AN ITERATIVE SURE-LET APPROACH TO SPARSE RECONSTRUCTION

Feng Xue[a], Thierry Blu[b], Runle Du[a] and Jiaqi Liu[a]

[a]National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics, Beijing, 100076, China

[b]Dept. of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

ABSTRACT

Sparsity-promoting regularization is often formulated as ℓ_ν-penalized minimization ($0 < \nu \leq 1$), which can be efficiently solved by iteratively reweighted least squares (IRLS). The reconstruction quality is generally sensitive to the value of regularization parameter. In this work, for accurate recovery, we develop two data-driven optimization schemes based on minimization of Stein's unbiased risk estimate (SURE). First, we propose a recursive method for computing SURE for a given IRLS iterate, which enables us to unbiasedly evaluate the reconstruction error, and select the optimal value of regularization parameter. Second, for fast optimization, we parametrize each IRLS iterate as a linear combination of few elementary functions (LET), and solve the linear weights by minimizing SURE. Numerical experiments show that iterating this process leads to higher reconstruction accuracy with remarkably faster computational speed than standard IRLS.

Index Terms— Sparse reconstruction, iteratively reweighted least squares (IRLS), Stein's unbiased risk estimate (SURE), linear expansion of thresholds (LET)

1. INTRODUCTION

Consider the standard estimation problem: find a good estimate of $\mathbf{x} \in \mathbb{R}^N$ from the following linear model [1]:

$$\mathbf{y} = \mu + \epsilon, \qquad \mu = \mathbf{A}\mathbf{x} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the observed data or the response vector, $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_N] \in \mathbb{R}^{M\times N}$ is a deterministic design matrix with the column vectors $\mathbf{A}_n$ representing predictors or features, $\epsilon \in \mathbb{R}^M$ is a vector of i.i.d. centered Gaussian random variable with known variance $\sigma^2 > 0$.

In many real applications, e.g. model/feature selection [2], signal recovery [3] and compressed sensing [4], it is preferable to promote the sparsity of the unknown vector $\mathbf{x}$, which is often formulated as ℓ_ν-penalized minimization [1,5]:

$$\mathbf{P}: \quad \min_{\mathbf{x}} \underbrace{\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \cdot \|\mathbf{x}\|_\nu^\nu}_{\mathcal{L}(\mathbf{x})}$$

where the ℓ_ν-norm regularization with $0 < \nu \leq 1$ enhances the sparsity of $\mathbf{x}$, $\lambda > 0$ is a regularization parameter. In this work, we use iteratively reweighted least squares (IRLS) to solve (P), due to its superior convergence speed [1, 6].

For accurate sparse recovery, it is essential to select a proper value of the regularization parameter λ, to keep a good balance between data fidelity and sparsity enhancement. We denote the IRLS solution to (P) by $\widehat{\mathbf{x}}_\lambda$ and $\widehat{\mu}_\lambda = \mathbf{A}\widehat{\mathbf{x}}_\lambda$, to emphasize the strong dependencies of the estimates upon λ. There have been a number of criteria for this selection of λ, e.g. generalized cross validation [7], L-curve method [8] and discrepancy principle [9]. However, they have been only applied to linear estimates, rather than the non-linear sparse reconstruction considered here.

In this paper, we quantify the reconstruction accuracy by the expected prediction error (EPE) [2, 10, 11]:

$$\mathrm{EPE}(\widehat{\mu}_\lambda) = \frac{1}{M}\mathbb{E}\left\{\left\|\widehat{\mu}_\lambda - \mu\right\|_2^2\right\} \tag{2}$$

and attempt to select a value of λ, such that the corresponding IRLS solution $\widehat{\mu}_\lambda$ achieves minimum prediction error. Note that here we do not consider the estimation error $\|\widehat{\mathbf{x}}_\lambda - \mathbf{x}\|_2^2$, since EPE (2) is easier to manipulate and keep numerical stability [11], despite the fact that it gives only a partial account of the actual reconstruction quality [15]. See [10, 11] for the similar treatments.

Notice that EPE (2) is inaccessible due to the unknown μ. In practice, Stein's unbiased risk estimate (SURE) has been proposed as a statistical substitute for EPE [10, 12]:

$$\mathrm{SURE}(\widehat{\mu}_\lambda) = \frac{1}{M}\left\|\mathbf{A}\widehat{\mathbf{x}}_\lambda - \mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{M}\mathrm{Tr}\left(\mathbf{A}\mathbf{J}_\mathbf{y}(\widehat{\mathbf{x}}_\lambda)\right) - \sigma^2 \tag{3}$$

since it depends on the observed data $\mathbf{y}$ only. Here, $\mathbf{J}_\mathbf{y}(\widehat{\mathbf{x}}_\lambda) \in \mathbb{R}^{N\times M}$ is a Jacobian matrix defined as:

$$\left[\mathbf{J}_\mathbf{y}(\widehat{\mathbf{x}}_\lambda)\right]_{n,m} = \frac{\partial(\widehat{\mathbf{x}}_\lambda)_n}{\partial y_m}$$

Note that, here, we make it explicit that the IRLS solution $\widehat{\mathbf{x}}_\lambda$ depends on the observed data $\mathbf{y}$.

This paper is to optimize the sparse reconstruction by IRLS, based on minimization of SURE (3). Our main contributions are twofold. First, we develop a recursive SURE for IRLS, which finally provides a reliable estimate of the prediction error of the non-linear sparse reconstruction. The

optimal λ can then be identified by exhaustive search for minimum SURE. Furthermore, for fast optimization, we adopt a strategy very similar to [6, 13, 18]: approximate the sparse estimation process by a linear combination of few elementary functions (LET bases) with different but fixed λ, and solve for the linear weights (LET coefficients) by minimizing recursive SURE. Experimentally, IRLS iteration of the recursive SURE-LET process is much more efficient in solving sparse reconstruction problems, compared to the standard IRLS.

2. RECURSIVE SURE FOR IRLS ALGORITHMS

2.1. Basic scheme of IRLS algorithms

To solve $\mathbf{P}$ with given fixed λ, the IRLS algorithm updates $\mathbf{x}$ by solving the following data-dependent linear system of equations [3,5,6]:

$$\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}+\lambda\mathbf{W}^{(i-1)}\right)\mathbf{x}^{(i)}=\mathbf{A}^{\mathrm{T}}\mathbf{y} \tag{4}$$

at i-th iteration, where $\mathbf{W}^{(i-1)}\in\mathbb{R}^{N\times N}$ is a diagonal matrix and $\mathbf{W}^{(i-1)}_{n,n}=\nu\left(|x_n^{(i)}|^2+\beta\right)^{\frac{\nu}{2}-1}$ for $n=1,2,...,N$. β is a small constant [5]. (4) leads to the following update of $\mathbf{x}^{(i)}$:

$$\mathbf{x}^{(i)}=\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}+\lambda\mathbf{W}^{(i-1)}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y} \tag{5}$$

By (3), SURE of the i-th iterate $\mu^{(i)}=\mathbf{A}\mathbf{x}^{(i)}$ is:

$$\mathrm{SURE}(\mu^{(i)})=\frac{1}{M}\left\|\mathbf{A}\mathbf{x}^{(i)}-\mathbf{y}\right\|_2^2+\frac{2\sigma^2}{M}\mathrm{Tr}\left(\mathbf{A}\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})\right)-\sigma^2 \tag{6}$$

The SURE computation requires to evaluate $\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})$. Note that $\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})\neq\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}+\lambda\mathbf{W}^{(i-1)}\right)^{-1}\mathbf{A}^{\mathrm{T}}$, since (5) indicates that $\mathbf{x}^{(i)}$ is NOT a linear transformation of $\mathbf{y}$: $\mathbf{W}^{(i-1)}$ is constructed by $\mathbf{x}^{(i-1)}$, which is dependent on data $\mathbf{y}$.

2.2. Matrix splitting to solve IRLS iteration

It is not easy to directly compute $\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})$ by the closed-form solution (5). Instead, we apply the matrix-splitting (MS) scheme [14] to solve (4), which, as we will see later, enables us to evaluate $\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})$ in a recursive manner. This method splits the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}+\lambda\mathbf{W}^{(i-1)}$ as:

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}+\lambda\mathbf{W}^{(i-1)}=\underbrace{(\mathbf{D}+\lambda\mathbf{W}^{(i-1)})}_{\mathbf{P}_{(i-1)}}-\underbrace{(\mathbf{D}-\mathbf{A}^{\mathrm{T}}\mathbf{A})}_{\mathbf{Q}} \tag{7}$$

for any matrix $\mathbf{D}\in\mathbb{R}^{N\times N}$, and the solution $\mathbf{x}^{(i,j)}$ is generated as follows (fixed i, indexed by j):

$$\mathbf{x}^{(i,j)}=\mathbf{P}_{(i-1)}^{-1}\left(\mathbf{Q}\mathbf{x}^{(i,j-1)}+\mathbf{A}^{\mathrm{T}}\mathbf{y}\right) \tag{8}$$

provided that matrix $\mathbf{P}_{(i-1)}$ is invertible. The MS iteration (8) is convergent to the unique solution of (4) for any initial $\mathbf{x}^{(i,0)}$, if and only if the spectral radius $\rho(\mathbf{P}_{(i-1)}^{-1}\mathbf{Q})<1$ [14]. To guarantee the convergence and easy inversion of $\mathbf{P}^{(i-1)}$, we choose $\mathbf{D}=\alpha\mathbf{I}$ with $\alpha>\rho(\mathbf{A}^{\mathrm{T}}\mathbf{A})$ to satisfy that: (1) $\rho(\mathbf{P}_{(i-1)}^{-1}\mathbf{Q})<1$; (2) $\mathbf{P}_{(i-1)}$ is diagonal: its inverse is easy to compute.

2.3. Recursion of Jacobian matrix

In this part, we will see that the computation of Jacobian matrix becomes tractable by the matrix splitting.

Rewrite (8) as $\mathbf{v}=\mathbf{H}\mathbf{b}$ for brevity, where $\mathbf{v}=\mathbf{x}^{(i,j)}$, $\mathbf{H}=\mathbf{P}_{(i-1)}^{-1}$ and $\mathbf{b}=\mathbf{A}^{\mathrm{T}}\mathbf{y}+\mathbf{Q}\mathbf{x}^{(i,j-1)}$. The Jacobian matrix of (8) is derived as:

$$\begin{aligned}[\mathbf{J}_{\mathbf{y}}(\mathbf{v})]_{m,k} &= \sum_{n=1}^{N}H_{m,n}\frac{\partial b_n}{\partial y_k}+\sum_{n=1}^{N}b_n\frac{\partial H_{m,n}}{\partial y_k}\\ &= \left[\mathbf{H}\mathbf{J}_{\mathbf{y}}(\mathbf{b})\right]_{m,k}+b_m\frac{\partial H_{m,m}}{\partial y_k}\end{aligned} \tag{9}$$

for the (m,k)-th entry of Jacobian matrix, where $\mathbf{J}_{\mathbf{y}}(\mathbf{b})=\mathbf{A}^{\mathrm{T}}+\mathbf{Q}\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i,j-1)})$ by the property of Jacobian matrix.

Considering the second term of (9):

$$\begin{aligned}\frac{\partial H_{m,m}}{\partial y_k} &= \frac{\partial\left[\lambda\nu\left(|x_m^{(i-1)}|^2+\beta\right)^{\frac{\nu}{2}-1}+\alpha\right]^{-1}}{\partial y_k}\\ &= \underbrace{\frac{\partial\left[\lambda\nu\left(|x_m^{(i-1)}|^2+\beta\right)^{\frac{\nu}{2}-1}+\alpha\right]^{-1}}{\partial\left|x_m^{(i-1)}\right|}\cdot\frac{\partial|x_m^{(i-1)}|}{\partial x_m^{(i-1)}}}_{e_m}\cdot\frac{\partial x_m^{(i-1)}}{\partial y_k}\end{aligned}$$

where e_m is given as:

$$e_m=\left[\lambda\nu\left(|x_m^{(i-1)}|^2+\beta\right)^{\frac{\nu}{2}-1}+\alpha\right]^{-2}\cdot\lambda\nu(2-\nu)\cdot\left(|x_m^{(i-1)}|^2+\beta\right)^{\frac{\nu}{2}-2}\cdot x_m^{(i-1)}$$

then, the second term of (9) becomes:

$$\begin{aligned}b_m\frac{\partial H_{m,m}}{\partial y_k} &= \underbrace{b_m e_m}_{c_m}\frac{\partial x_m^{(i-1)}}{\partial y_k}=[\underbrace{\mathrm{diag}(\mathbf{c})}_{\mathbf{C}}]_{m,m}\left[\mathbf{J}_{\mathbf{y}}(\mathbf{x}_m^{(i-1)})\right]_{m,k}\\ &= \left[\mathbf{C}\mathbf{J}_{\mathbf{y}}(\mathbf{x}_m^{(i-1)})\right]_{m,k}\end{aligned}$$

where diagonal matrix $\mathbf{C}=\mathrm{diag}(\mathbf{c})\in\mathbb{R}^{N\times N}$ with (m,m)-th entry $C_{m,m}=b_m e_m$ for $m=1,2,...,N$. Finally, (9) becomes:

$$\mathbf{J}_{\mathbf{y}}\left(\mathbf{x}^{(i,j)}\right)=\mathbf{P}_{(i-1)}^{-1}\mathbf{A}^{\mathrm{T}}+\mathbf{P}_{(i-1)}^{-1}\mathbf{Q}\mathbf{J}_{\mathbf{y}}\left(\mathbf{x}^{(i,j-1)}\right)+\mathbf{C}\mathbf{J}_{\mathbf{y}}\left(\mathbf{x}^{(i-1)}\right) \tag{10}$$

The key equation (10) expresses the recursion of the Jacobian matrix. For (i,j)-th iterate, SURE is:

$$\mathrm{SURE}(\mu^{(i,j)})=\frac{1}{M}\left\|\mathbf{A}\mathbf{x}^{(i,j)}-\mathbf{y}\right\|_2^2+\frac{2\sigma^2}{M}\mathrm{Tr}\left(\mathbf{A}\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i,j)})\right)-\sigma^2 \tag{11}$$

where $\mathbf{x}^{(i)}=\mathbf{x}^{(i,\infty)}=\mathbf{x}^{(i+1,0)}$ and $\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})=\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i,\infty)})=\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i+1,0)})$, assuming that matrix splitting is converged at $j=\infty$.

2.4. Summary of IRLS with matrix-splitting strategy

Finally, we summarize the proposed IRLS-MS-SURE algorithm as **Algorithm 1**, which enables us to solve $\mathbf{P}$ with a prescribed value of λ, and simultaneously evaluate the SURE during the IRLS iterations.

To find the optimal value of λ, an intuitive idea is to repeatedly implement **Algorithm 1** for various tentative values of λ, then, the minimum SURE indicates the optimal λ (see Fig.2-(1) for example). This *grid search* has been frequently used in [11, 15–17], despite of its high computational cost.

Algorithm 1: SURE evaluation for IRLS-MS

Input: $\mathbf{y}$, $\mathbf{A}$, λ, ν, α, β, initial $\mathbf{x}^{(0)}$
Output: reconstructed $\widehat{\mathbf{x}}_\lambda$, $\widehat{\mu}_\lambda$, and SURE($\widehat{\mu}_\lambda$)
for $i = 1, 2, ...$ *(IRLS iteration)* **do**
  for $j = 0, 1, 2, ...$ *(MS iteration)* **do**
    1 compute $\mathbf{x}^{(i,j)}$ by (8);
    2 update $\mathbf{J}_\mathbf{y}(\mathbf{x}^{(i,j)})$ by (10);
    3 compute SURE of $\mu^{(i,j)}$ by (11);
  end
end

3. ITERATIVE SURE-LET RECONSTRUCTION

Now, for fast optimization, based on **Algorithm 1**, we adopt our previously used strategy in [6, 13, 18], which decomposes each IRLS iterate (5) into a linear combination of elementary thresholding functions—Linear Expansion of Thresholds (LET):

$$\mathbf{x}^{(i)} = \sum_{k=1}^{K} a_k \underbrace{\left(\mathbf{A}^\mathrm{T}\mathbf{A} + \lambda_k \mathbf{W}^{(i-1)}\right)^{-1}\mathbf{A}^\mathrm{T}\mathbf{y}}_{\mathbf{x}_k^{(i)}}; \quad \mu^{(i)} = \sum_{k=1}^{K} a_k \underbrace{\mathbf{A}\mathbf{x}_k^{(i)}}_{\mu_k^{(i)}} \tag{12}$$

Thus, the update $\mathbf{x}^{(i)}$ is a linear combination (by LET coefficients a_k) of a number of LET bases $\mathbf{x}_k^{(i)}$. Each LET basis $\mathbf{x}_k^{(i)}$ is constructed by different but fixed λ_k. Now, the optimization problem becomes to determining the optimal LET coefficients a_k instead of unique non-linear parameter λ. Substituting (12) into (6), we have:

$$\mathrm{SURE}(\mu^{(i)}) = \frac{1}{M}\left\|\sum_{k=1}^{K} a_k \mathbf{A}\mathbf{x}_k^{(i)} - \mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{M}\sum_{k=1}^{K} a_k \mathrm{Tr}\left(\mathbf{A}\mathbf{J}_\mathbf{y}(\mathbf{x}_k^{(i)})\right) - \sigma^2 \tag{13}$$

where $\mathbf{x}_k^{(i)}$ and $\mathbf{J}_\mathbf{y}(\mathbf{x}_k^{(i)})$ are again obtained by MS iteration (8) and (10), respectively, with λ_k for $k = 1, 2, ..., K$.

The SURE (13) is a quadratic functional of a_k: minimizing SURE w.r.t. a_k boils down to solving the following linear system of equations:

$$\sum_{k'=1}^{K} \underbrace{\frac{1}{M}\mu_{k'}^{(i)\mathrm{T}}\mu_k^{(i)}}_{\mathbf{M}_{k,k'}} a_{k'} = \underbrace{\frac{1}{M}\left(\mathbf{y}^\mathrm{T}\mu_k^{(i)} - \sigma^2 \mathrm{Tr}\left(\mathbf{J}_\mathbf{y}(\mu_k^{(i)})\right)\right)}_{c_k} \tag{14}$$

for $k = 1, 2, ..., K$. These equations can be summarized in matrix form as $\mathbf{M}\mathbf{a} = \mathbf{c}$, where $\mathbf{M} = [\mathbf{M}_{k,k'}]_{k,k'=1,2,...,K}$ and $\mathbf{c} = [c_1, c_2, ..., c_K]^\mathrm{T}$.

The underlying principle of the SURE-LET approach is that different values of λ_k capture various features of the data $\mathbf{x}$: smaller λ reveals more details of signal, whereas larger λ yields smoother data but with more noise suppression. The SURE-LET method consists in finding the best combination of the candidates $\mu_k^{(i)}$ in terms of SURE, which is automatically done by solving (14). The optimal linear coefficients a_k control the best balance between data fidelity and regularization enforcement. In practice, the number of LET bases K is very small (typically, less than 10), which dramatically reduces the problem dimension. Therefore, we expect the SURE-LET update (12) to achieve smaller SURE with faster computational speed, though it is not an exact solution to (P) with any value of λ. The proposed method is summarized in **Algorithm 2**.

Algorithm 2: SURE-LET within IRLS-MS algorithms

Input: $\mathbf{y}$, $\mathbf{A}$, ν, α, β, initial $\mathbf{x}^{(0)}$, λ_k for $k = 1, 2, ..., K$
Output: reconstructed $\widehat{\mathbf{x}}_\lambda$, $\widehat{\mu}_\lambda$, and SURE($\widehat{\mu}_\lambda$)
for $i = 1, 2, ...$ *(IRLS iteration)* **do**
  for $j = 0, 1, ...$ *(MS iteration)* **do**
    (1) update $\mathbf{x}_k^{(i,j)}$ and $\mu_k^{(i,j)}$ by (8) with λ_k for $k = 1, 2, ..., K$;
    (2) update of $\mathbf{J}_\mathbf{y}(\mathbf{x}_k^{(i,j)})$ by (10) with λ_k for $k = 1, 2, ..., K$;
  end
  1 update $\mathbf{x}_k^{(i,\infty)} = \mathbf{x}_k^{(i)} = \mathbf{x}_k^{(i+1,0)}$ and $\mathbf{J}_\mathbf{y}(\mathbf{x}_k^{(i,\infty)}) = \mathbf{J}_\mathbf{y}(\mathbf{x}_k^{(i)}) = \mathbf{J}_\mathbf{y}(\mathbf{x}_k^{(i+1,0)})$;
  2 build and solve (14) for a_k;
  3 update $\mathbf{x}^{(i)}$ and $\mu^{(i)}$ by (12);
  4 compute SURE of $\mu^{(i)}$ by (13).
end

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we are going to solve (P) with $\nu = 0.5$ by IRLS, and present the results of the proposed recursive SURE (i.e. **Algorithm 1**) and iterative SURE-LET algorithm (i.e. **Algorithm 2**).

4.1. Experimental setting

To demonstrate the wide applicability of our proposed approaches, we consider a random numerical example: we randomly generate the matrix $\mathbf{A} \in \mathbb{R}^{300\times500}$, and set $\mathbf{x} \in \mathbb{R}^{500}$ as a sparse vector with very few non-zeros (in this example, 10 non-zeros). Then, we add the noise ϵ with noise variance σ^2 to obtain the observed data $\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon$, such that the input SNR is 10dB[1].

4.2. Recursive SURE for IRLS-MS with fixed λ

First, we apply **Algorithm 1** to solve **P** with fixed $\lambda = 1$. Fig.1 shows the IRLS-MS convergence. The objective value of $\mathcal{L}(\mathbf{x}^{(i)})$ keeps decreasing to converge, as shown in Fig.1-(1). Fig.1-(2) shows the evolutions of SURE and true EPE

[1]Input signal-to-noise ratio (SNR) is defined as: $10\log_{10}\left(\frac{\|\mu\|_2^2}{\|\mathbf{y}-\mu\|_2^2}\right) = 10\log_{10}\left(\frac{\|\mu\|_2^2}{M\sigma^2}\right)$ in dB.

during the iterations. We can see that the SURE is always a reliable substitute for EPE.

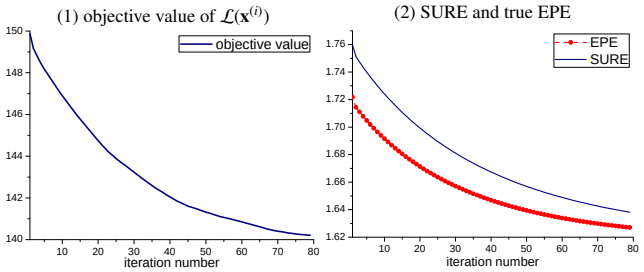

Fig. 1. The convergence of IRLS-MS with fixed $\lambda = 1.0$.

4.3. Optimization of sparse reconstruction

In this work, the optimization of sparse reconstruction can be performed by *grid search* or iterative SURE-LET. For grid search, we repeatedly implement **Algorithm 1** for 50 tentative values of λ (logarithmically spaced), and obtain the corresponding SURE. Fig.2-(1) shows the relation between SURE and λ, where the minimum point indicates optimal value of λ. For iterative SURE-LET, we set $K = 3$ regularization parameters (i.e. three LET bases): $\lambda_1 = 1$, $\lambda_2 = 10$ and $\lambda_3 = 100$, and solve three LET coefficients by minimization of recursive SURE for each update. Fig.2-(2) shows that the iterative SURE-LET produces the optimal reconstruction in ONE implementation of IRLS-MS, which achieves smaller SURE with much faster convergence speed, compared to basic IRLS-MS (see Fig.1).

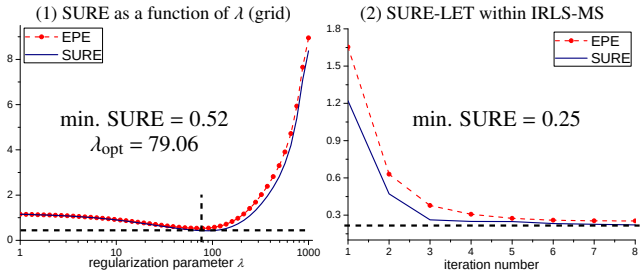

Fig. 2. Optimizations by grid search and iterative SURE-LET.

Fig. 3 shows two fractions of reconstructed signal $\widehat{\mathbf{x}}$ for the comparison between grid search and iterative SURE-LET method.

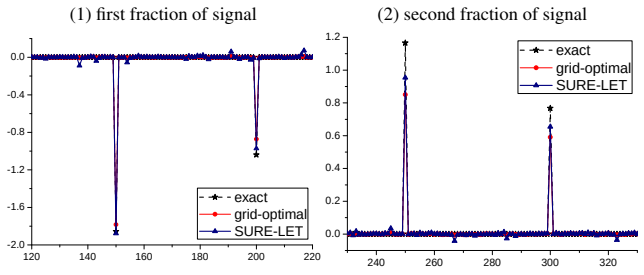

Fig. 3. Two small fractions of the signal reconstruction.

Table 1 reports the errors and computational time of grid search and iterative SURE-LET method. The errors of $\mathbf{x}$ and μ are defined as: $\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2/N$ and $\|\widehat{\mu} - \mu\|_2^2/M$, respectively.

Table 1. Comparisons between grid search and iterative SURE-LET

methods	grid search	iterative SURE-LET
error of $\mathbf{x}$	1.21×10^{-3}	$\mathbf{5.34 \times 10^{-4}}$
error of μ	0.52	**0.25**
time (in sec.)	852.32	**23.50**

From Fig.3 and Table 1, we can see that the iterative SURE-LET produces more accurate reconstruction. The remarkably improved computational efficiency is due to the following several facts: (1) grid search requires 50 times of implementations of IRLS-MS with various λ; (2) SURE-LET greatly accelerates the convergence speed of IRLS-MS, and complete the optimization in ONE execution; (3) SURE-LET for each IRLS update finally boils down to solving a 3-order (i.e., $K = 3$) linear system of equations (14), which costs negligible time.

5. CONCLUSIONS

SURE has been proven a powerful tool to select regularization parameter [11, 15]. In this paper, to solve ℓ_v-minimization problem, we proposed a recursive method for computing SURE for a given IRLS iterate, which always keeps statistical unbiasedness w.r.t. the prediction error during the iterations. It enables us to accurately estimate the prediction loss, without referring to the true unknown data. Furthermore, for fast optimization of sparse reconstruction, we represent each IRLS iterate as a linear combination of elementary functions (LET basis) with optimal weights (LET coefficients) obtained by minimizing recursive SURE.

Theoretical derivations in this work related to the SURE evaluation and SURE-LET framework can be extended, in principle, to other types of regularizers and regularized iterative reconstruction algorithms. We would also like to emphasize that not limited to the simple example shown here, the developed recursive SURE and iterative SURE-LET can be applied to many practical applications, e.g. MRI reconstruction [11] and image deconvolution [6, 13].

6. ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 61401013 and the Hong Kong Research Grant Council (CUHK410012). The authors are obliged to Hanjie PAN for many useful discussions on this article.

7. REFERENCES

[1] D. Ba, B. Babadi, P.L. Purdon, and E.N. Brown, "Convergence and stability of iteratively re-weighted least squares algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 183–195, 2014.

[2] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, vol. 2, Springer, 2009.

[3] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C Sinan Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

[4] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[5] O. Batu and M. Cetin, "Parameter selection in sparsity-driven SAR imaging," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 3040–3050, 2011.

[6] Hanjie Pan and Thierry Blu, "An iterative linear expansion of thresholds for ℓ_1-based image restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3715–3728, 2013.

[7] G.H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[8] Per Christian Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review*, vol. 34, no. 4, pp. 561–580, 1992.

[9] V.A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, New York, 1984.

[10] Hui Zou, Trevor Hastie, Robert Tibshirani, et al., "On the degrees of freedom of the lasso," *The Annals of Statistics*, vol. 35, no. 5, pp. 2173–2192, 2007.

[11] S. Ramani, Zhihao Liu, J. Rosen, J. Nielsen, and J.A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3659–3672, 2012.

[12] F. Xue and T. Blu, "A novel SURE-based criterion for parametric PSF estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 595–607, 2015.

[13] F. Xue, F. Luisier, and T. Blu, "Multi-Wiener SURE-LET deconvolution," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1954–1968, 2013.

[14] Richard S. Varga, *Matrix Iterative Analysis*, vol. 27, Springer Series in Computational Mathematics, 2000.

[15] C. Vonesch, S. Ramani, and M. Unser, "Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint," in *15th IEEE International Conference on Image Processing*, 2008, pp. 665–668.

[16] R. Giryes, M. Elad, and Y.C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 407–422, 2010.

[17] J. Shi, W. Yin, and S. Osher, "A new regularization path for logistic regression via linearized Bregman," *Rice CAAM Techical Report TR12-24*, 2012.

[18] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778–2786, 2007.