

Iterative Reweighted l_1 Penalty Regression Approach for Line Spectral Estimation

Fei Ye, Xian Luo, Wanzhou Ye

College of Science, Shanghai University, Shanghai, China

Email: wzhy@shu.edu.cn

How to cite this paper: Ye, F., Luo, X. and Ye, W.Z. (2018) Iterative Reweighted l_1 Penalty Regression Approach for Line Spectral Estimation. *Advances in Pure Mathematics*, 8, 155-167.

<https://doi.org/10.4236/apm.2018.82008>

Received: January 14, 2018

Accepted: February 23, 2018

Published: February 26, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we proposed an iterative reweighted l_1 penalty regression approach to solve the line spectral estimation problem. In each iteration process, we first use the ideal of Bayesian lasso to update the sparse vectors; the derivative of the penalty function forms the regularization parameter. We choose the anti-trigonometric function as a penalty function to approximate the l_0 norm. Then we use the gradient descent method to update the dictionary parameters. The theoretical analysis and simulation results demonstrate the effectiveness of the method and show that the proposed algorithm outperforms other state-of-the-art methods for many practical cases.

Keywords

Line Spectral Estimation, Penalty Regression, Bayesian Lasso, Iterative Reweighted Approach

1. Introduction

Spectral estimation technology is widely used in the fields of electronic countermeasures, radar, sonar and mobile communication. In this paper, we mainly consider the line spectral estimation in compressed sensing. Considering the problem of the line spectral estimation using a pre-specified discrete Fourier transform matrix, the sparse solution we obtained may not close to the real sparse vectors when the true frequency components may not lie on the pre-specified frequency grid. This error, referred as grid mismatch, results in performance degradation or even recovery failure. Therefore, in this paper, we treat the dictionary parameters as the unknown variable along with the sparse signal, and complete the optimization of the dictionary parameters when we estimate the sparse vector through the iterative way.

Rather than applying the traditional compressed sensing theory, an increasing

number of scholars have concentrated on the grid mismatch problem instead. For example, work [1] focused on the impact of the basis mismatch on the reconstruction error by treating the error as a perturbation between the presumed and the actual dictionaries. In work [2], to handle the grid mismatch, the true dictionary is approximated as a summation of a presumed dictionary and a structured parametrized matrix via the Taylor expansion. A highly coherent dictionary was used to approximate the real dictionary in [3], and a class of greedy algorithms that use the technique of band exclusion was proposed. On the other hand, in [4] [5] [6], the grid mismatch problem was studied by proposing an atomic norm-minimization approach to handle an infinite dictionary with continuous atoms. Bayesian statistics has also been applied to solve the grid mismatch problem. In work [7] and [8], Bayesian approaches were proposed to iteratively refine the dictionary by treating the sparse signals as hidden variables. The work [8] used a generalized expectation-maximization (EM) algorithm to solve the dictionary parameters and determine the sparse vector. Work [9] proved that the problem of compressed sensing using a logarithmic penalty can be transformed into an iterative reweighted l_2 norm regression problem by providing a special surrogate function.

In addition, we analyze the first-order optimal condition of the original problem and then prove that the problem can be transformed into a series of reweighted lasso [10] problems by using the iterative method. In each step of the iteration, the derivative of the anti-triangular penalty function forms the weight of the l_1 norm. Compared with the algorithms proposed in [9] and [11], our method is more adaptive with regard to the choice of penalty function and the calculation method for the weight, additionally, the sparse effect of the l_1 norm is also better than the l_2 norm. However, it is well known that there is no explicit solution to the problem of the l_1 penalty regression. In this study, we use a Bayesian lasso approach to determine the optimal solution for each step.

The remainder of the paper is organized as follows. Section 2 is the description of the line spectral estimation problem, which we formulate as the penalty least squares problem with dictionary parameters. In Section 3, we provide a theoretical analysis and propose the iterative reweighted l_1 algorithm. In Section 4, we present several sets of numerical experiments to demonstrate that the iterative reweighted l_1 method is better than other state-of-the-art algorithms in many cases. Section 5 concludes the paper and provides some ideas for future work.

2. Model Overview

2.1. Line Spectral Estimation

Assume the line spectral estimation problem where the observed signal is a summation of a number of complex sinusoids:

$$y_n = \sum_{j=1}^k \beta_j e^{-in\theta_j} + \varepsilon_n, \quad n = 1, \dots, N \quad (1)$$

And we write it in the form of a matrix expression:

$$Y = X(\theta)\beta + \varepsilon \tag{2}$$

where $Y^T = [y_1, \dots, y_N]$ represents the observed value. $\theta^T = [\theta_1, \dots, \theta_k]$ is unknown parameters represent the frequency. $X(\theta)_{N \times k}$ ($N \ll k$) determined by parameter θ . The covariate in the model $\beta^T = [\beta_1, \dots, \beta_k]$ represents the amplitude of the corresponding frequency. $\varepsilon^T = [\varepsilon_1, \dots, \varepsilon_n]$ represents a random error term, assuming that they are independent.

2.2. Penalty Least Squares Regression

In the process of signal reconstruction, the dimension of Y is much smaller than the number of measurements ($N \ll k$). Since the signal is sparse, the Equation (2) would be transformed into an optimization problem (3):

$$\begin{aligned} \min \|\beta\|_0 \\ \text{s.t } Y = X(\theta)\beta + \varepsilon \end{aligned} \tag{3}$$

where $\|\beta\|_0$ stands for the number of the non-zero components of $\|\beta\|$. The optimization (3), however, is an NP-hard problem (which is difficult to find the solution in polynomial time). We can transform optimization (3) into a penalty least squares problem:

$$\begin{aligned} \min G(\beta) \\ \text{s.t } Y = X(\theta)\beta + \varepsilon \end{aligned} \tag{4}$$

The optimization (4) can be formulated as an unconstrained optimization problem by removing the constraint and adding a penalty term to the objective function:

$$\min_{\beta, \theta} H(\theta, \beta) = \|Y - X(\theta)\beta\|_2^2 + \lambda \sum_{i=1}^k G(\beta_i) \tag{5}$$

where λ represents the adjustable penalty parameter. Different penalty functions form different regularization parameters in the iterative process. We find that the penalty function of the inverse trigonometric function has better properties than other common penalty function such as logarithmic penalty function. In the next section we propose an iterative reweighted l_1 sparse algorithm with anti-trigonometric function penalties.

3. Iterative Reweighted l_1 Sparse Algorithm

3.1. Algorithm Description

We now develop an iterative reweighted algorithm for joint dictionary parameter learning and sparse signal recovery. Consider the line spectral estimation with anti-trigonometric penalty function:

$$\min_{\beta, \theta} H(\theta, \beta) = \|Y - X(\theta)\beta\|_2^2 + \lambda \sum_{i=1}^k \arctan(\varphi|\beta_i|) \tag{6}$$

We consider the first derivative of the problem (6). Since the absolute value is involved, we summarize the following derivative functions:

$$\frac{\partial H(\theta, \beta)}{\partial \beta_i} = \begin{cases} 2(X'(\theta)X(\theta)\beta - X'(\theta)Y)_i + \frac{\lambda\varphi}{1 + \varphi^2|\beta_i|^2}, & \beta_i > 0 \\ 2(X'(\theta)X(\theta)\beta - X'(\theta)Y)_i - \frac{\lambda\varphi}{1 + \varphi^2|\beta_i|^2}, & \beta_i < 0 \\ 2(X'(\theta)X(\theta)\beta - X'(\theta)Y)_i - \lambda C_{|\beta_i|}, & \beta_i = 0 \end{cases} \quad (7)$$

$$\frac{\partial H(\theta, \beta)}{\partial \theta} = \beta' \left(\frac{\partial X'(\theta)}{\partial \theta} X(\theta) + X'(\theta) \frac{\partial X(\theta)}{\partial \theta} \right) \beta - 2Y' \frac{\partial X'(\theta)}{\partial \theta} \beta \quad (8)$$

The penalty function $\arctan(\varphi|\beta_i|)$ cannot be guided at zero, $C_{|\beta_i|}$ represents its sub-gradient at zero which is a set of real number: $a \leq C_{|\beta_i|} \leq b$

$$a = \lim_{\beta_i \rightarrow 0^-} \frac{\arctan(\varphi|\beta_i|) - \arctan(0)}{\beta_i - 0} = -\varphi$$

$$b = \lim_{\beta_i \rightarrow 0^+} \frac{\arctan(\varphi|\beta_i|) - \arctan(0)}{\beta_i - 0} = \varphi \quad (9)$$

If we have the iteration value of step t : (θ^t, β^t) , combine with (7) we can estimate β^{t+1} by solving the next weighted lasso problem:

$$G(\beta^{t+1} | \theta^t, \beta^t) = \min_{\beta} \|Y - X(\theta)\beta^{t+1}\|_2^2 + \sum_{i=1}^k \frac{\lambda\varphi}{1 + \varphi^2|\beta_i^t|^2} |\beta_i^{t+1}| \quad (10)$$

Here we use the ideal of Bayesian lasso [11] (which we will briefly introduce later in this article) to find β^{t+1} . Similar to (7), the first derivative of (10) can be summarized as:

$$\frac{\partial G(\beta^{t+1} | \theta^t, \beta^t)}{\partial \beta_i^{t+1}} = \begin{cases} 2(X'(\theta)X(\theta)\beta - X'(\theta)Y)_i + \frac{\lambda\varphi}{1 + \varphi^2|\beta_i^t|^2}, & \beta_i > 0 \\ 2(X'(\theta)X(\theta)\beta - X'(\theta)Y)_i - \frac{\lambda\varphi}{1 + \varphi^2|\beta_i^t|^2}, & \beta_i < 0 \\ 2(X'(\theta)X(\theta)\beta - X'(\theta)Y)_i - \frac{\lambda\varphi}{1 + \varphi^2|\beta_i^t|^2} \hat{C}_{|\beta_i|}, & \beta_i = 0 \end{cases} \quad (11)$$

$\hat{C}_{|\beta_i|}$ represents the sub-gradient of $|\beta_i^{t+1}|$ at zero which is also a set of real number: $\hat{a} \leq \hat{C}_{|\beta_i|} \leq \hat{b}$

$$\hat{a} = \lim_{\beta_i^{t+1} \rightarrow 0^-} \frac{|\beta_i^{t+1}| - 0}{\beta_i^{t+1} - 0} = -1$$

$$\hat{b} = \lim_{\beta_i^{t+1} \rightarrow 0^+} \frac{|\beta_i^{t+1}| - 0}{\beta_i^{t+1} - 0} = 1 \quad (12)$$

The next step is to find θ^{t+1} . In this situation we do not need to calculate the optimal solution, instead we are going to find the estimation θ^{t+1} which satisfied:

$$\|Y - X(\theta^{t+1})\beta^{t+1}\|_2^2 \leq \|Y - X(\theta^t)\beta^{t+1}\|_2^2 \quad (13)$$

The stop condition of the algorithm is controlled by tolerance value ω_1, ω_2 . In this paper we set the tolerance value equals 0.02 in the numerical simulation. Based on the discussion above, we summarise our algorithm as follows:

Algorithm 1

Require:

Initial value β^0, θ^0 and prescribed tolerance value ω_1, ω_2
 Iterative step size $d = D$ for θ^t ;

Ensure:

- Estimated value of the optimal solution β^*, θ^* ;
- 1: **while** $\beta^{t+1} - \beta^t > \omega_1$ or $\theta^{t+1} - \theta^t > \omega_2$ **do**
 - 2: Based on the θ^t and β^t , we construct the weighted Lasso problem $G(\beta^{t+1}|\theta^t, \beta^t)$ (5), then using Gibbs sampling to estimate β^{t+1} .
 - 3: Based on the estimate β^{t+1} and objective function $H(\beta^{t+1}, \theta)$, calculate the gradient of $\theta: \Delta$.
 - 4: $\hat{\theta} = \theta^t + d * \Delta$
 - 5: **while** $H(\hat{\theta}, \beta^{t+1}) > H(\theta^t, \beta^{t+1})$ **do**
 - 6: $d = d/2$
 - 7: $\hat{\theta} = \theta^t + d * \Delta$
 - 8: **end while**
 - 9: $d = D$
 - 10: $\theta^{t+1} = \hat{\theta}$
 - 11: **end while**
-

3.2. Theoretical Analysis

First we want to prove that the objective function (6) is guaranteed to be non-increasing at each iteration:

$$H(\theta^t, \beta^t) \geq H(\theta^t, \beta^{t+1}) \geq H(\theta^{t+1}, \beta^{t+1}) \tag{14}$$

Since we obtain θ^{t+1} by the gradient descent method, it is obvious that $H(\theta^t, \beta^{t+1}) \leq H(\theta^{t+1}, \beta^{t+1})$.

On the other hand, we prove $H(\theta^t, \beta^t) \geq H(\theta^t, \beta^{t+1})$ using the next lemma which has been introduced by [12]:

LEMMA: Given that the adjust parameter $\varphi > 0$, then we have the following inequality:

$$\arctan(\varphi|\beta_i^t|) - \arctan(\varphi|\beta_i^{t+1}|) \geq \frac{\varphi}{1 + \varphi^2|\beta_i^t|^2} (|\beta_i^t| - |\beta_i^{t+1}|) \tag{15}$$

proof:

We first denote $f(x) = \arctan(x)$ and let $x \geq 0$, then by the mean value theorem we have: $f(x_1) - f(x_2) = f'(\zeta)(x_1 - x_2)$, where ζ between x_1 and x_2 .

Since $f(x)$ is an increasing function and $f'(x)$ is a decreasing function, the following inequality: $f(x_1) - f(x_2) \geq f'(x_1)(x_1 - x_2)$ is always holds for any non-negative value x_1 and x_2 . If we let $x_1 = \varphi|\beta_i^t|$ and $x_2 = \varphi|\beta_i^{t+1}|$, the inequality (15) would be certainly proved.

Next we consider the following equality:

$$\begin{aligned}
 & H(\theta^t, \beta^t) - H(\theta^t, \beta^{t+1}) \\
 &= \|Y - X(\theta^t)\beta^t\|_2^2 + \lambda \sum_{i=1}^k \arctan(\varphi|\beta_i^t|) - \|Y - X(\theta^t)\beta^{t+1}\|_2^2 \\
 &\quad - \lambda \sum_{i=1}^k \arctan(\varphi|\beta_i^{t+1}|) \\
 &= \|X(\theta^t)\beta^t\|_2^2 + \|X(\theta^t)\beta^{t+1}\|_2^2 - 2(Y - X(\theta^t)\beta^t)' X(\theta^t)\beta^t \\
 &\quad + 2(Y - X(\theta^t)\beta^{t+1})' X(\theta^t)\beta^{t+1} + \lambda \sum_{i=1}^k (\arctan(\varphi|\beta_i^t|) - \arctan(\varphi|\beta_i^{t+1}|)) \\
 &= \|X(\theta^t)\beta^{t+1} - X(\theta^t)\beta^t\|_2^2 + 2(X(\theta^t)\beta^{t+1} - X(\theta^t)\beta^t)' X(\theta^t)\beta^t \\
 &\quad - 2(Y - X(\theta^t)\beta^t)' X(\theta^t)\beta^t + 2(Y - X(\theta^t)\beta^{t+1})' X(\theta^t)\beta^{t+1} \\
 &\quad + \lambda \sum_{i=1}^k (\arctan(\varphi|\beta_i^t|) - \arctan(\varphi|\beta_i^{t+1}|)) \\
 &= \|X(\theta^t)\beta^{t+1} - X(\theta^t)\beta^t\|_2^2 + 2[(Y - X(\theta^t)\beta^{t+1})X(\theta^t)(\beta^{t+1} - \beta^t)] \\
 &\quad + \lambda \sum_{i=1}^k (\arctan(\varphi|\beta_i^t|) - \arctan(\varphi|\beta_i^{t+1}|))
 \end{aligned} \tag{16}$$

using the lemma above we can yields:

$$\begin{aligned}
 & H(\theta^t, \beta^t) - H(\theta^t, \beta^{t+1}) \\
 &\geq \|X(\theta^t)\beta^{t+1} - X(\theta^t)\beta^t\|_2^2 + 2[(Y - X(\theta^t)\beta^{t+1})X(\theta^t)(\beta^{t+1} - \beta^t)] \\
 &\quad + \lambda \sum_{i=1}^k \frac{\varphi}{1 + \varphi^2|\beta_i^t|^2} (|\beta_i^t| - |\beta_i^{t+1}|)
 \end{aligned} \tag{17}$$

where β_i^{t+1} is the optimal solution of problem (10), which satisfied:

$$\frac{\partial G(\beta^{t+1} | \theta^t, \beta^t)}{\partial \beta_i^{t+1}} = 0 \tag{18}$$

Substituting (18) to (17), we show that we have the inequality:

$$H(\theta^t, \beta^t) \leq H(\theta^t, \beta^{t+1}) \tag{19}$$

in different situations:

When: $\beta_i^{t+1} \geq 0$:

$$2[(Y - X(\theta^t)\beta^{t+1})' X(\theta^t)(\beta^{t+1} - \beta^t)] = \lambda \sum_{i=1}^k \frac{\varphi}{1 + \varphi^2|\beta_i^t|^2} (\beta_i^{t+1} - \beta_i^t) \tag{20}$$

then:

$$\begin{aligned}
 & H(\theta^t, \beta^t) - H(\theta^t, \beta^{t+1}) \\
 &\geq \|X(\theta^t)\beta^{t+1} - X(\theta^t)\beta^t\|_2^2 + \lambda \sum_{i=1}^k \frac{\varphi}{1 + \varphi^2|\beta_i^t|^2} (\beta_i^{t+1} - \beta_i^t + |\beta_i^t| - |\beta_i^{t+1}|)
 \end{aligned} \tag{21}$$

By the fact that $\beta_i^{t+1} - \beta_i^t + |\beta_i^t| - |\beta_i^{t+1}| \geq 0$ on the condition of $\beta_i^{t+1} > 0$, the inequality (19) has proved to be correct.

When: $\beta_i^{t+1} \leq 0$:

$$2 \left[(Y - X(\theta^t)\beta^{t+1})' X(\theta^t)(\beta^{t+1} - \beta^t) \right] = -\lambda \sum_{i=1}^k \frac{\varphi}{1 + \varphi^2 |\beta_i^t|^2} (\beta_i^{t+1} - \beta_i^t) \quad (22)$$

then:

$$\begin{aligned} & H(\theta^t, \beta^t) - H(\theta^t, \beta^{t+1}) \\ & \geq \|X(\theta^t)\beta^{t+1} - X(\theta^t)\beta^t\|_2^2 + \lambda \sum_{i=1}^k \frac{\varphi}{1 + \varphi^2 |\beta_i^t|^2} (\beta_i^t - \beta_i^{t+1} + |\beta_i^t| - |\beta_i^{t+1}|) \end{aligned} \quad (23)$$

By the fact that $\beta_i^t - \beta_i^{t+1} + |\beta_i^t| - |\beta_i^{t+1}| \geq 0$ on the condition of $\beta_i^{t+1} < 0$, the inequality (19) has proved to be correct.

When: $\beta_i^{t+1} = 0$:

The set of sub-gradient at β_i^{t+1} (which is $[\hat{a}, \hat{b}]$) should contains 0 when β_i^{t+1} is the critical point, thus we have the following inequality:

$$-2(X'(\theta)Y)_i - \frac{\varphi\lambda}{1 + \varphi^2 |\beta_i^t|^2} \leq 0, -2(X'(\theta)Y)_i + \frac{\varphi\lambda}{1 + \varphi^2 |\beta_i^t|^2} \geq 0 \quad (24)$$

Consider the inequality we want to prove, we can easily formulate:

$$\begin{aligned} & H(\theta^t, \beta^t) - H(\theta^t, \beta^{t+1}) \\ & \geq \|X(\theta^t)\beta^t\|_2^2 + \sum_{i=1}^k \frac{\lambda\varphi}{1 + \varphi^2 |\beta_i^t|^2} (|\beta_i^t| - \beta_i^t) \geq 0 \end{aligned} \quad (25)$$

The discussion above proves that we can ensure that the function value keeps non-increasing at each iteration. In addition we want to illustrate that β^{t+1} convergence to β^* and θ^{t+1} convergence to θ^* on the limit situation $\beta^t \rightarrow \beta^*$ and $\theta^t \rightarrow \theta^*$.

As β^{t+1} is the optimal solution of (10), we define: $\hat{\beta}^{t+1} = \lim_{\beta \rightarrow \beta^*, \theta \rightarrow \theta^*} \beta^{t+1}$.

Consider the first-order condition of β^{t+1} which satisfied:

$$\lim_{\beta \rightarrow \beta^*, \theta \rightarrow \theta^*} \frac{\partial G(\beta | \beta^t, \theta^t)}{\partial \beta_i^{t+1}} = 0 \quad (26)$$

On the other hand we have the following conclusion in the situation of $\beta^* \neq 0$:

$$\frac{\partial H(\beta, \theta)}{\partial \beta_i^*} = 0 \quad (27)$$

Compare the above two equalities we can conclude that $\hat{\beta}^{t+1} = \beta^*$.

As for the situation when $\beta^{t+1} = 0$, consider the inequality from (18):

$$\begin{aligned} & -2(X'(\theta)Y)_i - \frac{\varphi\lambda}{1 + \varphi^2 |\beta_i^*|^2} \leq 0 \\ & -2(X'(\theta)Y)_i + \frac{\varphi\lambda}{1 + \varphi^2 |\beta_i^*|^2} \geq 0 \end{aligned} \quad (28)$$

Since $\frac{\varphi\lambda}{1 + \varphi^2 |\beta_i^*|^2} \leq \varphi\lambda$, we can easily conclude:

$$\begin{aligned} -2(X'(\theta)Y)_i - \phi\lambda &\leq 0 \\ -2(X'(\theta)Y)_i + \phi\lambda &\geq 0 \end{aligned} \tag{29}$$

After the discussion before we can summarize that $\hat{\beta}^{t+1}$ always satisfied the first-order condition (7) when $\beta^t \rightarrow \beta^*$ and $\theta^t \rightarrow \theta^*$. Thus we demonstrate the limit of: $\beta^{t+1} \rightarrow \beta^*$. When $\theta^t \rightarrow \theta^*$, consider the gradient of θ^t :

$$\lim_{\theta \rightarrow \theta^*} \frac{\partial H(\beta^{t+1}, \theta)}{\partial \theta^t} = 0 \tag{30}$$

Thus the value of θ^{t+1} remains θ^* .

3.3. The Bayesian Lasso

In this article we use the ideal of Bayesian lasso to estimate the optimal solution of problem (10).

Assuming that the prior distribution of the parameter β follows the Laplace distribution:

$$f(\beta_i) = \frac{\lambda}{4\sigma^2} \exp\left(-\frac{\lambda}{2\sigma^2}|\beta_i|\right) \tag{31}$$

Combined with the likelihood function we can get the posterior probability:

$$f(\beta | Y) \propto \exp\left(-\frac{1}{2\sigma^2}\|Y - X\beta\|_2^2 - \sum_{i=1}^k \frac{\lambda}{2\sigma^2}|\beta_i|\right) \tag{32}$$

Solving problem (10) is equivalent to solving the maximal probability of posterior probability, which we can obtain from Gibbs sampling [13].

As Laplace distribution is difficult to directly derive intuitive full condition posterior distribution, the following integral (33) provides an effective solution:

$$\frac{a}{2} \exp(-a|z|) = \int_0^\infty \frac{1}{(2\pi v)^{1/2}} \exp\left(-\frac{z^2}{2v}\right) \times \frac{a^2}{2} \exp\left(-\frac{a^2 v}{2}\right) dv \tag{33}$$

Using the above integral we can rewritten the Laplace prior distribution $\beta_i \sim \frac{a}{2} \exp(-a|z|)$ by introducing the intermediate parameter v :

$$\begin{aligned} f(\beta_i) &\sim N(0, v_i) \\ f(v_i) &\sim \exp\left(-\frac{a^2}{2}v_i\right) \end{aligned} \tag{34}$$

In problem (32) we let $a = \frac{\lambda}{2\sigma^2}$.

Then we can motivate the following hierarchical Bayesian lasso model:

$$\begin{aligned} L(Y | \beta, X) &\sim N(X\beta, \sigma^2 I) \\ f(\beta | V) &\sim N(0, D_v) \\ f(v_i) &\sim \exp\left(-\frac{\lambda^2}{8\sigma^4}v_i\right) \end{aligned} \tag{35}$$

where $V = [v_1, \dots, v_k]$ corresponds to $\beta = [\beta_1, \dots, \beta_k]$ and $D_v = \text{diag}[v_1, \dots, v_k]$.

For Bayesian inference, the full condition distribution of β and v is:

$$\begin{aligned} f(\beta | V, Y) &\sim N(A^{-1}X^T Y, \sigma^2 A^{-1}) \\ f\left(\frac{1}{v_i} | \beta_i, Y\right) &\sim I-G(a^*, b^*) \end{aligned} \tag{36}$$

where $A^{-1} = X^T X + D_v$, the distribution $I-G(a^*, b^*)$ represent inverse Gaussian distribution with $a^* = \frac{\lambda^*}{\beta_i}$, $b^* = (\lambda^*)^2$ and $\lambda^* = \frac{\lambda}{2\sigma^2}$. The density of inverse

Gaussian as follows:

$$f(x, a^*, b^*) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(-\frac{-b^*(x-a^*)^2}{2(a^*)^2 x}\right) \tag{37}$$

By repeated sampling, we will form a Markov chain contains a series of point:

$$(\beta_1^t, v_1), (\beta_2^t, v_2), \dots, (\beta_m^t, v_m).$$

Since each iteration will lead to a Markov chain, we will get a long sample of β through the whole algorithm:

$$\begin{aligned} \beta_1^1, \dots, \beta_i^1, \dots, \beta_k^1 &\rightarrow \beta^2, \theta^2 \\ \beta_1^2, \dots, \beta_i^2, \dots, \beta_k^2 &\rightarrow \beta^3, \theta^3 \\ &\vdots \\ \beta_1^t, \dots, \beta_i^t, \dots, \beta_k^t &\rightarrow \beta^*, \theta^* \end{aligned} \tag{38}$$

And we have demonstrated that $H(\theta^t, \beta^t) \geq H(\theta^*, \beta^{t+1}) \geq \dots \geq H(\theta^*, \beta^*)$ in the above. The simulation results are shown in the next section.

4. Simulation Results

In this section, we carry out a series of experiments to illustrate the performance of our proposed l_1 iterative reweighted algorithm (denoted to as l_1 -IR). In our simulations, we compare our proposed algorithm with other existing state-of-the-art methods, including the sparse Bayesian learning with dictionary refinement algorithm (denoted as DicRefCS) [7], the sparse Bayesian learning with dictionary estimation (denoted as SBL-DE) [8], and especially the super-resolution iterative reweighted algorithm (denoted as SR -IR) [11].

In order to control the noise level at some of the experiments, we first give the definition of observation quality by the peak-signal-to-noise ratio ($PSNR$):

$$PSNR \equiv 10 \log_{10} (1/\sigma^2) \tag{39}$$

where σ^2 represents the variance of noise. We calculate the signal-to-noise ratio ($PSNR$) to complete the recovery effect comparison:

$$RSNR = 20 \log_{10} \left(\frac{\|\beta^*\|_2^2}{\|\beta^* - \hat{\beta}\|_2^2} \right) \tag{40}$$

where β^* represents the original sparse signal and $\hat{\beta}$ represents the signal recovered by the algorithm. Parameters λ and φ have the same effect as regularization parameters, we choose $\varphi=1$ and select optimal parameters λ by cross validation.

In the following, we examine the behaviour of respective algorithms under different scenarios. First we control the noise level $PSNR = 20$, which means $\sigma^2 = 0.01$. **Figure 1** shows the changes in $PSNR$ of the various algorithms with the change of sparseness S (the sparseness S represents the number of non-zero values in original sparse signal). We keep $K = 64$ and the sparseness S values range from 1 to 10. We set the number of measurements $N = 20$ in order to control the variables. Each data point is averaged by repeated tests.

Figure 1 indicates that the performance of the l_1 iterative reweighted algorithm is better than other state-of-the-art methods. With a low sampling number ($N = 20$), the performances of the DicRefCS and SBL-DE algorithms deteriorated quickly and failed when the sparseness S was higher than 5. On the other hand, our proposed algorithm still performed well, even when the sparseness S reached 8. It is worth mentioning that the $SR-IR$ algorithm also performed better than the other two Bayesian algorithms but not as well as the $L1-IR$ algorithm. The reasons are that the l_1 norm is sparser than the l_2 norm and we chose a better weight function.

Next, we illustrate the influence of the sample size N on the recovery effect using another set of experiments. We keep $K = 64$, the sparseness $S = 3$ and the $PSNR$ maintains at 20. We change the number of measurements N from 6 to 32. For each N , **Figure 2** shows the performance of respective algorithms.

It can be observed from **Figure 2** that, with the increase in N , the recovery performance of the respective algorithms is increased to a relatively high level but our proposed algorithm outperforms the other methods for a small number

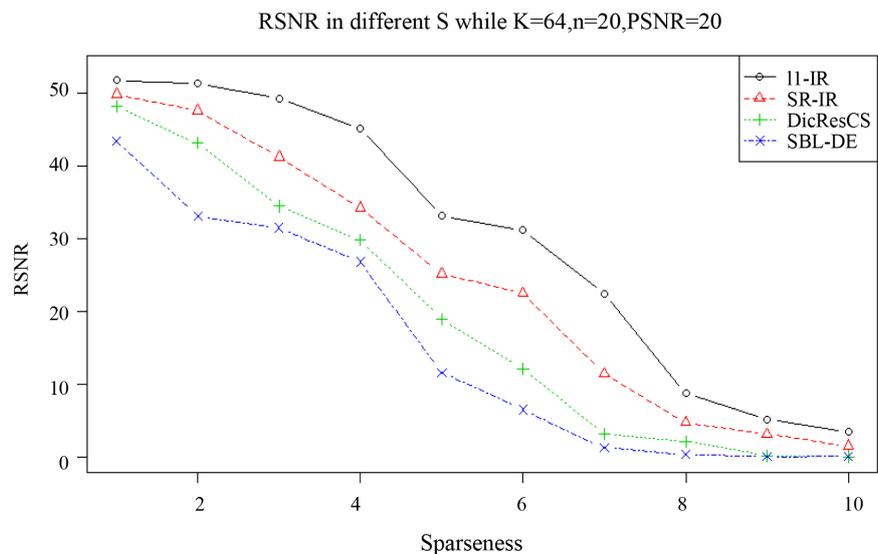


Figure 1. RSNR in different S while $K = 64$, $n = 20$, $PSNR = 20$.

of measurements. This means that our method performs better for the recovery of sparse signals when the number of samples is limited.

Our last experiment tested the recovery performance of respective algorithm under different noise level $PSNR$. According to the definition of $PSNR$, we set the variance of noise from 0.01 to 1, which makes the $PSNR$ changed from 20 to 0. At this experiment use signals of length $K = 64$ contains $S = 3$ complex sinusoids and set the number of measurements $N = 24$. We take 20 points evenly between 0 and 20 and record the performance of respective algorithm in each point. Each data point is averaged by repeated tests.

Relatively speaking, the $L1-IR$ algorithm and the $SR-IR$ algorithm were more stable than the other algorithms as the noise level increased. **Figure 3** indicates that the $PSNR$ of all algorithms deteriorated quickly when the noise was very strong. In the case of relatively strong noise, some trials lead to failure results and the failure rate increases with an increasing noise level. Here, a trial was considered successful if the $PSNR$ was higher than 15 dB. **Table 1** shows the success rate of the algorithms for different levels of $PSNR$ (at each level, we repeated 20 trials) in the last experiment. The success rate of DicResCS and SBL-DE is obviously less than that of $L1-IR$ and $SR-IR$ when σ^2 is large than 0.2.

The validity, superiority, and stability of the $L1-IR$ algorithm are illustrated by these experiments, indicating that the algorithm is worth applying in some practical cases.

5. Conclusion

In this paper, we treated the real dictionary parameters as unknown variables, and studied the line spectral estimation problem with unknown dictionary parameters. Based on the ideal of Bayesian lasso and the analysis of the first-order

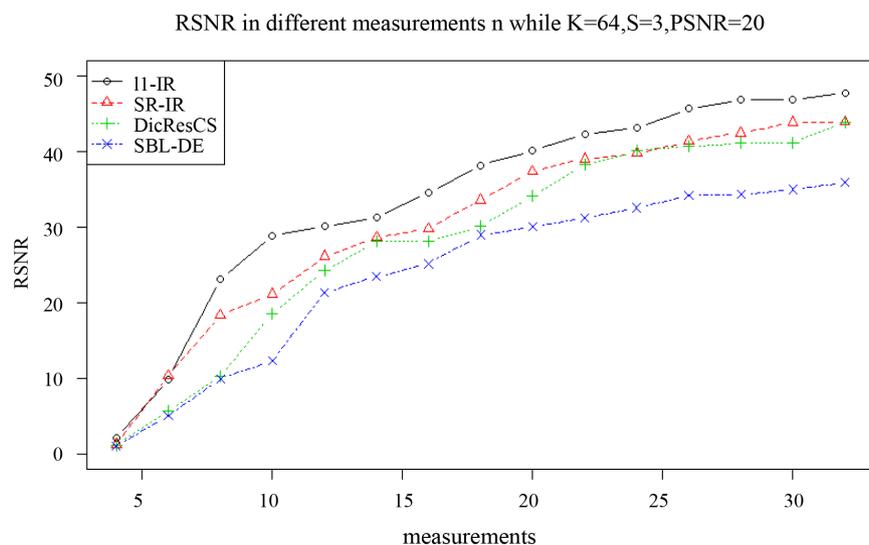


Figure 2. RSNR in different n while $K = 64$, $S = 3$, $PSNR = 20$.

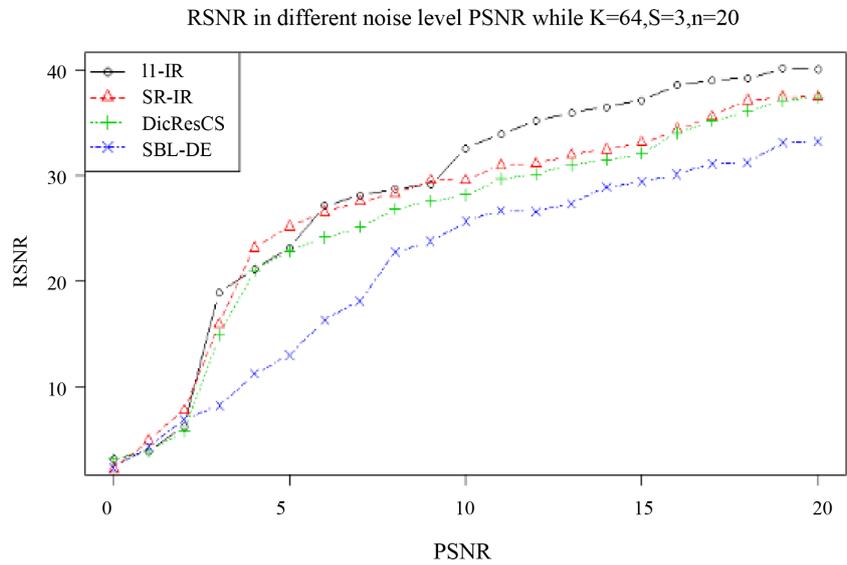


Figure 3. RSNR in different noise level while K = 64, S = 3, n = 20.

Table 1. Successful rate of different algorithms.

		$RSNR/\sigma^2/1$					
		0/1	3/0.5	6.98/0.2	10/0.1	13.1/0.05	20.0/0.1
7* algorithm	SBL-DE	45%	75%	80%	95%	100%	100%
	DisRefCS	40%	60%	75%	80%	90%	100%
	SR-IR	55%	80%	95%	100%	100%	100%
	L1-IR	60%	75%	95%	95%	100%	100%

condition of the optimal solution, we proposed an iterative reweighted l_1 penalty regression algorithm. We proved that in each step of the iterative process, the function value is continuously reduced until the approximate solution of the real sparse vector is obtained. The numerical results in Section 4 illustrated that the performance of our algorithm is better than other state-of-the-art algorithms in some cases. The disadvantage is that our method is more time-consuming, partly because in each sampling step it is necessary to ensure convergence, resulting in a sampling length that cannot be effectively reduced. Future studies will focus on this problem.

References

- [1] Chi, Y.J., Scharf, L.L., Pezeshki, A. and Calderbank, A.R. (2011) Sensitivity to Basis Mismatch in Compressed Sensing. *IEEE Transactions on Signal Process*, **59**, 2182-2195. <https://doi.org/10.1109/TSP.2011.2112650>
- [2] Tan, Z., Yang, P. and Nehorai, A. (2014) Joint Sparse Recovery Method for Compressed Sensing with Structured Dictionary Mismatches. *IEEE Transactions on Signal Process*, **62**, 4997-5008. <https://doi.org/10.1109/TSP.2014.2343940>
- [3] Fannjiang, A. and Liao, W. (2014) Coherence Pattern-Guided Compressive Sensing with Unresolved Grids. *SIAM Journal on Imaging Sciences*, **5**, 179-202.

- <https://doi.org/10.1137/110838509>
- [4] Duarte, M.F. and Baraniuk, R.G. (2013) Spectral Compressive Sensing. *Applied and Computational Harmonic Analysis*, **35**, 111-129.
<https://doi.org/10.1016/j.acha.2012.08.003>
- [5] Candes, E. and Fernandez-Granda, C. (2014) Towards a Mathematical Theory of Super-Resolution. *Communications on Pure and Applied Mathematics*, **67**, 906-956.
<https://doi.org/10.1002/cpa.21455>
- [6] Tang, G., Bhaskar, B.N., Shah, P. and Recht, B. (2013) Compressed Sensing Off the Grid. *IEEE Transactions on Information Theory*, **59**, 7465-7490.
<https://doi.org/10.1109/TIT.2013.2277451>
- [7] Hansen, T.L., Badiu, M.A., Fleury, B.H. and Rao, B.D. (2014) A Sparse Bayesian Learning Algorithm with Dictionary Parameter Estimation. *Sensor Array and Multichannel Signal Processing Workshop*, A Coruna, 22-25 June 2014, 385-388.
<https://doi.org/10.1109/SAM.2014.6882422>
- [8] Hu, L., Shi, Z., Zhou, J. and Fu, Q. (2012) Compressed Sensing of Complex Sinusoids: An Approach Based on Dictionary Refinement. *IEEE Transactions on Signal Process*, **60**, 3809-3822. <https://doi.org/10.1109/TSP.2012.2193392>
- [9] Fang, J., Li, J., Shen, Y., Li, H. and Li, S. (2014) Super-Resolution Compressed Sensing: An Iterative Reweighted Algorithm for Joint Parameter Learning and Sparse Signal Recovery. *IEEE Transactions on Signal Process*, **21**, 761-765.
<https://doi.org/10.1109/LSP.2014.2316004>
- [10] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the royal Statistical Society. Series B*, **58**, 276-288.
- [11] Fang, J., Wang, F., Shen, Y. and Li, H. (2016) Super-Resolution Compressed Sensing for Line Spectral Estimation: An Iterative Reweighted Approach. *IEEE Transactions on Signal Process*, **64**, 4649-4672. <https://doi.org/10.1109/TSP.2016.2572041>
- [12] Lai, M.J., Xu, Y. and Yin, W. (2013) Improved Iteratively Reweighted Least Squares for Unconstrained Smoothed l_q Minimization Estimation. *SIAM Journal on Numerical Analysis*, **51**, 927-957. <https://doi.org/10.1137/110840364>
- [13] Andrieu, C., Freitas, N. and Doucet, A. (2003) An Introduction to MCMC for Machine Learning. *Machine Learning*, **50**, 5C43.