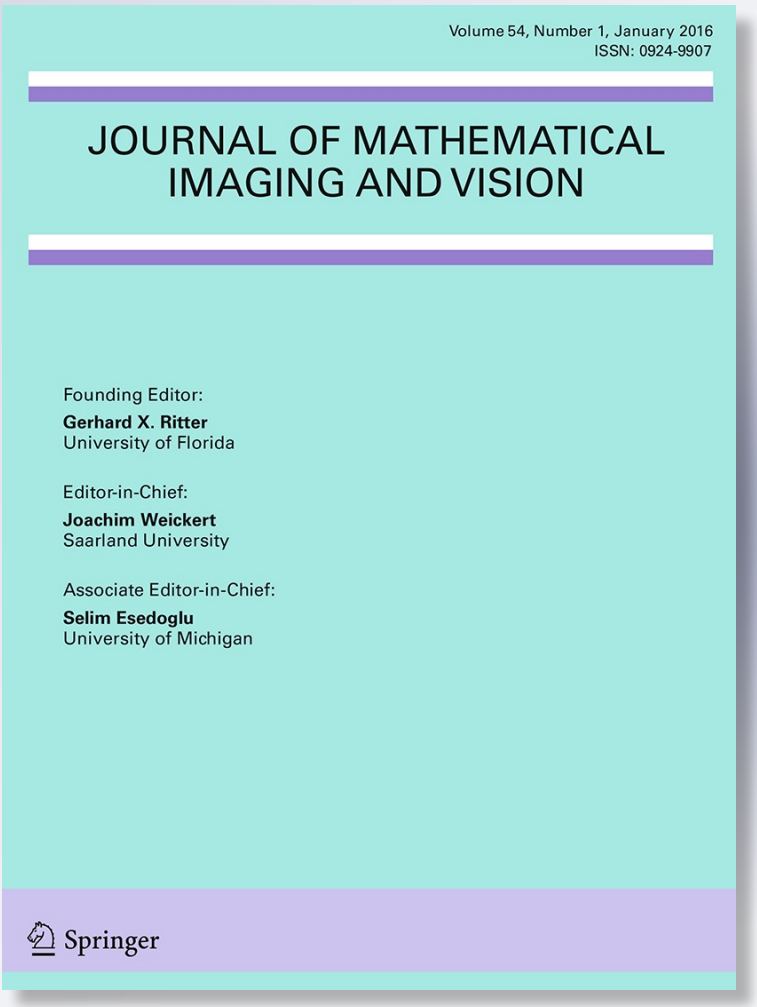
SURE-Type Functionals as Criteria for Parametric PSF Estimation

Feng Xue, Jiaqi Liu, Shenghai Jiao, Shengdong Liu, Min Zhao & Zhenhong Niu

Journal of Mathematical Imaging and Vision

ISSN 0924-9907
Volume 54
Number 1

J Math Imaging Vis (2016) 54:78-105
DOI 10.1007/s10851-015-0590-z



Springer

J Math Imaging Vis (2016) 54:78–105
DOI 10.1007/s10851-015-0590-z

SURE-Type Functionals as Criteria for Parametric PSF Estimation

Feng Xue[1] · Jiaqi Liu[1] · Shenghai Jiao[1] · Shengdong Liu[1] · Min Zhao[1] · Zhenhong Niu[1]

Received: 13 September 2014 / Accepted: 19 June 2015 / Published online: 7 July 2015

Abstract Point spread function (PSF) estimation plays an important role in blind image deconvolution. This paper proposes two novel criteria for parametric PSF estimation, based on Stein's unbiased risk estimate (SURE), namely, prediction-SURE and its variant. We theoretically prove the SURE-type functionals incorporating exact (complementary) smoother filtering as the valid criteria for PSF estimation. We also provide the theoretical error analysis for the regularizer approximations, by which we show that the proposed frequency-adaptive regularization term yields more accurate PSF estimate than others. In particular, the proposed SURE-variant enables us to avoid estimation of noise variance, which is a key advantage over the traditional SURE-like functional. Finally, we propose an efficient algorithm for the minimizations of the criteria. Not limited to the examples we show in this paper, the proposed SURE-based framework has a great potential for other imaging applications, provided the parametric PSF form is available.

Keywords Blind deconvolution · Parametric PSF estimation · Prediction-SURE · SURE-variant

1 Introduction

1.1 Problem Statement

In this paper, we consider the following observation model

$$\mathbf{y} = \mathbf{H}_0\mathbf{x} + \mathbf{b}; \qquad \mu = \mathbf{H}_0\mathbf{x}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^N$ is the observed data of the original (unknown) $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{H}_0 \in \mathbb{R}^{N\times N}$ is a ground truth (unknown) convolution matrix constructed by PSF 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 $\mathbf{x}$, from the observed data $\mathbf{y}$ only.

1.2 Related Works

This problem has been an important image processing topic for several decades, and recent contributions can be found in [2,9,16,22,23]. Examples of real applications include medical imaging [19], fluorescence microscopy [28], astronomical imaging [20], remote sensing [26], and infrared detection [35]. Blind deconvolution is a highly ill-posed problem. To tackle the ill-posedness, it is often to apply regularization or Bayesian approaches, which enforce certain regularity conditions on the original image and point spread function (PSF), and formulate deconvolution as a constrained optimization problem [9,22]. People may refer to [4,7,15,21] for a comprehensive review.

In our preceding works [32–34], we apply the separate strategy where PSF estimation is followed by non-blind deconvolution, and focus on the parametric PSF estimation. Refer to [34] and references therein for relevant literature review. A number of methods have been proposed for parametric PSF estimation, e.g., GCV [27], kurtosis [17], DL1C [10], and APEX [8]. Refer to [34] for these methods and their limitations.

1.3 SURE-Based Approach

Recently, Stein's unbiased risk estimate (SURE), which was first proposed by Stein [29], has been revitalized in signal processing field.

✉ Feng Xue
fxue2012@gmail.com; 18910571983@163.com

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

SURE is an unbiased estimate of the mean squared error (MSE) under additive Gaussian noise assumption. Mathematically, considering the linear model (1) with exactly known $\mathbf{H}_0$, we denote a function (or processing) by $\mathbf{f}$, applied to the observed data $\mathbf{y}$. The standard MSE, referring to the estimation of $\mathbf{x}$, is defined as [13]

$$\text{MSE} = \frac{1}{N}\left\|\mathbf{f}(\mathbf{y}) - \mathbf{x}\right\|^2, \tag{2}$$

where $\mathbf{f}(\mathbf{y})$ is an estimate of $\mathbf{x}$ by processing $\mathbf{f}$. Correspondingly, the standard SURE is formulated as [13]

$$\text{SURE} = \frac{1}{N}\left(\left\|\mathbf{f}(\mathbf{y})\right\|^2 - 2\mathbf{y}^{\mathrm{T}}\mathbf{H}_0^{-\mathrm{T}}\mathbf{f}(\mathbf{y}) + 2\sigma^2 \mathrm{div}_{\mathbf{y}}\left(\mathbf{H}_0^{-\mathrm{T}}\mathbf{f}(\mathbf{y})\right) + \|\mathbf{x}\|^2\right), \tag{3}$$

where the divergence is $\mathrm{div}_{\mathbf{y}}\mathbf{v} = \sum_{n=1}^{N} \frac{\partial v_n}{\partial y_n}$. The statistical unbiasedness of the SURE w.r.t. MSE has been proved in [13]. As SURE depends on the observed $\mathbf{y}$ only[1], it can be used as a practical substitute for the true MSE. For signal processing, SURE is often used as a criterion for denoising [12] and non-blind deconvolution [13,25,31]. The key advantage of the SURE-based approach is that it does not require any prior knowledge of the original image $\mathbf{x}$.

For blind deconvolution, in our preceding works [32–34], we proposed a modified version of SURE as a criterion for parametric PSF estimation. However, there are three main limitations with the methods.

- There is no theoretical analysis to guarantee that applying some regularizer as approximation, the SURE-type minimization will yield accurate PSF estimate. This is a rather strong claim that needs an error analysis for the approximation.
- The regularization terms used in [32–34] are not adaptive to various types of images, and thus, they cannot consistently yield the good PSF estimates for a wide range of natural images. Is it possible to propose an adaptive regularizer to improve this estimation accuracy?
- The computation of the SURE-like estimate requires the knowledge of the noise variance, which is, unfortunately, often unknown in practice. One has to estimate it in advance. Is it possible to develop a novel criterion, which does not depend on noise variance?

1.4 Our Contributions

As a continuation of the preceding works [32–34], the present paper reports the recent research progress to address the foregoing problems. Correspondingly, the key contributions of this work are as follows.

- We provide a theoretical error analysis to directly link the approximations of (complementary) smoother filtering to the PSF estimation accuracy.
- To adapt to any type of natural images, we propose a frequency-adaptive regularization term, which yields more accurate PSF estimate than others, especially under high noise levels.
- We propose a variant of SURE as a novel criterion without need of knowledge of noise variance, and theoretically prove it as a valid objective functional for PSF estimation.

In addition, by the theoretical analysis and the cross-validation experiments (see Sect. 6.6), we demonstrate the wide applicability and robustness of the proposed criteria: even if the retrieved PSF does not belong to the same family as the underlying true one, the proposed SURE-functionals always naturally find the best estimate within the assumed parametric form.

1.5 Paper Organization

Based on prediction-MSE, Sect. 2 provides a theoretical error analysis for the regularizer approximation, and proposes a frequency-adaptive regularizer. In Sect. 3, we particularly derive a prediction-SURE for the proposed data-dependent regularizer. Section 4 proposes a variant of SURE that is independent of noise variance and proves it as a valid criterion for PSF estimation. In Sect. 5, we provide an analysis to evaluate the reliability of the estimators (both prediction-SURE and SURE-variant), and summarize the SURE-type framework. In Sect. 6, we exemplify the proposed framework with two particular types of PSF, and report the experimental results for thorough discussions. Some concluding remarks are finally given in Sect. 7.

1.6 Additional Remarks

Throughout this paper, we use boldface lowercase letters, e.g., $\mathbf{x} \in \mathbb{R}^N$, to denote N-dimensional real vectors, where N is typically the number of pixels in an image. The n-th element of $\mathbf{x}$ is written as x_n. 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 $\mathbf{H}$. The notation $\mathbf{H}_{\mathbf{s}}$ is used, if $\mathbf{H}$ is represented by a small number of parameters, say parameter vector $\mathbf{s}$. 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 $\mathbf{H}$.

In this paper, we always assume periodic boundary condition for convolution and deconvolution, so that the linear

[1] The last term $\|\mathbf{x}\|^2$ of (3) is a constant irrelevant to the optimization of function $\mathbf{f}$.

filtering can be computed by discrete Fourier transform (DFT). The convolution matrix (e.g., $\mathbf{H}_0$ in (1)) is circulant and can, therefore, be diagonalized by DFT. We denote the frequency representation of convolution matrix $\mathbf{H}$ by $H(\omega)$, and particularly, $H_{\mathbf{s}}(\omega)$, if $\mathbf{H}$ is described by the parameter vector $\mathbf{s}$. Denote the DFT coefficients of data, say $\mathbf{x}$, by $X(\omega)$.

In the presentation, we will switch repeatedly between frequency domain and matrix language. An important reason for this alternation of viewpoints is that the frequency expressions are easy to understand and manipulate, but less compact and concise than matrix notations. We also want to emphasize in this way the close links between the two domains.

2 Regularizer Approximation and Error Analysis

Based on the linear model (1), the main purpose of this paper is to estimate the matrix $\mathbf{H}_0$, from the observed data $\mathbf{y}$. In this part, we will explain why the standard MSE/SURE fails to estimate $\mathbf{H}_0$, and then, summarize the main result of our previous work [34], which is a fundamental of this paper.

2.1 Theoretical Background

2.1.1 Why Standard MSE/SURE Fails?

We restrict ourselves to linear processing $\mathbf{f}$, denoted by matrix $\mathbf{U} \in \mathbb{R}^{N\times N}$. Then, the standard MSE (2) can be written as

$$\text{MSE} = \frac{1}{N}\left\|\mathbf{U}\mathbf{y} - \mathbf{x}\right\|^2. \tag{4}$$

If the original data $\mathbf{x}$ is exactly known, the expected MSE can be used as an *oracle*[2] criterion of PSF estimation, stated as follows.

Theorem 1 *Consider the linear function* $\mathbf{U}$ *in* (4) *as exact Wiener filtering, defined as*

$$\begin{aligned}\mathbf{U} &= \mathbf{S}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{S}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}\right)^{-1} \Longleftrightarrow U(\omega)\\ &= \frac{H^*(\omega)}{|H(\omega)|^2 + \sigma^2/S_x(\omega)},\end{aligned} \tag{5}$$

where $\mathbf{S}_{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}$, $\mathbf{I}$ *is identity in matrix notation.* $H(\omega)$ *and* $U(\omega)$ *are the frequency representations of matrices* $\mathbf{H}$ *and* $\mathbf{U}$. $S_x(\omega)$ *denotes the power spectral density (PSD) of original image* $\mathbf{x}$. *Then, minimizing the expected MSE over* $\mathbf{H}$ *yields that*

$$\underbrace{\frac{H^*(\omega)}{|H(\omega)|^2 + \sigma^2/S_x(\omega)}}_{U(\omega)} = \underbrace{\frac{H_0^*(\omega)}{|H_0(\omega)|^2 + \sigma^2/S_x(\omega)}}_{U_0(\omega)},$$

and therefore, $H(\omega) = H_0(\omega)$ *for* $\forall\omega$.

Proof First, consider the minimization of expected MSE over all possible linear processings $\mathbf{U}$:

$$\min_{\mathbf{U}} \frac{1}{N}\mathbb{E}\left\{\left\|\mathbf{U}\mathbf{y} - \mathbf{x}\right\|^2\right\}.$$

Replacing $\mathbf{y}$ by $\mathbf{H}_0\mathbf{x} + \mathbf{b}$ as (1), the expected MSE becomes

$$\begin{aligned}\text{Expected MSE} &= \frac{1}{N}\mathbb{E}\left\{\left\|\mathbf{U}\mathbf{y} - \mathbf{x}\right\|^2\right\}\\ &= \frac{1}{N}\mathbb{E}\left\{\left\|\mathbf{U}(\mathbf{H}_0\mathbf{x} + \mathbf{b}) - \mathbf{x}\right\|^2\right\}\\ &= \frac{1}{N}\mathbb{E}\left\{\left\|(\mathbf{U}\mathbf{H}_0 - \mathbf{I})\mathbf{x} + \mathbf{U}\mathbf{b}\right\|^2\right\}\\ &= \frac{1}{N}\mathbb{E}\left\{\left\|(\mathbf{U}\mathbf{H}_0 - \mathbf{I})\mathbf{x}\right\|^2\right\} + \frac{1}{N}\mathbb{E}\left\{\left\|\mathbf{U}\mathbf{b}\right\|^2\right\}\\ &= \frac{1}{N}\mathrm{Tr}\left((\mathbf{U}\mathbf{H}_0 - \mathbf{I})\mathbf{S}_{\mathbf{x}}(\mathbf{U}\mathbf{H}_0 - \mathbf{I})^{\mathrm{T}}\right)\\ &\quad + \frac{1}{N}\sigma^2\mathrm{Tr}\left(\mathbf{U}\mathbf{U}^{\mathrm{T}}\right),\end{aligned}$$

where Tr denotes matrix trace, $\mathbf{I}$ is identity matrix. The covariance matrices are $\mathbf{S}_{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}$, $\sigma^2\mathbf{I} = \mathbb{E}\{\mathbf{b}\mathbf{b}^{\mathrm{T}}\}$. Thus, the minimization over $\mathbf{U}$ yields that $\left(\mathbf{H}_0\mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}\right)\mathbf{U}_0 - \mathbf{S}\mathbf{H}_0^{\mathrm{T}} = \mathbf{0}$, which implies that

$$\mathbf{U}_0 = \mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}}\left(\mathbf{H}_0\mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}\right)^{-1}$$

is a global minimizer of expected MSE.

If we base our linear processing $\mathbf{U}$ as (5), the minimization of expected MSE over $\mathbf{H}$ becomes

$$\mathbf{S}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{S}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}\right)^{-1} = \mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}}\left(\mathbf{H}_0\mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}\right)^{-1},$$

from which we conclude that $\mathbf{H} = \mathbf{H}_0$. It is equivalent to $H(\omega) = H_0(\omega)$ in frequency domain. □

This theorem demonstrates that incorporating exact Wiener filtering (5), the minimization of expected MSE could produce exact PSF estimation. However, the standard MSE (4) cannot be accessed due to the unknown $\mathbf{x}$. Following (3), the corresponding SURE for this case is expressed as

$$\text{SURE} = \frac{1}{N}\left(\left\|\mathbf{U}\mathbf{y}\right\|^2 - 2\mathbf{y}^{\mathrm{T}}\mathbf{H}_0^{-\mathrm{T}}\mathbf{U}\mathbf{y} + 2\sigma^2\mathrm{Tr}\left(\mathbf{H}_0^{-\mathrm{T}}\mathbf{U}\right) + \|\mathbf{x}\|^2\right). \tag{6}$$

We observe that the standard SURE is also inaccessible in practice, since $\mathbf{H}_0$ is unknown. Thus, SURE fails to be a practical criterion for PSF estimation.

[2] *Oracle* means that this criterion is not accessible in practice, due to the unknown $\mathbf{x}$ in (4).

2.1.2 Prediction-MSE—An Oracle Criterion for PSF Estimation

Instead of the standard MSE and SURE, [34] proposed the following predictive version of MSE, namely prediction-MSE (P-MSE):

$$\text{P-MSE} = \frac{1}{N}\left\|\mathbf{U}\mathbf{y} - \mathbf{H}_0\mathbf{x}\right\|^2 \tag{7}$$

and prove its expectation as an *oracle*[3] criterion for PSF estimation. The term *prediction-MSE* stems from model selection problem [1], for the use of predicting the value of $\mu = \mathbf{H}_0\mathbf{x}$, instead of the value of $\mathbf{x}$[4]. Nevertheless, it should be re-emphasized that we are interested in estimating $\mathbf{H}_0$, rather than the value of μ.

Theorem 2.1 of [34] shows that given $\mathbf{U}$ as an exact smoother matrix (described below), the solution $\mathbf{H}$ that minimizes the expected prediction-MSE has the same magnitude as the true matrix $\mathbf{H}_0$ in frequency domain. For ease of the following discussions, let us restate the theorem and its proof here.

Theorem 2 *Consider the linear function* $\mathbf{U}$ *in* (7) *defined as*

$$\begin{aligned}\mathbf{U} &= \mathbf{H}\mathbf{S}_\mathbf{x}\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{S}_\mathbf{x}\mathbf{H}^\mathrm{T} + \sigma^2\mathbf{I}\right)^{-1} \Longleftrightarrow U(\omega)\\ &= \frac{|H(\omega)|^2}{|H(\omega)|^2 + \sigma^2/S_x(\omega)},\end{aligned} \tag{8}$$

where all the notations have been described in Theorem 1. Then, minimizing the expected prediction-MSE over $\mathbf{H}$ *yields that*

$$\underbrace{\frac{|H(\omega)|^2}{|H(\omega)|^2 + \sigma^2/S_x(\omega)}}_{U(\omega)} = \underbrace{\frac{|H_0(\omega)|^2}{|H_0(\omega)|^2 + \sigma^2/S_x(\omega)}}_{U_0(\omega)},$$

and therefore, $|H(\omega)| = |H_0(\omega)|$ *for* $\forall\omega$.

Proof First, consider the minimization of expected P-MSE over all possible linear processings $\mathbf{U}$:

$$\min_{\mathbf{U}} \frac{1}{N}\mathbb{E}\left\{\left\|\mathbf{U}\mathbf{y} - \mathbf{H}_0\mathbf{x}\right\|^2\right\}.$$

Replacing $\mathbf{y}$ by $\mathbf{H}_0\mathbf{x} + \mathbf{b}$ as (1), and applying the similar derivation with proof of Theorem 1, the expected P-MSE finally becomes

$$\begin{aligned}\text{Expected P-MSE} = {}& \frac{1}{N}\mathrm{Tr}\left((\mathbf{U} - \mathbf{I})\mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T}(\mathbf{U} - \mathbf{I})^\mathrm{T}\right)\\ &+ \frac{1}{N}\sigma^2\mathrm{Tr}\left(\mathbf{U}\mathbf{U}^\mathrm{T}\right),\end{aligned} \tag{9}$$

where all the notations were described in Theorem 1. Thus, the minimization over $\mathbf{U}$ yields that $(\mathbf{U}_0 - \mathbf{I})\mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T} + \sigma^2\mathbf{U}_0 = \mathbf{0}$, which implies that

$$\mathbf{U}_0 = \mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T}\left(\mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T} + \sigma^2\mathbf{I}\right)^{-1}$$

is a global optimizer of expected prediction-MSE.

If we base our linear processing $\mathbf{U}$ as (8), the minimization of expected P-MSE over $\mathbf{H}$ becomes

$$\mathbf{H}\mathbf{S}_\mathbf{x}\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{S}_\mathbf{x}\mathbf{H}^\mathrm{T} + \sigma^2\mathbf{I}\right)^{-1} = \mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T}\left(\mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T} + \sigma^2\mathbf{I}\right)^{-1},$$

from which we conclude that $\mathbf{H}\mathbf{S}_\mathbf{x}\mathbf{H}^\mathrm{T} = \mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T}$. It is equivalent to $|H(\omega)| = |H_0(\omega)|$ in frequency domain. □

The frequency description of Theorem 2 and the phase limitation have been mentioned in Corollary 2.1 of [34]. See Fig. 4 for experimental demonstration of this theorem.

Here, it should be noted that the expected-MSE minimization could recover the whole information (both magnitude and phase of PSF) (see Theorem 1), whereas the prediction-MSE criterion cannot detect the phase of PSF (see Theorem 2). However, unlike the inaccessible standard MSE/SURE, the prediction-MSE can be unbiasedly estimated by a statistical substitute (i.e., prediction-SURE), that is accessible in practice (see Sect. 3). Moreover, if the frequency response of the PSF is positive with zero phase, the prediction-MSE minimization succeeds in estimating the accurate PSF. Hence, we consider only zero-phase blur models in this paper. Fortunately, many real-life blurs—linear motion, out-of-focus, and atmospheric turbulence blurs—have zero phase, this assumption is rather unrestrictive.

2.2 Error Analysis for the Regularization Approximation

Note that the exact $U(\omega)$ in (8) cannot be used in practice, since $\sigma^2/S_x(\omega)$ is unknown. It is crucial to find a good regularization term $\lambda R(\omega)$ to approximate $\sigma^2/S_x(\omega)$, i.e.,

$$\begin{aligned}&\mathbf{U}_{\mathbf{H},\lambda} = \mathbf{H}\mathbf{R}^{-1}\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^\mathrm{T} + \lambda\mathbf{I}\right)^{-1}\\ &\Longleftrightarrow U_{H,\lambda}(\omega) = \frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda R(\omega)},\end{aligned} \tag{10}$$

where λ is a regularization parameter. $\mathbf{U}_{\mathbf{H},\lambda}$ is termed as *smoother matrix* in model selection problem [14]. Similarly, we call $U_{H,\lambda}(\omega)$ as *smoother filtering*. The filter $U(\omega)$ in (8) is called *exact smoother filtering*. Thus, by the regularization approximation (10), we formulate the PSF estimation

[3] *Oracle* means that this criterion is not accessible in practice, due to the unknown $\mathbf{H}_0\mathbf{x}$ in (7).

[4] Following the convention of [14], we refer to estimation of $\mathbf{x}$ as *estimation*, and to estimation of $\mu = \mathbf{H}_0\mathbf{x}$ as *prediction*.

as minimizing $\|\mathbf{U}_{\mathbf{H},\lambda}\mathbf{y} - \mathbf{H}_0\mathbf{x}\|^2$ over both $\mathbf{H}$ and λ in practice (for the moment, despite of the unknown $\mathbf{H}_0\mathbf{x}$). Thus, the next problem naturally arises: how is the PSF estimation accuracy, if we use $\mathbf{U}_{\mathbf{H},\lambda}$ (10) instead of exact $\mathbf{U}$ (8)?

Since Theorem 2 theoretically guarantees the exact estimation of magnitude $|H_0(\omega)|$ by using exact smoother filtering (8), it will serve as a benchmark for evaluating the error of PSF estimation by the approximation (10). The following proposition provides an error analysis that directly links the approximation accuracy of $\sigma^2/S_x(\omega)$ to the error of PSF estimation.

Proposition 1 *Considering the minimization of prediction-MSE of* $\mathbf{U}_{\mathbf{H},\lambda}$*:*

$$\min_{\mathbf{H},\lambda} \|\mathbf{U}_{\mathbf{H},\lambda}\mathbf{y} - \mathbf{H}_0\mathbf{x}\|^2,$$

where $\mathbf{U}_{\mathbf{H},\lambda}$ *is defined as* (10)*, the estimation error of* $|H_0(\omega)|$ *satisfies the following inequality*

$$\left||H(\omega)|^2 - |H_0(\omega)|^2\right| \le C_1 \cdot \delta_\omega + C_2, \quad \forall\omega, \tag{11}$$

where C_1 *and* C_2 *are two constants,* δ_ω *denotes the approximation error of regularizer:* $\left|\lambda R(\omega) - \frac{\sigma^2}{S_x(\omega)}\right| \le \delta_\omega$ *for* $\forall\omega$*.* (11) *can be simplified under the following two special cases:*

- *If* $\lambda R(\omega) = \sigma^2/S_x(\omega)$*, then the constant* $C_1 = 0$*;*
- *For non-parametric setting of PSF, the constant* $C_2 = 0$*.*

Proof (upper-bound analysis) From the proof of Theorem 2, we can see that $\mathbf{U}_0 = \mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^{\mathrm{T}}\left(\mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}\right)^{-1}$ is a global minimizer of the prediction-MSE, which indicates that $\|\mathbf{U}\mathbf{y} - \mathbf{H}_0\mathbf{x}\|^2 \ge \|\mathbf{U}_0\mathbf{y} - \mathbf{H}_0\mathbf{x}\|^2$ for any matrix $\mathbf{U}$. By triangular inequality, we have

$$\begin{aligned}\|\mathbf{U}_0\mathbf{y} - \mathbf{H}_0\mathbf{x}\| &\le \underbrace{\|\mathbf{U}\mathbf{y} - \mathbf{H}_0\mathbf{x}\|}_{p(\mathbf{U})} \\ &= \|\mathbf{U}\mathbf{y} - \mathbf{U}_0\mathbf{y} + \mathbf{U}_0\mathbf{y} - \mathbf{H}_0\mathbf{x}\| \\ &\le \underbrace{\|\mathbf{U}\mathbf{y} - \mathbf{U}_0\mathbf{y}\| + \|\mathbf{U}_0\mathbf{y} - \mathbf{H}_0\mathbf{x}\|}_{q(\mathbf{U},\mathbf{U}_0)}.\end{aligned}$$

Here, $q(\mathbf{U}, \mathbf{U}_0)$ is an upper bound of $p(\mathbf{U})$, and satisfies that $q(\mathbf{U}_0, \mathbf{U}_0) = p(\mathbf{U}_0)$. By *Majorization–Minimization* framework [3,24], minimizing $p(\mathbf{U})$ is equivalent to minimizing the upper bound $q(\mathbf{U}, \mathbf{U}_0)$. Note that the second term of $q(\mathbf{U}, \mathbf{U}_0)$ is constant, provided that the optimizer $\mathbf{U}_0$ is fixed. Thus, we have

$$\begin{aligned}&\min_{\mathbf{U}} \underbrace{\|\mathbf{U}\mathbf{y} - \mathbf{H}_0\mathbf{x}\|^2}_{\text{P-MSE}} \Longleftrightarrow \min_{\mathbf{U}} p(\mathbf{U}) \Longleftrightarrow \min_{\mathbf{U}} q(\mathbf{U}, \mathbf{U}_0) \\ &\Longleftrightarrow \min_{\mathbf{U}} \|\mathbf{U}\mathbf{y} - \mathbf{U}_0\mathbf{y}\| \Longleftrightarrow \min_{\mathbf{U}} \|\mathbf{U}\mathbf{y} - \mathbf{U}_0\mathbf{y}\|^2\end{aligned}$$

which means that the prediction-MSE minimization is essentially matching $\mathbf{U}$ to the global optimizer $\mathbf{U}_0$.

Considering $\mathbf{U}$ as $\mathbf{U}_{\mathbf{H},\lambda}$ defined by (10), by the frequency representation, we have

$$\begin{aligned}\|(\mathbf{U} - \mathbf{U}_0)\mathbf{y}\|^2 &= \sum_\omega \left|(U(\omega) - U_0(\omega))Y(\omega)\right|^2 \\ &\le Y^\star \cdot \sum_\omega \left|U(\omega) - U_0(\omega)\right|^2,\end{aligned}$$

where $Y^\star = \max_\omega |Y(\omega)|^2$, and

$$\begin{aligned}U(\omega) &= \frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda R(\omega)} = \frac{V(\omega)}{V(\omega)+1} \text{ with } V(\omega) \\ &= \frac{|H(\omega)|^2}{\lambda R(\omega)} \ge 0, \ \forall\omega \\ U_0(\omega) &= \frac{|H_0(\omega)|^2}{|H_0(\omega)|^2 + \frac{\sigma^2}{S_x(\omega)}} = \frac{V_0(\omega)}{V_0(\omega)+1} \text{ with } V_0(\omega) \\ &= |H_0(\omega)|^2 \frac{S_x(\omega)}{\sigma^2} \ge 0, \ \forall\omega.\end{aligned}$$

Then, by substitution, we have

$$\begin{aligned}\|(\mathbf{U} - \mathbf{U}_0)\mathbf{y}\|^2 &\le Y^\star \cdot \sum_\omega \left|U(\omega) - U_0(\omega)\right|^2 \\ &= Y^\star \cdot \sum_\omega \left(\frac{V(\omega)}{V(\omega)+1} - \frac{V_0(\omega)}{V_0(\omega)+1}\right)^2 \\ &= Y^\star \cdot \sum_\omega \left(\frac{V(\omega) - V_0(\omega)}{(V(\omega)+1)\cdot(V_0(\omega)+1)}\right)^2 \\ &\le Y^\star \cdot \sum_\omega \left(V(\omega) - V_0(\omega)\right)^2 \\ &\le Y^\star \cdot A^\star \cdot \sum_\omega \left(|H(\omega)|^2 - |H_0(\omega)|^2 \frac{S_x(\omega)}{\sigma^2}\lambda R(\omega)\right)^2,\end{aligned}$$

where $A^\star = \max_\omega \left(\lambda R(\omega)\right)^{-1}$. The second last inequality is due to the fact that $(V(\omega)+1)\cdot(V_0(\omega)+1) \ge 1, \ \forall\omega$.

Finally, by *Majorization–Minimization* again, we obtain

$$\begin{aligned}&\min_{\mathbf{U}} \underbrace{\|\mathbf{U}\mathbf{y} - \mathbf{H}_0\mathbf{x}\|^2}_{\text{P-MSE}} \\ &\Longleftrightarrow \min_H \underbrace{\sum_\omega \left(|H(\omega)|^2 - |H_0(\omega)|^2 \frac{S_x(\omega)}{\sigma^2}\lambda R(\omega)\right)^2}_{J(H)}\end{aligned} \tag{12}$$

• *Non-parametric case* For the non-parametric setting of $H(\omega)$, $H(\omega)$ has the full degree of freedom. The solution to (12), i.e., the minimizer of $J(H)$, is obviously

$$|H^\star(\omega)|^2 = |H_0(\omega)|^2 \frac{S_x(\omega)}{\sigma^2}\lambda R(\omega) = |H_0(\omega)|^2\Big(1 + \frac{S_x(\omega)}{\sigma^2}\underbrace{\Big(\lambda R(\omega) - \frac{\sigma^2}{S_x(\omega)}\Big)}_{\Delta_1\omega}\Big), \quad \forall\omega, \tag{13}$$

where $\Delta\omega$ denotes the approximation error of $\lambda R(\omega)$ w.r.t. $\sigma^2/S_x(\omega)$. And the minimum value of $J(H)$ is $J_1^\star = J(H^\star) = 0$. Moreover, one can easily verify that $H^\star(\omega)$ is also the minimizer of $\big(|H(\omega)|^2 - |H_0(\omega)|^2\lambda R(\omega)\frac{S_x(\omega)}{\sigma^2}\big)^2$—each summation term of $J(H)$ for any fixed ω. It implies that not only $J(H^\star) = 0$, but each term of $J(H^\star)$ is exactly zero as well.

1. Ideally, if $\lambda R(\omega) = \frac{\sigma^2}{S_x(\omega)}$, such that $\Delta_1\omega = 0$, (13) leads to $|H^\star(\omega)| = |H_0(\omega)|$, which coincides with Theorem 2.
2. In practice, we use the regularizer $\lambda R(\omega) \neq \frac{\sigma^2}{S_x(\omega)}$, (13) yields that

$$\Big||H^\star(\omega)|^2 - |H_0(\omega)|^2\Big| = |H_0(\omega)|^2\frac{S_x(\omega)}{\sigma^2}\big|\Delta_1\omega\big| \leq \underbrace{\max_\omega\Big(|H_0(\omega)|^2\frac{S_x(\omega)}{\sigma^2}\Big)}_{C^\star}\cdot\underbrace{\max_\omega\big|\Delta_1\omega\big|}_{\delta_\omega}, \quad \forall\omega.$$

- *Parametric case* In a parametric setting, $H(\omega)$ is determined by a few parameters $\mathbf{s}$, denoted by $H_\mathbf{s}(\omega)$. The minimization problem (12) becomes

$$\min_\mathbf{s}\underbrace{\sum_\omega\Big(|H_\mathbf{s}(\omega)|^2 - |H_{\mathbf{s}_0}(\omega)|^2\frac{S_x(\omega)}{\sigma^2}\lambda R(\omega)\Big)^2}_{J(\mathbf{s})}, \tag{14}$$

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

1. Ideally, if $\lambda R(\omega) = \frac{\sigma^2}{S_x(\omega)}$, the minimization (14) becomes

$$\min_\mathbf{s}\sum_\omega\Big(|H_\mathbf{s}(\omega)|^2 - |H_{\mathbf{s}_0}(\omega)|^2\Big)^2 \tag{15}$$

   – If the retrieved $H_\mathbf{s}(\omega)$ has the same parametric form with the true $H_{\mathbf{s}_0}(\omega)$, the minimization (15) 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 2.
   – If the retrieved $H_\mathbf{s}(\omega)$ and the true $H_{\mathbf{s}_0}(\omega)$ belong to different parametric functions (see cross-validation experiments in Sect. 6.6), 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), (15) does not lead to $|H_{\mathbf{s}^\star}(\omega)| = |H_{\mathbf{s}_0}(\omega)|$ for $\forall\omega$ and $J^\star = 0$. Instead, a small error $\Delta_2\omega$ should be allowed

$$|H_{\mathbf{s}^\star}(\omega)|^2 = |H_{\mathbf{s}_0}(\omega)|^2 + \Delta_2\omega, \quad \forall\omega$$

to guarantee the existence of optimal $\mathbf{s}^\star$[5], and the residual error is

$$J_2^\star = J(\mathbf{s}^\star) = \sum_\omega|\Delta_2\omega|^2 \geq \max_\omega|\Delta_2\omega|^2 > 0 = J_1^\star$$

Finally, we obtain the following inequality:

$$\Big||H_{\mathbf{s}^\star}(\omega)|^2 - |H_{\mathbf{s}_0}(\omega)|^2\Big| \leq \max_\omega\big|\Delta_2(\omega)\big| \leq \sqrt{J_2^\star}, \quad \forall\omega$$

2. In practice, we use the regularizer $\lambda R(\omega) \neq \frac{\sigma^2}{S_x(\omega)}$. 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 (14) always yields

$$|H_{\mathbf{s}^\star}(\omega)|^2 = |H_0(\omega)|^2\Big(1 + \frac{S_x(\omega)}{\sigma^2}\Delta_1\omega\Big) + \Delta_2(\omega), \quad \forall\omega$$

and

$$J_2^\star = J(\mathbf{s}^\star) = \sum_\omega|\Delta_2\omega|^2 \geq \max_\omega|\Delta_2\omega|^2 > 0 = J_1^\star$$

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

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

It should be noted that for parametric case with $\lambda R(\omega) \neq \sigma^2/S_x(\omega)$, the minimum value of J, denoted by $J_2^\star = J(\mathbf{s}^\star)$, is always greater than $J_1^\star = J(H^\star)$—the minimum value attained under non-parametric condition. The reason is that due to the limited degrees of freedom by the PSF parametrization, Eq. (13) is by no means satisfied for $\forall\omega$. One can understand the PSF parametrization as a constraint of the

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

optimization (14), whereas problem (12) is unconstrained optimization under non-parametric setting.

Summarizing all the cases leads to the inequality (11), where either C_1 or C_2 vanishes for some special cases. □

Note that Theorem 2 is merely a special case of Proposition 1 under non-parametric condition with $\lambda R(\omega) = \sigma^2/S_x(\omega)$. Proposition 1 deals with much more general problem settings: it shows that the smaller the error δ_ω is, the more accurate the estimate $|H(\omega)|$ is. Furthermore, Proposition 1 also reveals the underlying philosophy of the prediction-MSE minimization: it naturally finds the best approximation of $|H_0(\omega)|$. The inequality (11) always holds, even if the retrieved $H(\omega)$ has a different parametric form from the underlying $H_0(\omega)$. See the above proof and cross-validation experiments in Sect. 6.6.

2.3 Choices of Regularizer $R(\omega)$ and the Comparisons

In the preliminary works of [32,33], the authors used a constant λ (i.e. $R(\omega) = 1$) to replace $\sigma^2/S_x(\omega)$ and obtained the following approximation:

$$U_{H,\lambda}(\omega) = \frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda}. \tag{16}$$

Note that this simple approximation (16) is not sufficiently accurate for the frequency-varying term $\sigma^2/S_x(\omega)$. Consider the basic observation that for natural images with strong low frequencies and weak high frequencies, $\sigma^2/S_x(\omega)$ increases as the frequency ω goes up, since the noise content is often relatively flat with frequency. For this reason, [34] proposed the following approximation:

$$U_{H,\lambda}(\omega) = \frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda\omega^2}, \tag{17}$$

where $R(\omega) = \omega^2$. It is desired to have $\sigma^2/S_x(\omega) \approx \lambda\omega^2$ for some λ.

However, both the above approximations cannot cope with various types of natural images. For example, if the image contains abundant high-frequency information (e.g. *Noise* and *Crowd* in Fig. 3), $S_x(\omega)$ does not drop rapidly at high frequencies. Consequently, one can observe the slow varying of the term $\sigma^2/S_x(\omega)$ with ω, which makes a constant λ of (16) to be a sufficiently accurate approximation of $\sigma^2/S_x(\omega)$, whereas the regularization (17) is not a good choice. Conversely, if the image, like *Bridge* and *Coco* in Fig. 3, is dominantly occupied by homogeneous regions with very few edges and simple patterns, $\sigma^2/S_x(\omega)$ will increase very sharply with the increasing frequency. Thus, $\sigma^2/S_x(\omega)$ cannot be accurately approximated by (16) or (17) in this case.

Hence, both options of (16) and (17) are not adaptive to various types of images, for a consistently good approximation of exact smoother filtering (8). To avoid the disadvantage, we now propose the following frequency-adaptive smoother filtering:

$$\begin{aligned} &\mathbf{U}_{\mathbf{H},\lambda} = \mathbf{H}\mathbf{S}_\mathbf{y}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{S}_\mathbf{y}\mathbf{H}^{\mathrm{T}} + \lambda\mathbf{I}\right)^{-1} \\ \Longleftrightarrow\ &U_{H,\lambda}(\omega) = \frac{|H(\omega)|^2}{|H(\omega)|^2 + \lambda/S_y(\omega)}, \end{aligned} \tag{18}$$

where $\mathbf{R}^{-1} = \mathbf{S}_\mathbf{y}$, $S_y(\omega) = |Y(\omega)|^2$ is the power spectral density of the observed data $\mathbf{y}$, $Y(\omega)$ is the DFT of data $\mathbf{y}$. Though $\mathbf{y}$ is different from the original $\mathbf{x}$, $S_y(\omega)$ shows the similar trend with $S_x(\omega)$ along the frequency ω. Thus, with proper value of parameter λ, $\lambda/S_y(\omega)$ would be a better approximation of $\sigma^2/S_x(\omega)$ for $\forall\omega$.

According to Proposition 1, we evaluate the approximation performance of $\lambda R(\omega)$ by $\delta_\omega = \max_\omega \left|\lambda R(\omega) - \sigma^2/S_x(\omega)\right|$, which can be regarded as ℓ^∞-norm metric. For fair comparison of the above three choices of $R(\omega)$, we optimize the parameter λ by δ_ω-minimization for each choice of $R(\omega)$, i.e.,

$$\lambda_{\mathrm{opt}} = \arg\min_\lambda \underbrace{\max_\omega \left|\lambda R(\omega) - \sigma^2/S_x(\omega)\right|}_{\delta_\omega(\lambda)}$$

for $R(\omega) = 1, \omega^2$ and $1/S_y(\omega)$, respectively. Finally, we compare δ_ω to see which one is the best approximation.

Figure 1 shows the approximations of $\lambda_{\mathrm{opt}} R(\omega)$, taking *Coco* and *Noise* for example. For a 2-D frequency domain of $R(\omega_1, \omega_2)$ and $\sigma^2/S_x(\omega_1, \omega_2)$, we only show a line with $\omega_2 = 0$, i.e., $R(\omega_1, 0)$ and $\sigma^2/S_x(\omega_1, 0)$ varying with ω_1. The same applies to other figures.

The approximation error δ_ω is reported in Table 1. We can see that

- The approximation of $R(\omega) = 1$ performs best for *Noise*, but worst for *Coco*;
- On average, the approximation by $R(\omega) = 1/S_y(\omega)$ achieves a smallest error δ_ω.

As claimed by Proposition 1, the approximation error δ_ω determines the value of minimum prediction-MSE and the error of PSF estimation. We expect the proposed frequency-adaptive regularizer (18) to obtain more accurate PSF estimate. See Sect. 6.2 for further discussions.

3 Prediction-SURE: An Unbiased Estimate of the Prediction-MSE

As mentioned above, the *oracle* prediction-MSE can be unbiasedly estimated by a statistical substitute—prediction-SURE. Theorem 3.1 of [34] formulates the prediction-SURE

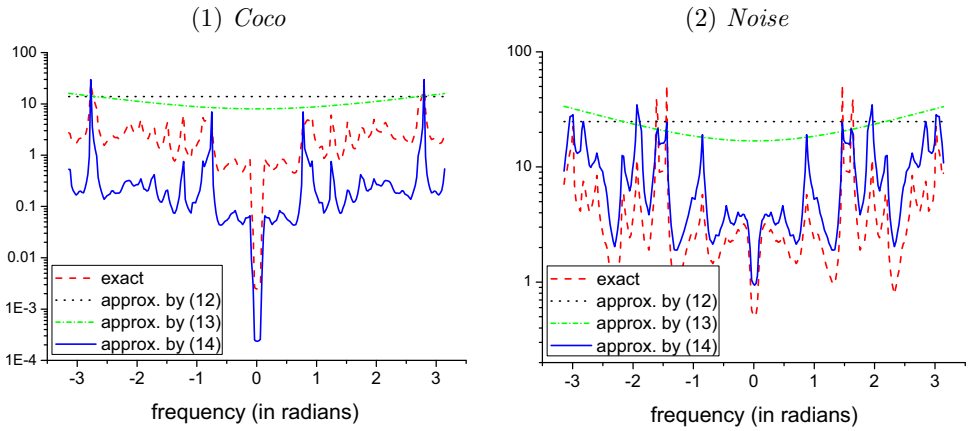


Fig. 1 Approximations of $\sigma^2/S_x(\omega)$ using $\lambda_{\text{opt}}R(\omega)$, where $R(\omega) = 1, \omega^2$ and $1/S_y(\omega)$

Table 1 The approximation errors δ_ω of $\lambda R(\omega)$ to $\sigma^2/S_x(\omega)$

Approximations	λ (16)	$\lambda\omega^2$ (17)	$\lambda/S_y(\omega)$ (18)
Coco	15.44	15.26	6.26
Noise	27.49	31.58	29.90

for the general function $\mathbf{f}(\mathbf{y})$. Let us restate it here as the following lemma.

Lemma 1 *Considering the linear model* (1)*, the following random variable:*

$$\epsilon = \frac{1}{N}\left\|\mathbf{f}(\mathbf{y}) - \mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\text{div}_{\mathbf{y}}\big(\mathbf{f}(\mathbf{y})\big) - \sigma^2$$

is an unbiased estimator of the PMSE, i.e., $\mathbb{E}\{\epsilon\} = \frac{1}{N}\mathbb{E}\big\{\|\mathbf{f}(\mathbf{y}) - \mathbf{H}_0\mathbf{x}\|^2\big\}$.

See Appendix B in [34] for the complete proof. We can see that unlike the standard SURE (3), the prediction-SURE depends on only the measurements $\mathbf{y}$, and thus, is accessible in practice.

Now, we consider the function $\mathbf{f}(\cdot)$ as the smoother matrix $\mathbf{U}_{\mathbf{H},\lambda}$ given as (16)–(18).

- if $\mathbf{U}_{\mathbf{H},\lambda}$ is given as (16) or (17), which is independent of data $\mathbf{y}$, then, $\text{div}_{\mathbf{y}}\big(\mathbf{U}_{\mathbf{H},\lambda}\mathbf{y}\big) = \text{Tr}(\mathbf{U}_{\mathbf{H},\lambda})$ as we derived in [34].
- if $\mathbf{U}_{\mathbf{H},\lambda}$ is given as (18), then, $\text{div}_{\mathbf{y}}\big(\mathbf{U}_{\mathbf{H},\lambda}\mathbf{y}\big) \neq \text{Tr}(\mathbf{U}_{\mathbf{H},\lambda})$, due to the dependence of $\mathbf{U}_{\mathbf{H},\lambda}$ on $\mathbf{y}$. The following theorem formulates the prediction-SURE for this particular case.

Theorem 3 (one-dimensional)

Consider the linear model (1) *with one-dimensional data* $\mathbf{x}$ *and* $\mathbf{y}$*. Given the linear smoother filtering* $\mathbf{U}_{\mathbf{H},\lambda}$ *as* (18)*, the following random variable:*

$$\epsilon = \frac{1}{N}\left\|\mathbf{U}_{\mathbf{H},\lambda}\mathbf{y} - \mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\text{Tr}\big(\mathbf{U}_{\mathbf{H},\lambda} + \mathbf{Q}_{\mathbf{H},\lambda}\big) - \sigma^2$$

is an unbiased estimator of the P-MSE (7)*. By the assumptions of this paper (see the end of Sect.* 1*), the matrix* $\mathbf{Q}_{\mathbf{H},\lambda}$ *is diagonalized by DFT with the following frequency coefficients*[6]*:*

$$Q_{H,\lambda}(\omega) = \frac{|H(\omega)|^2\lambda}{\left(|H(\omega)|^2 + \frac{\lambda}{|Y(\omega)|^2}\right)^2 \cdot |Y(\omega)|^2}$$

See Appendix 1 for the proof. The following corollary, which is naturally derived from Theorem 3, extends one-dimensional to two-dimensional case.

Corollary 1 (two-dimensional) *Consider the linear model* (1) *with two-dimensional data* $\mathbf{x}$ *and* $\mathbf{y}$*. Given the linear smoother filtering* $\mathbf{U}_{\mathbf{H},\lambda}$ *as*

$$U_{H,\lambda}(\omega_1,\omega_2) = \frac{|H(\omega_1,\omega_2)|^2}{|H(\omega_1,\omega_2)|^2 + \lambda/S_y(\omega_1,\omega_2)},$$

the following random variable:

$$\epsilon = \frac{1}{N}\left\|\mathbf{U}_{\mathbf{H},\lambda}\mathbf{y} - \mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\text{Tr}\big(\mathbf{U}_{\mathbf{H},\lambda} + \mathbf{Q}_{\mathbf{H},\lambda}\big) - \sigma^2$$

is an unbiased estimator of the P-MSE (7)*, i.e.,* $\mathbb{E}\{\epsilon\} = \mathbb{E}\big\{\text{P-MSE}\big\}$*. The frequency representation of matrix* $\mathbf{Q}_{\mathbf{H},\lambda}$ *is*

$$Q_{H,\lambda}(\omega_1,\omega_2) = \frac{|H(\omega_1,\omega_2)|^2\lambda}{\left(|H(\omega_1,\omega_2)|^2 + \frac{\lambda}{|Y(\omega_1,\omega_2)|^2}\right)^2 \cdot |Y(\omega_1,\omega_2)|^2}$$

See Appendix 2 for the proof, most of which is based on the proof of Theorem 3. This two-dimensional result can be readily used in this paper to compute the prediction-SURE for image processing.

[6] The practical computation can be fully performed in Fourier domain. The introduction of matrix $\mathbf{Q}$ is for sake of concise expression by linear algebra language.

4 A Variant of Prediction-SURE: Free of Estimation of Noise Variance

Now, we have established the prediction-SURE framework: the PSF is estimated by minimizing the prediction-SURE w.r.t. $\mathbf{H}$ and λ. However, it requires the knowledge of noise variance σ^2. In many scenarios, the prediction-SURE cannot be readily used, when σ^2 is unknown.

4.1 Derivation of a SURE-Variant

Now, we propose a variant of the prediction-SURE, which will be independent of σ^2. First, we consider $\mathbf{U}$ as the exact smoother filtering (8), and denote the linear estimate by $\widehat{\mu} = \mathbf{U}\mathbf{y}$. Then, by Lemma 1, the prediction-SURE of $\widehat{\mu}$ is

$$\mathrm{SURE}(\widehat{\mu}) = \frac{1}{N}\left\|\mathbf{U}\mathbf{y} - \mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\mathrm{Tr}(\mathbf{U}) - \sigma^2$$

Instead of the estimate $\widehat{\mu}$, we consider the following linear combination

$$\begin{aligned}\overline{\mu} &= \alpha\mathbf{y} + (1-\alpha)\widehat{\mu} = \alpha\mathbf{y} + (1-\alpha)\mathbf{U}\mathbf{y}\\ &= \underbrace{\left(\alpha\mathbf{I} + (1-\alpha)\mathbf{U}\right)}_{\overline{\mathbf{U}}}\mathbf{y}\end{aligned}$$

weighted by a parameter α. Then, similar to $\widehat{\mu}$, the prediction-SURE of the estimate $\overline{\mu}$ is

$$\mathrm{SURE}(\overline{\mu}) = \frac{1}{N}\left\|\overline{\mathbf{U}}\mathbf{y} - \mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\mathrm{Tr}(\overline{\mathbf{U}}) - \sigma^2$$

To make $\mathrm{SURE}(\overline{\mu})$ be independent of σ^2, let $\mathrm{Tr}(\overline{\mathbf{U}}) = 0$, which yields that $\alpha = -\frac{\mathrm{Tr}(\mathbf{U})}{N-\mathrm{Tr}(\mathbf{U})}$. Thus, the matrix associated with $\overline{\mu}$, i.e., $\overline{\mathbf{U}}$, has trace 0. The estimate $\overline{\mu}$, therefore, is called *nil-trace linear estimate*. Thus, the prediction-SURE procedure on this estimate becomes

$$\begin{aligned}\mathrm{SURE}(\overline{\mu}) &= \frac{1}{N}\left\|\overline{\mathbf{U}}\mathbf{y} - \mathbf{y}\right\|^2 - \sigma^2\\ &= \frac{1}{N}\left\|(1-\alpha)(\mathbf{U}-\mathbf{I})\mathbf{y}\right\|^2 - \sigma^2\\ &= \frac{1}{N}(1-\alpha)^2\left\|(\mathbf{U}-\mathbf{I})\mathbf{y}\right\|^2 - \sigma^2\\ &= \frac{N}{(N-\mathrm{Tr}(\mathbf{U})^2}\left\|(\mathbf{I}-\mathbf{U})\mathbf{y}\right\|^2 - \sigma^2\\ &= \frac{N\left\|(\mathbf{I}-\mathbf{U})\mathbf{y}\right\|^2}{(\mathrm{Tr}(\mathbf{I}-\mathbf{U}))^2} - \sigma^2\\ &= N\underbrace{\frac{\left\|\mathbf{M}\mathbf{y}\right\|^2}{(\mathrm{Tr}(\mathbf{M}))^2}}_{\overline{\epsilon}} - \sigma^2 \qquad (\text{denoting } \mathbf{I}-\mathbf{U} = \mathbf{M}).\end{aligned} \tag{19}$$

Thus, we obtain a new objective functional $\overline{\epsilon}$, the minimization of which is equivalent to that of the prediction-SURE of $\overline{\mu}$[7]. Since $\overline{\epsilon}$ is derived from the SURE, restricted to a sub-class of linear estimate, $\overline{\epsilon}$ is called *SURE-variant*. The following theorem verifies the expectation of $\overline{\epsilon}$ as a valid criterion for PSF estimation.

Theorem 4 *If the equality*

$$\mathbf{M} = \mathbf{I} - \mathbf{U} = \sigma^2\left(\mathbf{H}\mathbf{S}_\mathbf{x}\mathbf{H}^\mathrm{T} + \sigma^2\mathbf{I}\right)^{-1} \tag{20}$$

holds, minimization of $\mathbb{E}\{\overline{\epsilon}\}$ (19) *over* $\mathbf{H}$ *yields the exact frequency magnitude of PSF:* $|H(\omega)| = |H_0(\omega)|$.

Proof First, we perform the minimization of $\overline{\epsilon}$ (19) over any matrix $\mathbf{M} \in \mathbb{R}^{N\times N}$. Substituting $\mathbf{y} = \mathbf{H}_0\mathbf{x} + \mathbf{b}$ into (19), we develop $\overline{\epsilon}$ as

$$\begin{aligned}\overline{\epsilon} &= \frac{1}{\left(\mathrm{Tr}(\mathbf{M})\right)^2}\left\|\mathbf{M}(\mathbf{H}_0\mathbf{x} + \mathbf{b})\right\|^2\\ &= \frac{1}{\left(\mathrm{Tr}(\mathbf{M})\right)^2}\Big(\mathrm{Tr}\left(\mathbf{M}\mathbf{H}_0\mathbf{x}\mathbf{x}^\mathrm{T}\mathbf{H}_0^\mathrm{T}\mathbf{M}^\mathrm{T}\right) + \mathrm{Tr}\left(\mathbf{M}\mathbf{b}\mathbf{b}^\mathrm{T}\mathbf{M}^\mathrm{T}\right)\\ &\qquad +2\mathbf{b}^\mathrm{T}\mathbf{M}\mathbf{H}_0\mathbf{x}\Big).\end{aligned}$$

The expected value of $\overline{\epsilon}$ is

$$\begin{aligned}\mathbb{E}\{\overline{\epsilon}\} &= \frac{1}{\left(\mathrm{Tr}(\mathbf{M})\right)^2}\Big(\mathrm{Tr}\left(\mathbf{M}\mathbf{H}_0\mathbb{E}\{\mathbf{x}\mathbf{x}^\mathrm{T}\}\mathbf{H}_0^\mathrm{T}\mathbf{M}^\mathrm{T}\right)\\ &\qquad +\mathrm{Tr}\left(\mathbf{M}\mathbb{E}\{\mathbf{b}\mathbf{b}^\mathrm{T}\}\mathbf{M}^\mathrm{T}\right)\Big)\\ &= \frac{1}{\left(\mathrm{Tr}(\mathbf{M})\right)^2}\Big(\mathrm{Tr}\left(\mathbf{M}\mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T}\mathbf{M}^\mathrm{T}\right) + \sigma^2\mathrm{Tr}\left(\mathbf{M}\mathbf{M}^\mathrm{T}\right)\Big)\\ &= \frac{1}{\left(\mathrm{Tr}(\mathbf{M})\right)^2}\mathrm{Tr}\Big(\mathbf{M}(\underbrace{\mathbf{H}_0\mathbf{S}_\mathbf{x}\mathbf{H}_0^\mathrm{T} + \sigma^2\mathbf{I}}_{\mathbf{A}})\mathbf{M}^\mathrm{T}\Big).\end{aligned}$$

The problem becomes

$$\min_{\mathbf{M}} \frac{\mathrm{Tr}\left(\mathbf{M}\mathbf{A}\mathbf{M}^\mathrm{T}\right)}{\left(\mathrm{Tr}(\mathbf{M})\right)^2},$$

where $\mathbf{A} \succ 0$. The optimization is equivalent to the following problem:

$$\min_{\mathbf{M}} \mathrm{Tr}\left(\mathbf{M}\mathbf{A}\mathbf{M}^\mathrm{T}\right) \quad \text{s.t.} \quad \mathrm{Tr}(\mathbf{M}) = 1. \tag{21}$$

Lagrangian method formulates it as

$$\min_{\mathbf{M}} \underbrace{\mathrm{Tr}\left(\mathbf{M}\mathbf{A}\mathbf{M}^\mathrm{T}\right) - 2\beta\left(\mathrm{Tr}(\mathbf{M}) - 1\right)}_{J(\mathbf{M})},$$

[7] The last term σ^2, though unknown, is a constant irrelevant to the minimization procedure.

where β is a Lagrangian multiplier. By variation method, we have

$$\begin{aligned}\delta J &= J(\mathbf{M}+\delta\mathbf{M}) - J(\mathbf{M}) \\ &= 2\mathrm{Tr}(\delta\mathbf{M}\mathbf{A}\mathbf{M}^{\mathrm{T}}) - 2\beta\mathrm{Tr}(\delta\mathbf{M}) + \mathcal{O}\big(\mathrm{Tr}(\delta\mathbf{M})\big) \\ &= \mathrm{Tr}\Big(\delta\mathbf{M}\big(2\mathbf{A}\mathbf{M}^{\mathrm{T}} - 2\beta\mathbf{I}\big)\Big) + \mathcal{O}\big(\mathrm{Tr}(\delta\mathbf{M})\big).\end{aligned}$$

Hence, the solution of $\mathbf{M}$ should satisfy

$$2\mathbf{A}\mathbf{M}^{\mathrm{T}} - 2\beta\mathbf{I} = 0 \Longrightarrow \mathbf{M} = \beta\mathbf{A}^{-\mathrm{T}}.$$

Substituting $\mathbf{A} = \mathbf{H}_0\mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}$, we obtain

$$\mathbf{M} = \beta\big(\mathbf{H}_0\mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}\big)^{-\mathrm{T}} = \beta\big(\mathbf{H}_0\mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}\big)^{-1},$$

where β can be any constant for the moment.

On the other hand, if we restrict the linear processing to $\mathbf{M} = \sigma^2\big(\mathbf{H}\mathbf{S}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}\big)^{-1}$, then, we have

$$\sigma^2\big(\mathbf{H}\mathbf{S}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I}\big)^{-1} = \beta\big(\mathbf{H}_0\mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}\big)^{-1}$$

which yields $\mathbf{H}\mathbf{S}_{\mathbf{x}}\mathbf{H}^{\mathrm{T}} = \mathbf{H}_0\mathbf{S}_{\mathbf{x}}\mathbf{H}_0^{\mathrm{T}}$ and the constant $\beta = \sigma^2$. In frequency domain, it is equivalent to $|H(\omega)| = |H_0(\omega)|$. □

The proof reveals that the SURE-variant minimization is essentially equivalent to matching the filtering $M(\omega) = \sigma^2(|H(\omega)|^2 S_x(\omega) + \sigma^2)^{-1}$ in frequency domain. Recalling the definition of exact smoother filtering as Eq. (8), we have a connection that $M(\omega) + U(\omega) = 1$ for $\forall\omega$ or $\mathbf{M} + \mathbf{U} = \mathbf{I}$. Hence, we call $M(\omega)$ as *complementary smoother filtering*.

Theorem 4 shows that incorporating exact complementary smoother filtering (20), minimization of SURE-variant yields the exact estimate of PSF magnitude. Compared to the traditional SURE, a key advantage of the SURE-variant is that $\overline{\epsilon}$ does not depend on noise variance σ^2: we do not need to estimate σ^2 in advance.

4.2 Error Analysis for the Approximation of Complementary Smoother Filtering

By analogy to Sect. 2.2, due to the unknown $\sigma^2/S_x(\omega)$, we practically use the following regularization

$$\begin{aligned}\mathbf{M}_{\mathbf{H},\lambda} = \lambda\big(\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^{\mathrm{T}} + \lambda\mathbf{I}\big)^{-1} &\Longleftrightarrow M_{H,\lambda}(\omega) \\ &= \frac{\lambda R(\omega)}{|H(\omega)|^2 + \lambda R(\omega)} \qquad (22)\end{aligned}$$

to approximate the exact one (20). Hence, we apply $\mathbf{M}_{\mathbf{H},\lambda}$ in the SURE-variant $\overline{\epsilon}$, instead of the exact but inaccessible $\mathbf{M}$ (20). Thus, PSF estimation is formulated as minimization of $\overline{\epsilon}$ w.r.t. $\mathbf{H}$ and λ.

Similar to Proposition 1, the following proposition provides a close link between the approximation error of regularization and PSF estimation error.

Proposition 2 *Considering the minimization of SURE-variant of* $\mathbf{M}_{\mathbf{H},\lambda}$*:*

$$\min_{\mathbf{H},\lambda} \frac{\big\|\mathbf{M}_{\mathbf{H},\lambda}\mathbf{y}\big\|^2}{\big(\mathrm{Tr}\big(\mathbf{M}_{\mathbf{H},\lambda}\big)\big)^2},$$

where $\mathbf{M}_{\mathbf{H},\lambda}$ *is defined as* (22)*, the estimation error of* $|H_0(\omega)|$ *satisfies the following inequality*

$$\Big||H(\omega)|^2 - |H_0(\omega)|^2\Big| \le C_1 \cdot \delta_\omega + C_2,$$

where C_1 *and* C_2 *are two constants,* δ_ω *denotes the approximation error of regularizer:* $\big|\lambda R(\omega) - \frac{\sigma^2}{S_x(\omega)}\big| \le \delta_\omega$ *for* $\forall\omega$*. This inequality can be simplified under the following two special cases:*

- *If* $\lambda R(\omega) = \sigma^2/S_x(\omega)$*, then the constant* $C_1 = 0$*;*
- *For non-parametric setting of PSF, the constant* $C_2 = 0$*.*

Proof (upper-bound analysis) From the proof of Theorem 4, the minimization of expected SURE-variant is finally equivalent to

$$\min_{\mathbf{M}} \mathrm{Tr}\big(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathrm{T}}\big) - 2\sigma^2\big(\mathrm{Tr}(\mathbf{M}) - 1\big)$$

which can be expressed in frequency domain (ignoring the last constant):

$$\min_{M} \sum_{\omega} M^2(\omega)\big(\underbrace{|H_0(\omega)|^2 S_x(\omega) + \sigma^2}_{A(\omega)}\big) - 2\sigma^2 \cdot M(\omega).$$

It is equivalent to

$$\min_{M} \sum_{\omega} M^2(\omega)\frac{1}{M_0(\omega)} - 2M(\omega),$$

where $M_0(\omega) = \big(|H_0(\omega)|^2\frac{S_x(\omega)}{\sigma^2} + 1\big)^{-1}$. Then, the minimization problem becomes (ignoring the last constant)

$$\min_{M} \sum_{\omega} \frac{1}{M_0(\omega)}\Big(M(\omega) - M_0(\omega)\Big)^2.$$

Since $M_0(\omega) \le 1$ for $\forall\omega$, we have

$$\underbrace{\sum_{\omega} \frac{1}{M_0(\omega)}\Big(M(\omega) - M_0(\omega)\Big)^2}_{p(M)} \le \underbrace{\sum_{\omega}\Big(M(\omega) - M_0(\omega)\Big)^2}_{q(M,M_0)}$$

which indicates that $q(M, M_0)$ is an upper bound of $p(M)$, and $p(M_0) = q(M_0, M_0) = 0$. By *Majorization–Minimization*, we establish the following equivalence

$$\text{Minimizing SURE-variant} \iff \min_M \sum_\omega \Big(M(\omega) - M_0(\omega)\Big)^2$$

Considering $M(\omega)$ given by (22), i.e., $M(\omega) = \big(|H(\omega)|^2 \frac{1}{\lambda R(\omega)} + 1\big)^{-1}$. By the substitution, denoting $V_0(\omega) = |H_0(\omega)|^2 \frac{S_x(\omega)}{\sigma^2}$ and $V(\omega) = |H(\omega)|^2 \frac{1}{\lambda R(\omega)}$, we have

$$\begin{aligned}
\sum_\omega \Big(M(\omega) - M_0(\omega)\Big)^2 &= \sum_\omega \left(\frac{1}{V(\omega)+1} - \frac{1}{V_0(\omega)+1}\right)^2 \\
&= \sum_\omega \left(\frac{V(\omega) - V_0(\omega)}{(V(\omega)+1)\cdot(V_0(\omega)+1)}\right)^2 \\
&\le \sum_\omega \Big(V(\omega) - V_0(\omega)\Big)^2 \\
&\le A^\star \cdot \sum_\omega \left(|H(\omega)|^2 - |H_0(\omega)|^2 \frac{S_x(\omega)}{\sigma^2}\lambda R(\omega)\right)^2,
\end{aligned}$$

where $A^\star = \max_\omega(\lambda R(\omega))^{-1}$. The second last inequality is due to the fact that $(V(\omega)+1)\cdot(V_0(\omega)+1) \ge 1, \forall\omega$. Finally, by *Majorization–Minimization* again, the SURE-variant minimization is finally equivalent to

$$\min_H \sum_\omega \left(|H(\omega)|^2 - |H_0(\omega)|^2 \frac{S_x(\omega)}{\sigma^2}\lambda R(\omega)\right)^2.$$

Now, we are solving exactly the same problem with Proposition 1 (see Eq. (12)). Refer to the proof of **Proposition 1** to complete the discussion of all aspects. □

Similar to the relation between Theorem 1 and Proposition 1, it is noted that Theorem 4 stands for a special case of Proposition 2 under non-parametric condition with $\lambda R(\omega) = \sigma^2/S_x(\omega)$. Proposition 2 covers much more general problem settings and draws the same conclusion with Proposition 1.

Similar to (16)–(18), the regularizer $R(\omega)$ can be selected as $1, \omega^2$ or $1/S_y(\omega)$. Section 2.3 has demonstrated that the following frequency-adaptive regularizer:

$$\begin{aligned}
\mathbf{M}_{\mathbf{H},\lambda} = \lambda\big(\mathbf{H}\mathbf{S}_\mathbf{y}\mathbf{H}^\mathrm{T} + \lambda\mathbf{I}\big)^{-1} &\iff M_{H,\lambda}(\omega) \\
&= \frac{\lambda}{|H(\omega)|^2 S_y(\omega) + \lambda} \qquad (23)
\end{aligned}$$

is superior to other choices in terms of approximation accuracy.

5 SURE-Type Framework

5.1 A Unified Analysis of Statistical Reliability

Theorems 2 and 4 verified expected prediction-MSE and expected SURE-variant as valid criteria for PSF estimation. However, in practice, we are minimizing prediction-SURE and SURE-variant, rather than their expected values. It is necessary to analyze the discrepancy between a random variable (e.g., prediction-SURE or SURE-variant) and its expectation (e.g., expected prediction-MSE and expected SURE-variant) for more rigorous treatment.

Let $\widehat{\xi}$ denote a random variable, and ξ denote its expectation $\xi = \mathbb{E}\{\widehat{\xi}\}$. By Theorems 2 and 4, if exact (complimentary) smoother filtering is applied, an *oracle* is supposed to determine the exact estimate of $\mathbf{H}_0$, i.e., $\mathbf{H}_0 = \arg\min_\mathbf{H} \xi(\mathbf{H})$, and correspondingly, we have $\mathbf{H}^\star = \arg\min_\mathbf{H} \widehat{\xi}(\mathbf{H})$ in practice.

More formally, the associated *oracle* inequality states that with high probability $\xi(\mathbf{H}^\star) \le \xi(\mathbf{H}_0) + \gamma$. That is to say for $\forall\alpha \in (0,1)$, there is a positive number γ, s.t.

$$\mathbb{P}\big(\xi(\mathbf{H}^\star) - \xi(\mathbf{H}_0) \ge \gamma\big) \le \alpha$$

Notice that $\widehat{\xi}(\mathbf{H}^\star) \le \widehat{\xi}(\mathbf{H}_0)$, we can write

$$\begin{aligned}
\mathbb{P}\big(\xi(\mathbf{H}^\star) - \xi(\mathbf{H}_0) \ge \gamma\big) &\le \mathbb{P}\Big(\big|\widehat{\xi}(\mathbf{H}_0) - \xi(\mathbf{H}_0)\big| \ge \frac{\gamma}{2}\Big) \\
&\quad + \mathbb{P}\Big(\big|\widehat{\xi}(\mathbf{H}^\star) - \xi(\mathbf{H}^\star)\big| \ge \frac{\gamma}{2}\Big) \\
&\le \sum_\mathbf{H} \mathbb{P}\Big(\big|\widehat{\xi}(\mathbf{H}) - \xi(\mathbf{H})\big| \ge \frac{\gamma}{2}\Big)
\end{aligned}$$

considering all the convolution matrix $\mathbf{H}$. Then, by Chebyshev's inequality, we obtain

$$\begin{aligned}
\mathbb{P}\big(\xi(\mathbf{H}^\star) - \xi(\mathbf{H}_0) \ge \gamma\big) &\le \sum_\mathbf{H} \frac{4}{\gamma^2}\mathbb{E}\Big\{\big|\widehat{\xi}(\mathbf{H}) - \xi(\mathbf{H})\big|^2\Big\} \\
&= \frac{4}{\gamma^2}\sum_\mathbf{H} \mathcal{R}\big(\widehat{\xi}(\mathbf{H})\big),
\end{aligned}$$

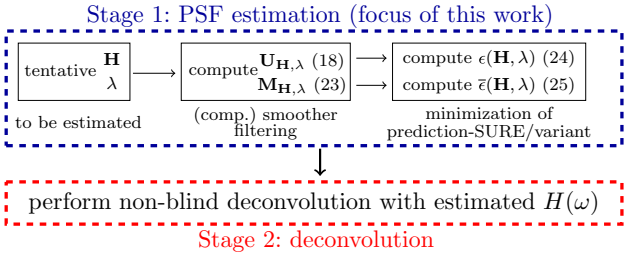


Fig. 2 The flowchart of PSF estimation: joint minimization of the SURE-type functionals over $\mathbf{H}$ and λ, as shown in (24) and (25)

Fig. 3 Original images. **a** *Bridge* 512 × 512; **b** *Coco* 256 × 256; **c** *Crowd* 512 × 512; **d** *Noise* 256 × 256

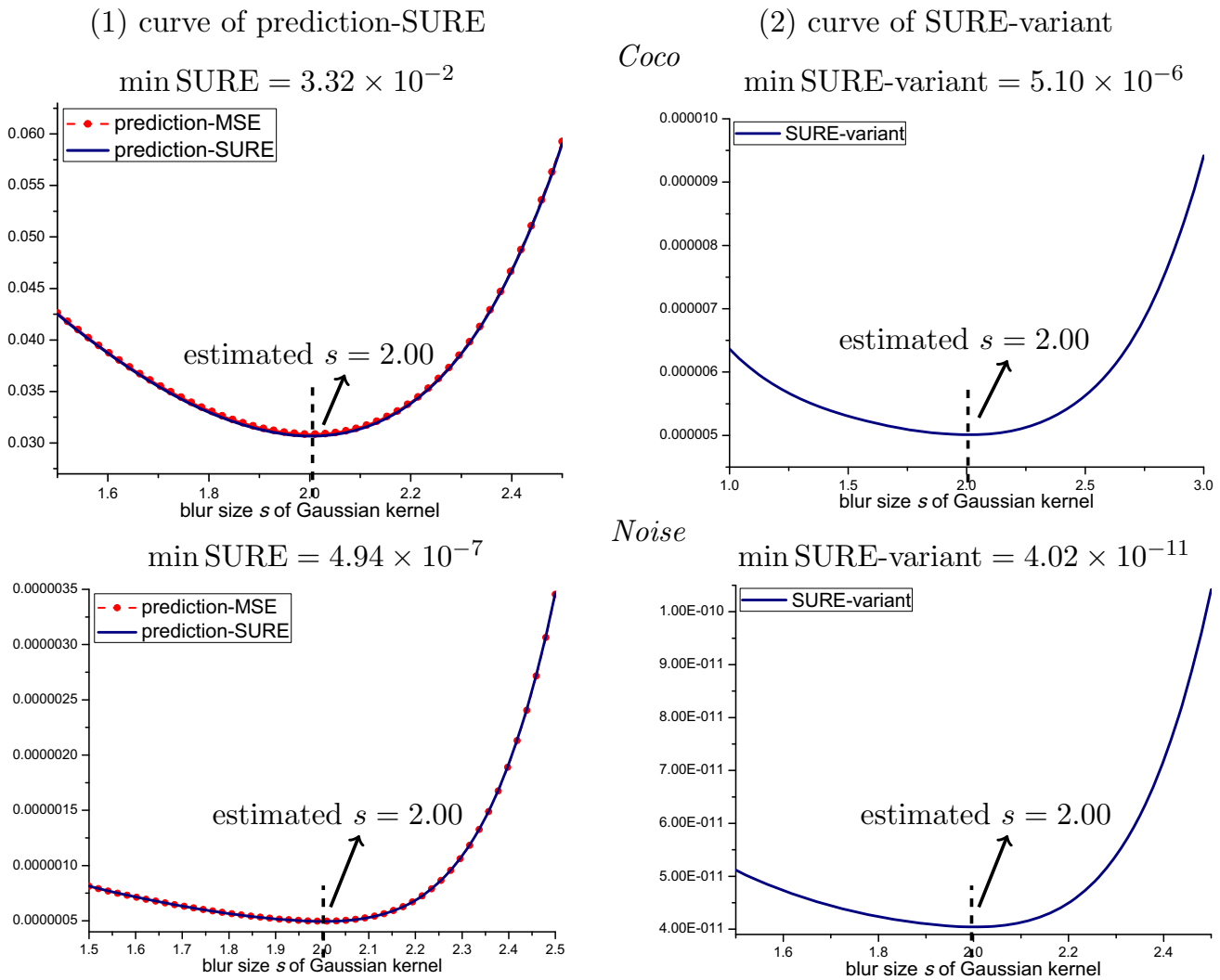


Fig. 4 The prediction-SURE/SURE-variant minimization with exact (complementary) smoother filtering under **Gaussian** kernel: *Coco* and *Noise*, BSNR = 40 dB.

where $\mathcal{R}$ is the statistical risk of a quadratic loss function

$$\mathcal{R}\big(\widehat{\xi}(\mathbf{H})\big) := \mathbb{E}\Big\{\big|\widehat{\xi}(\mathbf{H}) - \xi(\mathbf{H})\big|^2\Big\}.$$

The variance of SURE, i.e., $\mathcal{R}$, has been discussed in [18,25,30]. Especially, [18] proposed an upper bound of the variance of SURE w.r.t. the expectation. Generally, the variance of SURE is inversely proportional to the sample number

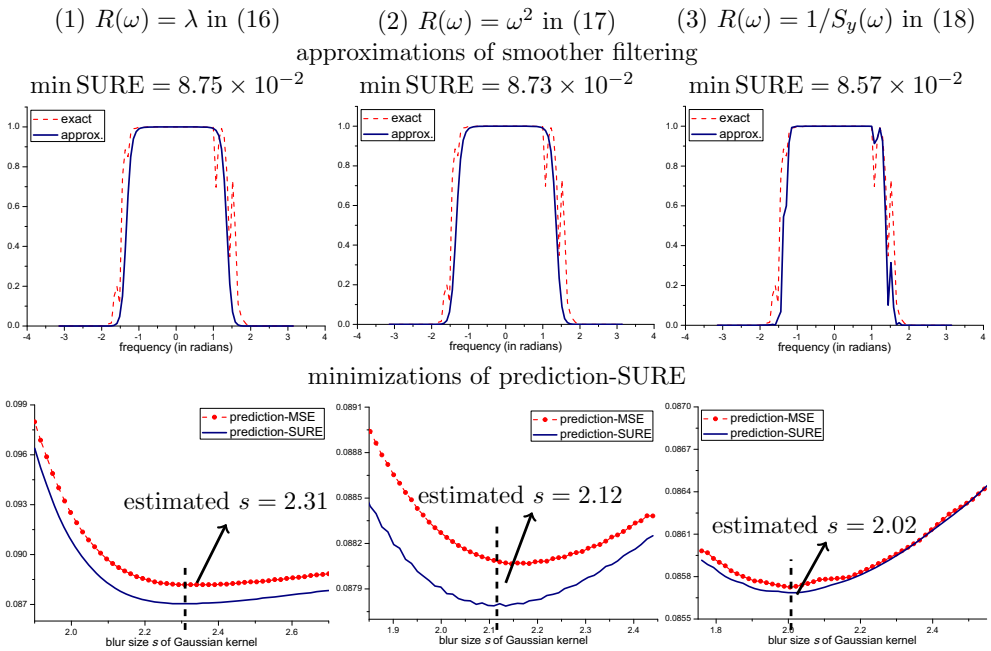


Fig. 5 The prediction-SURE minimization under **Gaussian** kernel: *Coco*, BSNR = 20 dB

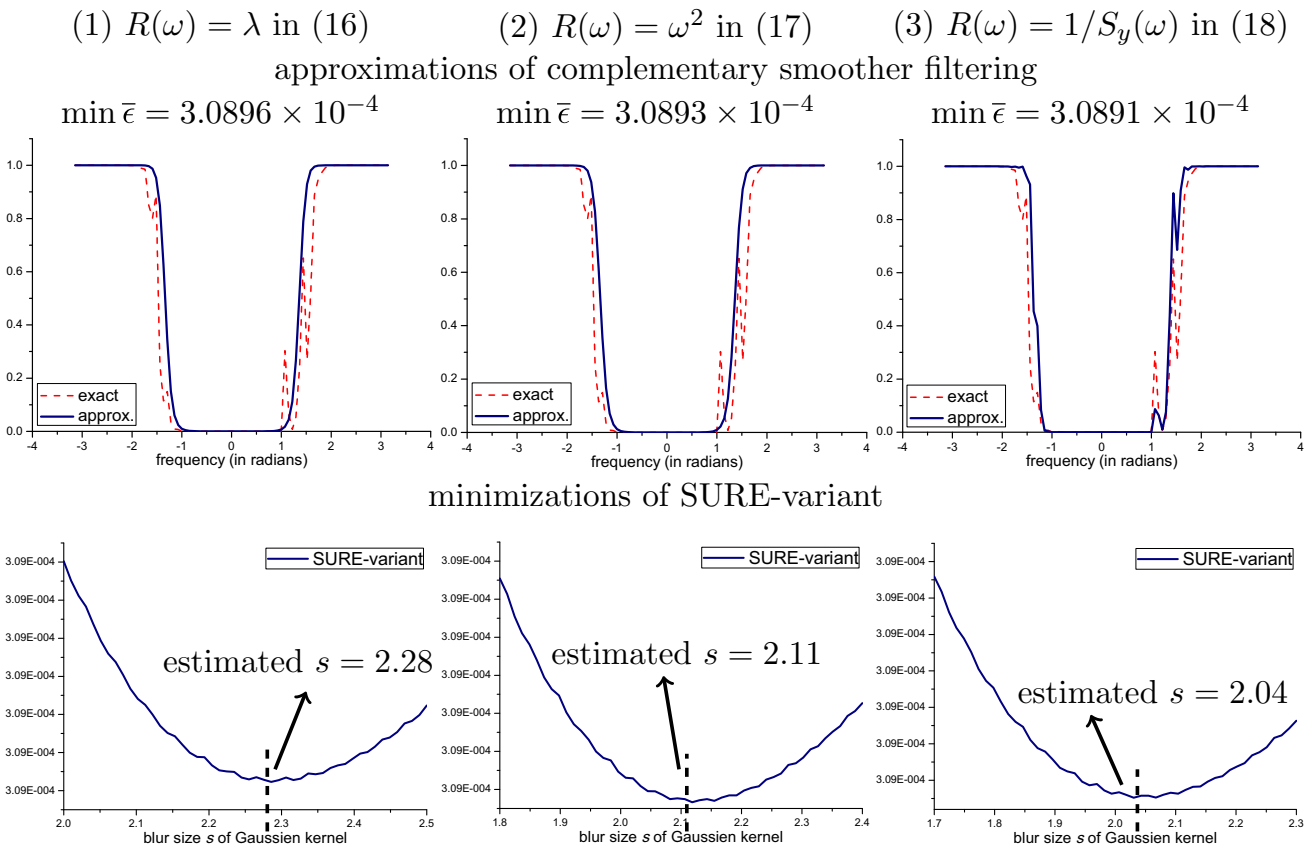


Fig. 6 The SURE-variant minimization under **Gaussian** kernel: *Coco*, BSNR = 20 dB

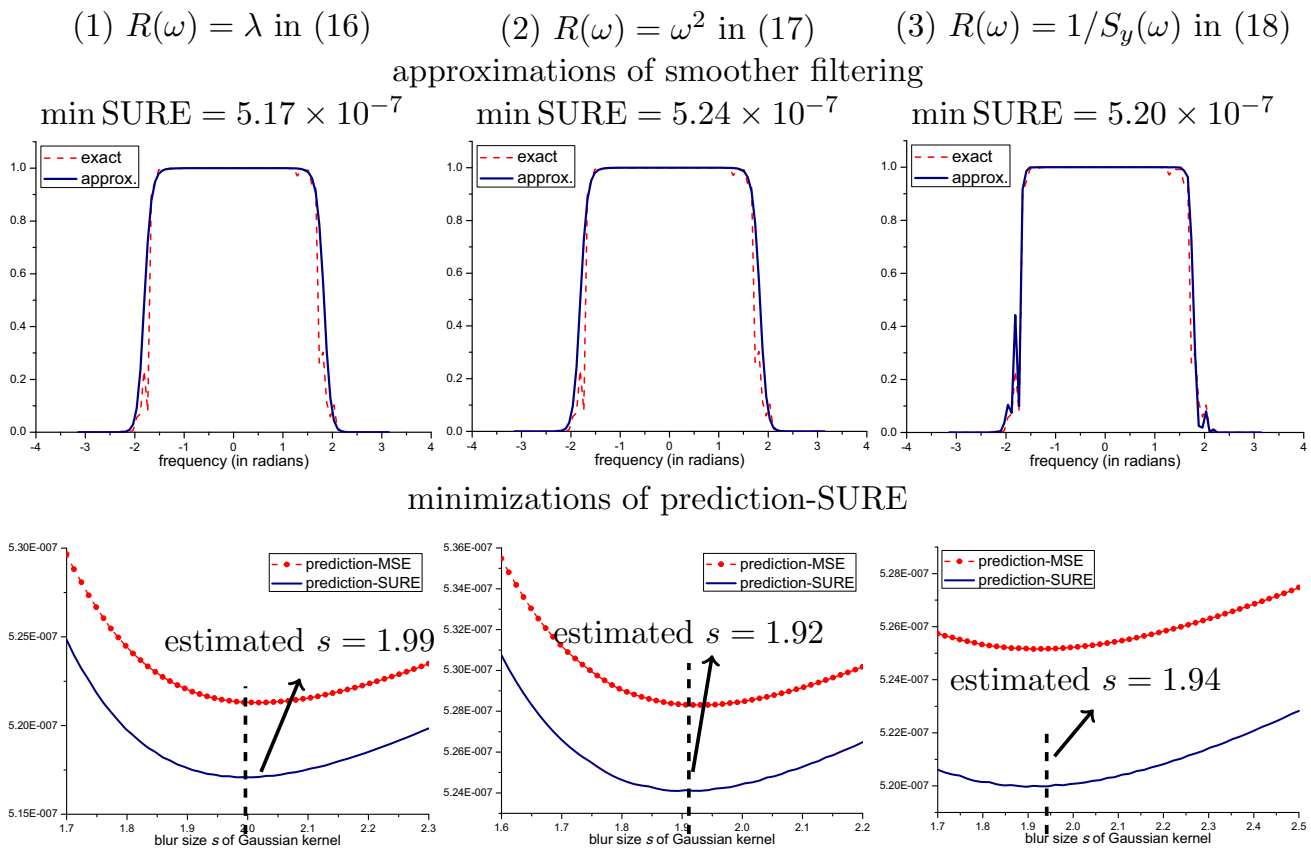


Fig. 7 The prediction-SURE minimization under **Gaussian** kernel: *Noise*, BSNR = 40 dB

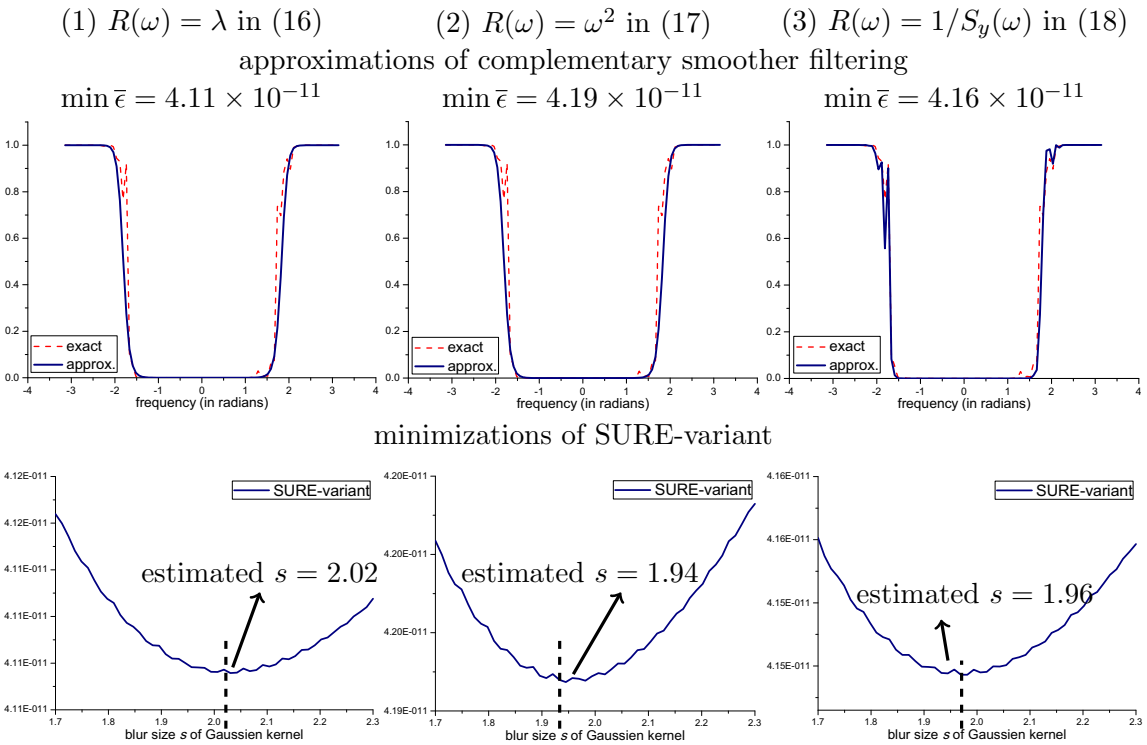


Fig. 8 The SURE-variant minimization under **Gaussian** kernel: *Noise*, BSNR = 40 dB

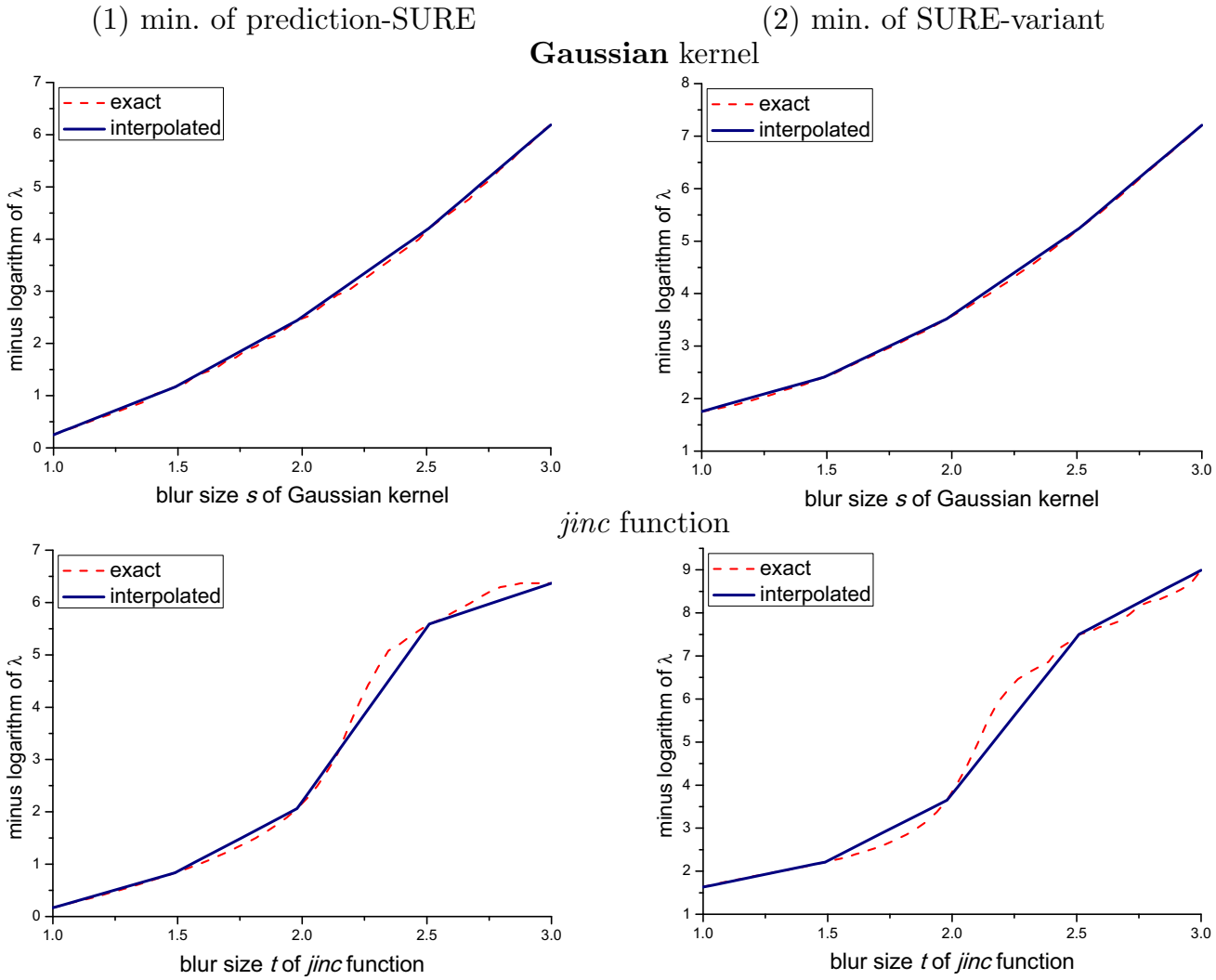


Fig. 9 Linear interpolation of the function between s and $\log_{10}\lambda$, under **Gaussian** and *jinc* kernels: *House*, BSNR = 40 dB

N. In image processing, the variance is very small due to the large number of pixels [18,25].

5.2 Short Summary

As summary, we formulate the PSF estimation as

- minimizing the prediction-SURE over both $\mathbf{H}$ and λ, i.e.,

$$\min_{\mathbf{H},\lambda} \underbrace{\frac{1}{N}\left\|\mathbf{U}_{\mathbf{H},\lambda}\mathbf{y}-\mathbf{y}\right\|^2+\frac{2\sigma^2}{N}\mathrm{Tr}\left(\mathbf{U}_{\mathbf{H},\lambda}+\mathbf{Q}_{\mathbf{H},\lambda}\right)-\sigma^2}_{\text{prediction-SURE: }\epsilon(\mathbf{H},\lambda)} \tag{24}$$

- minimizing the SURE-variant over both $\mathbf{H}$ abd λ

$$\min_{\mathbf{H},\lambda} \underbrace{\frac{\left\|\mathbf{M}_{\mathbf{H},\lambda}\mathbf{y}\right\|^2}{(\mathrm{Tr}(\mathbf{M}_{\mathbf{H},\lambda}))^2}}_{\text{SURE-variant: }\bar{\epsilon}(\mathbf{H},\lambda)} \tag{25}$$

Here, $\mathbf{U}_{\mathbf{H},\lambda}$ and $\mathbf{M}_{\mathbf{H},\lambda}$ are given by (18) and (23), respectively. $\mathbf{Q}_{\mathbf{H},\lambda}$ is specified in Theorem 3. The prediction-SURE/SURE-variant framework is summarized as Fig. 2. In addition, if the PSF is represented by a number of parameters $\mathbf{s}$, the minimizations become with respect to $\mathbf{s}$ and λ.

By Propositions 1 and 2, the solution $H(\omega)$ to (24) or (25) satisfies (11), which implies that the estimation error of $|H_0(\omega)|$ is upper bounded by δ_ω—the approximation error of $\lambda/S_y(\omega)$.

By Theorems 2 and 4, the minimization of prediction-SURE or SURE-variant is essentially matching of exact (complementary) smoother filtering. Also notice that the matching of the exact complementary smoother filtering

$M_0(\omega)$ (minimizing SURE-variant) is equivalent to matching of the exact smoother filtering $U_0(\omega)$ (minimizing prediction-SURE), since $M(\omega) = M_0(\omega)$ implies $1 - M(\omega) = 1 - M_0(\omega)$, i.e., $U(\omega) = U_0(\omega)$ for $\forall\omega$.

6 Experimental Results and Discussions

6.1 Experimental Setting

The prediction-SURE/SURE-variant framework is especially applicable for parametric PSF estimation, where the PSF is given in a specific parametric form with a small number of unknown parameters. Now, we exemplify this approach with two typical parametrized PSF:

- **Gaussian** kernel, with an unknown parameter—blur size s:

$$h(\mathbf{i}; s) = C \cdot \exp\left(-\frac{i_1^2 + i_2^2}{2s^2}\right) \tag{26}$$

- *jinc* function[8], with an unknown scaling factor t:

$$h(\mathbf{i}; t) = C \cdot \left[\frac{2J_1(r/t)}{r/t}\right]^2, \tag{27}$$

where $J_1(\cdot)$ is first-order Bessel function of first kind, the radius $r = \sqrt{i_1^2 + i_2^2}$. The constant C in (26) and (27) is a normalization factor, s.t. $\sum_{\mathbf{i}} h(\mathbf{i}) = 1$.

Gaussian kernel has been widely used in many applications, see [26,34,35] for example. The *jinc* function is often used to describe optical diffraction-limited condition, where the parameter t is related to wave number and aperture diameter of optical imaging system [5].

Here, we stress that the proposed SURE-type criteria are also applicable for other blur kernels, not necessarily circularly symmetric kernels, for example, [33] estimated the blur length and direction of linear motion, based on the prediction-SURE principle.

We perform the following synthetic experiments: the test images displayed in Fig. 3 are blurred by Gaussian kernel (26) with the true $s_0 = 2.0$ and *jinc* function (27) with the true $t_0 = 2.0$, respectively, and subsequently corrupted by white Gaussian noise corresponding to BSNR = 40, 30, 20, 10 dB[9].

[8] The terminology *jinc* is due to the structural similarity to *sinc* function [6].

[9] BSNR (blurr signal-to-noise ratio) is defined as $\text{BSNR} = 10\log_{10}\left(\frac{\|\mathbf{H_0x} - \text{mean}(\mathbf{H_0x})\|^2}{N\sigma^2}\right)$ [24,34].

6.2 Comparisons Between Different Regularization Terms

First, Fig. 4 shows that if we use the exact (complementary) smoother filtering (8) or (20), the PSF is exactly estimated for **Gaussian** kernel, and for *jinc* kernel (omitted here). Thus, Theorem 2 and Theorem 4 have been experimentally verified.

Due to the unknown $\sigma^2/S_x(\omega)$ in the exact (complementary) smoother filtering, Sects. 2 and 4 discussed the various approximations by the regularization term $\lambda R(\omega)$ for $R(\omega) = 1, \omega^2$, or $1/S_y(\omega)$. Now, we are going to show the experimental results and compare the three regularizers.

For *Coco* image, Fig. 5 shows the minimizations of prediction-SURE and the approximations of smoother filtering. Figure 6 shows the minimizations of SURE-variant and the approximations of complementary smoother filtering. We can see that for this simple image, the rapidly changing spectrum $\sigma^2/S_x(\omega)$ cannot be well approximated by a constant λ, whereas the choice $R(\omega) = \omega^2$ or $1/S_y(\omega)$ is more suitable. Hence, the resulted estimations of (17) and (18) are more accurate than that of (16).

Figures 7 and 8 shows the counter-example of *Noise* image. For this complicated image without significant decrease at high frequencies, the slowly varying spectrum $\sigma^2/S_x(\omega)$ is perfectly approximated by a constant λ (16): it produces more accurate results than using $R(\omega) = \omega^2$ or $1/S_y(\omega)$.

Overall, Figs. 4, 5, 6, 7, and 8 show that the following three indicators are closely linked: minimum value of prediction-SURE/SURE-variant, approximation accuracy of (complementary) smoother filtering, and PSF estimation accuracy. Fig. 4 shows the smallest value of minimum prediction-SURE/SURE-variant, as it uses exact (complementary) smoother filtering to obtain the exact PSF estimate. For different regularizers, the smaller value of minimum prediction-SURE/SURE-variant implies the better approximation of (complementary) smoother filtering and the more accurate estimate of PSF. From the results, we can observe that the proposed adaptive regularizer (18) yields consistently better PSF estimations for very different types of images.

6.3 An Efficient Algorithm for the Minimizations (24) and (25)

Considering the prediction-SURE/SURE-variant minimization over two scalar variables (PSF parameter s and regularization parameter λ), the most straightforward way is to perform exhaustive search over all the possible values of s and λ in a certain range. Numerically, if we take 100 discrete points for s and λ to process the image, the exhaustive search requires $100 \times 100 = 10^4$ times of computing the SURE-like functional.

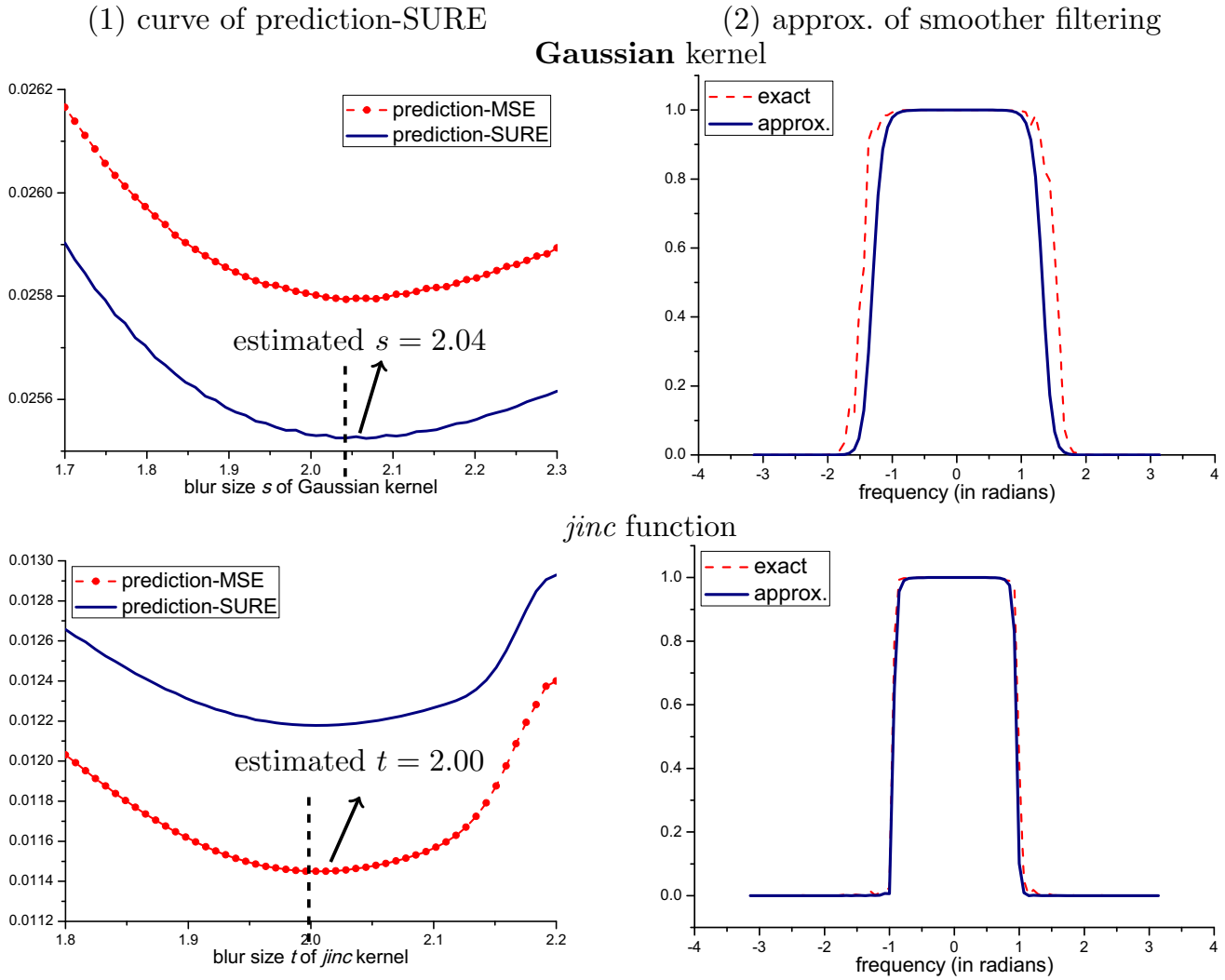


Fig. 10 The prediction-SURE minimization under **Gaussian** and *jinc* kernels: *Coco*, BSNR = 40 dB

The exhaustive search is essentially equivalent to what follows: for each fixed s, obtain the corresponding optimal λ by minimizing the SURE-type estimate, and then, insert the function $\lambda = \lambda(s)$ into the functional. Thus, the minimization has to be performed over only one variable s. Experimentally, we found an approximate piecewise linear relation between the logarithm of λ and the associated optimal blur size s. Thus, it is reasonable to perform a linear interpolation of the function $s = s(\log_{10}\lambda)$ over very few sampling points.

We show the shape of this function (see the dashed curves) and its linear interpolation (see the navy blur curve) in Fig. 9 for the Gaussian and *jinc* kernels. We can observe that the linear interpolation over very few sampling points suffices to approximate the function between s and $-\log_{10}\lambda$. The high approximation accuracy enables us to develop a considerably more efficient algorithm than exhaustive search (see Algorithm 1). Compared to the exhaustive search, Algorithm 1 requires only $5 \times 100 + 100 = 600$ computations.

Algorithm 1 : Algorithm by approximating the function $\lambda = \lambda(s)$

Input: $\epsilon(s, \lambda)$ or $\overline{\epsilon}(s, \lambda)$: objective function given as (24) or (25);
Output: optimal λ and s

1: take five sample-values $s_k = 1.0,\ 1.5,\ 2.0,\ 2.5,\ 3.0$ for $k = 1, 2, ..., 5$;
2: for each s_k, find the optimal λ_k corresponding to the minimum ϵ or $\overline{\epsilon}$:

$$\min_{\lambda_k} \epsilon(s_k, \lambda_k) \quad \text{or} \quad \min_{\lambda_k} \overline{\epsilon}(s_k, \lambda_k)$$

3: use linear interpolation to establish $\lambda = \lambda(s)$, based on the five pairs (s_k, λ_k);
4: incorporate $\lambda = \lambda(s)$ into $\epsilon(s, \lambda)$ or $\overline{\epsilon}(s, \lambda)$, and minimize $\epsilon(s, \lambda(s))$ or $\overline{\epsilon}(s, \lambda(s))$ over s.

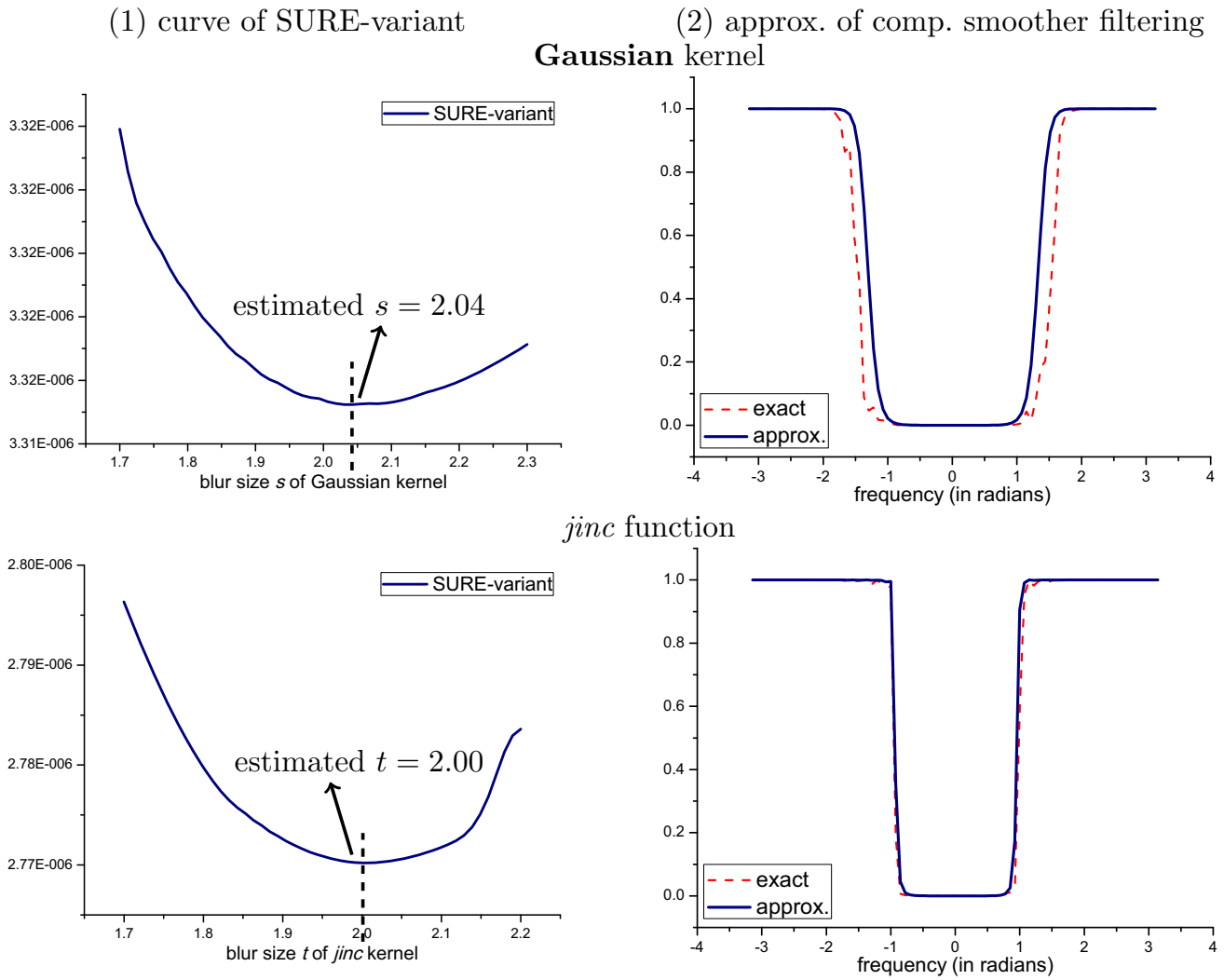


Fig. 11 The SURE-variant minimization under **Gaussian** and *jinc* kernels: *Coco*, BSNR = 40 dB

Finally, we would like to note that thanks to the quadratic nature, the SURE-type functionals can be directly computed in the Fourier domain: there is no need to transform the Fourier coefficients back to the image domain. Thus, the computational complexity is further greatly reduced.

6.4 Experimental Results of Parametric PSF Estimation

In this part, we present the minimizations of the prediction-SURE (24) and the SURE-variant (25), using the proposed adaptive regularizer (18).

Taking *Coco* and BSNR = 40 dB for example, Fig. 10 shows the estimated s and t by the prediction-SURE minimization, and the approximations of smoother filtering, for both Gaussian and *jinc* kernels. We can see that:

- the retrieved $s = 2.04$ and $t = 2.00$ are very close to true value 2.0, as stated in Theorem 2;
- the prediction-SURE is a reliable estimate of prediction-MSE, as stated in Theorem 3;
- the retrieved s and t, with the optimal λ match the exact smoother filtering, where the red dashed curve denotes exact $U_0(\omega)$, the navy blue curve denotes approximated $U_{H,\lambda}(\omega)$.

Figure 11 shows the estimated s and t by the SURE-variant minimization, and the approximations of complementary smoother filtering, for both Gaussian and *jinc* kernels. We can draw the same conclusions as Fig. 10. Theorem 4 has now been verified by this experiment.

Tables 2 and 3 report the complete estimation results by the proposed prediction-SURE (denoted by '→(24)' in tables) and SURE-variant minimization (denoted by '→(25)' in tables) under various noise levels, and present the comparisons with other state-of-the-art methods, including GCV [27], kurtosis [17], and DL1C [10]. To emphasize the necessity of the frequency-adaptive regularizer (18), we also

Table 2 The estimated parameter s of **Gaussian**

Image	*Bridge*			*Coco*			*Crowd*			*Noise*		
BSNR	30	20	10	30	20	10	30	20	10	30	20	10
GCV	2.15	2.15	2.23	2.09	2.08	2.15	1.85	1.82	1.77	1.84	1.77	1.71
Kurtosis	1.95	1.94	1.90	1.92	2.07	2.11	1.96	1.90	1.92	1.84	1.92	1.95
DL1C	2.40	2.55	2.73	2.33	2.55	2.70	1.91	1.87	1.83	1.82	1.89	1.94
(16)	2.45	2.52	2.64	2.43	2.54	2.63	2.30	2.45	2.67	**1.99**	**2.00**	**1.99**
(17)	2.08	2.15	2.23	2.11	2.09	2.11	2.05	2.10	2.13	1.90	1.87	1.84
→ (24)	1.97	**1.98**	**2.02**	**2.03**	1.96	1.94	**1.98**	**2.02**	2.05	1.93	1.94	1.96
→ (25)	**1.98**	**2.02**	2.04	2.04	**1.98**	**2.02**	**1.98**	**1.98**	**2.02**	1.94	1.94	1.94

Table 3 The estimated parameter t of *jinc*

Image	*Bridge*			*Coco*			*Crowd*			*Noise*		
BSNR	30	20	10	30	20	10	30	20	10	30	20	10
GCV	1.94	1.94	**1.95**	1.87	1.89	1.92	1.77	1.82	1.89	1.89	1.91	1.96
Kurtosis	1.80	1.83	1.88	1.74	1.90	1.95	1.85	1.89	1.94	1.81	1.90	1.97
DL1C	2.04	2.18	2.24	2.06	2.07	2.10	1.90	1.87	1.81	1.84	1.89	1.94
(16)	2.11	2.20	2.31	2.15	2.31	2.46	2.05	2.13	2.11	**2.00**	**1.99**	**2.00**
(17)	2.03	2.09	2.12	2.05	2.12	2.21	**2.00**	2.02	2.05	2.01	1.98	1.98
→ (24)	1.95	**2.03**	**2.05**	**2.01**	**2.02**	**2.04**	1.97	1.98	2.01	2.01	**1.99**	1.98
→ (25)	**1.98**	2.05	**2.05**	2.02	2.03	**2.04**	**2.00**	**2.00**	**2.00**	2.02	1.98	1.98

Table 4 PSNR comparison (**Gaussian** kernel with true $s_0 = 2.0$)

Image	*Bridge*			*Coco*			*Crowd*		
BSNR	30	20	10	30	20	10	30	20	10
GCV	33.15	32.26	30.85	33.15	**31.90**	30.16	21.33	20.30	18.94
Kurtosis	33.88	**32.50**	30.95	33.29	**31.91**	**30.23**	21.72	20.47	**19.09**
DL1C	27.32	27.12	27.62	29.15	26.22	26.42	21.57	20.41	19.01
(16)	26.05	27.61	28.47	26.79	26.39	27.10	19.10	17.86	16.99
(17)	33.86	32.26	30.85	32.98	**31.89**	**30.23**	21.75	**20.58**	**19.14**
→ (24)	**33.96**	**32.55**	**31.07**	**33.46**	**31.95**	**30.25**	**21.86**	**20.62**	**19.16**
→ (25)	**33.99**	**32.57**	**31.07**	**33.43**	**31.97**	**30.30**	**21.86**	**20.60**	**19.15**
NBD	*34.02*	*32.57*	*31.07*	*33.49*	*31.99*	*30.35*	*21.89*	*20.62*	*19.16*

Table 5 PSNR comparison (*jinc* kernel with true $t_0 = 2.0$)

Image	*Bridge*			*Coco*			*Crowd*		
BSNR	30	20	10	30	20	10	30	20	10
GCV	31.34	31.01	**30.26**	30.80	30.21	**28.88**	19.00	18.84	18.16
Kurtosis	31.04	30.78	**30.21**	30.05	30.20	**28.88**	19.24	18.96	18.21
DL1C	**31.44**	30.35	29.98	31.15	**30.32**	**28.91**	19.37	18.93	18.08
(16)	30.75	30.11	29.71	28.92	27.10	26.94	**19.50**	18.94	**18.16**
(17)	**31.43**	31.02	**30.23**	31.20	**30.26**	28.80	**19.62**	**19.16**	**18.27**
→ (24)	**31.39**	**31.12**	**30.27**	**31.31**	**30.33**	**28.91**	**19.55**	**19.12**	**18.26**
→ (25)	**31.47**	**31.12**	**30.27**	**31.30**	**30.32**	**28.91**	**19.62**	**19.14**	**18.25**
NBD	*31.50*	*31.12*	*30.28*	*31.31*	*30.34*	*28.82*	*19.62*	*19.16*	*18.28*

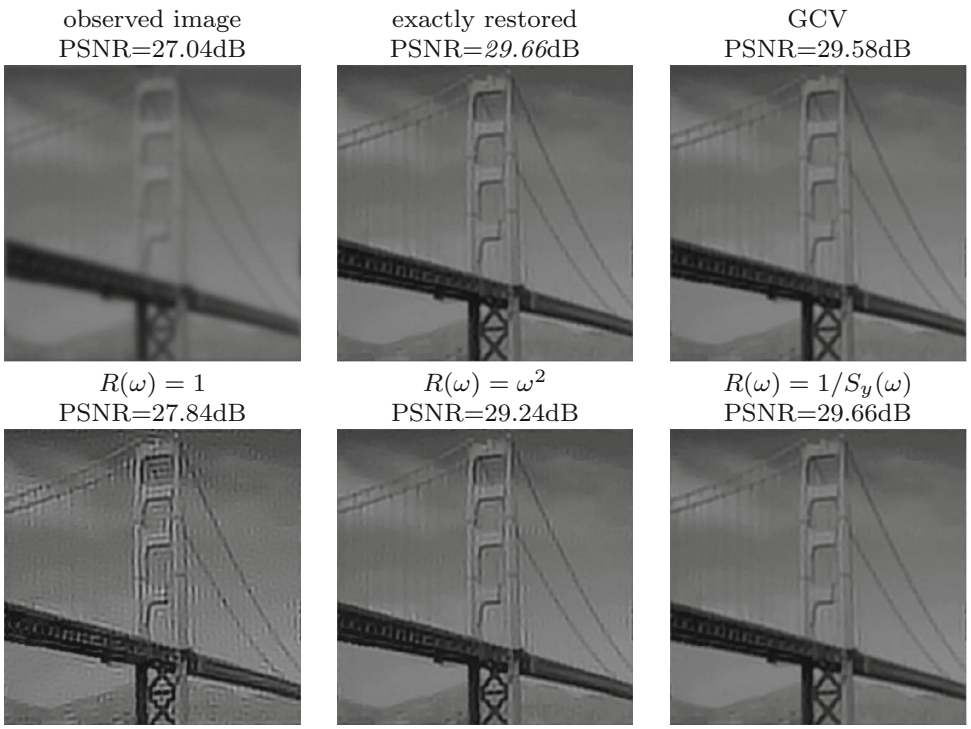


Fig. 12 *Top-left jinc*-blurred image *Bridge* (true $t_0 = 3.0$, BSNR = 20 dB); 'exactly restored' is the non-blind deconvolution with $t_0 = 3.0$. The estimated t by GCV, (16), (17), and (18) are 2.83, 3.45, 3.22, and 3.04, respectively

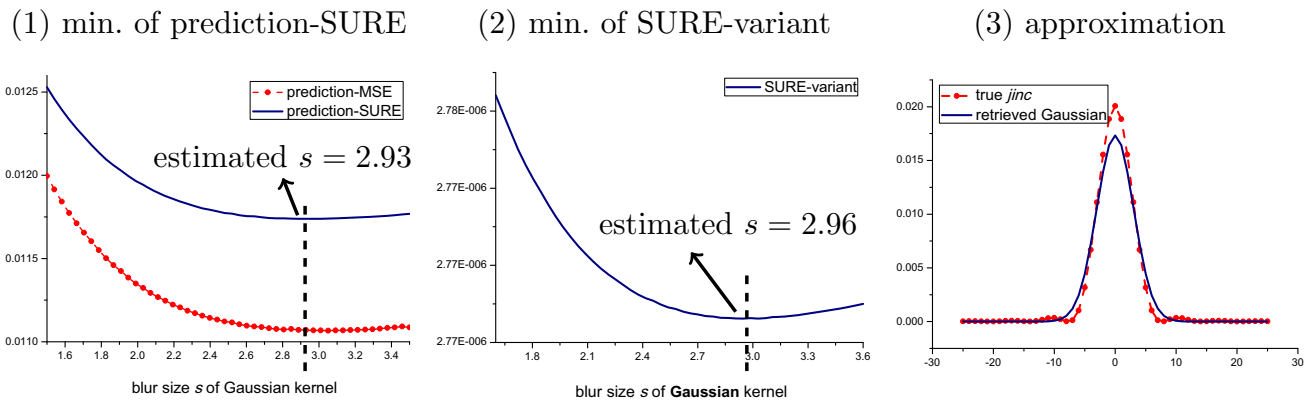


Fig. 13 The prediction-SURE/SURE-variant minimization under *jinc* function with $t_0 = 2.0$ as the true blur: *Crowd*, BSNR = 40 dB

Fig. 14 *Left jinc*-blurred image *Crowd* (BSNR = 30 dB); *Middle* Deconvolution by exact *jinc* function with $t_0 = 2.0$; *Right* Deconvolution by the retrieved Gaussian PSF with $s = 2.93$

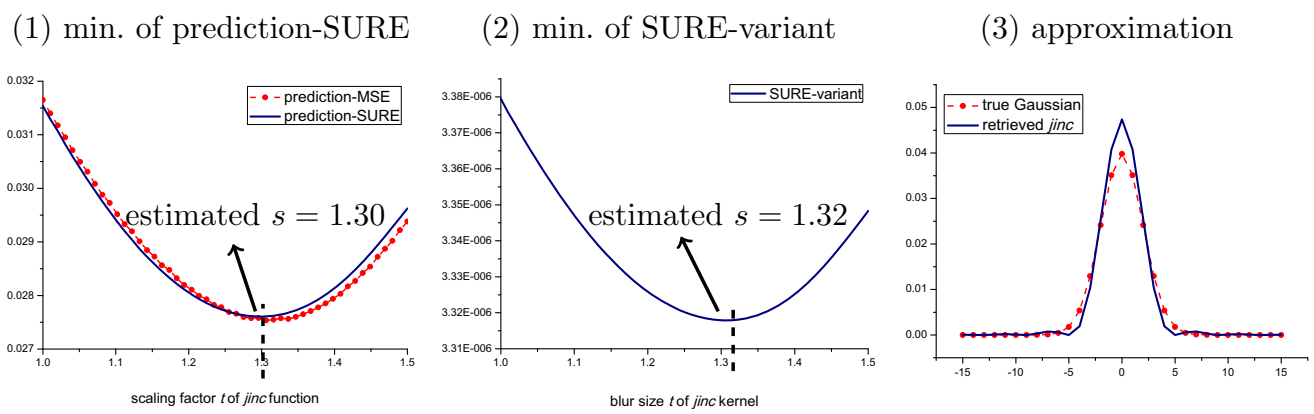


Fig. 15 The prediction-SURE/SURE-variant minimization under Gaussian function with $s_0 = 2.0$ as the true blur: *Coco*, BSNR = 40 dB

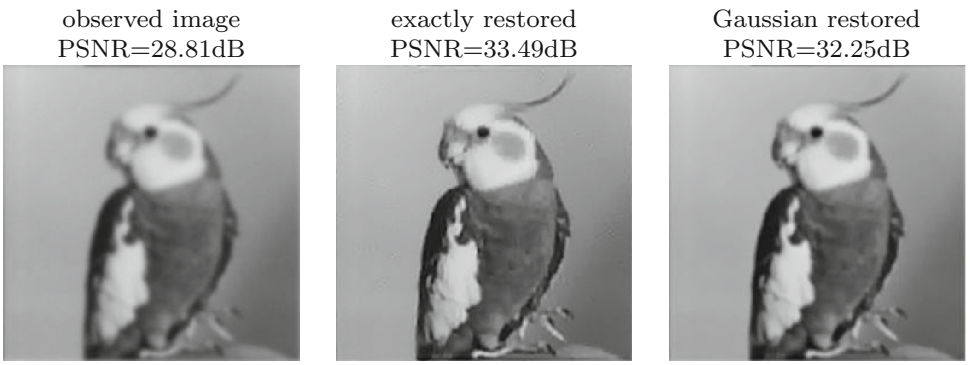


Fig. 16 *Left* **Gaussian**-blurred image *Coco* (BSNR = 30 dB); *Middle* Deconvolution by exact **Gaussian** function with $s_0 = 2.0$; *Right* Deconvolution by the retrieved *jinc* kernel with $s = 1.30$

provide the results using the previous two choices (16) and (17) for comparison. We evaluate the estimation performance by the error $e = |s - s_0|$, and highlight the best results by boldface. We can see that

- Based on the above discussions, there is no surprise that a constant λ as the regularizer in (16) would give the best estimate for *Noise* image.
- Using $R(\omega) = 1$ or ω^2, the estimated parameters often tend to be less accurate under higher noise level (see Fig. 4 for example), whereas $R(\omega) = 1/S_y(\omega)$ yields much more consistent estimates for various noise levels.

Overall, the estimated parameters by $R(\omega) = 1/S_y(\omega)$ are very close to the true values s_0 and t_0, and outperforms the other competitors in average for natural images.

6.5 Influence of Accuracy of PSF Estimation upon Deconvolution Performance

As emphasized above, we separate the PSF estimation and deconvolution. It is important to evaluate the influence of the PSF accuracy upon the deconvolution performance. For fair comparisons in this part, with the estimated PSF by various methods, we always perform the same BM3D deconvolution algorithm [11]. Thus, the deconvolution quality completely depends on the PSF estimation. We evaluate the deconvolution performance in terms of peak signal-to-noise ratio (PSNR):

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{255^2}{\|\widehat{\mathbf{x}} - \mathbf{x}\|^2 / N} \right),$$

where $\widehat{\mathbf{x}}$ is the image deconvolved using the estimated PSF parameters.

Tables 4 and 5 show the deconvolution results and comparisons, where the best PSNR results within a 0.1 dB margin are highlighted. We can see that the estimated blur kernels by the proposed methods achieve better deconvolution performance. Moreover, in most cases, our blind deconvolution achieves a PSNR gain that is within 0.2 dB of the gain of the non-blind algorithm [11] (reported in italic number of '*NBD*' rows in Tables 4 and 5, where the PSF is exactly known). Figure 12 shows a visual example of a severely blurring case, where the estimated blur size $t = 3.04$ by the SURE-variant minimization. This is very similar to the result of prediction-SURE, and thus, we omit it here.

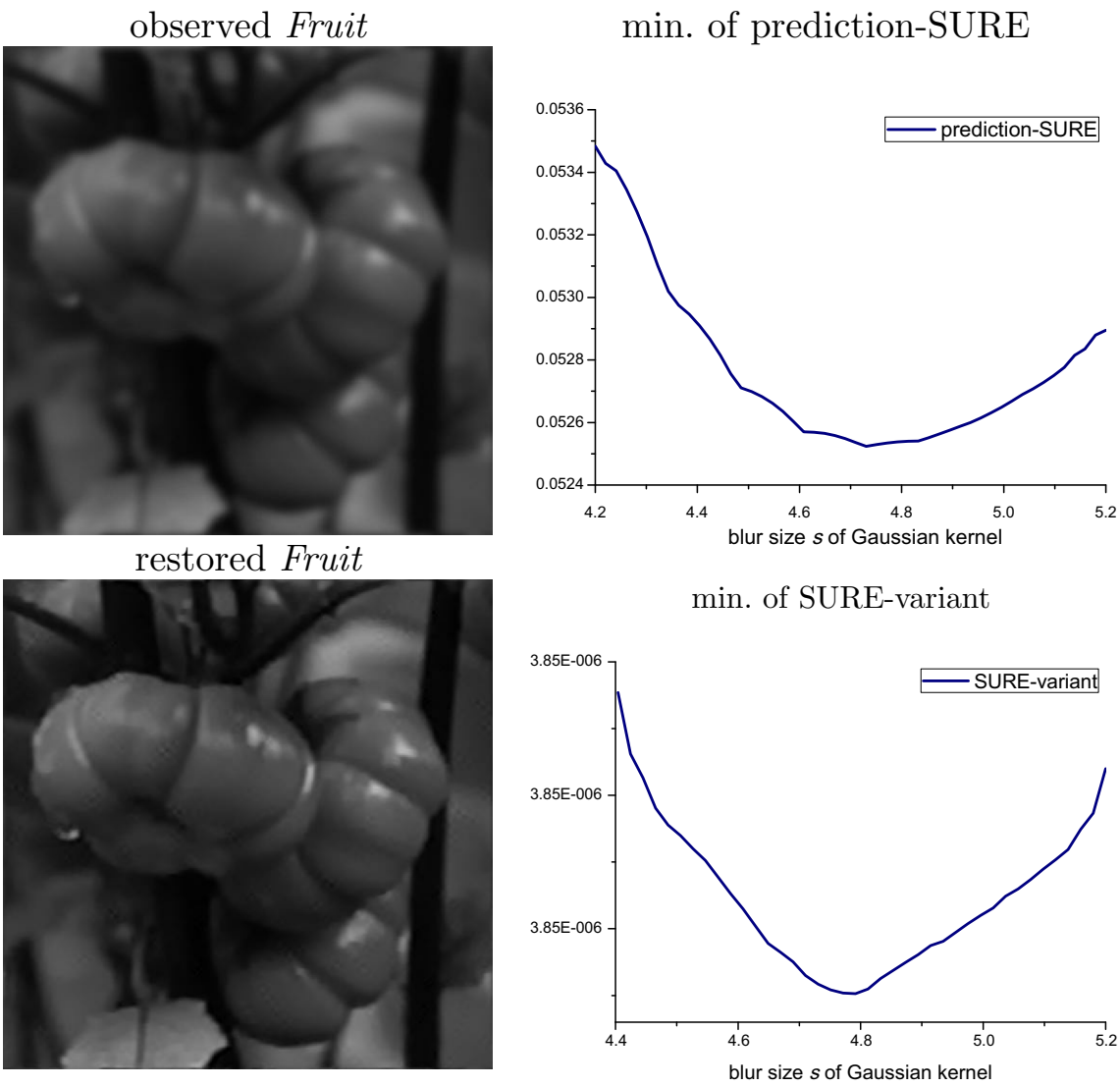


Fig. 17 The restoration of real image: *Fruit*

6.6 Cross-Validation Experiments

The proposed approach is not limited to any specific forms of PSF. To demonstrate this robustness to the PSF assumption, we now perform the cross-validation experiments. For example, in the diffraction-limited condition, where the underlying true PSF is *jinc* function, we assume Gaussian kernel as the retrieved PSF, and estimate the blur size s from the *jinc*-blurred image. Fig. 13-(1), (2) show the minimizations of prediction-SURE and SURE-variant for this case; Fig. 13-(3) shows that the retrieved Gaussian kernel with $s = 2.93$ is also a good approximation of the *jinc* function with $t_0 = 2.0$.

Furthermore, it is also interesting to evaluate the approximation accuracy of the retrieved Gaussian kernel (26) to the true *jinc* function (27) (assuming known in synthetic experiments). Hence, we perform the following approximations for the comparison:

1. Least-square fitting of **Gaussian** kernel to *jinc* function with t_0:

$$s_1 = \arg\min_s \sum_{\mathbf{i}} \left| h(\mathbf{i}; s) - h(\mathbf{i}; t_0) \right|^2.$$

2. Gaussian deconvolution of the *jinc*-blurred image with minimum standard MSE:

$$s_2 = \arg\min_s \|\widehat{\mathbf{x}}_s - \mathbf{x}\|^2,$$

where $\widehat{\mathbf{x}}_s$ denotes the restored image by the BM3D algorithm [11], using Gaussian kernel with size s.

For *jinc* function with $t_0 = 2.0$ (the case of Fig. 13), we obtain that $s_1 = 2.85$ and $s_2 = 2.88$, which are very close to the blindly estimated $s = 2.93$ and 2.96 by the SURE-type minimizations. It implies that the estimated blur

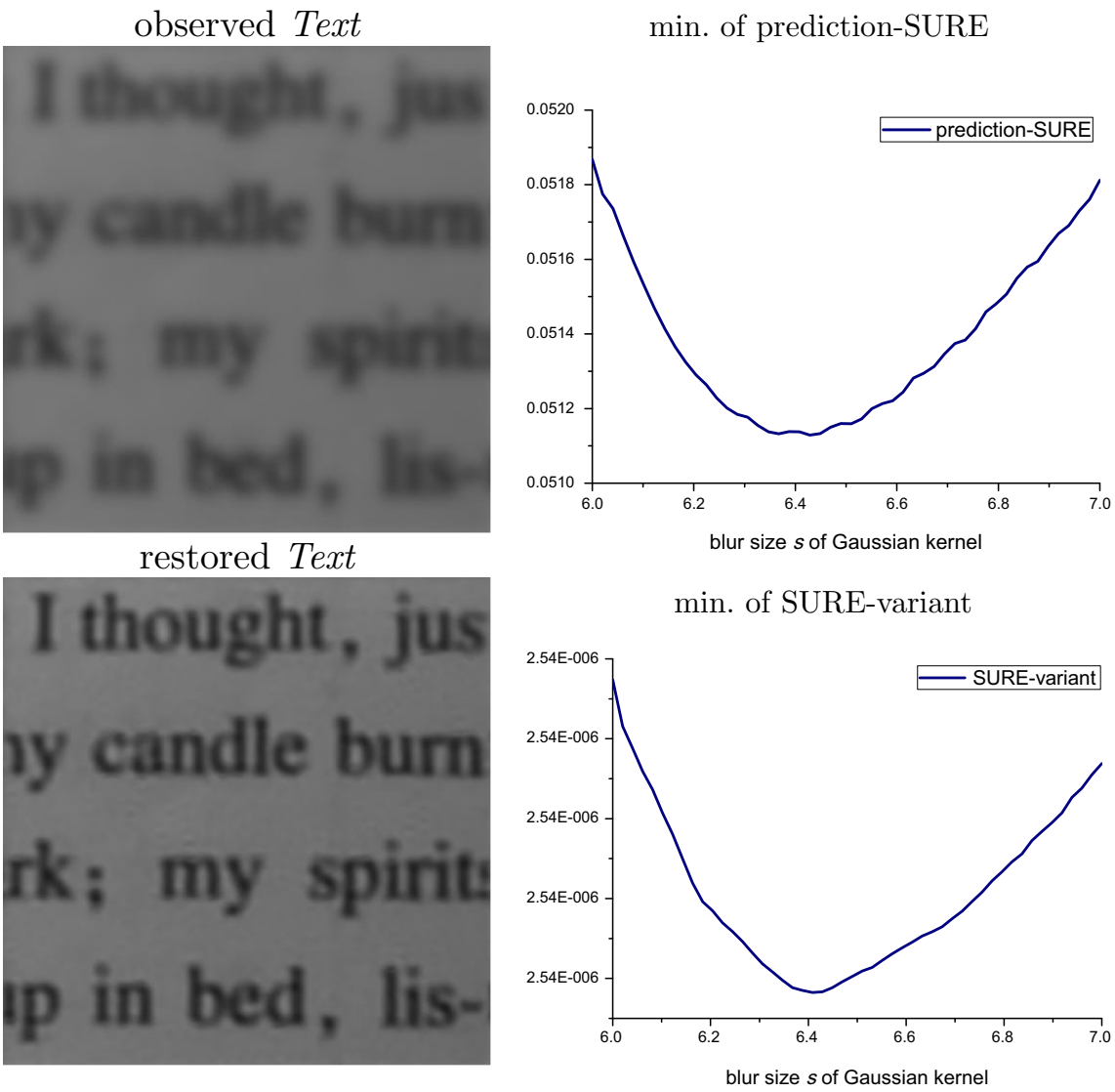


Fig. 18 The restoration of real image: *Text*

size by the SURE-type minimizations provides the nearly optimal approximations of **Gaussian** kernel to true *jinc* function, and also yields the best deconvolution performance using Gaussian kernel. Figure 14 shows a visual comparison of deconvolution performances by BM3D algorithm [11]. Though the blindly restored image using Gaussian kernel is not satisfactory (the PSNR difference w.r.t. the exact deconvolution is almost 1dB), it has already been the best performance reached by Gaussian kernel.

Figure 15 shows the converse validation: assuming *jinc* function as the retrieved PSF, to deconvolve the Gaussian blurred image. We also found that the estimated *jinc* function with $t = 1.30$ is a good approximation of Gaussian kernel with $s_0 = 2.0$. Figure 16 shows the corresponding visual example. We can draw the similar conclusions to Figs. 13 and 14.

As summary, the above cross-validation experiments imply that the SURE-type minimizations behave like the accurate approximation of the assumed PSF to the underlying true PSF, even if the two PSF have different parametric forms. In other words, the SURE-type minimizations naturally find the closest blur kernel to the underlying true one, within the presumed function of PSF, no matter both the retrieved and true PSF belong to the same function family or not.

6.7 Application to Real Images

In our last set of experiments, we perform experiments with two real camera-captured images: *Fruit* and *Text*, shown in Figs. 17 and 18. For both real images, we assume the underlying PSF as Gaussian function. For *Fruit*, the estimated blur

size is 4.72 by prediction-SURE and 4.78 by SURE-variant. For *Text*, the estimated result is 6.43 and 6.41 by the two criteria, respectively. Then, we apply BM3D algorithm [11] to perform deconvolution, with the estimated PSF. Figures 17 and 18 shows the restorations with the estimated PSF by prediction-SURE, which achieved great improvements of visual quality. Since the estimation results by the two criteria are very similar, we omit the SURE-variant results here.

7 Conclusions

In this paper, we proposed two novel criteria for PSF estimation—prediction-SURE and SURE-variant. We have shown that based on a linear smoother filtering with a frequency-adaptive regularization term, the SURE-type minimizations yield highly accurate parametric PSF estimation, and outperform other regularizers. In addition, one can readily apply the SURE-variant without the knowledge of noise variance.

The examples of the blur kernel listed in the paper are but the exemplifications of the proposed framework for parametric PSF estimation. It is worth noting that the SURE-type minimization itself, which does not specify any particular parametric form of PSF as shown in (24) and (25), is applicable for a great variety of PSF models. Furthermore, the cross-validation experiments showed that the proposed framework is robust to the PSF assumption: it naturally seeks the best approximation of the underlying true PSF, if the retrieved function belongs to different family of the true one. There is huge potential to develop specific algorithms for various applications, based on the SURE-type criterion.

Acknowledgments This work was supported by the National Natural Science Foundation of China (No. 61401013). The authors would like to thank the coordinating editor and anonymous reviewers for their insightful comments, and also obliged to Hanjie PAN for many useful discussions on this paper.

Appendix 1: Proof of Theorem 3

Proof Based on Lemma 1, the prediction-SURE can be written as

$$\epsilon = \frac{1}{N}\left\|\mathbf{U}\mathbf{y}-\mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\mathrm{div}_{\mathbf{y}}\left(\mathbf{U}\mathbf{y}\right) - \sigma^2.$$

We omit the subscript $\mathbf{U}_{\mathbf{H},\lambda}$ as $\mathbf{U}$ for brevity in this proof.

Now, we are going to compute the divergence term—$\mathrm{div}_{\mathbf{y}}\left(\mathbf{U}\mathbf{y}\right)$. Notice that the matrix $\mathbf{U}$ is the function of $\mathbf{y}$, hence, $\mathrm{div}_{\mathbf{y}}\left(\mathbf{U}\mathbf{y}\right) \neq \mathrm{Tr}(\mathbf{U})$.

First, by definition of divergence term, we have

$$\mathrm{div}_{\mathbf{y}}\left(\mathbf{U}\mathbf{y}\right) = \mathrm{Tr}(\mathbf{U}) + \underbrace{\left(\frac{\partial \mathrm{diag}(\mathbf{U})}{\partial \mathbf{y}}\right)^{\mathrm{T}}}_{\alpha}\mathbf{y}, \tag{28}$$

where the vector $\mathrm{diag}(\mathbf{U}) \in \mathbf{R}^N$ consists of the diagonal element $U_{n,n}$ of matrix $\mathbf{U}$:

$$\frac{\partial \mathrm{diag}(\mathbf{U})}{\partial \mathbf{y}} = \left[\frac{\mathrm{d}U_{1,1}(\mathbf{y})}{\mathrm{d}y(0)}, \frac{\mathrm{d}U_{2,2}(\mathbf{y})}{\mathrm{d}y(1)}, \ldots, \frac{\mathrm{d}U_{N,N}(\mathbf{y})}{\mathrm{d}y(N-1)}\right]^{\mathrm{T}}$$

Notice that under periodic boundary condition, $\mathbf{U}$ is circulant matrix, whose diagonal element is a constant given by

$$\begin{aligned} U_{n,n}(\mathbf{y}) &= \underbrace{u_N(0)}_{\text{filter}} \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\underbrace{\frac{|H(k)|^2}{|H(k)|^2+\frac{\lambda}{|Y(k)|^2}}}_{U(k)}, \qquad \text{for } n = 0, 1, ..., N-1. \end{aligned}$$

The second equality is from inverse Fourier transform. Now, we consider the n-th element of α:

$$\begin{aligned} \alpha_n &= \frac{\partial U_{n,n}(\mathbf{y})}{\partial y(n)} = \frac{\partial U_{n,n}(\mathbf{y})}{\partial |Y(k)|^2}\cdot\frac{\partial |Y(k)|^2}{\partial y(n)} \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\frac{\partial \frac{|H(k)|^2}{|H(k)|^2+\frac{\lambda}{|Y(k)|^2}}}{\partial |Y(k)|^2}\cdot\frac{\partial |Y(k)|^2}{\partial y(n)}, \end{aligned}$$

where

$$\frac{\partial\left(\frac{|H(k)|^2}{|H(k)|^2+\frac{\lambda}{|Y(k)|^2}}\right)}{\partial |Y(k)|^2} = \frac{|H(k)|^2\lambda}{\left(|H(k)|^2+\frac{\lambda}{|Y(k)|^2}\right)^2\cdot|Y(k)|^4}$$

and for any fixed n:

$$\begin{aligned} \frac{\partial |Y(k)|^2}{\partial y(n)} &= \frac{\partial\left|\sum_{l=0}^{N-1} y(l)e^{-j\frac{2\pi kl}{N}}\right|^2}{\partial y(n)} \\ &= \frac{\partial\left(\left|\sum_{l=0}^{N-1} y(l)\cos\frac{2\pi kl}{N}\right|^2 + \left|\sum_{l=0}^{N-1} y(l)\sin\frac{2\pi kl}{N}\right|^2\right)}{\partial y(n)}, \end{aligned}$$

where the two terms are

$$\frac{\partial\left(\sum_{l=0}^{N-1} y(l)\cos\frac{2\pi kl}{N}\right)^2}{\partial y(n)} = 2\cos\frac{2\pi nk}{N}\left(\sum_{l=0}^{N-1} y(l)\cos\frac{2\pi lk}{N}\right) = 2\cos\frac{2\pi nk}{N}\mathcal{R}\{Y(k)\}$$

and

$$\frac{\partial\left(\sum_{l=0}^{N-1} y(l)\sin\frac{2\pi kl}{N}\right)^2}{\partial y(n)} = 2\sin\frac{2\pi nk}{N}\left(\sum_{l=0}^{N-1} y(l)\sin\frac{2\pi lk}{N}\right) = -2\sin\frac{2\pi nk}{N}\mathcal{I}\{Y(k)\}$$

Hence, α_n becomes

$$\begin{aligned}\alpha_n &= \frac{2}{N}\sum_{k=0}^{N-1}\underbrace{\frac{|H(k)|^2\lambda}{\left(|H(k)|^2+\frac{\lambda}{|Y(k)|^2}\right)^2\cdot|Y(k)|^4}}_{V(k)\text{:real-valued}}\cdot\left(\mathcal{R}\{Y(k)\}\cos\frac{2\pi nk}{N}-\mathcal{I}\{Y(k)\}\sin\frac{2\pi nk}{N}\right)\\ &= \frac{2}{N}\sum_{k=0}^{N-1}\mathcal{R}\{\underbrace{V(k)Y(k)}_{G(k)}\}\cos\frac{2\pi nk}{N}-\frac{2}{N}\sum_{k=0}^{N-1}\mathcal{I}\{\underbrace{V(k)Y(k)}_{G(k)}\}\sin\frac{2\pi nk}{N}\\ &= 2\mathcal{R}\{g(n)\},\end{aligned}$$

where $g(n)$ is the inverse Fourier transform of $G(k) = V(k)Y(k)$: $g(n) = \frac{1}{N}\sum_{k=0}^{N-1} G(k)e^{j\frac{2\pi nk}{N}}$. Since $V(k)$ is real-valued, and $Y(k)$ is Hermitian symmetric, $g(n)$ is real-valued. Thus, we have $\alpha_n = 2\mathcal{R}\{g(n)\} = 2g(n)$.

Finally, the second term of (28) becomes

$$\begin{aligned}\alpha^{\mathrm{T}}\mathbf{y} &= 2(\mathbf{F}G)^{\mathrm{T}}\mathbf{F}Y = 2G^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}Y = 2G^{\mathrm{T}}Y\\ &= \sum_{k=0}^{N-1} V(k)Y^*(k)Y(k) = \sum_{k=0}^{N-1} V(k)|Y(k)|^2\\ &= \sum_{k=0}^{N-1}\underbrace{\frac{|H(k)|^2\lambda}{\left(|H(k)|^2+\frac{\lambda}{|Y(k)|^2}\right)^2\cdot|Y(k)|^2}}_{Q(k)},\end{aligned}$$

where $\mathbf{F}$ is 1-D DFT matrix. The proof is completed. □

Appendix 2: Proof of Corollary 1

Proof The proof is very similar to that of Theorem 3 (see Appendix 1). The only difference from one-dimensional case is how to compute $\frac{\partial|Y(k_1,k_2)|^2}{\partial y(n_1,n_2)}$:

$$\frac{\partial|Y(k_1,k_2)|^2}{\partial y(n_1,n_2)} = \frac{\partial\left|\sum_{l_1=0}^{M-1}\sum_{l_2=0}^{N-1} y(l_1,l_2)e^{-j\frac{2\pi k_1 l_1}{M}}e^{-j\frac{2\pi k_2 l_2}{N}}\right|^2}{\partial y(n_1,n_2)}.$$

The numerator is

$$\begin{aligned}&\left|\sum_{l_1=0}^{M-1}\sum_{l_2=0}^{N-1} y(l_1,l_2)e^{-j\frac{2\pi k_1 l_1}{M}}e^{-j\frac{2\pi k_2 l_2}{N}}\right|^2\\ &= \left[\sum_{l_1=0}^{M-1}\sum_{l_2=0}^{N-1} y(l_1,l_2)\Big(\underbrace{\cos\frac{2\pi k_1 l_1}{M}\cos\frac{2\pi k_2 l_2}{N}-\sin\frac{2\pi k_1 l_1}{M}\sin\frac{2\pi k_2 l_2}{N}}_{A(l_1,l_2)}\Big)\right]^2\\ &\quad+\left[\sum_{l_1=0}^{M-1}\sum_{l_2=0}^{N-1} y(l_1,l_2)\Big(\underbrace{\cos\frac{2\pi k_1 l_1}{M}\sin\frac{2\pi k_2 l_2}{N}+\sin\frac{2\pi k_1 l_1}{M}\cos\frac{2\pi k_2 l_2}{N}}_{B(l_1,l_2)}\Big)\right]^2.\end{aligned} \tag{29}$$

The derivatives of both the terms of (29) are

$$\begin{aligned}&\frac{\partial\left[\sum_{l_1=0}^{M-1}\sum_{l_2=0}^{N-1} y(l_1,l_2)A(l_1,l_2)\right]^2}{\partial y(n_1,n_2)}\\ &= 2\Big(\underbrace{\sum_{l_1=0}^{M-1}\sum_{l_2=0}^{N-1} y(l_1,l_2)A(l_1,l_2)}_{\mathcal{R}\{Y(k_1,k_2)\}}\Big)\times\Big(\underbrace{\cos\frac{2\pi k_1 n_1}{M}\cos\frac{2\pi k_2 n_2}{N}-\sin\frac{2\pi k_1 n_1}{M}\sin\frac{2\pi k_2 n_2}{N}}_{A(n_1,n_2)}\Big)\end{aligned}$$

and

$$\frac{\partial\Big[\sum_{l_1=0}^{M-1}\sum_{l_2=0}^{N-1} y(l_1,l_2)B(l_1,l_2)\Big]^2}{\partial y(n_1,n_2)} = 2\Big(\underbrace{\sum_{l_1=0}^{M-1}\sum_{l_2=0}^{N-1} y(l_1,l_2)B(l_1,l_2)}_{-\mathcal{I}\{Y(k_1,k_2)\}}\Big) \times\Big(\underbrace{\cos\frac{2\pi k_1 n_1}{M}\sin\frac{2\pi k_2 n_2}{N}+\sin\frac{2\pi k_1 n_1}{M}\cos\frac{2\pi k_2 n_2}{N}}_{B(n_1,n_2)}\Big).$$

Hence, α_{n_1,n_2} is

$$\begin{aligned}\alpha_{n_1,n_2} &= \frac{2}{MN}\sum_{k_1=0}^{M-1}\sum_{k_2=0}^{N-1}\underbrace{\frac{|H(k_1,k_2)|^2\lambda}{\Big(|H(k_1,k_2)|^2+\frac{\lambda}{|Y(k_1,k_2)|^2}\Big)^2\cdot|Y(k_1,k_2)|^4}}_{V(k_1,k_2)\text{:real-valued}} \\ &\quad\cdot\Big(\mathcal{R}\{Y(k_1,k_2)\}A(n_1,n_2)-\mathcal{I}\{Y(k_1,k_2)\}B(n_1,n_2)\Big) \\ &= \frac{2}{MN}\sum_{k_1=0}^{M-1}\sum_{k_2=0}^{N-1}\mathcal{R}\Big\{\underbrace{V(k_1,k_2)Y(k_1,k_2)}_{G(k_1,k_2)}\Big\}A(n_1,n_2) \\ &\quad-\frac{2}{MN}\sum_{k_1=0}^{M-1}\sum_{k_2=0}^{N-1}\mathcal{I}\Big\{\underbrace{V(k_1,k_2)Y(k_1,k_2)}_{G(k_1,k_2)}\Big\}B(n_1,n_2) \\ &= 2\mathcal{R}\big\{g(n_1,n_2)\big\},\end{aligned}$$

where $g(n_1,n_2)$ is the inverse Fourier transform of $G(k_1,k_2) = V(k_1,k_2)Y(k_1,k_2)$: $g(n_1,n_2) = \frac{1}{MN}\sum_{k_1=0}^{M-1}\sum_{k_2=0}^{N-1} G(k_1,k_2)e^{j\frac{2\pi n_1 k_1}{M}}e^{j\frac{2\pi n_2 k_2}{N}}$. Since $V(k_1,k_2)$ is real-valued, and $Y(k_1,k_2)$ is Hermitian symmetric, $g(n_1,n_2)$ is real-valued. Thus, we have $\alpha_{n_1,n_2} = 2\mathcal{R}\{g(n_1,n_2)\} = 2g(n_1,n_2)$.

Finally, the second term of (28) becomes

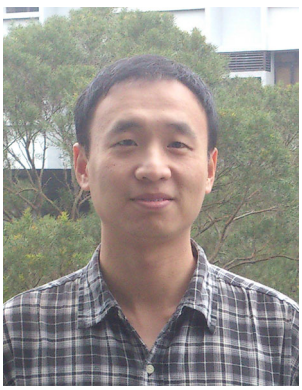
$$\begin{aligned}\alpha^{\mathrm{T}}\mathbf{y} &= 2(\mathbf{F}G)^{\mathrm{T}}\mathbf{F}Y = 2G^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}Y = 2G^{\mathrm{T}}Y \\ &= \sum_{k_1=0}^{M-1}\sum_{k_2=0}^{N-1}V(k_1,k_2)Y^*(k_1,k_2)Y(k_1,k_2) \\ &= \sum_{k_1=0}^{M-1}\sum_{k_2=0}^{N-1}V(k_1,k_2)|Y(k_1,k_2)|^2 \\ &= \sum_{k_1=0}^{M-1}\sum_{k_2=0}^{N-1}\underbrace{\frac{|H(k_1,k_2)|^2\lambda}{\Big(|H(k_1,k_2)|^2+\frac{\lambda}{|Y(k_1,k_2)|^2}\Big)^2\cdot|Y(k_1,k_2)|^2}}_{Q(k_1,k_2)},\end{aligned}$$

where $\mathbf{F}$ is 2-D DFT matrix. The proof is completed. □

References

1. Allen, D.M.: Mean square error of prediction as a criterion for selecting variables. Technometrics **13**(3), 469–475 (1971)
2. Babacan, S., Molina, R., Katsaggelos, A.: Variational Bayesian blind deconvolution using a total variation prior. IEEE Trans. Image Process. **18**(1), 12–26 (2009)
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. In: Blind image deconvolution: theory and applications, pp. 1–23. CRC Press (2007)
5. Born, M., Wolf, E., Bhatia, A.: Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. Cambridge University Press, Cambridge (1999)
6. Cao, Q.: Generalized Jinc functions and their application to focusing and diffraction of circular apertures. J. Opt. Soc. Am. A **20**(4), 661–667 (2003)
7. Carasso, A.: Linear and nonlinear image deblurring: a documented study. SIAM J. Numer. Anal. **36**, 1659–1689 (1999)
8. Carasso, A.: The APEX method in image sharpening and the use of low exponent Lévy stable laws. SIAM J. Appl. Math. **63**, 593–618 (2002)
9. Chan, T., Wong, C.: Total variation blind deconvolution. IEEE Trans. Image Process. **7**(3), 370–375 (1998)
10. Chen, F., Ma, J.: An empirical identification method of Gaussian blur parameter for image deblurring. IEEE Trans. Signal Process. **57**(7), 2467–2478 (2009)
11. Danielyan, A., Katkovnik, V., Egiazarian, K.: BM3D frames and variational image deblurring. IEEE Trans. Image Process. **21**(4), 1715–1728 (2012)
12. Donoho, D., Johnstone, J.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**(3), 425–455 (1994)
13. Eldar, Y.: Generalized SURE for exponential families: applications to regularization. IEEE Trans. Signal Process. **57**(2), 471–481 (2009)
14. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning, vol. 2. Springer, New York (2009)
15. Kundur, D., Hatzinakos, D.: Blind image deconvolution. IEEE Signal Process. Mag. **13**(3), 43–64 (1996)
16. Liao, H., Ng, M.K.: Blind deconvolution using generalized cross-validation approach to regularization parameter estimation. IEEE Trans. Image Process. **20**(3), 3005–3019 (2011)
17. Li, D., Simske, S.: Atmospheric turbulence degraded-image restoration by kurtosis minimization. IEEE Geosci. Remote Sens. Lett. **6**(2), 244–247 (2009)
18. Luisier, F.: The SURE-LET approach to image denoising. Ph.D. Thesis, École Polytechnique Fédérale de Lausanne (2010)
19. Michailovich, O., Tannenbaum, A.: Blind deconvolution of medical ultrasound images: a parametric inverse filtering approach. IEEE Trans. Image Process. **16**(12), 3005–3019 (2007)
20. Moffat, A.: A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry. Astron. Astrophys. **3**, 455–461 (1969)
21. Molina, R., Núñez, J., Cortijo, F., Mateos, J.: Image restoration in astronomy: a Bayesian perspective. IEEE Signal Process. Mag. **18**(2), 11–29 (2001)
22. Molina, R., Mateos, J., Katsaggelos, A.: Blind deconvolution using a variational approach to parameter, image, and blur estimation. IEEE Trans. Image Process. **15**(12), 3715–3727 (2006)
23. 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)

24. 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)
25. Pesquet, J., Benazza-Benyahia, A., Chaux, C.: A SURE approach for digital signal/image deconvolution problems. IEEE Trans. Signal Process. **57**(12), 4616–4632 (2009)
26. Poropat, G.: Effect of system point spread function, apparent size, and detector instantaneous field of view on the infrared image contrast of small objects. Opt. Eng. **32**(10), 2598–2607 (1993)
27. Reeves, S.J., Mersereau, R.M.: Blur identification by the method of generalized cross-validation. IEEE Trans. Image Process. **1**(3), 301–311 (1992)
28. Sarder, P., Nehorai, A.: Deconvolution methods for 3-D fluorescence microscopy images. IEEE Signal Process. Mag. **23**(3), 32–45 (2006)
29. Stein, C.: Estimation of the mean of a multivariate normal distribution. Ann. Stat. **9**(6), 1135–1151 (1981)
30. Vaiter, S., Deledalle, C., Peyré, G., Fadili, J., Dossal, C.: Local behavior of sparse analysis regularization: applications to risk estimation. Appl. Comput. Harmon. Anal. **35**(3), 433–451 (2012)
31. Vonesch, C., Ramani, S., Unser, M.: Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint. In: 15th IEEE International Conference on Image Processing, pp. 665–668 (2008)
32. Xue, F., Blu, T.: SURE-based blind Gaussian deconvolution. In: 2012 IEEE Statistical Signal Processing Workshop (SSP), pp. 452–455. Ann Arbor, MI, USA, Aug 2012
33. Xue, F., Blu, T.: SURE-based motion estimation. In: 2012 IEEE International Conference on Signal Processing, Communications and Computing, pp. 373–377. Hong Kong, Aug 2012
34. Xue, F., Blu, T.: A novel SURE-based criterion for parametric PSF estimation. IEEE Trans. Image Process. **24**(2), 595–607 (2015)
35. Xue, F., Yagola, A.G.: Analysis of point-target detection performance based on ATF and TSF. Infrared Phys. Technol. **52**(5), 166–173 (2009)



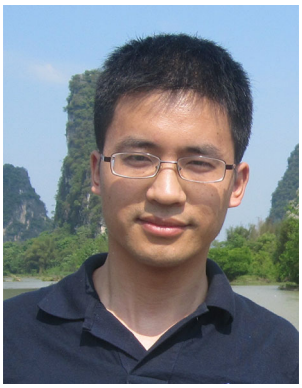
Feng Xue was born in Shandong, China, in 1983. He received his Bachelor degree in electronic science and technology and Master degree in optical engineering from Harbin Institute of Technology, Harbin, China, in 2005 and 2007, respectively. In June 2013, he obtained his Ph.D. degree in electronic engineering from The Chinese University of Hong Kong, Hong Kong. He is now with the National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics, Beijing, China. His research interests include image deconvolution, sparse reconstruction, and statistical learning.

Jiaqi Liu was born in Hunan, China, in 1963. In March 2007, he received his Ph.D. degree of Circuit and Systems from Beihang University, Beijing, China. His research area is signal processing and target recognition. He currently serves as vice director and leading research fellow of National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics.

Shenghai Jiao was born in Shandong, China, in 1976. He received his Master degree from China Academy of Launch Vehicle Technology, Beijing, China, in 2002. Now he is a senior engineer in the National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics. His research interest is design of light vehicles.

Shengdong Liu was born in Hunan, China, in 1977. He received his Master degree in Engineering from Beijing Institute of Technology, Beijing, China, in 2003. Now he is a senior engineer in the National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics. His research interests include image processing, object detection, and tracking.

Min Zhao was born in Shandong, China, in 1965. In August 2005, he received his Ph.D. degree from China Academy of Launch Vehicle Technology, Beijing, China. His research interests include system design of flight vehicles and the related engineering analysis. He has a long professional career in designing launch vehicles, and is now a chief engineer of China Academy of Launch Vehicle Technology.



Zhenhong Niu was born in Handan, China, in 1979. He received his Master degree in Engineering from Graduate University of Chinese Academy of Sciences, Beijing, China, in 2006. Now he is a senior engineer in the National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics. His research interests consist of physics and electronic system design.