cn).
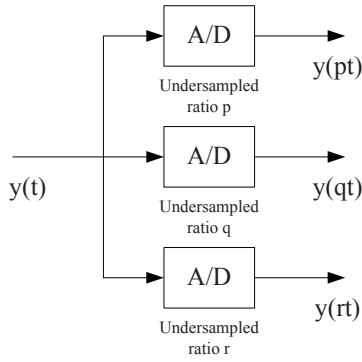
Fig. 1. The process diagram of the proposed sampling scheme.

The observation model can be written more compactly as

$$\boldsymbol{y} = \boldsymbol{\Phi}(\omega)\boldsymbol{s}, \tag{5}$$

where $\boldsymbol{\Phi}(\omega) = \begin{bmatrix} \boldsymbol{\phi}(\omega_1) & \boldsymbol{\phi}(\omega_2) & \cdots & \boldsymbol{\phi}(\omega_N) \end{bmatrix}$, $\boldsymbol{\phi}(\omega) = \begin{bmatrix} e^{j\omega t_1} & \cdots & e^{j\omega t_M} \end{bmatrix}^T$ and $\boldsymbol{s} = \begin{bmatrix} \tilde{c}_1 & \cdots & \tilde{c}_N \end{bmatrix}^T$ is a K-sparse vector. In general, $M < N$ is set and (5) is solved as a problem of sparse recovery. However, the property of $\boldsymbol{\Phi}$ as a CS matrix is difficult to certify in theory. To improve the probability of success, we utilize more samples to form multiple tasks and synthesize the effects of these tasks, namely

$$\boldsymbol{y}_l = \boldsymbol{\Phi}_l\boldsymbol{s}_l, \ l = 1, 2, \cdots, L, \tag{6}$$

where

$$\boldsymbol{\Phi}_l = \begin{bmatrix} 1 & \cdots & 1 \\ e^{j\omega_1(t_{l+1}-t_l)} & \cdots & e^{j\omega_N(t_{l+1}-t_l)} \\ \vdots & \ddots & \vdots \\ e^{j\omega_1(t_{l+M-1}-t_l)} & \cdots & e^{j\omega_N(t_{l+M-1}-t_l)} \end{bmatrix}, \tag{7}$$

$\boldsymbol{y}_l = \begin{bmatrix} y(t_l) & y(t_{l+1}) & \cdots & y(t_{l+M-1}) \end{bmatrix}^T$ and $\boldsymbol{s}_l = \begin{bmatrix} \tilde{c}_1 e^{j\omega_1 t_l} & \cdots & \tilde{c}_N e^{j\omega_N t_l} \end{bmatrix}^T$. The total number of samples is $L + M - 1$. Note that all of $\boldsymbol{s}_l$ share the same sparsity profile and $\boldsymbol{\Phi}_l$ repeat after a certain period. M is expected to be as large as possible, but an appropriate value of M must ensure $\boldsymbol{\Phi}_l$ not to contain duplicate rows. The joint estimation can achieve satisfactory results as shown in Section IV, even though $\boldsymbol{\Phi}_l$ may not have good property.

The most widely used criterion to evaluate the property of a CS matrix is *restricted isometry property* (RIP). The CS matrix $\boldsymbol{\Phi}$ has the (k, δ)-RIP if

$$(1 - \delta) \|\boldsymbol{x}\|_2^2 \le \|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \le (1 + \delta) \|\boldsymbol{x}\|_2^2 \tag{8}$$

holds for all k-sparse vectors $\boldsymbol{x}$, $\|\boldsymbol{x}\|_2$ denotes ℓ_2-norm of $\boldsymbol{x}$ [16]. The smallest δ for (k, δ)-RIP is the restricted isometry constant (RIC) δ_k. A small δ_k implies good performance when recovering a k-sparse signal. Let $\boldsymbol{\Phi}$ be a matrix with ℓ_2-normalized columns $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \cdots, \boldsymbol{\varphi}_N$, i.e., $\|\boldsymbol{\varphi}_n\|_2 = 1$ for $n = 1, 2, \cdots, N$, the condition (8) is equivalent to that the Gram matrix $\boldsymbol{\Phi}_\mathcal{K}^H \boldsymbol{\Phi}_\mathcal{K}$ of every column submatrix $\boldsymbol{\Phi}_\mathcal{K}(\mathcal{K} \subset \{1, 2, \cdots, N\}, |\mathcal{K}| \le k)$ has all its eigenvalues in the interval $[1 - \delta_k, 1 + \delta_k]$, where $[*]^H$ denotes the conjugate transpose operation.

Next we give an example in order to clearly illustrate our sampling scheme. If the three undersampled ratios are $p = 9, q = 10$ and $r = 11$, and the number of discrete grid points is $N = 100$, the configuration of the samples is

$$\overbrace{\begin{pmatrix} y(9) & y(10) & y(11) & \cdots \\ y(10) & y(11) & y(18) & \cdots \\ y(11) & y(18) & y(20) & \cdots \\ y(18) & y(20) & y(22) & \cdots \\ y(20) & y(22) & y(27) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}}^{L} \Big\} M. \tag{9}$$

When 50 tasks are used, i.e., $L = 50$, it is better to choose $M = 27$ to prevent the corresponding sensing matrices from having duplicate rows. The sensing matrix $\boldsymbol{\Phi}_l$ amounts to picking partial rows from the $N \times N$ Fourier matrix. The sensing property of this deterministic partial Fourier matrix approximates a random partial Fourier matrix, which has been proven to be appropriate as a CS matrix [17].

We select $\boldsymbol{\Phi}_1$ to present the statistical RIP of $\boldsymbol{\Phi}_l$ intuitively, the maximum and minimum eigenvalues of its Gram matrices are plotted. These eigenvalues of a random partial Fourier matrix are also plotted for comparison. The data are obtained from $k^2 N$ sub-Gram matrices for each k. The solid lines sketch the average values of maximum and minimum eigenvalues of all sub-Gram matrices and the dashed lines sketch the limiting values. Fig. 2 shows that the eigenvalues of $\boldsymbol{\Phi}_1$'s sub-Gram matrices distribute slightly further away from 1 than the random partial Fourier matrix. In Section IV we will see that the probability of success increases significantly when more tasks are introduced.

III. PROPOSED ALGORITHM

In this section, a complex-valued multitask algorithm based on variational Bayesian inference is proposed to solve the above model. In [18], the multitask Bayesian CS algorithm utilized empirical Bayesian analysis to recover multiple real-valued sparse solutions. We also address the problem within the hierarchical Bayesian framework. Assume the measurement noise to be independent and complex Gaussian with zero-mean and variance equal to β^{-1}, the model (6) can be rewritten as

$$\boldsymbol{y}_l = \boldsymbol{\Phi}_l\boldsymbol{s}_l + \boldsymbol{\epsilon}_l, \ l = 1, 2, \cdots, L. \tag{10}$$

The likelihood function for the parameters $\boldsymbol{s}$ and β may be expressed as

$$p\left(\boldsymbol{y}_l|\boldsymbol{s}_l, \beta\right) = (\pi/\beta)^{-M} \exp\left(-\beta \|\boldsymbol{y}_l - \boldsymbol{\Phi}_l\boldsymbol{s}_l\|_2^2\right). \tag{11}$$

The hierarchical Gaussian prior is typically imposed on $\boldsymbol{s}_l$ in sparse Bayesian leaning to induce sparsity. Denote the prior variance of the i-th element of $\boldsymbol{s}_l$ as α_i^{-1}, the prior distribution of $\boldsymbol{s}_l$ is

$$p\left(\boldsymbol{s}_l|\boldsymbol{\alpha}\right) = \frac{1}{\pi^N |\boldsymbol{A}|^{-1}} \exp\left(-\boldsymbol{s}_l^H \boldsymbol{A}\boldsymbol{s}_l\right), \tag{12}$$

(a) $\boldsymbol{\Phi}_1$

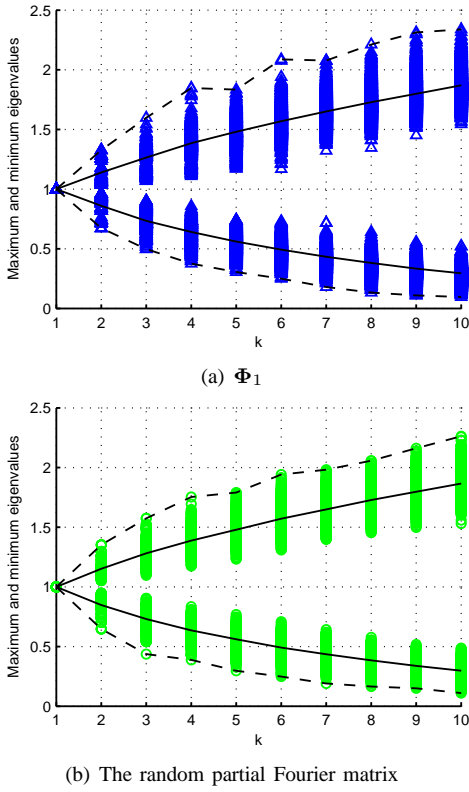(b) The random partial Fourier matrix

Fig. 2. Maximum and minimum eigenvalues of sub-Gram matrices for different k. (a) $\boldsymbol{\Phi}_1$; (b) The random partial Fourier matrix.

where $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_N \end{bmatrix}^T$ and $\boldsymbol{A} = \text{diag}(\boldsymbol{\alpha})$. Gamma priors are placed on the hyperparameters $\boldsymbol{\alpha}$, and similarly on the noise precision β, i.e.,

$$p(\boldsymbol{\alpha}|a,b) = \prod_{i=1}^{N} \text{Gamma}(\alpha_i|a,b), \quad (13)$$

$$p(\beta|c,d) = \text{Gamma}(\beta|c,d). \quad (14)$$

The parameters a, b, c and d are typically set to very small values (e.g., $a = b = c = d = 10^{-6}$), which amounts to assuming uninformative priors for $\boldsymbol{\alpha}$ and β [19].

Define $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_L \end{bmatrix}$ and $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}_1 & \cdots & \boldsymbol{s}_L \end{bmatrix}$, the joint probability of data, parameters and hyperparameters is

$$p(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{\alpha}, \beta) = p(\boldsymbol{Y}|\boldsymbol{S}, \beta) \cdot p(\boldsymbol{S}|\boldsymbol{\alpha}) \cdot p(\boldsymbol{\alpha}|a,b) \cdot p(\beta|c,d)$$

$$= \prod_{l=1}^{L} p(\boldsymbol{y}_l|\boldsymbol{s}_l, \beta) \cdot \prod_{l=1}^{L} p(\boldsymbol{s}_l|\boldsymbol{\alpha}) \cdot p(\boldsymbol{\alpha}|a,b) \cdot p(\beta|c,d). \quad (15)$$

By applying the variational expectation maximization (EM) algorithm [20] and the above equations, the posterior distributions of $\boldsymbol{S}$, $\boldsymbol{\alpha}$ and β can be approximately calculated as

$$\ln q(\boldsymbol{S}) = \langle \ln p(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{\alpha}, \beta) \rangle_{q(\boldsymbol{\alpha})q(\beta)} + \text{const}, \quad (16)$$

$$\ln q(\boldsymbol{\alpha}) = \langle \ln p(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{\alpha}, \beta) \rangle_{q(\boldsymbol{S})q(\beta)} + \text{const}, \quad (17)$$

and

$$\ln q(\beta) = \langle \ln p(\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{\alpha}, \beta) \rangle_{q(\boldsymbol{S})q(\boldsymbol{\alpha})} + \text{const}, \quad (18)$$

where $\langle * \rangle_{q(x)}$ is the expectation with respect to $q(x)$.

Substituting (15) into (16), after some arrangement we find that the vector $\boldsymbol{s}_l$ obeys a complex Gaussian distribution, i.e.,

$$q(\boldsymbol{s}_l) = \mathcal{CN}(\boldsymbol{s}_l|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l), \quad l = 1, 2, \cdots, L. \quad (19)$$

The mean $\boldsymbol{\mu}_l$ and covariance matrix $\boldsymbol{\Sigma}_l$ are given by

$$\boldsymbol{\mu}_l = \langle \beta \rangle \boldsymbol{\Sigma}_l \boldsymbol{\Phi}_l^H \boldsymbol{y}_l, \quad (20)$$

$$\boldsymbol{\Sigma}_l = \left(\langle \beta \rangle \boldsymbol{\Phi}_l^H \boldsymbol{\Phi}_l + \langle \boldsymbol{A} \rangle \right)^{-1}. \quad (21)$$

According to (15) and (17), it can be shown that the posterior density of $\boldsymbol{\alpha}$ is

$$q(\boldsymbol{\alpha}) = \prod_{i=1}^{N} \text{Gamma}\left(\alpha_i|\tilde{a}, \tilde{b}_i\right), \quad (22)$$

where

$$\tilde{a} = a + L, \quad (23)$$

$$\tilde{b}_i = b + \left\langle \sum_{l=1}^{L} |s_{l,i}|^2 \right\rangle, \quad (24)$$

$s_{l,i}$ is the i-th element of $\boldsymbol{s}_l$. Similarly, we obtain

$$q(\beta) = \text{Gamma}\left(\beta|\tilde{c}, \tilde{d}\right), \quad (25)$$

where

$$\tilde{c} = c + LM, \quad (26)$$

$$\tilde{d} = d + \left\langle \sum_{l=1}^{L} \|\boldsymbol{y}_l - \boldsymbol{\Phi}_l \boldsymbol{s}_l\|_2^2 \right\rangle. \quad (27)$$

Utilizing the property of Gamma distribution, the required expected values can be computed as

$$\langle \alpha_i \rangle = \frac{\tilde{a}}{\tilde{b}_i}, \quad (28)$$

$$\langle \beta \rangle = \frac{\tilde{c}}{\tilde{d}}. \quad (29)$$

Based on the above results, the procedure of the algorithm can be summarized as follows:

1) Set the iteration count to 0. Initialize $\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l, \boldsymbol{\alpha}$ and β.

2) According to (22)-(29) and the current estimated values of $\boldsymbol{\mu}_l$ and $\boldsymbol{\Sigma}_l$, update the posterior distributions of $\boldsymbol{\alpha}$ and β.

3) According to (19)-(21) and the current posterior densities of $\boldsymbol{\alpha}$ and β, update the the estimated values of $\boldsymbol{\mu}_l$ and $\boldsymbol{\Sigma}_l$.

4) Return to Step 2) until the iteration count reaches the maximum value.

After using this algorithm, the lowest several valleys of $\boldsymbol{\alpha}$ indicate the positions of the frequencies contained in the signal. When the frequencies do not fall onto the grid, the algorithm often finds the nearest grid point. So the closest interval of the true frequencies can not be too small.

IV. SIMULATION RESULTS

We make experiments to verify the effect of multiple tasks relative to single task. The signals contain $K = 3$ frequency components with amplitudes of 0.2, 0.4 and 0.8 respectively and random phase angles. The three undersampled ratios are set to $p = 9$, $q = 10$ and $r = 11$. According to the analysis in Section II, $M = 27$ and $N = 100$ are fixed. We

set the frequencies to 0.178, 0.353 and 0.372, respectively. Complex white Gaussian noise at SNR=20dB is added to the measurements. The power spectra with respect to different numbers of tasks $L = 1$, 10 and 30 are plotted. Meanwhile, in spite of impracticality of random sampling, we construct multiple random samples in the program to compare with our method. As shown in Fig. 3, the performance of estimation is improved as the number of tasks L. When only one task is utilized, the first frequency component is not really obvious. When $L = 1$ and $L = 10$, the last two frequency components are not clearly distinguished. When $L = 30$, the proposed method achieves the same effects with random sampling.
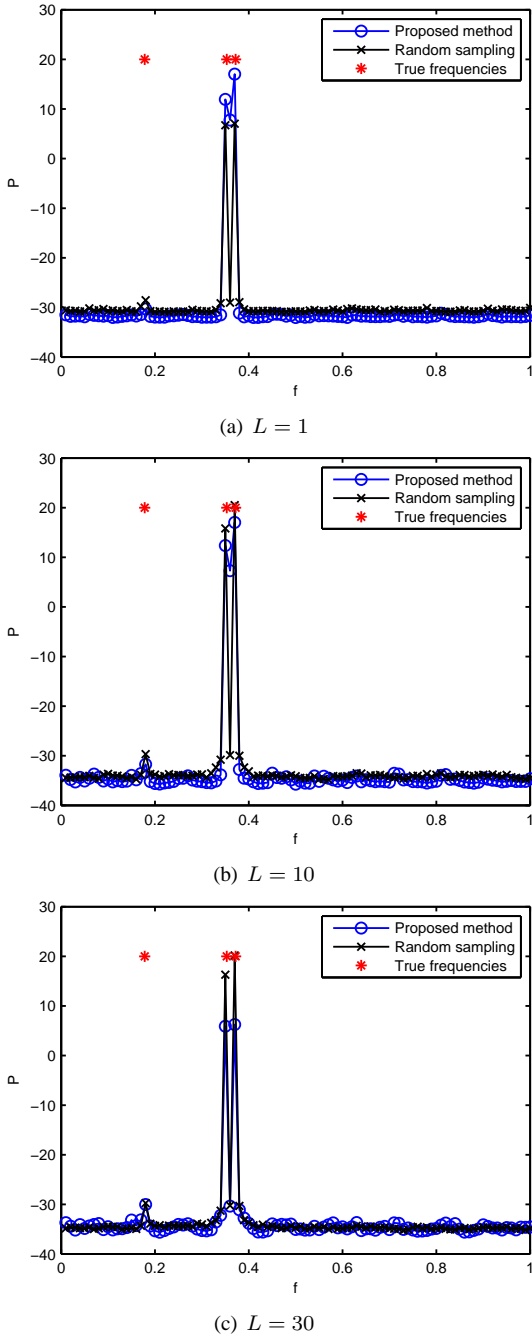
Then we test the performance of the proposed method in different noisy environment. The three undersampled ratios are set to $p = 7$, $q = 8$ and $r = 9$. $M = 32$, $N = 100$ and $L = 30$ are fixed and the SNR varies from 10dB to 30dB. The signals contain $K = 3$ frequency components with random amplitudes and random phase angles. To keep it simple, we assume K is known, so the K frequencies corresponding to maximum K peaks in power spectrum are estimated results. If the deviation of all estimated frequencies from true frequencies are within $0.5/N$, we say this trial is successful. The probabilities of success are obtained from 500 trials for each SNR. We compare the success probabilities of the proposed method with that of random sampling. The MUSIC algorithm using normal sampling is also considered for comparison. The same number of samples are used for our method and MUSIC. As shown in Fig. 4, the proposed method and MUSIC have approximately the same probabilities of success for different SNRs, which are slightly lower than that of random sampling.
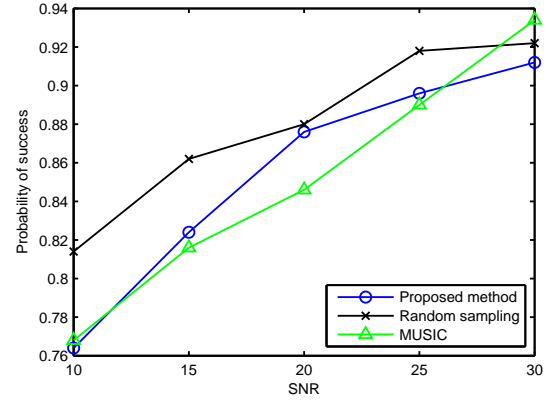

Fig. 4. The probabilities of success for different SNRs.

V. CONCLUSION

In this letter, we proposed a deterministic sampling scheme to replace the unpractical random sampling. Three sub-Nyquist analog-to-digital converters which have coprime undersampled ratios are shown to be enough to estimate the spectra of frequency-sparse signals. The property of the corresponding CS matrices is verified through numerical simulations. Then an algorithm based on variational Bayesian inference is proposed to solve the multitask model. Simulations show that this method possesses as good performance as conventional MUSIC with normal sampling. We believe that this method can improve the practicability of CS in line spectral estimation.

Fig. 3. The estimated power spectra with respect to different numbers of tasks L. (a) $L = 1$; (b) $L = 10$; (c) $L = 30$.

REFERENCES

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

[2] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, 2011.

[3] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.

[4] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6576–6601, 2014.

[5] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[6] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 575–585, 2007.

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

[8] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4658–4671, 2012.

[9] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3809–3822, 2012.

[10] J. Fang, J. Li, Y. Shen, H. Li, and S. Li, "Super-resolution compressed sensing: An iterative reweighted algorithm for joint parameter learning and sparse signal recovery," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 761–765, 2014.

[11] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[12] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[13] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139–3153, 2015.

[14] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[15] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[16] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[17] E. J. Candès, J. Romberg, and T. 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.

[18] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 92–106, 2009.

[19] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.

[20] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.