TELKOMNIKA, Vol. 11, No. 6, June 2013, pp. 3242 ~ 3250
e-ISSN: 2087-278X

Monte-Carlo SURE for Choosing Regularization Parameters in Image Deblurring

Yubing Han*, Kelan Wang, Mengna Xu
School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China, 210094
*Corresponding author, e-mail: hanyb@mail.njust.edu.cn

Abstract

Parameter choice is crucial to regularization-based image deblurring. In this paper, a Monte Carlo method is used to approximate the optimal regularization parameter in the sense of Stein's unbiased risk estimate (SURE) which has been applied to image deblurring. The proposed algorithm is suitable for the exact deblurring functions as well as those of not being expressed analytically. We justify our claims by presenting experimental results for SURE-based optimization with two different regularization algorithms of Tikhonov and total variation regularization. Experiment results show the validity of the proposed algorithm, which has similar performance with the minimum MSE.

Keywords: *Monte-Carlo, Stein's unbiased risk estimate (SURE), image deblurrin, Tikhonov regularization, total variation*

Copyright © 2013 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Image deblurring is very common in image processing field. However, image deblurring is a ill-posed inverse problem. The concept of ill-posed problems goes back to Hadamard in the beginning of this century [1]. Hadamard essentially defined a problem to be ill-posed if the solution is not unique or it is not continuous function of the data, if an arbitrarily small perturation of the data can cause an arbitrarily large perturbation of the solution. A popular strategy for solving inverse problems is to use regularization techniques. Regularization method is a useful strategy which can stabilize the problem and to obtain a useful and stable solution. However, when applying this method, the user is faced with the difficult task of adjusting regularization parameter to obtain best performance.

Generally, the effect of reconstructed image is measured by minimizing mean squared error (MSE), as we all know, the MSE depends on the original signal which is generally is unavailable or unknown a priori, a practical approach is to replace the true MSE by some estimate in the sense of Stein's unbiased risk estimate (SURE), which depends on the given data and provides a mean for unbiased risk estimate of the true MSE [2, 3]. In recent years, the SURE criterion has been employed in variety of denosing problems for choosing regularization parameters, in that case of denosing algorithms are not being expressed analytically. It has been demonstrated that Monte Carlo method is practicable in calculation of SURE [4]. However, its application is not limited to denosing case. In this paper, we extend the SURE method to a much broader class of problems.

This paper is organized as follows. In section 2, we introduce the image degraded model and regularization method. Section 3 describes gradient descent method and image reconstruction by an iterative algorithm based on Tikhonov and total variation (TV) regularization. In section 4, we extend Monte-Carlo SURE technique to image deblurring problems. In section 5, we present experimental results and demonstrate numerically that SURE, computed using the Monte-Carlo strategy, faithfully approach the true MSE curve. Finally the conclusion is given in section 6.

2. Problem Formulation

It is well known that signals are inevitably degraded during acquisition, transmission and storage process. In most cases, there are two kinds of degraded factors, one is the deterministic

Received January 18, 2013; Revised April 7, 2013; Accepted April 22, 2013

factors, such as the defects of camera itself, the defocus blur, the motion blur, and the atmospheric disturbances, which are mainly caused by image acquisition system, the other is random factors, such as photoelectric noise, channel noise and so on. In general, we assume the noise follows a certain probability distribution, such as Gaussian distribution.

Let $\mathbf{X} \in \mathbf{R}^{L_1 \times L_2}$ be the ideal discrete signal of a continuous scene. $\mathbf{Y} \in \mathbf{R}^{M_1 \times M_2}$ is the observed degraded signal. $L_1 \geq M_1, L_2 \geq M_2$ are the sizes of the original and observed signal, respectively. The degraded model of the general signal model is given by:

$$\mathbf{Y} = \mathbf{HX} + \mathbf{E} \tag{1}$$

where $\mathbf{E} \sim \mathbf{N(0,C)}$ is zero-mean white Gaussian noise with a variance of C, H is deterministic part of degraded model, which is assumed as a linear operator, and represents any kinds of distortion, blurring and downsample in the process of image acquisition.

In the variational framework [5, 6], the reconstruction signal is obtained in general by minimizing a cost functional of:

$$J(\mathbf{X}) = \frac{\mathbf{1}}{\mathbf{2}}\|\mathbf{Y} - \mathbf{HX}\|_2^2 + \lambda \mathbf{R}(\mathbf{X}) \tag{2}$$

where $\| \ \|^2$ denotes the Euclidean norm, $\frac{\mathbf{1}}{\mathbf{2}}\|\mathbf{Y} - \mathbf{HX}\|_2^2$ is the data fidelity term that measures the consistency of X to the given data Y, and $\mathbf{R}(\mathbf{X})$ is a suitable regularization function that often penalized the lack of smoothness in X. The determination of regularization parameter λ is an important task and the main goal in this work is to optimize λ given Y [6-9].

3. Image Reconstruction based on Regularization

When the linear operator H is a blur or convolution operator, reconstruction the original signal X from the observation is called deconvolution or deblurring. Regularization method is crucial to image deblurring processing. There are two issues to be considered in regularization, the type of regularization term and the selection of regularization parameter, they all have close connection with the effect of the restored image. In this paper, we mainly discuss two regularization methods: Tikhonov regularization and TV regularization

3.1. Tikhonov Regularization

Generally we can choose regularization function as $R(\mathbf{X}) = \iint \varphi(|\nabla \mathbf{X}|)dxdy$, and let $\varphi(\mathrm{s}) = \mathrm{s}^2$, where ∇X is the gradient of X, then we have $R(\mathbf{X}) = \iint_\Omega |\nabla \mathbf{X}|^2 dxdy$, which is Tikhonov regularization [10, 11].

$$J(\mathbf{X}) = \frac{\mathbf{1}}{\mathbf{2}}\|Y - \mathbf{HX}\|_2^2 + \lambda \iint |\nabla \mathbf{X}|^2 dxdy \tag{3}$$

The Euler-Lagrange equation of (3) is:

$$\mathbf{H}^{\mathrm{T}}\mathbf{HX} - 2\lambda \Delta \mathbf{X} = \mathbf{H}^{\mathrm{T}}\mathbf{Y} \tag{4}$$

Where Δ is the Laplacian operator.

3.2. Total Variation (TV) Regularization

When $\varphi(\mathrm{s}) = \mathrm{s}$, then regularization function becomes $R(\mathbf{X}) = \iint_\Omega |\nabla \mathbf{X}| dxdy$, this is total variation (TV) regularization [12].

$$J(\mathbf{X}) = \frac{1}{2}\|Y - \mathbf{HX}\|_2^2 + \lambda \iint |\nabla \mathbf{X}| dxdy \tag{5}$$

The Euler-Lagrange equation of (5) is:

$$\mathbf{H}^T\mathbf{HX} - \lambda\, div\left(\frac{\nabla \mathbf{X}}{|\nabla \mathbf{X}|}\right)\mathbf{X} = \mathbf{H}^T\mathbf{Y} \tag{6}$$

If $|\nabla \mathbf{X}| \approx 0$, the diffusion coefficient $\frac{1}{|\nabla \mathbf{X}|} = \frac{1}{\sqrt{\mathbf{X}_x^2 + \mathbf{X}_y^2}} \approx +\infty$. That means, in the flat region of image, adding a great smoothing effect will lead to a bad staircase effect. In order to overcome this phenomenon, we introduce a parameter ε, so:

$$\hat{\mathbf{X}}_\lambda = \arg\min_{\mathbf{X}} J(\mathbf{X}) = \arg\min_{\mathbf{X}} \left\{ \frac{1}{2}\|\mathbf{Y} - \mathbf{HX}\|_2^2 + \lambda \iint |\nabla^\varepsilon \mathbf{X}| dxdy \right\} \tag{7}$$

And the corresponding Euler-Lagrange equation is:

$$\mathbf{H}^T\mathbf{HX} - \lambda\, div\left(\frac{\nabla \mathbf{X}}{|\nabla^\varepsilon \mathbf{X}|}\right)\mathbf{X} = \mathbf{H}^T\mathbf{Y} \tag{8}$$

Where $|\nabla^\varepsilon u| = \sqrt{u_x^2 + u_y^2 + \varepsilon^2}$.

3.3. Gradient Descent Method

Large-scale equation problem is inevitable in image restoration processing. The dimension of some matrix is too large, for example, if the size of a given image is 256×256, and then the size of operator H is $256^2 \times 256^2$. Therefore, some direct methods such as Gauss elimination method and LU decomposition can not be applied in practice because large matrix can not be stored in our computer. The iterative algorithms can avoid the decomposition of matrix and the amount of storage is less than that of direct methods. In this paper, we restore image by gradient descent method which is a type of iterative method. Gradient descent is also known as steepest descent method. In this iterative algorithm, the inverse operation of H can be avoided and some prior knowledge of the solution can be effectively combined in the iterative process.

A simple form of the gradient descent method is given by:

$$\mathbf{X}^{k+1} = \mathbf{X}^k - \tau \nabla J(\mathbf{X}^k) \tag{9}$$

Where k is the number of iteration. τ is the iterative step, which is a small enough to ensure the convergence of iterative algorithm. $\mathbf{X}^k$ is the estimate of $\mathbf{X}$ after k times iterative calculation and $-\nabla J(\mathbf{X}^k)$ is the negative gradient of $J(\mathbf{X})$ at $\mathbf{X}^k$.

$$\nabla J(\mathbf{X}^k) = \frac{\partial J(\mathbf{X})}{\partial(\mathbf{X})}\Big|_{\mathbf{X}=\mathbf{X}^k} \tag{10}$$

For the regularization function (2), the negative gradient is given by:

$$-\frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{H}^T\mathbf{Y} - \mathbf{H}^T\mathbf{HX} + \lambda L(\mathbf{X})\mathbf{X} \tag{11}$$

Where $L(\mathbf{X})$ is the differential operator of $R(\mathbf{X})$, iterative algorithm is as follows:

$$\mathbf{X}^k = \mathbf{X}^k - \tau \frac{\partial J(\mathbf{X})}{\partial(\mathbf{X})}\Big|_{\mathbf{X}=\mathbf{X}^{k-1}} = \mathbf{X}^{k-1} + \tau\left[\mathbf{H}^T\mathbf{Y} - \mathbf{H}^T\mathbf{HX}^{k-1} + \lambda L(\mathbf{X}^{k-1})\mathbf{X}^{k-1}\right] \tag{12}$$

4. Regularization Parameter Determination
4.1. Choosing Regurarization Parameter Based On The Minimum SURE

For the degraded model, the probability density of the observed Y can be expressed as the exponential distribution [2].

$$f(\mathbf{Y}|\mathbf{X}) = b(\mathbf{Y}) exp\{\mathbf{X}^T \varphi(\mathbf{Y}) - g(\mathbf{X})\} \tag{13}$$

Where $b(\mathbf{Y}) = \frac{1}{\sqrt{(2\pi)^n det(\mathbf{C})}} exp\left\{-\frac{1}{2}\mathbf{Y}^T\mathbf{C}^{-1}\mathbf{Y}\right\}$, $\varphi(\mathbf{Y}) = \mathbf{H}^T\mathbf{C}^{-1}\mathbf{Y}$, $g(\mathbf{X}) = \frac{1}{2}\mathbf{X}^T\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\mathbf{X}$. Apparently, a sufficient statistics for estimating X is given by $\mathbf{u} = \varphi(\mathbf{Y})$. Therefore, any reasonable of Y will be a function only of u. More specifically, from the Rao-Blackwell theorem [14], it follows that if $\hat{\mathbf{X}}_\lambda$ is an estimate of X which is not a function only of u, then the estimate $\mathbf{E}(\hat{\mathbf{X}}_\lambda|\mathbf{u})$ has lesser or equal MSE than that of $\hat{\mathbf{X}}_\lambda = h_\lambda(\mathbf{u})$, therefore, in the sequel, we only consider methods that depend on the data via u. Where $h_\lambda(\mathbf{u})$ is a function of u that depends on the observations Y and the subscript λ denotes that the estimation is related to regularization parameters. For the estimate $\hat{\mathbf{X}}_\lambda = h_\lambda(\mathbf{u})$, MSE is defined as:

$$\frac{1}{N^2} E\left\{\left\|\mathbf{X} - \hat{\mathbf{X}}_\lambda\right\|^2\right\} = \frac{1}{N^2} E\left\{\|\mathbf{X} - h_\lambda(\mathbf{u})\|^2\right\} \tag{14}$$

$$\lambda^* = \underset{\lambda}{\operatorname{argmin}}\ E\left\{\frac{1}{N^2}\left\|\mathbf{X} - \hat{\mathbf{X}}_\lambda\right\|^2\right\} = \underset{\lambda}{\operatorname{argmin}}\ E\left\{\frac{1}{N^2}\|\mathbf{X} - h_\lambda(\mathbf{u})\|^2\right\} \tag{15}$$

The sufficient statistics u lies in the range space $\aleph = \Re(\mathbf{H}^\mathbf{T})$, so $\hat{\mathbf{X}}_\lambda = h_\lambda(\mathbf{u})$ also belongs to this space. Denote by $\mathbf{P} = \mathbf{H}^\mathrm{T}(\mathbf{H}\mathbf{H}^\mathrm{T})^+\mathbf{H}$ the orthogonal projection onto $\aleph$, where H is rank-deficient. Then,

$$E\left\{\left\|\mathbf{X} - \hat{\mathbf{X}}_\lambda\right\|^2\right\} = E\left\{\left\|\mathbf{P}\mathbf{X} - \mathbf{P}\hat{\mathbf{X}}_\lambda\right\|^2\right\} + E\left\{\left\|(\mathbf{I}-\mathbf{P})\mathbf{X} - (\mathbf{I}-\mathbf{P})\hat{\mathbf{X}}_\lambda\right\|^2\right\} = E\left\{\left\|\mathbf{P}\mathbf{X} - \mathbf{P}\hat{\mathbf{X}}_\lambda\right\|^2\right\} + E\left\{\|(\mathbf{I}-\mathbf{P})\mathbf{X}\|^2\right\} \tag{16}$$

If $\hat{\mathbf{X}}_\lambda$ lies in the space $\aleph$, then $(\mathbf{I}-\mathbf{P})\hat{X}_\lambda = 0$ .and $(\mathbf{I}-\mathbf{P})\mathbf{X}$ is constant and independent of $\hat{\mathbf{X}}_\lambda$. Therefore, in this case, it is sufficient to obtain the estimate of the first term for optimizing $\hat{\mathbf{X}}_\lambda$.

Next we consider the specific Gaussian distribution expression. Suppose H has the singular value decomposition $\mathbf{H} = \mathbf{U}\Sigma\mathbf{Q}^\mathbf{T}$ for some unitary matrices U and Q. Let H has rank r so that Σ is a $\mathrm{n}\times\mathrm{m}$ diagonal matrix whose the first r diagonal elements are equal to $\sigma_i^2 > 0$. Projection matrix is $\mathbf{P} = \mathbf{V}\mathbf{V}^\mathbf{T}$, where V equals to the first r columns of Q, and let $\mathbf{X}' = \mathbf{V}^\mathbf{T}\mathbf{X}$. The sufficient statistic for estimating $\mathbf{X}'$ is $\mathbf{u}' = \mathbf{V}^\mathbf{T}\mathbf{H}^\mathbf{T}\mathbf{C}^{-1}\mathbf{Y} = \mathbf{V}^\mathbf{T}\mathbf{u}$, and $\mathbf{u}'$ is a Gaussian random vector with a mean u' and a covariance $\mathbf{C}' = \mathbf{V}^\mathbf{T}\mathbf{H}^\mathbf{T}\mathbf{C}^{-1}\mathbf{H}\mathbf{V}$. Using the SVD decomposition of H, we have:

$$\begin{cases} \boldsymbol{\mu}' = \boldsymbol{\Lambda}[\mathbf{U}^\mathbf{T}\mathbf{C}^{-1}\mathbf{U}]_\mathbf{r}\mathbf{X}' \\ \mathbf{C}' = \boldsymbol{\Lambda}[\mathbf{U}^\mathbf{T}\mathbf{C}^{-1}\mathbf{U}]_\mathbf{r} \end{cases} \tag{17}$$

Where $\boldsymbol{\Lambda}$ is a $r\times r$ diagonal matrix with diagonal elements $\sigma_i^2 > 0$ and $[\mathbf{A}]_\mathbf{r}$ is the $r\times r$ top-left principle block of size r of the matrix A. Since $\mathbf{C} > \mathbf{0}$, $\mathbf{C}'$ is invertible. Therefore:

$$f(\mathbf{u}'|\mathbf{X}') = q(\mathbf{u}') exp\{\mathbf{X}'^\mathbf{T}\mathbf{u}' - \mathbf{g}(\mathbf{X}')\} \tag{18}$$

Where $\mathbf{u}'$ is the sufficient statistic of $\mathbf{X}'$, $f(\mathbf{u}'|\mathbf{X}')$ is exponential distribution, $g(\mathbf{X}') = \frac{1}{2}\mathbf{X}'^{\mathrm{T}}\boldsymbol{\Lambda}[\mathbf{U}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{U}]_{\mathrm{r}}\mathbf{X}'$, $q(\mathbf{u}') = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C}')}}\exp\{-(1/2)\mathbf{u}'^{\mathrm{T}}\mathbf{C}'^{-1}\mathbf{u}'\}$.

In order to estimate $\mathbf{E}\left\{\left\|\mathbf{PX} - \mathbf{P}\hat{\mathbf{X}}_\lambda\right\|^2\right\}$, we must compute:

$$\mathbf{E}\{\hat{\mathbf{X}}_\lambda^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{P}\mathbf{X}\} = \mathbf{E}\{\hat{\mathbf{X}}_\lambda^{\mathrm{T}}\mathbf{P}\mathbf{X}\} = \mathbf{E}\{\hat{\mathbf{X}}_\lambda^{\mathrm{T}}\mathbf{V}\mathbf{V}^{\mathrm{T}}\mathbf{X}\} = \mathbf{E}\{\hat{\mathbf{X}}_\lambda^{\mathrm{T}}\mathbf{V}\mathbf{X}'\} \tag{19}$$

Since $\hat{\mathbf{X}}_\lambda = h_\lambda(\mathbf{u}) = h_\lambda(\mathbf{Vu}')$ and $f(\mathbf{u}'|\mathbf{X}')$ is exponential families distribution. Let $k_\lambda(u') = \mathbf{V}^{\mathrm{T}}h_\lambda(\mathbf{V}u')$ is the estimation of $\hat{\mathbf{X}}_\lambda$, then:

$$\begin{aligned}
E\left\{h_\lambda^{\mathrm{T}}(\mathbf{Vu}')\mathbf{VX}'\right\} &= E\left\{k_\lambda^{\mathrm{T}}(\mathbf{u}')\mathbf{X}'\right\} \\
&= -E\left\{Tr\left(\frac{\partial k_\lambda(\mathbf{u}')}{\partial \mathbf{u}'}\right)\right\} - E\left\{k_\lambda^{\mathrm{T}}(\mathbf{u}')\frac{\partial \ln q(\mathbf{u}')}{\partial \mathbf{u}'}\right\} \\
&= -E\left\{Tr\left(\mathbf{V}\frac{\partial h_\lambda^{\mathrm{T}}(\mathbf{Vu}')}{\partial \mathbf{u}'}\right)\right\} - E\left\{h_\lambda^{T}(\mathbf{Vu}')\mathbf{V}\frac{\partial \ln q(\mathbf{u}')}{\partial \mathbf{u}'}\right\} \\
&= -E\left\{Tr\left(\mathbf{VV}^{\mathrm{T}}\frac{\partial h_\lambda^{\mathrm{T}}(\mathbf{u})}{\partial \mathbf{u}}\right)\right\} - E\left\{h_\lambda^{\mathrm{T}}(\mathbf{u})\mathbf{V}\frac{\partial \ln q(\mathbf{u}')}{\partial \mathbf{u}'}\right\} \\
&= -E\left\{Tr\left(\mathbf{P}\frac{\partial h_\lambda^{\mathrm{T}}(\mathbf{u})}{\partial \mathbf{u}}\right)\right\} - E\left\{h_\lambda^{\mathrm{T}}(\mathbf{u})\mathbf{V}\frac{\partial \ln q(\mathbf{u}')}{\partial \mathbf{u}'}\right\}
\end{aligned} \tag{20}$$

So the unbiased estimate of the MSE $E\left\{\mathbf{P}X - \mathbf{P}\hat{X}_\lambda\right\}$ can be obtained by [2]:

$$S(h_\lambda(\mathbf{u})) = \left\|\mathbf{PX}\right\|^2 + \left\|\mathbf{P}h_\lambda(\mathbf{u})\right\|^2 + 2Tr\left(\mathbf{P}\frac{\partial h_\lambda^{\mathrm{T}}(\mathbf{u})}{\partial \mathbf{u}}\right) - 2h_\lambda^{\mathrm{T}}(\mathbf{u})\hat{\mathbf{X}}_{ML} \tag{21}$$

Where $\hat{\mathbf{X}}_{ML} = (\mathbf{H}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{H})^{+}\mathbf{H}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{Y}$ is the maximum likelihood estimation, and the $(\bullet)^{+}$ denotes the Moore-Penrose pseudo inverse.

4.2. Monte-Carlo Realization Of The Unbiased Risk Estimation

The crucial step for evaluating the SURE is the computation of $Tr\left(\mathbf{P}\frac{\partial h_\lambda^{T}(\mathbf{u})}{\partial u}\right)$; however, in most reconstruction algorithms (especially iterative algorithm), $h_\lambda(\mathbf{u})$ is not available explicitly, so we use Monte-Carlo technique to approximate the trace. To compute $Tr\left(\mathbf{P}\frac{\partial h_\lambda^{T}(\mathbf{u})}{\partial u}\right)$, we introduce the following theorem at first.

Theorem 1 [3]: let $b \in \mathbf{R}^N$ be a zero-mean i.i.d. random vector. Assume $h_\lambda(\mathbf{u})$ has the second-order Taylor expansion, so we have

$$Tr\left(\mathbf{P}\frac{\partial h_\lambda^{T}(\mathbf{u})}{\partial \mathbf{u}}\right) = \lim_{\varepsilon \to 0} E_{\mathbf{b}}\left\{\mathbf{b}^{T}\left(\frac{\mathbf{P}h_\lambda(\mathbf{u}+\varepsilon\mathbf{b}) - \mathbf{P}h_\lambda(\mathbf{u})}{\varepsilon}\right)\right\} \tag{22}$$

Proof: The second-order Taylor expansion of $h_\lambda(\mathbf{u}+\varepsilon\mathbf{b})$ can be written as:

$$h_\lambda(\mathbf{u}+\varepsilon\mathbf{b}) = h_\lambda(\mathbf{u}) + \varepsilon J_{h_\lambda}(\mathbf{u})\mathbf{b} + \varepsilon^2 e_{h_\lambda} \tag{23}$$

Where $J_{h_\lambda}(\mathbf{u})$ is the Jacobian matrix of $h_\lambda(\mathbf{u})$ evaluated at Y and $e_{\mathbf{h}_\lambda}$ represents the error vector. In this case, the components are bounded in the expectation sense, then:

$$E_b\left\{b^T\mathbf{P}\left(h_\lambda\left(\mathbf{u}+\boldsymbol{\varepsilon}\mathbf{b}\right)-h_\lambda\left(\mathbf{u}\right)\right)\right\}=\varepsilon E_{\mathbf{b}}\left\{b^T\mathbf{P}J_{h_\lambda}(\mathbf{u})\mathbf{b}\right\}+\varepsilon^2E_{\mathbf{b}}\left\{\mathbf{b}^T\mathbf{P}e_{h_\lambda}\right\}=\varepsilon tr\{\mathbf{P}J_{h_\lambda}(\mathbf{u})\}+\mathbf{C}\varepsilon^2 \tag{24}$$

Where $E_b\left\{b^T\mathbf{P}e_{h_\alpha}\right\}=C<+\infty$ because b has finite higher-order moments and the components of $e_{\mathbf{h}_\lambda}$ are bounded in the expectation sense. P is projection matrix. When $\varepsilon\to 0$, we have:

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}E_{\mathbf{b}}\left\{\mathbf{b}^T\mathbf{P}\left(h_\lambda(\mathbf{u}+\varepsilon\mathbf{b})-h_\lambda(\mathbf{u})\right)\right\}=tr\{\mathbf{P}J_{h_\lambda}(\mathbf{u})\}=Tr\left(\mathbf{P}\frac{\partial h_\lambda^T(\mathbf{u})}{\partial\mathbf{u}}\right) \tag{25}$$

So $Tr\left(\mathbf{P}\frac{\partial h_\lambda^T(\mathbf{u})}{\partial\mathbf{u}}\right)=\lim_{\varepsilon\to 0}E_{\mathbf{b}}\left\{\mathbf{b}^T\left(\frac{\mathbf{P}h_\lambda(\mathbf{u}+\varepsilon\mathbf{b})-\mathbf{P}h_\lambda(\mathbf{u})}{\varepsilon}\right)\right\}$. According to this theorem, we propose the following algorithm to approximate $\frac{1}{N^2}Tr\left(\mathbf{P}\frac{\partial h_\lambda(\mathbf{u})}{\partial\mathbf{u}}\right)$.

Step 1: generate a zero-mean i.i.d random vector $\mathbf{b}\in\mathbf{R}^N$ with unit variance.
Step 2: evaluate $h_\lambda(\mathbf{u})$.
Step 3: let $z=\mathbf{u}+\varepsilon\mathbf{b}$, and evaluate $h_\lambda(\mathbf{z})$.
Step 4: compute $\frac{1}{\varepsilon N^2}b^T\mathbf{P}\left(h_\lambda(\mathbf{z})-h_\lambda(\mathbf{u})\right)$ and $\frac{1}{N^2}Tr\left(\mathbf{P}\frac{\partial h_\lambda(\mathbf{u})}{\partial\mathbf{u}}\right)$.

5. Experimental Results

In this section, the standard cameraman and peppers images with size of 256×256 are adopted as test images. In degradation, 3×3 Gaussian kernel with a variance of 1 is used to blur the original images, and then the white Gaussian noise with a standard deviation of 0.1 is added to the blurred image.

Tikhonov regularization and TV regularization are adopted in deblurring algorithm. Figure 1(a) and Figure 3(a) are the original images; Figure 1(b) and Figure 3(b) are the degraded images; Figure 1(c) and Figure 3(c) are the restored images using Tikhonov regularization with optimized parameters; Figure 1(d) and Figure 3(d) are the restored images using TV regularization with optimized parameters. From the visual perspective, the image restored by Tikhonov regularization can not preserve the details of the edges of image, while the image restored by TV regularization is much better in keeping details of image, and the edge is more visible compared to Tikhonov regularization. The peak signal-to-noise ratio (PSNR) values are listed in Table 1 where the output PSNR based on true MSE and Monte-Carlo SURE respectively are given and the PSNR based on these method are similar. Figure 2 and Figure 4 are the SURE and MSE curves of cameraman and peppers respectively. We use Monte-Carlo SURE to select regularization parameter instead of MSE which has been described in section 4. It can be seen that the curves of SURE and MSE obtained by two reconstruction algorithms are very close. Now we compare the regularization parameter λ selected by SURE and MSE. For cameraman image in Tikhonov regularization, MSE and SURE reach the minimum almost at the same point $\lambda_{MSE}=\lambda_{SURE}=0.26$. In TV regularization, $\lambda_{MSE}=\lambda_{SURE}=0.036$. While for peppers image, the parameter λ selected by MSE and SURE has a small gap, when the Tikhonov regularization is used, the optimum regularization parameters are $\lambda_{MSE}=0.40$ and $\lambda_{SURE}=0.35$. For the TV regularization, the optimum regularization parameters are $\lambda_{MSE}=0.046$ and $\lambda_{SURE}=0.041$ respectively. From these results, we can see the effectiveness of the Monte-Carlo SURE algorithm and it is a good way to select the regularization parameter in image deblurring.

6. Conclusion

In this paper, we developed the unbiased estimate of the MSE in image deblurring by extending the SURE method, which is used to choose the optimal regularization parameter. Computation and application of SURE need to evaluate the trace, however, the computation of the trace may turn out to be nontrivial, especially when the deblurring reconstruction algorithm does not have explicit analytical form. In this paper, we use the Monte-Carlo method to solve this problem. The contribution of our work is that the Monte-Carlo SURE method is extended to the application of image debluuing. The advantage of this method for selecting parameters is that the MSE can be estimated purely from the measured data without need the knowledge of original image. Experiment results show that the optimal parameter obtained by Monte-Carlo SURE is perfectly agreed with the true minimum value of MSE.

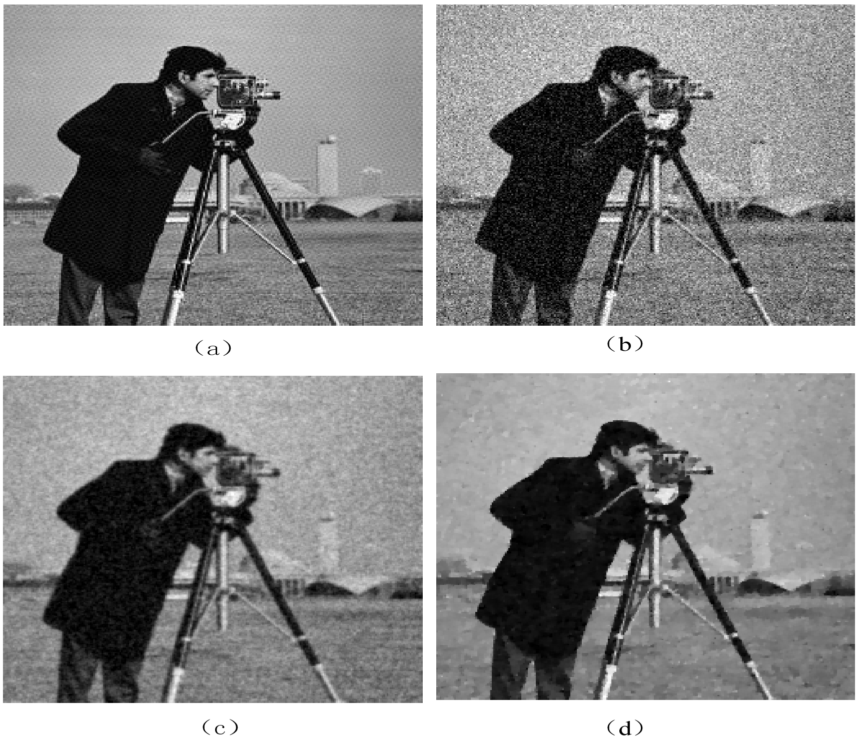

Figure 1. Visual Comparison of SURE-optimized Deblurred Results for Cameraman. (a) Original image. (b) Degraded image. (c) Restored image by using Tikhobov regularization. (d) Restored image by using TV regularization.

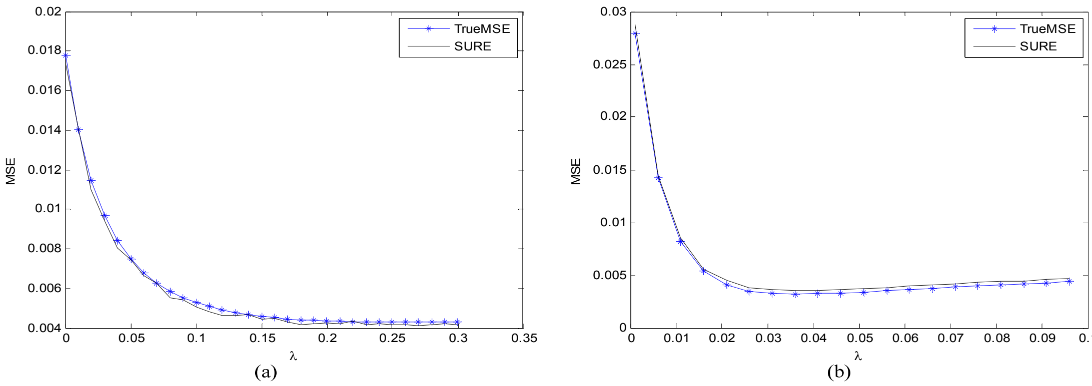

Figure 2. MSE(λ) and SURE(λ) for Cameraman. (a) MSE(λ) and SURE(λ) based on Tikhonov Regularization. (b) MSE(λ) and SURE(λ) based on TV Regularization.

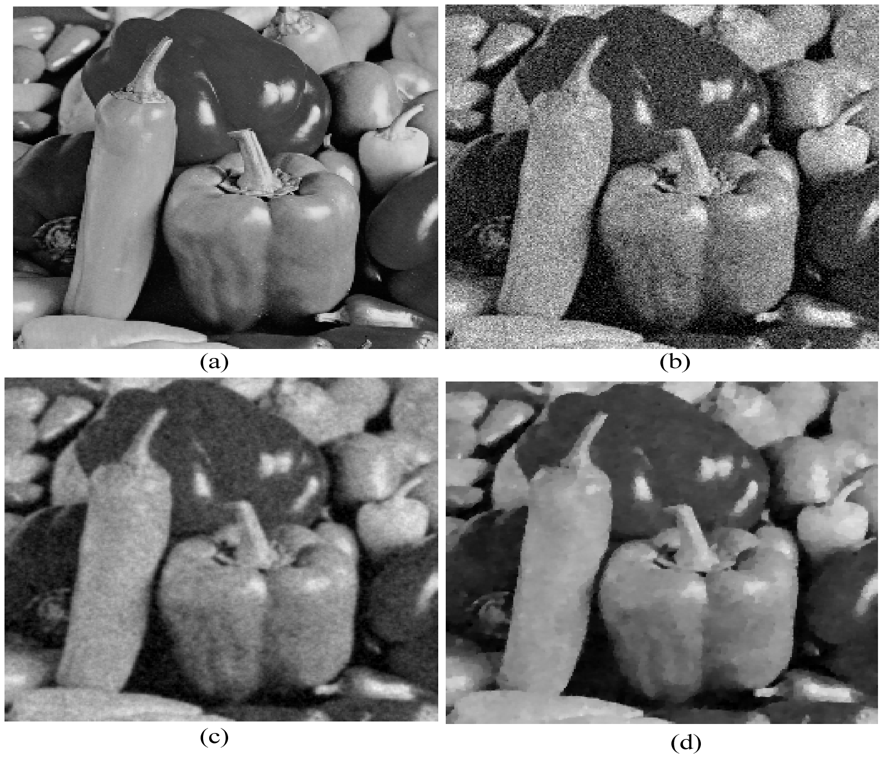

Figure 3. Visual Comparison of SURE-optimized Deblurred Results for Peppers. (a)Original image. (b) Degraded image. (c) Restored image by using Tikhonov regularization. (d) Restored image by using TV regularization.

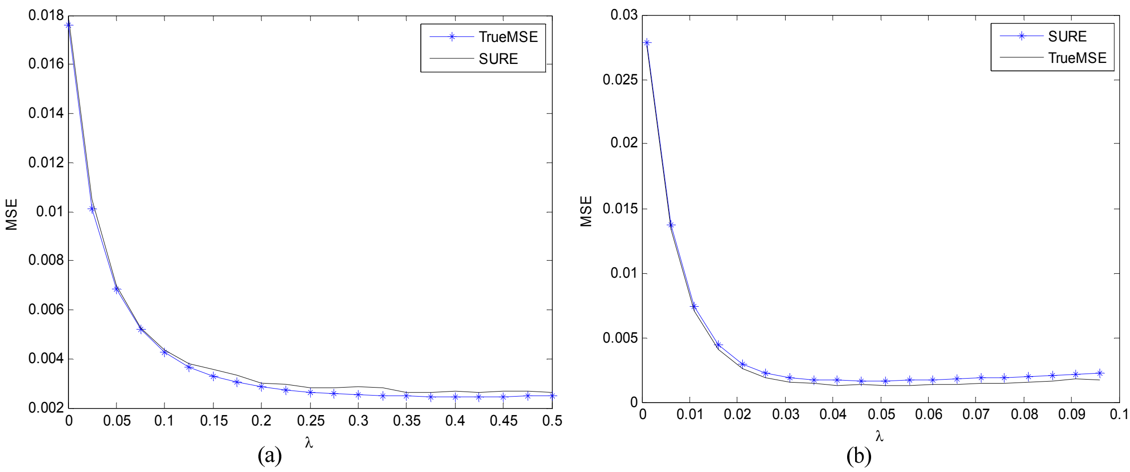

Figure 4. MSE(λ) and SURE(λ) for cameraman for peppers. (a) MSE(λ) and SURE(λ) based on Tikhonov regularization. (b) MSE(λ) and SURE(λ) based on TV regularization.

Table 1. Comparison of MSE and SURE in Terms of Output PSNR (dB)

Regularization method	Image	Input PSNR	Output PSNR based on MSE	Output PSNR based on SURE
Tikhonov	Cameraman	19.7608	23.7006	23.7006
	Peppers	19.9057	26.0399	26.0122
TV	Cameraman	19.7608	24.9044	24.9044
	Peppers	19.9057	27.8045	27.7497

Acknowledgement

This work is partially supported by the National Natural Science Foundation of China (11273017).

References

[1] Hadamard J. Lectures on on Cauchy's Problem in in Linear Partial Differential Equations. New Haven: Yale University Press; 1923.
[2] Eldar YC. Generalized SURE for Exponential Families: Applications to Regularization. *IEEE Trans. on Signal Processing*. 2009; 57(2): 471-481.
[3] Giryes R, Elad M, Eldar YC. The projected GSURE for automatic parameter tuning in iterative shrinkage methods. *Applied and Computational Harmonic Analysis*. 2011; 30(3): 407- 422.
[4] Ramani S, Blu T, Unser M. Monte-Carlo Sure: A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms. *IEEE Transactions on Image Processing*. 2008; 17(9): 1540-1554.
[5] Karl WC. *Regularization in image restoration and reconstruction*. in Handbook of Image and Video processing, A. Bovik, Ed., 2nd ed. New York: Elsevier. 2005: 183–202.
[6] Park SC, Park MK, Kang MG. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*. 2003; 20(3): 21–36.
[7] Girard DA. The fast Monte-Carlo Cross-Validation and CL procedures: comments, new results and application to image recovery problems. *Comput. Statist.*, 1995; 10: 205–231.
[8] Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*. 1979; 21: 215-223.
[9] Rice J. Choice of smoothing parameter in deconvolution problems. *Contemporary Math.*, 1986; 59: 137–151.
[10] Tikhonov AN, Goncharsky AV, Stepanov VV et al. Numerical Methods for the Solution of Ill-Posed Problems. Kluwer Academic Publishers. 1995.
[11] Tikhonov AN. Regularization of incorrectly posed problems. *Soviet athematical Doklady*. 1963; 4: 1624-1627.
[12] Young M. *The Technical Writer's Handbook*. Mill Valley, CA: University Science; 1989.
[13] Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*. 1992; 60(1): 259-268.
[14] Snyman JA. Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms. Springer Publishing; 2005.
[15] Kay S M. Fundamentals of Statistical Signal Processing: Estimation Theory, Upper Saddle River, NJ: Prentice Hall, Inc.; 1993.