JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **114**, 37–51 (1986)

Projection Filter, Wiener Filter, and Karhunen–Loève Subspaces in Digital Image Restoration

HIDEMITSU OGAWA

Tokyo Institute of Technology, Department of Computer Science, Ookayama, Meguro-ku, Tokyo 152, Japan

AND

ERKKI OJA

Kuopio University, Department of Applied Mathematics, 70211 Kuopio 21, Finland

Submitted by Shunji Osaki

In image processing by computer, the transformation from the original continuous-domain image to the degraded and sampled discrete observation image is usually modelled as a linear transformation with additive noise. The relation between two types of filters, the Wiener filter (WF) and the projection filter (PF), for the restoration of the original image from the observation is discussed. The latter is based on the same principle as pseudoinverse filtering but also suppresses the additive noise. The PF and the WF are shown to be closely related under a condition depending on the degradation-sampling operator and the Karhunen–Loève expansion for the family of original images. The relation between the PF and the Gauss–Markov estimator is also clarified. © 1986 Academic Press, Inc.

1. INTRODUCTION

In image processing applications such as biomedical and remote sensing applications, our observations are usually degraded images. With $f(x)$ denoting the original image intensity at the two-dimensional spatial position x and $f_d(x)$ denoting the degraded image, obtained from f by a linear mapping H, the image restoration problem is to estimate f from f_d. There are three kinds of problems: (1) the operator H usually has no inverse; (2) in estimating the original image from its degraded observation by using a digital computer, we can only use values of the degraded image at a finite number of sample points; and (3) in practice there will be quan-

0022-247X/86 $3.00

tization and other errors, which are conveniently modelled as independent additive noise. Thus the digital image restoration problem is to estimate a continuous-domain function from a finite dimensional vector made up from the sample values of the degraded image and corrupted by additive random noise.

A number of estimation criteria have been suggested (see, e.g., [2]) either for the continuous–continuous case, in which both f and f_d are continuous-domain images, or for the discrete–discrete case, when both are discretized images. They include minimum variance filtering, pseudoinverse filtering, maximum entropy filtering, maximum likelihood estimation, and Bayesian estimation. For computational reasons, most practical methods rely on least-squares criteria and linear filters. In the continuous–continuous case, Fourier transform techniques are widely used, while for the discrete–discrete case an approach based on matrix algebra is taken.

Estimation in the continuous-discrete image degradation model has received much less attention, although it is the most natural model of image formation in practice. An exception is the case of band-limited signal processing. In Hilbert space methodology, the central tool when considering band-limited signals is the reproducing kernel of the space of such functions, allowing, e.g., the convenient sampling expansion. In image processing applications, the space of degraded or transformed images f_d may well be "band-limited" but not necessarily with respect to the Fourier transform. For discrete sampling, it is only important that this space should have some reproducing kernel, but in what follows no assumptions are made regarding this kernel.

In the general solutions to discrete-continuous image restoration, the approach has been to choose a priori a set of basis functions, say, cubic or linear splines [2, 4], and to compute the best estimate to f in terms of these fixed basis functions. Henceforth, such basis functions will be called restoration functions. Determining them in an optimal way from the point of view of some optimization criterion instead of for computational ease only is an important theoretical problem which has not been properly addressed in signal restoration literature.

In the case of linear estimation without additive noise, the family of estimates for all the original images constitutes a linear subspace $\mathscr{L}$ of the space of images, spanned by the restoration functions. The best approximation to an image f by elements of $\mathscr{L}$ is the orthogonal projection of f onto $\mathscr{L}$, with the approximation criterion thus depending on the inner product of the image space. The optimal linear image restoration problem in the noiseless case can be formulated as follows: obtain the orthogonal projection of f onto a subspace $\mathscr{L}$ as a linear function of the vector g made up from values $\{f_d(x_m), m = 1, 2,..., M\}$ of the degraded observation f_d at a finite number of sample points $\{x_m, m = 1, 2,..., M\}$.

When there is additive noise, the restoration is a sum of two components: the "image component," which is a linear function of the degraded and sampled image only, and the "noise component" which results from the same linear operator applied to the additive noise. A reasonable criterion is that the estimation of the image component only, regardless of noise, should reduce to the above problem. Then there is the extra problem of minimizing the noise component or maximizing the signal-to-noise ratio in the restored image.

In Section 2, the notation for the continuous-discrete image degradation model and the image and noise statistics are given. After that, the Wiener or minimum-variance filter is shortly reviewed in Section 3. In Section 4, those subspaces of the space of images f are characterized onto which the orthogonal projection of any f can be obtained as a linear mapping of the degraded and sampled vector g. The maximal subspace among them is determined. This provides the optimal set of restoration functions. The projection is optimal for each individual image f instead of being optimal in a statistical sense only. It is shown how the variance of the additive noise can be minimized in this case, leading to the so-called projection filter, earlier suggested by one of the present authors [7]. A relationship of the projection filter estimate and the Gauss–Markov estimate in statistical linear models is discussed.

A relation between the Gauss–Markov estimators and the pseudoinverse estimators has been given earlier by [11]. The estimators are equal if the columns of the coefficient matrix in the linear model are linear combinations of the eigenvectors of the noise covariance matrix. The purpose of Section 5 is to establish a somewhat similar relation between the Wiener filter and the projection filter. A condition is given under which an orthogonal projection can be computed on the subspace of Wiener estimates. It turns out that this condition depends on the relationship of the degradation mapping and the Karhunen–Loève (KL) subspaces, spanned by eigenfunctions of the covariance kernel of images f: all the restoration functions must be linear combinations of the eigenfunctions.

2. The Continuous-Discrete Model

Throughout this paper, we resort to the following notation for the linear image degradation-sampling model:

$$g = Af \tag{1}$$

with A a linear mapping on $\mathscr{H}$, the separable real Hilbert space of continuous-domain images f, into $\mathscr{R}^M$, the vector space of the degraded and sampled images g. With the additive quantization and other noise the model becomes

$$g = Af + n \tag{2}$$

with n the random noise vector. The linear restoration mapping on $\mathscr{R}^M$ into $\mathscr{H}$ is denoted by B, hence

$$f_1 = Bg = BAf + Bn \tag{3}$$

with f_1 the restored image.

The following will be assumed of the family of images:

$$E(f) = m, \qquad E(\langle f-m, f-m \rangle) = R, \tag{4}$$

where m is the mean of f in $\mathscr{H}$ and R is the nonnegative covariance kernel of f, a linear operator on $\mathscr{H}$ into $\mathscr{H}$. The notation $\langle \cdot, \cdot \rangle$, indicates a *dyad*, defined as $\langle f_1, f_2 \rangle f_3 = (f_3, f_2) f_1$. The inner product of $\mathscr{H}$ is $(\cdot, \cdot)$ and the norm induced by this inner product is $\|\cdot\|$. For a finite-dimensional $\mathscr{H}$, say, $\mathscr{H} = \mathscr{R}^N$, m is an $N \times 1$ vector and R is an $N \times N$ matrix. We also assume that $E\,\|f-m\|^2 < \infty$, which implies that R has finite trace. For $\mathscr{H}$, the L_2 norm and inner product can be used as a concrete example, but $\mathscr{H}$ could be any separable Hilbert space.

The noise ensemble satisfies the following:

$$E(n) = 0, \qquad E(\langle n, n \rangle) = Q \tag{5}$$

with Q an $M \times M$ nonnegative matrix. We also assume $E(\langle n, f-m \rangle) = 0$.

If an unbiased estimator is desired, i.e.,

$$E(f_1) = m, \tag{6}$$

then the mapping from g to f_1 must be affine linear. Therefore, in practice (3) may be replaced by

$$f_1 = Bg + b. \tag{7}$$

Then $E(f_1) = BE(g) + b = BAm + b = m$ if and only if

$$b = (I - BA)\,m. \tag{8}$$

The restoration becomes then

$$\begin{aligned} f_1 &= Bg + (I - BA)\,m, \\ &= BAf + Bn + (I - BA)\,m, \\ &= m + BA(f-m) + Bn, \end{aligned}$$

or

$$\bar{f}_1 = f_1 - m = B[A(f-m)+n],$$
$$= B\bar{g},$$

with

$$\bar{g} = A(f-m)+n = A\bar{f}+n,$$
$$= g - Am. \tag{9}$$

If m is known, $\bar{g}$ can be computed from the discrete image g and $\bar{f}_1$ can be obtained by the linear restoration filter B. If m is not known and no reasonable approximation is available, then usually the covariance kernel R is unknown. Then R should be replaced by the correlation kernel in (4), b is put to zero, and the requirement of unbiasedness is dropped. Then the orthogonality of noise and image is replaced by the assumption that n and f are uncorrelated.

Throughout this paper, the following convention in notation will be used: (1) If $E(f)=m$ is known, then f, f_1, and g will denote the normalized zero-mean variables ($\bar{f}$, $\bar{f}_1$, and $\bar{g}$ in Eqs. (9)) and R will be the covariance. The estimates will be unconditionally unbiased. (2) If $E(f)=m$ is not known, then f and g stand for the original image and sampled image, respectively, $f_1 = Bg$, and R is the correlation kernel of f.

3. Wiener Restoration in the Continuous-Discrete Model

The Wiener filter (WF) is a linear restoration mapping such that the original image f and the restoration f_1 satisfy

$$E\,\|f-f_1\|^2 = \text{minimum}. \tag{10}$$

Let A^* denote the adjoint operator of A and $\operatorname{tr}(\cdot)$ the trace of an operator. The WF is provided by

LEMMA 1. *The functional* (10) *has a minimum value if and only if B satisfies*

$$B(ARA^* + Q) = RA^*. \tag{11}$$

In this case the minimum value is given by

$$E\,\|f-f_1\|^2 = \operatorname{tr}(R - BAR). \tag{12}$$

Proof. The proof of Eq. (11) is analogous to the proof given in [2, p. 133]. Equation (12) follows by direct substitution.

Equation (11) gives the standard form of the Wiener restoration filter. If $ARA^* + Q$ is nonsingular, then

$$B = RA^*(ARA^* + Q)^{-1}, \tag{13}$$

a well-known result. Note that $ARA^* + Q$ is an $M \times M$ matrix even if $\mathscr{H}$ is infinite dimensional.

For the discrete–discrete degradation model it is usually assumed that A is homogeneous and the image field f and noise vector n are stationary. When all block Toeplitz forms are approximated by block circulant matrices, the DFT can be used for the effective computation of (13) [2]. Good restoration results are obtained with the WF especially in favorable signal-to-noise ratio cases.

4. Projection Filter

Consider first the following model without additive noise:

$$g = Af, \tag{14}$$

$$f_1 = Bg = BAf. \tag{15}$$

Equation (15) implies that for any f, f_1 is in the subspace $\mathscr{R}(BA)$, the range space of the operator BA. In this subspace, there is exactly one point which is closest to the original image f: the orthogonal projection of f on $\mathscr{R}(BA)$. It would be the optimal restoration. However, it is possible that the projection cannot be obtained in practice, since the restorations must be linear operations on g. We may ask: what are the subspaces of $\mathscr{H}$ such that we can compute the orthogonal projection of any f on them as a linear operation on g? It turns out that among all the subspaces of $\mathscr{H}$ there is exactly one maximal subspace that satisfies this criterion.

In answering the question, we resort to the concept of the pseudoinverse of a linear operator [5]. It follows that if A is any linear bounded operator on a Hilbert space, with $\mathscr{R}(A)$ closed, then A has a unique pseudoinverse A^+ satisfying the well-known Penrose equations:

$$AA^+A = A, \qquad A^+AA^+ = A^+,$$
$$(AA^+)^* = AA^+, \qquad (A^+A)^* = A^+A. \tag{16}$$

Since A in (14) has a finite-dimensional range, it has a unique pseudoinverse. Especially, the projection operator with range $\mathscr{R}(A^*)$ is $P = A^+A$.

LEMMA 2. *Let $\mathscr{L}$ be a subspace of $\mathscr{H}$. We can obtain the orthogonal projection of any $f \in \mathscr{H}$ on $\mathscr{L}$ as a linear operation on $g = Af$ if and only if $\mathscr{L} \subseteq \mathscr{R}(A^*)$.*

Proof. Denote the orthogonal projection operator on $\mathscr{L}$ by $P_{\mathscr{L}}$. Then the projection is $P_{\mathscr{L}} f$, and it is obtained as some linear operation K on g if and only if

$$KA = P_{\mathscr{L}}. \tag{17}$$

Assume first that $\mathscr{L} \subseteq \mathscr{R}(A^*)$. This is equivalent to $P_{\mathscr{L}} A^+ A = P_{\mathscr{L}}$, with $A^+ A$ the orthogonal projection operator on $\mathscr{R}(A^*)$. But then $K = P_{\mathscr{L}} A^+$ is a solution to (17). Second, assume that (17) has a solution. Then

$$P_{\mathscr{L}} A^+ A = KAA^+ A = KA = P_{\mathscr{L}},$$

hence $\mathscr{L} \subseteq \mathscr{R}(A^*)$.

This result means that $\mathscr{R}(A^*)$ is the maximal subspace on which we can project f in such a way the projection can be computed as Bg with some known linear operator B. Hence, any function basis of $\mathscr{R}(A^*)$ provides the optimal restoration basis. Typically, by Eq. (14), A is an $M \times 1$ vector whose elements are kernel functions, e.g., convolution kernels. These M functions span $\mathscr{R}(A^*)$. Equivalently, because $\mathscr{R}(A^*) = \mathscr{R}(A^*A)$ for operators with finite-dimensional range, such an optimal basis is provided by the orthogonal eigenfunctions of the self-adjoint operator A^*A belonging to nonzero eigenvalues.

The mapping B which yields the orthogonal projection on $\mathscr{R}(A^*)$ is obtained from Eq. (17). It follows that

$$BA = A^+ A. \tag{18}$$

The pseudoinverse filter, $B = A^+$ [2] is a special case of (18). For instance, every operator of the form $B = (A^*SA)^+ A^*S$ with S positive definite also satisfies (18). It is intuitively clear that this non-uniqueness of the solution of (18) should be used somehow to optimize among all the possible solutions. This is exactly what is done in defining the projection filter for the case of additive noise.

Consider now the more realistic case that f is random with correlation kernel R and there is additive zero-mean noise with correlation matrix Q, according to Eq. (2). An estimation criterion

$$E_n \| A^+ Af - f_1 \|^2 = \text{minimum}, \tag{19}$$

where E_n denotes expectation with respect to noise n, was originally used in [6, 8] to define the projection filter (PF). In this paper we give another criterion, leading to the same result.

The estimate f_1 can be written as

$$f_1 = Bg = BAf + Bn,$$
$$= f_1^{(i)} + f_1^{(n)}, \tag{20}$$

with $f_1^{(i)}$ the "image component" and $f_1^{(n)}$ the "noise component." A reasonable estimation criterion is that $f_1^{(i)}$ is the orthogonal projection of f on the maximal computable subspace, i.e., on $\mathscr{R}(A^*)$, and the signal-to-noise ratio (SNR) defined as $E \|f_1^{(i)}\|^2/E \|f_1^{(n)}\|^2$ is maximum. The operator E can be either expectation over both f and n or the expectation over n only. The result of Lemma 3 will be the same in both cases. In the latter case,

$$\mathrm{SNR} = \|f_1^{(i)}\|^2/E_n \|f_1^{(n)}\|^2,$$

and the signal-to-noise ratio will be maximized then for each individual image f.

LEMMA 3. *Assume* $BA = A^+A$. *Then the signal-to-noise ratio* $E \|f_1^{(i)}\|^2/E \|f_1^{(n)}\|^2$ *is maximized if and only if*

$$\mathrm{tr}(BQB^*) = \text{minimum}. \tag{21}$$

Proof. We have

$$\mathrm{SNR} = \mathrm{tr}(BARA^*B^*)/\mathrm{tr}(BQB^*),$$
$$= \mathrm{tr}(A^+ARA^+A)/\mathrm{tr}(BQB^*).$$

Since the numerator is independent of B, this is maximized exactly when the denominator is minimized, yielding (21).

Also the criterion (19) yields Eq. (21), as shown in [6].

We might note the similarity of (18) and (21) with the formulation of the Gauss–Markov estimator or BLUE (best linear unbiased estimator) [1]. For the linear model $g = Af + n$ with f a deterministic parameter vector and n the zero-mean noise with covariance matrix Q, the BLUE for f, $f_1 = Bg$ is a solution to the problem

$$BA = I,$$
$$\mathrm{tr}(BQB^*) = \text{minimum}. \tag{22}$$

The condition $BA = I$ guarantees conditional unbiasedness and $\mathrm{tr}(BQB^*) = E_n \|f - f_1\|^2$.

There is an essential difference between the image degradation model given here in Eq. (2) and the conventional linear statistical model. While the latter is usually overdetermined, with the number of scalar measurements exceeding that of the parameters, the situation is opposite in Eq. (2). Hence, A can never be of full "column rank" and $BA = I$ never has a solution. Hence the usual requirement of conditional unbiasedness, $E_n(f_1 \mid f) = f$ for all f, cannot hold. It is replaced by the condition $f_1^{(i)} = E_n(f_1 \mid f) = A^+Af$ for all f, i.e., the average of the restoration over noise is equal to the orthogonal projection of the original image f on the maximal computable subspace $\mathscr{R}(A^*)$. For this reason, the estimate obtained by the PF could also be termed the best linear projection estimate (BLPE).

The problem (18), (21) was solved in [6] and the result is

LEMMA 4. *The solution of* (18) *and* (21) *is*

$$B = A^+A[A^*G^+A]^+ A^*G^- + D(I - GG^-), \tag{23}$$

with D arbitrary,

$$G = AA^* + Q, \tag{24}$$

and G^- any operator satisfying,

$$GG^-G = G. \tag{25}$$

The solution (23) allows many special cases. One such special case is implied by the assumption that Q is positive definite. Then we have

LEMMA 5. *If $Q > 0$, then* (23) *is equal to*

$$B = (A^*Q^{-1}A)^+ A^*Q^{-1}. \tag{26}$$

Proof. Now $G = AA^* + Q > 0$, hence $G^- = G^+ = G^{-1}$, and Eq. (23) yields

$$B = A^+A(A^*G^{-1}A)^+ A^*G^{-1}.$$

This implies further

$$B = (A^*G^{-1}A)^+ (A^*G^{-1}A)(A^*G^{-1}A)^+ A^*G^{-1} = (A^*G^{-1}A)^+ A^*G^{-1},$$

because in general $A^+A = (A^*SA)^+ (A^*SA)$ for a positive definite operator S. We now make use of the identity

$$(Q + AA^*)^{-1} = Q^{-1} - Q^{-1}A(I + A^*Q^{-1}A)^{-1} A^*Q^{-1},$$

which is established by multiplying both sides by $Q + AA^*$. Substituting this above yields:

$$\begin{aligned} B &= \{A^*[Q^{-1} - Q^{-1}A(I + A^*Q^{-1}A)^{-1} A^*Q^{-1}] A\}^+ A^*G^{-1}, \\ &= \{A^*Q^{-1}A - A^*Q^{-1}A(I + A^*Q^{-1}A)^{-1} A^*Q^{-1}A\}^+ A^*G^{-1}, \\ &= \{A^*Q^{-1}A(I + A^*Q^{-1}A)^{-1}\}^+ A^*G^{-1}, \\ &= (A^*Q^{-1}A)^+ (I + A^*Q^{-1}A) A^*G^{-1}, \\ &= (A^*Q^{-1}A)^+ A^*Q^{-1}(AA^* + Q) G^{-1}, \\ &= (A^*Q^{-1}A)^+ A^*Q^{-1}, \end{aligned}$$

which concludes the proof.

Note that (26) has the form of the Gauss–Markov estimator [1]. However, $A^*Q^{-1}A$ is an operator on $\mathscr{H}$ into $\mathscr{H}$ and it has no inverse if dim $\mathscr{H} > M$, which is the case in the continuous-discrete degradation model. The operator B may be computed in some cases by the singular value decomposition technique applied to the operator A. Since A has finite range, it has only a finite set of singular values which are the square roots of the eigenvalues of the $M \times M$ matrix AA^*. In a discrete–discrete case, the computational problem of the projection filter has been studied in [10].

5. *PF*, *WF*, AND THE *KL* SUBSPACES

The questions that are addressed in this section are: (1) When do the Wiener filter, given by (11), and the projection filter, given by (23), coincide? (2) If they do not coincide, when can we compute the orthogonal projection of any f on the subspace of the Wiener estimates?

If case (1) follows, then the Wiener filter is optimal by two criteria. First, it is the minimum variance filter with respect to both signal and noise, and second, the image component in the estimate is the orthogonal projection of f on $\mathscr{R}(A^*)$ for each image f separately, with the signal-to-noise ratio minimized. Case (2) means the following: the image component in Wiener restoration is $f_1^{(i)} = BAf$ and the noise component is Bn, with the operator B given in (11). The orthogonal projection of f on $\mathscr{R}(BA)$ is $BA(BA)^+ f$. If we can express $BA(BA)^+$ in the form

$$BA(BA)^+ = XA, \tag{27}$$

for some operator X, then using X as the restoration filter yields

$$Xg = XAf + Xn = BA(BA)^+ f + Xn. \tag{28}$$

In this decomposition, the image component will be $BA(BA)^+ f$ and the noise component Xn. It follows that

$$\|XAf-f\| \leqslant \|BAf-f\| \qquad \text{for each } f. \tag{29}$$

Hence, if noise n is not taken into account, the restoration is improved in terms of the norm of $\mathscr{H}$.

However, since the Wiener filter is optimal in the sense that the total estimation error variance $E\|f-f_1\|^2 = E\|f-f_1^{(i)}\|^2 + E\|f_1^{(n)}\|^2$ is minimized, bringing $f_1^{(i)}$ closer to f means that the variance of the noise component will be increased. The above procedure means a trade-off between correcting the degradation and reducing the additive noise.

The question of the equivalence of the two filter types means that the operator sets $\mathrm{WF} = \{B \mid B(ARA^* + Q) = RA^*\}$ and $\mathrm{PF} = \{B \mid BA = A^+A$ and $\mathrm{tr}(BQB^*)$ is minimum$\}$ are the same. Here we restrict to the case when both WF and PF contain only one element, i.e., the two filters are both unique. A sufficient condition for the uniqueness is that either ARA^* or Q is nonsingular. In this situation, the following two theorems hold:

THEOREM 1. *Assume ARA^* is nonsingular. Then*

(i) $PF = \{A^+\}$.

(ii) $WF = \{RA^*(ARA^* + Q)^{-1}\}$.

(iii) If $Q \neq 0$, then $PF \neq WF$.

(iv) *If $Q = 0$, then $PF = WF$ if and only if*

$$\mathscr{R}(RA^*) \subseteq \mathscr{R}(A^*). \tag{30}$$

(v) *With $B = RA^*(ARA^* + Q)^{-1}$, the operator equation $BA(BA)^+ = XA$ has a solution X if and only if* (30) *holds, and the unique solution is*

$$X = BA(BA)^+ A^+. \tag{31}$$

THEOREM 2. *Assume Q is nonsingular and $A \neq 0$. Then*

(i) $PF = \{(A^*Q^{-1}A)^+ A^*Q^{-1}\}$.

(ii) $WF = \{RA^*(ARA^* + Q)^{-1}\}$.

(iii) $PF \neq WF$.

(iv) *With $B = RA^*(ARA^* + Q)^{-1}$, the operator equation $BA(BA)^+ = XA$ has a solution X if* (30) *holds, and the solution is*

$$X = BA(BA)^+ A^+ + C(I - AA^+) \tag{32}$$

with C an arbitrary linear operator on $\mathscr{R}^M$ into $\mathscr{H}$.

Proof of Theorem 1. (i) If ARA^* is nonsingular, then the rank of A is M, implying $AA^+ = I$. But then the operator equation $BA = A^+A$, which is always satisfied by the PF, has a unique solution $B = A^+$.

(ii) This is trivially true.

(iii) Assume that WF = PF, i.e.,

$$A^+ = RA^*(ARA^* + Q)^{-1}.$$

This implies $RA^* = A^+(ARA^* + Q)$, hence $ARA^* = AA^+ARA^* + AA^+Q = ARA^* + Q$. Thus $Q = 0$. The result follows by contraposition.

(iv) Now PF = WF is equivalent to $A^+ARA^* = RA^*$. But this is equivalent to $\mathscr{R}(RA^*) \subseteq \mathscr{R}(A^*)$.

(v) The equation $BA(BA)^+ = XA$ has solutions in X if and only if $BA(BA)^+ A^+A = BA(BA)^+$, in which case the general solution is given by Eq. (32). Now $BA(BA)^+ A^+A = BA(BA)^+$ is equivalent to $A^+ABA(BA)^+ = BA(BA)^+$ which is easily shown to be equivalent to $A^+ABA = BA$. Assume first that $A^+ABA = BA$ holds. Then substituting B from (ii) yields $A^+ARA^*(ARA^* + Q)^{-1} A = RA^*(ARA^* + Q)^{-1} A$ which multiplied on the right by $RA^*(ARA^*)^{-1} (ARA^* + Q)$ yields $A^+ARA^* = RA^*$. This implies (30). Assume then (30), hence $A^+ARA^* = RA^*$. This implies directly $A^+ARA^*(ARA^* + Q)^{-1} A = RA^*(ARA^* + Q)^{-1} A$, which is equivalent to $A^+ABA = BA$. It remains to show that the solution X is indeed given by (31). This follows from the general solution (32) when $AA^+ = I$ is substituted. This concludes the proof of Theorem 1.

Proof of Theorem 2. (i) Since Q is always nonnegative, this follows from Lemma 5.

(ii) This is trivially true.

(iii) Assume that $PF = WF$, i.e.,

$$(A^*Q^{-1}A)^+ A^*Q^{-1} = RA^*(ARA^* + Q)^{-1}.$$

This implies $RA^* = (A^*Q^{-1}A)^+ A^*Q^{-1}(ARA^* + Q) = (A^*Q^{-1}A)^+ (A^*Q^{-1}A) RA^* + (A^*Q^{-1}A)^+ A^* = A^+ARA^* + (A^*Q^{-1}A)^+ A^*$, since it follows that $(A^*Q^{-1}A)^+ (A^*Q^{-1}A) = (A^*Q^{-1/2}Q^{-1/2}A)^+ A^*Q^{-1/2} Q^{-1/2}A = (Q^{-1/2}A)^+ Q^{-1/2}A$ which is the orthogonal projection operator on $\mathscr{R}(A^*Q^{-1/2}) = \mathscr{R}(A^*)$, hence equal to A^+A. Further, we obtain from the above that $ARA^* = ARA^* + A(A^*Q^{-1}A)^+ A^*$, hence $A(A^*Q^{-1}A)^+ A^* = 0$. But then $(A^*Q^{-1}A)(A^*Q^{-1}A)^+ (A^*Q^{-1}A) = A^*Q^{-1}A = 0$, hence $A = 0$. The result follows by contraposition.

(iv) According to the proof of (v) in Theorem 1, it is sufficient to show that (30) implies $A^+ABA = BA$. Now (30) implies $A^+ARA^* =$

RA^*, hence $A^+ARA^*(ARA^*+Q)^{-1}A=RA^*(ARA^*+Q)^{-1}A$, hence $A^+ABA=BA$. This concludes the proof of Theorem 2.

The condition expressed in Eq. (30) of Theorems 1 and 2 can be given a convenient interpretation in terms of a well-known expansion for stationary second-order stochastic processes, the Karhunen–Loève expansion (see, e.g., [9]). The eigenfunctions and eigenvalues of the covariance kernel R are given by

$$Ru_i=\lambda_i u_i,\ i=1, 2,..., \tag{33}$$

with the $u_i \in \mathscr{H}$ and the λ_i are real numbers. By a Karhunen–Loève subspace of R we mean any subspace of $\mathscr{H}$ that is spanned by a subset of the u_i functions. We have

THEOREM 3. *$\mathscr{R}(RA^*)\subseteq\mathscr{R}(A^*)$ if and only if $\mathscr{R}(A^*)$ is a Karhunen–Loève subspace of the covariance kernel R.*

Proof. It is clear that if $\mathscr{R}(A^*)$ is spanned by a set of eigenfunctions of R, then $\mathscr{R}(RA^*)$ is a subspace of $\mathscr{R}(A^*)$. To prove the converse, assume $\mathscr{R}(RA^*)\subseteq\mathscr{R}(A^*)$. Let P be the orthogonal projection operator onto $\mathscr{R}(A^*)$. Since $R^*=R$, $\mathscr{R}(A^*)$ is a reducing subspace of R [3]. Hence we have $RP=PR$. Consider now the operator $R_1=RP$. Since R_1 is a selfadjoint operator with finite dimensional range, there exists an orthonormal basis $\{v_1,...,\}$ of $\mathscr{H}$ such that

$$R_1 v_n=\mu_n v_n.$$

Assume that exactly the first s eigenvalues $\mu_1,..., \mu_s$ are nonzero. It follows that

$$\mu_n v_n=R_1 v_n=RPv_n=PRv_n,$$

which by definition of P must be in $\mathscr{R}$ (A^*). This yields $v_n\in\mathscr{R}(A^*)$ for $n=1,..., s$. Then v_n is also an eigenfunction of R since

$$Rv_n=RPv_n=R_1 v_n=\mu_n v_n.$$

We consider now the following two cases:

(a) $s=\dim\mathscr{R}(A^*)$. Then $\mathscr{R}(A^*)$ is spanned by the functions $v_1,..., v_s$ which are also eigenfunctions of R.

(b) $s<\dim\mathscr{R}(A^*)$. Let now $\mathscr{M}$ be a subspace spanned by the $v_1,..., v_s$. Then $\mathscr{M}$ is also a subspace of $\mathscr{R}(A^*)$ and the orthogonal complement $\mathscr{M}^\perp$ is the null space of R_1. This shows that for any orthonormal basis $y_{s+1},..., y_d$ of $\mathscr{R}(A^*) \ominus \mathscr{M}$, with $d = \dim \mathscr{R}(A^*)$, it follows that $R_1 y_j = 0$. Since $y_j \subseteq \mathscr{R}(A^*)$, we have

$$Ry_j = RPy_j = R_1 y_j = 0,$$

showing that y_j is in fact an eigenfunction of R corresponding to eigenvalue zero. When we let $y_j = v_j$ for $j \leqslant s$, then $\mathscr{R}(A^*)$ is spanned by the basis $y_1,..., y_d$ which are eigenfunctions of R. This concludes the proof of Theorem 3.

In practice, if A^*g is given by

$$(A^*g)(x) = \sum_{i=1}^{M} g_i h_i(x), \tag{34}$$

for any real vector $g = (g_1,..., g_M)^T$ in $\mathscr{R}^M$, then $\mathscr{R}(A^*)$ is spanned by the functions h_i. Equivalently, the h_i functions could be eigenfunctions of A^*A belonging to nonzero eigenvalues. If

$$h_i(x) = \sum_{j=1}^{M} \gamma_{ji} u_j(x) \tag{35}$$

for some scalars γ_{ji}, then (30) is satisfied, and it is possible to obtain the orthogonal projection of f on the subspace of Wiener estimates.

Acknowledgments

This work was done during the second author's (EO) stay at the Department of Computer Science, Tokyo Institute of Technology. He is grateful for financial help from the Matsumae International Foundation, the Wihuri Foundation, and the Academy of Finland, during his visit.

References

1. A. Albert, "Regression and the Moore–Penrose Pseudoinverse," Academic Press, New York/London, 1972.
2. H. C. Andrews and B. R. Hunt, "Digital Image Restoration," Prentice–Hall, Englewood Cliffs, N.J., 1977.
3. S. L. Campbell and C. D. Meyer, Jr., "Generalized Inverses of Linear Transformations," Pitman, London/San Francisco, 1979.
4. W. Freiberger and U. Grenander, A method in pattern theory, *in* "Proc. 2nd Scand. Conf. on Image Analysis," Helsinki, Finland, June 15–17, 1981, pp. 11–17.
5. C. W. Groetsch, "Generalized Inverses of Linear Operators," Dekker, New York/Basel, 1977.
6. N. Nakamura and H. Ogawa, Optimum digital image restoration under additive noises, *Trans. IECE Japan* **J67–D**, No. 5 (1984), 563–570. [Japanese]

7. H. OGAWA, Optimum digital image restoration , *in* "Proc. Int. Conf. on Advances in Information Science and Technology," Calcutta, India, Jan. 11–14, 1982, pp. 8.1–8.7.
8. H. OGAWA AND N. NAKAMURA, Projection Filter restoration of degraded images, *in* "Proc. 7th Int. Conf. on Pattern Recognition," Montreal, Canada, July 30–Aug. 2, 1984.
9. E. OJA, "Subspace Methods of Pattern Recognition," Research Stud., Letchworth, Wiley, New York, 1983.
10. E. OJA AND H. OGAWA, Parametric projection filter for image and signal restoration, 1984, submitted for publication.
11. G. ZYSKIND, On canonical forms, nonnegative covariance matrices, and best and simple least squares estimators in linear models, *Ann. of Math. Statist.* **38** (1967), 1092–1109.