A RECURSIVE PREDICTIVE RISK ESTIMATE FOR PROXIMAL ALGORITHMS

Feng Xue, Runle Du and Jiaqi Liu

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

ABSTRACT

For accurate signal reconstruction, proximal gradient methods generally require proper selection of regularization parameter. In this paper, we develop two data-driven optimization schemes, based on minimization of unbiased predictive risk estimate (UPRE). First, we propose a recursive UPRE to estimate the prediction error during the proximal iterations, which can be used to optimize the regularization parameter. Second, for fast optimization, we parametrize each proximal iterate as a linear combination of few elementary functions (LET), and solve the linear weights by minimizing recursive UPRE. We further exemplify the proposed approaches with the basic iterative shrinkage/thresholding (IST) algorithms for ℓ_1-minimization. Numerical experiments show that iterating this process leads to higher reconstruction accuracy with remarkably faster computational speed than standard IST.

Index Terms— Proximal gradient methods, unbiased predictive risk estimate (UPRE), linear expansion of thresholds (LET), iterative shrinkage/thresholding (IST)

1. INTRODUCTION

In many applications in signal processing, e.g. signal recovery [1] and compressed sensing [2], there is a need for solving the following linear inverse problem [3]:

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

where $\mathbf{y} \in \mathbb{R}^M$ is the observed data, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a deterministic matrix, $\epsilon \in \mathbb{R}^M$ is a vector of i.i.d. centered Gaussian random variable with variance $\sigma^2 > 0$. The goal is to reconstruct the original data $\mathbf{x} \in \mathbb{R}^N$ from measurements $\mathbf{y}$.

As a standard technique, regularization-based approaches typically formulate the problem as:

$$\text{(P1)}: \quad \min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 + \lambda \cdot g(\mathbf{x})$$

where $g(\mathbf{x})$ is a regularization term, λ is a regularization parameter.

Proximal gradient methods have been an appealling choice for solving (P1) [4]. In general, the solution to (P1), denoted by $\widehat{\mathbf{x}}_\lambda$, is very sensitive to the value of λ. Hence, proximal algorithms often require appropriate selection of regularization parameter λ, to achieve a good trade-off between data fidelity and regularity enforcement. There have been a number of criteria for this selection of λ, e.g. generalized cross validation [5], L-curve method [6] and discrepancy principle [7], of which the most commonly used choice is mean squared error (MSE): $\text{MSE} = \mathbb{E}\{\|\widehat{\mathbf{x}}_\lambda - \mathbf{x}\|_2^2\}/N$. Due to the inaccessible $\mathbf{x}$ in MSE, the Stein's unbiased risk estimate (SURE) was proposed in [8] as a statistical estimate of MSE

However, SURE is valid only if $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is invertible [3]. [3, 9] developed regularized or projected SURE for *iterative shrinkage/thresholding* (IST) to overcome this limitation. In this paper, for stable and easy manipulation, we consider the expected prediction error (EPE) instead [10]:

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

where $\widehat{\mu}_\lambda = \mathbf{A}\widehat{\mathbf{x}}_\lambda$.

Since true μ is unknown in practice, unbiased predictive risk estimate (UPRE) is a statistical substitute for the prediction error [11, 12]:

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

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

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

In this work, motivated by the works of [3, 9], we propose two optimization methods for accurate reconstruction, based on minimization of UPRE (3). Extending IST, we develop a recursive UPRE for the general proximal algorithms, and the optimal λ can be recognized by exhaustive search for the minimum UPRE. Moreover, for fast optimization, we adopt a very similar strategy to [13–15]: approximate the sparse estimation process by a linear combination of few elementary functions (LET bases) with different but fixed λ, and solve the linear weights (LET coefficients) by minimizing recursive UPRE. Experimentally, proximal iteration of the recursive UPRE-LET process remarkably improves the reconstruction performance in terms of both estimation error and computational time, compared to standard IST.

2. RISK ESTIMATE FOR PROXIMAL ALGORITHMS

2.1. Basic scheme of proximal algorithms

Not limited to (P1), we consider the following generic optimization problem [4]:

$$\text{(P2):} \quad \min_{\mathbf{x}} f(\mathbf{x}) + \lambda \cdot g(\mathbf{x})$$

where $f(\mathbf{x})$ is twice differentiable, $g(\mathbf{x})$ is at least subdifferentiable. Proximal gradient methods solve (P2), by the following iterate [4]:

$$\mathbf{x}^{(i+1)} = \text{prox}_{t\lambda g}\Big(\underbrace{\mathbf{x}^{(i)} - t\nabla f(\mathbf{x}^{(i)})}_{\mathbf{u}^{(i)}}\Big) \tag{4}$$

where t is a step size, the proximal operator $\text{prox}_g(\cdot)$ denotes the proximal point of $(\cdot)$ w.r.t. the function g [1, 4]. Correspondingly, $\mu^{(i)}$ is updated by $\mu^{(i)} = \mathbf{A}\mathbf{x}^{(i)}$. By (3), the UPRE of $\mu^{(i)}$ can be expressed in terms of $\mathbf{x}^{(i)}$:

$$\text{UPRE}(\mu^{(i)}) = \frac{1}{M}\left\|\mathbf{A}\mathbf{x}^{(i)} - \mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{M}\text{Tr}\Big(\mathbf{A}\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})\Big) - \sigma^2 \tag{5}$$

2.2. Recursion of Jacobian matrix

The UPRE computation requires to evaluate $\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})$. From (4), we have the Jacobian matrix of $\mathbf{x}^{(i+1)}$ is:

$$\begin{aligned}\Big[\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i+1)})\Big]_{m,n} &= \frac{\partial[\text{prox}_{t\lambda g}(\mathbf{u}^{(i)})]_m}{\partial y_n} = \sum_{s=1}^{N} \underbrace{\frac{\partial[\text{prox}_{t\lambda g}(\mathbf{u}^{(i)})]_m}{\partial u_s^{(i)}}}_{\mathbf{P}_{m,s}^{(i)}} \cdot \frac{\partial u_s^{(i)}}{\partial y_n} \\ &= \Big[\mathbf{P}^{(i)}\mathbf{J}_{\mathbf{y}}(\mathbf{u}^{(i)})\Big]_{m,n}\end{aligned}$$

Finally, we obtain that the recursion of Jacobian matrix for the update (4) is:

$$\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i+1)}) = \mathbf{P}^{(i)}\Big(\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)}) - t\mathbf{H}^{(i)}\Big) \tag{6}$$

where the matrix $\mathbf{H}^{(i)} \in \mathbb{R}^{N\times M}$ is given as $\mathbf{H}^{(i)} = \frac{\partial^2 f(\mathbf{x}^{(i)})}{\partial \mathbf{x}^{(i)}\partial \mathbf{y}}$.

2.3. Summary of the proposed algorithm

Now, we summarize the proposed risk estimation for proximal algorithms as **Algorithm 1**, which enables us to solve (P2) with a prescribed value of λ, and simultaneously evaluate the UPRE during the proximal iterations.

3. A PROXIMAL UPRE-LET APPROACH

3.1. Related works

The proposed UPRE evaluation (i.e. **Algorithm 1**) can be used to optimize λ of (P2). [3, 9] discussed a number of optimization procedures, among which the most typical methods include:

Algorithm 1: UPRE evaluation for proximal algorithms

Input: $\mathbf{y}$, $\mathbf{A}$, σ^2, λ, t, initial $\mathbf{x}^{(0)}$
Output: reconstructed $\widehat{\mathbf{x}}_\lambda$, $\widehat{\mu}_\lambda$, and UPRE($\widehat{\mu}_\lambda$)
for $i = 1, 2, \ldots$ **do**
  1 update $\mathbf{x}^{(i)}$ by (4);
  2 update $\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)})$ by (6);
  3 compute UPRE of $\mu^{(i)}$ by (5);
end

- *Global method* [9]: an exhaustive search, which repeatedly implements **Algorithm 1** with various tentative values of λ, and choose one with minimum risk estimate.
- *Greedy method* [3]: during each iteration, to perform exhaustive search for updating λ and $\mathbf{x}$ alternatively, by minimizing the evolved risk estimator.

3.2. Recursive UPRE-LET for proximal algorithms

Both global and greedy methods require the exhaustive search for the optimization, which is rather time consuming. Now, based on **Algorithm 1**, we propose a novel optimization procedure, to substantially improve the computational speed.

We adopt a very similar strategy to [13–15], which decomposes each proximal iterate (4) into a linear combination of elementary functions—Linear Expansion of Thresholds (LET):

$$\mathbf{x}^{(i+1)} = \sum_{k=1}^{K} a_k \cdot \underbrace{\text{prox}_{t\lambda_k g}\Big(\mathbf{x}^{(i)} - t\nabla f(\mathbf{x}^{(i)})\Big)}_{\mathbf{x}_k^{(i+1)}}; \ \mu^{(i+1)} = \sum_k a_k \underbrace{\mathbf{A}\mathbf{x}_k^{(i+1)}}_{\mu_k^{(i+1)}} \tag{7}$$

i.e., the update $\mathbf{x}^{(i+1)}$ is a linear combination (by LET coefficients a_k) of a number of LET bases $\mathbf{x}_k^{(i+1)}$, which are updated with different but fixed λ_k, individually.

By the LET strategy (7), the optimization problem becomes finding optimal LET coefficients a_k instead of the nonlinear parameter λ. Substituting (7) into (5), we have:

$$\text{UPRE}(\mu^{(i)}) = \frac{1}{M}\Big\|\sum_{k=1}^{K} a_k\mathbf{A}\mathbf{x}_k^{(i)} - \mathbf{y}\Big\|_2^2 + \frac{2\sigma^2}{M}\sum_{k=1}^{K} a_k\text{Tr}\Big(\mathbf{A}\mathbf{J}_{\mathbf{y}}(\mathbf{x}_k^{(i)})\Big) - \sigma^2 \tag{8}$$

where the Jacobian matrix $\mathbf{J}_{\mathbf{y}}(\mathbf{x}_k^{(i)})$ is evolved as:

$$\mathbf{J}_{\mathbf{y}}(\mathbf{x}_k^{(i+1)}) = \mathbf{P}_k^{(i)}\Big(\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)}) - t\mathbf{H}^{(i)}\Big), \quad \text{for } k = 1, 2, \ldots, K \tag{9}$$

by (6). Here, the matrix $\mathbf{P}$ depends on different regularization parameter λ_k.

The UPRE (8) is a quadratic functional of a_k: minimizing UPRE w.r.t. a_k boils down to solving the following linear system of equations:

$$\sum_{k'=1}^{K} \underbrace{\frac{1}{M}\mu_{k'}^{(i)\text{T}}\mu_k^{(i)}}_{\mathbf{M}_{k,k'}} a_{k'} = \underbrace{\frac{1}{M}\Big(\mathbf{y}^{\text{T}}\mu_k^{(i)} - \sigma^2\text{Tr}\Big(\mathbf{J}_{\mathbf{y}}(\mu_k^{(i)})\Big)\Big)}_{c_k} \tag{10}$$

for $k = 1, 2, ..., K$. These equations can be summarized in matrix form as $\mathbf{Ma} = \mathbf{c}$, 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 UPRE-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 UPRE-LET method consists in finding the best combination of the candidates $\mu_k^{(i)}$ in terms of UPRE, which is automatically done by solving (10). 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 UPRE-LET update (7) to achieve smaller prediction error with faster computational speed, though it is not an exact solution to (P2) with any value of λ. The proposed method is summarized in **Algorithm 2**.

Algorithm 2: UPRE-LET within proximal algorithms

Input: $\mathbf{y}$, $\mathbf{A}$, σ^2, t, initial $\mathbf{x}^{(0)}$, λ_k for $k = 1, 2, ..., K$
Output: reconstructed $\widehat{\mathbf{x}}_\lambda$, $\widehat{\mu}_\lambda$, and UPRE($\widehat{\mu}_\lambda$)
for $i = 1, 2, ...$ **do**
  1 update $\mathbf{x}_k^{(i)}$ and $\mu_k^{(i)}$ by (7) for $k = 1, 2, ..., K$;
  2 update of $\mathbf{J}_{\mathbf{y}}(\mathbf{x}_k^{(i)})$ by (9) for $k = 1, 2, ..., K$;
  3 build and solve (10) for a_k;
  4 update $\mathbf{x}^{(i)}$ and $\mu^{(i)}$ by (7);
  5 compute UPRE of $\mu^{(i)}$ by (8);
end

4. IST — A TYPICAL EXEMPLIFICATION

4.1. Basic scheme of IST

In this section, we exemplify Algorithms 1 and 2 with a typical class of proximal algorithms—iterative shrinkage/thresholding (IST), which is, typically, used to solve the following ℓ_1-minimization problem:

$$(\mathrm{P3}): \quad \min_{\mathbf{x}} \underbrace{\frac{1}{2}\|\mathbf{Ax}-\mathbf{y}\|_2^2 + \lambda \cdot \|\mathbf{x}\|_1}_{\mathcal{L}(\mathbf{x})}$$

for sparse reconstruction of $\mathbf{x}$. IST algorithm updates $\mathbf{x}^{(i)}$ as [14]:

$$\mathbf{x}^{(i+1)} = \mathcal{T}_{t\lambda}\Big(\underbrace{\mathbf{x}^{(i)} - t(\mathbf{A}^{\mathrm{T}}\mathbf{Ax}^{(i)} - \mathbf{A}^{\mathrm{T}}\mathbf{y})}_{\mathbf{u}^{(i)}}\Big) \tag{11}$$

where $\mathcal{T}_T(\cdot)$ is a point-wise soft-thresholding function: $\mathcal{T}_T(\cdot) = \mathrm{sign}(\cdot)(|\cdot| - T)_+$ [14]. Here, the parameter λ is regarded as a threshold, to decide which components of $\mathbf{u}^{(i)}$ are to be set to zero, i.e., the level of sparsity.

4.2. Recursive risk estimate for IST

Again, the recursions of Jacobian matrix and UPRE for IST are given by (6) and (5), respectively, where $\mathbf{H}^{(i)} = \mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{J}_{\mathbf{y}}(\mathbf{x}^{(i)}) - \mathbf{A}^{\mathrm{T}}$, $\mathbf{P}^{(i)}$ becomes diagonal with (n,n)-th element:

$$P_{n,n}^{(i)} = \begin{cases} 1, & \text{if } |u_n^{(i)}| > t\lambda \\ 0, & \text{if } |u_n^{(i)}| \le t\lambda \end{cases}$$

We now apply the UPRE-LET strategy (i.e. **Algorithm 2**) to optimize the IST update (11). Here, we set K different but fixed λ_k:

$$\mathbf{x}^{(i+1)} = \sum_{k=1}^{K} a_k \underbrace{\mathcal{T}_{t\lambda_k}\Big(\mathbf{x}^{(i)} - t(\mathbf{A}^{\mathrm{T}}\mathbf{Ax}^{(i)} - \mathbf{A}^{\mathrm{T}}\mathbf{y})\Big)}_{\mathbf{x}_k^{(i+1)}} \tag{12}$$

which implies that the candidates $\mathbf{x}_k^{(i+1)}$ have different levels of sparsity. The coefficients a_k obtained by solving (10) automatically constitute the optimal sparsity with minimum UPRE.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we are going to solve (P3) by IST, and present the results of the proposed recursive UPRE (i.e. **Algorithm 1**) and proximal UPRE-LET algorithm (i.e. **Algorithm 2**).

5.1. 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 zero-mean noise ϵ with noise variance σ^2 to obtain the observed data $\mathbf{y} = \mathbf{Ax} + \epsilon$, such that the input SNR is 10dB[1]. We set step size $t = 10^{-4}$.

5.2. Convergence of IST with fixed λ

We apply **Algorithm 1** to solve (P3) with fixed λ. Fig.1 shows the convergence of IST with $\lambda = 1$. The objective value of $\mathcal{L}(\mathbf{x}^{(i)})$ keeps decreasing to converge, shown in Fig.1-(1). Fig.1-(2) shows the evolutions of UPRE and true EPE during the iterations. We can see that the UPRE is always a reliable substitute for EPE.

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

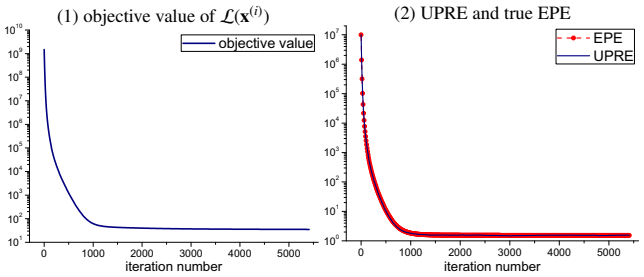

Fig. 1. The convergence of IST with fixed $\lambda = 1.0$.

5.3. Optimization of IST algorithms

In this part, we compare our proposed UPRE-LET method to global search. For global search, we implement IST for 50 tentative values of λ, and obtain the UPRE of corresponding $\widehat{\mu}_\lambda$. Fig.2-(1) shows the relation between UPRE and λ, where the minimum point corresponds to optimal value of λ. For UPRE-LET, we set $K = 3$ regularization parameters: $\lambda_1 = 1$, $\lambda_2 = 10$ and $\lambda_3 = 100$. Fig.2-(2) shows that by UPRE-LET, we obtain the optimal reconstruction in ONE implementation of IST, which achieves smaller UPRE with much faster convergence speed, compared to basic IST (shown in Fig.1).

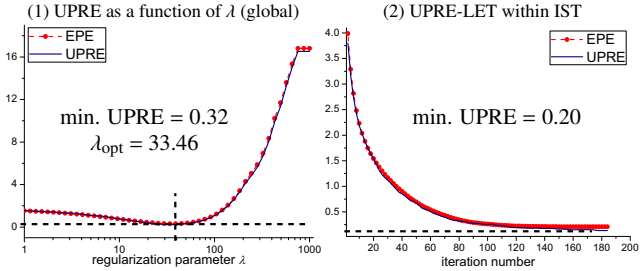

Fig. 2. Optimization of IST: global vs. UPRE-LET.

Fig. 3 shows two fractions of reconstructed signal for the comparison between global optimal and UPRE-LET method.

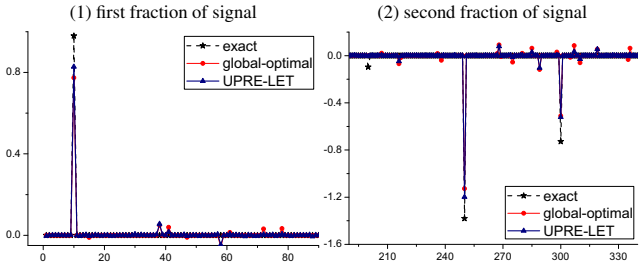

Fig. 3. Two small fractions of the signal reconstruction.

Table 1 reports the errors and computational time of global, greedy and the proposed UPRE-LET methods. 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.

We can see that the proposed approach produces more accurate reconstruction. The remarkably improved computational efficiency is due to the following several facts: (1) global method requires 50 times of implementations of IST with various λ; (2) greedy method needs to perform exhaustive search to optimize λ for each IST update; (3) UPRE-LET greatly accelerates the convergence speed of IST, and complete the optimization in ONE execution; (4) UPRE-LET for each IST update finally boils down to solving a 3-order (i.e., $K = 3$) linear system of equations (10), which costs negligible time, compared to exhaustive search.

Table 1. Comparisons between various methods

methods	global	greedy	UPRE-LET
error of $\mathbf{x}$	7.29×10^{-4}	7.31×10^{-4}	$\mathbf{4.23\times10^{-4}}$
error of μ	0.32	0.33	**0.20**
time (in sec.)	1835.65	165.15	**5.30**

6. CONCLUSIONS

In this paper, we propose a predictive risk estimate for a general proximal gradient methods, which is recursively evaluated during the proximal iterations. Moreover, we propose a UPRE-LET strategy within proximal algorithms, and demonstrated the superior performance by more accurate reconstruction, faster convergence speed and computational time.

Besides the basic IST, this work, in principle, can be extended to more complicated proximal algorithms, e.g. FISTA [16] and ADMM [4]. In addition, not limited to the simple numerical example shown in this paper, the proposed approach has a great potential for many real applications, e.g. compressed sensing and image deconvolution.

Theoretical derivations in this work related to the evaluation of Jacobian matrix and linear parametrization strategy can be extended, in principle, to other types of regularizers and regularized iterative reconstruction algorithms.

7. ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China under Grant No. 61401013.

8. REFERENCES

[1] P.L. Combettes and V.R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[2] 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.

[3] 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, 2011.

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

[5] 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.

[6] 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.

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

[8] Y.C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 471–481, 2009.

[9] C. Vonesch, S. Ramani, and M. Unser, "Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint," in *Proceedings of the 15th IEEE International Conference on Image Processing,* San Diego CA, USA, October 12-15, 2008, pp. 665–668.

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

[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] Youzuo Lin, Brendt Wohlberg, and Hongbin Guo, "UPRE method for total variation parameter selection," *Signal Processing*, vol. 90, no. 8, pp. 2546–2551, 2010.

[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] 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.

[15] 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.

[16] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.