ELSEVIER

Signal Processing 82 (2002) 1773–1787

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

A unified method for optimizing linear image restoration filters

Masashi Sugiyama *, Hidemitsu Ogawa

Department of Computer Science, Tokyo Institute of Technology, 2-12-1, O-okayama, Meguro-ku, Tokyo 152-8552, Japan

Abstract

Image restoration from degraded images lies at the foundation of image processing, pattern recognition, and computer vision, so it has been extensively studied. A large number of image restoration filters have been devised so far. It is known that a certain filter works excellently for a certain type of original image or degradation. However, the same filter may not be suitable for other images, so the selection of filters is exceedingly important in practice. Moreover, if a filter includes adjustable parameters such as the regularization parameter or threshold, its restoration performance relies heavily on the choice of the parameter values. In this paper, we therefore discuss the problem of optimizing the filter type and parameter values. Our method is based on the subspace information criterion (SIC), which is an unbiased estimator of the expected squared error between the restored and original images. Since SIC is applicable to any linear filters, one can optimize the filter type and parameter values in a consistent fashion. Our emphasis in this article is laid on the practical concerns of SIC, such as the noise variance estimation, computational issues, and comparison with existing methods. Specifically, we derive an analytic form of the optimal parameter values for the moving-average filter, which will greatly reduce the computational cost. Experiments with the regularization filter show that SIC is comparable to existing methods in the small degradation case, and SIC tends to outperform existing methods in the severe degradation case.

Keywords: Image restoration; Denoising; Linear filter; Subspace information criterion (SIC); Moving-average filter

1. Introduction

The goal of image restoration is to recover the original clean image given a degraded image. This topic lies at the foundation of image processing, pattern recognition, and computer vision, so it has been extensively studied. So far, a large number of image restoration filters have been devised and their properties are investigated thoroughly, e.g., the inverse filter [22], the Wiener filter [17], the regularization filter [20,28,10], the moving-average filter [9], the parametric Wiener filter [18], and the band-pass filter [9].

It is empirically known that a certain filter works excellently for a certain type of original image or degradation while it may not be suitable for other images, i.e., there is no universally optimal filter. Therefore, the choice of the filter is as important as the development of new efficient filters. Furthermore, if a filter includes parameters to be adjusted, its restoration performance relies heavily on the

* Corresponding author. Tel.: +81-3-5734-2190; fax: +81-3-5734-2949.

E-mail addresses: sugi@og.cs.titech.ac.jp (M. Sugiyama), ogawa@og.cs.titech.ac.jp (H. Ogawa).

URL: http://ogawa-www.cs.titech.ac.jp/~sugi/

0165-1684/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0165-1684(02)00339-0

choice of the parameter values. For example, the regularization filter includes the regularization parameter, the moving-average filter has degrees of freedom in the choice of the weight pattern, and the band-pass filter contains the bandwidth to be determined.

For this reason, we discuss the problem of optimizing the filter type or parameter values. Specifically, we would like to optimize the filter so that the expected squared error (ESE) of the restored image is minimized. This is not a straightforward task since the unknown original image itself is required to directly calculate ESE. A general approach to the task is deriving an alternative criterion to ESE, and determining the filter type or parameter values so that the criterion is optimized [20,28,10]. The most crucial point in this approach is how well the alternative criterion approximates the original ESE. This topic is also a traditional concern in the communities of statistics and machine learning, and it has been extensively studied [1,6,14,16,21,23]. Most of the methods proposed so far proved their usefulness in the asymptotic sense. However, in practice, we are interested in the case with finite samples.

So far, an estimator of ESE called the *subspace information criterion* (SIC) was proposed [25,26]. Among several other interesting theoretical properties, SIC is shown to be an unbiased estimator of ESE with finite samples. SIC has been successfully applied to image restoration problems [24]. However, the study still lacks some practical concerns, such as the noise variance estimation, computational issues, and comparison with existing methods. In this paper, we therefore extensively investigate the characteristics of SIC. Specifically, we derive an analytic form of the optimal parameter value for the moving-average filter, which will greatly reduce the computational cost. Experiments with the regularization filter show that SIC works as well as existing methods in the small degradation case, and SIC tends to outperform existing methods in the severe degradation case. As a conclusion, this paper provides a powerful tool for optimizing linear image restoration filters.

The rest of this paper is organized as follows. In Section 2, the image restoration problem is mathematically formulated. In Section 3, an unbiased estimator of ESE called the SIC is reviewed. Section 4 gives a practical procedure for optimizing linear filters by SIC. Some practical concerns are also discussed here. In Section 5, computational issues are discussed, where SIC is applied to the moving-average filter and an analytic form of the optimal parameter value is derived. In Section 6, the performance of SIC is experimentally investigated through computer simulations. Finally, Section 7 gives concluding remarks and future prospects.

2. Problem formulation

In this section, we formulate the problem of image restoration.

Let $f(x,y)$ be an unknown original image in a real functional Hilbert space H_1. Let $g(x,y)$ be an observed image in a real functional Hilbert space H_2. Here, the domain of $f(x,y)$ or $g(x,y)$ can be continuous or discrete, and H_2 can be different from H_1. We assume that the dimension of H_2 is finite. Let us consider a standard case that the observed image g is given by

$$g = Af + n, \tag{1}$$

where A is a *known* operator from H_1 to H_2, and $n(x,y)$ is an additive noise in H_2. We assume that the mean noise is zero. A is called the *observation operator* and it expresses, e.g., the sampling or blur process. Let $\hat{f}(x,y)$ be a restored image in H_1. If a *restoration filter* is denoted by X, then $\hat{f}$ is expressed as

$$\hat{f} = Xg. \tag{2}$$

In this paper, we restrict ourselves to a linear filter X. Note that in the simplest case that the domains of H_1 and H_2 are both discrete, f, g, and $\hat{f}$ can be expressed by (vertically re-arranged) vectors, and A and X can be expressed by matrices. The above formulation is summarized in Fig. 1.

We evaluate the quality of the restored image $\hat{f}$ by the *expected squared error* (ESE):

$$\mathrm{ESE}[X] = E_n\|\hat{f} - f\|^2, \tag{3}$$

where E_n denotes the expectation over the noise and $\|\cdot\|$ denotes the norm in H_1. The norm is typically

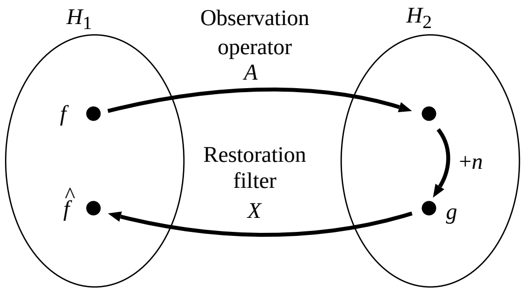


Fig. 1. Formulation of image restoration problem. f is the unknown original image, A is the observation operator, g is the observed image, n is the additive noise with mean zero, X is a linear restoration filter and $\hat{f}$ is a restored image.

defined by [1]

$$\|\hat{f}-f\|^2=\int(\hat{f}(x,y)-f(x,y))^2\,\mathrm{d}x\,\mathrm{d}y. \tag{4}$$

The goal of image restoration considered in this paper is to obtain the optimally restored image $\hat{f}$ that minimizes ESE given the observed image g.

3. Subspace information criterion for image restoration

Since ESE defined by Eq. (3) includes the unknown original image f, it cannot be directly calculated. In this section, we review an unbiased estimator of ESE called the SIC, which can be calculated without the original image f.

In the derivation of SIC, the following conditions are assumed.

(1) A linear filter X_u that gives an unbiased estimate $\hat{f}_u$ of the original image f is available:

$$\hat{f}_u = X_u g, \tag{5}$$

where $\hat{f}_u$ satisfies

$$E_n \hat{f}_u = f. \tag{6}$$

(2) The noise covariance operator Q is known.

[1] Rigorously speaking, the inner product is first defined and then the norm is derived from the inner product. Eq. (4) was actually derived from the inner product defined by

$$\langle f,g\rangle=\int f(x,y)g(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

These assumptions may not be practical in some cases. In the next section, we will come back to the assumptions and give practical methods for estimating X_u and Q. Until then, we assume that X_u and Q are available.

It follows from Eq. (3) that ESE can be decomposed as follows (see e.g. [11,13]).

$$\begin{aligned}\mathrm{ESE}[X] &= E_n\|\hat{f}-E_n\hat{f}+E_n\hat{f}-f\|^2\\ &= E_n\|\hat{f}-E_n\hat{f}\|^2+2E_n\langle \hat{f}-E_n\hat{f},E_n\hat{f}-f\rangle\\ &\quad + E_n\|E_n\hat{f}-f\|^2\\ &= E_n\|\hat{f}-E_n\hat{f}\|^2+\|E_n\hat{f}-f\|^2,\end{aligned} \tag{7}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in H_1. The first and second terms in Eq. (7) are called the *variance term* and the *bias term*, respectively.

It follows from Eqs. (2) and (1) that the variance term is expressed as

$$\begin{aligned}E_n\|\hat{f}-E_n\hat{f}\|^2 &= E_n\|Xg-E_nXg\|^2\\ &= E_n\|X(Af+n)-E_nX(Af+n)\|^2\\ &= E_n\|Xn\|^2\\ &= \mathrm{tr}(XQX^*),\end{aligned} \tag{8}$$

where X^* denotes the adjoint of X, $\mathrm{tr}(\cdot)$ denotes the trace of an operator, and Q is the noise covariance operator. Since we assumed that Q is known, we can calculate the variance term by Eq. (8).

On the other hand, the bias term $\|E_n\hat{f}-f\|^2$ is inaccessible since both $E_n\hat{f}$ and f are unavailable. Our key idea here is to use the unbiased estimate $\hat{f}_u$ for estimating the bias term, i.e., the bias term is roughly approximated by the squared distance between $\hat{f}$ and $\hat{f}_u$, which is accessible (Fig. 2):

$$\|\hat{f}-\hat{f}_u\|^2. \tag{9}$$

Indeed, it follows from Eqs. (6), (2), and (5) that the bias term is expressed as

$$\begin{aligned}&\|E_n\hat{f}-f\|^2\\ &\quad=\|\hat{f}-\hat{f}_u\|^2-\|\hat{f}-\hat{f}_u\|^2+\|E_n\hat{f}-f\|^2\\ &\quad=\|\hat{f}-\hat{f}_u\|^2-\|E_n(\hat{f}-\hat{f}_u)-E_n(\hat{f}-\hat{f}_u)\\ &\qquad+\hat{f}-\hat{f}_u\|^2+\|E_n\hat{f}-E_n\hat{f}_u\|^2\end{aligned}$$

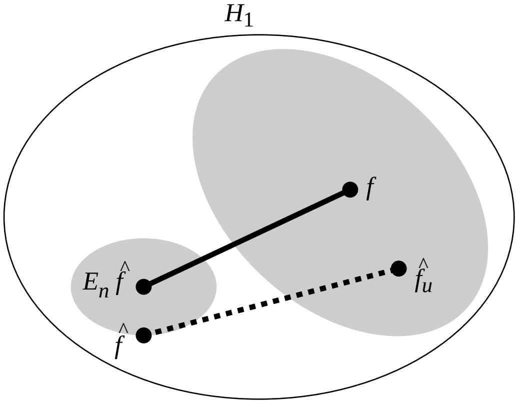


Fig. 2. Basic idea of SIC. The unknown bias term $\|E_n\hat{f}-f\|^2$ (depicted by the solid line) can be roughly estimated by $\|\hat{f}-\hat{f}_u\|^2$ (depicted by the dotted line), which is accessible.

$$\begin{aligned}&=\|Xg-X_ug\|^2-\|E_n(\hat{f}-\hat{f}_u)\|^2\\&\quad+2\langle E_n(\hat{f}-\hat{f}_u),\\&\quad E_n(\hat{f}-\hat{f}_u)-(\hat{f}-\hat{f}_u)\rangle-\|E_n(\hat{f}-\hat{f}_u)\\&\quad-(\hat{f}-\hat{f}_u)\|^2+\|E_n(\hat{f}-\hat{f}_u)\|^2\\&=\|(X-X_u)g\|^2+2\langle E_n(\hat{f}-\hat{f}_u),\\&\quad E_n(\hat{f}-\hat{f}_u)-(\hat{f}-\hat{f}_u)\rangle\\&\quad-\|E_n(\hat{f}-\hat{f}_u)-(\hat{f}-\hat{f}_u)\|^2.\end{aligned}\tag{10}$$

The second and third terms in Eq. (10) cannot be directly evaluated since they include an inaccessible term $E_n(\hat{f}-\hat{f}_u)$, so we will average out the second and third terms in Eq. (10) over the noise. Then the second term vanishes and it follows from Eqs. (2), (5), and (1) that the third term yields

$$\begin{aligned}&E_n(-\|E_n(\hat{f}-\hat{f}_u)-(\hat{f}-\hat{f}_u)\|^2)\\&\quad=-E_n\|E_n(X-X_u)g-(X-X_u)g\|^2\\&\quad=-E_n\|E_n(X-X_u)(Af+n)\\&\qquad-(X-X_u)(Af+n)\|^2\\&\quad=-E_n\|(X-X_u)n\|^2\\&\quad=-\mathrm{tr}((X-X_u)Q(X-X_u)^*).\end{aligned}\tag{11}$$

Then we have the following SIC:

$$\begin{aligned}\mathrm{SIC}[X]&=\|(X-X_u)g\|^2-\mathrm{tr}((X-X_u)Q(X-X_u)^*)\\&\quad+\mathrm{tr}(XQX^*).\end{aligned}\tag{12}$$

The validity of SIC as an approximation to ESE is theoretically substantiated by the fact that SIC is an unbiased estimator of ESE for any linear filter X [24]

$$E_n\mathrm{SIC}[X]=\mathrm{ESE}[X].\tag{13}$$

4. Application of SIC to filter optimization in practice

Although SIC derived in the previous section does not include the unknown original image f, it still includes the factors which are sometimes unknown: the filter X_u that gives an unbiased estimate $\hat{f}_u$ of the original image f and the noise covariance operator Q (see Section 3). Here, we show their practical estimation methods.

Let A^* be the adjoint operator of A. When the range of A^* is H_1 (i.e., A^*A is non-singular), the filter X_u is given by

$$X_u=A^\dagger,\tag{14}$$

where † denotes the generalized inverse (see, e.g. [3,4,12]). It is known that $A^\dagger$ gives the best linear unbiased estimator (BLUE), i.e., it gives the minimum variance estimator among the unbiased ones (see e.g. [3]). How to apply SIC when A^*A is singular is discussed in the article [27], so we do not go into the detail.

When the noise covariance operator Q is given in the form

$$Q=\sigma^2I,\tag{15}$$

where σ^2 is the (unknown) noise variance and I denotes the identity operator, a robust estimate of σ^2 is given by (see [8])

$$\hat{\sigma}^2=\left(\frac{\mathrm{MAD}}{0.6745}\right)^2,\tag{16}$$

where MAD is the median of the magnitudes of all the coefficients at the finest decomposition scale of the wavelet transform. It is empirically known that Eq. (16) gives a very accurate estimate of σ^2 in practice. Other methods for estimating σ^2 are given in, e.g., Refs. [29,25].

Making naive use of SIC, suitable filter type and parameter values can be obtained as follows. First, a finite set $\mathscr{X}$ of filters of different type or filters with different parameter values are prepared. For each filter in the set $\mathscr{X}$, the value of SIC is calculated. Then the filter $\hat{X}$ that minimizes SIC is selected:

$$\hat{X} = \mathop{\mathrm{argmin}}_{X \in \mathscr{X}} \mathrm{SIC}[X]. \tag{17}$$

The selected filter $\hat{X}$ is expected to be the best in terms of ESE.

5. Analytic solution for moving-average filter

The naive filter optimization method given by Eq. (17) is general and applicable to any linear filters. The more filter candidates we have, the better filter we obtain. Therefore, we would like to have as many filters as possible in the set $\mathscr{X}$. However, the number of filters in the set $\mathscr{X}$ should be kept as small as possible from the viewpoint of the computational cost.

For a set $\mathscr{X}$ of filters, there are probably two schemes to reduce the computational cost:

(1) Utilize the calculation results of SIC for other filters. For example, when we have two filters X_1 and X_2, calculating SIC$[X_2]$ based on the calculation result of SIC$[X_1]$ may be computationally more efficient than straightforwardly calculating SIC$[X_1]$ and SIC$[X_2]$ separately.

(2) Remove filters from the set $\mathscr{X}$ that are found to be worse than others. For example, if we can prove that X_1 is better than X_2, we can omit calculating SIC$[X_2]$.

In this paper, we focus on Scheme 2. The study along Scheme 1 is reserved for the future.

A promising approach to Scheme 2 is solving the optimization problem analytically. For example, analytically obtaining the optimal value of a continuous-valued parameter corresponds to reducing an infinite number of candidates to one.[2] This will efficiently reduce the computational cost. In this section, we consider the problem of optimizing parameters in the moving-average filter [9], and derive an analytic form of the optimal parameter value.

[2] Note that in this paper, filters with different parameter values are regarded as *different* filters since they are different operators from the mathematical point of view.

First, we describe the setting. Let $H_1 = H_2$ be sets of discrete images of size $D \times D$, i.e., $f(x, y)$ and $g(x, y)$ are defined on $\{1, 2, \ldots, D\} \times \{1, 2, \ldots, D\}$. The inner product in H_1 ($=H_2$) is defined by

$$\langle f, g \rangle = \sum_{x,y=1}^{D} f(x, y) g(x, y). \tag{18}$$

We assume that the observation operator A is identity:

$$A = I. \tag{19}$$

In this case, the filter X_u defined by Eq. (5) is also given by the identity operator, according to Eq. (14):

$$X_u = I. \tag{20}$$

We also assume that the noise covariance operator Q is given by Eq. (15).

The moving-average filter X restores the image by the weighted average over nearby pixels:

$$\hat{f}(x, y) = \frac{1}{C} \sum_{i,j=-W}^{W} w_{i,j} h(x - i, y - j), \tag{21}$$

where C is the normalizing constant defined by

$$C = \sum_{i,j=-W}^{W} w_{i,j}. \tag{22}$$

$h(x, y)$ is the same image as $g(x, y)$ but surrounded by mirrored images, i.e., $h(x, y)$ is defined on

$$\{1 - W, 2 - W, \ldots, D + W\} \times \{1 - W, 2 - W, \ldots, D + W\} \tag{23}$$

and $h(x, y)$ is defined by

$$h(x, y) = g(x', y'), \tag{24}$$

where

$$x' = \begin{cases} 2 - x : & 1 - W \leqslant x \leqslant 0, \\ x : & 1 \leqslant x \leqslant D, \\ 2D - x : & D + 1 \leqslant x \leqslant D + W, \end{cases} \tag{25}$$

$$y' = \begin{cases} 2 - y : & 1 - W \leqslant y \leqslant 0, \\ y : & 1 \leqslant y \leqslant D, \\ 2D - y : & D + 1 \leqslant y \leqslant D + W. \end{cases} \tag{26}$$

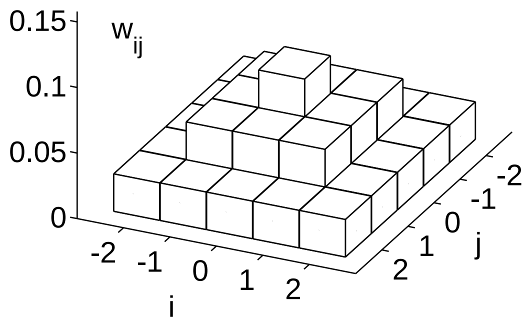


Fig. 3. Weight pattern (Window size $W = 2$).

The integer W ($\geqslant 1$) is called the *window size*, and the set $\{w_{i,j}\}_{i,j=-W}^{W}$ of scalars is called the *weight pattern*. We define the weight pattern like a pyramid shape [3] (see Fig. 3), i.e.,

$$w_{i,j} = (W - \max(|i|, |j|))w + 1, \tag{27}$$

where w is called the *weight coefficient*.

From here on, we focus on a fixed window size W and derive an analytic form of the weight coefficient w that minimizes SIC. As a result of a series of straightforward calculations, SIC for the moving-average filter can be expressed as follows: [4]

$$\mathrm{SIC}(w) = \frac{aw^2 + bw + c}{(w + d)^2} - \sigma^2 D^2, \tag{28}$$

where a, b, c, and d are the constants defined by

$$a = \sum_{x,y=1}^{D} t(x, y)^2 + 2\sigma^2 D^2 W s, \tag{29}$$

$$b = 2\sum_{x,y=1}^{D} t(x, y)u(x, y) + 2\sigma^2 D^2 (1 + Wd)s, \tag{30}$$

$$c = \sum_{x,y=1}^{D} u(x, y)^2 + 2\sigma^2 D^2 sd, \tag{31}$$

$$d = \frac{3(2W + 1)}{W(2W - 1)}, \tag{32}$$

$$s = \frac{3}{W(2W + 1)(2W - 1)}, \tag{33}$$

$$t(x, y) = s \sum_{i,j=-W}^{W} (W - \max(|i|, |j|))h(x - i, y - j) - g(x, y), \tag{34}$$

$$u(x, y) = s \sum_{i,j=-W}^{W} h(x - i, y - j) - dg(x, y). \tag{35}$$

From Eq. (28), we immediately have the following theorem.

Theorem 1. *For a fixed W,* SIC *has the minimum with respect to w if and only if*

$$2ad - b \neq 0, \tag{36}$$

$$ad^2 - bd + c > 0. \tag{37}$$

If the above conditions hold, the minimizer $\hat{w}$ of SIC is given by

$$\hat{w} = \frac{2c - bd}{2ad - b}. \tag{38}$$

A proof of Theorem 1 is omitted since it can be immediately obtained from Eq. (28) with elementary calculation. We attempted computer simulations thousands of times with different conditions, and empirically confirmed that conditions (36) and (37) always hold. Therefore, the verification of the conditions may be practically omitted.

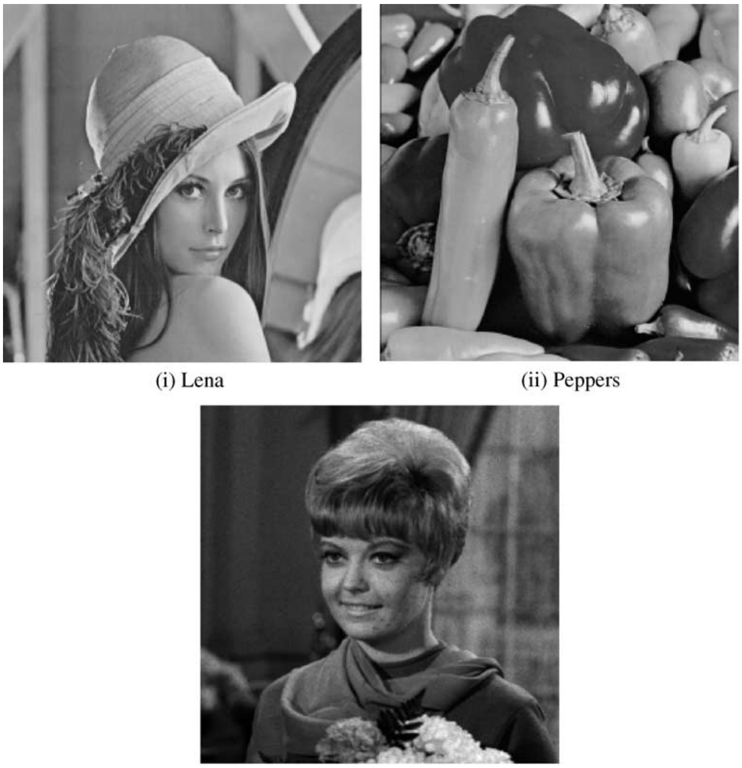
In this section, we derived an analytic form of the optimal parameter value in the moving-average filter. The result is rather simple, but is a meaningful step towards, the development of filter optimization theories. We expect that it is possible to derive analytic forms of the optimal parameter values for various filters in the same way. Further investigation along this line will be highly important since a collection of such results forms an excellent set of filters that is practically extremely useful.

6. Computer simulations

Finally, the effectiveness of SIC is demonstrated through computer simulations.

[3] It is possible to consider other parameterizations. However, we do not go into the detail here since they may be solved similarly.

[4] Derivation is outlined in Appendix A.



(i) Lena (ii) Peppers (iii) Girl

Fig. 4. Original images.

6.1. Regularization filter

Let $H_1 = H_2$ be a set of discrete images of size 256×256, i.e., $D=256$. The inner product in H_1 $(=H_2)$ is defined by

$$\langle f, g\rangle = \sum_{x,y=1}^{256} f(x, y)g(x, y). \tag{39}$$

As the original image, let us employ three gray-scale images (i) *Lena*, (ii) *Peppers*, and (iii) *Girl* displayed in Fig. 4. The pixel values $\{f(x, y)\}_{x,y=1}^{256}$ of the images are integers in $[0, 255]$.

Let the degradation operator A be a horizontal blur given by

$$[Af](x, y) = \frac{1}{2b+1}\sum_{i=-b}^{b} f(x + i \bmod 256, y), \tag{40}$$

where b expresses the blur level. We attempt $b=8$ and 16. We suppose that the noises $\{n(x, y)\}_{x,y=1}^{256}$ are independently drawn from the normal distribution with mean zero and variance σ^2. We attempt $\sigma^2 = 1$ and 4.

In this experiment, we would like to compare the performance of the proposed method with that of existing methods. For this reason, we will use the regularization filter for restoration since there exist a number of parameter optimization methods tailored for the regularization filter. The regularization filter X is defined as the minimizer of

$$\|AXg - g\|^2 + \alpha\|Xg\|^2, \tag{41}$$

where α is a positive scalar called the *regularization parameter*, which is to be determined. The regularization filter X is given by

$$X = (A^*A + \alpha I)^{-1}A^*. \tag{42}$$

The regularization parameter α is selected from

$$\{10^{-5}, 10^{-4}, 10^{-3}, \ldots, 10^{3}\}, \tag{43}$$

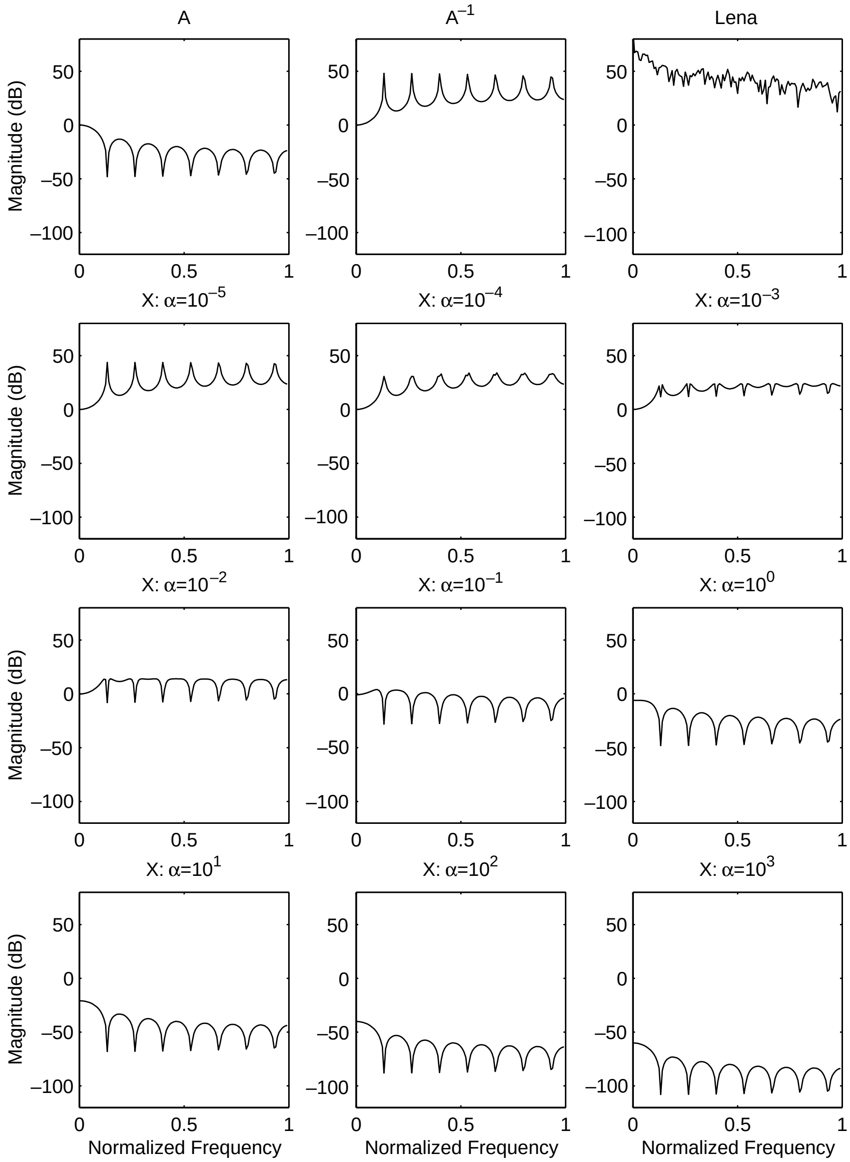


Fig. 5. Frequency responses of the blur operator A, the inverse of the blur operator, the original image *Lena*, and the regularization filters X with different values of the regularization parameter α in Eq. (43). The horizontal axis denotes the normalized frequency and the vertical axis denotes the magnitude of the frequency response.

i.e., Eq. (42) is substituted into SIC and SIC is regarded as a function of the regularization parameter α. We choose the one that minimizes SIC as the best regularization parameter.

Fig. 5 depicts the frequency responses of the blur operator A with the blur level $b = 8$, the inverse of the blur operator, the original image *Lena*, and the regularization filters with different values of the regularization parameter α in Eq. (43). This shows that the blur operator A is a low-pass filter, and the inverse of the blur operator amplifies high-frequency components. The regularization filter also amplifies high-frequency components for small α, and it tends to be a low-pass filter as α gets large.

The experiments are repeated 100 times for each original image, each blur lever b, and each noise variance σ^2, changing the noise $\{n(x,y)\}_{x,y=1}^{256}$ in each trial. The noise variance σ^2 is estimated by Eq. (16), where we use Daubechies's compactly supported wavelets with external phase and four vanishing moments [7].

As mentioned above, there are a large number of parameter optimization methods for the regularization filter. Here we compare the proposed SIC with some of the representative methods: Mallows's C_L [14], the leave-one-out cross-validation (CV) [19], the generalized cross-validation (GCV) [12], the network information criterion (NIC) [15], a Bayesian information criterion (ABIC) [2], and Vapnik's measure (VM) [5]. We measure the actual error of the restored image $\hat{f}$ by the following mean squared error (MSE):

$$\mathrm{MSE} = \frac{1}{D^2}\sum_{x,y=1}^{D}(\hat{f}(x,y) - f(x,y))^2. \tag{44}$$

In the following experiments, the value of SIC is divided by D^2 so that it is consistent with the above MSE.

Because of the lack of space, we only show the simulation results with Lena in the easiest and hardest cases in detail. The results for the easiest case with $(b,\sigma^2) = (8,1)$ are displayed in Fig. 6, and the results for the hardest cases with $(b,\sigma^2) = (16,4)$ are displayed in Fig. 7. The top-left graphs show the mean value of each criterion corresponding to each α. The error bar shows the standard error. Note that the vertical scale of MSE, SIC, C_L, CV, and VM are the same. The top-center graphs show the distribution of the selected value α by each criterion. 'OPT' denotes the optimal selection that minimizes MSE. The top-right graphs show the distribution of the obtained MSE by each criterion. Examples of observed image and restored image by SIC are displayed below the graphs.

The top-left graphs show that SIC gives a very good estimate of MSE and its accuracy is remarkable. Consequently, SIC always specifies the best regularization parameter that minimizes MSE. In the easiest case with $(b,\sigma^2) = (8,1)$ (Fig. 6), all methods except GCV and ABIC work excellently. However, when it comes to the hardest case with $(b,\sigma^2) = (16,4)$ (Fig. 7), SIC tends to outperform other methods. In Fig. 7, the variance of SIC is rather large. Now we investigate a cause of the large variance. SIC is expressed as

$$\mathrm{SIC}[X] = \|\hat{f}\|^2 + 2\langle \hat{f}, \hat{f}_u\rangle + \|\hat{f}_u\|^2 + 2\sigma^2\mathrm{tr}(XX_u^*) - \sigma^2\mathrm{tr}(X_uX_u^*). \tag{45}$$

In Eq. (45), $\|\hat{f}_u\|^2$ and $\sigma^2\mathrm{tr}(X_uX_u^*)$ are irrelevant to the regularization parameter α. Let SICr (SIC relevant) be SIC without the irrelevant terms

$$\mathrm{SICr}[X] = \|\hat{f}\|^2 + 2\langle \hat{f}, \hat{f}_u\rangle + 2\sigma^2\mathrm{tr}(XX_u^*). \tag{46}$$

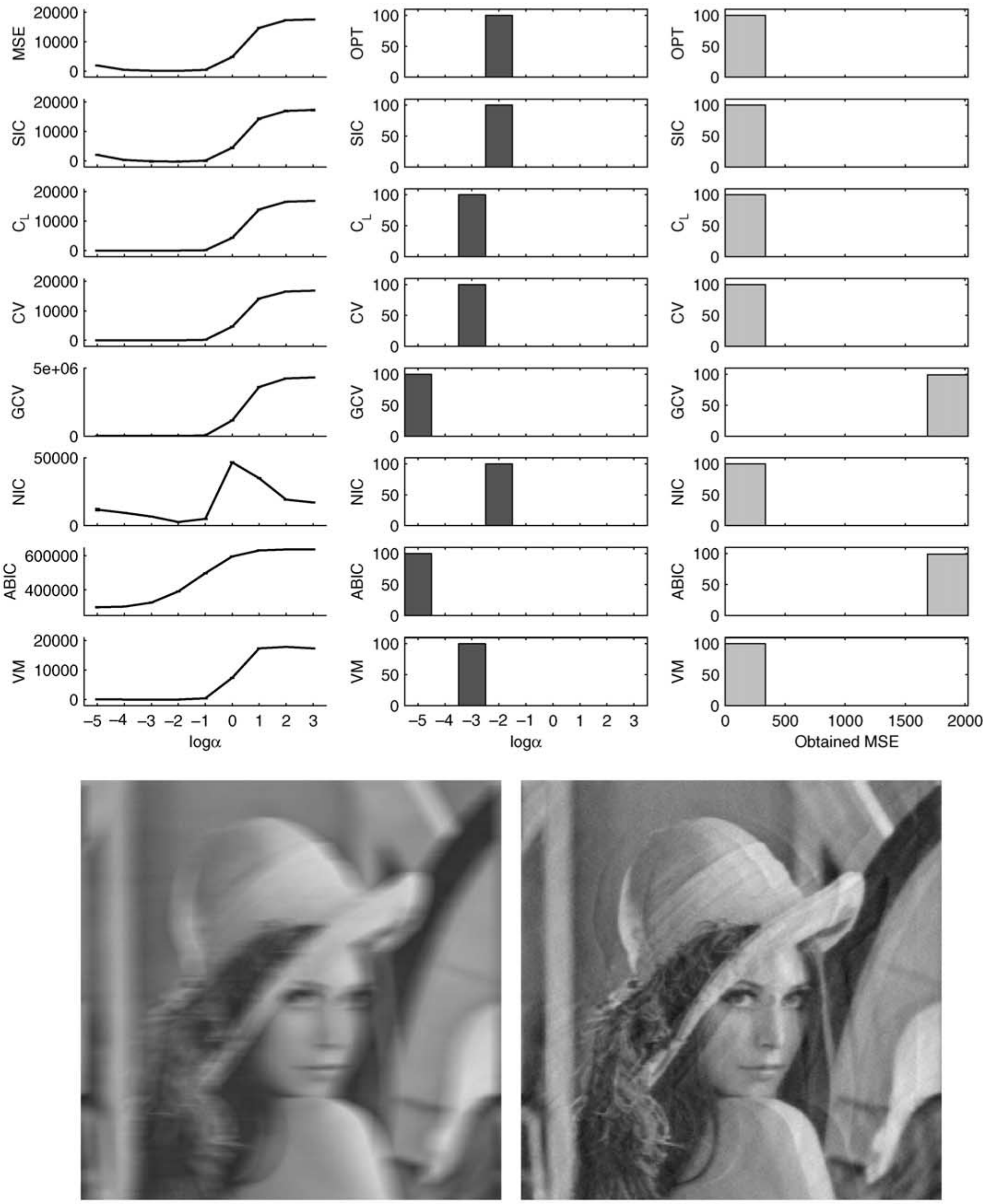
Fig. 8 depicts the values of MSE, SIC, and SICr when $(b,\sigma^2) = (16,4)$. The vertical scale of the three graphs is the same. The graphs show that the variance of SICr is very small and SICr approximates MSE very accurately, though it is biased. This implies that the large variance of SIC is dominated by the variance of the irrelevant terms $\|\hat{f}_u\|^2$ and $\sigma^2\mathrm{tr}(X_uX_u^*)$. Therefore, we come to the conclusion that the large variance of SIC does not degrade the performance of SIC.

Simulation results with *Peppers* and *Girl* were almost the same as those of Lena, so we only show the images in Fig. 9.

6.2. Moving-average filter

The previous simulation demonstrated the effectiveness of SIC. Now we experimentally investigate the performance of Theorem 1.

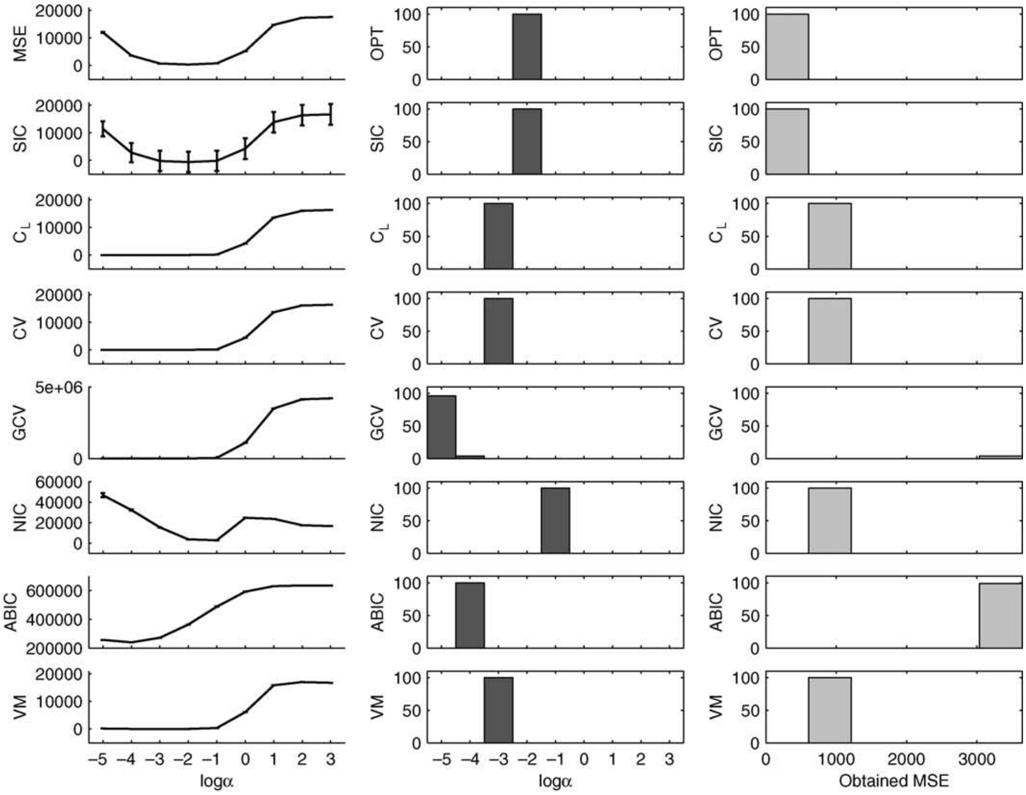
Let us again employ (i) *Lena*, (ii) *Peppers*, and (iii) *Girl* displayed in Fig. 4 as the original image. Let the observation operator A be an identity operator. The noises $\{n(x,y)\}_{x,y=1}^{256}$ are independently drawn from the normal distribution with mean zero and variance σ^2. In this case, the noise covariance operator Q is given by $\sigma^2 I$. We attempt the noise variance $\sigma^2 =$ 900, 1600, and 2500. The noise variance is estimated by the same method as that used in Section 6.1. We use the moving-average filter with the pyramid-shaped weight pattern (see Section 5). The window size W is selected from $\{1,2,3,4,5\}$, and the weight coefficient w is determined by Eq. (38). We measure the actual error of the restored image $\hat{f}$ by MSE defined by Eq. (44). For reference, we also examine the



Observed ‘Lena’ (MSE=467)

SIC restoration (MSE=132)

Fig. 6. Simulation results with Lena in the easiest case ($b = 8$ and $\sigma^2 = 1$). Mean value of each criterion (top-left), distribution of selected parameter value (top-center), distribution of actual MSE (top-right), observed image (bottom-left), and restored image by SIC (bottom-right).



Observed 'Lena' (MSE=943)

SIC restoration (MSE=327)

Fig. 7. Simulation results with Lena in the hardest case ($b = 16$ and $\sigma^2 = 4$). SIC outperforms other methods. The obtained errors by GCV are not properly plotted in the graph since they are such large values as around 1.2×10^4.

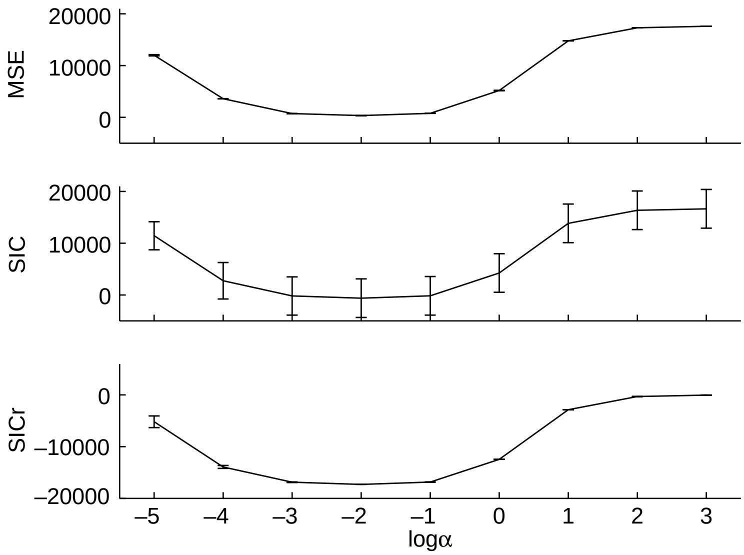


Fig. 8. SIC and SIC without irrelevant terms ($\mathrm{SIC_r}$) when $(b, \sigma^2) = (16, 4)$.

Observed 'Peppers' (MSE=1028)

SIC restoration (MSE=331)

Observed 'Girl' (MSE=381)

SIC restoration (MSE=157)

Fig. 9. Results with Peppers and Girl ($b = 16$ and $\sigma^2 = 4$).

Table 1
Maximum relative error by SIC over 100 trials

	$\sigma^2 = 900$	$\sigma^2 = 1600$	$\sigma^2 = 2500$
Lena	2.5×10^{-7}	9.4×10^{-3}	1.3×10^{-7}
Peppers	0	1.1×10^{-2}	1.5×10^{-7}
Girl	8.8×10^{-3}	0	1.2×10^{-2}

weight coefficient $\tilde{w}$ that minimizes MSE. It is not possible to analytically obtain $\tilde{w}$, so we search it by the gradient descent method with an accuracy of 10^{-2}.

The experiments are repeated 100 times for each original image and each noise variance, changing the noise in each trial.

Let us define the relative error of the restored image obtained by SIC as follows.

$$\frac{\mathrm{MSE}_{\mathrm{SIC}} - \mathrm{MSE}_{\mathrm{OPT}}}{\mathrm{MSE}_{\mathrm{SIC}}}, \tag{47}$$

where $\mathrm{MSE}_{\mathrm{SIC}}$ denotes MSE of the restored image by SIC and $\mathrm{MSE}_{\mathrm{OPT}}$ denotes MSE of the optimally restored image. Table 1 displays the maximum value of the relative error over 100 trials. The table shows that the relative error is very small so Theorem 1 works excellently in practice.

7. Conclusions and future works

The subspace information criterion (SIC) is an unbiased estimator of the expected squared error (ESE) for any linear filters. Computer simulations demonstrated that SIC gives a very accurate estimate of the actual squared error under various conditions, and the analytic form of the optimal parameter values for the moving-average filter works excellently in practice.

This paper provided a powerful tool for filter optimization. However, there is still plenty of room for further investigation.

In Section 5, we mentioned the computational issues of selecting filters from given candidates. As pointed out there, methods for efficiently comparing a large number of filters (possible, an infinitely large number) are indispensable for practical use. Developing methods for calculating values of SIC using the calculation results of SIC for other filters, or methods for removing worse filters from the candidates (e.g., obtaining an analytic form of a parameter) will be promising directions in this line.

In Section 2, we stated that our goal is to find a restored image that minimizes the ESE. However, the potential and essential problem is that ESE can be different from humans' visual system. Therefore, it is extremely important to find an alternative error measure that reflects humans' visual system well. We expect that SIC will still play a critical role in this future scenario, which is supported by the fact that SIC is valid for any weighted norms. That is, if an alternative error measure is expressed in the form of the weighted norm, SIC can be straightforwardly applied without any modifications. For example, we can incorporate the prior knowledge that humans' visual system is sensitive to the error around the edges by putting higher weight on the edge regions. Finding an appropriate weight pattern would be a promising approach to this important and essential problem.

Another challenging scenario is to consider the case where the observation operator is not available. For such a case, it is practically important to devise a general method for estimating the observation operator. It still remains to see whether SIC works well for estimated observation operators.

We entirely focused on linear filters. Generalizing SIC so that non-linear filters can be dealt with is also a considerably important direction for the future.

Finally, as well as the reinforcement of the filter optimization theories, the development of new efficient restoration filters is extremely essential.

Acknowledgements

The author would like to thank Daisuke Imaizumi and Toshihisa Tanaka for their assistance. Our gratitude also goes to anonymous reviewers for their helpful comments on the manuscript.

Appendix A. Derivation of Eq. (28)

Substituting $A = I$ and $Q = \sigma^2 I$ into Eq. (12), we have

$$\mathrm{SIC}[X] = \sum_{x,y=1}^{D} (\hat{f}(x,y) - g(x,y))^2 + 2\sigma^2 \mathrm{tr}(X) - \sigma^2 D^2. \tag{A.1}$$

From Eq. (27), the normalizing constant C defined by Eq. (22) is expressed as

$$C = \frac{1}{s}(w + d), \tag{A.2}$$

where s and d are defined by Eqs. (33) and (32), respectively. From Eqs. (21), (27), and (A.2), the first term in Eq. (A.1) is expressed as

$$\begin{aligned}
&\sum_{x,y=1}^{D} (\hat{f}(x,y) - g(x,y))^2 \\
&\quad = \sum_{x,y=1}^{D} \left(\frac{1}{C} \sum_{i,j=-W}^{W} w_{i,j} h(x-i, y-j) - g(x,y) \right)^2 \\
&\quad = \frac{1}{C^2} \sum_{x,y=1}^{D} \left(\sum_{i,j=-W}^{W} ((W - \max(|i|,|j|))w + 1) \right. \\
&\qquad\qquad \left. \times\, h(x-i, y-j) - Cg(x,y) \right)^2 \\
&\quad = \frac{1}{(w+d)^2} \sum_{x,y=1}^{D} (t(x,y)w + u(x,y))^2, \qquad \text{(A.3)}
\end{aligned}$$

where $t(x,y)$ and $u(x,y)$ are defined by Eqs. (34) and (35), respectively. From Eqs. (27) and (A.2) the second term in Eq. (A.1) is expressed as

$$\begin{aligned}
2\sigma^2 \mathrm{tr}(X) &= 2\sigma^2 \sum_{x,y=1}^{D} \frac{1}{C} w_{0,0} \\
&= 2\sigma^2 \sum_{x,y=1}^{D} \frac{1}{C} (Ww + 1) \\
&= \frac{2\sigma^2 D^2}{C^2} (Ww + 1)C \\
&= \frac{2\sigma^2 D^2 s}{(w+d)^2} (Ww^2 + (1 + Wd)w + d).
\end{aligned} \tag{A.4}$$

Substituting Eqs. (A.3) and (A.4) into Eq. (A.1), we have Eq. (28). □

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control AC-19 (6) (1974) 716–723.
[2] H. Akaike, Likelihood and the Bayes procedure, in: N.J. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith (Eds.), Bayesian Statistics, University Press, Valencia, 1980, pp. 141–166.
[3] A. Albert, Regression and the Moore–Penrose Pseudoinverse, Academic Press, New York, London, 1972.
[4] A. Ben-Israel, T.N.E. Greville, Generalized Inverses: Theory and Applications, Wiley, New York, 1974.
[5] V. Cherkassky, X. Shao, F.M. Mulier, V.N. Vapnik, Model complexity control for regression using VC generalization bounds, IEEE Trans. Neural Networks 10 (5) (1999) 1075–1089.
[6] P. Craven, G. Wahba, Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation Numer. Math. 31 (1979) 377–403.
[7] I. Daubechies, Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.
[8] D.L. Donoho, Wavelet shrinkage and W.V.D.: a 10-minute tour, in: International Conference on Wavelets and Applications, Toulouse, France, 1992. URL http://www-stat.stanford.edu/~donoho/Reports/index.html
[9] E.R. Dougherty, C.R. Giardina, Image Processing—Continuous to Discrete, Vol. 3, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1987.
[10] N.P. Galatsanos, A.K. Katsaggelos, Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation, IEEE Trans. Image Process. 1 (3) (1992) 322–336.
[11] S. Geman, E. Bienenstock, R. Doursat, Neural networks and the bias/variance dilemma, Neural Comput. 4 (1) (1992) 1–58.
[12] C.W. Groetsch, Generalized Inverse of Linear Operators: Representation and Approximation, Marcel Dekker, Inc, New York, Basel, 1977.
[13] T. Heskes, Bias/variance decompositions for likelihood-based estimators, Neural Comput. 10 (6) (1998) 1425–1433.
[14] C.L. Mallows, Some comments on C_P, Technometrics 15 (4) (1973) 661–675.
[15] N. Murata, Bias of estimators and regularization terms, in: Proceedings of 1998 Workshop on Information-Based Induction Sciences (IBIS'98), Izu, Japan, 1998, pp. 87–94.
[16] N. Murata, S. Yoshizawa, S. Amari, Network information criterion—determining the number of hidden units for an artificial neural network model, IEEE Trans. Neural Networks 5 (6) (1994) 865–872.
[17] H. Ogawa, E. Oja, Projection filter, Wiener filter, and Karhunen-Loève subspaces in digital image restoration, J. Math. Anal. Appl. 114 (1) (1986) 37–51.
[18] E. Oja, H. Ogawa, Parametric projection filter for image and signal restoration, IEEE Trans. Acoust. Speech Signal Process. ASSP-34 (6) (1986) 1643–1653.

[19] M.J.L. Orr, Introduction to radial basis function networks, Tech. rep., Center for Cognitive Science, University of Edinburgh, 1996. URL http://www.anc.ed.ac.uk/~mjo/papers/intro.ps.gz
[20] S.J. Reeves, R.M. Mersereau, Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration, Opt. Eng. 29 (5) (1990) 446–454.
[21] J. Rissanen, Modeling by shortest data description, Automatica 14 (1978) 465–471.
[22] A. Rosenfeld, A.C. Kak, Digital Picture Processing, Vol. 1, Academic Press, New York, 1982.
[23] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.
[24] M. Sugiyama, D. Imaizumi, H. Ogawa, Subspace information criterion for image restoration—optimizing parameters in linear filters, IEICE Trans. Inform. Systems E84-D (9) (2001) 1249–1256.
[25] M. Sugiyama, H. Ogawa, Subspace information criterion for model selection, Neural Comput. 13 (8) (2001) 1863–1889.
[26] M. Sugiyama, H. Ogawa, Theoretical and experimental evaluation of the subspace information criterion, Mach. Learning 48 (1/2/3) (2002) 25–50.
[27] A. Tanaka, H. Imai, M. Miyakoshi, Choosing the parameter of image restoration filters by modified subspace information criterion, IEICE Trans. Fundam. E85-A (5) (2002) 1104–1110.
[28] A.M. Thompson, J.C. Brown, J.W. Kay, D.M. Titterington, A study of methods of choosing the smoothing parameter in image restoration by regularization, IEEE Trans. Pattern Anal. Mach. Intell. 13 (4) (1991) 326–339.
[29] H. Wahba, Spline Model for Observational Data, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1990.