OPTIMIZATION OF REGULARIZATION PARAMETER FOR SPARSE RECONSTRUCTION BASED ON PREDICTIVE RISK ESTIMATE

Feng Xue[a], Hanjie Pan[b], Xin Liu[a], Hongyan Liu[a] and Jiaqi Liu[a]

[a]National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics, Beijing, 100076, China
[b]Audiovisual Communications Laboratory, School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ABSTRACT

Recently, ℓ_1-based sparse reconstruction has shown its superior performance in terms of accuracy and interpretation. The reconstruction quality is generally sensitive to the value of regularization parameter. In this work, we develop two data-driven optimization schemes based on minimization of predicted Stein's unbiased risk estimate (p-SURE)—statistically equivalent to predicted mean squared error (p-MSE). First, we propose a recursive evaluation of p-SURE for any given regularization parameter, where the optimal value is identified by the minimum p-SURE. Second, for fast optimization, we perform alternating optimization between regularization parameter and solution during the ℓ_1-based iterative algorithms. We exemplify the proposed methods with a typical *alternating direction method of multipliers* (ADMM) for both *analysis* and *synthesis* formulations. Numerical experiments show that the proposed methods lead to highly accurate estimate of regularization parameter and nearly optimal reconstruction.

Index Terms— Sparse reconstruction, predicted Stein's unbiased risk estimate (p-SURE), alternating direction method of multipliers (ADMM), analysis vs. synthesis

1. INTRODUCTION

Consider the standard estimation problem: find a good estimate of $\mathbf{x}_0 \in \mathbb{R}^N$ from the following linear model [1]:

$$\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \epsilon \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the observed data, $\mathbf{H} \in \mathbb{R}^{M\times N}$ is a deterministic observation matrix, $\epsilon \in \mathbb{R}^M$ is a vector of i.i.d. Gaussian random variable with zero mean and known variance $\sigma^2 > 0$.

In many real applications, e.g. model/feature selection [2], signal recovery [1] and compressed sensing [3], it is preferable to promote the sparsity of the unknown vector $\mathbf{x}_0$, which is often formulated as the following ℓ_1-penalized unconstrained minimizations [4]:

$$\textit{analysis formulation:} \quad \min_{\mathbf{x}} \underbrace{\frac{1}{2}\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|_2^2 + \lambda\cdot\left\|\mathbf{D}\mathbf{x}\right\|_1}_{\mathcal{L}_1(\mathbf{x})} \tag{2}$$

or

$$\textit{synthesis formulation:} \quad \min_{\mathbf{c}} \underbrace{\frac{1}{2}\left\|\mathbf{y}-\mathbf{H}\mathbf{R}\mathbf{c}\right\|_2^2 + \lambda\cdot\left\|\mathbf{c}\right\|_1}_{\mathcal{L}_2(\mathbf{c})} \tag{3}$$

where it is believed that the original data $\mathbf{x}_0$ has a sparse representation in transform domain $\mathbf{D} \in \mathbb{R}^{L\times N}$, i.e. the transform coefficients $\mathbf{c}_0 = \mathbf{D}\mathbf{x}_0$ are sparse in $\mathbf{D}$-domain. $\mathbf{R} \in \mathbb{R}^{N\times L}$ is a reconstruction operator that satisfies $\mathbf{R}\mathbf{D} = \mathbf{I}_N$ for perfect reconstruction. Both formulations are fundamentally distinct; their connections have been discussed in [4]. In this paper, we choose *alternating direction method of multipliers* (ADMM) to solve both formulations [5].

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 enforcement [6, 7]. We denote the solutions of (2) and (3) by $\widehat{\mathbf{x}}_\lambda$ and $\widehat{\mathbf{c}}_\lambda$, 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 [8], L-curve method [9] and discrepancy principle [10]. 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 predicted mean squared error (p-MSE) [2, 11, 12]:

$$\text{p-MSE} = \begin{cases} \frac{1}{M}\mathbb{E}\left\{\left\|\mathbf{H}\widehat{\mathbf{x}}_\lambda - \mathbf{H}\mathbf{x}_0\right\|_2^2\right\} & \text{(analysis)} \\ \frac{1}{M}\mathbb{E}\left\{\left\|\mathbf{H}\mathbf{R}\widehat{\mathbf{c}}_\lambda - \mathbf{H}\mathbf{R}\mathbf{c}_0\right\|_2^2\right\} & \text{(synthesis)} \end{cases} \tag{4}$$

and attempt to select a value of λ, such that the corresponding solutions $\widehat{\mathbf{x}}_\lambda$ and $\widehat{\mathbf{c}}_\lambda$ achieve minimum p-MSE. Note that here we do not consider the standard MSE $\|\widehat{\mathbf{x}}_\lambda - \mathbf{x}_0\|_2^2$, since p-MSE (4) is easier to manipulate and keep numerical stability. See [12] for the similar treatments.

Notice that p-MSE (4) is inaccessible due to the unknown $\mathbf{x}_0$ and $\mathbf{c}_0$. In practice, predicted Stein's unbiased risk estimate (p-SURE) has been proposed as a statistical substitute for p-

MSE [11, 12]:

$$\text{p-SURE} = \begin{cases} \frac{1}{M}\left\|\mathbf{H}\widehat{\mathbf{x}}_\lambda - \mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{M}\mathrm{Tr}\Big(\mathbf{H}\mathbf{J}_\mathbf{y}(\widehat{\mathbf{x}}_\lambda)\Big) - \sigma^2 & \text{(analysis)} \\ \frac{1}{M}\left\|\mathbf{H}\mathbf{R}\widehat{\mathbf{c}}_\lambda - \mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{M}\mathrm{Tr}\Big(\mathbf{H}\mathbf{R}\mathbf{J}_\mathbf{y}(\widehat{\mathbf{c}}_\lambda)\Big) - \sigma^2 & \text{(synthesis)} \end{cases} \tag{5}$$

since it depends on the observed data $\mathbf{y}$ only. Here, $\mathbf{J}_\mathbf{y}(\cdot)$ is a Jacobian matrix defined as:

$$\Big[\mathbf{J}_\mathbf{y}(\cdot)\Big]_{n,m} = \frac{\partial(\cdot)_n}{\partial y_m}$$

This paper is to optimize the regularization parameter λ for both *analysis* and *synthesis* formulations, based on minimization of p-SURE (5). Our main contributions are twofold. First, we develop a recursive evaluation of p-SURE for the ADMM iterate, which finally provides a reliable estimate of the p-MSE for the non-linear sparse reconstruction. The optimal λ can then be identified by exhaustive search for minimum p-SURE. Furthermore, for fast optimization, we perform alternating update between regularization parameter and solution during the ADMM iteration, which yields very accurate estimate of parameter.

2. RECURSIVE EVALUATION OF P-SURE FOR ADMM

2.1. Basic scheme of ADMM for ℓ_1-minimization

From the viewpoint of mathematical notation, one can treat *synthesis* formulation as a special case of *analysis* version corresponding to $\mathbf{D} = \mathbf{I}$. For a general purpose, we consider the following minimization problem:

$$\min_{\mathbf{x}} \underbrace{\frac{1}{2}\left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2^2 + \lambda \cdot \left\|\mathbf{D}\mathbf{x}\right\|_1}_{\mathcal{L}(\mathbf{x})} \tag{6}$$

with $\mathbf{A} \in \mathbb{R}^{M\times N}$ and $\mathbf{D} \in \mathbb{R}^{L\times N}$. By augmented Lagrangian, ADMM solves the problem by the following steps [5]:

$$\begin{cases} \mathbf{x}^{(i+1)} &= \Big(\mathbf{D}^\mathrm{T}\mathbf{D} + \tau\mathbf{A}^\mathrm{T}\mathbf{A}\Big)^{-1}\Big(\tau\mathbf{A}^\mathrm{T}\mathbf{y} + \mathbf{D}^\mathrm{T}(\mathbf{z}^{(i)} - \tau\mathbf{u}^{(i)})\Big) \\ \mathbf{z}^{(i+1)} &= \mathcal{T}_{\tau\lambda}\Big(\mathbf{D}\mathbf{x}^{(i+1)} + \tau\mathbf{u}^{(i)}\Big) \\ \mathbf{u}^{(i+1)} &= \mathbf{u}^{(i)} + \frac{1}{\tau}\Big(\mathbf{D}\mathbf{x}^{(i+1)} - \mathbf{z}^{(i+1)}\Big) \end{cases} \tag{7}$$

By (5), p-SURE of the i-th iterate is:

$$\text{p-SURE} = \frac{1}{M}\left\|\mathbf{A}\mathbf{x}^{(i)} - \mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{M}\mathrm{Tr}\Big(\mathbf{A}\mathbf{J}_\mathbf{y}(\mathbf{x}^{(i)})\Big) - \sigma^2 \tag{8}$$

The computation of p-SURE requires to evaluate $\mathbf{J}_\mathbf{y}(\mathbf{x}^{(i)})$, which can be obtained in a recursive manner.

2.2. Recursion of Jacobian matrix

Regarding $\mathbf{J}_\mathbf{y}(\mathbf{z}^{(i+1)})$, let $\mathbf{v}^{(i)} = \mathbf{D}\mathbf{x}^{(i+1)} + \tau\mathbf{u}^{(i)}$, we have:

$$\begin{aligned} \Big[\mathbf{J}_\mathbf{y}(\mathbf{z}^{(i+1)})\Big]_{m,n} &= \frac{\partial[\mathcal{T}_{\tau\lambda}(\mathbf{v}^{(i)})]_m}{\partial y_n} = \sum_{s=1}^{N} \underbrace{\frac{\partial[\mathcal{T}_{\tau\lambda}(\mathbf{v}^{(i)})]_m}{\partial v_s^{(i)}}}_{[\mathbf{P}_{\tau\lambda}^{(i)}]_{m,s}} \cdot \frac{\partial v_s^{(i)}}{\partial y_n} \\ &= \Big[\mathbf{P}_{\tau\lambda}^{(i)}\mathbf{J}_\mathbf{y}(\mathbf{v}^{(i)})\Big]_{m,n} \end{aligned}$$

where $\mathbf{J}_\mathbf{y}(\mathbf{v}^{(i)}) = \mathbf{D}\mathbf{J}_\mathbf{y}(\mathbf{x}^{(i+1)}) + \tau\mathbf{J}_\mathbf{y}(\mathbf{u}^{(i)})$, and $\mathbf{P}_{\tau\lambda}^{(i)} \in \mathbb{R}^{L\times L}$ is diagonal matrix with the diagonal element is:

$$\Big[\mathbf{P}_{\tau\lambda}^{(i)}\Big]_{l,l} = \begin{cases} 1, & \text{if } \Big|(\mathbf{D}\mathbf{x})_l^{(i+1)} + \tau u_l^{(i)}\Big| > \tau\lambda \\ 0, & \text{if } \Big|(\mathbf{D}\mathbf{x})_l^{(i+1)} + \tau u_l^{(i)}\Big| \le \tau\lambda \end{cases}$$

Thus, the recursions of Jacobian matrices are:

$$\begin{cases} \mathbf{J}_\mathbf{y}(\mathbf{x}^{(i+1)}) &= \Big(\mathbf{D}^\mathrm{T}\mathbf{D} + \tau\mathbf{A}^\mathrm{T}\mathbf{A}\Big)^{-1}\Big(\tau\mathbf{A}^\mathrm{T} + \mathbf{D}^\mathrm{T}\mathbf{J}_\mathbf{y}(\mathbf{z}^{(i)}) - \tau\mathbf{D}^\mathrm{T}\mathbf{J}_\mathbf{y}(\mathbf{u}^{(i)})\Big) \\ \mathbf{J}_\mathbf{y}(\mathbf{z}^{(i+1)}) &= \mathbf{P}_{\tau\lambda}^{(i)}\Big(\mathbf{D}\mathbf{J}_\mathbf{y}(\mathbf{x}^{(i+1)}) + \tau\mathbf{J}_\mathbf{y}(\mathbf{u}^{(i)})\Big) \\ \mathbf{J}_\mathbf{y}(\mathbf{u}^{(i+1)}) &= \mathbf{J}_\mathbf{y}(\mathbf{u}^{(i)}) + \frac{1}{\tau}\Big(\mathbf{D}\mathbf{J}_\mathbf{y}(\mathbf{x}^{(i+1)}) - \mathbf{J}_\mathbf{y}(\mathbf{z}^{(i+1)})\Big) \end{cases} \tag{9}$$

where $\mathbf{J}_\mathbf{y}(\mathbf{x}^{(i)})$ and $\mathbf{J}_\mathbf{y}(\mathbf{u}^{(i)})$ of (9) are obtained by the basic property of Jacobian matrix.

2.3. Summary of ADMM with p-SURE evaluation

Now, we summarize the proposed algorithm as **Algorithm 1**, which enables us to solve (2) and (3) with a prescribed value of λ, and simultaneously evaluate the p-SURE during the ADMM iterations.

Algorithm 1: p-SURE evaluation for ADMM

Input: $\mathbf{y}$, $\mathbf{A}$, $\mathbf{D}$, λ, τ, initial $\mathbf{x}^{(0)}$, $\mathbf{z}^{(0)}$ and $\mathbf{u}^{(0)}$
Output: reconstructed $\widehat{\mathbf{x}}_\lambda$ and p-SURE($\widehat{\mathbf{x}}_\lambda$)
for $i = 1, 2, ...$ *(ADMM iteration)* **do**
  1 compute $\mathbf{x}^{(i)}$ by (7);
  2 update $\mathbf{J}_\mathbf{y}(\mathbf{x}^{(i)})$ by (9);
  3 compute p-SURE of $\mathbf{x}^{(i)}$ by (8);
end

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 p-SURE indicates the optimal λ. This *global search* has been frequently used in [6, 7, 12].

2.4. Applications to *analysis* and *synthesis* formulations

The above discussions cover both *analysis* and *synthesis* formulations with the following replacements of notations in (6):

$$\begin{cases} \textit{analysis}: & \mathbf{A} = \mathbf{H} \\ \textit{synthesis}: & \mathbf{A} = \mathbf{H}\mathbf{R},\ \mathbf{D} = \mathbf{I},\ \mathbf{x} = \mathbf{c} \end{cases}$$

To exemplify the proposed method, we consider a typical wavelet-based signal reconstruction, where $\mathbf{D}$ and $\mathbf{R}$ denote wavelet decomposition and reconstruction, respectively.

2.4.1. *Orthogonal decimated wavelet transform (DWT)*

If $\mathbf{D}$ is orthogonal (e.g. Haar DWT), then, $\mathbf{D}^{\mathrm{T}} = \mathbf{D}^{-1} = \mathbf{R}$. The update of $\mathbf{c}^{(i)}$ of *synthesis* is:

$$\begin{aligned}\mathbf{c}^{(i+1)} &= \left(\mathbf{I}+\tau\mathbf{R}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{R}\right)^{-1}\left(\tau\mathbf{R}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{y}+(\mathbf{z}^{(i)}-\tau\mathbf{u}^{(i)})\right)\\ &= \left(\mathbf{D}(\mathbf{I}+\tau\mathbf{H}^{\mathrm{T}}\mathbf{H})\mathbf{D}^{\mathrm{T}}\right)^{-1}\left(\tau\mathbf{D}\mathbf{H}^{\mathrm{T}}\mathbf{y}+(\mathbf{z}^{(i)}-\tau\mathbf{u}^{(i)})\right)\\ &= \mathbf{D}\underbrace{\left(\mathbf{I}+\tau\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\left(\tau\mathbf{H}^{\mathrm{T}}\mathbf{y}+\mathbf{D}^{\mathrm{T}}(\mathbf{z}^{(i)}-\tau\mathbf{u}^{(i)})\right)}_{\mathbf{x}^{(i+1)}\text{ in } analysis \text{ formulation}}\end{aligned}$$

We can see that both versions of ADMM exactly coincides for each step. From the computational aspect of ADMM, it is also verified that both *analysis* and *synthesis* formulations are equivalent for orthogonal $\mathbf{D}$, as claimed in [4].

2.4.2. *Redundant wavelet transform (RWT)*

If $\mathbf{D}$ denotes RWT, then, the ADMM for both formulations is given by:

$$\begin{cases}\mathbf{x}^{(i+1)} = \left(\mathbf{D}^{\mathrm{T}}\mathbf{D}+\tau\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\left(\tau\mathbf{H}^{\mathrm{T}}\mathbf{y}+\mathbf{D}^{\mathrm{T}}(\mathbf{z}^{(i)}-\tau\mathbf{u}^{(i)})\right)\\ \mathbf{c}^{(i+1)} = \left(\mathbf{I}+\tau\mathbf{R}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{R}\right)^{-1}\left(\tau\mathbf{R}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{y}+(\mathbf{z}^{(i)}-\tau\mathbf{u}^{(i)})\right)\end{cases}$$

It is obviously that $\mathbf{c}^{(i+1)} \neq \mathbf{D}\mathbf{x}^{(i+1)}$, which implies that both formulations are not equivalent, even if $\mathbf{D}$ is 1-tight frame [4].

3. SURE-BASED ALTERNATING OPTIMIZATION WITHIN ADMM

Note that the global search for the optimal λ requires repeated implementations of the ADMM, which is rather computationally expensive. It is possible to reduce the computational complexity, if the optimization of λ can be completed within ONE execution of the iterative algorithm (ADMM). A possible solution is to alternatively optimize the parameter $\lambda^{(i)}$ (by the p-SURE minimization) and update the solution $\mathbf{x}^{(i)}$ (using $\lambda^{(i)}$) within the i-th iterate of ADMM—so-called *alternating optimization*, summarized in Fig.1.

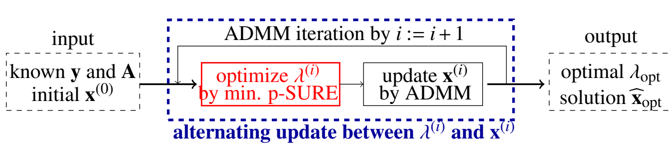

Fig. 1. Alternating optimization between parameter $\lambda^{(i)}$ and solution $\mathbf{x}^{(i)}$ within i-th step of ADMM.

Let us consider the highlighted step of Fig.1—optimization of $\lambda^{(i)}$ by minimizing p-SURE:

$$\lambda^{(i)} = \arg\min_{\lambda} \underbrace{\frac{1}{M}\left\|\mathbf{A}\mathbf{x}_{\lambda}^{(i)}-\mathbf{y}\right\|_2^2+\frac{2\sigma^2}{M}\mathrm{Tr}\left(\mathbf{A}\mathbf{J}_{\mathbf{y}}(\mathbf{x}_{\lambda}^{(i)})\right)-\sigma^2}_{\text{p-SURE of } \mathbf{x}^{(i)}} \quad (10)$$

given the previous step $\mathbf{x}^{(i-1)}$. Here, $\mathbf{x}_{\lambda}^{(i)}$ denotes the i-th update by tentative λ. And then, compute $\mathbf{x}^{(i)}$ using the optimal $\lambda^{(i)}$. Since the p-SURE is not a simple function of λ, a straightforward method is to perform exhaustive search for the optimal value of $\lambda^{(i)}$ during each iterate.

Algorithm 2: Alternating optimization within ADMM

Input: $\mathbf{y}$, $\mathbf{A}$, $\mathbf{D}$, τ, initial $\mathbf{x}^{(0)}$, $\mathbf{z}^{(0)}$ and $\mathbf{u}^{(0)}$, $i = 1$
Output: optimal λ_{opt}, $\widehat{\mathbf{x}}_{\mathrm{opt}}$ and p-SURE($\widehat{\mathbf{x}}_{\mathrm{opt}}$)
while *stopping criterion is not met* **do**
  1 optimize $\lambda^{(i)}$ by (10);
  2 update $\mathbf{x}^{(i)}$ using $\lambda^{(i)}$ by (7);
  3 compute $\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})$ by (9);
  4 compute p-SURE of $\mathbf{x}^{(i)}$ by (8);
  5 $i := i+1$;
end

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we are going to solve both *analysis* and *synthesis* formulations by ADMM, and present the results of the proposed algorithms.

4.1. Experimental setting

We consider a 1-D signal deconvolution problem, where $\mathbf{H} \in \mathbb{R}^{256\times256}$ denotes a convolution matrix constructed by a Gaussian blur, $\mathbf{x}_0 \in \mathbb{R}^{256}$ as a simple 1-D signal (shown in Fig.7). Then, we add the noise ϵ with noise variance σ^2 to obtain the observed data $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \epsilon$, such that the blur-SNR is 30dB[1]. We consider $\mathbf{D}$ as Haar DWT and RWT, respectively.

4.2. Global optimization of λ

First, we apply **Algorithm 1** to solve both formulations with fixed $\lambda = 1$ under Haar DWT and RWT. Figs.2–3 show the ADMM convergence. We can also see that the p-SURE is always a reliable substitute for p-MSE during the iterations.

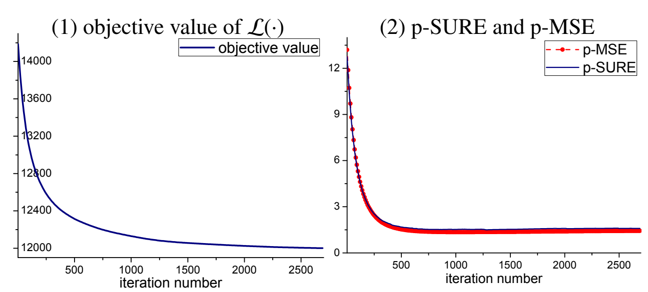

Fig. 2. The convergence of ADMM with fixed $\lambda = 1.0$ (Haar DWT).

[1] Blur signal-to-noise ratio (BSNR) is defined as: $10\log_{10}(\|\mathbf{H}\mathbf{x}_0 - \mathrm{mean}(\mathbf{H}\mathbf{x}_0)\|_2^2/(M\sigma^2))$ in dB.

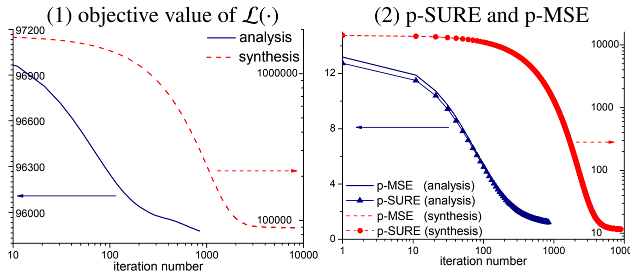

Fig. 3. The convergence of ADMM with fixed $\lambda = 1.0$ (Haar RWT).

We repeatedly implement **Algorithm 1** for various λ, and show the global search for optimal λ in Fig.4.

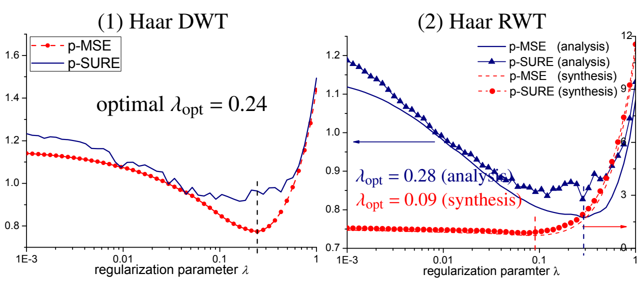

Fig. 4. p-SURE as a function of λ: global search for optimal λ.

4.3. Alternating optimization between $\lambda^{(i)}$ and $\mathbf{x}^{(i)}$

For Haar DWT, we implement **Algorithm 2** to perform alternating optimization, and show the results in Fig.5. Fig.6 shows the results of Haar RWT in *analysis*. To save the page space, we omit the Haar RWT of *synthesis* here. Refer to Table 1 for the complete comparisons.

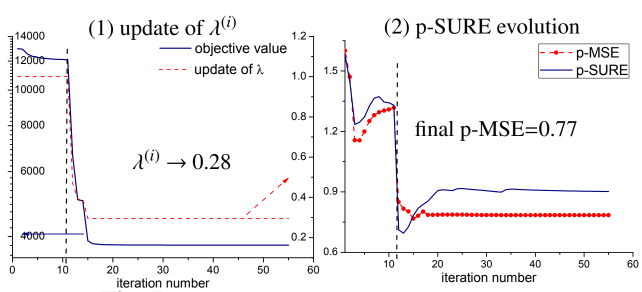

Fig. 5. Alternating optimization (Haar DWT).

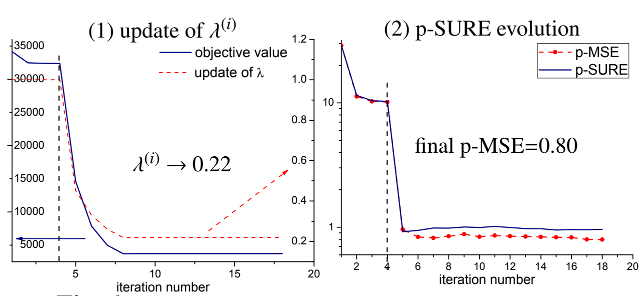

Fig. 6. Alternating optimization (Haar RWT–analysis).

Fig.7 shows two fractions of reconstructed signal $\widehat{\mathbf{x}}$ by Haar DWT and RWT, with p-SURE optimized λ.

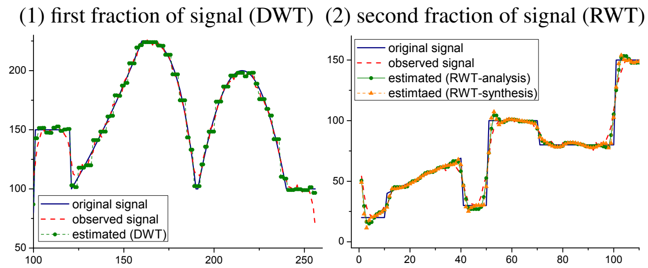

Fig. 7. Two fractions of the signal reconstruction.

Table 1 reports the results and computational time of both optimizations ('ALT' in this table denotes alternating optimization). The MSE and p-MSE are defined as: MSE $= \|\widehat{\mathbf{x}} - \mathbf{x}_0\|_2^2/N$ and p-MSE $= \|\mathbf{H}\widehat{\mathbf{x}} - \mathbf{H}\mathbf{x}_0\|_2^2/M$, respectively.

Table 1. Comparisons between global and alternating opt.

methods	est. λ	MSE	p-MSE	time (in sec.)
global (DWT)	0.24	19.05	0.77	5459
ALT (DWT)	0.28	20.21	0.77	13
global (RWT-analysis)	0.28	26.58	0.77	13406
ALT (RWT-analysis)	0.22	14.05	0.80	19
global (RWT-synthesis)	0.09	21.38	0.72	53331
ALT (RWT-synthesis)	0.07	20.19	0.74	49

From Table 1, we can see that the alternating optimization yields nearly optimal reconstruction with 100–1000 times faster speed. The remarkably improved computational efficiency is due to the following two reasons: (1) global search requires 50 times of implementations of ADMM with various λ; (2) alternating strategy greatly accelerates the convergence speed of ADMM (usually less than 100 steps to converge), and complete the optimization in ONE execution.

5. CONCLUSIONS

In this paper, to solve both ℓ_1-based *analysis* and *synthesis* formulations, we proposed two optimization schemes based on a recursive evaluation of predicted-SURE for ADMM algorithm. The alternating optimization yields nearly optimal value of λ within ONE execution of ADMM.

We would like to emphasize that not limited to the simple 1-D example shown here, the developments of this paper can be applied to many practical applications, e.g. compressed sensing [3] and image deconvolution [1].

6. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under grant number 61401013. The authors are obliged to Dr. Thierry BLU of The Chinese University of Hong Kong for many useful discussions on this article.

7. REFERENCES

[1] 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.

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

[3] 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.

[4] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, pp. 947–968, 2007.

[5] N. Parikh and S. S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] F. Xue, Anatoly G. Yagola, J.Q. Liu, and G. Meng, "Recursive SURE for iterative reweighted least square algorithms," *Inverse Problems in Science and Engineering*, vol. 24, no. 4, pp. 625–646, 2016.