A NOVEL GCV-BASED CRITERION FOR PARAMETER SELECTION IN IMAGE DECONVOLUTION

Feng Xue[a], *Thierry Blu*[b], *Jiaqi Liu*[a] *and Xia Ai*[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

A proper selection of regularization parameter is essential for regularization-based image deconvolution. The main contribution of this paper is to propose a new form of generalized cross validation (GCV) as a criterion for this optimal selection. Incorporating a nil-trace non-linear estimate, we develop this new GCV based on Stein's unbiased risk estimate (SURE)—an unbiased estimate of mean squared error (MSE). The key advantage of this GCV over SURE is that it does not require the knowledge of noise variance. We exemplify this criterion with both Tikhonov regularization and ℓ_1-based sparse deconvolution. In particular, we develop a recursive evaluation of GCV for the ℓ_1-estimate based on iterative soft-thresholding (IST) algorithm. Numerical experiments demonstrate the nearly optimal parameter selection and negligible loss of its resultant deconvolution quality.

Index Terms— Image deconvolution, regularization parameter, generalized cross validation (GCV), Stein's unbiased risk estimate (SURE).

1. INTRODUCTION

Image deconvolution attempts to solve the original image $\mathbf{x}_0 \in \mathbb{R}^N$ from the following linear observation model [1–3]:

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

where $\mathbf{y} \in \mathbb{R}^N$ is the observed image, $\mathbf{H} \in \mathbb{R}^{N\times N}$ denotes a convolution matrix, $\epsilon \in \mathbb{R}^N$ denotes an additive white Gaussian noise with variance $\sigma^2 > 0$. Regularization technique formulates the estimation of $\mathbf{x}_0$ as [1–3]:

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

where $\lambda > 0$ is a regularization parameter that balances the data fidelity and the regularity enforcement $\mathcal{J}(\mathbf{D}\mathbf{x})$.

Note that the solution to (2), denoted by $\widehat{\mathbf{x}}$, in general strongly depends on the value of λ [3–5]. The Stein's unbiased risk estimate (SURE) has been developed as a popular criterion for this selection (if matrix $\mathbf{H}$ is invertible) [6, 7]:

$$\text{SURE} = \frac{1}{N}\left\|\widehat{\mathbf{x}}-\mathbf{H}^{-1}\mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{N}\text{Tr}\Big(\mathbf{H}^{-\mathrm{T}}\mathbf{J}_{\mathbf{y}}(\widehat{\mathbf{x}})\Big) - \frac{\sigma^2}{N}\text{Tr}\Big(\mathbf{H}^{-1}\mathbf{H}^{-\mathrm{T}}\Big) \tag{3}$$

which is a statistical substitute for the mean squared error (MSE):

$$\text{MSE} = \frac{1}{N}\mathbb{E}\Big\{\left\|\widehat{\mathbf{x}}-\mathbf{x}_0\right\|_2^2\Big\} \tag{4}$$

Here, $\mathbf{J}_{\mathbf{y}}(\widehat{\mathbf{x}}) \in \mathbb{R}^{N\times N}$ in (3) is a Jacobian matrix defined as [4]:

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

which measures the sensitivity to change of $\widehat{\mathbf{x}}$, which is determined by $\mathbf{y}$.

Note that SURE (3) depends on $\mathbf{y}$ only, and thus, can be practically used instead of MSE (4). Recently, SURE has been extensively applied as an optimization tool for wavelet-based denoising [8, 9], Wiener-type filtering [7, 10, 11] and ℓ_1-based sparse deconvolution [3–5]. However, the SURE requires the knowledge of noise variance σ^2 (see (3)), which is often unknown in practice. The purpose of this paper is to propose a novel criterion that is independent of σ^2, such that the selected parameter λ leads to nearly optimal deconvolution performance in terms of MSE.

2. A NOVEL SURE-BASED NON-LINEAR GCV

2.1. The original GCV

Limited to a linear estimate $\widehat{\mathbf{x}} = \mathbf{U}\mathbf{y}$, the original form of GCV, first proposed in [12], is given as:

$$\text{GCV} = \frac{\left\|(\mathbf{I}-\mathbf{U}\mathbf{H})\mathbf{y}\right\|_2^2}{\Big(\text{Tr}(\mathbf{I}-\mathbf{U}\mathbf{H})\Big)^2}$$

which has been proved as a modified version of predicted-SURE in [13], where the p-SURE for $\widehat{\mathbf{x}} = \mathbf{U}\mathbf{y}$ is:

$$\text{p-SURE} = \frac{1}{N}\left\|\mathbf{U}\mathbf{y}-\mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{N}\text{Tr}\Big(\mathbf{U}\Big) - \sigma^2$$

which is an unbiased estimate of predicted-MSE:

$$\text{p-MSE} = \frac{1}{N}\mathbb{E}\Big\{\left\|\mathbf{H}\widehat{\mathbf{x}}-\mathbf{H}\mathbf{x}_0\right\|_2^2\Big\}$$

However, this traditional GCV (related to predicted-SURE) has two main drawbacks:

(1) it can be applied to linear estimate only;

(2) it is somewhat a modified measure of predicted-MSE, which, compared to MSE, only takes a partial account for the actual deconvolution performance [4, 5].

Now, we adopt a method similar to [13] to develop a new form of GCV, corresponding to SURE (rather than p-SURE). We are then able to optimize parameters for any (non-linear) estimate $\widehat{\mathbf{x}}$, without the knowledge of noise variance.

2.2. A novel SURE-based non-linear GCV

For any (non-linear) estimate $\widehat{\mathbf{x}}$, the SURE is given by (3). We now consider the following linear combination:

$$\overline{\mathbf{x}} = \alpha \mathbf{H}^{-1}\mathbf{y} + (1-\alpha)\widehat{\mathbf{x}} = \widehat{\mathbf{x}} + \underbrace{\alpha(\mathbf{H}^{-1}\mathbf{y} - \widehat{\mathbf{x}})}_{\text{bias of } \overline{\mathbf{x}} \text{ w.r.t. } \widehat{\mathbf{x}}}$$

weighted by a parameter α. Then, similar to (3), the SURE of the estimate $\overline{\mathbf{x}}$ is:

$$\text{SURE}(\overline{\mathbf{x}}) = \frac{1}{N}\left\|\overline{\mathbf{x}} - \mathbf{H}^{-1}\mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{N}\text{Tr}\Big(\mathbf{H}^{-\text{T}}\mathbf{J}_{\mathbf{y}}(\overline{\mathbf{x}})\Big) - \frac{\sigma^2}{N}\text{Tr}\Big(\mathbf{H}^{-1}\mathbf{H}^{-\text{T}}\Big)$$

To make SURE($\overline{\mathbf{x}}$) to be independent of σ^2, we let

$$\text{Tr}\Big(\mathbf{H}^{-\text{T}}\mathbf{J}_{\mathbf{y}}(\overline{\mathbf{x}})\Big) = 0$$

which yields that

$$\alpha = \frac{\text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{J}_{\mathbf{y}}(\widehat{\mathbf{x}}))}{\text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{J}_{\mathbf{y}}(\widehat{\mathbf{x}})) - \text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{H}^{-1})}$$

Thus, the matrix associated with $\overline{\mathbf{x}}$, i.e. $\mathbf{H}^{-\text{T}}\mathbf{J}_{\mathbf{y}}(\overline{\mathbf{x}})$, has trace 0. The estimate $\overline{\mathbf{x}}$, therefore, is called *nil-trace (non-linear) estimate*. Thus, the SURE procedure on this estimate becomes:

$$\begin{aligned}
\text{SURE} &= \frac{1}{N}\left\|\overline{\mathbf{x}} - \mathbf{H}^{-1}\mathbf{y}\right\|_2^2 - C \\
&= \frac{1}{N}\left\|(1-\alpha)(\widehat{\mathbf{x}} - \mathbf{H}^{-1}\mathbf{y})\right\|_2^2 - C \\
&= \frac{1}{N}(1-\alpha)^2\left\|\widehat{\mathbf{x}} - \mathbf{H}^{-1}\mathbf{y}\right\|_2^2 - C \\
&= \frac{1}{N}\left(\frac{\text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{H}^{-1})}{\text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{J}_{\mathbf{y}}(\widehat{\mathbf{x}})) - \text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{H}^{-1})}\right)^2\left\|\widehat{\mathbf{x}} - \mathbf{H}^{-1}\mathbf{y}\right\|_2^2 - C \\
&\propto \underbrace{\frac{\left\|\widehat{\mathbf{x}} - \mathbf{H}^{-1}\mathbf{y}\right\|_2^2}{\Big(\text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{J}_{\mathbf{y}}(\widehat{\mathbf{x}})) - \text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{H}^{-1})\Big)^2}}_{\text{GCV}} \qquad (5)
\end{aligned}$$

where the constant $C = \frac{\sigma^2}{N}\text{Tr}\Big(\mathbf{H}^{-1}\mathbf{H}^{-\text{T}}\Big)$.

Thus, we obtain a new objective functional—non-linear GCV (5), the minimization of which is equivalent to that of the SURE of $\overline{\mathbf{x}}$. Note that there is no need of noise variance σ^2 to compute (5).

For fixed regularizer $\mathcal{J}(\mathbf{Dx})$, the GCV can be used to optimize the parameter λ in (2), instead of SURE. An intuitive idea for this optimization is to repeatedly evaluate this GCV for the estimates $\widehat{\mathbf{x}}$ obtained by various tentative values of λ, then, the minimum GCV indicates the optimal λ (see Fig.1 for example). This *global search* has been frequently used in [3, 5, 14].

3. APPLICATION TO IMAGE DECONVOLUTION

Now, we exemplify the GCV criterion with two typical deconvolution methods: Tikhonov regularization and ℓ_1-based sparse estimation.

3.1. Tikhonov regularization

Considering the regularizer $\mathcal{J}(\mathbf{Dx}) = \frac{1}{2}\|\mathbf{Dx}\|_2^2$, Tikhonov regularization gives the following estimate [10]:

$$\widehat{\mathbf{x}} = \underbrace{\Big(\mathbf{H}^{\text{T}}\mathbf{H} + \lambda\mathbf{D}^{\text{T}}\mathbf{D}\Big)^{-1}\mathbf{H}^{\text{T}}}_{\mathbf{U}}\mathbf{y}$$

where $\mathbf{D}$ denotes discrete Laplacian operator in this paper. Due to the linearity of Tikhonov estimate, we have $\mathbf{J}_{\mathbf{y}}(\widehat{\mathbf{x}}) = \mathbf{U}$, and the GCV becomes:

$$\text{GCV} = \frac{\left\|(\mathbf{U} - \mathbf{H}^{-1})\mathbf{y}\right\|_2^2}{\Big(\text{Tr}(\mathbf{H}^{-\text{T}}(\mathbf{U} - \mathbf{H}^{-1}))\Big)^2} \qquad (6)$$

3.2. ℓ_1-based sparse deconvolution

A typical wavelet-based *synthesis formulation* for the ℓ_1-based sparse deconvolution is given as [1, 2]:

$$\widehat{\mathbf{c}} = \min_{\mathbf{c}} \underbrace{\frac{1}{2}\left\|\mathbf{HRc} - \mathbf{y}\right\|_2^2 + \lambda \cdot \|\mathbf{c}\|_1}_{\mathcal{L}(\mathbf{c})} \qquad (7)$$

to solve the problem $\mathbf{y} = \mathbf{HRc}_0 + \epsilon$. Here, $\mathbf{R}$ denotes a wavelet reconstruction, $\mathbf{c}$ denotes the wavelet coefficients. We apply a basic iterative soft-thresholding (IST) algorithm to solve (7), which updates $\mathbf{c}$ as [2]:

$$\mathbf{c}^{(i+1)} = \mathcal{T}_{t\lambda}\Big(\underbrace{\mathbf{c}^{(i)} - t\Big(\mathbf{R}^{\text{T}}\mathbf{H}^{\text{T}}\mathbf{HRc}^{(i)} - \mathbf{R}^{\text{T}}\mathbf{H}^{\text{T}}\mathbf{y}\Big)}_{\mathbf{u}^{(i)}}\Big) \qquad (8)$$

where t is a step size. The GCV of the ℓ_1-estimate $\widehat{\mathbf{c}}$ can be evaluated in a recursive manner. For the update $\mathbf{c}^{(i)}$, the GCV is:

$$\text{GCV}(\mathbf{c}^{(i)}) = \frac{\left\|\mathbf{c}^{(i)} - \mathbf{R}^{-1}\mathbf{H}^{-1}\mathbf{y}\right\|_2^2}{\Big(\text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{R}^{-\text{T}}\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})) - \text{Tr}(\mathbf{H}^{-\text{T}}\mathbf{R}^{-\text{T}}\mathbf{R}^{-1}\mathbf{H}^{-1})\Big)^2}$$

where the recursion of Jacobian matrix is [15]:

$$\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i+1)}) = \mathbf{P}^{(i)}\Big(\underbrace{\Big(\mathbf{I} - t\mathbf{R}^{\text{T}}\mathbf{H}^{\text{T}}\mathbf{HR}\Big)}_{\mathbf{A}}\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)}) + t\mathbf{R}^{\text{T}}\mathbf{H}^{\text{T}}\Big) \qquad (9)$$

with the diagonal element of $\mathbf{P}^{(i)} \in \mathbb{R}^{N\times N}$:

$$\left[\mathbf{P}^{(i)}\right]_{n,n} = \begin{cases} 1, & \text{if } |u_n^{(i)}| > t\lambda \\ 0, & \text{if } |u_n^{(i)}| \le t\lambda \end{cases}$$

Thus, the final GCV of $\widehat{\mathbf{c}}$ is obtained by the IST iteration and the recursion of Jacobian matrix until convergence (see Fig.3-(1) for example).

To simplify this example, we in this paper only consider decimated Haar[1], which implies that $\mathbf{R}^{-1} = \mathbf{R}^{\mathrm{T}} = \mathbf{D}$, where $\mathbf{D}$ denotes Haar decomposition. Thus, the GCV becomes:

$$\mathrm{GCV}(\mathbf{c}^{(i)}) = \frac{\left\|\mathbf{c}^{(i)} - \mathbf{D}\mathbf{H}^{-1}\mathbf{y}\right\|_2^2}{\left(\mathrm{Tr}(\mathbf{H}^{-\mathrm{T}}\mathbf{R}\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})) - \mathrm{Tr}(\mathbf{H}^{-\mathrm{T}}\mathbf{H}^{-1})\right)^2} \tag{10}$$

3.3. Regularization of GCV

Note that unlike the original GCV, the SURE-based GCV in (6) and (10) needs to compute the inverse $\mathbf{H}^{-1}$, which may cause numerical instability for the ill-conditioned convolution matrix $\mathbf{H}$. In practice, for stable computation, we use $\mathbf{H}_\beta^{-1}$ to replace $\mathbf{H}^{-1}$:

$$\mathbf{H}_\beta^{-1} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \beta\mathbf{I})^{-1}\mathbf{H}^{\mathrm{T}}$$

with a regularization parameter β. In this paper, we empirically choose $\beta = 10^{-4}$, which guarantees both numerical stability and optimality.

3.4. Monte-Carlo for practical computation

Note that the difficulty for computing GCV lies in the trace terms of (6) and (10). For 2-D case, due to the limited computational resources (e.g. RAM), it is impractical to store the huge matrices and compute the trace.

For the Tikhonov regularization, the trace term of (6) can be easily computed in Fourier domain. For the ℓ_1-estimate (10), Monte-Carlo (MC) simulation provides an alternative way to compute the trace $\mathrm{Tr}(\mathbf{H}^{-\mathrm{T}}\mathbf{R}\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)}))$ by the following fact [9]:

$$\mathrm{Tr}(\mathbf{H}^{-\mathrm{T}}\mathbf{R}\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})) = \mathbb{E}\Big\{\mathbf{n}_0^{\mathrm{T}}\mathbf{H}^{-\mathrm{T}}\mathbf{R}\underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{c}}^{(i)}}\Big\} \tag{11}$$

where $\mathbf{n}_0 \sim \mathcal{N}(0, \mathbf{I}_N)$. By (9), with this input noise $\mathbf{n}_0$, the noise evolution during IST update is:

$$\underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i+1)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{c}}^{(i+1)}} = \mathbf{P}^{(i)}\mathbf{A}\underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{c}}^{(i)}} + t\mathbf{P}^{(i)}\underbrace{\mathbf{D}\mathbf{H}^{\mathrm{T}}\mathbf{n}_0}_{\mathbf{n}_1} \tag{12}$$

where $\mathbf{A}$ is expressed in (9). Notice that $\mathbf{n}_1$, $\mathbf{n}_{\mathbf{c}}^{(i)}$ and $\mathbf{H}^{-\mathrm{T}}\mathbf{R}\mathbf{n}_{\mathbf{c}}^{(i)}$ can be computed by Fourier and wavelet transforms, without the storage of huge matrices. The MC evaluation is summarized as **Algorithm 1**.

[1]Please keep in mind that the purpose of this paper is to demonstrate the effectiveness of the proposed GCV as a novel criterion for parameter selection, rather than to achieve the best deconvolution quality. The simple decimated Haar should suffice.

Algorithm 1: MC for GCV evaluation (for ℓ_1-estimate)

for $i = 1, 2, \ldots$ *(IST iteration)* **do**
  1 compute $\mathbf{c}^{(i)}$ by (8);
  2 compute $\mathbf{n}_1$ and $\mathbf{n}_{\mathbf{c}}^{(i)}$ by (12);
  3 compute the trace of $\mathbf{H}^{-\mathrm{T}}\mathbf{R}\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})$ by (11);
  4 compute GCV of i-th iterate by (10);
end

4. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we are going to implement the proposed GCV-based parameter selection method for both examples mentioned above. We stress here that we perform SURE/MSE minimization to obtain the optimal value of λ, denoted as λ_{opt}, as the benchmark for comparison with the resultant λ_{GCV} by the proposed method. We use PSNR[2] to evaluate deconvolution quality, where the best $\mathrm{PSNR}_{\mathrm{opt}}$ and $\mathrm{PSNR}_{\mathrm{GCV}}$ denote the resultant quality with the parameters λ_{opt} and λ_{GCV}, respectively. The PSNR loss is defined as their difference.

We consider two typical test images *Cameraman* and *House*, blurred by the following two kernels, commonly used in the literature [1, 10]:

- **Type-I:** rational filtering: $h(i,j) = K\cdot(1+i^2+j^2)^{-1}$;
- **Type-II:** Gaussian kernel: $h(i,j) = K\cdot\exp\left(-\frac{i^2+j^2}{2s^2}\right)$ with $s = 2.0$;

Here, K is a normalization factor, s.t. $\sum_{i,j} h(i,j) = 1$. The noise levels we consider here correspond to BSNR[3] being from 40dB to 10dB.

4.1. Tikhonov regularization

Fig.1 shows the results of two degradation cases. We can see that the GCV has the same trend as MSE/SURE for various values of λ, and yields very similar results with MSE/SURE minimization.

We perform the global optimization of λ (as shown in Fig.1) for various BSNR, and show the results in Fig.2. We can see that the GCV selected values of λ are always very close to that by MSE/SURE selection for various images, blurs and noise levels.

[2]For a typical 8-bit grayscale image, peak signal-to-noise ratio (PSNR) of any estimate $\widehat{\mathbf{x}}$ w.r.t. true $\mathbf{x}_0$ is defined as: $10\log_{10}\left(\frac{255^2}{\|\widehat{\mathbf{x}}-\mathbf{x}_0\|_2^2/N}\right)$ in dB [1, 11].

[3]Blur signal-to-noise ratio (BSNR) is defined as: $10\log_{10}\left(\frac{\|\mathbf{H}\mathbf{x}_0 - \mathrm{mean}(\mathbf{H}\mathbf{x}_0)\|_2^2}{N\sigma^2}\right)$ in dB [1, 11].

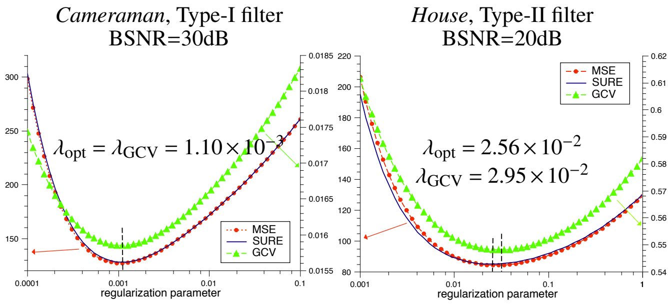

Fig. 1. Estimated λ by GCV (Tikhonov).

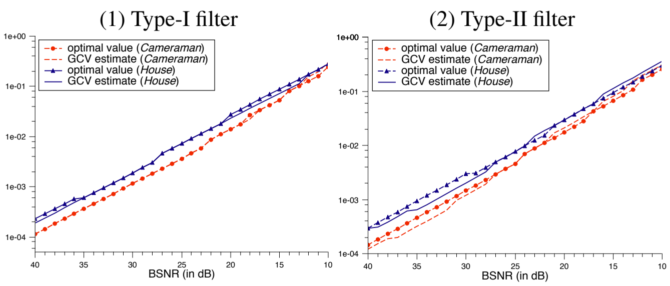

Fig. 2. Optimal values of λ under various BSNR (Tikhonov).

4.2. ℓ_1-based sparse deconvolution

We use IST (8) to obtain the ℓ_1-based estimate. Fig.3-(1) shows the recursive SURE and GCV with fixed $\lambda = 0.1$ during the iterations. Repeatedly implementing the IST for various values of λ (until convergence), we obtain the SURE and GCV of the corresponding ℓ_1-based estimate, shown in Fig.3-(2).

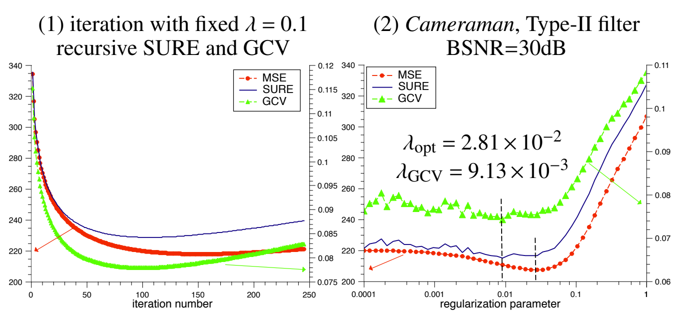

Fig. 3. Estimated λ by GCV (ℓ_1-estimate).

We perform the global optimization of λ for various BSNR, and show the results in Fig.4.

Finally, to demonstrate the influence of estimation accuracy of λ upon the deconvolution quality, we evaluate the PSNR loss, which has been defined at the beginning of Section 4. The results are shown in Fig.5. We can see that the PSNR loss is always within 0.03dB for Tikhonov and 0.3dB for ℓ_1-estimate, which indicates that the GCV selection could yield nearly optimal performance, without the knowledge of noise level.

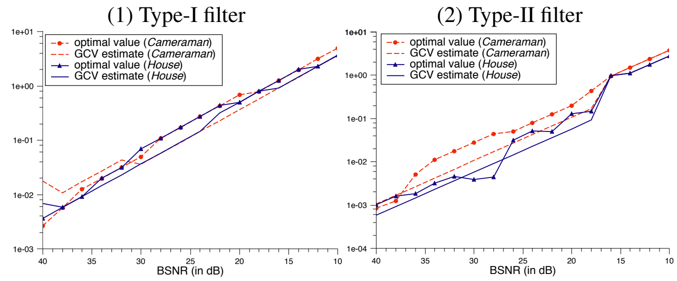

Fig. 4. Optimal values of λ under various BSNR (ℓ_1-estimate).

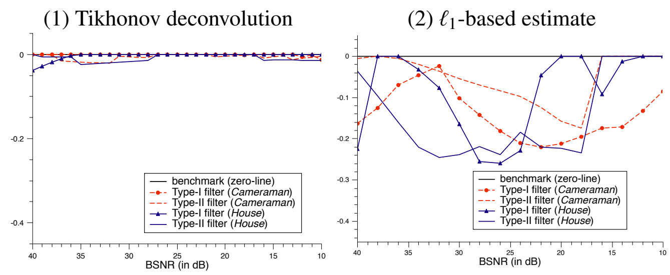

Fig. 5. PSNR loss by GCV (Tikhonov and ℓ_1-estimate).

Fig.6 shows a visual example, which demonstrates that the GCV selection yields very similar result with MSE/SURE optimization.

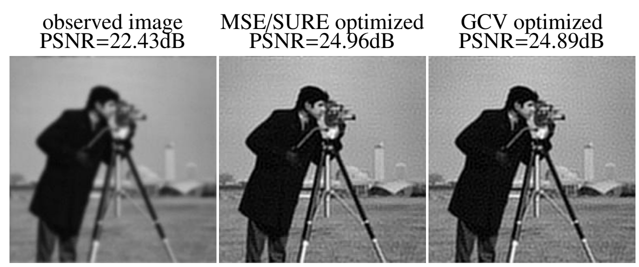

Fig. 6. A visual example of *Cameraman*.

5. CONCLUSIONS

In this paper, we proposed a novel SURE-based non-linear GCV as a criterion for selection of regularization parameter. Without knowledge of noise variance, this method yields negligible loss of deconvolution quality, compared to the MSE-based optimization. Not limited to image recovery, this proposed GCV can be efficiently used for many applications, e.g. sparse reconstruction, regression analysis and model selection.

6. ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under grant number 61401013 and the Hong Kong Research Grant Council (CUHK14210617).

7. REFERENCES

[1] 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.

[2] 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.

[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, 2010.

[4] 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.

[5] 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.

[6] Charles M Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, pp. 1135–1151, 1981.

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

[8] David L Donoho and Iain M Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[9] 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.

[10] 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.

[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] 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.

[13] Ker-Chau Li, "From Stein's unbiased risk estimates to the method of generalized cross validation," *Annals of Statistics*, vol. 13, no. 4, pp. 1352–1377, 1985.

[14] 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.

[15] Feng Xue, Runle Du, and Jiaqi Liu, "A recursive predictive risk estimate for proximal algorithms," in *Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing,* Shanghai, China, March 20-25, 2016, pp. 4498–4502.