SUBSPACE INFORMATION CRITERION FOR IMAGE RESTORATION — MEAN SQUARED ERROR ESTIMATOR FOR LINEAR FILTERS

Masashi Sugiyama, Daisuke Imaizumi, Hidemitsu Ogawa

Department of Computer Science, 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 that a certain error measure such as the mean squared error (MSE) between the restored image and original image is minimized. However, this is not generally possible since the unknown original image itself is required for evaluating MSE. In this article, we derive a criterion called the subspace information criterion (SIC) for linear filters. SIC gives an unbiased estimate of the expected MSE. By the use of SIC, we give a procedure for optimizing the parameters of the moving-average filter, i.e., the window size and weight pattern. Computer simulations show that SIC gives a very accurate estimate of MSE in various situations, and the proposed procedure actually gives the optimal parameter that minimizes MSE.

Keywords: Image restoration, Mean squared error, Subspace information criterion, Moving-average filter.

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, band-width, and regularization factors [7, 3, 10, 1, 2]. The restoration properties of the filters depend heavily on the values of these parameters.

The quality of restored images is generally evaluated by the *mean squared error* (MSE) between the restored image and original image. If the parameters of filters 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.

In this article, we derive an estimate of MSE called the *subspace information criterion* (SIC) for linear filters, which is originated in the statistical model selection criterion [8, 9]. SIC can be calculated without the original image. The quality of SIC as an approximation of MSE is theoretically substantiated by the fact that SIC is an unbiased estimate of the expected MSE over the noise. Moreover, computer simulations demonstrate that SIC gives a very accurate estimate of MSE.

Since SIC is a good approximation of MSE, it can be used as a substitute for MSE. That is, if the parameters of a filter are determined so that SIC is minimized, then the filter is expected to provide the optimal restoration property. In this article, we will optimize the parameters of the *moving-average filter*, i.e., the window size and weight pattern.

2. PROBLEM FORMULATION

In this section, we formulate the problem of image restoration following the reference [6].

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 observed image g is given as

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

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 as

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

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

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

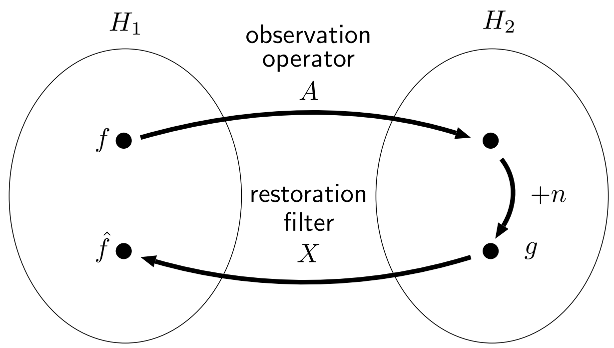


Fig. 1. Formulation of image restoration problem. f is the unknown original image. A is the observation operator. g is the observed image. n is the additive noise. X is a restoration filter. $\hat{f}$ is a restored image.

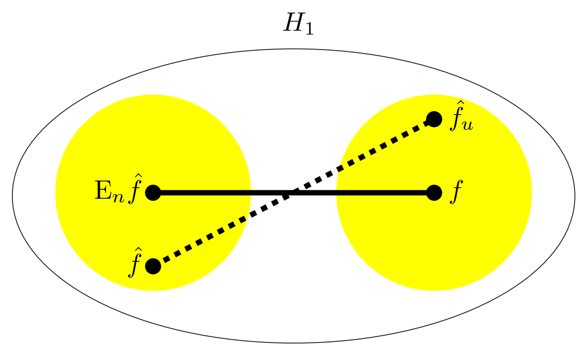


Fig. 2. Basic idea of SIC. The solid line denotes the bias of $\hat{f}$. It can be roughly estimated by the dotted line, which can be calculated (see the text for detail).

where $\|\cdot\|$ denotes the norm in H_1. The norm is typically defined as

$$\|\hat{f} - f\|^2 = \int \Big(\hat{f}(x,y) - f(x,y)\Big)^2 dxdy. \quad (4)$$

Then the problem of image restoration considered in this article is to obtain 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 estimate 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:

$$\mathrm{E}_n n = 0, \quad (5)$$

   where E_n is 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:

$$\mathrm{E}_n \hat{f}_u = f, \quad (6)$$

   where

$$\hat{f}_u = X_u g. \quad (7)$$

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

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

$$\begin{aligned}\mathrm{E}_n\mathrm{MSE}[X] &= \mathrm{E}_n\|\hat{f} - \mathrm{E}_n\hat{f} + \mathrm{E}_n\hat{f} - f\|^2 \\ &= \mathrm{E}_n\|\hat{f} - \mathrm{E}_n\hat{f}\|^2 + 2\mathrm{E}_n\langle \hat{f} - \mathrm{E}_n\hat{f}, \mathrm{E}_n\hat{f} - f\rangle \\ &\quad + \mathrm{E}_n\|\mathrm{E}_n\hat{f} - f\|^2 \\ &= \mathrm{E}_n\|\hat{f} - \mathrm{E}_n\hat{f}\|^2 + \|\mathrm{E}_n\hat{f} - f\|^2, \end{aligned} \quad (8)$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in H_1. The first and second terms in Eq.(8) 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 (5) that the variance of $\hat{f}$ is expressed as

$$\begin{aligned}\mathrm{E}_n\|\hat{f} - \mathrm{E}_n\hat{f}\|^2 &= \mathrm{E}_n\|Xg - \mathrm{E}_n Xg\|^2 \\ &= \mathrm{E}_n\|X(Af+n) - \mathrm{E}_n X(Af+n)\|^2 \\ &= \mathrm{E}_n\|Xn\|^2 \\ &= \mathrm{tr}\,(XQX^*)\,, \end{aligned} \quad (9)$$

where X^* denotes the adjoint of X, and $\mathrm{tr}\,(\cdot)$ denotes the trace of an operator. It follows from Eqs.(6), (2), and (7) that the bias of $\hat{f}$ is expressed as

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

The second and third terms in Eq.(10) can not be directly evaluated since they include an unknown term $\mathrm{E}_n(\hat{f} - \hat{f}_u)$, 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), (7), (1), and (5) that the third term yields

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

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:*

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

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

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

$$\mathrm{E}_n\mathrm{SIC}[X] = \mathrm{E}_n\mathrm{MSE}[X]. \qquad (13)$$

(Proof) It follows from Eqs.(12), (1), (5), (2), and (7) that the first term in SIC yields

$$\begin{aligned}
&\mathrm{E}_n\|(X - X_u)g\|^2 \\
&\quad = \mathrm{E}_n\|(X - X_u)(Af + n)\|^2 \\
&\quad = \|(X - X_u)Af\|^2 + \mathrm{E}_n\|(X - X_u)n\|^2 \\
&\quad = \|(X - X_u)\mathrm{E}_n g\|^2 \\
&\qquad + \mathrm{tr}\left((X - X_u)Q(X - X_u)^*\right) \\
&\quad = \|\mathrm{E}_n\hat{f} - f\|^2 + \mathrm{tr}\left((X - X_u)Q(X - X_u)^*\right). \qquad (14)
\end{aligned}$$

It follows from Eqs.(12), (14), (9), and (8) that

$$\begin{aligned}
\mathrm{E}_n\mathrm{SIC}[X] &= \|\mathrm{E}_n\hat{f} - f\|^2 + \mathrm{tr}\left(XQX^*\right) \\
&= \mathrm{E}_n\mathrm{MSE}[X], \qquad (15)
\end{aligned}$$

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 SIC

In this section, we give a method for optimizing the parameters of the *moving-average filter* [3], 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

$$\{1, 2, \ldots, D\} \times \{1, 2, \ldots, D\}. \qquad (16)$$

Let us define the norm in H_1 as

$$\|f\|^2 = \frac{1}{D^2} \sum_{x,y=1}^{D} \left(f(x, y)\right)^2. \qquad (17)$$

Then Eq.(3) yields a typical definition of MSE in the discrete case:

$$\mathrm{MSE}[X] = \frac{1}{D^2} \sum_{x,y=1}^{D} \left(\hat{f}(x, y) - f(x, y)\right)^2. \qquad (18)$$

Let I be the identity operator on H_1 $(= H_2)$. We assume that the observation operator A and the noise covariance operator Q are given as

$$A = I, \qquad (19)$$

$$Q = \sigma^2 I, \qquad (20)$$

where $\sigma^2 > 0$. In this case, the observed image g is given as

$$g = f + n. \qquad (21)$$

This implies that g itself is an unbiased estimate of the original image f:

$$\mathrm{E}_n g = f. \qquad (22)$$

For this reason, we use the identity operator as X_u:

$$X_u = I. \qquad (23)$$

Note that in the current setting, SIC agrees with the traditional C_P-statistics [4, 5].

4.2. Moving-average filter

The moving-average filter (MAF) restores the image by the weighted sum of the brightness of nearby pixels:

$$\hat{f}(x, y) = \frac{1}{C_{x,y}} \sum_{i,j} w_{i,j}\, g(x - i, y - j), \qquad (24)$$

where $\sum_{i,j}$ is taken over integers i and j such that

$$-W \le i,\ j \le W, \qquad (25)$$

$$1 \le x - i\ , y - j \le D. \qquad (26)$$

The integer W (≥ 0) is called the *window size*, and the set $\{w_{i,j}\}_{i,j=-W}^{W}$ of scalars is called the *weight pattern*. The scalar $C_{x,y}$ is defined as

$$C_{x,y} = \sum_{i,j} w_{i,j}, \qquad (27)$$

where $\sum_{i,j}$ is taken over Eqs.(25) and (26), which depends on x and y. We assume that $C_{x,y}$ is not zero for any x and y.

In the case of MAF, the window size W and weight pattern $\{w_{i,j}\}_{i,j=-W}^{W}$ are the parameters.

4.3. SIC for moving-average filter

By the use of SIC, the parameters can be optimized as follows. First, a set $\mathcal{M}$ of filters with different values of the parameters is prepared.

$$\mathcal{M} = \{X\}. \tag{28}$$

Then SIC is calculated for each filter X in the set $\mathcal{M}$, and the filter $\hat{X}$ that minimizes SIC is selected:

$$\hat{X} = \operatorname*{argmin}_{X \in \mathcal{M}} \mathrm{SIC}[X], \tag{29}$$

where SIC in the current setting is given as

$$\mathrm{SIC}[X] = \frac{1}{D^2} \sum_{x,y=1}^{D} \left(\hat{f}(x,y) - g(x,y)\right)^2 + 2\sigma^2 \mathrm{tr}\,(X) - \sigma^2. \tag{30}$$

The values of the parameters in $\hat{X}$ is expected to be the best. Indeed, the expectation is theoretically supported by Theorem 1, and experimentally demonstrated in Section 5.

Ignoring the effect of verge pixels, we have

$$\mathrm{tr}\,(X) = \frac{1}{D^2} \sum_{x,y=1}^{D} \frac{w_{0,0}}{C_{x,y}} \approx \frac{w_{0,0}}{C}, \tag{31}$$

where the constant C is defined as

$$C = \sum_{i,j=-W}^{W} w_{i,j}. \tag{32}$$

Then SIC is approximated as

$$\mathrm{SIC}[X] \approx \frac{1}{D^2} \sum_{x,y=1}^{D} \left(\hat{f}(x,y) - g(x,y)\right)^2 + \frac{2\sigma^2 w_{0,0}}{C} - \sigma^2, \tag{33}$$

which can be calculated efficiently.

5. COMPUTER SIMULATIONS

In this section, the effectiveness of SIC for MAF is demonstrated through computer simulations.

Let us consider (i) *Lena*, (ii) *Peppers*, and (iii) *Girl* shown in Fig. 3 as original images. The size D of the images is 256 and the values $\{f(x,y)\}_{x,y=1}^{256}$ of the brightness are integers from 0 to 255. We suppose that the noises $\{n(x,y)\}_{x,y=1}^{256}$ are independently subject to the same normal distribution with mean zero and variance σ^2. In this case, the noise covariance operator Q is given by Eq.(20). We calculate SIC by Eq.(33). As candidates of the parameters of MAF, we consider six different window sizes W:

$$W = 0, 1, \ldots, 5. \tag{34}$$

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

(a) Rhombus pattern:

$$w_{i,j}^{(a)} = \max(0, W + 1 - |i| - |j|). \tag{35}$$

(b) Pyramid pattern:

$$w_{i,j}^{(b)} = W + 1 - \max(|i|, |j|). \tag{36}$$

(c) Gauss pattern:

$$w_{i,j}^{(c)} = \frac{1}{2\pi(W/2)^2} \exp\left(-\frac{i^2 + j^2}{2(W/2)^2}\right). \tag{37}$$

The above weight patterns for $W = 2$ are illustrated in Fig. 4.

Figs. 5, 6, and 7 display the simulation results when the noise variance σ^2 is 900, 1600, and 2500, respectively. The top rows show the degraded images $\{g(x,y)\}_{x,y=1}^{256}$. Their MSEs measured by Eq.(18) are described below the images. The middle rows show the values of MSE and SIC corresponding to each filter. The horizontal axis denotes the window size W. The bottom rows show the restored images 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 type of the original image, noise variance, window size, and weight pattern. The restored images in the bottom rows show that the filter parameter that minimizes SIC actually minimizes MSE, i.e., the optimal filter parameter can be obtained by SIC.

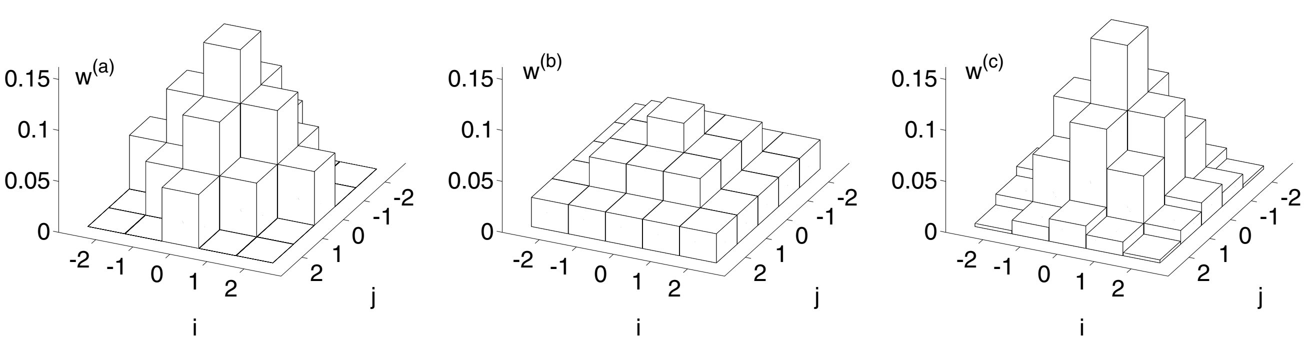
6. CONCLUSIONS

We derived an unbiased estimate of the expected mean squared error for linear filters. We named the estimate the subspace information criterion (SIC) following model selection publications. By the use of SIC, we proposed a procedure for optimizing the parameters of the moving-average filter. Computer simulations showed that SIC gives a very accurate estimate of MSE in various situations, and the proposed procedure actually gave the optimal parameter that minimizes MSE.

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

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

Fig. 3. Original images.



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

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

7. 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 Transactions on Image Processing*. (To appear)

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

[4] C. L. Mallows, Choosing a Subset Regression, *Presented at the Central Regional Meeting of the Institute of Mathematical Statistics*, Manhattan, Kansas, 1964.

[5] C. L. Mallows, Some Comments on C_P, *Technometrics*, vol. 15, no. 4, pp. 661–675, 1973.

[6] H. Ogawa, Image and Signal Restoration, *The Journal of the Institute of Electronics, Information and Communication Engineers*, vol. 71, no. 5–8, pp. 491–197, 593–601, 739–748, 828–835, 1988.

[7] E. Oja and H. Ogawa, Parametric Projection Filter for Image and Signal Restoration, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 6, pp. 1643–1653, 1986.

[8] M. Sugiyama and H. Ogawa, Subspace Information Criterion for Model Selection, *Neural Computation*, 2001. (To appear)

[9] 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)

[10] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 326–339, 1991.

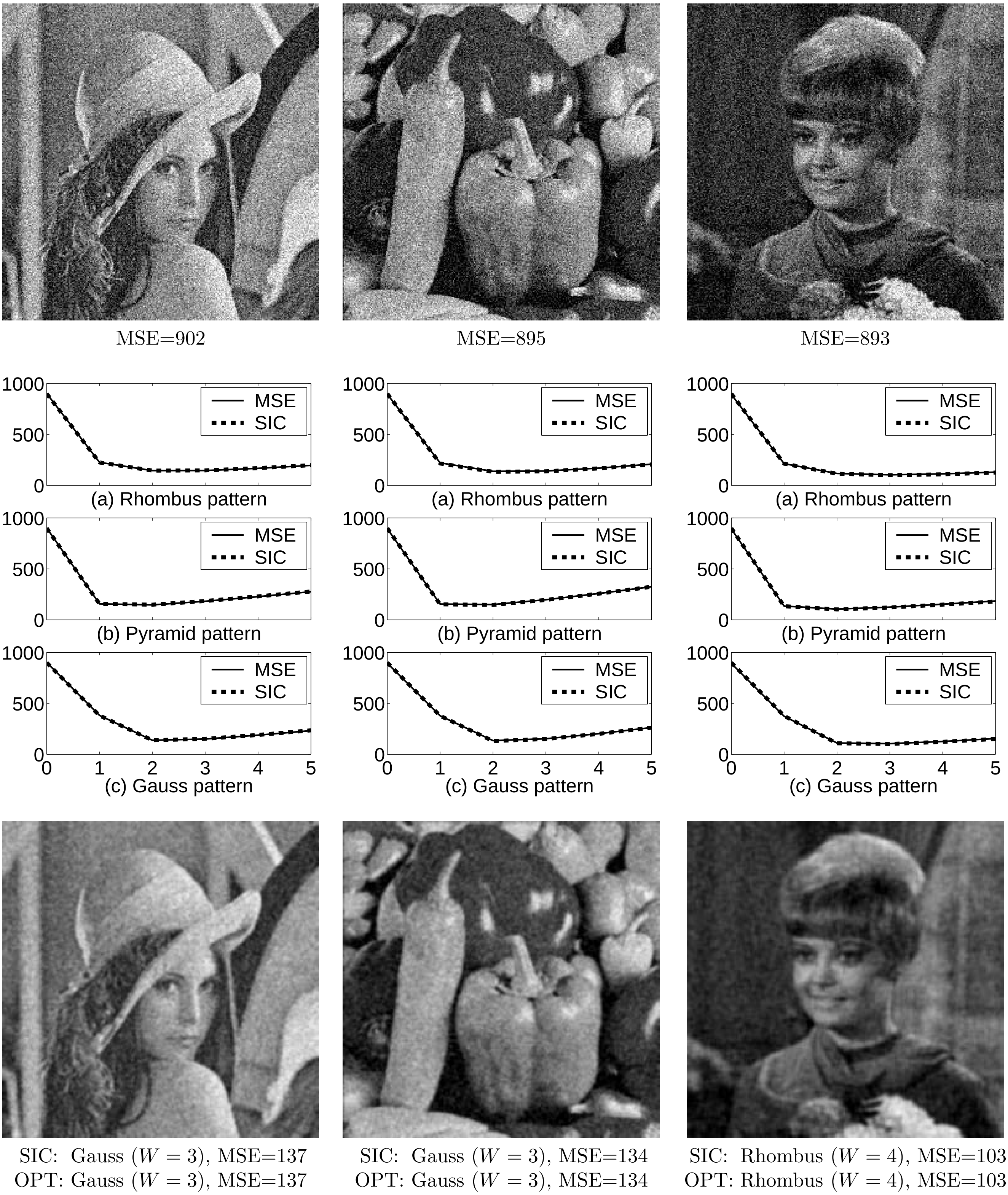
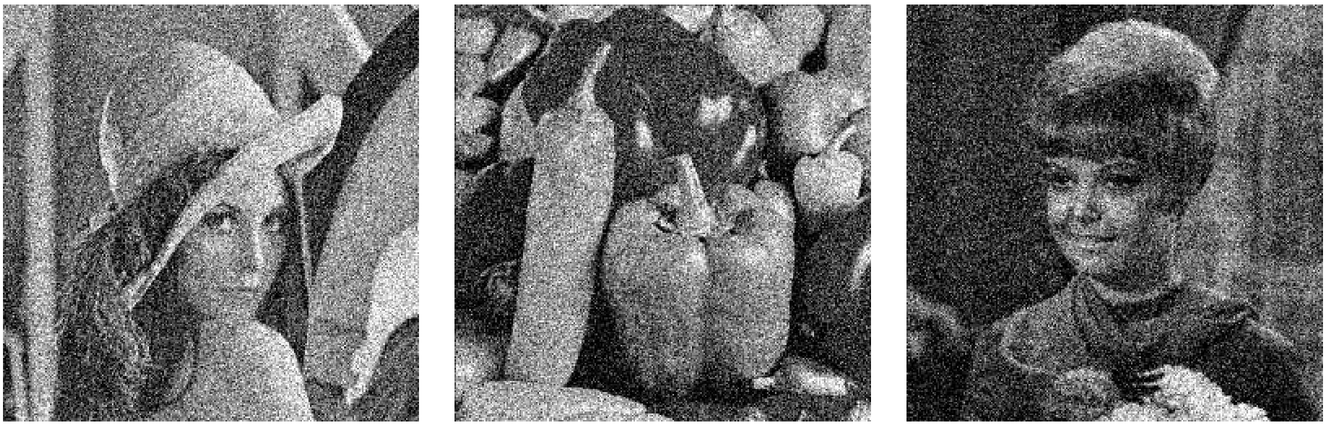
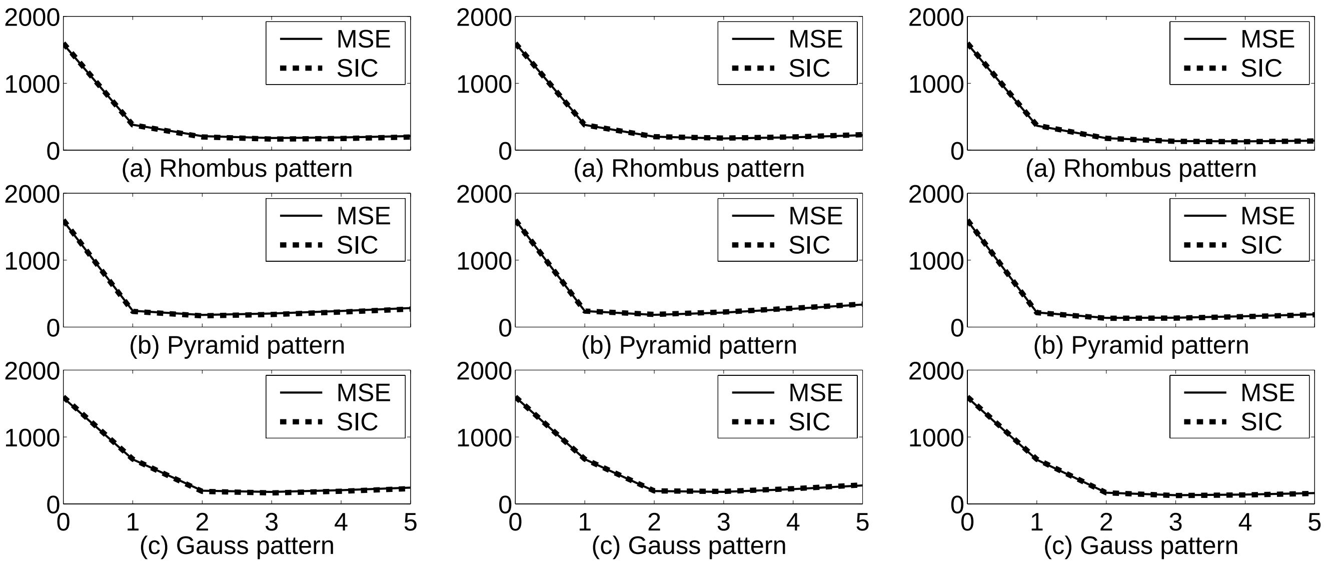


Fig. 5. Simulation results when $\sigma^2 = 900$. The top row shows the degraded images. Their MSEs are described below. The middle row shows the values of MSE and SIC corresponding to each filter. The horizontal axis denotes the window size W. The bottom row shows the restored images by SIC. Selected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



MSE=1587 MSE=1598 MSE=1590



SIC: Gauss ($W = 4$), MSE=179
OPT: Gauss ($W = 4$), MSE=179

SIC: Rhombus ($W = 4$), MSE=176
OPT: Rhombus ($W = 4$), MSE=176

SIC: Gauss ($W = 4$), MSE=129
OPT: Gauss ($W = 4$), MSE=129

Fig. 6. Simulation results when $\sigma^2 = 1600$. The top row shows the degraded images. Their MSEs are described below. The middle row shows the values of MSE and SIC 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.

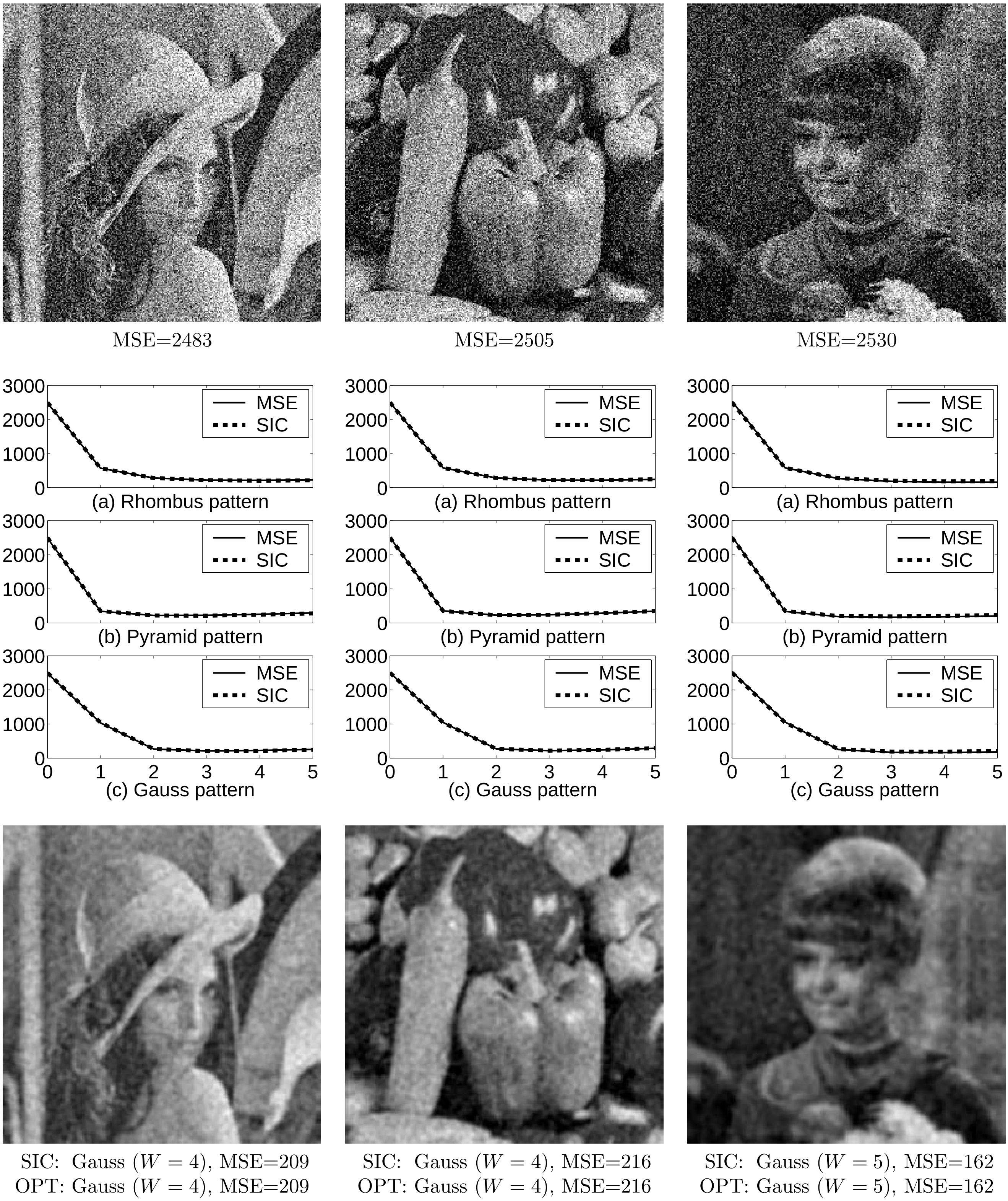


Fig. 7. Simulation results when $\sigma^2 = 2500$. The top row shows the degraded images. Their MSEs are described below. The middle row shows the values of MSE and SIC corresponding to each filter. The horizontal axis denotes the window size W. The bottom row shows the restored images by SIC. Selected filter parameters and MSEs of the restored images are described below. 'OPT' indicates the optimal parameters that minimizes MSE.