OPTICAL SCIENCES

THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA

TECHNICAL REPORT NO. 23

OPTIMUM, NON-LINEAR PROCESSING OF NOISY IMAGES

B. Roy Frieden

March 6, 1968

OPTICAL SCIENCES

The University of Arizona

Tucson, Arizona

Technical Report No. 23

OPTIMUM, NON-LINEAR PROCESSING OF NOISY IMAGES

B. Roy Frieden

March 6, 1968

ABSTRACT

It has been traditional to constrain image processing to linear operations upon the image. This is a realistic limitation of analog processing. In this paper, the calculus of variations is used to find the optimum, generally non-linear, processor of a noisy image. In general, such processing requires the use of an electronic computer. The criterion of optimization is that expectation $\langle |O_j - \bar{O}_j|^K \rangle$ be a minimum. Subscript j denotes the spatial frequency ω_j at which the unknown object spectrum O is to be restored, $\bar{O}$ denotes the optimum restoration by this criterion, and K is an even power at the user's discretion. A further generality is to allow the image-forming phenomenon to obey an arbitrary law $I_j = L(\tau_j, O_j, N_j)$. Here, τ_j denotes the intrinsic system characteristic (usually the optical transfer function), and N_j represents a noise function. The optimum $\bar{O}_j$ is found to be the root of a finite polynomial. When the particular value $K = 2$ is used, the root $\bar{O}_j$ is known analytically, along with the expected, mean-square, minimal error due to its use. When $K = 2$, processor $\bar{O}_j$ has the added significance of minimizing the total mean-square restoration error over the *spatial* object. This error may be further minimized by choice of an optimal processing bandwidth. Particular processors $\bar{O}_j$ are found for the "image recognition" problem and for the case of a "white" object region. The latter case is numerically simulated.

CONTENTS

BACKGROUND

The problem of estimating an object scene, given its image, is receiving accelerated attention.[1-10] A processed image cannot, strictly speaking, convey more information about the object than is already contained in the detected image.[11] However, it *can* be a closer facsimile to the object, both in the mean-square sense[3] and *to* the visual sense.[4] For these reasons, image processing can be a useful tool whenever decisions about an unknown object must be based on its experimental image, such as in astronomical and biological research, and the reconnaissance problem.

It has been shown[5-9] that the restoration can be perfect if (1) the image detection is perfect, i.e. noiseless, and (2) the image is formed by a linear process, where (3) the linear process is exactly specified by a known transfer function. Of course, none of these assumptions holds true in practice, for example if the image is a high-contrast photograph (the usual case of interest).

It is the goal of this paper to optimize the restoration *in the general absence* of assumptions (1) through (3). Hence, we treat the object scene, image-forming device(s), and noise as randomly fluctuating parameters. Under these conditions, the best that can be hoped for is a *statistically* optimum, albeit imperfect, restoration process. In order to seek this type of solution, it must be assumed that the statistics of all the fluctuating parameters are known, perhaps due to long-term observation of their behavior.

This problem is analogous to the "signal processing" problem of electrical engineering.[12] However, past solutions have usually been limited by the demands of analog processing, which constrains the processing to physically realizable operations: the linear (Weiner) filter,[12] the power law filter,[13] or other specialized types. By contrast, the development of electronic computers with sufficiently large memory banks has lately made possible the *digital* processing of images. The limitation of physical realizability is now removed, so that the processing may follow any prescribed law, such as the "optimal" laws derived in this paper.

As noted above, the optimum processor will be constructed, in part, from the known distribution function for the object statistics. In the following section we discuss the statistical format of optical objects in general, and find those object distributions which can be expected to physically occur. These distributions are later used to apply the optimal processor to specific cases.

STATISTICAL PROPERTIES OF OBJECTS

Definitions

The problem under consideration is estimation of an optical object $o(\vec{x})$, which is related to its spectrum $O(\vec{\omega})$ through

$$o(\vec{x}) = (2\pi)^{-1} \iint_{-\infty}^{\infty} d\vec{\omega}\, O(\vec{\omega}) e^{i\vec{\omega}\cdot\vec{x}} \equiv F^{-1}\{O\}; \quad i = (-1)^{1/2} \quad . \tag{1}$$

As usual, F^{-1} denotes the inverse Fourier transform. Hence, the problem of estimating $O(\vec{x})$ is equivalent to estimation of $O(\vec{\omega})$.

By the Fourier integral theorem[14]

$$O(\vec{\omega}) = (2\pi)^{-1} \iint_{-\infty}^{\infty} d\vec{x}\, o(\vec{x}) e^{-i\vec{\omega}\cdot\vec{x}} \equiv F\{o\} \ . \tag{2}$$

We assume $O(\vec{\omega})$ to be represented with sufficient accuracy by a finite array of values

$$O_1,\ O_2, \ldots,\ O_j, \ldots,\ O_J \tag{3a}$$

corresponding to spatial frequencies

$$\vec{\omega}_1,\ \vec{\omega}_2, \ldots,\ \vec{\omega}_j, \ldots,\ \vec{\omega}_J \quad . \tag{3b}$$

This subdivision of frequencies is assumed to be so fine that the discrete array (3a) causes negligible error when used to find $o(\vec{x})$ by Eq. (1).

"Foreknowledge" of object statistics

Definition of foreknowledge--It is reasonable to expect the quality of a restoration to vary directly as the experimenter's "foreknowledge" about $\mathcal{O}$-information that is at hand in addition to the detected image. As an extreme example, if the object is known to be bounded by a given, finite area, and if assumptions (1) through (3) above are true, the restoration can be made perfect.[5-9] A less extreme example is treated later.

Definition of object "class"--In this study, we assume foreknowledge of object statistics. Physically, this means that the unknown object is known to belong to a "class" of objects, i.e. a group of objects which are characterized by a known statistical behavior.

As an example, we might consider the class "views of the desert from altitude 5,000 ft. at 12m. on sunny days." Such fixed "seeing" conditions result in a fixed relation between each brightness fluctuation and its frequency of occurrence over all objects in the class.

Independent statistics at each frequency

We first consider the simplest case, where fluctuations at one frequency do not correlate with those at any other frequency. An object class is then defined by a fixed probability density[15] for object values $\mathcal{O}$ at each $\vec{\omega}_j$, denoted as

$$p_{\mathcal{O}}(\mathcal{O}_j), \quad j = 1,2,\ldots, J \quad . \tag{4}$$

Quantity $p_{\mathcal{O}}$ has the usual meaning for a probability density, that

$$p_{\mathcal{O}}(\mathcal{O}_j')\, d\mathcal{O}_j^{(re)}\, d\mathcal{O}_j^{(im)} \tag{5a}$$

represents the probability that

$$\mathcal{O}_j^{(re)} \le \mathcal{O}_j'^{(re)} \le \mathcal{O}_j^{(re)} + d\mathcal{O}_j^{(re)} \tag{5b}$$

and

$$\mathcal{O}_j^{(im)} \le \mathcal{O}_j'^{(im)} \le \mathcal{O}_j^{(im)} + d\mathcal{O}_j^{(im)} \quad . \tag{5c}$$

Superscripts (re), (im) denote real and imaginary parts.

We now establish the object statistics for two important limits of object class; those of complete determinacy, i.e. many identical views of the same object, and of Gaussian randomness. The latter might result when considering a very broad class of objects, such as views from a satellite of all portions of the earth. These two object classes are later used to find specific forms for the optimum processor.

Deterministic limit--In the limit of many photos of the same object, defining a very narrow class of objects, the probability density becomes infinitely sharp about one sequence of object values. Then

$$p_{\mathcal{O}}(\mathcal{O}_j) = \delta(\mathcal{O}_j - \mathcal{O}_{oj}), \tag{6a}$$

where δ is the Dirac delta function[16] and $\mathcal{O}_o$ is the object in question. Eq. (6a) states that there is no chance that the jth value of $\mathcal{O}$ in any one view of $\mathcal{O}_o$ can be anything but $\mathcal{O}_{oj}$.

Eq. (6a) can be generalized to the situation where L (finite) known objects are repetitively viewed in unknown order. If P_ℓ denotes the probability of occurrence of the ℓth object,

$$p_O(O_j) = P_1\delta(O_j - O_{1j}) + \ldots + P_L\delta(O_j - O_{Lj}) \quad . \tag{6b}$$

This type of object statistics occurs in the "image recognition" problem, which is considered later on.

Gaussian statistics--At the opposite extreme from the preceding, we now suppose object $o(\vec{x})$ to be chosen with complete randomness at each $\vec{x}$. In the absence of any information about "class," this would be the necessary assumption for the experimenter to make. If, in addition, the object is bandwidth-limited, it is known that[17] p_O is Gaussian:

$$p_O(O_j) = (2\pi\sigma_{Oj}{}^2)^{-1}\, e^{-|O_j - \langle O_j\rangle|^2/2\sigma_{Oj}{}^2} \quad . \tag{7a}$$

Here, the real and imaginary parts of O_j have the same variance, σ_{Oj}, and $\langle O_j\rangle$ is the (complex) mean. If σ_{Oj} is constant with j, the object has a "white" power spectrum.

It is interesting, and useful, to note that Gaussian statistics can approach deterministic "statistics." As $\sigma_{Oj} \to 0$,

$$p_O(O_j) \to \delta(O_j - \langle O_j\rangle). \tag{7b}$$

This may be shown by letting $\sigma_{Oj} \to 0$ in Eq. (7a) and using the defining properties[16] of the δ-function. Zero variance then represents the "class" of one object.

General, or Markov, object statistics

In the space domain--By visual inspection of his immediate surroundings, the reader may notice a simple, but important, property of everyday objects:

They are very redundant. For example, most object scenes contain large areas of virtually uniform brightness. In this case, knowledge that the brightness at point $\vec{x}_j$ is o_1 implies, with high certainty, that the brightness at nearest-neighbors $\vec{x}_{j\pm1}$ is also o_1. Undoubtedly, foreknowledge of this type can be used to advantage in restoring the object.

Because we shall be working in the frequency domain, the situation of a generally redundant object *spectrum* is treated next.

In the frequency domain--The general case arises when M + 1 parameters $O_j, O_{j(1)}, \ldots, O_{j(M)}$ are statistically dependent, but any other O_k is statistically independent of these parameters. Subscripts j(1),..., j(M) denote frequencies associated with frequency j in this way. They might, for example, be nearest neighbors to $\vec{\omega}_j$. The object (spectral) scene is then called a Markov[18] information array of order M. A complete description of the statistical behavior of O_j thereby requires knowledge of the *joint* probability density[15] $p_O(O_j, O_{j(1)}, \ldots O_{j(M)})$ for O_j and its M dependent neighbors.

The manner by which parameters $O_{j(1)}, \ldots, O_{j(M)}$ affect the statistics for O_j is reflected in the conditional probability law[15] $p(O_j|O_{j(1)}, \ldots, O_{j(M)})$. This is the probability density for O_j *if* values $O_{j(1)}, \ldots, O_{j(M)}$ are known. For later use, we note the connection between the joint and conditional probabilities to be[15]

$$p_O(O_j, O_{j(1)}, \ldots, O_{j(M)}) = p(O_{j(1)}, \ldots O_{j(M)}) \times p(O_j|O_{j(1)}, \ldots, O_{j(M)}) \quad . \qquad (8)$$

Use of the Markov description in image processing--The experimenter is assumed to be able to specify a joint probability law

$$p_o(O_j, O_{j(1)}, \ldots, O_{j(M)}) \qquad (9)$$

at each j. This probability law defines the object class in its most general aspect.

It is intuitive, from (8), that as M grows, a monotonically increasing amount of object foreknowledge is required. This suggests that as M grows the quality of the restoration should likewise increase monotonically. This suspicion will be borne out by means of a specific example.

The "disjoint" object--By stating that M = 0 the experimenter admits complete ignorance of the dependent object statistics over neighboring frequencies. In this case the joint probability for any L + 1 neighboring parameters O_j, $O_{j(1)}, \ldots, O_{j(L)}$ is strictly *dis*joint,[15] obeying

$$p_o(O_j, O_{j(1)}, \ldots, O_{j(L)}) = p_o(O_j)\, p_o(O_{j(1)}) \ldots p_o(O_{j(L)}) \quad . \qquad (10)$$

If, in reality, M > 0 the experimenter is now ignoring some object information. However, it will be seen that the processing can still be optimized, albeit to a lesser extent than if the actual value of M were used.

The highly redundant object--If a class of objects is sufficiently narrow in scope, the occurrence of any one set of values $O_{j(1)}, \ldots, O_{j(M)}$ implies *one* corresponding value for O_j. The conditional probability for O_j then degenerates to the description of a single value,

$$p_o(O_j | O_{j(1)}, \ldots, O_{j(M)}) = \delta[O_j - g_j(O_{j(1)}, \ldots, O_{j(M)})] \quad . \qquad (11)$$

Function g_j will generally vary with j.

An example of (11) is a "white" spectral region of extent M + 1. Here

$$O_j = O_{j(1)} = \dots = O_{j(M)} \quad , \tag{12a}$$

defining a

$$g_j(O_{j(1)}, \dots, O_{j(M)}) = O_{j(1)} \quad . \tag{12b}$$

This important case is used below in a numerical simulation of optimal processing.

When a disjoint object is highly redundant, it obeys Eq. (6a).

THE DETECTED IMAGE

The general law of image formation

The unknown object value O_j, after relay through an array of lenses and/or other relays of information, is detected as a noisy image $I(\vec{\omega}_j)$. We can specify the noiseless effect of this information relay, which might well be non-linear, by a system parameter $\tau(\vec{\omega}_j)$, and the noise of detection by a function $N(\vec{\omega}_j)$. For brevity, we denote these parameters as I_j, τ_j and N_j. In general, the value of I at *one* frequency is a function of τ, O and N at *all* spatial frequencies:

$$I_j = I(\tau_j, \tau_{j-1}, \tau_{j+1}, \ldots, O_j, O_{j-1}, O_{j+1}, \ldots, N_j, N_{j-1}, N_{j+1}, \ldots) \quad . \tag{13a}$$

A less general situation is where

$$I_j = L(\tau_j, O_j, N_j) \tag{13b}$$

alone. Operation L represents the physical law of image formation. For mathematical convenience, the optimal processor is found below for *all* cases where image formation takes on the somewhat restricted form (13b).

Important examples of laws L are next considered.

Linear image formation

The forms most frequently encountered are

$$I_j = \tau_j O_j + N_j \tag{14a}$$

and

$$I_j = \tau_j O_j N_j \quad . \tag{14b}$$

In these cases, τ represents a conventional transfer function, i.e. one which multiplies the object spectrum. The noise in (14a) is said to be "additive," in (14b) "multiplicative."

Case (14a) arises, for example, when an aerial image[19] is detected by a phototube whose noise limitation is predominantly a dark-current.[20] Case (14b) is the well-known representation of low-contrast, photographic image formation.[21] Here, I corresponds to photographic density, and N represents a transfer function of emulsion granularity.

Because of its mathematical convenience, law (14a) will be frequently assumed, below, in derivation of specific optimal processors.

Transfer functions are statistical parameters

Since τ_j represents a physical property of a real entity, it cannot be determined with arbitrary precision. When speaking of lens systems, τ_j seems to be determinable to about 3%.[22] Hence, any assumed value of τ_j suffers an unknown, statistical fluctuation from its true value.

The actual probability density $p(\tau_j)$ may be established quite simply, when values τ_j are determined from experimental use of the sampling theorem[23]

$$\tau_j = \text{constant} \sum_{m,n} s(\vec{r}_{mn}) e^{-i\pi \vec{r}_{mn} \cdot \vec{\omega}_j} \quad . \tag{15}$$

Here, s is the measured point spread function and the $\vec{r}_{mn}$ are sampling points. If the detected values s at points r_{mn} suffer from additive, statistically independent errors, $p(\tau_j)$ is known to be Gaussian.[17]

CLASSICAL IMAGE PROCESSING

The simplest method of processing the image is to assume the linear law (14a) of image formation, and to define the processed image as

$$\hat{O}_j = I_j/\tau_j \quad . \tag{16}$$

This method has been used quite successfully.[5,6] However, processor (16) is less than optimal since it ignores the possibility of minimizing the effects of noise upon the quality of restoration; it assumes the law *L* of image formation to be linear, which is often a poor approximation (e.g., in high-contrast photographs[24]); it ignores the possible use of statistical object information to aid the restoration process; and it assumes that each τ_j is precisely known.

We now derive a method of processing which takes into account all the preceding phenomena.

OPTIMAL IMAGE PROCESSING

Criterion of optimization

Let the processed image now be denoted as $\bar{O}_j$, j = 1,..., J. We seek to optimize $\bar{O}_j$ at each such j. Therefore, j should now be regarded as an arbitrary but *fixed* (frequency) value.

We shall seek the processor which is optimum for a given class of objects, the class being specified by known statistics (9). As previously discussed, the noise statistics and system parameter statistics are also assumed to be known.

The criterion which shall be used is that the expected *mean-power error of restoration* at the arbitrary point j

$$\varepsilon_j^{(K)} \equiv \langle \, |O_j - \bar{O}_j|^K \rangle = \text{a minimum} \quad . \tag{17}$$

The expectation $\langle \; \rangle$ is to be over many representative restorations of many typical members of the object class.

Possible powers K

Although the derivation below allows K to remain a general parameter, it is important to discuss the effect of particular values of K upon the expected error $\varepsilon_j^{(K)}$.

For any even K the minimum attained in (17) is zero if and *only* if $\bar{O}_j = O_j$ at each restoration. Therefore, a small minimum for $\varepsilon_j^{(K\ \text{even})}$ rigorously implies a good restoration (on the average). By contrast, if K is odd the minimum can be zero even though $\bar{O}_j$ differs widely from O_j for every restoration (positive and negative errors cancelling each other). Thus, K must be even for criterion (17) to be meaningful.

Among all even powers, the value K = 2 is the traditional choice. This is mainly because of mathematical tractability, but also because least-squares processing oftentimes results in the need for a linear filter, and linear filters are physically realizable. Of course, physical realizability is not of overriding importance in this study of optimal digital processing.

On the other hand, the use of higher K-values might often result in a smaller maximum error $|O_j - \bar{O}_j|$, over trial restorations, than for K = 2. Such control over the maximum error has proven to be advantageous in other problems.[25]

It might also prove useful to process one image with successive values of K = 2, 4,.... Each value of K would result in a different restoration, and this might be of use for purposes of object identification.

For these reasons, and the simple expedient of preserving the most general derivation, K is left as an experimental parameter in criterion (17). The special case K = 2, corresponding to optimum mean-square (ms) processing, will be considered in special cases.

Derivation of optimal processor

Image dependence--In seeking to optimize the resemblance between $\bar{O}$ and O at any one point j, we must allow $\bar{O}_j$ to be a function of all the observables which contain any information about O_j. According to the Markov property (8), the statistics of O_j depend upon the statistics of parameters O_k, k = j(1),..., j(M). When this is combined with the general law of image formation (13b), we see that all possible observable information about the unknown O_j is contained in the M + 1 image values I_k, k = j, j(1),..., j(M). Therefore, we let $\bar{O}_j$ be a general function

$$\bar{O}_j = \bar{O}_j(I_j, I_{j(1)}, \ldots, I_{j(M)}) \quad . \tag{18}$$

Criterion (17) is now

$$\varepsilon_j^{(K)} = \langle\, |O_j - \bar{O}_j(I_j, I_{j(1)}, \ldots, I_{j(M)})|^K \rangle = \text{min.} \tag{19}$$

Explicit dependence upon the statistics--In general, quantities τ_k, $k = j, j(1), \ldots, j(M)$ are statistically independent of the O_k, and both of these sets of parameters are statistically independent of the N_k. This independence is the result of differing physical laws for the phenomena τ, O and N. Therefore, Eq. (19) may be cast in terms of independent statistics p_τ, p_O and p_N for system parameter, object, and noise, respectively:

$$\int \cdots \int d\vec{\tau}\, d\vec{O}\, d\vec{N}\; |O_j - \bar{O}_j(\vec{I})|^K p_\tau(\vec{\tau}) p_O(\vec{O}) p_N(\vec{N}) = \text{minimum} \quad . \tag{20a}$$

For brevity, we adopt the notation

$$d\vec{a} = da_j da_{j(1)} \cdots da_{j(M)}, \quad \vec{a} = a_j, a_{j(1)} \cdots, a_{j(M)}, \quad \text{for } a = \tau, O \text{ or } N \quad . \tag{20b}$$

The integration in (20a) is over all possible values of the statistical parameters.

Solution by calculus of variations--The particular function $\bar{O}_j(\vec{I})$ satisfying criterion (20a) is the sought-after optimal processor. Although at this point the problem looks hopelessly difficult, a change of variables allows all the $d\vec{\tau}$ and $d\vec{O}$ integrations in (20a) to be carried through, greatly simplifying the problem.

Change of variables--Because the $\vec{I}$ in the argument of $\bar{\mathcal{O}}_j$ are themselves functions of the independent parameters $\vec{\tau}$, $\vec{\mathcal{O}}$ and $\vec{N}$ (through Eq. 13b), and because $\bar{\mathcal{O}}_j$ is an unknown function, as it stands Eq. (20a) cannot be simplified by successive quadratures. However, if we let parameters $\vec{I}$ replace the $\vec{N}$ as statistical coordinates, all terms in the integrand of (20a) may be integrated through $d\vec{\tau}d\vec{\mathcal{O}}$. Therefore, let Eq. (13b) be considered as an equation of transformation for defining new statistical coordinates $\vec{I}$. In addition, let the remaining variables $\vec{\tau}$ and $\vec{\mathcal{O}}$ be identically transformed to the primed variables

$$\tau_k' = \tau_k,\ \mathcal{O}_k' = \mathcal{O}_k,\ k = j,\ j\text{-}1, \ldots,\ j\text{-}M \quad . \tag{21}$$

The total Jacobian of the transformation is defined[26] as

$$J = \det \left| \frac{\vec{I},\ \vec{\mathcal{O}}',\ \vec{\tau}'}{\vec{N},\ \vec{\mathcal{O}},\ \vec{\tau}} \right| \quad . \tag{22a}$$

Since the (implied) matrix in (22a) is upper triangular, the determinant may be simply evaluated as the chain product[27]

$$J = f_j f_{j(1)} \cdots f_{j(M)} \tag{22b}$$

where

$$f_k \equiv f(I_k,\ \mathcal{O}_k,\ \tau_k) = \frac{\partial I_k^{(re)}}{\partial N_k^{(re)}} \frac{\partial I_k^{(im)}}{\partial N_k^{(im)}} - \frac{\partial I_k^{(re)}}{\partial N_k^{(im)}} \frac{\partial I_k^{(im)}}{\partial N_k^{(re)}} ,$$

$$k = j,\ j(1), \ldots j(M). \tag{22c}$$

Examining Eqs. (22b) and (22c), and using identities (21), we see that J is a real, known function of the new variables, i.e.

$$J = J(\vec{\tau}', \vec{O}', \vec{I}) \ . \tag{23a}$$

Also, with law L known it is possible to formally solve transformation Eq. (13b) for N_k in terms of the new statistical variables:

$$N_k = N(\tau_k', O_k', I_k) \ , \ k = j, \, j(1), \ldots, \, j(M). \tag{23b}$$

Transformations (21) are also used in (23b).

Transformed equation--By expressing the old differential in terms of the new,

$$d\vec{\tau} d\vec{O} d\vec{N} = |J|^{-1} \, d\vec{\tau}' d\vec{O}' d\vec{I} \ , \tag{24}$$

and with substitution (23b), Eq. (20a) becomes

$$\int \ldots \int d\vec{\tau}' d\vec{O}' d\vec{I} \, |J(\vec{\tau}', \vec{O}', \vec{I})|^{-1} |O_j' - \bar{O}_j(\vec{I})|^K \times \\ p_\tau(\vec{\tau}) p_O(\vec{O}') p_N(\vec{\tau}', \vec{O}', \vec{I}) = \text{minimum} \ . \tag{25}$$

Reduced equation--Eq. (25) may be integrated $d\vec{\tau}' d\vec{O}'$. After using the binomial theorem to expand out $|O_j' - \bar{O}_j|^K$, termwise integration results in the reduced equation

$$\int \cdots \int d\vec{I} \sum_{i,i'} \binom{K/2}{i} \binom{K/2}{i'} (-1)^{i+i'} \bar{O}_j(\vec{I})^i \bar{O}_j^*(\vec{I})^{i'} \times q^{(i,i')}(\vec{I}) = \text{minimum}. \qquad (26a)$$

The indicated summation is independently formed over $i,i' = 0,1,\ldots, K/2$; the asterisk denotes complex conjugate; and the first bracketed quantities are binomial coefficients. Also, the $q^{(i,i')}(\vec{I})$ are functions of the known statistics:

$$q^{(i,\, i')}(\vec{I}) = \int \cdots \int d\vec{\tau}' d\vec{O}' \; |J(\vec{\tau}', \vec{O}', \vec{I})|^{-1} \times O_j'^{K/2 - i} \; O_j'^{*K/2} \; p_\tau(\vec{\tau}') \; p_O(\vec{O}') \; p_N(\vec{\tau}', \vec{O}', \vec{I}). \qquad (26b)$$

Solution as root of polynomial--The integrand in Eq. (26a) is noted to be a function of variables $\vec{I}$ alone, there being no derivatives d/dI_k. In this simple case the calculus of variations[28] shows that the integral (26a) is a minimum if and only if the *integrand* is a minimum at each $\vec{I}$. Then $\bar{O}_j$ may simply be regarded as a variable which is to minimize the integrand, and which may be found by the usual method of differential calculus: $\partial/\partial\bar{O}_j$ (integrand) = 0. Since $\bar{O}_j$ is complex it has two degrees of freedom, so that in addition $\partial/\partial\bar{O}_j^*$ (integrand) = 0. Each of these equalities results in the requirement

$$\sum_{i,i'} \binom{K/2}{i} \binom{K/2}{i'} (-1)^{i+i'} i q^{(i,i')}(\vec{I}) \bar{O}_j(\vec{I})^{i-1} \bar{O}_j^*(\vec{I})^{i'} = 0 \,. \qquad (27)$$

This is the solution for the optimum processor due to a general law L of image formation.

Immediately apparent properties of $\bar{O}_j$

Because Eq. (27) is merely a polynomial of degree K - 1 in the real and imaginary parts of the unknown $\bar{O}_j$, it may be generally solved for $\bar{O}_j$ by numerical means. For each sequence $\vec{I}$ that is actually observed, a solution to (27) may be numerically generated by the proper subroutine for extracting roots from a polynomial. This procedure can conceivably be operated in "real time," by direct linkage between the detector of observables $\vec{I}$ and an electronic computer which extracts the appropriate root from Eq. (27) as the $\vec{I}$ are detected.

In the special case of K = 2, representing mean-square processing, the solution to Eq. (27) and the resulting minimal error $\varepsilon_j^{(2)}$ are both analytically known (as found below). This is of definite advantage for understanding the effects upon $\bar{O}_j$ and $\varepsilon_j^{(2)}$ of varying parameters in the distributions p_τ, p_O and p_N.

The $\vec{I}$-dependence of the optimal $\bar{O}_j$ is through coefficients $q^{(i,i')}(\vec{I})$. Noting the dependence (26b) of $q^{(i,i')}$ upon input functions p_τ, p_O, and p_N, $\bar{O}_j$ cannot be generally expected to be linear in observables $\vec{I}$. Hence, *linear processing is not generally optimal processing*. This subject is further pursued below.

Because it effects an analytic solution to Eq. (27), the case K = 2 is considered from this point on.

Optimal mean-square processing

With K = 2, polynomial (27) is linear in O_j^* so that the solution is immediate:

$$\bar{O}_j(\vec{I}) = q^{(1,0)*}(\vec{I})/q^{(1,1)}(\vec{I}) \quad . \tag{28}$$

The resulting minimum error may also be analytically found. By substitution of processor (28) into error experssion (26a), we find

$$\varepsilon_j^{(2)} = \int \ldots \int d\vec{I}[q^{(0,0)}(\vec{I}) - |q^{(1,0)}(\vec{I})|^2 / q^{(1,1)}(\vec{I})] \quad . \tag{29}$$

In the special case of M = 0, L linear, and $\vec{\tau}$ non-statistical, solution (28) has been derived before.[29] We wish to emphasize, however, that due to the general nature of L allowed in solution (28), it is also applicable to cases where the image formation and detection are inherently *non*-linear.

We note, from Eqs. (26b) and (28), that optimal processing is not generally linear processing even when K = 2.

Solution (28) is now applied to some important cases.

Linear image formation with additive noise--*General solution*--By substitution of Eq. (14a) into Eqs. (22), we find that Jacobian

$$J = 1 \quad . \tag{30a}$$

Further substitution of Eqs. (14a) and (30a) into Eqs. (26b) and (28) yields the optimum ms processor for this case:

$$\bar{O}_j(\vec{I}) = \frac{\int \ldots \int d\vec{\tau}' d\vec{O}' O_j' p_\tau(\vec{\tau}') p_O(\vec{O}') p_N(\vec{I} - \vec{\tau}'\vec{O}')}{\int \ldots \int d\vec{\tau}' d\vec{O}' p_\tau(\vec{\tau}') p_O(\vec{O}') p_N(\vec{I} - \vec{\tau}' \vec{O}')} \tag{30b}$$

where, for brevity, we adopt notation

$$(\vec{I} - \vec{\tau}'\vec{O}') \equiv (I_j - \tau_j' O_j', \ldots, I_{j(M)} - \tau_{j(M)}' O_{j(M)}') \quad . \tag{30c}$$

For well-behaved functions p_τ, p_O, and p_N, processor $\bar{O}(\vec{I})$ can be numerically or analytically generated from (30b).

Case of a deterministic transfer function--If the fluctuations in $\vec{\tau}'$ are so small as to be negligible with respect to those of $\vec{O}'$, the indication is that each τ_k' has been measured very accurately. The deterministic limit [see the discussion of Eq. (6a)]

$$p_\tau(\vec{\tau}') = \delta(\vec{\tau}' - \vec{\tau}_0) \tag{31a}$$

may then be used in Eq. (30b). Each τ_k' now has a precise value, τ_{0k}. The result is a processor

$$\bar{O}_j(\vec{I}) = \frac{\int \cdots \int d\vec{O}' O_j' p_O(\vec{O}') p_N(\vec{I} - \vec{\tau}_o \vec{O}')}{\int \cdots \int d\vec{O}' p_O(\vec{O}') p_N(\vec{I} - \vec{\tau}_o \vec{O})} \quad . \tag{31b}$$

The only remaining integrations are the $d\vec{O}'$.

Classical limit--If the noise is known to be zero at $\vec{\omega}_j$, so that p_N is separable as

$$p_N(\vec{I} - \vec{\tau}_o \vec{O}') = \delta(I_j - \tau_{oj} O_j') p_N(I_{j(1)} - \tau_{oj(1)} O_{j(1)}', \ldots), \tag{32a}$$

then by substitution of Eq. (32a) into Eq. (31b),

$$\bar{O}_j(\vec{I}) = I_j / \tau_{oj} \quad . \tag{32b}$$

This is just the classical, linear processing method (16). Classical processing is therefore optimal processing when the image formation is linear, with a precisely known transfer function, and when the noise is additive and known to be zero. With any departure from these conditions, the use of processor (32b) must be less than optimum.

OPTIMAL PROCESSING IN SPACE DOMAIN

A method has been established for minimizing the error of restoration in the frequency domain. However, the aim of image processing is to optimally restore *spatial* objects, $o(\vec{x})$. It is therefore necessary to find the effect of optimal processor $\bar{O}_j$ upon the quality of the spatial restoration. We find it convenient to measure this "quality" by the following criterion.

Mean-square error of spatial restoration

Let $\hat{O}(\vec{\omega})$ represent the restored value of $O(\vec{\omega})$ by any restoration process. For example, $\hat{O}$ may be $\bar{O}$. Then, by use of Eq. (1) the corresponding restoration in the space domain is

$$\hat{o}(\vec{x}) = (2\pi)^{-1} \iint_{-\vec{\Omega}}^{\vec{\Omega}} d\vec{\omega}\hat{O}(\vec{\omega})e^{i\vec{\omega}\cdot\vec{x}}, \quad \vec{\Omega} \equiv (\Omega_1, \Omega_2) \quad . \tag{33}$$

(We now take points $\vec{\omega}_j$ so close together that they define a continuous variable $\vec{\omega}$.) That frequencies Ω_1, Ω_2 are necessarily finite is due to the well-known limitation of band-limited[19] image formation. By use of Eq. (1) we can form the error of restoration $\Delta o(\vec{x}) \equiv o(\vec{x}) - \hat{o}(\vec{x})$:

$$\Delta O(x) = (2\pi)^{-1} \iint_{-\vec{\Omega}}^{\vec{\Omega}} d\vec{\omega}[O(\vec{\omega}) - \hat{O}(\vec{\omega})]e^{i\vec{\omega}\cdot\vec{x}} + (2\pi)^{-1} \left(\iint_{-\infty}^{-\vec{\Omega}} + \iint_{\vec{\Omega}}^{\infty} \right) d\vec{\omega}O(\vec{\omega})e^{i\vec{\omega}\cdot\vec{x}} \quad . \tag{34}$$

In analogy with criterion (17) for restoration quality in the frequency domain, we measure the restoration quality, over the *entire* spatial object, by the single parameter

$$E^{(2)} \equiv \left\langle \iint_{-\infty}^{\infty} d\vec{x}[\Delta o(\vec{x})]^2 \right\rangle \quad . \tag{35}$$

Use of the special power 2 in criterion (35) allows $E^{(2)}$ to be related to $\varepsilon^{(2)}$, as is shown below.

By substitution of Eq. (34) into Eq. (35), we find that

$$E^{(2)} = \iint_{-\vec{\Omega}}^{\vec{\Omega}} d\vec{\omega} \left\langle |O(\vec{\omega}) - \hat{O}(\vec{\omega})|^2 \right\rangle + \left(\iint_{-\infty}^{-\vec{\Omega}} + \iint_{\vec{\Omega}}^{\infty} \right) d\vec{\omega} \left\langle |O(\vec{\omega})|^2 \right\rangle \tag{36}$$

holds true exactly.

<u>An important correspondence</u>

Eq. (36) shows that $E^{(2)}$ is minimized by the choice of $\hat{O}$ such that $\langle |O(\vec{\omega}) - \hat{O}(\vec{\omega})|^2 \rangle$ is a minimum at each $\vec{\omega}$. But this is precisely the definition (17) of the optimum *mean-square* processor $\bar{O}(\vec{\omega})$. We have therefore established the important fact that *optimum mean-square processing in the frequency domain results in a minimal error of restoration* $E^{(2)}$ *over the entire spatial object*. This is a further virtue of using K = 2 in criterion (17). Mean-square processors (28), (30b), and (31b) now have added significance.

The quantity $\langle |O(\vec{\omega})|^2 \rangle$ in Eq. (36) is usually called the "power spectrum" for the object, and we denote it as $\phi_O(\vec{\omega})$. Quantity ϕ_O is, of course, independent of any processing scheme. By this change of notation, and the substitution $\hat{O} = \bar{O}$ in Eq. (36), the minimum value of $E^{(2)}$ due to choice of a processor is simply

$$E_m^{(2)} = \iint_{-\vec{\Omega}}^{\vec{\Omega}} d\vec{\omega}\varepsilon^{(2)}(\vec{\omega}) + \left(\iint_{-\infty}^{-\vec{\Omega}} + \iint_{\vec{\Omega}}^{\infty}\right) d\vec{\omega}\phi_o(\vec{\omega}) \quad . \tag{37}$$

All quantities on the right-hand side may be assumed as known to the user (constituting "foreknowledge," as previously discussed).

Optimal processing bandwidth

Derivation--The only free parameters remaining in Eq. (37) are Ω_1 and Ω_2, which define a "processing bandwidth" $\vec{\Omega}$ by Eq. (33). From the form of Eq. (37), an optimum $\vec{\Omega}$ exists such that contributions toward $E_m^{(2)}$ of $\varepsilon^{(2)}(\vec{\omega})$ are balanced against those from $\phi_o(\vec{\omega})$, resulting in a minimum for $E^{(2)}$. We now find this optimum $\vec{\Omega}$ in the special, but usual, case where object class, noise and system parameter are isotropic in $\vec{\omega}$, i.e.

$$\varepsilon^{(2)}(\vec{\omega}) = \varepsilon^{(2)}(\omega), \ \phi_o(\vec{\omega}) = \phi_o(\omega), \text{ and } \vec{\Omega} = \Omega \quad . \tag{38a}$$

Under conditions (38a), error (37) becomes

$$E_m^{(2)} = 2\pi \int_0^{\Omega} d\omega \ \omega\varepsilon^{(2)}(\omega) + 2\pi \int_{\Omega}^{\infty} d\omega \ \omega\phi_o(\omega) \quad . \tag{38b}$$

We now use the ordinary rules of calculus to find a minimum in $E_m^{(2)}$. Setting $d\,E_m^{(2)}/\,d\Omega = 0$, we find that Ω must satisfy

$$\varepsilon^{(2)}(\Omega) - \phi_o(\Omega) = 0. \tag{39}$$

Since the functions $\varepsilon^{(2)}(\Omega)$ and $\phi_o(\omega)$ are assumed known [see Eq. (29)], the optimum Ω may be found as a root of Eq. (39). In the special case $\phi_o(\omega)$ = constant $\equiv c$ of a "white" power spectrum, the optimum Ω is simply that value for which $\varepsilon^{(2)}(\Omega) = c$.

Simple example--With $\hat{O} \neq \bar{O}$ in general, Eq. (39) may be replaced by

$$\langle |O(\Omega) - \hat{O}(\Omega)|^2 \rangle - \phi_O(\Omega) = 0 \quad . \tag{40}$$

I.e., Eq. (40) defines the bandwidth Ω for minimizing $E^{(2)}$ when *any* processing law $\hat{O}$ is used. It may be shown that Eq. (40) also holds in the case where all parameters are one-dimensional (rather than radial), and symmetric in ω.

It is then possible to find the optimum Ω for this very simple case: linear law (14a) of image formation; non-statistical transfer function[19]

$$\tau(\omega) = \begin{cases} 1 - |\omega|/\Omega_o & \text{for } |\omega| \le \Omega_o \\ 0 & \text{for } |\omega| > \Omega_o \end{cases} \tag{41}$$

due to one-dimensional, diffraction-limited optics; use of the classical (rather than optimal) processor (16); and constant power spectra ϕ_O and ϕ_N. By combining Eqs. (14a), (16), (40), and (41), we find the optimum bandwidth for classical processing to be

$$\Omega = \Omega_o(1 - \phi_N/\phi_O) \quad . \tag{42}$$

The variation of Ω with signal-to-noise ratio ϕ_O/ϕ_N in (42) agrees with intuition: The allowable Ω increases monotonically with ϕ_O/ϕ_N, approaching a maximum value of Ω_o as $\phi_O/\phi_N \to \infty$ (the noiseless case). Because any image is bandlimited[19] to frequency Ω_o, Ω_o is indeed the largest possible processing frequency.

In the other extreme, as $\phi_O/\phi_N \to 1$, Eq. (42) makes $\Omega \to 0$. When the noise is so great that it equals the signal, processing by classical formula (16) is ruled out.

APPLICATIONS

In the two general applications described below, we use optimum ms processing of a linear image (14a), with non-statistical τ. Therefore, in each case the starting point is processing formula (31b).

Image recognition problem

Optimum processor--Suppose the object class contains a finite number L of known members. An example is the "character recognition" problem, where L = 26 for English letters. The detected image must then identify with one of the L objects. The problem we tackle is to find the optimum processor for making the required identification.

Let the known transfer function be τ_o, the known L objects $\mathcal{O}_{\ell k}$, ℓ = 1, 2,..., L. (As usual, k denotes frequency.) The *noiseless* image $I_{\ell k}$ corresponding to $\mathcal{O}_{\ell k}$ is given as[19]

$$I_{\ell k} = \tau_{ok}\mathcal{O}_{\ell k} \quad . \tag{43}$$

Since all the right-hand quantities are known, the $I_{\ell k}$ are likewise known. Therefore, identification of the detected image with one $\mathcal{O}_\ell$ is equivalent to identification with one I_ℓ.

Because the objects $\mathcal{O}_\ell$ are a discrete set with a finite number of members, $p_{\mathcal{O}}$ is a Dirac "comb function" [see Eq. (6b)]

$$p_{\mathcal{O}}(\vec{\mathcal{O}}') = P_1\delta(\vec{\mathcal{O}}' - \vec{\mathcal{O}}_1) + \ldots + P_L\delta(\vec{\mathcal{O}}' - \vec{\mathcal{O}}_L) \quad , \tag{44a}$$

where

$$P_1 + P_2 + \ldots + P_L = 1 \quad . \tag{44b}$$

Each P_ℓ is the known, finite probability associated with the occurrence of its O_ℓ. Thus, Eq. (44b) states that one of the L objects must be present.

The optimum ms processor may now be found. By substitution of Eq. (44a) into processor (31b), we find

$$\bar{O}_j(\vec{I}) = \frac{1}{\tau_{oj}} \frac{\sum_{\ell=1}^{L} P_\ell I_{\ell j} p_N(\vec{I} - \vec{I}_\ell)}{\sum_{\ell=1}^{L} P_\ell p_N(\vec{I} - \vec{I}_\ell)} . \tag{45}$$

Discussion--The behavior of optimum ms processor (45) may be described as follows:

(1) For general p_N, $\bar{O}_j$ is generally *non-linear* in the observables, $\vec{I}$.

(2) $\bar{O}_j$ is essentially the quotient of two weighted series of noise distributions.

(3) As the noise approaches zero (and is known to do so), $\bar{O}_j(\vec{I}) \to O_{\ell' j}$, where ℓ' is the object that is actually present. This is a self-consistency test for the processor.

(4) As one $P_{\ell'} \to 1$, i.e. the appearance of one object becomes increasingly probable, Eq. (44b) indicates that the $P_\ell \to \delta_{\ell\ell'}$, where δ is the Kronecker delta-function. Using this limit in Eq. (45) we find that $\bar{O}_j(\vec{I}) \to O_{\ell' j}$. Hence, the restoration is perfect, and is independent of the observables $\vec{I}$. In a less extreme case, where (say) $P_{\ell'} = 0.9$, the tendency $\bar{O}_j(\vec{I}) \to O_{\ell' j}$ remains, with near independence of the observables $\vec{I}$.

(5) The most random[30] noise case arises when the noise is uniform and independent at each $\vec{\omega}_j$. This situation is approximated by a noise distribution

$$p_N(\vec{I} - \vec{I}_\ell) = \begin{cases} \text{constant } b^{M+1} \text{ for } |I_k - I_{\ell k}| \le (2b)^{-1} \\ 0 \text{ for } |I_k - I_{\ell k}| > (2b)^{-1} \quad , \end{cases} \tag{47a}$$

$k = j, j(1), \ldots, j(M)$, b small,

at each $\vec{\omega}_j$. Use of (47a) in (45) results in the degenerate processor

$$\bar{O}_j(\vec{I}) = \sum_{\ell=1}^{L} P_\ell O_{\ell j} \equiv \langle O_j \rangle \quad . \tag{47b}$$

Or, the optimum processor is simply the average value of O_j over the L objects, *independent* of observables $\vec{I}$. Upon reflection, this is as it should be. Since practically every value of noise is now equally likely [including very large values, according to Eq. (47a)], observables $\vec{I}$ actually yield no information about $O_{\ell j}$. Therefore, the optimum estimate of $O_{\ell j}$ can only be based upon the object foreknowledge, and $\langle O_j \rangle$ in particular.

Decision theory aspect--Although processor (45) yields the best estimate of the unknown object, it cannot by itself identify this estimate $\bar{O}$ with one of the O_ℓ. A "decision-making" function of observables $\vec{I}$ is needed for making this identification.

Derivation--Because τ' is a non-statistical quantity of the problem, the optimum estimate $\bar{I}_j$ of the noiseless *image* obeys

$$\bar{I}_j(\vec{I}) = \tau_{oj} \bar{O}_j(\vec{I}) \quad . \tag{48a}$$

Let $I_{\ell'}$ denote the true image under observation. Because of criterion (19) and Eq. (48a)

$$\sum_{j=1}^{J} \langle |I_{\ell' j} - \bar{I}_j(\vec{I})|^2 \rangle = \text{minimum} \tag{48b}$$

where, from Eqs. (45) and (48a), the optimum image estimate is

$$\bar{I}_j(\vec{I}) = \frac{\sum_{\ell=1}^{L} P_\ell I_{\ell j} p_N(\vec{I} - \vec{I}_\ell)}{\sum_{\ell=1}^{L} P_\ell p_N(\vec{I} - \vec{I}_\ell)} \quad . \tag{48c}$$

A decision function may now be formed. According to identity (48b), on the average

$$\sum_{j=1}^{J} |I_{\ell' j} - \bar{I}_j(\vec{I})|^2 < \sum_{j=1}^{J} |I_{\ell''} - \bar{I}_j(\vec{I})|^2 \quad . \tag{48d}$$

Here $I_{\ell'}$ is the unknown image and $I_{\ell''}$ is any of the remaining (L-1) images. It is therefore logical to construct a decision function as

$$D(\ell) = \sum_{j=1}^{J} g[|I_{\ell j} - \bar{I}_j(\vec{I})|^2], \; \ell = 1,2,\ldots, L, \tag{49a}$$

where the ℓ resulting in the smallest D is accepted as identifying the true image. Function g [ ] may be imagined to be unity, but for generality it is considered as any monotonically increasing function.

An optimal decision function?--Given one set of restored values $\bar{I}_j(\vec{I})$, $j = 1, 2, \ldots, J$, the value of ℓ that minimizes D must depend upon the function g chosen. For example, if $g = \log_e$, any $D(\ell)$ is strongly negative if $I_{\ell j} \cong \bar{I}_j(\vec{I})$ at *but one* value of j. This tendency biases decisions in favor of an ℓ for which there are one or two near coincidences $I_{\ell j} \cong \bar{I}_j(\vec{I})$, almost regardless of the differences $I_{\ell j} - \bar{I}_j(\vec{I})$ at all other values of j.

Since the decision depends upon the form of function g, we can conceive of an optimal function g such that the probability of making a correct decision is maximized. To find this function g would constitute a major advance in image recognition.

Comparison with classical decision function--If the noise distribution p_N is Gaussian, the classical function for making a decision ℓ is[1]

$$C(\ell) = \sum_{j=1}^{J} |I_{\ell j} - I_j|^2, \quad \ell = 1,2, \ldots, L \quad . \tag{50}$$

Again, the smallest C is accepted as identifying the true image.

It can be shown that if p_N in Eq. (48c) is Gaussian, with L = 2 (a "binary" decision problem), classical function $C(\ell)$ has the same probability of success as our function $D(\ell)$ with the choice $g = \log_e$.

Comparing Eqs. (49a) and (50), we see that $C(\ell)$ is formed from noisy data I_j, while $D(\ell)$ is formed from optimally processed data $\bar{I}_j$. We may therefore suspect that use of $D(\ell)$ will more usually lead to successful decisions than will use of $C(\ell)$. However, the preceding paragraph does not support this hypothesis (although neither does it disprove it). Further research on this hypothesis is indicated.

Processing a white object region

Derivation of optimal processor--It has previously been suggested that knowledge of the statistical behavior of O over neighboring frequencies $\vec{\omega}$ can improve the quality of restoration. We now show this to be true in the case (12) where O is known to be constant (but with unknown value) over M + 1 adjacent frequencies $\vec{\omega}_{j+k}$, k = 0, ±1, ±2, ..., {M'}; where {M'} = ±M/2 for M even, or {M'} = ±(M - 1)/2, - (M + 1)/2, for M odd. Thus, O_j is centered in the white region. In addition, the object statistics at $\vec{\omega}_j$ are to be Gaussian, and the additive noise N_j is Gaussian and independent over the $\vec{\omega}_{j+k}$. All the Gaussian parameters are assumed known.

With these conditions substituted into formula (31b), the optimal ms processor is found to be

$$\bar{O}_j(\vec{I}) = \frac{\sigma_N^2 \langle O_j \rangle + \sigma^2_{Oj} \sum_{k=0}^{\{M'\}} \tau_{j+k}{}^* I_{j+k}}{\sigma_N^2 + \sigma^2_{Oj} \sum_{k=0}^{\{M'\}} |\tau_{j+k}|^2} . \quad (51a)$$

The expected error is found by using Eq. (51a) and criterion (19), with $K = 2$,

$$\epsilon_j^{(2)} = \frac{2\sigma_N^2 \sigma_{Oj}^2}{\sigma_N^2 + \sigma_{Oj}^2 \sum_{k=0}^{\{M'\}} |\tau_{j+k}|^2} . \quad (51b)$$

In the preceding, σ_N and σ_{Oj} are the variances for noise and object statistics, respectively, $\langle O_j \rangle$ is the mean value of O_j, and the mean value of the noise is known to be zero.

Discussion--The most important feature of processor (51a) is that it is linear in the observables $\vec{I}$. Hence, (51a) is nearly as simple to use as the classical processor (16).

The dependence of $\bar{O}_j$ upon signal and noise parameters σ_{Oj}, $\langle O_j \rangle$, σ_N is appropriate to its "optimal" nature. As either $\sigma_{Oj} \to 0$ [limit of essentially one object; see Eq. (7b)], or $\sigma_N \to \infty$, $\bar{O}_j(\vec{I}) \to \langle O_j \rangle$. In the former case, O_j must uniquely be its mean value, so that $\langle O_j \rangle$ is the proper limit for $\bar{O}_j(\vec{I})$. [Note also the consistent behavior, $\epsilon_j^{(2)} \to 0$, of Eq. (52b).]

In the latter case, since $\sigma_N \to \infty$, all readings $\vec{I}$ are submerged in noise. Hence readings $\vec{I}$ contain no information about O_j, and processor $\bar{O}_j$ accordingly ignores this data. Since the only remaining knowledge about O_j which can be of aid is the known mean value, $\langle O_j \rangle$, $\bar{O}_j$ takes on this value. Error expression (52b) is consistent with this interpretation, yielding $\varepsilon_j^{(2)} \to 2\sigma_{Oj}^2$. Hence, the expected error grows with the ability of O_j to statistically vary from $\langle O_j \rangle$.

To what extent does object foreknowledge aid the restoration?--In this problem, object parameters $\langle O_j \rangle$, σ_{Oj} and M are assumed known to the user. The effect of each of these quantities upon the quality of restoration can be measured by error expression (51b).

Since parameter $\langle O_j \rangle$ does not enter into Eq. (51b), the magnitude of $\langle O_j \rangle$ does not affect the restoration quality. This agrees with intuition: It does not require more knowledge to know that $\langle O_j \rangle$ is large, than that $\langle O_j \rangle$ is small; hence, variation in the known value $\langle O_j \rangle$ cannot affect the quality of restoration.

The smaller σ_{Oj} is, the more specific is the user's knowledge about the unknown object [see Eq. (7b)]. Eq. (51b) shows that $\varepsilon_j^{(2)}$ monotonically decreases as σ_{Oj} decreases.

The greater M is, the more information about joint object distribution $p_O(O_j, \ldots, O_{j+M})$ is required. Since all terms in Eq. (51b) are positive, $\varepsilon_j^{(2)}$ monotonically decreases as M increases.

In summary, when optimally processing a white object region, the processing quality monotonically increases with the user's foreknowledge of the object statistics.

Weiner filter limit--We examine the particular case M = 0 of a "disjoint" object. Eqs. (51) now describe optimal ms processing under conditions of linear image formation, Gaussian object and noise distributions, and a non-statistical transfer function. Because M = 0, the object is no longer "white."

Under these circumstances, and in the particular case of $\langle O_j \rangle = 0$, the optimum, ms, linear filter has been shown[3] to resemble the classical Weiner filter.[12] The identical result follows from Eqs. (51), furnishing a corroboration of Eqs. (51). It should be added that if $\langle O_j \rangle \neq 0$, Eq. (51a) (with M = 0) must replace the Weiner formula as the optimal ms processor.

<u>Numerical simulation</u>

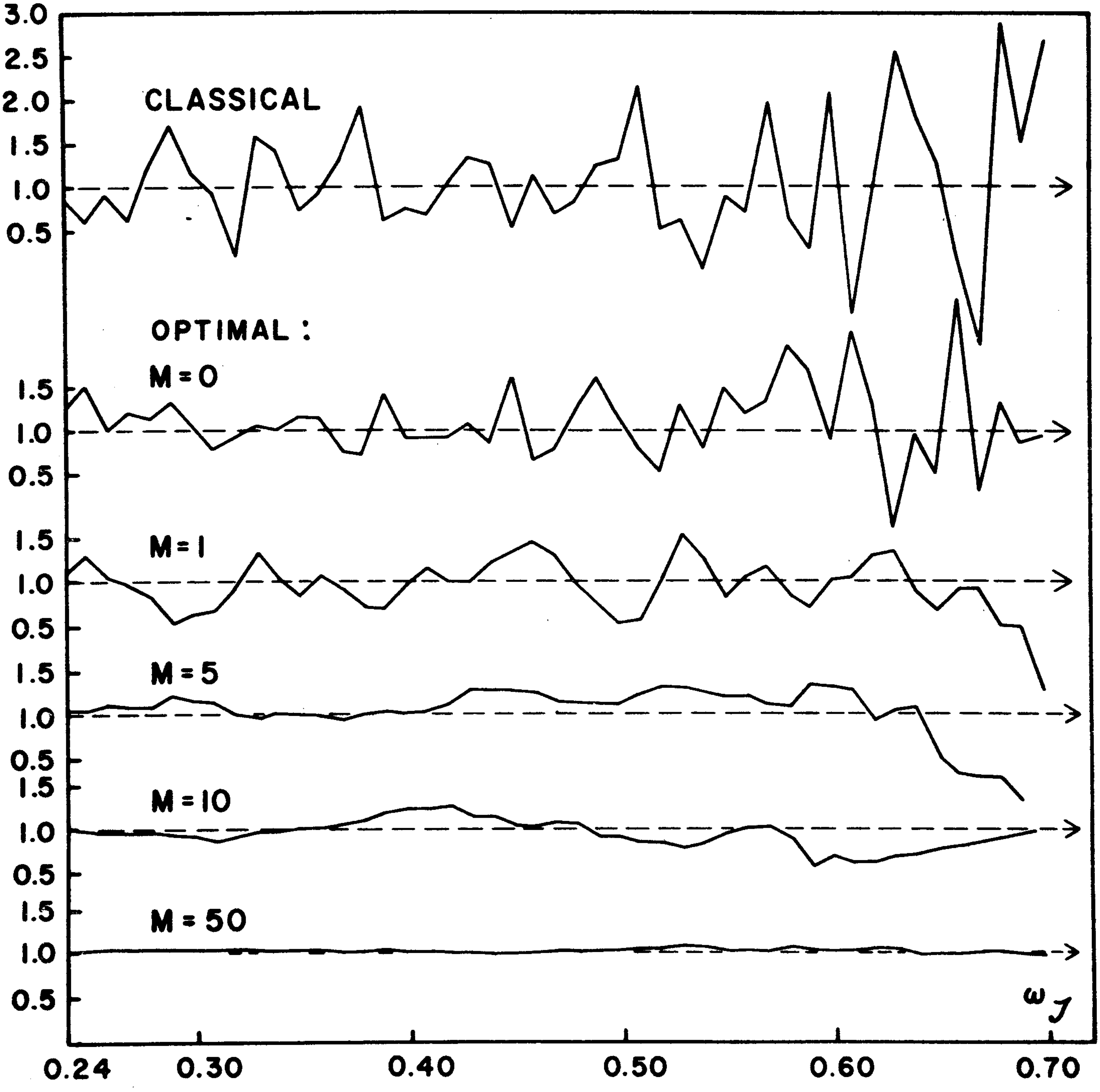
Because processor (51a) is an analytical function of observables $\vec{I}$, it is convenient to use in a numerical simulation of optimal processing. Simulation was effected upon a CDC 6400 electronic computer.

Image values were generated from law (14a), with parameters N_j, O_j, and τ_j determined in the following way. Values N_j were made to be independently, Gaussian random by use of a (uniformly) random number generator and an auxiliary subroutine[31] for conversion to Gaussian randomness. For easy interpretation of the resulting restorations, one unknown object value (of unity) was assumed to extend over all frequencies involved. Finally, the known transfer function was taken as

$$\tau(\vec{\omega}_j) = \begin{cases} \pi^{-1}\, 2\, [\cos^{-1}(\omega_j) - \omega_j(1-\omega_j{}^2)^{\frac{1}{2}}], & |\omega_j| \leq 1 \\ 0 \text{ for } |\omega_j| > 1 \end{cases} \qquad (52)$$

which is due to circular, diffraction-limited optics.[19]

Each set of noisy image values was independently processed according to classical law (16), and optimal ms law (51a) with successive values M = 0, 1, 5, 10, and 50. Results are shown in the figure below for one typical set of image values.



Classical vs. optimal processing in the case of a constant, or "white," object spectrum over M + 1 adjacent frequencies. A noisy image was generated upon an electronic computer according to law (14a) for linear image formation with additive noise. The latter was made to be Gaussian random by use of the computer's random number generator and an associated subroutine. The noisy image was independently processed by classical formula (16), and by optimal formula (51a) with successive values M = 0, 1, 5, 10 and 50. The actual object, $O_j = 1.0$, is plotted as a dashed line for comparison with each restoration. Graphical indications are that (a) optimal processing offers quantitative improvement over classical processing, and (b) optimal processing improves as the "object foreknowledge" (in this case, M) increases. The processor for M = 0 is a close relative of the Weiner filter.

All restorations are over the frequency band $0.24 \le \omega_j \le 0.70$ (normalized relative to the optical cutoff). Numerical values $\sigma_{Oj} = 1.0$, $\langle O_j \rangle = 1.4$ and $\sigma_N = 0.2$ are taken. The actual object, $O_j = 1.0$, is plotted as a dashed line for comparison with each restoration.

The top graph shows the classical restoration resulting from use of processor (16) at each ω_j. The quality of resotration is observed to be quite poor, especially at higher frequencies. By combining Eqs. (14a) and (16) we can see why:

$$\hat{O}_j = O_j + N_j/\tau_j \quad . \tag{53}$$

As $\omega_j \to 1$ (cutoff), $\tau_j \to 0$. Hence, the noise contribution to each $\hat{O}_j$ grows uncontrollably as $\omega_j \to 1$.

The remaining five graphs, identified by values M, show the virtue of processing according to the optimal processor (51a), and in particular, the effect of increasing foreknowledge M upon the quality of the optimal restoration. As M increases, these curves gradually flatten and smooth toward the actual object $O_j = 1$. This is also the tendency predicted by error formula (51b).

SUMMARY

Digital, as compared to analog, processing allows a noisy image to be manipulated in arbitrary fashion. Hence, it is possible to digitally process the the image in the generally non-linear manner required by a criterion of *optimal* processing, for example.

One such criterion is (19). This criterion demands that, at each spatial frequency, the expected mean power error between restoration and original object be a minimum. To attain this goal, the optimum processor at any frequency $\vec{\omega}_j$ is allowed to be an arbitrary function of *all* parameters which can possibly influence the detected image at $\vec{\omega}_j$. These parameters are the object O, noise N, and image-forming characteristic τ, at each of $\vec{\omega}_j$, $\vec{\omega}_{j(k)}$, k = 1, ..., M. It is tacitly assumed that the joint probability distributions P_O, P_N and P_τ over these $\vec{\omega}$ are known to the experimenter. It is this "foreknowledge," i.e. knowledge aside from observation of the image, which allows the processing to be optimized. Examples p_O are discussed between Eqs. (4) and (12). One example of p_τ is discussed at Eq. (15).

In order for criterion (19) to be enforced, it is necessary to make an assumption regarding the law of image formation, i.e. the dependence of detected image I_j upon parameters τ, O and N. It is a fortunate aspect of the optimization problem that a nearly general law $L(\tau_j, O_j, N_j)$ of image formation (13b) can be assumed throughout. Particular forms of L need only be substituted into the *solution* for the optimal processor. Hence, the solution holds for a wide class of image-forming situations, in particular those for which L is as yet unknown. Some simple laws L are discussed at Eqs. (14).

In Eqs. (20), criterion (19) is cast in terms of the given probability distributions. The optimum processor is found by treating its definition, Eqs. (20), as an extremum problem. For any sequence of observed $\vec{I}$, the optimum

processor is found to be a root of polynomial Eq. (27). Since this polynomial is of finite degree, Eq. (27) is relatively easy to solve by numerical methods.

When the optimum mean *square* processor is desired, the solution to Eq. (27) is known analytically as processor (28). Also, the expected error resulting from use of processor (28) is known analytically as Eq. (29).

The effect of processing in the frequency domain upon the restored *spatial* object is studied. A central result is that the mean square error of restoration over the spatial object is a minimum $E_m^{(2)}$ when optimum mean square processing $\bar{O}_j$ in the *frequency* domain is employed. This result justifies the use of processors $\bar{O}_j$ in practice.

It is found that error $E_m^{(2)}$ may be further minimized through choice of the processing bandwidth. Eq. (39) yields the optimum bandwidth Ω for this purpose. In the special case of classical image processing (16), one-dimensional, diffraction-limited optics, and constant power spectra for noise and object, the simple solution (42) results.

The remainder of the paper is devoted to special uses of the optimum ms processor (28). In all cases, a law (14a) of linear image formation is assumed for mathematical convenience.

In these cases the most general solution is processor (30b). When, in addition, the transfer function is known with sufficient accuracy to be regarded as non-statistical, the solution is processor (31b).

Solution (31b) is applied to the image recognition problem. Because the possible object values are now discrete and finite at each $\vec{\omega}_j$, the solution is processor (45). Some interesting properties of this processor are discussed. Processor (45) is used to form a "decision function" (49a) for establishing which object (or image) is being detected. An "optimal" decision function is defined, but not explicitly found. Comparison is made with a "classical"[1] decision function (50), in the case of a "binary" object.

Finally, solution (31b) is applied to the case where M + 1 neighboring frequency points are known to share a *common* object value, defining the "white" object region. The general solution is processor (51a), with expected error (51b). The effect of various values of M upon the quality of typical restorations (51a) is graphically shown in the figure. (The case $M = 0$ is essentially that of Weiner filtering.) These restorations are of noisy image data generated by use of an electronic computer. The virtue of using as many adjacent image values I_{j+k}, $k = 0, \pm 1, \ldots, \{M'\}$ as is possible to restore each *single* object value O_j is evident in these plots.

REFERENCES

1. James L. Harris, "Resolving power and decision theory," J. Opt. Soc. Am 54(5):606-611, May 1964.

2. C. K. Rushforth, "Restoration of optical patterns" (abstr.), J. Opt. Soc. Am. 55(5):600-601, WF14, May 1965.

3. Carl W. Helstrom, "Image restoration by method of least squares," J. Opt. Soc. Am. 57(3):297-303, March 1967.

4. Robert Nathan, "Digital video-data handling," J. Opt. Soc. Am 57(4):575-578, FA2, April 1967.

5. H. Wolter in Progress in Optics, vol. I. Emil Wolf, Ed. (Amsterdam, North-Holland Publ. Co.; New York, Interscience Publishers, 1961).

6. James L. Harris, "Diffraction and resolving power," J. Opt. Soc. Am. 54(7): 931-936, July 1964.

7. Casper W. Barnes, "Object restoration in a diffraction-limited imaging system," J. Opt. Soc. Am. 56(5):575-578, May 1966.

8. G. J. Buck, J. J. Gustincic, IEEE Trans. on Antennas and Propagation AP-15:376, 1967.

9. B. Roy Frieden, "Band-unlimited reconstruction of optical objects and spectra," J. Opt. Soc. Am. 57(8):1013-1019, 1967.

10. The preceding is a representative, rather than a complete, list of the published research on this problem. Also, a great amount of the work in this field is either proprietary or classified, and hence unpublished. Related work up to 1960 by the electrical engineers is referenced by David Middleton, An Introduction to Statistical Communication Theory (McGraw-Hill, New York, 1960), pp. 1104-1109.

11. The processed image is, essentially, a transformation of statistical coordinates. Information remains invariant under such a transformation. See, e.g., Stanford Goldman, Information Theory (Prentice-Hall, Inc., New York, 1955), pp. 152-153.

12. Wilbur B. Davenport, Jr., and William L. Root, Random Signals and Noise (McGraw-Hill, New York, 1958).

13. Ref. 12, pp. 251, 280.

14. Ronald N. Bracewell, The Fourier Transform and Its Applications (McGraw-Hill, New York, 1965), p. 7.

15. Or, "zero-memory" in electrical engineering parlance. Eq. (10) is discussed in, e.g., Emanuel Parzen, Modern Probability Theory and Its Applications (John Wiley & Sons, New York, 1960).

16. Ref. 14, Ch. 5.

17. Ref. 11, pp. 114-116.

18. Ref. 10, p. 45.

19. Edward L. O'Neill, Introduction to Statistical Optics (Addison-Wesley Publ. Co., Reading, Mass., 1963).

20. This and other examples are discussed by P. B. Felgett and Edward H. Linfoot, Proc. Roy. Soc. A., 247:393, 1955.

21. Edward H. Linfoot, Fourier Methods in Optical Image Evaluation (Focal Press, London, 1964), p. 43.

22. See, e.g., Robert R. Shannon and A. H. Newman, "An instrument for measurement of the optical transfer function," Appl. Opt. 2(4):365-369, April 1963.

23. B. Roy Frieden, "Image evaluation by use of the sampling theorem," J. Opt. Soc. Am. 56(10):1355-1362, Oct. 1966.

24. Ref. 21, p. 45.

25. The Chebyshev criterion requires that the maximum error be a minimum. See, e.g., Richard Barakat, J. Opt. Soc. Am. 52(9):985-991, Sept. 1962.

26. Ref. 11, p. 135.

27. Robert R. Stoll, Linear Algebra and Matrix Theory (McGraw-Hill, New York, 1952), p. 101.

28. Eq. (26a) is a Euler-Lagrange problem where the integrand is not a function of the derivatives $\partial \bar{O}_j(\vec{I})/\partial I_{j-k}$, $\partial \bar{O}_j^*(\vec{I})/\partial I_{j-k}$, $k = 0, 1, \ldots, M$. The Euler-Lagrange equation then simplifies to the condition that the integrand be a minimum at each $\vec{I}$. The general problem is treated, e.g., by Harry Lass, Elements of Pure and Applied Mathematics (McGraw-Hill, New York, 1957), p. 291.

29. See, e.g., John M. Wozencraft and Irwin M. Jacobs, Principles of Communication Engineering (John Wiley & Sons, Inc., New York, 1965), p. 585.

30. A measure of randomness is the "entropy." The latter is maximized by a constant probability distribution. See C. E. Shannon, Bell Syst. Tech. J. 27:629, 1948.

31. This subroutine is based upon result 10.5 in Ref. 15, p. 334.

jal