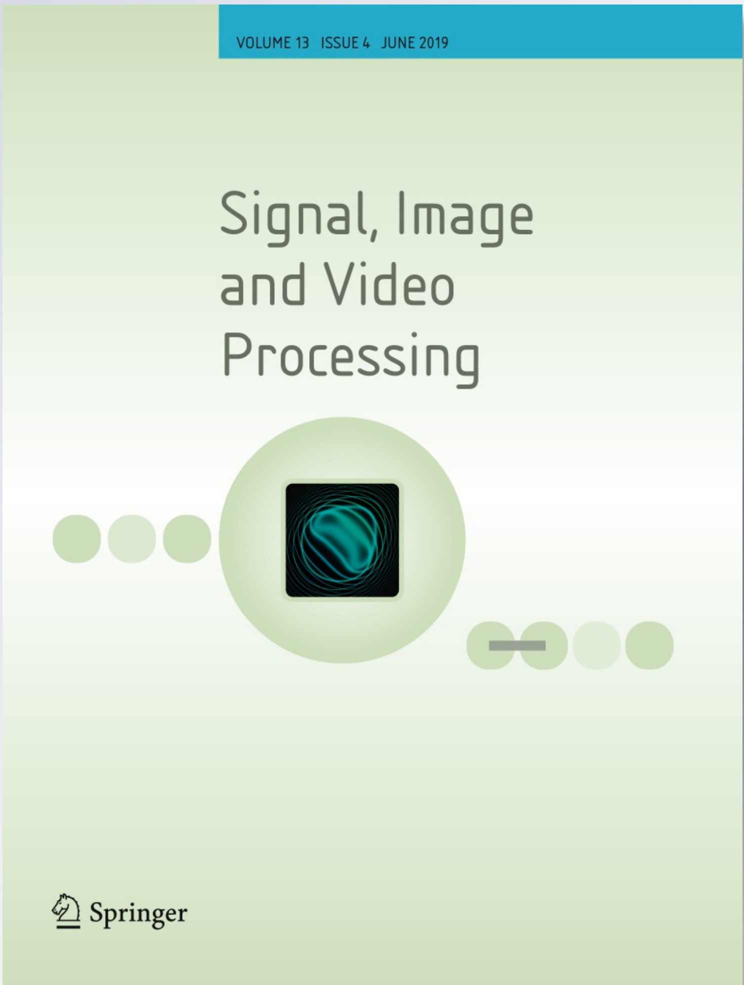
Parametric PSF estimation based on predicted-SURE with $$\ell _1$$ ℓ 1 -penalized sparse deconvolution

Feng Xue, Jiaqi Liu & Xia Ai

Signal, Image and Video Processing

ISSN 1863-1703
Volume 13
Number 4

SIViP (2019) 13:635-642
DOI 10.1007/s11760-018-1391-9



Springer

Signal, Image and Video Processing (2019) 13:635–642
https://doi.org/10.1007/s11760-018-1391-9

ORIGINAL PAPER

Parametric PSF estimation based on predicted-SURE with ℓ_1-penalized sparse deconvolution

Feng Xue[1] · Jiaqi Liu[1] · Xia Ai[1]

Received: 28 October 2017 / Revised: 23 October 2018 / Accepted: 8 November 2018 / Published online: 19 November 2018

Abstract
Point spread function (PSF) estimation plays an important role in blind image deconvolution. It has been shown that incorporating Wiener filter, minimization of the predicted Stein's unbiased risk estimate (p-SURE)—unbiased estimate of predicted mean squared error—could yield an accurate PSF estimate. In this paper, we provide a theoretical analysis for the PSF estimation error, which shows that the better deconvolution leads to more accurate PSF estimate. It motivates us to incorporate an ℓ_1-penalized sparse deconvolution into the p-SURE minimization, instead of the Wiener-type filtering. In particular, based on FISTA—one of the most popular iterative ℓ_1-solvers, we evaluate the p-SURE for each update, by Jacobian recursion and Monte Carlo simulation. Numerical results of both synthetic and real experiments demonstrate the improvements in PSF estimate, and therefore, deconvolution performance.

Keywords Blind deconvolution · Parametric PSF estimation · Predicted Stein's unbiased risk estimate (p-SURE) · ℓ_1-based sparse deconvolution · Fast iterative soft-thresholding algorithm (FISTA)

1 Introduction

Problem statement—In many applications, e.g., medical imaging [14], fluorescence microscopy [1], infrared imaging [30] and photography [24], the observed images are often degraded by the blurring effect and measurement noise (e.g., photon counting and readout noise) [1,30]. This image degradation is often mathematically expressed by the linear model [11,22,24,27]:

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

where $\mathbf{y} \in \mathbb{R}^N$ is the observed data of the original (unknown) $\mathbf{x}_0 \in \mathbb{R}^N$, $\mathbf{H}_0 \in \mathbb{R}^{N\times N}$ is a ground truth (unknown) convolution matrix constructed by point spread function (PSF) $\mathbf{h}_0$, the vector $\mathbf{b} \in \mathbb{R}^N$ is a zero-mean additive white Gaussian noise with variance σ^2. Blind image deconvolution attempts to estimate the original image $\mathbf{x}_0$, from the blurred image $\mathbf{y}$ only.

Related works—Blind deconvolution has been an important image processing topic for several decades [15]. Regularization is a standard technique to solve the ill-posed problem. It incorporates *a priori* information of the original image $\mathbf{x}_0$ and PSF $\mathbf{h}_0$ and formulates deconvolution process as a constrained optimization problem [13,24]. People may refer to [4,15] for a comprehensive review.

The key difficulty of this problem lies in the PSF estimation, since the deconvolution performance strongly depends on the accuracy of the PSF estimate [17]. In this paper, we focus on the parametric PSF estimation, where the PSF is of known parametric form with a small number of *unknown* parameters,[1] denoted by a parameter vector $\mathbf{s}$.

SURE-based approach [25]—In [25], we proved that incorporating the Wiener-type filtering (expressed in frequency domain):

$$\widehat{X}_{\mathbf{s}}(\omega) = \frac{H_{\mathbf{s}}^*(\omega)}{|H_{\mathbf{s}}(\omega)|^2 + \lambda\|\omega\|_2^2} \cdot Y(\omega) \tag{2}$$

This work was supported by the National Natural Science Foundation of China (No. 61401013).

✉ Feng Xue
fxue@hotmail.com

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

1 See [13,25] for a few parametric examples.

the minimization of predicted-MSE [25,31]:

$$\text{p-MSE}(\mathbf{s}) = \frac{1}{N}\left\|\mathbf{H_s}\widehat{\mathbf{x}}_\mathbf{s} - \mathbf{H}_0\mathbf{x}_0\right\|_2^2 \tag{3}$$

leads to very accurate estimate of PSF parameter $\mathbf{s}$, where $\mathbf{H_s}$ and $\widehat{\mathbf{x}}_\mathbf{s}$ are obtained using the tentative PSF parameters $\mathbf{s}$. Refer to Theorem 2.1 and Corollary 2.1 in [25] for the detailed proof.

Note that the p-MSE is not accessible in practice, due to the unknown $\mathbf{H}_0$ and $\mathbf{x}_0$. Hence, [23,25,31] proposed a predicted-SURE:

$$\text{p-SURE}(\mathbf{s}) = \frac{1}{N}\left\|\mathbf{H_s}\widehat{\mathbf{x}}_\mathbf{s} - \mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{N}\mathrm{Tr}\big(\mathbf{H_s}\mathbf{J_y}(\widehat{\mathbf{x}}_\mathbf{s})\big) - \sigma^2 \tag{4}$$

as a reliable substitute for p-MSE, which can be computed in practice. Here, N stands for the pixel number of the image, $\mathbf{y}$ is the blurred and noisy image, $\widehat{\mathbf{x}}_\mathbf{s}$ denotes the deconvolution estimate using $\mathbf{H_s}$. The Jacobian matrix $\mathbf{J_y}(\widehat{\mathbf{x}}_\mathbf{s})$ is defined as [25,31]:

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

See Theorem 3.1 of [25] for the proof of the unbiasedness of p-SURE w.r.t. p-MSE.

The extensive tests of [25] demonstrated the superior performance of p-SURE minimization (with Wiener-type filtering) to other methods, including GCV [20], kurtosis [12], DL1C [9] and APEX [7].

Our contributions—In this paper, we further develop the p-SURE framework by two contributions. First, we provide a theoretical analysis for the PSF estimation error, which shows that better deconvolution leads to more accurate PSF estimate. It motivates us to incorporate the ℓ_1-penalized sparse deconvolution into p-SURE, since the sparse deconvolution is generally better than Wiener-type filtering. Second, we propose a recursion of Jacobian matrix and Monte Carlo simulation, to facilitate the p-SURE minimization with the ℓ_1-estimate.

Additional remarks—Throughout this paper, we use boldface lowercase letters, e.g., $\mathbf{x} \in \mathbb{R}^N$, to denote N-dimensional real vectors. The linear (matrices) and nonlinear transformations $\mathbb{R}^N \to \mathbb{R}^M$ are denoted by boldface uppercase letters, e.g., $\mathbf{H} \in \mathbb{R}^{M\times N}$. $\mathbf{H}^\mathrm{T} \in \mathbb{R}^{N\times M}$ denotes the transpose of matrix $\mathbf{H}$. Also note that we use the subscript $(\cdot)_0$ to denote the true ("ground truth") quantity of $(\cdot)$; for example, matrix $\mathbf{H}_0$ is the true quantity of $\mathbf{H}$.

We always assume periodic boundary condition for linear filtering, which can be efficiently computed by discrete Fourier transform (DFT). We denote the frequency representation of convolution matrix $\mathbf{H}$ by $H(\omega)$. Denote the DFT coefficients of image data, say $\mathbf{x}$, by $X(\omega)$. In the presentation, we will switch repeatedly between frequency domain and matrix language, to emphasize in this way the close links between the two domains.

2 Error analysis of PSF estimation

2.1 The p-MSE minimization

The following theorem shows that the p-MSE minimization yields exact estimate of $\mathbf{H}_0$.

Theorem 1 *Minimizing the following p-MSE:*

$$\min_{\mathbf{H}} \left\|\mathbf{H}\mathbf{x}_0 - \mathbf{H}_0\mathbf{x}_0\right\|_2^2 \tag{5}$$

yields $H(\omega) = H_0(\omega)$ for $\forall\omega$.

Proof (5) is equivalent to the following problem expressed in frequency domain:

$$\sum_{\omega} \min_{H(\omega)} \left|H(\omega) - H_0(\omega)\right|^2 \cdot \left|X_0(\omega)\right|^2$$

which indicates that the minimizer is $H(\omega) = H_0(\omega)$ for $\forall\omega$, when the p-MSE vanishes. □

2.2 Error analysis for the PSF estimation

Since the exact $\mathbf{x}_0$ in (5) is unknown, we use the estimate $\widehat{\mathbf{x}}$ instead in practice. The following proposition states that the PSF estimation error is upper bounded by the estimation error of $\mathbf{x}_0$.

Proposition 1 *Considering the minimization of p-MSE over* $\mathbf{H}$*:* $\min_\mathbf{H} \|\mathbf{H}\widehat{\mathbf{x}} - \mathbf{H}_0\mathbf{x}_0\|_2^2$ *with some estimate* $\widehat{\mathbf{x}}$*, the estimation error of* $H_0(\omega)$ *satisfies:*

$$\left|H(\omega) - H_0(\omega)\right| \le C_1 \cdot \delta_\omega + C_2, \quad \forall\omega \tag{6}$$

where C_1 and C_2 are two constants, δ_ω denotes the estimation error of $\widehat{\mathbf{x}}$: $\left|\frac{X_0(\omega) - \widehat{X}(\omega)}{\widehat{X}(\omega)}\right| \le \delta_\omega$ for $\forall\omega$. Furthermore, (6) can be simplified for the following two special cases:

- *If $\widehat{X}(\omega) = X_0(\omega)$, then the constant $C_1 = 0$;*
- *For non-parametric setting of PSF, the constant $C_2 = 0$.*

Proof (upper-bound analysis)
The p-MSE can be expressed in frequency domain:

$$\min_H \underbrace{\sum_{\omega} \left|H(\omega)\widehat{X}(\omega) - H_0(\omega)X_0(\omega)\right|^2}_{J(H)} \tag{7}$$

- *Non-parametric case* For the non-parametric setting of PSF, $H(\omega)$ has the full degree of freedom. The solution to (7), i.e., the minimizer of $J(H)$, is obviously:

$$H^\star(\omega) = H_0(\omega)\frac{X_0(\omega)}{\widehat{X}(\omega)} = H_0(\omega)\Bigg(1 + \underbrace{\frac{X_0(\omega) - \widehat{X}(\omega)}{\widehat{X}(\omega)}}_{\Delta_1\omega}\Bigg) \quad (8)$$

where $\Delta_1\omega$ denotes the estimation error of $\widehat{X}(\omega)$ w.r.t. $X_0(\omega)$. And the minimum value of $J(H)$ is $J_1^\star = J(H^\star) = 0$.

1. Ideally, if $\widehat{X}(\omega) = X_0(\omega)$, such that $\Delta_1\omega = 0$, (8) leads to $H^\star(\omega) = H_0(\omega)$, which coincides with Theorem 1.
2. In practice, the estimate $\widehat{X}(\omega) \neq X_0(\omega)$, we obtain from (8):

$$\big|H^\star(\omega) - H_0(\omega)\big| \le \underbrace{\max_\omega\Big(|H_0(\omega)|\Big)}_{C^\star} \cdot \underbrace{\max_\omega\big|\Delta_1\omega\big|}_{\delta_\omega}, \quad \forall\omega$$

- *Parametric case* In a parametric setting of PSF, $H(\omega)$ is determined by a few PSF parameters $\mathbf{s}$, denoted by $H_\mathbf{s}(\omega)$. The problem (7) becomes:

$$\min_\mathbf{s} \underbrace{\sum_\omega \big|H_\mathbf{s}(\omega)\widehat{X}(\omega) - H_{\mathbf{s}_0}(\omega)X_0(\omega)\big|^2}_{J(\mathbf{s})} \quad (9)$$

where $\mathbf{s}_0$ denotes the true PSF parameter. Let $\mathbf{s}^\star$ denote the minimizer of $J(\mathbf{s})$.

1. Ideally, if $\widehat{X}(\omega) = X_0(\omega)$, (9) becomes:

$$\min_\mathbf{s} \sum_\omega \big|H_\mathbf{s}(\omega) - H_{\mathbf{s}_0}(\omega)\big|^2 \cdot \big|X_0(\omega)\big|^2 \quad (10)$$

   - If the retrieved $H_\mathbf{s}(\omega)$ has the same parametric form with the true $H_{\mathbf{s}_0}(\omega)$, (10) leads to $\mathbf{s}^\star = \mathbf{s}_0$ and the minimum value of $J(\mathbf{s})$ is $J_2^\star = J(\mathbf{s}^\star) = J(\mathbf{s}_0) = 0 = J_1^\star$. It coincides with Theorem 1.
   - If the retrieved $H_\mathbf{s}(\omega)$ and the true $H_{\mathbf{s}_0}(\omega)$ belong different parametric functions, due to the limitation of degree of freedom of $H_\mathbf{s}(\omega)$ (i.e., the small dimension of parameter vector $\mathbf{s}$, compared to the number of frequency samples ω), a small error $\Delta_2\omega$ should be allowed: $H_{\mathbf{s}^\star}(\omega) = H_{\mathbf{s}_0}(\omega) + \Delta_2\omega$, to guarantee the existence of optimal $\mathbf{s}^\star$,[2] and the residual error is:

$$\begin{aligned} J_2^\star &= J(\mathbf{s}^\star) = \sum_\omega |\Delta_2\omega|^2 \cdot |X_0(\omega)|^2 \\ &\ge \min_\omega |X_0(\omega)|^2 \cdot \sum_\omega |\Delta_2\omega|^2 > 0 = J_1^\star \end{aligned}$$

Finally, we obtain the following inequality:

$$\begin{aligned} \big|H_{\mathbf{s}^\star}(\omega) - H_{\mathbf{s}_0}(\omega)\big| &\le \max_\omega \big|\Delta_2(\omega)\big| \le \sqrt{\sum_\omega |\Delta_2\omega|^2} \\ &\le \sqrt{\frac{J_2^\star}{\min_\omega |X_0(\omega)|^2}}, \quad \forall\omega \end{aligned}$$

2. In practice, the estimate $\widehat{X}(\omega) \neq X_0(\omega)$. No matter if the retrieved $H_\mathbf{s}(\omega)$ and the true $H_{\mathbf{s}_0}(\omega)$ belong to the same function family, the minimization (9) always yields:

$$H_{s^\star}(\omega) = H_0(\omega)\Bigg(1 + \underbrace{\frac{X_0(\omega) - \widehat{X}(\omega)}{\widehat{X}(\omega)}}_{\Delta_1\omega}\Bigg) + \Delta_2\omega; \quad \forall\omega$$

and

$$\begin{aligned} J_2^\star &= J(\mathbf{s}^\star) = \sum_\omega |\Delta_2\omega|^2 \cdot |X_0(\omega)|^2 \\ &\ge \min_\omega |X_0(\omega)|^2 \cdot \sum_\omega |\Delta_2\omega|^2 > 0 = J_1^\star \end{aligned}$$

where a small error $\Delta_2\omega$ should be allowed for the existence of optimal $\mathbf{s}^\star$. Finally, we obtain the following inequality:

$$\begin{aligned} \big|H_{\mathbf{s}^\star}(\omega) - H_0(\omega)\big| &= \big|H_0(\omega)\cdot\Delta_1\omega + \Delta_2\omega\big| \\ &\le \big|H_0(\omega)\cdot\Delta_1\omega\big| + \big|\Delta_2\omega\big| \\ &\le \underbrace{\max_\omega\big(|H_0(\omega)|\big)}_{C^\star} \cdot \underbrace{\max_\omega|\Delta_1\omega|}_{\delta_\omega} + \max_\omega|\Delta_2\omega| \\ &\le C^\star \cdot \delta_\omega + \sqrt{\frac{J_2^\star}{\min_\omega |X_0(\omega)|^2}}, \quad \forall\omega \end{aligned}$$

Summarizing all the cases leads to the inequality (6), where either C_1 or C_2 vanishes for some special cases. □

The above theorem implies that the more accurate estimate of $\mathbf{x}_0$ produces better estimate of $\mathbf{H}_0$. Therefore, it is important to find better estimate of $\mathbf{x}_0$.

3 The p-SURE minimization with ℓ_1-penalized sparse deconvolution

3.1 The proposed formulation

In [25], we use Wiener-type filtering as the estimate $\widehat{\mathbf{x}}$. To improve the accuracy of PSF estimation, we now propose to use the wavelet-based sparse deconvolution in p-MSE/SURE, since it is generally better than Wiener filtering in terms of deconvolution performance [2,16].

Given the fixed tentative parameters $\mathbf{s}$ and λ, the sparse deconvolution is often formulated by the following typical ℓ_1-based *synthesis formulation* [3,18]:

$$\widehat{\mathbf{c}}_{\mathbf{s},\lambda} = \arg\min_\mathbf{c} \frac{1}{2}\big\|\mathbf{y} - \mathbf{H}_\mathbf{s}\mathbf{R}\mathbf{c}\big\|_2^2 + \lambda\big\|\mathbf{c}\big\|_1 \quad (11)$$

[2] The optimal solution $\mathbf{s}^\star$ may not be unique. The uniqueness of the solution depends on the specific parametric form of PSF.

where $\mathbf{R}$ denotes 2-D wavelet reconstruction, $\widehat{\mathbf{c}}_{\mathbf{s},\lambda}$ is the wavelet coefficients. Then, the ℓ_1-estimated image is given as $\widehat{\mathbf{x}}_{\mathbf{s},\lambda} = \mathbf{R}\widehat{\mathbf{c}}_{\mathbf{s},\lambda}$.

We incorporate this ℓ_1-estimate into p-SURE and formulate the PSF parameter estimation as:

$$\min_{\mathbf{s},\lambda} \underbrace{\frac{1}{N}\left\|\mathbf{H}_{\mathbf{s}}\mathbf{R}\widehat{\mathbf{c}}_{\lambda,\mathbf{s}} - \mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\mathrm{Tr}\big(\mathbf{H}_{\mathbf{s}}\mathbf{R}\mathbf{J}_{\mathbf{y}}(\widehat{\mathbf{c}}_{\lambda,\mathbf{s}}) - \sigma^2(\mathbf{s},\lambda)}_{\text{p-SURE}} \tag{12}$$

3.2 Recursive evaluation of p-SURE for FISTA

To solve (12), we need to compute the solution $\widehat{\mathbf{c}}_{\mathbf{s},\lambda}$, its Jacobian matrix and p-SURE. First, to find the solution $\widehat{\mathbf{c}}_{\mathbf{s},\lambda}$, many optimization algorithms can be used for solving (11), e.g., [8,10,19,21]. In this paper, we choose FISTA—one of the most popular ℓ_1-solvers. The iteration scheme of FISTA is [3] (ignoring the subscripts $\mathbf{s}$ and λ for brevity in this subsection):

$$\begin{cases} \mathbf{c}^{(i)} &= \mathcal{T}_{\lambda/L}\Big(\mathbf{z}^{(i)} - \frac{1}{L}\mathbf{D}\mathbf{H}^{\mathrm{T}}\big(\mathbf{H}\mathbf{R}\mathbf{z}^{(i)} - \mathbf{y}\big)\Big) \\ t^{(i+1)} &= \frac{1+\sqrt{1+4(t^{(i)})^2}}{2} \\ \mathbf{z}^{(i+1)} &= \mathbf{c}^{(i)} + \nu^{(i)}\big(\mathbf{c}^{(i)} - \mathbf{c}^{(i-1)}\big) \end{cases} \tag{13}$$

where $\mathcal{T}_{\lambda/L}(\cdot)$ denotes the soft-thresholding function with the threshold λ/L, $\nu^{(i)} = \frac{t^{(i)}-1}{t^{(i+1)}}$, $\mathbf{D}$ denotes wavelet decomposition, L is a Lipschitz continuous constant. In particular, for orthonormal transform, it is easy to show that $L = 1$, if the convolution kernel $\mathbf{h}$ is normalized [3]. In the remainder of this paper, we use $L = 1$.

Regarding the Jacobian matrix of $\widehat{\mathbf{c}}$, we now propose a recursive evaluation based on FISTA, i.e., computing $\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})$ and p-SURE of $\mathbf{c}^{(i)}$ by

$$\text{p-SURE} = \frac{1}{N}\left\|\mathbf{H}\mathbf{R}\mathbf{c}^{(i)} - \mathbf{y}\right\|_2^2 + \frac{2\sigma^2}{N}\mathrm{Tr}\big(\mathbf{H}\mathbf{R}\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})\big) - \sigma^2 \tag{14}$$

during each iterate, until convergence.

Rewriting (13) as ($L = 1$):

$$\mathbf{c}^{(i)} = \mathcal{T}_{\lambda}\Big(\underbrace{\overbrace{\big(\mathbf{I} - \mathbf{D}\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{R}\big)}^{\mathbf{B}}\mathbf{z}^{(i)} + \mathbf{D}\mathbf{H}^{\mathrm{T}}\mathbf{y}}_{\mathbf{u}^{(i)}}\Big)$$

by the basic calculus and the property of Jacobian matrix [31], we obtain:

$$\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i+1)}) = \mathbf{P}^{(i)}\big(\mathbf{B}\mathbf{J}_{\mathbf{y}}(\mathbf{z}^{(i)}) + \mathbf{D}\mathbf{H}^{\mathrm{T}}\big)$$

Here, the matrix $\mathbf{P}^{(i)}$ is diagonal with the diagonal elements given as:

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

The Jacobian matrix of $\mathbf{z}^{(i+1)}$ is:

$$\mathbf{J}_{\mathbf{y}}(\mathbf{z}^{(i+1)}) = \mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)}) + \nu^{(i)}\Big(\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)}) - \mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i-1)})\Big)$$

Finally, the recursions of Jacobian matrices for the FISTA update (13) are:

$$\begin{cases} \mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)}) &= \mathbf{P}^{(i)}\big(\mathbf{B}\mathbf{J}_{\mathbf{y}}(\mathbf{z}^{(i)}) + \mathbf{D}\mathbf{H}^{\mathrm{T}}\big) \\ \mathbf{J}_{\mathbf{y}}(\mathbf{z}^{(i+1)}) &= \mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)}) + \nu^{(i)}\big(\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)}) - \mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i-1)})\big) \end{cases} \tag{15}$$

3.3 The computational issues

We can see that the key computational issue of the p-SURE lies in Jacobian matrices and matrix trace. However, if both $\mathbf{x}$ and $\mathbf{y}$ are the typical 256×256 grayscale images, the dimensions of $\mathbf{H}$, $\mathbf{R}$ and $\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})$ will be $65{,}536 \times 65{,}536$. It is generally difficult to explicitly store and compute such large matrices, due to the limited computational resources.

In [5], the authors applied the Monte Carlo (MC) simulation to compute the trace of matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, based on the fact that $\mathrm{Tr}(\mathbf{A}) = \mathbb{E}\{\mathbf{n}_0^{\mathrm{T}}\mathbf{A}\mathbf{n}_0\}$, for a random vector $\mathbf{n}_0 \sim \mathcal{N}(0, \mathbf{I}_N)$, if $\mathbf{A}\mathbf{n}_0$ can be computed without explicit expression of the matrix $\mathbf{A}$. The expectation can be approximately computed by averaging a number of random realizations of $\mathbf{n}_0$ (Q realizations in this paper). The procedure is shown in Fig. 1.

Now, we use this procedure to compute the p-SURE for FISTA. For the input noise $\mathbf{n}_0 \sim \mathcal{N}(0, \mathbf{I}_N)$, multiplying $\mathbf{n}_0$ on both sides of (15), we obtain:

$$\begin{cases} \underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{c}}^{(i)}} &= \mathbf{P}^{(i)}\mathbf{B}\underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{z}^{(i)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{z}}^{(i)}} + \mathbf{P}^{(i)}\underbrace{\mathbf{D}\mathbf{H}^{\mathrm{T}}\mathbf{n}_0}_{\mathbf{n}_1} \\ \underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{z}^{(i+1)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{z}}^{(i+1)}} &= (1+\nu^{(i)})\underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{c}}^{(i)}} - \nu^{(i)}\underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i-1)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{c}}^{(i-1)}} \end{cases} \tag{16}$$

Then, the trace of term of p-SURE becomes:

$$\mathrm{Tr}(\mathbf{H}\mathbf{R}\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})) = \mathbb{E}\big\{\mathbf{n}_0^{\mathrm{T}}\mathbf{H}\mathbf{R}\underbrace{\mathbf{J}_{\mathbf{y}}(\mathbf{c}^{(i)})\mathbf{n}_0}_{\mathbf{n}_{\mathbf{c}}^{(i)}}\big\} \tag{17}$$

We can see that the recursion of Jacobian matrix boils down to a simple evolution of the random noise $\mathbf{n}_{\mathbf{c}}^{(i)}$ and $\mathbf{n}_{\mathbf{z}}^{(i)}$. In addition, we have three remarks to facilitate the MC-based computations of (16): for any input vector $\mathbf{v}$,

- the outcome of $\mathbf{R}\mathbf{v}$, $\mathbf{H}\mathbf{v}$ and the combinations of the operations can be efficiently computed by the wavelet and Fourier transforms.

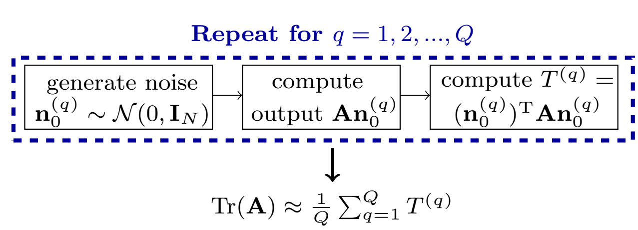


Fig. 1 MC simulation to evaluate matrix trace

- it is easy to compute $\mathbf{P}^{(i)}\mathbf{v}$ for diagonal matrix $\mathbf{P}^{(i)}$ by the point-wise product $[\mathbf{P}^{(i)}\mathbf{v}]_n = \mathbf{P}^{(i)}_{n,n}v_n$ for $n = 1, 2, \ldots, N$.
- $\mathbf{J}_\mathbf{y}(\mathbf{c}^{(i)})\mathbf{v}$ and $\mathbf{J}_\mathbf{y}(\mathbf{z}^{(i)})\mathbf{v}$ should be computed without explicit expressions of the Jacobian matrices. To this end, $\mathbf{c}^{(0)}$ and $\mathbf{z}^{(0)}$ can be initialized as $\mathbf{c}^{(0)} = \mathbf{z}^{(0)} = \mathbf{Dy}$, such that $\mathbf{J}_\mathbf{y}(\mathbf{c}^{(0)}) = \mathbf{J}_\mathbf{y}(\mathbf{z}^{(0)}) = \mathbf{D}$, which can be performed by wavelet transform.

Thus, all the noise evolutions in (16) can be performed without explicit matrix computation: the MC simulation enables us to avoid the explicit Jacobian recursion of (15). The procedure is depicted in Fig. 2.

3.4 Summary

Now we are able to evaluate p-SURE for the ℓ_1-estimate $\widehat{\mathbf{x}}_\lambda$ with a certain value of λ. It is then incorporated into the p-SURE minimization for the PSF estimation, whose procedure is shown in Fig. 3.

To minimize (12), we suggest to first find the optimal λ with each tentative fixed $\mathbf{s}$, and then, perform the exhaustive search of $\mathbf{s}$ with the corresponding optimal λ.

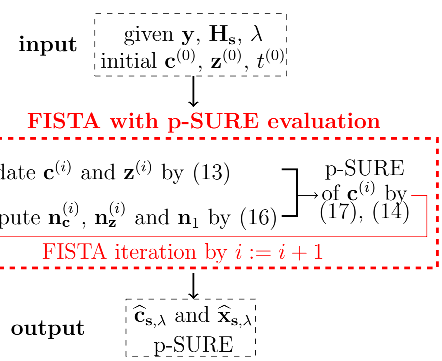


Fig. 2 p-SURE-MC evaluation for FISTA

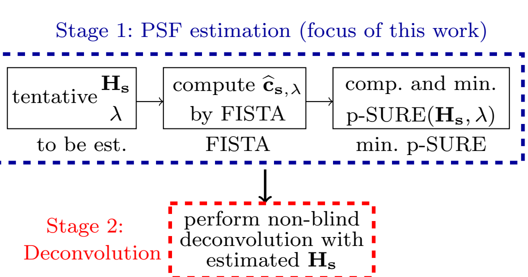


Fig. 3 The flowchart of PSF estimation: joint minimization of the p-SURE over $\mathbf{H}_\mathbf{s}$ and λ, as shown in (12), where the value of p-SURE is obtained by Fig. 2

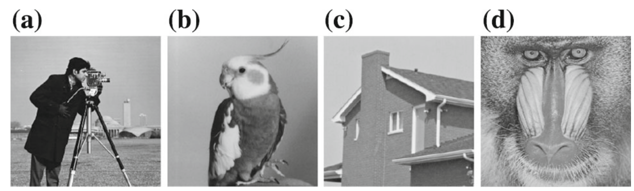


Fig. 4 Original test images: **a** *Cameraman* 256×256; **b** *Coco* 256×256; **c** *House* 256×256; **d** *Mandrill* 512×512

4 Experimental results and discussion

4.1 Experimental setting

The test dataset contains four 8-bit images of size 256×256 or 512×512 displayed in Fig. 4, covering a wide range of natural images.

We exemplify the proposed approach with two typical parametrized PSF's, which have been frequently used many practical applications:

- **Gaussian** kernel, with an unknown parameter—blur variance s^2 [25,30]:

$$h(i, j; s) = K \cdot \exp\left(-\frac{i^2 + j^2}{2s^2}\right) \qquad (18)$$

- *jinc* function, with an unknown scaling factor s [6]:

$$h(i, j; s) = K \cdot \left[\frac{2J_1(r/s)}{r/s}\right]^2 \qquad (19)$$

where $J_1(\cdot)$ is first-order Bessel function of first kind, the radius $r = \sqrt{i^2 + j^2}$. The constant K in (18) and (19) is a normalization factor, s.t. $\sum_{i,j} h(i, j) = 1$.

The Gaussian function is often used to model or approximate many blur kernels [25,30]. The *jinc* function describes the typical imaging pattern by optical diffraction, which often appears in optical apparatus, e.g., microscope and telescope [6]. The blur size s in (18) and (19) (which controls the blurring degree) is the unknown PSF parameter to be estimated.

Now, we perform the following synthetic experiments: the test images shown in Fig. 4 are blurred by Gaussian kernel (18) or *jinc* function (19) with the true value $s_0 = 2.0$, and corrupted by white Gaussian noise with various variances corresponding to blur signal-to-noise ratio (BSNR) of 40, 30, 20, 10 dB, which is defined as (in dB) [25]:

$$\mathrm{BSNR} = 10 \times \log_{10}\left(\frac{\|\mathbf{H}_0\mathbf{x}_0 - \mathrm{mean}(\mathbf{H}_0\mathbf{x}_0)\|_2^2}{N\sigma^2}\right)$$

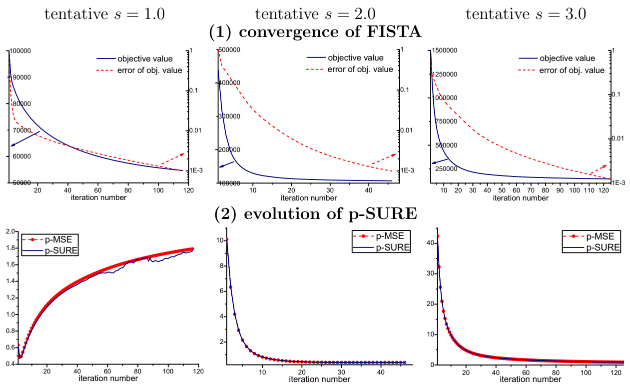


Fig. 5 Example: *Cameraman*, **Gaussian** with $s_0 = 2.0$, BSNR = 30 dB, fixed $\lambda = 0.01$

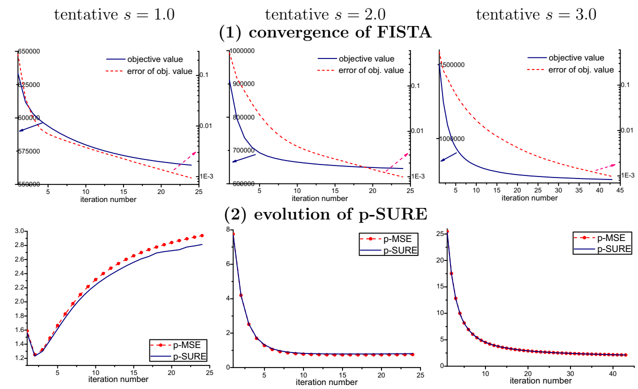


Fig. 6 Example: *Coco*, *jinc* with $s_0 = 2.0$, BSNR = 20 dB, fixed $\lambda = 0.1$

4.2 The convergence of FISTA for fixed λ and tentative s

First, we need to verify the proposed recursive evaluation of p-SURE for the ℓ_1-penalized estimate (shown in Fig. 2). For any tentative $\mathbf{H_s}$ and fixed λ, FISTA attempts to minimize

$$\mathcal{L}(\mathbf{c}^{(i)}) = \frac{1}{2}\|\mathbf{H}_s\mathbf{R}\mathbf{c}^{(i)} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{c}^{(i)}\|_1$$

We terminate the iteration by the stopping criterion $e^{(i)} = \frac{|\mathcal{L}(\mathbf{c}^{(i+1)} - \mathbf{c}^{(i)})|}{\mathcal{L}(\mathbf{c}^{(i)})} \leq 10^{-3}$.

Figures 5 and 6 show the convergence of FISTA for tentative $\mathbf{H_s}$ and fixed λ, and the variation of p-SURE during the iterations. The meaning of the vertical axis is given by the legend on the corner of each sub-figure. We can see that: (1) the objective value $\mathcal{L}(\mathbf{c}^{(i)})$ keeps decreasing until convergence; (2) p-SURE is always very close to the p-MSE during the iterations, which demonstrates that the p-SURE is a reliable estimate of p-MSE.

We also observe that in the left-down sub-figures of Figs. 5 and 6, the p-MSE (and the p-SURE) starts to increase after a few iterations. It is mainly because the FISTA aims at minimizing the objective functional (11), rather than the predicted error w.r.t. the true image $\mathbf{x}_0$. Hence, the objective value $\mathcal{L}(\mathbf{c}^{(i)})$ is monotonically decreasing (shown in top-left sub-figures); however, it is not guaranteed for the predicted error. That is to say, the FISTA only seeks the minimizer of the given functional, not a minimizer of the predicted error.[3]

[3] This similar phenomenon is also frequently encountered in non-blind sparse deconvolution (where $\mathbf{H}$ is exactly known), especially when the regularization parameter λ is very small. Refer to Fig. 2-(1) of [26], Fig. 2-(2) of [29] and Fig. 2-(2) of [28] for the typical examples with ℓ_1 or total variation as regularizer, we can also see that the predicted error increases after a few iterations, since the iterative algorithm used aims at minimizing a given functional, not the predicted or reconstructed error. That is also why there is generally no need to set a very strict stopping criterion for convergence in practice, since the exact solution at the convergence usually does not have the good restoration quality, especially when λ is small [3,18].

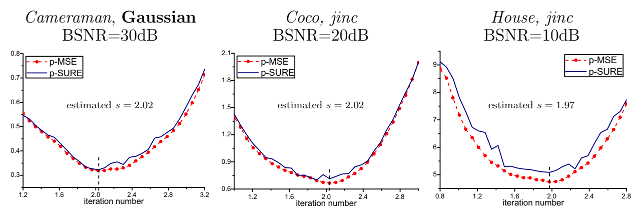


Fig. 7 Three examples for PSF estimation by p-SURE minimization with ℓ_1-estimate (the true $s_0 = 2.0$)

4.3 PSF estimation by p-SURE minimization with ℓ_1-estimate

The proposed method of Fig. 2 has been verified and can be safely incorporated into the PSF estimation of Fig. 3. Figure 7 shows three cases of p-SURE minimization with ℓ_1-estimate. We can see that: (1) the proposed method yields very accurate PSF estimate; (2) p-SURE is always very close to p-MSE for any cases.

Now, we are going to compare this method to our previous work [25]—the p-SURE minimization with Wiener filtering. It should be noted that there is no need to present the results of other PSF estimation methods, including GCV [20], kurtosis [12], DL1C [9] and APEX [7], since the extensive tests have shown that they are inferior to our previous work [25] in terms of estimation accuracy.

Table 1 reports the PSF estimation results and presents the comparisons under various images and noise levels. The proposed approach is denoted by 'ℓ_1-est.', whereas the previous work of [25] is denoted by 'Wiener'. We can see that the estimated parameters by the ℓ_1-estimate are very close to the true values s_0, and outperforms the previous work using Wiener filtering [25] in average.

After the PSF estimation, we apply the SURE-LET algorithm to perform non-blind deconvolution, using the estimated PSF [27]. Figure 8 shows four visual examples, which demonstrate the negligible PSNR loss of the blind restoration, compared to exact deconvolution.

Table 1 The estimated parameters s of **Gaussian** and *jinc* ($s_0 = 2.0$)

Blur kernel		**Gaussian** with $s_0 = 2.0$				*jinc* with $s_0 = 2.0$			
BSNR (in dB)		40	30	20	10	40	30	20	10
C–man	*Wiener*	2.06	1.96	1.97	2.09	1.97	2.01	2.06	2.06
	ℓ_1*-est.*	2.03	1.97	1.98	2.04	1.97	1.99	2.02	2.02
House	*Wiener*	2.21	2.11	2.24	2.43	2.01	2.06	2.14	2.09
	ℓ_1*-est.*	2.03	2.04	2.03	2.06	2.01	2.02	2.04	2.06
Coco	Wiener	2.14	2.12	2.17	2.32	2.01	2.05	2.09	2.22
	ℓ_1*-est.*	2.04	2.05	2.03	2.08	2.01	2.04	2.04	2.08
Mandrill	*Wiener*	1.97	1.94	1.93	1.86	1.99	1.99	1.97	1.93
	ℓ_1*-est.*	2.02	2.01	1.97	1.90	1.99	2.00	2.01	2.03

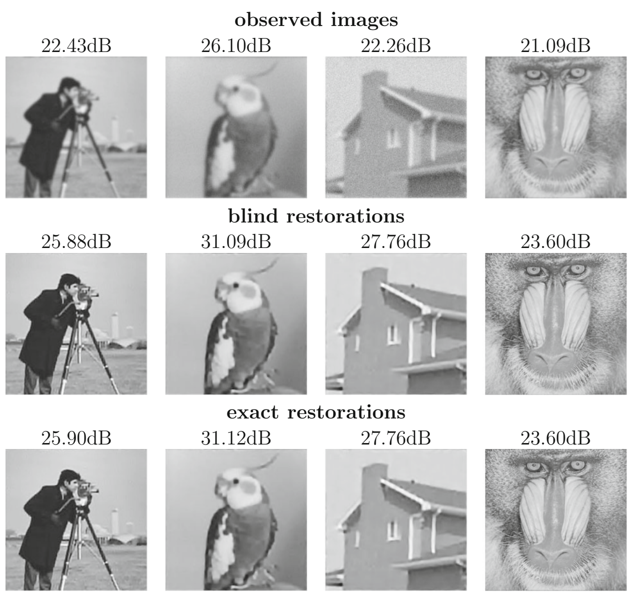


Fig. 8 Visual examples of blind deconvolution: (1) *Cameraman*, **Gaussian**, BSNR = 30 dB; (2) *Coco*, *jinc*, BSNR = 20 dB; (3) *House*, *jinc*, BSNR = 10 dB; (4) *Mandrill*, **Gaussian**, BSNR = 40 dB

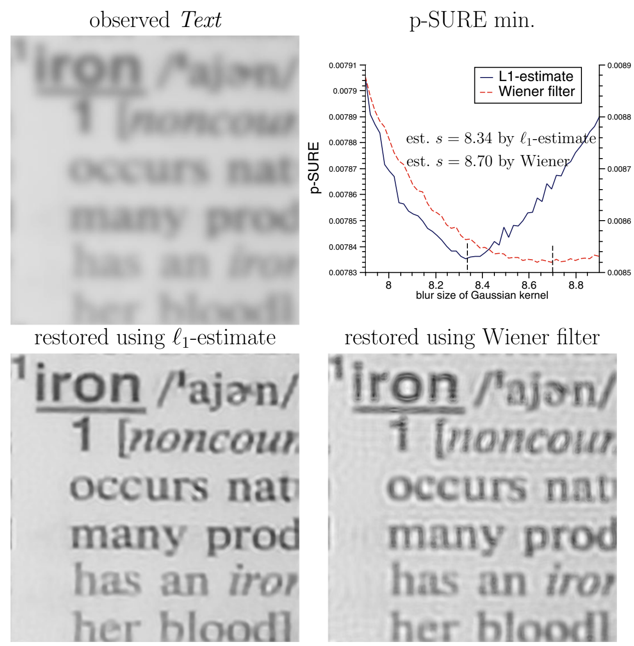


Fig. 9 Restoration of *Text* by the proposed method: the estimated **Gaussian** blur size is 8.34 by ℓ_1-based estimate and 8.70 by Wiener filtering [25]

4.4 Application to real image

In the last set of experiments, we apply the proposed method to real images: *Text* and *Fruit*, shown in Figs. 9 and 10. They were captured by a digital camera, blurred due to out-of-focus. There are no exact expressions of the PSF. We assume the underlying (unknown) PSF as **Gaussian** for *Text* and *jinc* function for *Fruit*, estimate the blur size s, and perform SURE-LET deconvolution using the estimated PSF [27].

Figures 9 and 10 show the variations of p-SURE with the blur size s, using ℓ_1-estimate and Wiener filtering [25], respectively.[4] We can see that the p-SURE with Wiener filtering [25] tends to produce larger blur size s than using ℓ_1-estimate. Thus, the restored images by [25] exhibit more ringing effect in the homogeneous regions and edges due to the over-deconvolution (see the smooth regions in *Fruit*, the background of *Text*, edges around the black characters in *Text*). As a comparison, the restored images by ℓ_1-estimate have better visual quality.

Notice that for fair comparison, both methods use the same SURE-LET deconvolution [27]: they only differ in the estimated PSF. Therefore, it is easy to recognize that the improvement in the restoration performance is mainly due to the more accurate PSF estimation, though the true PSF is unknown.

5 Conclusions

In this paper, we incorporated the ℓ_1-penalized sparse deconvolution into p-SURE, which has been demonstrated to yield more accurate PSF estimate, compared to using Wiener-type filter [25]. The Jacobian recursion and MC simulation developed in this paper can be, in principle, extended to other regularizers and iterative algorithms for the SURE

[4] We cannot compute p-MSE, since the original image $\mathbf{x}_0$ and true PSF $\mathbf{H}_0$ are unknown in the real experiments.

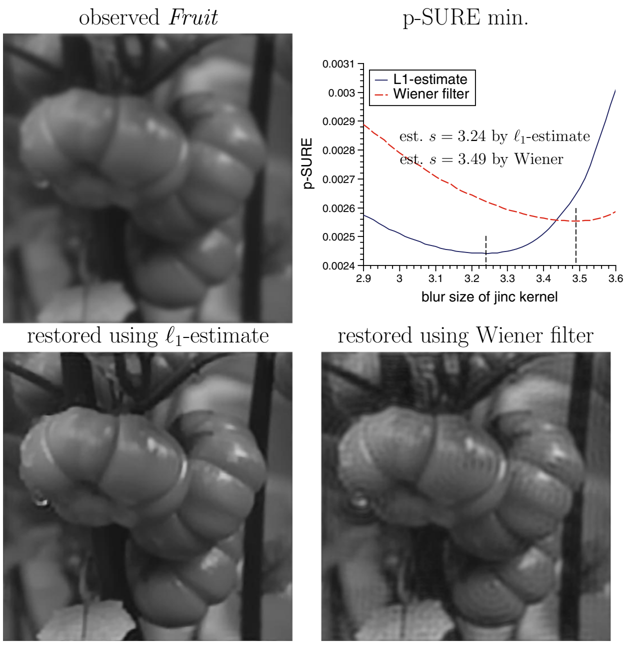


Fig. 10 Restoration of *Fruit* by the proposed method: the estimated *jinc* blur size is 3.24 by ℓ_1-based estimate and 3.49 Wiener filtering [25]

evaluation. There is also a huge potential to develop specific algorithms for various applications, e.g., fluorescence microscopy [13], based on this criterion consisting of sparse estimate.

References

1. Bal, U., Engin, M., Utzinger, U.: A multiresolution approach for enhancement and denoising of microscopy images. Signal Image Video Process. **9**, 787–799 (2015)
2. Baloch, G., Ozkaramanli, H.: Image denoising via correlation-based sparse representation. Signal Image Video Process. **11**, 1501–1508 (2017)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
4. Bishop, T., Babacan, S., Amizic, B., Katsaggelos, A., Chan, T., Molina, R.: Blind image deconvolution: problem formulation and existing approaches. Blind Image Deconvolution Theory & Applications (2007)
5. Blu, T., Luisier, F.: The SURE-LET approach to image denoising. IEEE Trans. Image Process. **16**(11), 2778–2786 (2007)
6. Born, M., Wolf, E., Bhatia, A.: Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. Cambridge University Press, Cambridge (1999)
7. Carasso, A.: The APEX method in image sharpening and the use of low exponent Lévy stable laws. SIAM J. Appl. Math. **63**(2), 593–618 (2003)
8. Chen, D.: Inext alternating direction method based on Newton descent algorithm with application to Poisson image deblurring. Signal Image Video Process. **11**, 89–96 (2017)
9. Chen, F., Ma, J.: An empirical identification method of Gaussian blur parameter for image deblurring. IEEE Trans. Signal Process. **57**(7), 2467–2478 (2009)
10. Dell'Acqua, P.: ν acceleration of statistical iterative methods for image restoration. Signal Image Video Process. **10**, 927–934 (2016)
11. Kerouh, F., Serir, A.: Wavelet-based blind blur reduction. Signal Image Video Process. **9**, 1587–1599 (2015)
12. Li, D., Simske, S.: Atmospheric turbulence degraded-image restoration by kurtosis minimization. IEEE Geosci. Remote Sens. Lett. **6**(2), 244–247 (2009)
13. Markham, J., Conchello, J.: Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur. J. Opt. Soc. Am. A **16**(10), 2377–2391 (1999)
14. Michailovich, O., Tannenbaum, A.: Blind deconvolution of medical ultrasound images: a parametric inverse filtering approach. IEEE Trans. Image Process. **16**(12), 3005–3019 (2007)
15. Molina, R., Núñez, J., Cortijo, F., Mateos, J.: Image restoration in astronomy: a Bayesian perspective. IEEE Signal Process. Mag. **18**(2), 11–29 (2001)
16. Nazzal, M., Ozkaramanli, H.: Wavelet domain dictionary learning-based single image super-resolution. Signal Image Video Process. **9**, 1491–1501 (2015)
17. Oliveira, J., Figueiredo, M., Bioucas-Dias, J.: Parametric blur estimation for blind restoration of natural images: linear motion and out-of-focus. IEEE Trans. Image Process. **23**(1), 466–477 (2014)
18. Pan, H., Blu, T.: An iterative linear expansion of thresholds for ℓ_1-based image restoration. IEEE Trans. Image Process. **22**(9), 3715–3728 (2013)
19. Ramirez, C., Argaez, M.: An ℓ_1 minimization algorithm for non-smooth regularization in image processing. Signal Image Video Process. **9**, 373–386 (2015)
20. Reeves, S.J., Mersereau, R.M.: Blur identification by the method of generalized cross-validation. IEEE Trans. Image Process. **1**(3), 301–311 (1992)
21. Shi, Y., Chang, Q., Yang, X.: A robust and fast combination algorithm for deblurring and denoising. Signal Image Video Process. **9**, 865–874 (2015)
22. Siadat, M., Aghazadeh, N., Öktem, O.: Reordering for improving global Arnoldi–Tikhonov method in image restoration problems. Signal Image Video Process. **12**, 497–504 (2018)
23. Stein, C.: Estimation of the mean of a multivariate normal distribution. Ann. Stat. **9**(6), 1135–1151 (1981)
24. Welk, M., Raudaschl, P., Schwarzbauer, T., Erler, M., Läuter, M.: Fast and robust linear motion deblurring. Signal Image Video Process. **9**, 1221–1234 (2015)
25. Xue, F., Blu, T.: A novel SURE-based criterion for parametric PSF estimation. IEEE Trans. Image Process. **24**(2), 595–607 (2015)
26. Xue, F., Blu, T., Liu, J., Ai, X.: Recursive evaluation of SURE for total variation denoising. In: Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1338–1342 (2018)
27. Xue, F., Luisier, F., Blu, T.: Multi-Wiener SURE-LET deconvolution. IEEE Trans. Image Process. **22**(5), 1954–1968 (2013)
28. Xue, F., Pan, H., Liu, X., Liu, H., Liu, J.: Optimization of regularization parameter for sparse reconstruction based on predictive risk estimate. In: Proceedings of 2017 IEEE International Conference on Image Processing (ICIP), pp. 1442–1446 (2017)
29. Xue, F., Pan, H., Wu, R., Liu, X., Liu, J.: Optimization of compound regularization parameters based on Stein's unbiased risk estimate. In: Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4591–4595 (2017)
30. Xue, F., Yagola, A.G.: Analysis of point-target detection performance based on ATF and TSF. Infrared Phys. Technol. **52**(5), 166–173 (2009)
31. Xue, F., Yagola, A.G., Liu, J., Meng, G.: Recursive SURE for iterative reweighted least square algorithms. Inverse Probl. Sci. Eng. **24**(4), 625–646 (2016)