Methods for Choosing the Regularization Parameter and Estimating the Noise Variance in Image Restoration and Their Relation

Nikolas P. Galatsanos, *Member, IEEE,* and Aggelos K. Katsaggelos, *Member, IEEE*

***Abstract*—The application of regularization to ill-conditioned problems necessitates the choice of a regularization parameter which trades fidelity to the data with smoothness of the solution. The value of the regularization parameter depends on the variance of the noise in the data. In this paper the problem of choosing the regularization parameter and estimating the noise variance in image restoration is examined. An error analysis based on an objective mean square error (MSE) criterion is used to motivate regularization. Two new approaches for choosing the regularization parameter and estimating the noise variance are proposed. The proposed and existing methods are compared and their relation to linear minimum mean square error (LMMSE) filtering is examined. Experiments are presented that verify the theoretical results.**

I. Introduction

For an $M \times M$ image degraded by a linear space invariant blur and independent identically distributed (IID) zero mean additive noise, the imaging equation is

$$g = Hf + n \tag{1}$$

where the vectors g, f, and n are $M^2 \times 1$ lexicographic orders of the degraded observed image, the original image, and the noise, respectively. The $M^2 \times M^2$ degradation matrix H is block Toeplitz and ill conditioned [1]. In the rest of this paper the circulant-to-Toeplitz approximation will be used along with the discrete Fourier transform (DFT) to simplify all computations [1], [8].

One of the approaches to obtain an estimate of f from (1) is to minimize the criterion

$$J(f) = \|g - Hf\|^2 \tag{2}$$

which results in the pseudo-inverse estimate

$$f^+ = (H^tH)^{-1}H^tg \tag{3}$$

where t denotes the transpose of a matrix or a vector. It is well known [1] that such an estimate is extremely noisy due to the amplification of the noise by the inverse of the ill-conditioned operator H.

Regularization is an effective method for obtaining satisfactory solutions to problems that involve inversion of ill-conditioned operators [32]. According to the regularization approach the solution of (1) is replaced by the minimization of

$$J_\lambda(f) = \|g - Hf\|^2 + \lambda \|Qf\|^2 \tag{4}$$

which yields the estimate

$$\hat{f}(\lambda) = (H^tH + \lambda Q^tQ)^{-1}H^tg = A(\lambda)g \tag{5}$$

where Q is the regularization operator [32]. The role of Q is two-fold [15]: a) to move the small eigenvalues of H away from zero while leaving the large eigenvalues unchanged, and b) to incorporate prior knowledge about f into the restoration process. The smoothness of f is the prior knowledge on which the selection of Q is usually based upon. Thus the criterion of (4) contains two terms; the first, $\|g - Hf\|$, expresses the fidelity to the available data g and the second, $\|Qf\|$, the smoothness of the estimate. Therefore, the regularization parameter λ represents the trade-off between fidelity to the data and smoothness of the estimate $\hat{f}(\lambda)$. The determination of the proper value of λ is an important problem and depends on the variance of the noise σ^2 and the properties of H, Q, and f. In this paper the problem of choosing the regularization parameter and estimating the noise variance is examined. A study on the stability of the estimate of the regularization parameter has recently appeared in [31]. Some of our observations and conclusions were also verified theoretically in [26] in an asymptotic sense, for specific forms of H, Q, and f.

More specifically, the rest of this paper is organized as follows: in Section II the norm of the difference of the regularized solution $\hat{f}(\lambda)$ from the original image f is analyzed as a function of the regularization parameter. The use of regularization is rigorously justified by showing that $\hat{f}(\lambda)$ yields a smaller error than f^+. In Section III existing methods for selecting λ are reviewed and the foundation is laid for the material to follow. In Section IV two new methods for choosing the regularization parameter are presented. In Section V the problem of estimating the noise variance is examined and two new methods are proposed. The proposed and existing methods for choosing the regularization parameter are compared in Section VI. It is shown that for a special choice of the regularization operator Q, all methods, except one, yield the same value

Manuscript received August 14, 1991; revised November 18, 1991.
N. P. Galatsanos is with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616.
A. K. Katsaggelos is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208.
IEEE Log Number 9200284.

for λ equal to the noise variance σ^2 [6], [7]. For this choice of Q the resulting regularized estimate $\hat{f}(\lambda)$ is the linear minimum mean square error (LMMSE) solution known as the Wiener filter [1]. In Section VII experiments are presented where the previous theoretical results are verified. Finally, Section VIII contains a discussion of the results, conclusions, and plans for future research.

II. Error Analysis

It is commonly assumed in image restoration that $\hat{f}(\lambda)$ is a much more preferable solution than f^+ [1]. However, no rigorous justification of this assumption, based on some objective criterion, has appeared in the image restoration literature. In this section an analysis is presented where the norm of the error between the original image f and the regularized estimate $\hat{f}(\lambda)$ is presented as a function of λ. This analysis justifies the use of regularization and rigorously shows that $\hat{f}(\lambda)$ is a better estimate than f^+.

Throughout this paper it is assumed that f in (1) is deterministic, the noise variance n is IID with variance σ^2, and H and Q are block-circulant matrices. From (3) and (5) we have

$$\hat{f}(\lambda) = (H^tH + \lambda Q^tQ)^{-1}H^tg = (H^tH + \lambda Q^tQ)^{-1}H^tHf^+ = P(\lambda)f^+. \quad (6)$$

Following similar steps as in [11] the mean square error (MSE) is given by

$$\begin{aligned} E[\|e(\lambda)\|^2] &= E[\|f - \hat{f}(\lambda)\|^2] \\ &= E[(f - P(\lambda)f^+)^t(f - P(\lambda)f^+)] \\ &= E[(f^+ - f)^tP(\lambda)^tP(\lambda)(f^+ - f)] \\ &\quad + E[(P(\lambda)f - f)^t(P(\lambda)f - f)] \quad (7) \end{aligned}$$

where $E[\cdot]$ denotes the expectation operator. Since $E[\hat{f}(\lambda)] = P(\lambda)f$, the first term of (7) is equal to the variance of $\hat{f}(\lambda)$ while the second term is equal to the bias of the estimate $\hat{f}(\lambda)$.

By using the circulant approximation for H and Q we can write in the DFT domain:

$$\text{Var}[\hat{f}(\lambda)] = \sigma^2 \sum_{i=1}^{M^2} \frac{|h_i|^2}{(|h_i|^2 + \lambda|q_i|^2)^2}$$

and

$$\text{Bias}[\hat{f}(\lambda)] = \sum_{i=1}^{M^2} \frac{|F_i|^2\lambda^2|q_i|^4}{(|h_i|^2 + \lambda|q_i|^2)^2} \quad (8)$$

where h_i and q_i are the eigenvalues of H and Q, respectively, and F_i is the ith component of the DFT of f. Clearly, the pseudo-inverse solution f^+, which corresponds to $\hat{f}(\lambda)$ with $\lambda = 0$, yields an error which is due only to the variance term and is equal to

$$E[\|e(0)\|^2] = \sigma^2 \sum_{i=1}^{M^2} \frac{1}{|h_i|^2}.$$

We now turn to the analysis of the variance and bias terms in (8). Taking their derivatives with respect to λ we obtain

$$\frac{\partial \text{Var}[\hat{f}(\lambda)]}{\partial\lambda} = -2\sigma^2 \sum_{i=1}^{M^2} \frac{|h_i|^2|q_i|^2}{(|h_i|^2 + \lambda|q_i|^2)^3} < 0 \quad (9)$$

$$\frac{\partial \text{Bias}[\hat{f}(\lambda)]}{\partial\lambda} = 2 \sum_{i=1}^{M^2} \frac{\lambda|F_i|^2|q_i|^4|h_i|^2}{(|h_i|^2 + \lambda|q_i|^2)^3} > 0. \quad (10)$$

From (8) and (9) we conclude that the variance is a strictly positive, monotonically decreasing function of λ, for $\lambda > 0$ and that $\text{Var}[\hat{f}(\infty)] = 0$. Similarly, from (8) and (10) we conclude that the bias is a positive, monotonically increasing function of λ, for $\lambda > 0$ and that $\text{Bias}[\hat{f}(0)] = 0$ and $\text{Bias}[\hat{f}(\infty)] = \|f\|^2$. Thus the total error as a function of λ is a sum of a monotonically decreasing and a monotonically increasing function of λ, and therefore, it has either one minimum or one maximum for $0 \le \lambda < \infty$.

Examining the derivative

$$\frac{\partial E[\|e(\lambda)\|^2]}{\partial\lambda} = 2 \sum_{i=1}^{M^2} \frac{|h_i|^2|q_i|^2(\lambda|F_i|^2|q_i|^2 - \sigma^2)}{(|h_i|^2 + \lambda|q_i|^2)^3}$$

it is clear that

$$\frac{\partial E[\|e(\lambda)\|^2]}{\partial\lambda} < 0, \quad \text{for } 0 < \lambda < \frac{\sigma^2}{(|F_i|^2|q_i|^2)_{\max}}$$

and

$$\frac{\partial E[\|e(\lambda)\|^2]}{\partial\lambda} > 0, \quad \text{for } \frac{\sigma^2}{(|F_i|^2|q_i|^2)_{\min}} < \lambda < \infty.$$

Therefore, the error function has a unique minimum for $0 \le \lambda < \infty$, which lies in the interval

$$\left[\frac{\sigma^2}{(|F_i|^2|q_i|^2)_{\max}}, \frac{\sigma^2}{(|F_i|^2|q_i|^2)_{\min}}\right].$$

The value of λ for which the minimum error is achieved is denoted by λ_{MSE}.

From the above it is clear that $E[\|e(\lambda)\|^2]$ is a decreasing function for $0 \le \lambda < \lambda_{\text{MSE}}$. Thus for any λ in this interval, $E[\|e(\lambda)\|^2] < E[\|e(0)\|^2]$, and therefore, $\hat{f}(\lambda)$ is a better estimate than f^+ in a MSE sense.

In Fig. 1 an example is shown using image data. The $\text{Var}[\hat{f}(\lambda)]$, $\text{Bias}[\hat{f}(\lambda)]$, and $E[\|e(\lambda)\|^2]$ terms are plotted using a log 10 scale. For this example, we have

$$E[\|e(0)\|^2] = \sigma^2 \sum_{i=1}^{M^2} \frac{1}{|h_i|^2} > \|f\|^2 = E[\|e(\infty)\|^2].$$

Thus the regularized solution $\hat{f}(\lambda)$ yields a smaller MSE than f^+ for any $\lambda > 0$. This case is common in many practical applications where H is ill conditioned.

From the error analysis presented above we conclude that the regularized solution $\hat{f}(\lambda)$ with $\lambda > 0$ introduces a bias in the estimate of f. However, this is a price well worth paying because in this way the total MSE can be reduced.

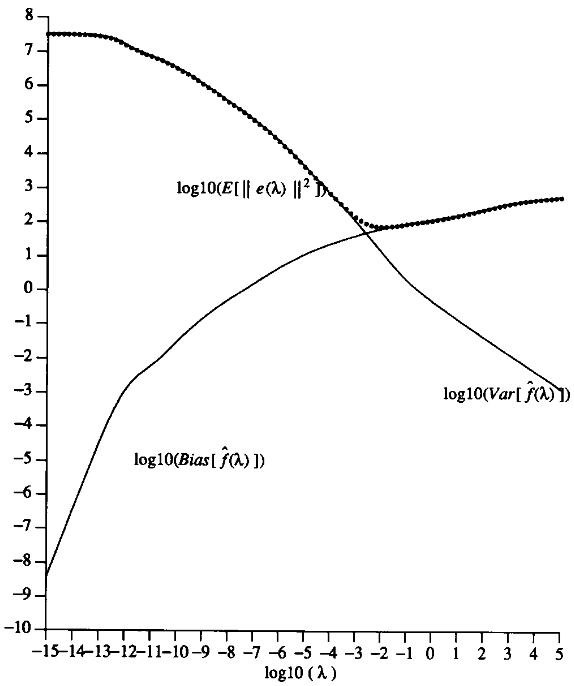


Fig. 1. Demonstration of the properties of $E[\|e(\lambda)\|^2]$. The Lena 128 by 128 image with 7 by 7 blur, Q, the 2-D Laplacian and SNR = 20 dB was used. Log10 (λ) is shown at the x-axis. Log10 of Var $[f(\lambda)]$, the Bias $[f(\lambda)]$ and $E[\|e(\lambda)\|^2]$ (dotted line) are shown on the y-axis.

III. Existing Methods for Choosing the Regularization Parameter

A. Methods That Require Knowledge of the Noise Variance σ^2

Hunt [12] was the first to address the problem of selecting the value of the regularization parameter in restoration problems. He used a deterministic constrained least squares (CLS's) framework. According to it, the parameter λ was selected such that the following equation holds:

$$\|g - H\hat{f}(\lambda)\|^2 = \|(I - HA(\lambda))g\|^2 = \|n\|^2 = \epsilon^2 = M^2\sigma^2. \quad (11)$$

Let us denote by λ_{CLS} such a parameter. Apart from the fact that the above choice of λ requires the prior knowledge of the noise variance, it has been observed and reported by a number of researchers (see, for example, [4]), that this choice of λ yields an oversmooth solution $\hat{f}(\lambda)$.

Using the CLS approach is equivalent to assuming that the error $(g - H\hat{f}(\lambda))$ is IID with variance σ^2. In other words, the ith component of the error $(g - H\hat{f}(\lambda))$ is Gaussian and $\|g - H\hat{f}(\lambda)\|^2$ is Chi-square distributed with variance σ^2 and M^2 degrees of freedom [33], [34]. In this context, but with a different application in mind, Wahba [36] and later Hall and Titterington [10] proposed in analogy to regression the notion of the *equivalent degrees of freedom* (EDF). The EDF takes into account the linear dependency between g and $\hat{f}(\lambda)$, therefore, $\|g - H\hat{f}(\lambda)\|^2$ is Chi-square distributed with variance σ^2 and trace $(I - HA(\lambda))$ degrees of freedom. Thus in this case the constraint equation to be used for computing λ is given by

$$\|g - H\hat{f}(\lambda)\|^2 = \|(I - HA(\lambda))g\|^2 = \sigma^2 \text{ trace } [I - HA(\lambda)]. \quad (12)$$

Very recently the EDF idea was experimentally tested for the image restoration problem [30].

Katsaggelos *et al.* [15], [16] used a set theoretic (ST) formulation for the image restoration problem. The ellipsoid

$$\|Qf\|^2 \leq E^2 \quad (13)$$

is used along with the ellipsoid defined by (11) by replacing the = sign by the $\leq$ sign in front of $\|n\|^2$ to specify a solution $\hat{f}(\lambda)$. The centers of ellipsoids bounding the intersection of these two ellipsoids are given by

$$\hat{f} = \left[p_1 \frac{H^tH}{\epsilon^2} + p_2 \frac{Q^tQ}{E^2}\right]^{-1} \cdot p_1 \frac{H^t}{\epsilon^2} g \quad (14)$$

where $p_1 + p_2 = 1$ and $p_1, p_2 \geq 0$. For $p_1 = p_2$, (14) has been derived in [14] using Miller's regularization approach, and it becomes identical in form with (5). In this case the regularization parameter is equal to

$$\lambda_{ST} = (\epsilon/E)^2. \quad (15)$$

It was shown, in [15], [16], using geometric arguments that

$$\lambda_{CLS} \geq \lambda_{ST}. \quad (16)$$

Clearly the ST method requires the prior knowledge of both ϵ and E, since although only their ratio is used in (15), a posterior check of the intersection of the two ellipsoids is required. Good estimates of f have been reported even when a loose bound is used in (13), such as $E^2 = \|Qg\|^2$. An iterative algorithm for the simultaneous restoration of the image and the determination of the regularization parameter by establishing tight bounds in (11) and (13) is presented in [13].

The minimization of the weighted error norm

$$E[\|He(\lambda)\|^2] = E[\|Hf - H\hat{f}(\lambda)\|^2]$$

has also been used as a criterion for choosing the regularization parameter [9], [25], [10]. The justification for this criterion is that data points which correspond to large eigenvalues of H are more reliable, and thus they are weighted heavier. This weighted error is often called *predicted mean square error* (PMSE) [25], [31]. It is easy to show that

$$E[\|He(\lambda)\|^2] = \|(I - HA(\lambda))g\|^2 + 2 \cdot \sigma^2[\text{trace } [HA(\lambda)] - M^2]. \quad (17)$$

Thus if σ^2 is known, (17) can be minimized directly to obtain the value of the regularization parameter. This value of λ is henceforth denoted by λ_{PMSE}. By using a similar analysis as in Section II it can be shown that the PMSE function has also a unique minimum for $0 \leq \lambda < \infty$.

B. Methods That Do Not Require Knowledge of the Noise Variance σ^2

In many practical situations the noise variance σ^2 is not known; therefore, the methods of subsection III-A cannot be used. Cross validation (CV) is a method that allows the selection of the regularization parameter when σ^2 is not known. CV has been studied by statisticians in the context of ill-posed problems for a long time (see, for example, [9], [27], [35], [38]). However, this mathematical tool has attracted the attention of the image recovery community only very recently [5], [23], [24], [26], [31]. It was shown experimentally that CV yields sharper estimates of the original image than the CLS method. The same observation was made by Craven and Wahba [3] in a very similar regularization problem, that of fitting noisy data with splines.

The ordinary cross validation (OCV) and the generalized cross-validation (GCV) functions [9], [38] can be derived from the *leave one out* principle (see [22] for a detailed step-by-step derivation). In [9] it has been shown that GCV has certain advantages over OCV. However, for image restoration where circulant matrices are used, the GCV and the OCV functions are the same [9]. The GCV function can be written as

$$\mathrm{CV}(\lambda) = \frac{\| (I - HA(\lambda))g \|^2}{[\mathrm{trace}\,(I - HA(\lambda))]^2}. \tag{18}$$

If the DFT domain is used, (18) becomes

$$\mathrm{CV}(\lambda) = \frac{\sum_{i=1}^{M^2} \frac{\lambda^2 |q_i|^4 |G_i|^2}{(|h_i|^2 + \lambda |q_i|^2)^2}}{\left(\sum_{i=1}^{M^2} \frac{\lambda |q_i|^2}{|h_i|^2 + \lambda |q_i|^2} \right)^2}. \tag{19}$$

The CV function is related to the PMSE function in (17). It was shown by Golub *et al.* [9] for 1-D signals and Solo [26] for 2-D signals that λ_{PMSE} is asymptotically equal (as $M^2 \to \infty$) to λ_{CV}, the value of λ that minimizes CV (λ).

IV. Proposed Methods for Choosing the Regularization Parameter

A. Methods That Require Knowledge of the Noise Variance σ^2

Direct minimization of the norm of the MSE in (7) with respect to λ requires knowledge of f, which is not available (see Bias $[\hat{f}(\lambda)]$ in (8)). This is very common in estimation problems where the MSE is used as the optimization criterion. A consequence of this difficulty (along with the asymptotic equivalence of λ_{PMSE} and λ_{CV}) is the popularity of the PMSE criterion for choosing the regularization parameter.

Another approach for evaluating λ by directly minimizing the MSE function is presented. The MSE function can be written as

$$E[\| e(\lambda) \|^2] = \| f \|^2 + E[\| \hat{f}(\lambda) \|^2] - 2E[f^t \hat{f}(\lambda)].$$

However, as was also observed in [25] and [26], since the term $\| f \|^2$ does not depend on λ, the minimization of $E[\| e(\lambda) \|^2]$ is equivalent to the minimization of

$$E[\| \hat{f}(\lambda) \|^2] - 2E[f^t \hat{f}(\lambda)]. \tag{20}$$

Using the circulant assumption and the DFT domain it is easy to show that

$$E[\| \hat{f}(\lambda) \|^2] = E\left[\sum_{i=1}^{M^2} \frac{|h_i|^2 |G_i|^2}{(|h_i|^2 + \lambda |q_i|^2)^2} \right] \tag{21}$$

and

$$\begin{aligned} E[f^t \hat{f}(\lambda)] &= E\left[\sum_{i=1}^{M^2} \frac{|h_i|^2 |F_i|^2}{(|h_i|^2 + \lambda |q_i|^2)} \right] \\ &= E\left[\sum_{i=1}^{M^2} \frac{|G_i|^2 - \sigma^2}{(|h_i|^2 + \lambda |q_i|^2)} \right]. \end{aligned} \tag{22}$$

The expectation on the right-hand side of (21) and (22) can be omitted, since $|G_i|^2$ is known. Thus the value of λ which minimizes (20), denoted by λ_{MSE}, can be computed directly from the data provided that σ^2 is known. Equations (20) and (22) also appeared in [25] and [26]. The asymptotic properties of the value of λ that minimizes (20) were studied in [26] for specific forms of H, Q, and f. It was claimed there that this approach is likely to offer the best choice of λ.

Substituting (21) and (22) into (20) and setting the derivative of (20) equal to zero yields

$$\sum_{i=1}^{M^2} \frac{\lambda^2 |q_i|^4 |G_i|^2}{(|h_i|^2 + \lambda |q_i|^2)^3} - \sigma^2 \sum_{i=1}^{M^2} \frac{\lambda |q_i|^2}{(|h_i|^2 + \lambda |q_i|^2)^2} = 0$$

or equivalently

$$\| Q^{-1}[(I - HA(\lambda))]^{3/2} g \|^2 = \sigma^2 \,\mathrm{trace}\,[Q^{-2}(I - HA(\lambda))^2]. \tag{23}$$

Equation (23) serves as the final equation for estimating λ_{MSE}. Its counterparts are (11) and (12) for respectively determining λ_{CLS} and λ_{EDF}. It is shown in Appendix-A that λ_{PMSE} satisfies

$$\| (HA(\lambda))^{1/2}(I - HA(\lambda))g \|^2 = \sigma^2 \cdot \mathrm{trace}\,[HA(\lambda)(I - HA(\lambda))]. \tag{24}$$

Because of the asymptotic equality of λ_{PMSE} and λ_{CV} shown in [9], [26], λ_{CV} also satisfies (24). In Section VII, experiments are presented that verify the above results.

B. Methods That Do Not Require Knowledge of the Noise Variance σ^2 [7]

Based on a Monte Carlo simulation study by Thompson *et al.* [30], it was observed that the GCV function may not have a unique minimum. It has also been observed that in some cases λ_{PMSE} results in severe undersmoothing of the estimate $\hat{f}(\lambda)$ [25] (for a detailed discussion see

Section VI). Thus in this subsection an alternative approach to choosing the regularization parameter without knowledge of σ^2 is presented. A stochastic formulation of the problem is used [4], [20], [37] and Gaussian *posteriors* and *priors* are assumed.

The *posterior is given by*

$$p(g \mid f, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{M^2/2} \cdot \exp\left[\left(\frac{-1}{2\sigma^2}\right)(g - Hf)^t(g - Hf)\right]. \tag{25}$$

In defining the *prior density* of the data we assume that $\lambda^{1/2}\, Qf$ is a white Gaussian signal with variance σ^2. This is a reasonable assumption if Q is chosen appropriately. Since most real life images have high local correlation, a local differential operator Q (like the 2-D Laplacian [12]) will yield an almost IID signal. The parameter λ can be viewed as a parameter that *scales* the variance of Qf to σ^2. In this context the regularization parameter is called the *hyperparameter* of the prior [4], [20]. Then it is straightforward to show that the prior of f is given by

$$p(f \mid \lambda, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{M^2/2} \det\,[\lambda Q^t Q]^{1/2} \cdot \exp\left[\left(\frac{-\lambda}{2\sigma^2}\right)(Qf)^t(Qf)\right] \tag{26}$$

where det $[A]$ denotes the determinant of A.

Using the previous assumptions, the hyperparameter can be estimated by maximizing a marginal likelihood function which is obtained by integrating out f from the joint posterior $p(g, f \mid \lambda, \sigma^2)$ [4]. This marginal likelihood function is given by

$$L(g \mid \lambda, \sigma^2) = \int_{-\infty}^{\infty} p(g \mid f, \sigma^2) \cdot p(f \mid \lambda, \sigma^2)\, df. \tag{27}$$

It is shown in Appendix B that a λ maximizing the likelihood function in (27), henceforth denoted by λ_{ML}, can be obtained by minimizing

$$ML(\lambda) = \frac{g^t(I - HA(\lambda))g}{(\det\,[I - HA(\lambda)])^{1/M^2}}. \tag{28}$$

Such a λ also satisfies

$$\|(I - HA(\lambda))^{1/2} g\|^2 = M^2\sigma^2 \tag{29}$$

as shown in Appendix B. Notice that the minimization of (28) as in the case of the CV method does not require knowledge of σ^2. An equation of similar form to (28) was derived by Wahba [37] for the problem of fitting noisy data with splines.

V. Noise Variance Estimation

Estimating the noise variance from the data g generated by the imaging model of (1) is a well-researched problem. The classical approach to this problem [1] is to use a smooth region of the image and estimate the noise variance from that region. In [21] the noise variance is estimated by first generating a sequence of variance estimates, by using various tesselations of the noisy image and then obtaining an estimate of it. Recursive noise estimators based on a maximum likelihood principle have been proposed in [2], [28], [29]. Iterative noise variance estimators based on the expectation-maximization algorithm have been proposed in [18], [19].

Wahba [36] proposed a single step noise variance estimator based on the assumption that $\lambda_{CV} \sim \lambda_{EDF}$, for the problem of computing confidence intervals for cross validated smoothing splines. More specifically λ_{CV} is used with (12), yielding the estimate

$$\hat{\sigma}^2_{EDF} = \frac{\|(I - HA(\lambda_{CV}))g\|^2}{\text{trace}\,[I - HA(\lambda_{CV})]}. \tag{30}$$

Theoretically, this estimator has not yet been justified [38]. In spite of this fact, experimental evidence presented by a number of researchers supports the choice of this estimator [5], [22], [31], [36].

Based on the relations shown in Sections III and IV, two new single-step noise variance estimators are proposed. The first estimator is based on the properties of the CV function. More specifically, λ_{CV} and λ_{PMSE} are asymptotically equivalent as shown in [9], [26], and therefore, (24) is also satisfied with λ_{CV}. Thus it can be used to estimate the noise variance, according to

$$\hat{\sigma}^2_{CV} = \frac{\|(HA(\lambda_{CV}))^{1/2}(HA(\lambda_{CV}) - I)g\|^2}{\text{trace}\,[HA(\lambda_{CV})(I - HA(\lambda_{CV}))]}. \tag{31}$$

This estimator has very similar form to the estimator of (30), the only difference being the multiplication with the matrix $HA(\lambda_{CV})$ in both the numerator and denominator.

The second estimator is based on the properties of the marginal likelihood function. As mentioned in the previous section, λ_{ML} satisfies (29). Therefore, the following estimate of the noise variance is proposed

$$\hat{\sigma}^2_{ML} = \left(\frac{1}{M^2}\right)\|(I - HA(\lambda_{ML}))^{1/2} g\|^2. \tag{32}$$

The assumption that Qf is an IID Gaussian signal is implied in the derivation of this estimator. Thus this estimator will perform poorly for a Q that does not whiten f "well enough." In contrast, a similar assumption is not necessary for the estimator of (31). Thus it is expected that $\hat{\sigma}^2_{CV}$ is more stable to the changes in the selection of Q. In Section VII, experiments are presented where these estimators are tested and the previous statement is verified.

VI. Comparison of the Values of the Regularization Parameter Obtained by Different Methods

In this section we compare the values of the regularization parameters obtained by the methods in Sections III and IV. In some cases a rigorous mathematical comparison could not be obtained and thus we resort to intuitive arguments backed by experiments. Since λ dictates the level of smoothness of $\hat{f}(\lambda)$, $\lambda_a > \lambda_b$ implies that $\hat{f}(\lambda_a)$ is smoother than $\hat{f}(\lambda_b)$. This does not reveal anything about the MSE of the estimate, which in any case is a debatable criterion for deciding image quality.

A. General Case (Any Q)

In order to compare the CLS and EDF methods, the properties of $\|(I - HA(\lambda))g\|$ as a function of λ need to be examined. It is easy to see using the DFT domain that

$$\frac{\partial \|(I - HA(\lambda))g\|^2}{\partial \lambda} = 2 \sum_{i=1}^{M^2} \frac{\lambda |q_i|^4 |h_i|^2 |G_i|^2}{(|h_i|^2 + \lambda |q_i|^2)^3} > 0.$$

Thus $\|(I - HA(\lambda))g\|^2$ is an increasing function of λ. Therefore, since $M^2 >$ trace $[(I - HA(\lambda)]$ we obtain that

$$\lambda_{\text{CLS}} > \lambda_{\text{EDF}}. \tag{33}$$

We next compare the λ's obtained by the ML and CLS methods. Equations (28) and (11), respectively, take the following forms in the DFT domain:

$$\sum_{i=1}^{M^2} \frac{\lambda_{\text{ML}} |q_i|^2 |G_i|^2}{(|h_i|^2 + \lambda_{\text{ML}} |q_i|^2)} = M^2 \sigma^2 \tag{34}$$

and

$$\sum_{i=1}^{M^2} \left[\frac{\lambda_{\text{CLS}} |q_i|^2 |G_i|^2}{(|h_i|^2 + \lambda_{\text{CLS}} |q_i|^2)^2} \right] = M^2 \sigma^2. \tag{35}$$

Now consider the left-hand sides of (34) and (35) as functions of λ. It is easy to verify that they are both increasing functions of λ. Furthermore, for the same value of λ the left-hand side of (35) is always smaller than the left-hand side of (34). Thus since the left-hand sides of (34) and (35) are equal, for $\lambda = \lambda_{\text{ML}}$ and $\lambda = \lambda_{\text{CLS}}$, respectively, it holds that

$$\lambda_{\text{CLS}} > \lambda_{\text{ML}}. \tag{36}$$

The comparison of λ_{MSE} and λ_{PMSE} can be based on the properties of the signal whose norm is minimized. In obtaining λ_{PMSE}, according to (17), the blurred error signal $\|He(\lambda)\|$ is used. If H is low pass, then greater importance is given at the low frequency error. Therefore, λ_{PMSE} will emphasize less high frequency fidelity than λ_{MSE}, resulting in

$$\lambda_{\text{MSE}} > \lambda_{\text{PMSE}}. \tag{37}$$

These observations agree with the observations of Rice [25]. In contrast, if H is high pass the opposite will apply. In most applications, however, blurs are low pass, thus (37) holds. Since λ_{CV} is asymptotically equal to λ_{PMSE} it is expected that, in general, $\lambda_{\text{MSE}} > \lambda_{\text{CV}}$.

The comparison of λ_{CV} and λ_{EDF} has been examined in [10], [31]. For high signal-to-noise ratios (SNR's) Thompson *et al.* [31] showed using a quadratic approximation that

$$\lambda_{\text{EDF}} \sim \lambda_{\text{CV}}. \tag{38}$$

It is easy to show by using (12), (23), (24), and (29) that when $H = Q = I$ and σ^2 is known:

$$\lambda_{\text{EDF}} = \lambda_{\text{PMSE}} = \lambda_{\text{MSE}} = \lambda_{\text{ML}} = \frac{1}{\text{SNR}}. \tag{39}$$

However, to our knowledge in the more general case no conclusions have been reached about the relation of λ_{CV} and λ_{ML}. The comparison of λ_{CV} and λ_{ML} does not always yield a clear-cut result. Our experiments indicate that their relation depends on the choice of Q (see next section for a more detailed discussion).

B. Special Case of $|q_i|^2 = |F_i|^{-2}$

A case of special interest is when the regularization operator is selected to satisfy $|q_i|^2 = |F_i|^{-2}$. In other words, if f is assumed to be stochastic this choice of Q implies that $Q^tQ = (\hat{R}_f)^{-1}$, where $\hat{R}_f$ is the periodogram-based estimate of the autocorrelation function of f [17].

In this case it is shown in Appendix C that

$$\lambda_{\text{MSE}} = \lambda_{\text{PMSE}} = \lambda_{\text{CV}} = \lambda_{\text{EDF}} = \lambda_{\text{ML}} = \lambda_{\text{ST}} = \sigma^2. \tag{40}$$

Thus for this choice of Q, λ_{CV} is *exactly equal* to λ_{PMSE} and not asymptotically equal as in the general case shown in [9] and also in [26]. It is shown in Appendix D that in this case

$$\lambda_{\text{CLS}} > \sigma^2. \tag{41}$$

The significance of this result is that if Qf is an IID signal *all methods* for choosing a regularization parameter, *except the CLS method*, yield the same value for the regularization parameter. Furthermore, for this case if $\hat{R}_f$ is replaced by $R_f = E[ff^t]$, the restored image $\hat{f}(\lambda)$ is equal to

$$\hat{f} = (H^tH + \sigma^2(R_f)^{-1})^{-1}H^tg \tag{42}$$

which is the *LMMSE* or the Wiener filter solution [1] to (1).

For this choice of Q it is further shown in Appendix E that $u = (I - A(\lambda))^{1/2}\, g$ is IID with variance σ^2. Thus the noise estimators of (31)–(33) are "identical," in the sense that they all express the relation

$$\|Bx\|^2 = \sigma_x^2 \text{ trace } (B^tB) \tag{43}$$

for x IID with variance σ_x^2 and any matrix B.

VII. Experimental Results

Experiments were performed to verify our theoretical results. In a first experiment the 128 by 128 "Lena" image was blurred by a 7 by 7 convolutional mask, and IID zero mean noise was added yielding SNR's of 10, 20, 30, and 40 dB. The weights of the convolutional mask used

TABLE I

(a) Results Obtained with the "Lena" 128 by 128 Image, a 7 by 7 Mask and SNR = 10 dB. The Values λ_{MSE1} and λ_{PMSE1} Were Computed Using the Original f Directly, the Values λ_{MSE2} and λ_{PMSE2} Using (23) and (24), Respectively, and the Values of $\hat{\sigma}^2_{EDF}$, $\hat{\sigma}^2_{CV}$, and $\hat{\sigma}^2_{ML}$ Using (30), (31) and (32), Respectively. (b) Results Obtained with the "Lena" 128 by 128 Image, a 7 by 7 Mask and SNR = 20 dB. The Values λ_{MSE1} and λ_{PMSE1} Were Computed Using the Original f Directly, the Values λ_{MSE2} and λ_{PMSE2} Using (23) and (24), Respectively, and the Values of $\hat{\sigma}^2_{EDF}$, $\hat{\sigma}^2_{CV}$, and $\hat{\sigma}^2_{ML}$ Using (30), (31) and (32), Respectively. (c) Results Obtained with the "Lena" 128 by 128 Image, a 7 by 7 Mask and SNR = 30 dB. The Values λ_{MSE1} and λ_{PMSE1} Were Computed Using the Original f Directly, the Values λ_{MSE2} and λ_{PMSE2} Using (23) and (24), Respectively, and the Values of $\hat{\sigma}^2_{EDF}$, $\hat{\sigma}^2_{CV}$, and $\hat{\sigma}^2_{ML}$ Using (30), (31) and (32), Respectively. (d) Results Obtained with the "Lena" 128 by 128 Image, a 7 by 7 Mask and SNR = 40 dB. The Values λ_{MSE1} and λ_{PMSE1} Were Computed Using the Original f Directly, the Values λ_{MSE2} and λ_{PMSE2} Using (23) and (24), Respectively, and the Values of $\hat{\sigma}^2_{EDF}$, $\hat{\sigma}^2_{CV}$, and $\hat{\sigma}^2_{ML}$ Using (30), (31) and (32), Respectively

	SNR = 10 dB		
	$Q = I$	$Q = Q_{2DL}$	$Q'Q = (\hat{R}_f)^{-1}$
λ_{CLS}	1.60E − 01	9.01E − 01	292.4
λ_{ST}	1.00E − 01	8.59E − 02	68.37
λ_{EDF}	5.67E − 02	1.26E − 01	67.63
λ_{MSE1}	9.83E − 02	1.30E − 01	59.75
λ_{MSE2}	1.05E − 01	1.25E − 01	88.01
λ_{PMSE1}	3.62E − 02	1.70E − 01	71.95
λ_{PMSE2}	3.68E − 02	1.60E − 01	71.46
λ_{CV}	3.54E − 02	1.57E − 01	71.35
λ_{ML}	1.12E − 02	2.87E − 01	67.64
σ^2	68.37	68.37	68.37
$\hat{\sigma}^2_{EDF}$	65.95	68.56	68.49
$\hat{\sigma}^2_{CV}$	65.85	69.02	68.21
$\hat{\sigma}^2_{ML}$	62.81	69.14	68.98

(a)

	SNR = 20 dB		
	$Q = I$	$Q = Q_{2DL}$	$Q'Q = (\hat{R}_f)^{-1}$
λ_{CLS}	3.00E − 02	4.76E − 02	28.55
λ_{ST}	1.00E − 02	8.59E − 03	6.837
λ_{EDF}	1.56E − 02	1.06E − 02	6.491
λ_{MSE1}	2.65E − 02	1.07E − 02	6.072
λ_{MSE2}	2.66E − 02	8.23E − 03	5.043
λ_{PMSE1}	1.12E − 02	1.09E − 02	6.452
λ_{PMSE2}	1.14E − 02	1.14E − 02	6.648
λ_{CV}	1.07E − 02	1.04E − 02	6.601
λ_{ML}	2.09E − 03	1.84E − 02	6.823
σ^2	6.837	6.837	6.837
$\hat{\sigma}^2_{EDF}$	6.556	6.845	6.839
$\hat{\sigma}^2_{CV}$	6.561	6.847	6.855
$\hat{\sigma}^2_{ML}$	5.954	7.014	6.864

(b)

	SNR = 30 dB		
	$Q = I$	$Q = Q_{2DL}$	$Q'Q = (\hat{R}_f)^{-1}$
λ_{CLS}	8.36E − 03	3.99E − 03	2.855
λ_{ST}	1.00E − 03	8.59E − 04	0.6837
λ_{EDF}	3.42E − 03	8.74E − 04	0.6920
λ_{MSE1}	5.08E − 03	1.02E − 03	0.6814
λ_{MSE2}	5.15E − 03	8.37E − 04	0.5252
λ_{PMSE1}	2.71E − 03	9.24E − 04	0.6753
λ_{PMSE2}	2.73E − 03	9.38E − 04	0.7093
λ_{CV}	2.54E − 03	9.25E − 04	0.6963
λ_{ML}	3.17E − 04	1.26E − 03	0.6687
σ^2	0.6837	0.6837	0.6837
$\hat{\sigma}^2_{EDF}$	0.6529	0.6867	0.6879
$\hat{\sigma}^2_{CV}$	0.6531	0.6896	0.6851
$\hat{\sigma}^2_{ML}$	0.5315	0.7055	0.6912

(c)

	SNR = 40 dB		
	$Q = I$	$Q = Q_{2DL}$	$Q'Q = (\hat{R}_f)^{-1}$
λ_{CLS}	1.90E − 03	4.38E − 04	0.3228
λ_{ST}	1.00E − 04	8.59E − 05	0.0683
λ_{EDF}	6.07E − 04	8.10E − 05	0.0665
λ_{MSE1}	8.02E − 04	9.09E − 05	0.0692
λ_{MSE2}	8.19E − 04	8.13E − 05	0.0694
λ_{PMSE1}	5.37E − 04	8.82E − 05	0.0693
λ_{PMSE2}	5.29E − 04	8.98E − 05	0.0680
λ_{CV}	4.86E − 04	9.24E − 05	0.0749
λ_{ML}	4.17E − 05	1.03E − 04	0.0690
σ^2	0.0683	0.0683	0.0683
$\hat{\sigma}^2_{EDF}$	0.0648	0.0695	0.0691
$\hat{\sigma}^2_{CV}$	0.0649	0.0696	0.0689
$\hat{\sigma}^2_{ML}$	0.0476	0.0699	0.0695

(d)

satisfy

$$h(i, j) = h(i, -j) = h(-i, j) = h(-i, j)$$

and

$$h(i, j) = h(j, i), \quad \text{for } i, j = -3, \cdots, 3.$$

The values in the upper-left quadrant of the mask (i.e., the lowest right element is $h(0, 0)$) are

$$\begin{matrix} 1.94E-02 & 1.98E-02 & 2.01E-02 & 2.03E-02 \\ 1.98E-02 & 2.04E-02 & 2.07E-02 & 2.08E-02 \\ 2.01E-02 & 2.07E-02 & 2.10E-02 & 2.11E-02 \\ 2.03E-02 & 2.08E-02 & 2.11E-02 & 2.12E-02. \end{matrix}$$

Three cases were examined using three different Q's as regularization operators. In the first case $Q = I$, was chosen where I is the identity matrix. In the second case $Q = Q_{2DL}$ was chosen, where Q_{2DL} is the 2-D Laplacian operator [12]. Finally, the case with $Q'Q = (\hat{R}_f)^{-1}$ was examined, where $\hat{R}_f$ represents the periodogram estimate of the autocorrelation of f. The values of the regularization parameter and the noise variance were found in all the above cases using all methods described in Sections III and IV, and they are tabulated in Table I(a)–(d). Plots of the MSE and PMSE as a function of λ are also shown in Figs. 2 and 3 for the SNR = 20-dB case.

From Table I(a)–(d), it is clear that the CLS method always yields the largest value for λ and thus an over regularized estimate. However, for $Q = I$ the difference between λ_{CLS} and $\lambda_{MSE1,2}$ is much smaller than for $Q = Q_{2DL}$ and $Q'Q = (\hat{R}_f)^{-1}$. Also, as expected, λ_{MSE} is in general larger than λ_{PMSE}. These values were computed using two methods. First, the original image f was used resulting in the λ_{MSE1} and λ_{PMSE1} values. Second, (23) and (24) were used with the knowledge of the noise var-

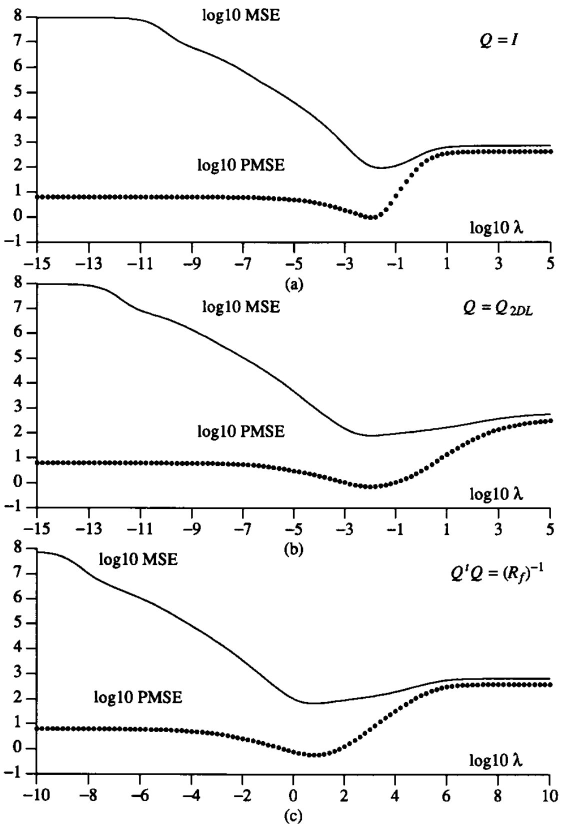


Fig. 2. PMSE (dotted line) and MSE (solid line) as function of λ for the 128 by 128 "Lena" image, with 7 by 7 blur, and SNR = 20 dB.

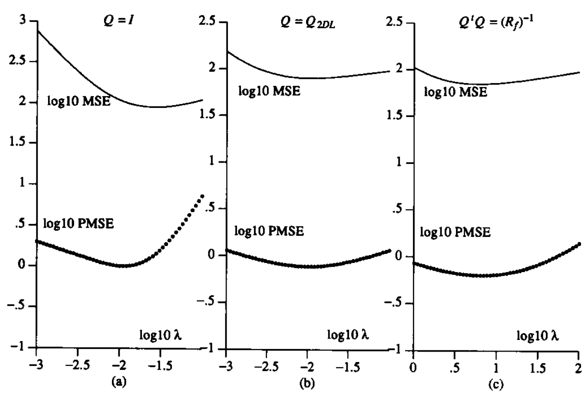


Fig. 3. Expanded plot of the area around the minima of Fig. 2(a)-(c).

iance σ^2, resulting in the λ_{MSE2} and λ_{PMSE2} values. By comparing the values obtained by both methods it was found that (23) and (25) yielded equally accurate estimates of λ_{MSE} and λ_{PMSE}, respectively. Also, as expected λ_{CV} was very close to λ_{PMSE}, thus verifying their asymptotic relation shown in [9], [26].

From Table I(a), (b), Figs. 2(a), (b), and 3(a), (b) it is easy to observe that for the "Lena" image with $Q = Q_{2DL}$ the values for λ obtained by the EDF, MSE, PMSE, CV, and ML methods are much closer to each other than with $Q = I$. This can be explained by the fact that the "Lena" image is highly correlated with low-pass characteristics, as is the case with most portrait or natural scenes images. Thus unlike f, $Q_{2DL}f$ is almost an IID signal, that is $Q_{2DL}^t Q_{2DL} \sim K(\hat{R}_f)^{-1}$, where K is a constant. Therefore, because of (40) the differences in the values of λ obtained by the EDF, MSE, PMSE, CV, and ML methods are small.

From the plots in Figs. 2(a)-(c), and 3(a)-(c) it is easy to notice that the curves of MSE and PMSE for $Q = Q_{2DL}$ and $Q^tQ = (\hat{R}_f)^{-1}$ "look alike." The MSE and PMSE curves have a minimum for almost the same values of λ and the curves around the minimum are flat. Thus in these cases the minimum error is very stable to changes in λ. In contrast, the curves for $Q = I$ show that values of λ_{MSE} and λ_{PMSE} are further apart and that the minimum error is more sensitive to changes in λ.

The ML method for $Q = I$ yields a value for λ which is significantly smaller than both the MSE- and PMSE-based methods. This can be explained by the fact that the ML method uses the assumption that $Q\hat{f}(\lambda_{ML})$ is IID. Then when $Q = I$, this yields that $\hat{f}(\lambda_{ML})$ is IID. For this to be true, the random noise must dominate $\hat{f}(\lambda_{ML})$ which requires an under regularized estimate. On the other hand, for $Q = Q_{2DL}$ the ML method yields values of λ larger than the MSE-, PMSE-, and CV-based methods. In other words, for high noise variance, $Q_{2DL}\hat{f}(\lambda)$ is a better approximation of an IID Gaussian, if $\hat{f}(\lambda)$ is a "relatively smooth" image which implies that $\lambda > \lambda_{MSE} \sim \lambda_{PMSE} \sim \lambda_{CV}$. However, as the noise power decreases this difference becomes smaller.

For $Q^tQ = (\hat{R}_f)^{-1}$, all experimental results shown in Table I(a)-(d) and in Figs. 2(c) and 3(c) agree with the theoretical results of Section VI-B. All methods (except CLS) yield values of λ that are almost equal to the noise variance.

Figs. 4 and 5 contain representative restored images corresponding to the experiments described by Table I(b) and (c), respectively. More specifically, for $Q = I$ the restored images with λ_{CLS}, λ_{CV}, and λ_{ML} are shown, while for $Q = Q_{2DL}$ the restored images with λ_{CLS} and λ_{MSE2} are shown. In both cases the restoration obtained with the use of the largest and smallest λ's were chosen to be shown. For $Q = I$, however, the range between λ_{CLS} and λ_{ML} was too large, so restorations with λ_{CV} are also shown.

From these images it is clear that λ_{CLS} oversmooths the restored images for all choices of Q, while λ_{ML} for $Q = I$ yields severe noise amplification. The effect of the type of the regularization operator can also be seen in the pattern of the noise in the restored images. $Q = I$ yields a grain-like noise pattern while Q = 2DL yields a block-like noise pattern. The size of the block of this noise pattern depends on the size of the region of support of Q.

The noise variance estimators were also tested using (30)-(32). For $Q = Q_{2DL}$ all methods produced extremely

Fig. 4. (a) Original Lena 128 by 128 image; (b) Distorted image; SNR = 20 dB and 7 by 7 blur. Restored images by (5): (c) $Q = I$, λ_{CLS}. (d) $Q = I$, λ_{CV}. (e) $Q = I$, λ_{ML}. (f) $Q = Q_{2DL}$, λ_{CLS}.

accurate estimates of σ^2. In contrast, for $Q = I$ the ML method yielded estimates significantly smaller than the actual values. This can be explained by the fact that in this case the assumption that $\hat{f}(\lambda)$ is IID Gaussian is embedded in this estimator. However, the original image is smooth; therefore, the estimator, in an effort to satisfy the IID assumption, attributes more variation from the observed data to f and less to the noise process.

In a second experiment a different image was used to test the validity of the above observations. A 128 by 128 version of the "Wedding" image from Kodak was utilized. The observed image data were obtained as in the first experiment (7 by 7 convolutional mask, IID zero mean noise of SNR's 10, 20, 30, and 40 dB). The three regularization operators used in the first experiment (I, Q_{2DL} and $(\hat{R}_f)^{-1/2}$) were also used and all methods for choosing λ and estimating the noise variance were tested. For this image, exactly the same observations as in the first experiment can be made, and the same conclusions were reached.

A third experiment was also performed. The clear objective of this experiment was to test the various methods

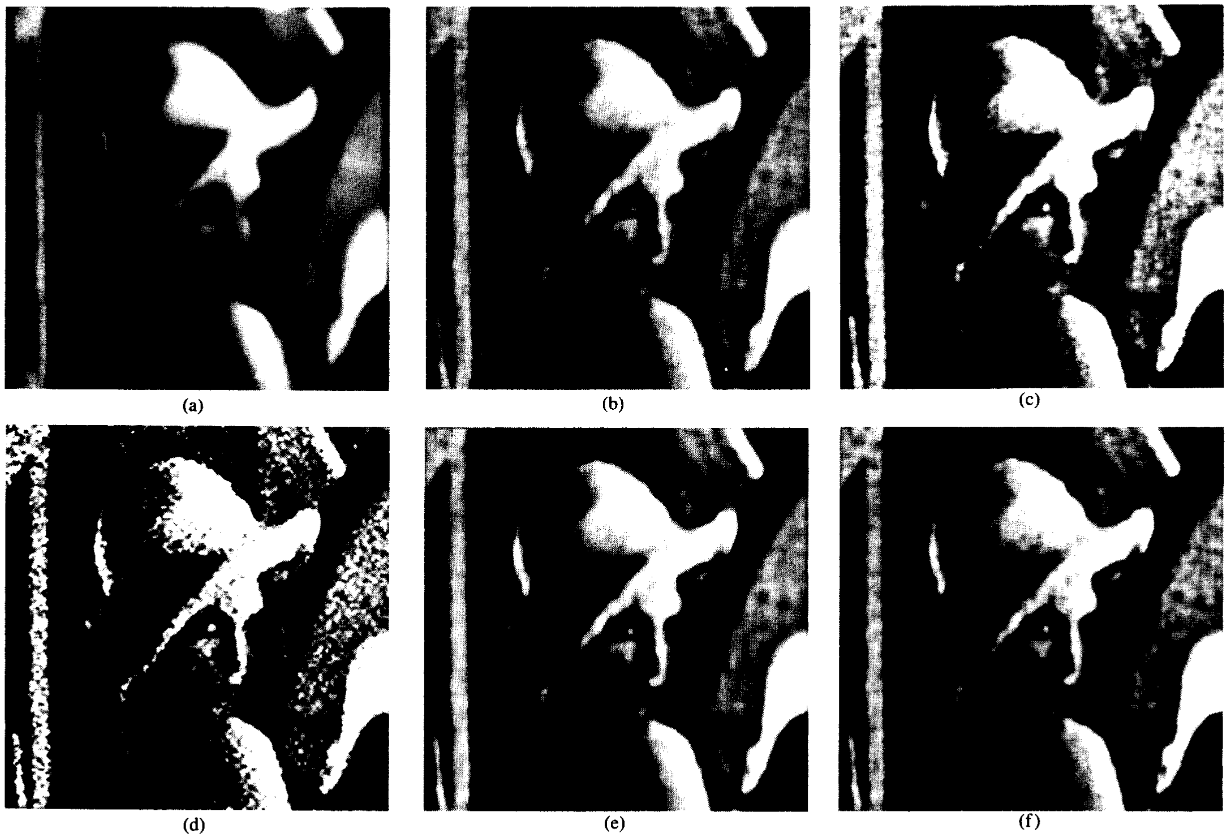


Fig. 5. (a) Distorted image; SNR = 30 dB and 7 by 7 blur. Restored images by (5): (b) $Q = I$, λ_{CLS}. (c) $Q = I$, λ_{CV}. (e) $Q = I$, λ_{ML}. (d) $Q = Q_{2DL}$, λ_{CLS}. (e) $Q = Q_{2DL}$, λ_{MSE2}.

for an image with different characteristics than the ones of natural scenes. This time a 128 by 128 synthetic grating image with vertical alternating black and white stripes of 8 pixels width was used. This image was chosen so as not to exhibit the correlation structure that natural image scenes have. In other words, an image which when operated on by the 2-D Laplacian operator does not produce an almost IID signal. The observed image data were generated as in experiments one and two. The resulting values of λ and $\hat{\sigma}^2$ were obtained by all methods as in the first and second experiments and are tabulated in Table II(a)–(d). MSE and PMSE plots similar to the ones in the first experiment are provided in Figs. 6 and 7 for SNR = 20 dB.

From Table II(a)–(d) and Figs. 6 and 7 it is observed that the CLS still yields the largest value of λ. Also, as in the previous two experiments λ_{MSE} is greater than λ_{MSE}. However, unlike the previous experiments, this time the discrepancy between λ_{MSE} and $\lambda_{PMSE/CV}$ is also significant for $Q = Q_{2DL}$. This is expected, since as also explained earlier, unlike the "Lena" and the "Wedding" images the grating image when operated by the 2-D Laplacian does not yield an almost IID signal. Therefore, for this image the case with $Q = Q_{2DL}$ does not resemble the case with $Q^tQ = (\hat{R}_f)^{-1}$ and thus does not yield approximately equal values of λ for the EDF, MSE, PMSE, CV, and ML methods.

The ML method produced similar results as in the previous two experiments (severe under regularization for $Q = I$ and over regularization which gradually decreases with the noise power for $Q = Q_{2DL}$). Thus the ML method seems to be more stable to the change in the image content than the MSE, PMSE, and CV methods.

The noise estimators in this experiment performed similarly to the ones in the previous experiments. Therefore, a MSE solution can still be obtained even when the noise variance is unknown.

VIII. Discussion and Conclusions

The assumptions used for the analysis in this paper were that the degradation and the regularization operators are circulant. Also, throughout this paper, with the exception of Section IV-B, it was assumed that only the noise n is a random signal. The latter assumption, unlike the former one, is not critical to the derivation of our results. That is, the original image f can also be assumed to be random, provided that it is independent of the noise n and its au-

TABLE II

(a) RESULTS OBTAINED WITH THE "GRATING" 128 BY 128 IMAGE, A 7 BY 7 MASK AND SNR = 10 dB. THE VALUES λ_{MSE1} AND λ_{PMSE1} WERE COMPUTED USING THE ORIGINAL f DIRECTLY, THE VALUES λ_{MSE2} AND λ_{PMSE2} USING (23) AND (24), RESPECTIVELY, AND THE VALUES OF $\hat{\sigma}^2_{EDF}$, $\hat{\sigma}^2_{CV}$, AND $\hat{\sigma}^2_{ML}$ USING (30), (31) AND (32), RESPECTIVELY. (b) RESULTS OBTAINED WITH THE "GRATING" 128 BY 128 IMAGE, A 7 BY 7 MASK AND SNR = 20 dB. THE VALUES λ_{MSE1} AND λ_{PMSE1} WERE COMPUTED USING THE ORIGINAL f DIRECTLY, THE VALUES λ_{MSE2} AND λ_{PMSE2} USING (23) AND (24), RESPECTIVELY, AND THE VALUES OF $\hat{\sigma}^2_{EDF}$, $\hat{\sigma}^2_{CV}$, AND $\hat{\sigma}^2_{ML}$ USING (30), (31) AND (32), RESPECTIVELY. (c) RESULTS OBTAINED WITH THE "GRATING" 128 BY 128 IMAGE, A 7 BY 7 MASK AND SNR = 30 dB. THE VALUES λ_{MSE1} AND λ_{PMSE1} WERE COMPUTED USING THE ORIGINAL f DIRECTLY, THE VALUES λ_{MSE2} AND λ_{PMSE2} USING (23) AND (24), RESPECTIVELY, AND THE VALUES OF $\hat{\sigma}^2_{EDF}$, $\hat{\sigma}^2_{CV}$ AND $\hat{\sigma}^2_{ML}$ USING (30), (31) AND (32), RESPECTIVELY. (d) RESULTS OBTAINED WITH THE "GRATING" 128 BY 128 IMAGE, A 7 BY 7 MASK AND SNR = 40 dB. THE VALUES λ_{MSE1} AND λ_{PMSE1} WERE COMPUTED USING THE ORIGINAL f DIRECTLY, THE VALUES λ_{MSE2} AND λ_{PMSE2} USING (23) AND (24), RESPECTIVELY, AND THE VALUES OF $\hat{\sigma}^2_{EDF}$, $\hat{\sigma}^2_{CV}$, AND $\hat{\sigma}^2_{ML}$ USING (30), (31) AND (32), RESPECTIVELY

	SNR = 10 dB	
	$Q = I$	$Q = Q_{2DL}$
λ_{CLS}	1.47E − 01	6.06E − 01
λ_{ST}	1.00E − 01	9.89E − 02
λ_{EDF}	4.90E − 02	4.54E − 02
λ_{MSE1}	8.33E − 02	6.73E − 02
λ_{MSE2}	8.38E − 02	5.34E − 02
λ_{PMSE1}	3.15E − 02	3.93E − 02
λ_{PMSE2}	3.18E − 02	3.61E − 02
λ_{CV}	3.01E − 02	3.54E − 02
λ_{ML}	1.09E − 02	1.83E − 01
σ^2	1621.1	1621.1
$\hat{\sigma}^2_{EDF}$	1560.8	1613.9
$\hat{\sigma}^2_{CV}$	1559.1	1611.4
$\hat{\sigma}^2_{ML}$	1479.6	1664.1

(a)

	SNR = 20 dB	
	$Q = I$	$Q = Q_{2DL}$
λ_{CLS}	2.60E − 02	2.09E − 02
λ_{ST}	1.00E − 02	9.89E − 03
λ_{EDF}	1.34E − 02	5.80E − 02
λ_{MSE1}	2.30E − 02	1.06E − 02
λ_{MSE2}	2.26E − 02	8.99E − 03
λ_{PMSE1}	9.64E − 03	5.97E − 03
λ_{PMSE2}	9.72E − 03	5.79E − 03
λ_{CV}	9.09E − 03	5.79E − 03
λ_{ML}	2.02E − 03	6.94E − 03
σ^2	162.1	162.1
$\hat{\sigma}^2_{EDF}$	153.3	161.9
$\hat{\sigma}^2_{CV}$	154.4	162.9
$\hat{\sigma}^2_{ML}$	140.1	162.4

(b)

	SNR = 30 dB	
	$Q = I$	$Q = Q_{2DL}$
λ_{CLS}	7.52E − 02	2.38E − 03
λ_{ST}	1.00E − 03	9.89E − 04
λ_{EDF}	3.32E − 03	5.56E − 04
λ_{MSE1}	5.55E − 03	1.01E − 03
λ_{MSE2}	5.28E − 03	8.82E − 04
λ_{PMSE1}	2.50E − 03	5.06E − 04
λ_{PMSE2}	2.51E − 03	5.34E − 04
λ_{CV}	2.28E − 03	5.30E − 04
λ_{ML}	3.06E − 04	6.17E − 04
σ^2	16.21	16.21
$\hat{\sigma}^2_{EDF}$	15.18	16.16
$\hat{\sigma}^2_{CV}$	15.21	16.13
$\hat{\sigma}^2_{ML}$	12.45	16.32

(c)

	SNR = 40 dB	
	$Q = I$	$Q = Q_{2DL}$
λ_{CLS}	2.01E − 03	3.10E − 04
λ_{ST}	1.00E − 04	9.89E − 05
λ_{EDF}	7.30E − 04	5.84E − 05
λ_{MSE1}	1.10E − 03	8.09E − 05
λ_{MSE2}	1.03E − 04	7.31E − 05
λ_{PMSE1}	5.72E − 04	5.24E − 05
λ_{PMSE2}	5.69E − 04	5.17E − 05
λ_{CV}	4.87E − 04	4.78E − 05
λ_{ML}	4.17E − 05	5.92E − 05
σ^2	1.621	1.621
$\hat{\sigma}^2_{EDF}$	1.461	1.571
$\hat{\sigma}^2_{CV}$	1.465	1.570
$\hat{\sigma}^2_{ML}$	1.095	1.619

(d)

tocorrelation function R_f is circulant. Then, all equations in this paper hold true if expectations are taken with respect to both f and n and F_i is replaced with r_i, where r_i is the ith eigenvalue of R_f. In this case the regularization operator in Section VI-B becomes $Q^tQ = (R_f)^{-1}$. Thus the periodogram estimate of f is not necessary to make the connection between the regularized and the LMMSE estimates.

From the analysis in Section VI and the experiments in Section VII the following conclusions have been drawn.

1) The CLS method, which has been widely used by the image restoration community always yields oversmoothed estimates.

2) The value of λ that minimizes the MSE criterion is larger than the one that minimizes the PMSE criterion.

3) For images that are highly correlated and exhibit low-pass characteristics the commonly used 2-D Laplacian represents a good choice of the regularization operator. In this case λ_{CV} yields a MSE very close to its minimum value not only with $Q = Q_{2DL}$ but with $Q = (R_f)^{-1/2}$ as well. However, for images that do not exhibit this behavior the 2-D Laplacian along with λ_{CV} yields an under-regularized estimate, with corresponding MSE larger than its minimum value. In this case it is preferable to use directly λ_{MSE} after the noise variance is estimated by the EDF or the CV estimators, instead of λ_{CV}.

4) The ML method is very sensitive to the choice of the regularization operator, but is less sensitive (in comparison to the MSE and PMSE/CV methods) to the characteristics of the image.

5) The new variance estimator $\hat{\sigma}^2_{CV}$ is at least as good as the $\hat{\sigma}^2_{EDF}$ estimator and its use is theoretically justified. The proposed $\hat{\sigma}^2_{ML}$ variance estimator is very sensitive to the regularization operator and in general does not yield as good an estimate as $\hat{\sigma}^2_{CV}$.

Some of the previous conclusions were also verified in

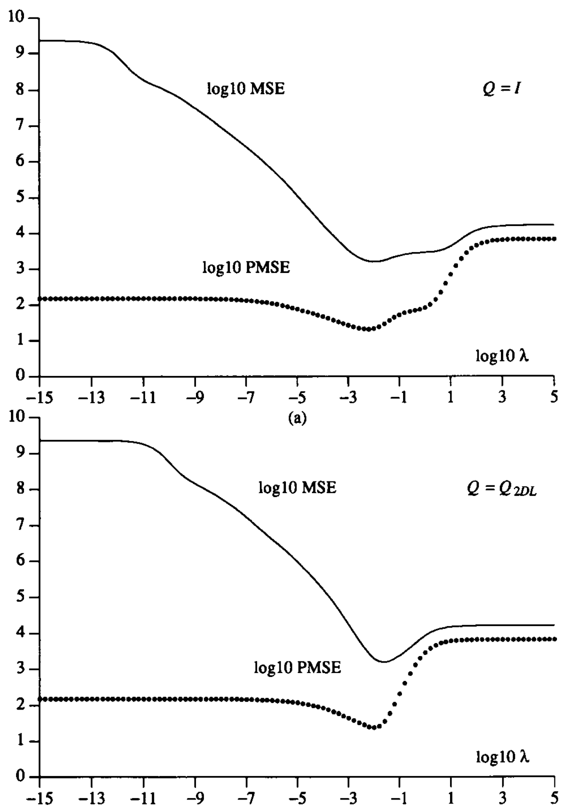


Fig. 6. PMSE (dotted line) and MSE (solid line) as function of λ for the 128 by 128 "grating" image, with 7 by 7 blur, and SNR = 20 dB.

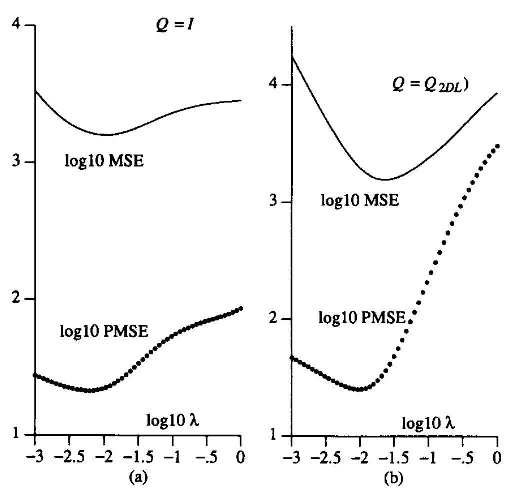


Fig. 7. An expanded plot of the area around the minima of Fig. 6(a) and (b) is shown.

[26] by examining asymptotic properties of the regularization parameter for certain forms of the eigenstructure of H, Q, and the covariance of f. The challenging problem that was not addressed in this paper is the extension of the regularization problem to the nonstationary image/space-varying degradation case. Then noncirculant operators are involved and the analysis cannot be carried out in the frequency domain.

Appendix A

In this Appendix the equivalence of minimizing $E[\|He(\lambda)\|^2]$ and satisfying (24) is shown. Equation (24) with $g = Hf + n$ yields

$$\begin{aligned} &E\,\|(HA(\lambda))^{1/2}(HA(\lambda) - I)g\|^2 - \sigma^2 \\ &\quad\cdot \text{trace}\,[HA(\lambda)(I - HA(\lambda))] \\ &\quad= \text{trace}\,[(HA(\lambda))^{1/2}(I - HA(\lambda))H[ff^t] \\ &\quad\quad H^t(I - HA(\lambda)^t)(HA(\lambda)^{1/2})^t \\ &\quad\quad + (HA(\lambda))^{1/2}(I - HA(\lambda)) \\ &\quad\cdot R_n(I - HA(\lambda)^t)((HA(\lambda))^{1/2})^t \\ &\quad\quad - \sigma^2 HA(\lambda)(I - HA(\lambda))] = 0. \end{aligned} \tag{A-1}$$

In the Fourier domain, (A-1) yields

$$\sum_{i=1}^{M^2}\left[\frac{\lambda^2|q_i|^4|h_i|^4|F_i|^2}{(|h_i|^2+\lambda|q_i|^2)^3} + \frac{\sigma^2\lambda^2|q_i|^4|h_i|^2}{(|h_i|^2+\lambda|q_i|^2)^3} - \frac{\sigma^2\lambda|q_i|^2|h_i|^2}{(|h_i|^2+\lambda|q_i|^2)^2}\right] = 0 \tag{A-2}$$

which yields

$$\sum_{i=1}^{M^2}\frac{|h_i|^4|q_i|^2(\lambda|q_i|^2|F_i|^2 - \sigma^2)}{(|h_i|^2+\lambda|q_i|^2)^3} = 0. \tag{A-3}$$

We now show that the value of λ minimizing $E[\|He(\lambda)\|^2]$ satisfies (A-3). For uncorrelated signal and noise we obtain

$$\begin{aligned} E(\|Hf - H\hat{f}\|^2) &= E(\|(I - HA(\lambda))Hf - HA(\lambda)n\|^2) \\ &= E(\|(I - HA(\lambda))Hf\|^2) \\ &\quad + E(\|HA(\lambda)n\|^2) \\ &= \text{trace}\,[(I - HA(\lambda))H[ff^t] \\ &\quad\cdot H^t(I - HA(\lambda))^t) + HA(\lambda)R_n(HA(\lambda))^t]. \end{aligned} \tag{A-4}$$

In the Fourier (A-4) becomes

$$E(\|Hf - H\hat{f}\|^2) = \sum_{i=1}^{M^2}\frac{\lambda^2|q_i|^4|h_i|^2|F_i|^2 + \sigma^2|h_i|^4}{(|h_i|^2+\lambda|q_i|^2)^2}. \tag{A-5}$$

Taking the derivative with respect to λ and setting it equal to zero yields

$$\frac{\partial}{\partial\lambda}(E(\|Hf - H\hat{f}\|^2)) = \sum_{i=1}^{M^2}\frac{|h_i|^4|q_i|^2(\lambda|q_i|^2|F_i|^2 - \sigma^2)}{(|h_i|^2+\lambda|q_i|^2)^3} = 0$$

which is identical to (A-3).

Appendix B

In this Appendix it is shown that maximizing $L(g|\lambda, \sigma^2)$ in (27) yields (29) and is equivalent to minimizing (28). The integrand in (27) is equal to

$$p(g \mid f, \sigma^2) \cdot p(f \mid \lambda, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{M^2} \cdot \det [\lambda Q^t Q]^{1/2} \cdot \exp\left[\left(\frac{-1}{2\sigma^2}\right)(f - \hat{f})^t (H^t H + \lambda Q^t Q)(f - \hat{f})\right] \cdot \exp\left[\left(\frac{-1}{2\sigma^2}\right) S(\lambda)\right] \quad \text{(B-1)}$$

with

$$\hat{f} = (H^t H + \lambda Q^t Q)^{-1} H^t g$$

$$A(\lambda) = (H^t H + \lambda Q^t Q)^{-1} H^t$$

and

$$S(\lambda) = g^t[I - HA(\lambda)]g = \|(I - HA(\lambda))^{1/2} g\|^2.$$

Integrating the multidimensional Gaussian f in (B-1) yields

$$L(g \mid \lambda, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{M^2/2} \det [\lambda Q^t Q]^{1/2} \cdot \det [H^t H + \lambda Q^t Q]^{-1/2} \cdot \exp\left[\frac{-S(\lambda)}{2\sigma^2}\right]. \quad \text{(B-2)}$$

Taking the logarithm of both sides of (B-2) yields

$$\log (L(g \mid \lambda, \sigma^2)) = \left(\frac{-M^2}{2}\right) \log (2\pi\sigma^2) + \left(\frac{1}{2}\right) \log (\det [\lambda Q^t Q]) - \left(\frac{1}{2}\right) \log (\det [H^t H + \lambda Q^t Q]) - \left(\frac{1}{2\sigma^2}\right) \cdot S(\lambda). \quad \text{(B-3)}$$

Taking the derivative of (B-3) and setting it equal to zero yields

$$\sigma^2 = \left(\frac{1}{M^2}\right) \cdot S(\lambda) = \left(\frac{1}{M^2}\right) \|(I - HA(\lambda))^{1/2} g\|^2. \quad \text{(B-4)}$$

Substituting (B-4) back into (B-3) yields

$$ML(\lambda) = \frac{g^t(I - HA(\lambda))g}{(\det [I - HA(\lambda)])^{1/M^2}} \quad \text{(B-5)}$$

which verifies the validity of the desired equations.

Appendix C

Throughout this Appendix we assume that $|E_i|^2 = (|q_i|^2)^{-1}$, where q_i are the eigenvalues of Q. This is equivalent to assuming that $Q^t Q = (R_f)^{-1}$ with $E[ff^t] = R_f$ if the periodogram estimate is used as the estimate of the power spectrum of f [17].

Proof of $\lambda_{ST} = \sigma^2$*:* For this choice of the regularization operator, the parameter E^2 in (13) for the set theoretic approach is equal to

$$E^2 = \|([ff^t])^{-1/2} f\|^2 = \sum_{i=1}^{M^2} (|F_i|^2)^{-1} |F_i|^2 = M^2$$

where F is the DFT of f. Thus from (13) and (15)

$$\lambda_{\text{ST}} = \sigma^2.$$

Proof of $\lambda_{\text{MSE}} = \lambda_{\text{PMSE}} = \lambda_{\text{EDF}} = \sigma^2$*:* Equation (1) in the DFT domain becomes

$$E[|G_i|^2] = |G_i|^2 = |h_i|^2 |F_i|^2 + \sigma^2 = |h_i|^2 |q_i|^{-2} + \sigma^2. \quad \text{(C-1)}$$

Substituting (C-1) into (23) results in

$$\sum_{i=1}^{M^2} \frac{\lambda |q_i|^2 |h_i|^2 (\lambda - \sigma^2)}{(|h_i|^2 + \lambda |q_i|^2)^3} = 0$$

which is satisfied for $\lambda_{\text{MSE}} = \sigma^2$ if $\lambda_{\text{MSE}} \neq 0$.

Substituting (C-1) into (24) results in

$$\sum_{i=1}^{M^2} \frac{\lambda |h_i|^4 |q_i|^2 (\lambda - \sigma^2)}{(|h_i|^2 + \lambda |q_i|^2)^3} = 0$$

which also yields $\lambda_{\text{PMSE}} = \sigma^2$ if $\lambda_{\text{PMSE}} \neq 0$.

Similarly, (12) in the DFT domain is equal to

$$\sum_{i=1}^{M^2} \frac{\lambda^2 |q_i|^4 |G_i|^2}{(|h_i|^2 + \lambda |q_i|^2)^2} - \sum_{i=1}^{M^2} \frac{\sigma^2 \lambda |q_i|^2}{(|h_i|^2 + \lambda |q_i|^2)} = 0.$$

Due to (C-1) it becomes

$$\sum_{i=1}^{M^2} \frac{\lambda |q_i|^2 |h_i|^2 (\lambda - \sigma^2)}{(|h_i|^2 + \lambda |q_i|^2)^2} = 0$$

which yields $\lambda_{\text{EDF}} = \sigma^2$.

Proof of $\lambda_{\text{CV}} = \sigma^2$*:* Taking the expectation of the CV function in the DFT domain and substituting $|q_i|^2 = (|F_i|^2)^{-1}$ yields

$$E[\text{CV}(\lambda)] = \left[\sum_{i=1}^{M^2} \frac{\lambda^2 |q_i|^2 |h_i|^2 + \sigma^2 \lambda^2 |q_i|^4}{(|h_i|^2 + \lambda |q_i|^2)^2}\right] \cdot \left[\sum_{i=1}^{M^2} \frac{\lambda |q_i|^2}{(|h_i|^2 + \lambda |q_i|^2)}\right]^{-2}. \quad \text{(C-2)}$$

The derivative of (C-2) with respect to λ after some algebra yields

$$2\left\{\left[\sum_{i=1}^{M^2}\frac{\lambda|q_i|^2(|h_i|^2+\sigma^2|q_i|^2)(|h_i|^2+\lambda|q_i|^2)-\lambda^2|q_i|^4(|h_i|^2+\sigma^2|q_i|^2)}{(|h_i^2+\lambda|q_i|^2)^3}\right]\left[\sum_{i=1}^{M^2}\frac{\lambda|q_i|^2}{(|h_i|^2+\lambda|q_i|^2)}\right]^2\right.$$
$$\left.-\left[\sum_{i=1}^{M^2}\frac{\lambda^2|q_i|^2(|h_i|^2+\sigma^2|q_i|^2)}{(|h_i|^2+\lambda|q_i|^2)^2}\right]\left[\sum_{i=1}^{M^2}\frac{|q_i|^2|h_i|^2}{(|h_i|^2+\lambda|q_i|^2)}\right]^2\left[\sum_{i=1}^{M^2}\frac{\lambda|q_i|^2}{(|h_i|^2+\lambda|q_i|^2)}\right]\right\}$$
$$\cdot\left[\sum_{i=1}^{M^2}\frac{\lambda|q_i|^2}{(|h_i|^2+\lambda|q_i|^2)}\right]^{-4}. \tag{C-3}$$

Setting $\lambda = \sigma^2$, the numerator of (C-3) yields

$$\left\{\left[\sum_{i=1}^{M^2}\frac{\sigma^2|q_i|^2|h_i|^2}{(|h_i|^2+\sigma^2|q_i|^2)^2}\right]\left[\sum_{i=1}^{M^2}\frac{\sigma^2|q_i|^2}{(|h_i|^2+\sigma^2|q_i|^2)}\right]^2\right.$$
$$-\left[\sum_{i=1}^{M^2}\frac{\sigma^2|q_i|^2|h_i|^2}{(|h_i|^2+\sigma^2|q_i|^2)^2}\right]$$
$$\left.\cdot\left[\sum_{i=1}^{M^2}\frac{\sigma^2|q_i|^2}{(|h_i|^2+\sigma^2|q_i|^2)}\right]^2\right\}=0$$

which completes our proof.

Proof of $\lambda_{\text{ML}} = \sigma^2$*:* Equation (29) is written in the DFT domain as

$$\sum_{i=1}^{M^2}\frac{\lambda|q_i|^2(|h_i|^2|F_i|^2+\sigma^2)}{(|h_i|^2+\lambda|q_i|^2)}=M^2\sigma^2. \tag{C-4}$$

For $|q_i|^2 = (|F_i|^2)^{-1}$ it becomes

$$\sum_{i=1}^{M^2}\frac{|h_i|^2(\lambda-\sigma^2)}{(|h_i|^2+\lambda|q_i|^2)}=0$$

which results in $\lambda_{\text{ML}} = \sigma^2$.

Appendix D

In this Appendix we show that for $|q_i|^2 = (|F_i|^2)^{-1}$ the value of λ obtained by the CLS method satisfies

$$\sigma^2 < \lambda_{\text{CLS}} \tag{D-1}$$

for $\lambda = \lambda_{\text{CLS}}$. Equation (11) yields

$$E\|g - H\hat{f}(\lambda)\|^2 - M^2\sigma^2 = 0 \tag{D-2}$$

or

$$\text{trace}\,[(I - HA(\lambda))H[f^tf]H^t(I - HA(\lambda)^t)$$
$$+ (I - HA(\lambda))R_n(I - HA(\lambda)^t) - I] = 0$$

or in the DFT domain

$$\sum_{i=1}^{M^2}\left[\frac{\lambda^2|q_i|^4|h_i|^2|F_i|^2}{(|h_i|^2+\lambda|q_i|^2)^2}+\frac{\sigma^2\lambda^2|q_i|^4}{(|h_i|^2+\lambda|q_i|^2)^3}\right.$$
$$\left.-\frac{\sigma^2(|h_i|^2+\lambda|q_i|^2)^2}{(|h_i|^2+\lambda|q_i|^2)^2}\right]=0. \tag{D-2}$$

With $|F_i|^2 = (|q_i|^2)^{-1}$, (D-2) yields

$$\sum_{i=1}^{M^2}\frac{|h_i|^2\lambda^2|q_i|^2}{(|h_i|^2+\lambda|q_i|^2)^2}$$
$$+\frac{\sigma^2\lambda^2|q_i|^4}{(|h_i|^2+\lambda|q_i|^2)^2}-M^2\sigma^2=0. \tag{D-3}$$

This equation is satisfied for $\lambda = \lambda_{\text{CLS}}$. If instead $\lambda = \sigma^2$ is used, the left-hand side of (D-3) yields

$$\sigma^2\left[\sum_{i=1}^{M^2}\left(\frac{\sigma^2|q_i|^2}{|h_i|^2+\sigma^2|q_i|^2}\right)-M^2\right]<0. \tag{D-4}$$

Consider now the function $\|g - H\hat{f}(\lambda)\|^2$. It has been shown in the beginning of Section IV that this is an increasing function of λ. Thus (D-4) yields that (D-1) must hold.

Appendix E

In this Appendix we show that for $(Q^tQ)^{-1} = ff^t$, $\lambda = \sigma^2$, where f is deterministic, and $u = (I - A(\lambda))^{1/2}g$ is IID with variance σ^2. Using the matrix inversion lemma [17] for $(I - A(\lambda))$ we obtain

$$(I - A(\lambda)) =$$
$$[I - H(H^tH + \sigma^2([ff^t])^{-1})^{-1}H^t] = \left(I + \frac{H[ff^t]H^t}{\sigma^2}\right)^{-1}. \tag{E-1}$$

By definition

$$E[uu^t] = E[(I - A(\lambda))^{1/2}gg^t((I - A(\lambda))^{1/2})^t]. \tag{E-2}$$

Since $(I - A(\lambda))$ and gg^t are symmetric, (E-1) can be written as

$$E[uu^t] = (I - A(\lambda))R_g \tag{E-3}$$

where

$$E[g^tg] = R_g = H[ff^t]H^t + \sigma^2 I. \tag{E-4}$$

Using (E-1)–(E-4) yields

$$E[uu^t] = \sigma^2. \tag{E-5}$$

Acknowledgement

N. P. Galatsanos would like to acknowledge Dr. H. Stark for the numerous conversations and the continuous encouragement during the course of writing this paper. Both authors would like to acknowledge Dr. T. Olsen of the Mathematics Department, Illinois Institute of Technology for pointing out [20] and Dr. V. Solo for providing them with [26].

References

[1] H. Andrews and B. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice Hall, 1977.

[2] D. L. Angwin and H. Kaufman, "Non-homogeneous image identification and restoration procedures," in *Digital Image Restoration*, A. K. Katsaggelos, ed., Springer Verlag New York: vol. 23, 1991.

[3] P. Craven and G. Wahba, "Smoothing noisy data with spline functions," *Numer. Math.*, vol. 31, pp. 377-403, 1975.

[4] G. Demoment, "Image reconstruction and restoration: Overview of common estimation problems," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 2024-2036, Dec. 1989.

[5] J. A. Fessler, "Nonparametric fixed-interval smoothing with vector splines," *IEEE Trans. Signal Processing*, vol. 39, pp. 852-859, Apr. 1991.

[6] N. P. Galatsanos, and A. K. Katsaggelos "Cross validation and other criteria for estimating the regularization parameter and the noise variance," in *Proc. IEEE ICASSP*, Toronto, Canada, May 1991, pp. 3021-3024.

[7] ——, "The use of cross-validation and maximum-likelihood in regularized image restoration," in *Proc. Conf. Information Sciences and Systems*, Mar. 20-22, 1991, pp. 459-464.

[8] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 725-730, Nov. 1972.

[9] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometr.*, vol. 21, no. 2, pp. 215-223, 1979.

[10] P. Hall and D. Titterington, "Common structure of techniques for choosing smoothing parameters in regression problems," *J. Roy. Stat. Soc. B*, vol. 49, no. 2, pp. 184-198, 1987.

[11] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometr.*, vol. 12, no. 1, pp. 55-67, 1970.

[12] B. R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Trans. Comput.*, vol. 22, pp. 805-812, Sept. 1973.

[13] M. G. Kang and A. K. Katsaggelos, "Simultaneous iterative image restoration and evaluation of the regularization parameter," to be published.

[14] A. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, "A general formulation of constrained iterative image restoration," in *Proc. ICASSP-85*, Mar. 1985, pp. 700-703.

[15] A. K. Katsaggelos, "Iterative image restoration algorithms," *Opt. Eng.*, vol. 28, no. 7, pp. 735-748, July 1989.

[16] A. K. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, "A regularized iterative image restoration algorithm," *IEEE Trans. Signal Processing*, vol. 39, pp. 914-929, Apr. 1991.

[17] S. M. Kay, *Modern Spectral Estimation: Theory and Applications*, Englewood Cliff, NJ: Prentice Hall, 1988.

[18] R. L. Lagendijk, J. Biemond, and D. E. Bokee, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 38, pp. 1180-1191, July 1990.

[19] K. T. Lay and A. K. Katsaggelos, "Image identification and restoration based on the expectation maximization algorithm," *Opt. Eng.*, vol. 29, no. 5, pp. 436-445, May 1990.

[20] G. Kitagawa and W. Gersch, "A smoothness priors long AR model method for spectral estimation," *IEEE Trans. Autom. Contr.*, vol. AC-30, pp. 57-65, Jan. 1985.

[21] P. Meer, J. Jolion, and A. Rosenfeld, "A fast parallel algorithm for blind estimation of noise variance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, pp. 216-222, Feb. 1990.

[22] D. Nychka, "Confidence intervals for smoothing splines," *J. Amer. Stat. Assoc.*, vol. 83, pp. 1134-1143, 1988.

[23] S. J. Reeves and R. M. Mersereau, "Optimal estimation of the regularization parameters and stabilizing functional for regularized image restoration," *Opt. Eng.*, vol. 29, pp. 446-454, May 1990.

[24] S. J. Reeves, "A cross-validation approach to image restoration and blur identification," Ph.D. dissertation, Georgia Inst. Technol., 1990.

[25] J. A. Rice, "Choice of smoothing parameter in deconvolution problems," in *Contemporary Mathematics*, vol. 59, J. S. Maron, ed. Providence, RI: American Math Soc., 1986.

[26] V. Solo, "Linear image restoration: An analytical study," *Tech. Rep.* 88-VS-May-1988, Dept. of ECE, The John Hopkins Univ., Baltimore, MD, May 1988.

[27] M. Stone, "Cross-validatory choice and assesment of statistical predictions," *J. Roy. Stattist. Soc.*, Series B, vol. 36, pp. 111-147, 1974.

[28] A. M. Tekalp, H. Kaufman, and J. W. Woods, "Identification of image and blur parameters for the restoration of noncausal blurs," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 34, pp. 963-972, Aug. 1986.

[29] A. M. Tekalp and H. Kaufman, "On statistical identification of a class of linear phase space-invariant blurs using non-minimum phase ARMA models," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 36, pp. 1360-1363, Aug. 1988.

[30] A. M. Thompson, J. W. Kay, and D. M. Titterington, "A cautionary note about the crossvalidatory choice," *J. Statist. Comput. Simul.*, vol. 33, pp. 199-216, 1989.

[31] 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. Pattern Analy. Mach. Intell.*, vol. 13, no. 4, pp. 326-339, Apr. 1991.

[32] A. Tikhonov and V. Arsenin, *Solution of Ill-Posed Problems*. New York: Wiley, 1977.

[33] H. J. Trussel, "Convergence criteria for iterative restoration methods," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-31, pp. 129-136, Feb. 1983.

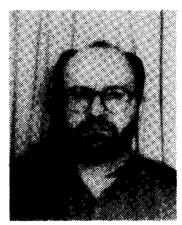
[34] H. J. Trussel and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 32, pp. 201-212, Apr. 1984.

[35] G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy," *SIAM, J. Numer. Anal.*, vol. 14, pp. 651-667, Sept. 1977.

[36] ——, "Bayesian 'confidence intervals' for the cross-validated smoothing splines," *J. Roy. Stat. Soc. B*, vol. 45, pp. 133-150, 1983.

[37] ——, "A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem," *Annals Statist.*, vol. 13, no. 4, pp. 1378-1402, 1985.

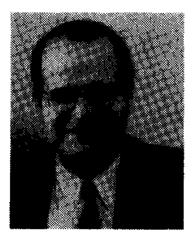
[38] ——, "*Spline models for observational data*," presented at SIAM-59, Philadelphia, PA, 1990.



Nikolas P. Galatsanos (M'89) was born in Athens, Greece in 1958. He received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1982, and the M.S. and the Ph.D. degrees in electrical engineering from the University of Wisconsin, Madison in 1984 and 1989, respectively.

Since 1989, he has been on the Faculty of the Department of Electrical and Computer Engineering at the Illinois Institute of Technology, Chicago, IL, where he is currently an Assistant Professor. His current research interests include image processing, and more specifically recovery, and compression of single and multichannel/frame images.

Dr. Galatsanos is a member of the OSA and the Technical Chamber of Greece.



Aggelos K. Katsaggelos (M'85) was born in Arnea, Greece, on April 17, 1956. He received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, Georgia, in 1981 and 1985, respectively.

From 1980 to 1985 he was a Research Assistant at the Digital Signal Processing Laboratory of the Electrical Engineering School at Georgia Tech. He is currently an Associate Professor in the Department of Electrical Engineering and Computer Science at Northwestern University, Evanston, IL. During the 1986-1987 academic year he was an Assistant Professor at Polytechnic University, Department of Electrical Engineering and Computer Science, Brooklyn, NY. His current research interests include signal and image processing, processing of image sequences, computational vision, and parallel implementation of signal processing algorithms.

Dr. Katsaggelos is an Ameritech Fellow and a member of the Associate Staff, Department of Medicine, at Evanston Hospital. He is also a member of SPIE, the Steering Committees of the IEEE Transactions on Medical Imaging, IEEE Transactions on Image Processing, the IEEE-CAS Technical Committee on Visual Signal Processing and Communications, the Technical Chamber of Commerce of Greece and Sigma Xi. He is an Associate Editor for the IEEE Transactions on Signal Processing and also the editor of the Springer-Verlag book *Digital Image Restoration*.