IEICE Transactions on Information and Systems,
vol.E84-D, no.9, pp.1249–1256, Sep. 2001.

Subspace Information Criterion for Image Restoration — Optimizing Parameters in Linear Filters

Masashi Sugiyama Daisuke Imaizumi Hidemitsu Ogawa

Department of Computer Science,
Graduate School of Information Science and Engineering,
Tokyo Institute of Technology.

2-12-1, O-okayama, Meguro-ku, Tokyo, 152-8552, Japan.

`sugi@og.cs.titech.ac.jp`
`http://ogawa-www.cs.titech.ac.jp/~sugi/`

Abstract

Most of the image restoration filters proposed so far include parameters that control the restoration properties. For bringing out the optimal restoration performance, these parameters should be determined so as to minimize a certain error measure such as the mean squared error (MSE) between the restored image and original image. However, this is not generally possible since the unknown original image itself is required for evaluating MSE. In this paper, we derive an estimator of MSE called the subspace information criterion (SIC), and propose determining the parameter values so that SIC is minimized. For any linear filter, SIC gives an unbiased estimate of the expected MSE over the noise. Therefore, the proposed method is valid for any linear filter. Computer simulations with the moving-average filter demonstrate that SIC gives a very accurate estimate of MSE in various situations, and the proposed procedure actually gives the optimal parameter values that minimize MSE.

Keywords

image restoration, mean squared error, subspace information criterion, moving-average filter, model selection.

1 Introduction

Image restoration from observed images is one of the most basic and important subjects in the fields of image processing, pattern recognition, and computer vision. So far, various image restoration filters have been proposed. Most of the filters include *parameters* that control the restoration properties, e.g., the window size and weight pattern in the moving-average filter [5], the band-width in the band-pass filter [5], the threshold in the wavelet thresholding filter [4, 3, 13], and the regularization factors in the regularization filters [8, 9, 12, 6]. The restoration properties of the filters depend heavily on the values of these parameters.

The quality of the restored images is generally evaluated by the *mean squared error* (MSE) between the restored image and original image. If the parameter values are determined so that MSE is minimized, then the optimal restoration performance is expected. However, this is not generally possible since the unknown original image itself is required for evaluating MSE. A general approach to the parameter optimization problem is to derive an estimator of MSE and determine the parameter values so that the estimator is minimized. So far, research based on this approach has been conducted, e.g., for the wavelet thresholding filters [4, 3, 13] and the regularization filters [9, 12, 6].

In this paper, we derive an estimator of MSE called the *subspace information criterion* (SIC) for linear filters, which is originated in the statistical model selection criterion [10, 11]. The quality of SIC as an approximation to MSE is theoretically substantiated by the fact that SIC gives an unbiased estimate of the expected MSE over the noise. We apply SIC to the moving-average filter, to which the existing methods described above can not be applied. Computer simulations demonstrate that SIC gives a very accurate estimate of MSE in various situations, and the proposed procedure actually gives the optimal parameter values that minimize MSE.

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. Note that 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, and the observed image g is given by

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

where A is an operator from H_1 to H_2, and $n(x, y)$ is an additive noise in H_2. A is called the *observation operator*. 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 by

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

We evaluate the goodness of the restored image $\hat{f}$ by the *mean squared error* (MSE):

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

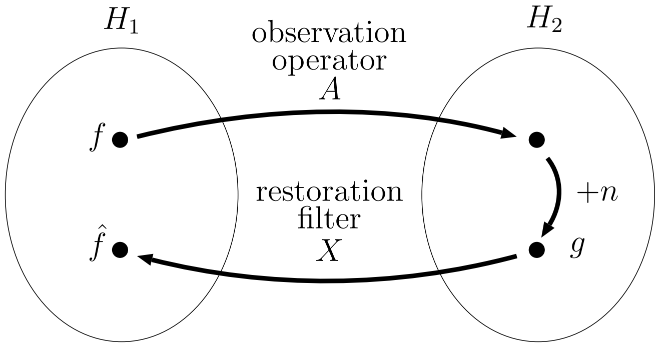


Figure 1: Formulation of image restoration problem. f is the unknown original image. A is the observation operator. g is the observed image. n is the additive noise. X is a restoration filter. $\hat{f}$ is a restored image.

where $\|\cdot\|$ denotes the norm in H_1. Then the problem of image restoration considered in this paper is formulated as the problem of obtaining the optimally restored image $\hat{f}$ that minimizes MSE from the observed image g. The above formulation is summarized in Fig. 1.

3 Subspace information criterion for image restoration

Since MSE includes the unknown original image f, it can not be directly evaluated. In this section, we derive an estimator of MSE called the *subspace information criterion* (SIC), which can be calculated without the original image f.

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

1. A filter X is linear.

2. The mean noise is zero:
$$\mathrm{E}_n n = 0, \tag{4}$$
where E_n denotes the expectation over the noise.

3. 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
$$\mathrm{E}_n \hat{f}_u = f. \tag{6}$$

A basic idea of SIC is that the unbiased estimate $\hat{f}_u$ is used for estimating MSE (Fig. 2).

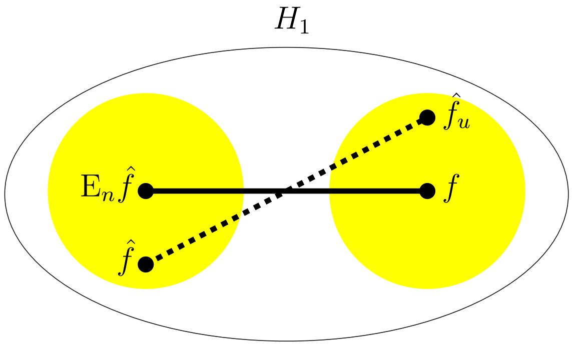


Figure 2: Basic idea of SIC. The solid line denotes the bias of $\hat{f}$. It can be roughly estimated by the dotted line, which can be calculated. (see the text for detail).

It follows from Eq.(3) that the expectation of MSE over the noise is decomposed as

$$\begin{aligned}
\mathrm{E}_n\mathrm{MSE}[X] &= \mathrm{E}_n\|\hat{f} - \mathrm{E}_n\hat{f} + \mathrm{E}_n\hat{f} - f\|^2 \\
&= \mathrm{E}_n\|\hat{f} - \mathrm{E}_n\hat{f}\|^2 + 2\mathrm{E}_n\langle \hat{f} - \mathrm{E}_n\hat{f}, \mathrm{E}_n\hat{f} - f\rangle + \mathrm{E}_n\|\mathrm{E}_n\hat{f} - f\|^2 \\
&= \mathrm{E}_n\|\hat{f} - \mathrm{E}_n\hat{f}\|^2 + \|\mathrm{E}_n\hat{f} - f\|^2, \qquad (7)
\end{aligned}$$

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

Let Q be the noise covariance operator. Then it follows from Eqs.(2), (1), and (4) that the variance of $\hat{f}$ is expressed by

$$\begin{aligned}
\mathrm{E}_n\|\hat{f} - \mathrm{E}_n\hat{f}\|^2 &= \mathrm{E}_n\|Xg - \mathrm{E}_nXg\|^2 \\
&= \mathrm{E}_n\|X(Af + n) - \mathrm{E}_nX(Af + n)\|^2 \\
&= \mathrm{E}_n\|Xn\|^2 \\
&= \mathrm{tr}\left(XQX^*\right), \qquad (8)
\end{aligned}$$

where X^* denotes the adjoint of X, and $\mathrm{tr}\left(\cdot\right)$ denotes the trace of an operator. It follows from Eqs.(6), (2), and (5) that the bias of $\hat{f}$ is expressed by

$$\begin{aligned}
\|\mathrm{E}_n\hat{f} - f\|^2 &= \|\hat{f} - \hat{f}_u\|^2 - \|\hat{f} - \hat{f}_u\|^2 + \|\mathrm{E}_n\hat{f} - f\|^2 \\
&= \|\hat{f} - \hat{f}_u\|^2 - \|\mathrm{E}_n(\hat{f} - \hat{f}_u) - \mathrm{E}_n(\hat{f} - \hat{f}_u) + \hat{f} - \hat{f}_u\|^2 + \|\mathrm{E}_n\hat{f} - \mathrm{E}_n\hat{f}_u\|^2 \\
&= \|Xg - X_ug\|^2 - \|\mathrm{E}_n(\hat{f} - \hat{f}_u)\|^2 + 2\langle \mathrm{E}_n(\hat{f} - \hat{f}_u), \mathrm{E}_n(\hat{f} - \hat{f}_u) - (\hat{f} - \hat{f}_u)\rangle \\
&\quad - \|\mathrm{E}_n(\hat{f} - \hat{f}_u) - (\hat{f} - \hat{f}_u)\|^2 + \|\mathrm{E}_n(\hat{f} - \hat{f}_u)\|^2 \\
&= \|(X - X_u)g\|^2 + 2\langle \mathrm{E}_n(\hat{f} - \hat{f}_u), \mathrm{E}_n(\hat{f} - \hat{f}_u) - (\hat{f} - \hat{f}_u)\rangle \\
&\quad - \|\mathrm{E}_n(\hat{f} - \hat{f}_u) - (\hat{f} - \hat{f}_u)\|^2. \qquad (9)
\end{aligned}$$

The second and third terms in Eq.(9) can not be directly evaluated since they include an inaccessible term $\mathrm{E}_n(\hat{f} - \hat{f}_u)$, so we will average out the second and third terms in Eq.(9)

over the noise. Then the second term vanishes and it follows from Eqs.(2), (5), (1), and (4) that the third term yields

$$\begin{aligned} \mathrm{E}_n(-\|\mathrm{E}_n(\hat{f}-\hat{f}_u)-(\hat{f}-\hat{f}_u)\|^2) &= -\mathrm{E}_n\|\mathrm{E}_n(X-X_u)g-(X-X_u)g\|^2 \\ &= -\mathrm{E}_n\|\mathrm{E}_n(X-X_u)(Af+n)-(X-X_u)(Af+n)\|^2 \\ &= -\mathrm{E}_n\|(X-X_u)n\|^2 \\ &= -\mathrm{tr}\left((X-X_u)Q(X-X_u)^*\right). \end{aligned} \tag{10}$$

Then we have the following criterion.

Definition 1 (Subspace information criterion) *The following functional* SIC *is called the subspace information criterion for a linear filter X:*

$$\mathrm{SIC}[X] = \|(X-X_u)g\|^2 - \mathrm{tr}\left((X-X_u)Q(X-X_u)^*\right) + \mathrm{tr}\left(XQX^*\right). \tag{11}$$

The goodness of SIC as an approximation to MSE is theoretically substantiated by the following theorem.

Theorem 1 *For any linear filter X,* SIC *gives an unbiased estimate of the expected* MSE *over the noise:*

$$\mathrm{E}_n\mathrm{SIC}[X] = \mathrm{E}_n\mathrm{MSE}[X]. \tag{12}$$

Proof: It follows from Eqs.(1), (4), (2), (5), and (6) that

$$\begin{aligned} \mathrm{E}_n\|(X-X_u)g\|^2 &= \mathrm{E}_n\|(X-X_u)(Af+n)\|^2 \\ &= \|(X-X_u)Af\|^2 + 2\mathrm{E}_n\langle (X-X_u)Af, (X-X_u)n\rangle \\ &\quad + \mathrm{E}_n\|(X-X_u)n\|^2 \\ &= \|(X-X_u)\mathrm{E}_n g\|^2 + \mathrm{tr}\left((X-X_u)Q(X-X_u)^*\right) \\ &= \|\mathrm{E}_n\hat{f}-f\|^2 + \mathrm{tr}\left((X-X_u)Q(X-X_u)^*\right). \end{aligned} \tag{13}$$

It follows from Eqs.(11), (13), (8), and (7) that

$$\begin{aligned} \mathrm{E}_n\mathrm{SIC}[X] &= \|\mathrm{E}_n\hat{f}-f\|^2 + \mathrm{tr}\left(XQX^*\right) \\ &= \mathrm{E}_n\mathrm{MSE}[X], \end{aligned} \tag{14}$$

which concludes the proof. ■

Based on Theorem 1, we will use SIC as a substitute for MSE in the following sections.

4 Optimization of moving-average filter by subspace information criterion

In this section, we give a method for optimizing the parameters of the *moving-average filter* [5], which is one of the classic but effective filters.

4.1 Setting

Let H_1 and H_2 be sets of discrete images of size $D \times D$, i.e., $f(x, y)$ and $g(x, y)$ are defined on

$$\{1, 2, \ldots, D\} \times \{1, 2, \ldots, D\}. \tag{15}$$

Let us define the inner product in H_1 by

$$\langle f, g \rangle = \sum_{x,y=1}^{D} f(x, y) g(x, y). \tag{16}$$

We adopt a typical definition of MSE in the discrete case:

$$\begin{aligned} \mathrm{MSE}[X] &= \frac{1}{D^2} \sum_{x,y=1}^{D} \left(\hat{f}(x, y) - f(x, y) \right)^2 \\ &= \frac{1}{D^2} \|\hat{f} - f\|^2. \end{aligned} \tag{17}$$

Let us assume that the observation operator A is non-singular. Then the filter X_u that gives an unbiased estimate of the original image f is given by

$$X_u = A^{-1}, \tag{18}$$

since it follows from Eqs.(1) and (4) that

$$\mathrm{E}_n A^{-1} g = \mathrm{E}_n A^{-1}(Af + n) = f. \tag{19}$$

4.2 Moving-average filter

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

$$\hat{f}(x, y) = \frac{1}{C} \sum_{i,j=-W}^{W} w_{i,j} h(x - i, y - j), \tag{20}$$

where W is a non-negative integer called the *window size*, and $\{w_{i,j}\}_{i,j=-W}^{W}$ is a set of scalars called the *weight pattern*. C is a normalizing constant defined by

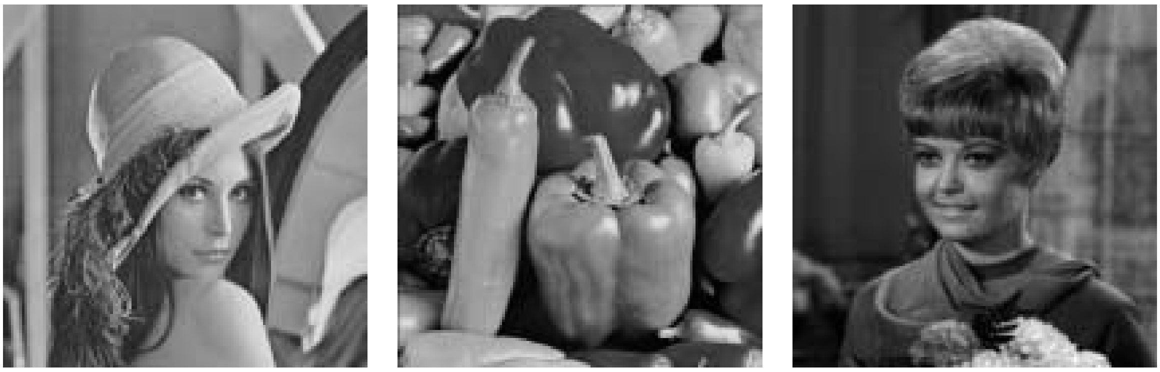
$$C = \sum_{i,j=-W}^{W} w_{i,j}, \tag{21}$$

which is assumed to be non-zero. $h(x, y)$ in Eq.(20) is the same image as $g(x, y)$ but surrounded by mirrored images, i.e., $h(x, y)$ is defined on

$$\{-W + 1, -W + 2, \ldots, D + W\} \times \{-W + 1, -W + 2, \ldots, D + W\}, \tag{22}$$

and it is defined by

$$h(x, y) = g(x', y'), \tag{23}$$



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

Figure 3: Original images.

where x' and y' are given by

$$x' = \begin{cases} 2-x & : -W+1 \leq x \leq 0, \\ x & : 1 \leq x \leq D, \\ 2D-x & : D+1 \leq x \leq D+W, \end{cases} \tag{24}$$

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

In the case of the moving-average filter, the window size W and weight pattern $\{w_{i,j}\}_{i,j=-W}^{W}$ are the parameters to be determined.

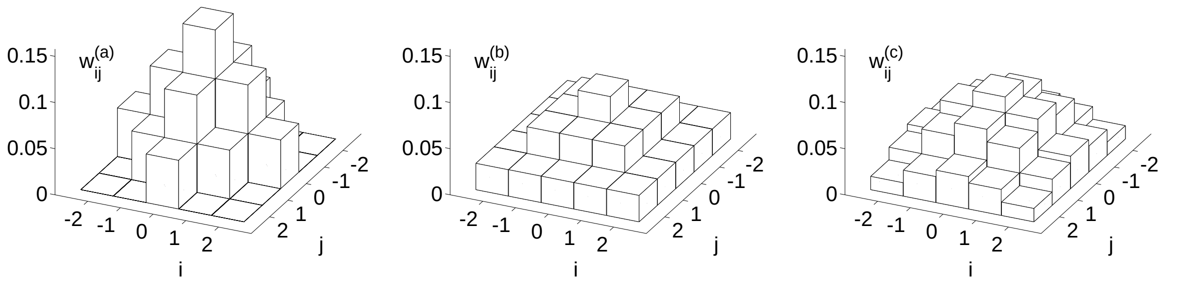
4.3 Parameter optimization by subspace information criterion

By using SIC given by Eq.(11), the filter parameters can be optimized as follows. First, a set $\mathcal{M}$ of filters with different parameter values is prepared. Then SIC is calculated for each filter X in the set $\mathcal{M}$, and the filter $\hat{X}$ that minimizes SIC is selected:

$$\hat{X} = \underset{X \in \mathcal{M}}{\operatorname{argmin}} \, \mathrm{SIC}[X], \tag{26}$$

where SIC in the current setting is given by

$$\begin{aligned} \mathrm{SIC}[X] = & \frac{1}{D^2} \sum_{x,y=1}^{D} \left(\hat{f}(x,y) - \hat{f}_u(x,y) \right)^2 \\ & + \frac{2}{D^2} \mathrm{tr} \left(XQ(A^{-1})^* \right) \\ & - \frac{1}{D^2} \mathrm{tr} \left(A^{-1}Q(A^{-1})^* \right). \end{aligned} \tag{27}$$



(a) Rhombus pattern (b) Pyramid pattern (c) Gauss pattern

Figure 4: Normalized weight patterns for the window size $W = 2$.

Note that SIC given by Eq.(11) is an estimator of $\|\hat{f} - f\|^2$ while the right-hand side of Eq.(27) is divided by D^2 since MSE defined by Eq.(17) is also divided by D^2. The parameter values in the selected filter $\hat{X}$ are expected to be the best. Indeed, the expectation is theoretically supported by Theorem 1, and experimentally demonstrated in Section 5.

Let I be the identity operator on H_1 and σ^2 be a positive scalar. When $A = I$ and $Q = \sigma^2 I$, SIC is reduced to a simpler expression:

$$\frac{1}{D^2}\sum_{x,y=1}^{D}\left(\hat{f}(x,y) - g(x,y)\right)^2 + \frac{2\sigma^2 w_{0,0}}{C} - \sigma^2. \tag{28}$$

In this case, SIC agrees with the traditional C_L-statistics [7].

5 Computer simulations

In this section, the effectiveness of SIC for the moving-average filter is demonstrated through computer simulations.

Let us employ (i) *Lena*, (ii) *Peppers*, and (iii) *Girl* displayed in Fig. 3 as the original image. The size D of the images is 256 and the pixel values $\{f(x,y)\}_{x,y=1}^{256}$ are integers in $[0, 255]$. Let A be an operator that multiplies the x-th column pixels by a_x:

$$[Af](x,y) = a_x f(x,y). \tag{29}$$

The multipliers $\{a_x\}_{x=1}^{256}$ are independently drawn from the uniform distribution on $(1,2)$. Since such an operator A is non-singular, the filter X_u that gives an unbiased estimate $\hat{f}_u$ of the original image f is given by Eq.(18). The noises $\{n(x,y)\}_{x,y=1}^{256}$ are independently drawn from the same normal distribution with mean zero and variance σ^2. In this case, the noise covariance operator Q is given by $Q = \sigma^2 I$. We attempt $\sigma^2 = 900$ and 2500.

Let us employ the restoration filter of the form $X = X'A^{-1}$, where X' is the moving-average filter. As candidates of the parameter values in the moving-average filter, we attempt the following window size W:

$$W = 0, 1, \ldots, 5. \tag{30}$$

For each W, we consider the following three weight patterns.

(a) Rhombus pattern:

$$w_{i,j}^{(a)} = \max(0, W + 1 - |i| - |j|). \tag{31}$$

(b) Pyramid pattern:

$$w_{i,j}^{(b)} = W + 1 - \max(|i|, |j|). \tag{32}$$

(c) Gauss pattern:

$$w_{i,j}^{(c)} = \frac{1}{2\pi(\frac{W+1}{2})^2} \exp\left(-\frac{i^2 + j^2}{2(\frac{W+1}{2})^2}\right). \tag{33}$$

The above weight patterns for $W = 2$ are illustrated in Fig. 4. We calculate the value of SIC for each window size and each weight pattern, and select the parameters that minimize SIC.

Figs. 5 and 6 display the simulation results when the noise variance σ^2 is 900 and 2500, respectively. The top rows show the degraded images $\{g(x, y)\}_{x,y=1}^{256}$. Their MSEs measured by Eq.(17) are described below the images. The middle rows show the values of MSE and SIC corresponding to each window size W and each weight pattern $\{w_{i,j}\}_{i,j=-W}^{W}$. The horizontal axis denotes the window size W. The bottom rows show the restored images $\{\hat{f}(x, y)\}_{x,y=1}^{256}$ obtained by SIC. Below the images, selected filter parameters and MSEs of the restored images are described. 'OPT' indicates the optimal parameters that minimize MSE.

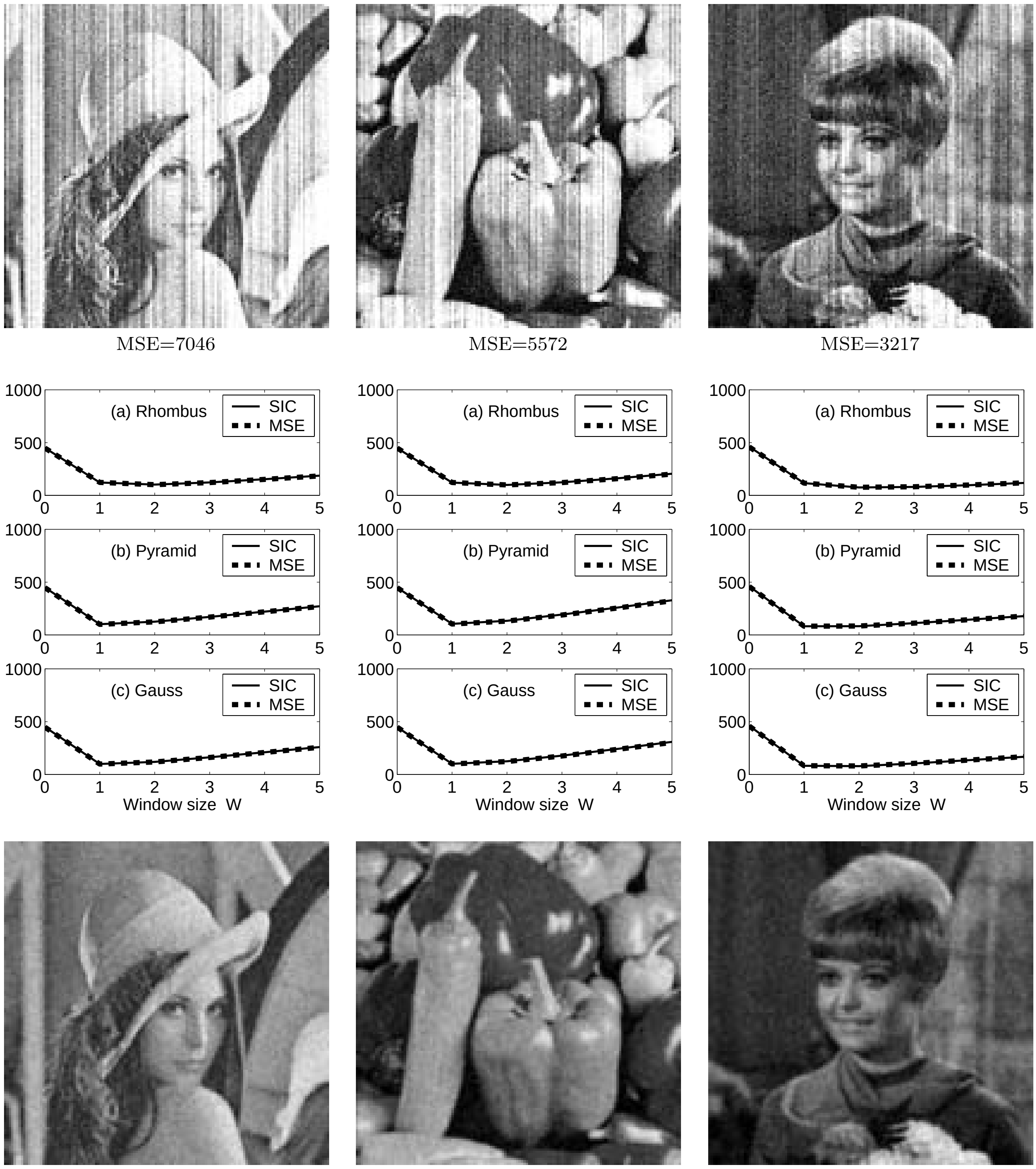
The graphs in the middle rows show that SIC gives a very accurate estimate of MSE irrespective of the original image, noise variance, window size, and weight pattern. The restored images in the bottom rows show that the filter parameters that minimize SIC actually minimize MSE, i.e., the optimal filter parameters can be obtained by SIC.

We also performed the same simulation with the noise drawn from the uniform distribution with mean zero and variance σ^2. The simulation results were the same, i.e., the optimal filter parameters can be obtained by SIC. This implies that the performance of SIC does not depend on the shape of the noise distribution.

6 Conclusions

We derived an unbiased estimator of the expected mean squared error for linear filters, which is named the subspace information criterion (SIC). We proposed determining the parameter values of the image restoration filters so that SIC is minimized. Computer simulations with the moving-average filter showed that SIC gives a very accurate estimate of MSE in various situations, and the optimal parameters that minimize MSE can be obtained by the proposed method.

SIC is valid for any linear filter. Applying SIC to other effective filters is prospective future work.



SIC: Gauss ($W = 1$), MSE=99
OPT: Gauss ($W = 1$), MSE=99

SIC: Rhombus ($W = 2$), MSE=99
OPT: Rhombus ($W = 2$), MSE=99

SIC: Rhombus ($W = 2$), MSE=77
OPT: Rhombus ($W = 2$), MSE=77

Figure 5: Simulation results when $\sigma^2 = 900$. The top row shows the degraded images. Their MSEs are described below. The middle row shows the values of SIC and MSE corresponding to each filter. The horizontal axis denotes the window size W. The bottom row shows the restored images by SIC. Selected filter parameters and MSEs of the restored images are described below. 'OPT' indicates the optimal parameters that minimizes MSE.

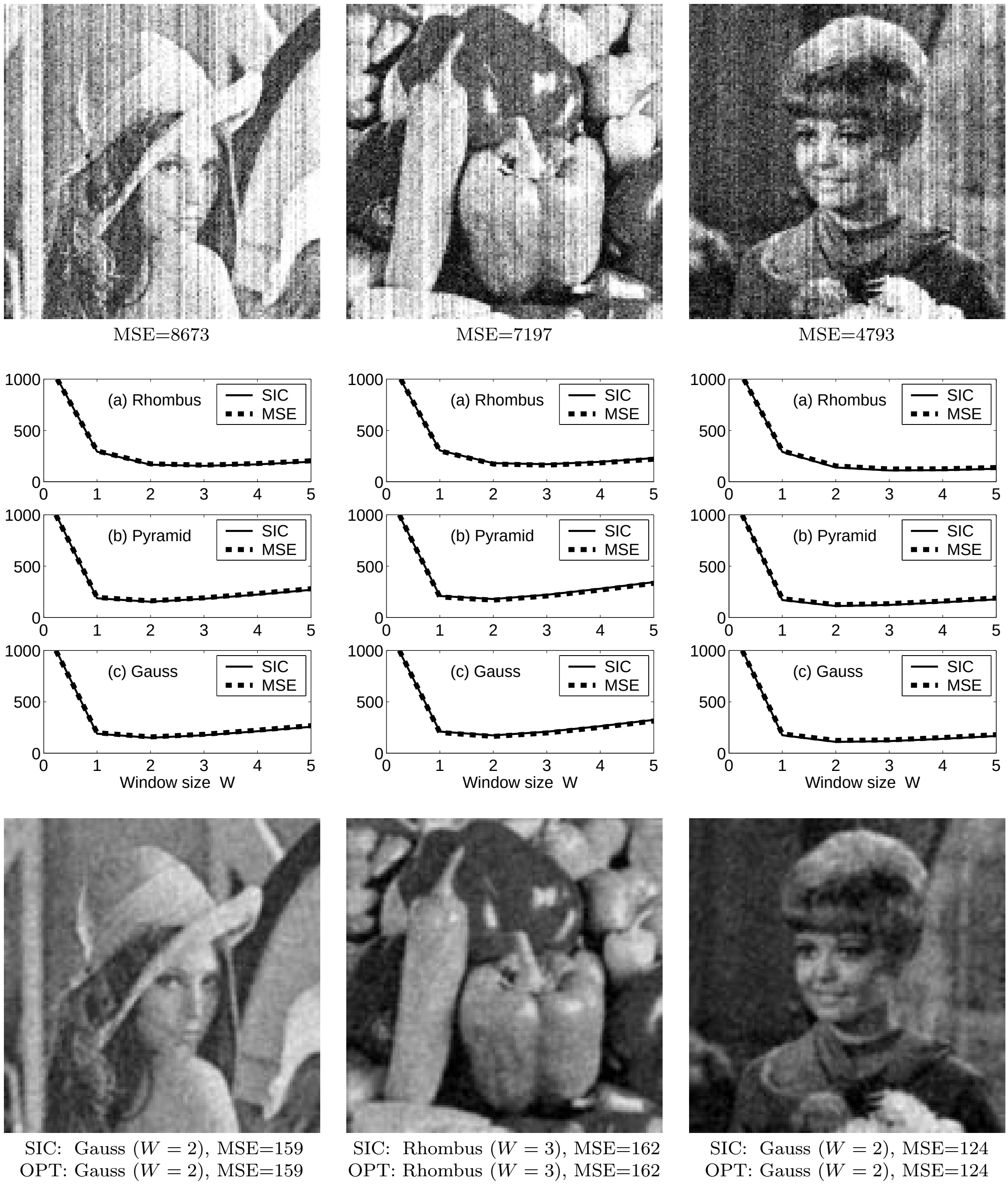


Figure 6: Simulation results when $\sigma^2 = 2500$. The top row shows the degraded images. Their MSEs are described below. The middle row shows the values of SIC and MSE corresponding to each filter. The horizontal axis denotes the window size W. The bottom row shows the restored images by SIC. Selected filter parameters and MSEs of the restored images are described below. 'OPT' indicates the optimal parameters that minimizes MSE.

References

[1] M. R. Banham and A. K. Katsaggelos, "Digital image restoration", IEEE Signal Processing Magazine, vol. 14, no. 2, pp. 24–41, 1997.

[2] M. Belge, M. Kilmer, and E. Miller, "Wavelet domain image restoration with adaptive edge-preserving regularization", IEEE Trans. on Image Processing, vol. 9, no. 4, pp. 597–608, 2000.

[3] S. G. Chang, B. Yu, and M. Vetterli, Adaptive wavelet thresholding for image denoising and compression", IEEE Trans. on Image Processing, vol. 9, no. 9, pp. 1532–1546, 2000.

[4] D. L. Donoho, "De-noising by soft thresholding", IEEE Trans. on Information Theory, vol. 41, no. 3, pp. 613–627, 1995.

[5] E. R. Dougherty and C. R. Giardina, Image Processing — Continuous to Discrete, vol. 1, Prentice-Hall, Inc., New Jersey, 1987.

[6] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation", IEEE Trans. on Image Processing, vol. 1, no. 3, pp. 322–336, 1992.

[7] C. L. Mallows, "Some comments on C_P", Technometrics, vol. 15, no. 4, pp. 661–675, 1973.

[8] E. Oja and H. Ogawa, "Parametric projection filter for image and signal restoration", IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-34, no. 6, pp. 1643–1653, 1986.

[9] S. J. Reeves and R. M. Mersereau, "Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration", Optical Engineering, vol. 29, no. 5, pp. 446–454, 1990.

[10] M. Sugiyama and H. Ogawa, "Subspace information criterion for model selection", Neural Computation, vol. 13, no. 8, 2001. (To appear)

[11] M. Sugiyama and H. Ogawa, "Theoretical and experimental evaluation of subspace information criterion", Machine Learning, Special Issue on New Methods for Model Selection and Model Combination, 2001. (To appear)

[12] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, "A study of methods of choosing the smoothing parameter in image restoration by regularization", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 13, no. 4, pp. 326–339, 1991.

[13] S. Zhong and V. Cherkassky, "Image denoising using wavelet thresholding and model selection", Proc. IEEE Intl. Conf. on Image Processing, vol. 3, pp. 262–265, 2000.