NASA-CR 122307

TWP

The Application of Signal Detection Theory to Optics

SEMI-ANNUAL STATUS REPORT

NASA Grant NGL 05-009-079

March 15, 1971--October 15, 1971

Carl W. Helstrom

Department of Applied Physics and Information Science

University of California, San Diego

La Jolla, California 92037

ABSTRACT

The restoration of images focused on a photosensitive surface is treated from the standpoint of maximum-likelihood estimation, taking into account the Poisson distributions of the observed data, which are the numbers of photoelectrons from various elements of the surface. A detector of an image focused on such a surface utilizes a certain linear combination of those numbers as the optimum detection statistic. Methods for calculating the false-alarm and detection probabilities are proposed. It is shown that measuring noncommuting observables in an ideal quantum receiver cannot yield a lower Bayes cost than that attainable by a system measuring only commuting observables.

N72-12658 (NASA-CR-122307) THE APPLICATION OF SIGNAL DETECTION THEORY TO OPTICS Semiannual Status Report, 15 Mar. - 15 Oct. 1971 C.W. Helstrom (California Univ.) Oct. 1971 50 p CSCL 20F G3/23

Unclas 10082

Restoration of Images on Photosensitive Surfaces

Methods for restoring images degraded by an optical system or by passage of the image-forming light through a turbulent medium often are based on a linear model of the optical process, in which the noise is independent of the desired image and combines with it additively. In these models it is necessary to postulate the statistical properties of the noise, which is usually assumed Gaussian, and to leave such parameters as its mean-square amplitude and its bandwidth to be measured separately. Such a treatment is incapable of determining the fundamental limitations on restoring degraded images.

Essaying to take into account the physical nature of the light and the statistical properties of the recording process, we have analyzed an imaging system in which the light from the object plane is focused on a photosensitive surface, from which it ejects photoelectrons. The surface is divided like a mosaic into a large number of small, insulated spots, from each of which the photocurrent can be measured. These measured values of the photocurrent constitute the data on which is to be based an estimate of the radiance of the object plane, which corresponds to the "true image". The numbers of photoelectrons emitted from the spots have Poisson distributions whose mean values are proportional to the instantaneous illuminance, which is the absolute square of the sum of the light fields created by object and background. These fields are spatio-temporal Gaussian random processes. The observed photocurrents are combined linearly to produce estimates of the object radiance, and the optimum linear combination determines the minimum mean-square error attainable by the restored image.

In a paper, "Linear Restoration of Incoherently Radiating Objects", attached to this report, it is shown that the formula for the minimum mean-square error has the same form as that obtained in Wiener filtering theory, but with the noise terms specified by the physical properties of the light fields and of the photosensitive surface. In particular, it is shown that under normal circumstances the shot noise arising from the stochastic nature of the photoelectron emissions far exceeds the noise associated with the random fluctuations of the object and background light fields. Image degradation by diffraction at the aperture and by passage of the object light through a random phase screen, representing a turbulent medium, are both analyzed.

As the shot noise predominates in causing random variations of the observed data, our knowledge of its properties can be made the basis of a technique for restoring images recorded by means of such a photosensitive mosaic. In developing the method, we describe the illuminance $I(\underset{\sim}{x})$ at point $\underset{\sim}{x}$ of the image plane as the convolution of a true or geometrical image $J(\underset{\sim}{x})$ with an incoherent point-spread function (psf),

$$I(\underset{\sim}{x}) = \int K(\underset{\sim}{x} - \underset{\sim}{y})\, J(\underset{\sim}{y})\, d^2\underset{\sim}{y}. \tag{1}$$

The psf will be assumed normalized so that

$$\int K(\underset{\sim}{x})\, d^2\underset{\sim}{x} = 1. \tag{2}$$

All integrals are carried out over the image plane, which will be taken finite for computational purposes. It is $J(\underset{\sim}{x})$ that we wish to estimate.

The image plane is a photosensitive surface divided like a mosaic into a large number of small spots of area s, which emit photoelectrons when the light impinges on them. The product WT of the bandwidth W of the incident light and the observation time T is supposed to be so large, $WT >> 1$, that the numbers n_i of photoelectrons from the spots are independent random variables with Poisson distributions. The mean value of n_i is given by

$$\underset{\sim}{E}(n_i) = \alpha\, s\, I(\underset{\sim}{x}_i), \tag{3}$$

where $\underset{\sim}{x}_i$ is the center of the i-th spot and α is a constant proportional to the quantum efficiency of the surface and to the observation time and inversely proportional to the quantum $h\nu$ of light energy. (More generally, $\alpha\ I(\underset{\sim}{x})$ and $\alpha\ J(\underset{\sim}{x})$ are averages over the spectrum of the incident light, weighted by $\eta(\nu)/h\nu$, where $\eta(\nu)$ is the quantum efficiency at frequency ν, the spectrum being assumed independent of $\underset{\sim}{x}$.)

Let $I_0 = J_0$ be the uniform illuminance of the image plane in the absence of a scene; we then put

$$J(\underset{\sim}{x}) = J_0 + j(\underset{\sim}{x}), \tag{4}$$

suppose J_0 known, and seek an estimate $\hat{j}(\underset{\sim}{x})$ of the difference $J(\underset{\sim}{x}) - J_0$. The likelihood ratio of the numbers n_i, i.e., the quotient of their joint probabilities in the presence and absence of an image, is then

$$\Lambda\{n_i\} = \prod_i \left[\frac{I(\underset{\sim}{x}_i)}{I_0}\right]^{n_i} \exp\left\{- \alpha\ s \sum_i [I(\underset{\sim}{x}_i) - I_0]\right\}, \tag{5}$$

with

$$I(\underset{\sim}{x}) - I_0 = \int K(\underset{\sim}{x} - \underset{\sim}{y})\ j(\underset{\sim}{y})\ d^2\underset{\sim}{y}. \tag{6}$$

We seek estimates of $j(\underset{\sim}{x})$ at sampling points $\underset{\sim}{x}_i$, and put $j(\underset{\sim}{x}_i) = \xi_i$, $I(\underset{\sim}{x}_i) = \eta_i + I_0$, approximating eq. (6) as

$$\eta_i = \sum_j K_{ij}\ \xi_j, \qquad K_{ij} = K(\underset{\sim}{x}_i - \underset{\sim}{x}_j)\ s. \tag{7}$$

We assume that the true image $J(\underset{\sim}{x})$ is a homogeneous spatial Gaussian random process with mean value J_0 and covariance

$$\underset{\sim}{E}[j(\underset{\sim}{x}_1)\ j(\underset{\sim}{x}_2)] = \varphi(\underset{\sim}{x}_1 - \underset{\sim}{x}_2), \tag{8}$$

and we form the matrix $\underset{\sim}{\varphi}$ whose elements are

$$\varphi_{ij} = \underset{\sim}{E}[\xi_i\ \xi_j] = \varphi(\underset{\sim}{x}_i - \underset{\sim}{x}_j), \tag{9}$$

and whose inverse is $\underset{\sim}{\mu} = \|\mu_{ij}\|$. The joint probability density function of the estimanda ξ_i is then

$$p(\{\xi_i\}) = M \exp\left(-\frac{1}{2}\sum_i\sum_j \mu_{ij}\, \xi_i\, \xi_j\right) \tag{10}$$

where M is a normalizing constant.

The maximum-likelihood estimate of the true image is given by the set of values ξ_i for which $\Lambda\{n_i\}\; p(\{\xi_i\})$ is maximum. Equivalently, we maximize the logarithm

$$\ln\,[\Lambda\{n_i\}\; p(\{\xi_i\})] = F(\{\xi_i\}) =$$

$$\sum_i [n_i\, \ln(1 + \frac{\eta_i}{I_0}) - \alpha s \eta_i] + \ln M - \frac{1}{2}\sum_i\sum_j \mu_{ij}\, \xi_i\, \xi_j, \tag{11}$$

where the η_i are expressed in terms of the ξ_j by eq. (7). Differentiating with respect to ξ_j, we get

$$\sum_i \left[\frac{n_i\, K_{ij}}{I_0 + \eta_i} - \alpha\, s\, K_{ij}\right] - \sum_m \mu_{jm}\, \xi_m = 0\,. \tag{12}$$

Solving these equations for ξ_m, we obtain

$$\begin{aligned}\xi_m &= \sum_j\sum_i \varphi_{mj}\, K_{ij} \left[\frac{n_i}{I_0 + \eta_i} - \alpha\, s\right]\\ &= \sum_i K'_{mi} \left[\frac{n_i}{I_0 + \eta_i} - \alpha\, s\right],\end{aligned} \tag{13}$$

where

$$K'_{mi} = \sum_j \varphi_{mj}\, K_{ij}. \tag{14}$$

If we define

$$K'(\underset{\sim}{x} - \underset{\sim}{y}) = \int \varphi(\underset{\sim}{x} - \underset{\sim}{u})\; K(\underset{\sim}{u} - \underset{\sim}{y})\; d^2\underset{\sim}{u} \tag{15}$$

and if we suppose the data n_i are given by $n_i = z_i\, s$, where the z_i can be considered as continuous random variables, $z_i = z(\underset{\sim}{x}_i)$, with conditional mean

$$\underset{\sim}{E}(z_i) = \alpha\, I(\underset{\sim}{x}_i), \tag{16}$$

we can convert eq. (13) into an integral equation,

$$\hat{j}(\underset{\sim}{x}) = \int K'(\underset{\sim}{x} - \underset{\sim}{y}) \left[\frac{z(\underset{\sim}{y})}{\hat{I}(\underset{\sim}{y})} - \alpha\right] d^2\underset{\sim}{y}, \tag{17}$$

$$\hat{I}(\underset{\sim}{y}) = \int K(\underset{\sim}{y} - \underset{\sim}{w})\ \hat{j}(\underset{\sim}{w})\ d^2\underset{\sim}{w} + I_0. \tag{18}$$

We can consider eq. (13) as a sampled version of eq. (17), with η_i given in terms of ξ_j by eq. (7), which is the sampled version of eq. (18). It is then no longer necessary for the n_i's to be integers. Equations (13) and (7) can be solved by iteration.

The function $F(\{\xi_i\})$ in eq. (11) may have many minima, and there is a danger that the iteration method may settle into the wrong one. It is therefore wise to start it as close to the absolute minimum as possible. One convenient starting point is the set of η_i obtained from the linear least-squares estimate of the ξ_j's. Designating these by circumflexes, we put

$$\hat{\eta}_i = \sum_j K_{ij}\ \hat{\xi}_j, \tag{19}$$

where

$$\hat{\xi}_i = \sum_j L_{ij}\ (n_j - \alpha s\ I_0)/\alpha s \tag{20}$$

with L_{ij} the least-squares estimating filter obtained by solving the Wiener filtering equations*

$$\sum_m K_{jm}\ \varphi_{mi} = \sum_m L_{im}\ \psi_{mj}, \tag{21}$$

$$\psi_{mj} = \sum_{k,n} K_{mk}\ K_{jn}\ \varphi_{kn} + (I_0/\alpha s)\delta_{mj}. \tag{22}$$

---

* C. W. Helstrom, "Image Restoration by the Method of Least Squares", J. Opt. Soc. Am. vol. 57, 297-303 (March, 1967).

The last term in eq. (22) represents the noise, which is assigned a mean-square value equal to the variance of the numbers n_i when the illuminance is I_0. The least-squares estimate will be most nearly optimum when the contrast is low, so that $I(\underset{\sim}{x}) \cong I_0$, and when the numbers n_i of counts from each element of the surface are large enough, $n_i >> 1$, so that the Poisson distribution can be approximated by a Gaussian with mean and variance given by eq. (3). Since the matrices K_{ij}, φ_{ij}, and L_{ij} have the Toeplitz form, finite Fourier transforms much simplify the computations involved in applying eqs. (19) - (22).

In the iterative solution negative values of $J_0 + \xi_m$ may be obtained. It is suggested that as these are unphysical, they be replaced by zero wherever they occur. There will then be no danger that $\hat{I}(\underset{\sim}{y})$ or $I_0 + \eta_i$ will vanish, since the kernel $K(\underset{\sim}{u})$ and the matrix elements k_{ij} are non-negative. If necessary, $J_0 + \xi_m$ could be constrained in the iteration to be at least equal to a positive number J_n, representing a uniform background illuminance.

False-Alarm and Detection Probabilities

of Photoelectric Images

1. False-Alarm Probability by the Method of Steepest Descent

For the detection of an optical image providing an illuminance $M_s(\underset{\sim}{x})$ at point $\underset{\sim}{x}$ of a photosensitive surface, in the presence of a background illuminance $M_0(\underset{\sim}{x})$, the optimum statistic has been shown to be*

$$g = \sum_i \ln \left[\frac{M_1(\underset{\sim}{x}_i)}{M_0(\underset{\sim}{x}_i)}\right] - \tau \int M_s(\underset{\sim}{x})\, d^2\underset{\sim}{x}, \tag{1.1}$$

$$M_1(\underset{\sim}{x}) = M_0(\underset{\sim}{x}) + M_s(\underset{\sim}{x}),$$

where $\underset{\sim}{x}_i$ is the point from which the i-th photoelectron is emitted and τ is a constant proportional to the quantum efficiency of the surface and such that the expected number n of photoelectrons during the observation interval is, in the presence of the image $M_s(\underset{\sim}{x})$,

$$\underset{\sim}{E}[n|H_1] = \tau \int M_1(\underset{\sim}{x})\, d^2\underset{\sim}{x}. \tag{1.2}$$

Integrals are taken over the entire photosensitive surface, and the sum in eq. (1.1) is taken over all the emitted photoelectrons.

Here H_0 will denote the hypothesis that the image is absent, background light alone being present; H_1 denotes the hypothesis that the image sought is present. The number n_i of photoelectrons from any small area δA at point $\underset{\sim}{x}_i$ of the surface has a Poisson distribution with mean values

$$\underset{\sim}{E}[n_i|H_0] = \tau M_0(\underset{\sim}{x}_i)\, \delta A, \tag{1.3}$$

$$\underset{\sim}{E}[n_i|H_1] = \tau M_1(\underset{\sim}{x}_i)\, \delta A. \tag{1.4}$$

The statistic g is compared with a decision level g_0; if $g > g_0$, the decision

* C. W. Helstrom, "The Detection and Resolution of Optical Signals", Trans. IEEE, vol. IT-10, 275-287 (October, 1964).

for hypothesis H_1--image present--is made. More generally $M_j(\underset{\sim}{x})$, $j = 0, 1$, is an average of the illuminance $J_j(\underset{\sim}{x}, \nu)$ as a function of frequency ν, weighted in accordance with the quantum efficiency $\eta(\nu)$ of the photosensitive surface,

$$M_j(\underset{\sim}{x}) = \int \eta(\nu)\, J_j(\underset{\sim}{x}, \nu)\, d\nu/h\nu, \quad j = 0, 1,$$

where h is Planck's constant. Then τ is the observation time.

The false-alarm probability Q_0 is

$$Q_0 = \Pr\{g > g_0 | H_0\}, \tag{1.5}$$

and the detection probability Q_d is

$$Q_d = \Pr\{g > g_0 | H_1\}. \tag{1.6}$$

The probability distributions of g under the two hypotheses are generally difficult to calculate, and we shall endeavor to determine Q_0 and Q_d from the Laplace transforms (moment-generating functions)

$$f_j(s) = \underset{\sim}{E}[e^{-sg} | H_j], \quad j = 0, 1, \tag{1.7}$$

of the distributions, which are given by

$$\begin{aligned} f_j(s) &= \exp[\mu_j(s)] \\ &= \exp\left\{\tau \int M_j(\underset{\sim}{x}) \left\{ \left[\frac{M_1(\underset{\sim}{x})}{M_0(\underset{\sim}{x})}\right]^{-s} - 1 \right\} d^2\underset{\sim}{x} + \tau\, s \int M_s(\underset{\sim}{x})\, d^2\underset{\sim}{x} \right\}, \\ &\qquad j = 0, 1; \end{aligned} \tag{1.8}$$

these are related by

$$f_1(s) = f_0(s - 1). \tag{1.9}$$

The false-alarm and detection probabilities are then, by the inversion formula for Laplace transforms,

$$Q_0(g_0) = \int_{c-i\infty}^{c+i\infty} \frac{1 - f_0(s)}{s}\, e^{sg_0}\, ds/2\pi i \tag{1.10}$$

and

$$Q_d(g_0) = \int_{c-i\infty}^{c+i\infty} \frac{1 - f_1(s)}{s} e^{sg_0} ds/2\pi i. \tag{1.11}$$

The contour of integration is a straight line parallel to and to the right of the imaginary axis in the complex s-plane. We concentrate here on the false-alarm probability; the technique for small values of the detection probability Q_d is quite similar.

Since the integrand of eq. (1.10) has no singularity at the origin--or indeed anywhere in the complex plane, as an analysis of eq. (1.8) shows--we can displace the contour to the left across the imaginary axis. Taking only the first term,

$$\int_C s^{-1} e^{sg_0} ds/2\pi i,$$

we can complete the contour around the left half-plane and show that this integral vanishes. Hence the false-alarm probability is $Q_0 = q(g_0)$, where

$$q(g) = -\int_C s^{-1} f_0(s) e^{gs} ds/2\pi i, \tag{1.12}$$

the contour C lying parallel to and to the left of the imaginary axis. We can write this as

$$\begin{aligned} q(g) &= -\int_C \exp[\mu_0(s) + gs - \ln s] ds/2\pi i \\ &= \int_C \exp \Phi(s) ds/2\pi i, \end{aligned} \tag{1.13}$$

where

$$\Phi(s) = \mu_0(s) + gs - \ln(-s) \tag{1.14}$$

is a complex phase.

We apply the method of steepest descent by looking for the point $s = s_0$ at which the phase $\Phi(s)$ is stationary, that is, where

$$\Phi'(s) = \mu_0'(s) + g - s^{-1} = 0, \quad s = s_0, \tag{1.15}$$

denoting by primes differentiations with respect to s. The contour C is shifted until it passes vertically through s_0. Expanding the phase $\Phi(s)$ in a power series about s_0, we write the integrand in eq. (1.13) as

$$\begin{aligned} \exp \Phi(s) &= \exp\left[\Phi(s_0) + \frac{1}{2}(s - s_0)^2 \Phi''(s_0) + \sum_{r=3}^{\infty} (r!)^{-1} (s - s_0)^r \Phi^{(r)}(s_0)\right] \\ &= \exp[\Phi(s_0) + \frac{1}{2}(s - s_0)^2 \Phi''(s_0)] \\ &\quad \times \left[1 + \sum_{r=3}^{\infty} (r!)^{-1} c_r (s - s_0)^r\right], \end{aligned} \tag{1.16}$$

where $\Phi^{(r)}(s) = d^r \Phi(s)/ds^r$, and where the coefficients c_r are derived from the series expansion of the exponential function,

$$\exp\left[\sum_{r=3}^{\infty} (r!)^{-1} (s - s_0)^r \Phi^{(r)}(s_0)\right] = 1 + \sum_{r=3}^{\infty} (r!)^{-1} c_r (s - s_0)^r. \tag{1.17}$$

Now we put for our variable of integration $s = s_0 + iy$, and integrating term by term on y we obtain

$$\begin{aligned} q(g) &= -[2\pi \Phi''(s_0)]^{-\frac{1}{2}} \exp \Phi(s_0) \\ &\quad \times \left[1 + \sum_{k=2}^{\infty} \frac{(-1)^k}{2^k k!} \frac{c_{2k}}{[\Phi''(s_0)]^k}\right]. \end{aligned} \tag{1.18}$$

In general, the series in brackets has only asymptotic validity and must be cut off when the terms, which at first decrease, begin to increase. The second derivative appearing here is, by eq. (1.15),

$$\Phi''(s_0) = \mu_0''(s_0) + s_0^{-2}. \tag{1.19}$$

A computer program to generate the coefficients c_r in eq. (1.17) from these derivatives is simple to write. If in general we wish to take the exponential function of a power series,

$$\sum_{r=0}^{\infty} (r!)^{-1}\, c_r\, x^r \;=\; \exp\left(\sum_{s=0}^{\infty} (s!)^{-1}\, b_s\, x^s\right),$$

we can find the c_r's by the recurrence relation

$$c_r \;=\; \sum_{s=0}^{r-1} \binom{r-1}{s}\, b_{r-s}\, c_s, \quad c_0 \;=\; \exp(b_0),$$

which comes from Euler's formula for the (r-1)-th derivative of $g'(x) \exp g(x)$, where $g(x)$ is the power series in the exponential function. Here $b_0 = b_1 = b_2 = 0$, $b_r = \Phi^{(r)}(s_0)$, $r \geq 3$.

The derivative $\mu_0^{(k)}(s)$ of the logarithm of the characteristic function appearing in $\Phi^{(k)}(s_0)$ can be written as

$$\mu_0^{(k)}(s) \;=\; d^k\, \mu_0(s)/ds^k \;=\; (-1)^k\, \tau \int M_0(\underset{\sim}{x}) \left[\ell n\left(\frac{M_1(\underset{\sim}{x})}{M_0(\underset{\sim}{x})}\right)\right]^k \left[\frac{M_1(\underset{\sim}{x})}{M_0(\underset{\sim}{x})}\right]^{-s} d^2\underset{\sim}{x}. \tag{1.20}$$

They will usually have to be integrated numerically, and a computer routine for generating all the necessary ones simultaneously will be expedient.

The approximation of the false-alarm probability Q_0 derived here is closely related to the Chernoff bound, which is based on the inequality

$$\exp s_1(g - g_0) \;\geq\; U(g - g_0), \quad s_1 > 0, \tag{1.21}$$

where U is the unit step function.* Thus

$$Q_0 \;=\; \int_{-\infty}^{\infty} U(g - g_0)\, p_0(g)\, dg \;\leq\;$$
$$\int_{-\infty}^{\infty} \exp s_1(g - g_0)\, p_0(g)\, dg \;=\; e^{-s_1 g_0}\, f_0(-s_1). \tag{1.22}$$

---

* H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations", Ann. Math. Stat. vol. 23, 493-507 (December, 1952).

The value of s_1 is then picked to minimize the bounding function. Putting $s = -s_1$, we must minimize

$$\ln f_0(s) + sg_0 = \mu_0(s) + sg_0;$$

this requires

$$\mu_0'(s) + g_0 = 0, \tag{1.23}$$

which except for the term $-s^{-1}$ is the same as eq. (1.15). If $s_0' < 0$ is the solution of eq. (1.23), the Chernoff bound asserts

$$Q_0(g_0) \leq \exp[\mu_0(s_0') + s_0'g_0], \tag{1.24}$$

and the exponent here is nearly equal to $\Phi(s_0')$. In fact, since eq. (1.22) gives a bound for all positive values of s_1, we can use instead of s_0' the saddlepoint s_0 obtained from eq. (1.15) and assert the upper bound

$$Q_0(g) < \exp[\mu_0(s_0) + s_0 g], \tag{1.25}$$

which will be nearly as tight as that in eq. (1.24). The method of steepest descent thus yields both an approximation and a bound to the false-alarm probability.

In determining the detectability of optical images, one fixes a false-alarm probability Q_0 and from it determines the decision level g_0, which is then used in calculating the detection probability $Q_d(g_0)$. It is therefore necessary to solve the pair of equations (1.15) and (1.18) for the value of g for which q(g) equals the pre-assigned false-alarm probability Q_0. Here Newton's method is most expedient. The function q(g) can, through eq. (1.15), be treated as a function of s_0, which we designate by $\bar{q}(s_0)$. Then if s_0^1 is a trial value of s_0, a new trial value is determined by the Newton equation

$$s_0 = s_0^1 + \frac{Q_0 - \bar{q}(s_0^1)}{\bar{q}'(s_0^1)}, \tag{1.26}$$

where the prime denotes differentiation with respect to s_0. In computing the derivative $\bar{q}'(s_0)$, the series in the bracket in eq. (1.18) can be replaced by 1, and the factor in front of the exponential function can be treated as constant; the rate of convergence to the solution will not be much affected. Thus we can use

$$\bar{q}'(s_0) \quad = \quad \Phi'(s_0)\ \bar{q}(s_0), \tag{1.27}$$

where in the differentiation of $\Phi(s)$ of eq. (1.14) g must be treated as a function of s_0,

$$\Phi'(s_0) \quad = \quad \mu_0'(s_0) + g + g's_0 - s_0^{-1} \quad = \quad g's_0 \tag{1.28}$$

by eq. (1.15). On the other hand, eq. (1.15) gives

$$\mu_0''(s_0) + g' + s_0^{-2} \quad = \quad 0. \tag{1.29}$$

Hence

$$\bar{q}'(s_0) \quad = \quad - \bar{q}(s_0)\ [s_0\mu_0''(s_0) + s_0^{-1}] \tag{1.30}$$

is to be used in the denominator of the Newton formula, eq. (1.26).

In the beginning of the iterative calculation of s_0, it will suffice to use for $\bar{q}(s_0)$ in eq. (1.26) the formula in eq. (1.18) with the bracketed series set equal to 1. At the end, as many terms of the series should be included as feasible. Rather than recalculating all the derivatives of $\Phi(s)$ at each new trial value of s_0 through eq. (1.20), it may be more expedient to determine them from the Taylor series

$$\Phi(s) \quad = \quad \sum_{r=0}^{\infty} (r!)^{-1}\ \Phi^{(r)}(s_0^1)\ (s - s_0^1)^r, \tag{1.31}$$

where s_0^1 is a suitable trial value at some stage of the calculation. The coefficients in the bracketed series in eq. (1.18) will not need to be accurately evaluated if the contribution of the variable terms of the series is small.

2. The Probability of Detection

For calculating the probability Q_d of detection, the method of steepest descent will be useful only for weak images, for which Q_d is not much larger than Q_0. The calculation will then be the same as what has just been described, except that $f_1(s) = f_0(s - 1)$ is used in place of $f_0(s)$. Eventually, as Q_d increases, the series in the bracket in eq. (1.18) will cease to converge properly, and this method must be abandoned.

For detection probabilities near 1, which are of most interest, an approximation method based on the analytical properties of $f_1(s)$ for large real values of s may be fruitful. In this analysis one may as well drop the constant last term

$$\gamma = \tau \int M_s(\underset{\sim}{x})\, d^2\underset{\sim}{x} \tag{2.1}$$

from the statistic g in eq. (1.1), using instead the modified statistic

$$g_1 = g + \gamma, \tag{2.2}$$

which is compared with the decision level $g_{10} = g_0 + \gamma$. The moment-generating function (m.g.f.) of g_1 is

$$f_{11}(s) = \underset{\sim}{E}[\exp(-sg_1)|H_1] = e^{-s\gamma} f_1(s).$$

The probability density function (p.d.f.) of g_1 has a delta function at the origin, for $g_1 = 0$ when no photoelectrons at all are emitted during the observation interval. The probability of this event is

$$P_1(0) = \exp\left[-\tau \int M_1(\underset{\sim}{x})\, d^2\underset{\sim}{x}\right] \tag{2.3}$$

under hypothesis H_1. The p.d.f. of g_1, therefore, has the form

$$p_1(g_1) = P_1(0)\,\delta(g_1) + p_1'(g_1). \tag{2.4}$$

The m.g.f. of the improper density function $p_1'(g)$ is

$$f_i(s) = \int_0^\infty p_1'(g)\, e^{-sg}\, dg =$$
$$P_1(0) \left\{ \exp\left[\tau \int M_0(\underset{\sim}{x}) \left[\frac{M_0(\underset{\sim}{x})}{M_1(\underset{\sim}{x})}\right]^{s-1} d^2\underset{\sim}{x} \right] - 1 \right\}. \tag{2.5}$$

The behavior of $p_1'(g)$ for small values of g, which are of principal concern when Q_d is near 1, is related to the behavior of $f_i(s)$ for large, real values of s. The detection probability is given by

$$Q_d = 1 - P_1(0) - \int_0^{g_{10}} p_1'(g)\, dg \tag{2.6}$$

and is therefore bounded above by $1 - P_1(0)$ independently of the decision level $g_{10} = g_0 + \gamma$.

Let us consider, for example, a uniform background $M_0(\underset{\sim}{x}) = M_0$ and a one-dimensional image of the form

$$M_s(x, y) = \begin{cases} A\, \varphi(x), & 0 < x < a, \\ 0\,, & x < 0,\ x > a, \end{cases} \tag{2.7}$$

with $\varphi(0) = 0$ and $\varphi(x)$ a monotone increasing function of x. The image covers a range of width b in the y-direction. Since it is the behavior of $M_s(\underset{\sim}{x})$ in the regions where it is nearly zero that matters in what follows, this form is sufficiently general for the time being.

The crucial term in eq. (2.5) is the integral in the exponent, which we write as

$$J(s) = \tau \int M_0(\underset{\sim}{x}) \left[\frac{M_0(\underset{\sim}{x})}{M_1(\underset{\sim}{x})}\right]^{s-1} d^2\underset{\sim}{x} = \tau\, b\, M_0 \int_0^a [1 + \lambda\, \varphi(x)]^{-(s-1)}\, dx,$$

$$\lambda = A/M_0. \tag{2.8}$$

We wish to determine the dependence of this integral, and hence also of $f_1(x)$, on s when $s >> 1$. Let us take as an example $\varphi(x) = x^\nu$, $\nu > 0$. Then we get

$$J(s) = \tau\, b\, M_0 \int_0^a (1 + \lambda\, x^\nu)^{-(s-1)}\, dx$$

$$\cong \tau\, b\, M_0 \int_0^\infty (1 + \lambda\, x^\nu)^{-(s-1)}\, dx; \tag{2.9}$$

the discrepancy introduced by extending the integral decreases exponentially with increasing s and can be neglected in this analysis. Evaluating the integral we get

$$J(s) \cong \tau\, b\, M_0\, \nu^{-1}\, \lambda^{-\nu^{-1}} \int_0^\infty y^{\nu^{-1}-1} (1 + y)^{-(s-1)}\, dy$$

$$= \tau\, b\, M_0\, \nu^{-1}\, \lambda^{-\nu^{-1}}\, \Gamma(\nu^{-1})\, \Gamma(s - 1 - \nu^{-1})/\Gamma(s - 1) \tag{2.10}$$

in terms of the gamma-function $\Gamma(x)$. If we now apply Stirling's asymptotic formula for the gamma-function, we find for $s \gg 1$

$$J(s) \sim \tau\, b\, M_0\, \nu^{-1}\, \lambda^{-\nu^{-1}}\, \Gamma(\nu^{-1})\, s^{-\nu^{-1}}, \quad s \gg 1. \tag{2.11}$$

Since $J(s) \ll 1$ when $s \gg 1$, the m.g.f. $f_i(s)$ in eq. (2.5) is asymptotically

$$f_i(s) \sim P_1(0)\, \tau\, b\, M_0\, \nu^{-1}\, \lambda^{-\nu^{-1}}\, \Gamma(\nu^{-1})\, s^{-\nu^{-1}}, \quad s \gg 1, \tag{2.12}$$

and its inverse Laplace transform is, for small values of g_1,

$$p_1'(g_1) \cong P_1(0)\, \tau\, b\, M_0\, \nu^{-1}\, \lambda^{-\nu^{-1}}\, g_1^{\nu^{-1}-1}. \tag{2.13}$$

The probability of detection is therefore approximately

$$Q_d \cong 1 - P_1(0) - P_1(0)\, \tau\, b\, M_0\, (g_{10}/\lambda)^{\nu^{-1}}. \tag{2.14}$$

We learn from this asymptotic analysis that the probability of detection depends strongly, when near 1, on the manner in which the image decreases to zero at its edge, which is reflected in the exponent ν in eq. (2.9) and the exponent ν^{-1} in eq. (2.14). If the image covers an infinite area, with $M_s(\underset{\sim}{x})$ decreasing to zero as $|\underset{\sim}{x}| \to \infty$, a different asymptotic analysis is required. There will now in general be zero probability of zero counts, $P_1(0) = 0$, and the delta-function vanishes from the origin.

As an example we take a circular Gaussian image,

$$M_S(\underset{\sim}{x}) = A \exp(-|\underset{\sim}{x}|^2/a^2), \tag{2.15}$$

to be detected against a uniform background M_0. The logarithm of the m.g.f. of g_1 becomes

$$\begin{aligned} \ln f_{11}(s) &= -\tau M_0 \int_0^\infty \int_0^{2\pi} [1 + \lambda \exp(-r^2/a^2)] \\ &\quad \times \left\{1 - [1 + \lambda \exp(-r^2/a^2)]^{-s}\right\} r dr d\varphi \\ &= -\bar{N} \int_0^\infty (1 + \lambda e^{-u}) [1 - (1 + \lambda e^{-u})^{-s}] du, \end{aligned} \tag{2.16}$$

where $\bar{N} = \tau M_0 \pi a^2$ is the mean number of counts under hypothesis H_0 from a circle of radius a. In eq. (2.16) we make the substitution

$$v = \ln(1 + \lambda e^{-u}) \tag{2.17}$$

to obtain*, with $b = \ln(1 + \lambda)$,

$$\begin{aligned} \ln f_{11}(s) &= -\bar{N} \int_0^b \frac{e^v(1 - e^{-sv})}{1 - e^{-v}} dv \\ &= -\bar{N} \int_0^b (e^v + 1) dv - \bar{N} \int_0^b \frac{e^{-v} - e^{-(s-1)v}}{1 - e^{-v}} dv \\ &= -\bar{N} (e^b - 1 + b) - \bar{N} \int_0^\infty \frac{e^{-v} - e^{-(s-1)v}}{1 - e^{-v}} + \bar{N} \int_b^\infty \frac{e^{-v} - e^{-(s-1)v}}{1 - e^{-v}} dv \\ &= -\bar{N} [\lambda + \ln(1 + \lambda)] - \bar{N} [\psi(s - 1) + C] - \bar{N} \ln(1 - e^{-b}) - \bar{N} \Delta J_\lambda(s) \\ &= -\bar{N} [\lambda + \ln \lambda + \psi(s - 1) + C] - \bar{N} \Delta J_\lambda(s), \end{aligned} \tag{2.18}$$

where $\psi(z)$ is the logarithmic derivative of the gamma function, C = 0.577215 is Euler's constant, and

---

* I.S. Gradshteyn and I. W. Ryzhik, Tables of Integrals, Series, and Products (Academic Press, New York, 1965), §3.311, eq. (6), p. 304.

$$\Delta J_\lambda(s) = \int_b^\infty \frac{e^{-(s-1)v}}{1 - e^{-v}}\, dv. \tag{2.19}$$

Since over $b < v < \infty$, $1 - e^{-b} < 1 - e^{-v} < 1$, we can bound $\Delta J_\lambda(s)$ by

$$\frac{e^{-(s-1)b}}{s - 1} < \Delta J_\lambda(s) < \left(\frac{1}{1 - e^{-b}}\right) \frac{e^{-(s-1)b}}{(s - 1)} \tag{2.20}$$

and we see that it decreases to zero exponentially as $s \to \infty$. Hence asymptotically

$$f_{11}(s) \sim \exp\{- \bar{N} [\lambda + \ell n\, \lambda + C + \psi(s - 1)]\}, \qquad s >> 1. \tag{2.21}$$

By using the asymptotic expansion* of the function $\psi(z)$,

$$\psi(z) \sim \ell n\, z - \frac{1}{2z} + O(z^{-2}), \quad R\ell\, z >> 1, \tag{2.22}$$

we obtain

$$\begin{aligned} f_{11}(s) &\sim \exp\{- \bar{N} [\lambda + C + \ell n(\lambda s) - \frac{3}{2s} + O(s^{-2})]\} \\ &= (\lambda s)^{-\bar{N}} e^{-\bar{N}(\lambda+C)} [1 - \frac{3\bar{N}}{2s} + O(s^{-2})]. \end{aligned}$$

The p.d.f. of the statistic g_1 is therefore, for $g_1 << \bar{N}$,

$$p_1(g_1) \sim \lambda^{-\bar{N}} e^{-\bar{N}(\lambda+C)} \Gamma(\bar{N})^{-1} g_1^{\bar{N}-1} [1 - \frac{3}{2} g_1 + O(g_1^2)],$$

and the probability of detection is

$$Q_d \sim 1 - \lambda^{-\bar{N}} e^{-\bar{N}(\lambda+C)} [\Gamma(\bar{N} + 1)]^{-1} g_{10}^{\bar{N}} (1 - \frac{3\bar{N}}{2(\bar{N} + 1)} g_{10} + O(g_{10}^2))$$

where g_{10} is the decision level on the statistic g_1.

For detection probabilities in the intermediate range where neither of the asymptotic methods just described will work, one can calculate the characteristic function $\underset{\sim}{E}(e^{i\omega g_1}|H_1) = f_1(-i\omega)$ of the distribution and take its inverse Fourier transform

$$p_1(g_1) = \int_{-\infty}^{\infty} f_1(-i\omega)\, e^{-i\omega g_1}\, d\omega/2\pi$$

* A. Erdélyi et al., Higher Mathematical Functions (McGraw-Hill, New York, 1953), vol. 1, §1.18, eq. (7), p. 47.

by one of the fast computer algorithms now available. This can then be integrated numerically to give the detection probability. Alternatively, the Gram-Charlier series can be tried, although it appears to be successful only for unimodal distributions that closely resemble the Gaussian.*

* C. W. Helstrom, Statistical Theory of Signal Detection (Pergamon Press, Oxford, 2nd ed., 1968), pp. 219-222.

Quantum Detection Theory

An eighty-page review article on quantum detection theory has been prepared for publication in Progress in Optics, E. Wolf, editor.

Quantum detection is a form of statistical hypothesis testing adapted to quantum-mechanical rather than classical laws of statistics. It has been formulated in terms of the conventional rules of quantum mechanics, one of which is that only observables associated with commuting operators can be measured simultaneously on the same system. The possibility of simultaneously measuring noncommuting observables has been studied in recent years, and it has been suggested that this would permit quantum receivers to attain lower average error probabilities or Bayes costs. In a paper attached to this report it is shown that no reduction in the minimum Bayes costs can be achieved in this manner. Measuring noncommuting observables requires an auxiliary apparatus that initially possesses no information about the state of the quantum receiver, and there is no way in which introducing such an apparatus can lead to a lower Bayes cost than the minimum attainable by measuring commuting observables on the receiver alone. Quantum-mechanical Cramér-Rao bounds on the mean square errors of unbiased estimates of parameters of a signal in a quantum receiver can likewise not be lowered by measuring noncommuting observables.

Noncommuting Observables in

Quantum Detection and Estimation Theory

Carl W. Helstrom*

Department of Applied Physics and Information Science

University of California, San Diego

La Jolla, California 92037

Abstract

In quantum detection theory the optimum detection operators must commute; admitting simultaneous approximate measurement of noncommuting observables cannot yield a lower Bayes cost. The lower bounds on mean square errors of parameter estimates predicted by the quantum-mechanical Cramér-Rao inequality can also not be reduced by such means.

* This research was supported by Grant NGL 05-009-079 from the National Aeronautics and Space Administration.

Quantum detection and estimation theory has been developed within the conventional framework of quantum mechanics, one of the principal tenets of which is that only observables associated with commuting operators can be simultaneously measured on the same system.[1-3] It has been suggested that this formulation is too restrictive, that noncommuting operators can be at least approximately measured on the same system, and that to include this possibility may permit more effective detection, as measured by a lower average Bayes cost.[4,5] We wish to show that no such improvement can be expected.

The simultaneous measurement of noncommuting observables has been treated by Gordon and Louisell.[6] In order to approximately measure certain such observables on a quantum-mechanical system S, it is made to interact for a time with a second system A, termed the apparatus. It was shown that a suitably defined ideal measurement yielding approximate values of the noncommuting observables can be based on the outcome of measurements of commuting observables on the apparatus A, or more generally on both S and A. What we must therefore do is apply quantum detection theory--with its restriction to commuting observables--to the combined system S + A.

Suppose we are to decide among M hypotheses $H_1, H_2, \ldots, H_M$. Under hypothesis H_j the density operator for the combined system at time t is $\rho_j^{S+A}(t)$ in the Schrödinger picture. If at an earlier time t_0 the density operator is $\rho_j^{S+A}(t_0)$, the two operators are related by[7]

$$\rho_j^{S+A}(t) = U(t, t_0)\, \rho_j^{S+A}(t_0)\, U^+(t, t_0), \tag{1}$$

with

$$U(t, t_0) = \exp\left[-i \int_{t_0}^{t} H dt'/\hbar\right], \tag{2}$$

where H is the Hamiltonian operator for the combined system S + A and $\hbar$ is

Planck's constant $h/2\pi$. The operator U is unitary; that is, with U^+ its Hermitian adjoint, UU^+ equals the identity operator $\underset{\sim}{1}$.

Let $\{\Pi_j\}$ be a set of commuting projection operators forming an M-fold resolution of the identity,

$$\sum_{j=1}^{M} \Pi_j = \underset{\sim}{1}. \tag{3}$$

On the combination S + A we are to measure these M projection operators at time t, and if the k-th yields the value 1, hypothesis H_k is selected as true.[1] The average cost is then

$$\bar{C} = \sum_{i=1}^{M}\sum_{j=1}^{M} \zeta_j C_{ij} \operatorname{Tr}[\rho_j^{S+A}(t)\, \Pi_i], \tag{4}$$

where ζ_j is the prior probability of hypothesis H_j and C_{ij} is the cost of choosing H_i when H_j is true. Let $\{\Pi_j(t)\}$ be the projection operators that minimize $\bar{C}$ when the system S + A is observed at time t; we call these optimum. Then by (1) the operators

$$\Pi_j(t_0) = U^+(t, t_0)\, \Pi_j(t)\, U(t, t_0) \tag{5}$$

will minimize $\bar{C}$ when S + A is observed at time t_0. Because of the unitarity of $U(t, t_0)$, the set $\{\Pi_j(t_0)\}$ also forms an M-fold resolution of the identity into commuting projection operators, and the $\Pi_j(t_0)$ are optimum at time t_0. Since the minimization is carried out over all possible M-fold resolutions of identity, the minimum Bayes cost $\bar{C}_{min}$ must be independent of the observation time t.

Now let us roll time back to an epoch t_0 before the system S has come into contact with the apparatus A. In the Schrödinger picture this amounts to applying the inverse unitary transformation $U^+(t, t_0)$ to the state vectors of the combined system S + A. Because S and A are independent at this time t_0, the density operators ρ_j^{S+A} must now have the factored form

$$\rho_j^{S+A}(t_0) = \rho_j^S(t_0)\, \rho^A(t_0), \quad j = 1, 2, \ldots M. \tag{6}$$

Furthermore, as the apparatus A before the interaction has no information about which hypothesis is true, $\rho^A(t_0)$ in (6) must be independent of j.

The Bayes cost is now

$$\bar{C} = \mathrm{Tr}\left[\rho^A(t_0) \sum_{i=1}^{M}\sum_{j=1}^{M} \zeta_j\, C_{ij}\, \rho_j^S(t_0)\, \Pi_i\right]. \tag{7}$$

Since S and A are completely uncoupled, and the state of A is independent of which hypothesis H_j is true, there is nothing to be gained by observing A. The optimum projection operators $\Pi_i(t_0)$ factor as $\Pi_i^S(t_0)\, \underset{\sim}{1}^A$, where $\underset{\sim}{1}^A$ is the identity operator for the apparatus A, and the set $\{\Pi_j^S(t_0)\}$ forms an M-fold resolution of the identity $\underset{\sim}{1}^S$ for the system S, minimizing the Bayes cost

$$\bar{C}_S = \sum_{i=1}^{M}\sum_{j=1}^{M} \zeta_j\, C_{ij}\, \mathrm{Tr}[\rho_j^S(t_0)\, \Pi_i]. \tag{8}$$

Since $\mathrm{Tr}[\rho^A\, \underset{\sim}{1}^A] = 1$, the minimum value of $\bar{C}_S$ is also the minimum value of $\bar{C}$ in (7) and equals the time-independent minimum Bayes cost $\bar{C}_{min}$. The decision among the M hypotheses made at time t_0 is based entirely on the measurement of commuting observables on system S.

Similar considerations apply to estimating the m parameters $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ of the density operator $\rho^S(\underset{\sim}{\theta})$ of a quantum-mechanical system S. A version of the Cramér-Rao inequality sets lower bounds to mean square errors of unbiased estimates of $\theta_1, \theta_2, \ldots, \theta_m$.[3] Let X_j be an operator whose measurement on S yields an unbiased estimate $\hat{\theta}_j$ of the j-th parameter; $\hat{\theta}_j$ must be an eigenvalue of X_j. Although in order to be measured simultaneously on the same system the operators X_j must commute, the analysis leading to the lower bounds given in [3] does not require commutativity of the operators X_j.

For the class of *commuting* operators yielding unbiased estimates of the parameters $\underset{\sim}{\theta}$ there will exist lower bounds on the mean square errors, and those will be greater than or at least equal to the bounds derived in [3].

Again including the possibility of measuring noncommuting observables cannot lead to lower bounds smaller than those in [3]. In order to measure such operators even approximately, a measuring apparatus A must be allowed to interact with the system S, and according to Gordon and Louisell's treatment of the process, commuting operators will at the end be measured on the combined system S + A.[6] In the Schrödinger picture the density operator $\rho^{S+A}(\underset{\sim}{\theta}, t)$ for S + A will have a time dependence similar to that in (1).

Referring to (7) of [3] we see that the symmetrized logarithmic derivatives (SLD) $L_j(t)$ appropriate for determining the Cramér-Rao lower bounds when the measurements of X_j are made at time t are related to those appropriate for measurements made at t_0 by

$$L_j(t) = U(t, t_0)\, L_j(t_0)\, U^+(t, t_0). \qquad (9)$$

Then (13) of [3] shows that the matrix $\underset{\sim}{A}$ that sets the lower bounds is independent of the time t of observation, again because of the unitarity of the operator $U(t, t_0)$.

Once more we move back to an epoch t_0 before the system and the apparatus have interacted. The density operator $\rho^{S+A}(\underset{\sim}{\theta}, t_0)$ factors as $\rho^S(\underset{\sim}{\theta}, t_0)\, \rho^A(t_0)$, where the density operator $\rho^A(t_0)$ of the apparatus A is independent of the estimanda $\underset{\sim}{\theta}$. The SLD operators for calculating the lower bounds are now the solutions of the operator equations

$$\partial\rho^S(\underset{\sim}{\theta}, t_0)/\partial\theta_n = \frac{1}{2}\,[\rho^S(\underset{\sim}{\theta}, t_0)\, L_n{}^S + L_n{}^S\, \rho^S(\underset{\sim}{\theta}, t_0)], \qquad (10)$$

and they act only on system S, commuting with ρ^A and all other operators on the apparatus A. When taking the trace over the states of A to form the elements of the matrix $\underset{\sim}{A}$, the density operator ρ^A is replaced by 1, and the lower bounds depend only on $\rho^S(\underset{\sim}{\theta}, t_0)$. Thus the apparatus A cannot help estimate the parameters $\underset{\sim}{\theta}$ of S with smaller mean square errors than the lower bounds calculated by the quantum-mechanical Cramér-Rao inequality as applied to the density operator of system S alone.

In [3, p. 238] lower bounds were calculated for unbiased estimates $\hat{m}_x$ and $\hat{m}_y$ of the components of the complex amplitude $\mu = m_x + im_y$ of a simple harmonic oscillator, which might represent a mode of the field in an ideal receiver in the presence of thermal noise. Those bounds are

$$\text{Var } \hat{m}_x \geq \frac{1}{2}(N + \frac{1}{2}), \qquad \text{Var } \hat{m}_y \geq \frac{1}{2}(N + \frac{1}{2}),$$

where N is the mean number of noise photons. The noncommutativity of the SLD's L_x and L_y used to derive these bounds does not invalidate them. It can be shown that if the mode is coupled with an ideal amplifier whose gain is high enough to raise the oscillator variables to the classical domain where they commute, error variances

$$\text{Var } \hat{m}_x = \text{Var } \hat{m}_y = \frac{1}{2}(N + 1)$$

can be attained [8]. It is unknown whether commuting operators can be found whose measurement will yield unbiased estimates $\hat{m}_x$ and $\hat{m}_y$ with variances lying between $\frac{1}{2}(N + \frac{1}{2})$ and $\frac{1}{2}(N + 1)$.

The measurements we need to make on a quantum-mechanical system S for testing hypotheses or estimating parameters will always have to be effected by means of an auxiliary apparatus A, and this apparatus, subject to thermal and quantum fluctuations of its own, will ordinarily introduce additional random uncertainties. Each measurement procedure will have to be analyzed to determine what error costs it entails. Detection theory and estimation theory seek lower bounds on these costs, and in doing so they minimize with respect to the entire class of possible detection or estimation operators that can be applied to the system. The resulting bounds are independent of the time of observation, and they cannot be reduced by using any auxiliary apparatus that initially posseses no information about the state of the system.

References

1. C. W. Helstrom, "Detection theory and quantum mechanics", Infor. and Control, vol. 10, pp. 254-291; March, 1967.
2. C. W. Helstrom, "Detection theory and quantum mechanics, II", Infor. and Control, vol. 13, pp. 156-171; August, 1968.
3. C. W. Helstrom, "The minimum variance of estimates in quantum signal detection", IEEE Trans. on Information Theory, vol. IT-14, pp. 234-242; March, 1968.
4. C. W. Helstrom, J. W. S. Liu, and J. P. Gordon, "Quantum-mechanical communication theory", Proceedings of the IEEE, vol. 58, pp. 1578-1598; October, 1970.
5. H. P. Yuen, R. S. Kennedy, and M. Lax, "On optimal quantum receivers for digital signal detection", Proceedings of the IEEE, vol. 58, pp. 1770-1773; October, 1970.
6. J. P. Gordon and W. H. Louisell, "Simultaneous measurement of noncommuting observables", in *Physics of Quantum Electronics*, P. L. Kelley, B. Lax, and P. E. Tannenwald, Eds. New York: McGraw-Hill, 1966, pp. 833-840.
7. W. H. Louisell, *Radiation and Noise in Quantum Electronics*. New York: McGraw-Hill, 1964, p. 225.
8. C. W. Helstrom, "Quantum detection theory", to appear in *Progress in Optics*, E. Wolf, Ed. Amsterdam: North-Holland Publishing Co.

Linear Restoration of Incoherently Radiating Objects

Carl W. Helstrom*

Department of Applied Physics and Information Science

University of California, San Diego

La Jolla, California 92037

ABSTRACT

Light from an incoherently radiating object and background light are focused onto a photosensitive mosaic, the currents from whose elements constitute the data on which is based a least-squares linear estimate of the radiance at points in the object. By comparison of the mean square error with that given by Wiener filtering theory, the equivalent noise spectral density for use in the latter is shown to consist of a shot-noise term and a term due to the random fluctuations of the incoherent light; the former predominates under most circumstances. Turbulent distortion of the image after passage of the rays through a random phase screen is also treated from this standpoint.

* This research was supported by Grant NGL 05-009-079 from the National Aeronautics and Space Administration.

The restoration of degraded images is often based on a linear model of image formation: a two-dimensional object function is pictured as having passed through a spatial filter, to whose output spatial random noise has been added; the sum of these constitutes the observed data.[1-7] In the approach known as Wiener filtering, the data are passed through a linear spatial filter whose transfer function is designed to reproduce the original object function within the smallest possible mean square error.[2] In actuality, most objects radiate incoherently, and it is not the light field they create that is wanted; that field is a complex spatio-temporal gaussian random process, whose instantaneous values are of no interest. Rather the observer wants to know the distribution of radiance across the object, and the radiance is related to the mutual coherence function, which is the statistical average of a quadratic functional of the light field. Furthermore, the linear filtering theory does not explicitly distinguish the two basic types of noise, that due to the inherent fluctuations of the light fields of object and background, and that associated with the process of recording the light. Instead it makes certain assumptions about the spatial spectrum of the noise and requires its strength to be measured separately. This theory is inadequate to determine the basic limitations set by nature on the restorability of degraded images.

A previous paper showed how the statistical properties of the information-bearing quadratic functional of the light field modify the usual Shannon formula for the information transfer from an incoherently radiating object.[8] Here we shall treat the linear processing of the image illuminance that seeks an estimate of the radiance distribution in a distant object plane. In order to incorporate the random behavior of the recording medium, we shall assume that the light from the object is focused on a photosensitive surface divided like a mosaic into many small, contiguous regions, the photoelectronic currents from each of which make up the observed data. These data are combined linearly to provide estimates of the radiance at discrete points of the object plane.

Only spatial filtering of this kind will be considered. Temporal filtering for reducing the estimation error on the basis of color differences between object and background will be disregarded. The object light and the background light will be assumed to pass through a narrowband filter, over whose passband both have uniform spectral densities.

We begin by working out the means and covariances of the photocurrents in terms of the mutual coherence function of the incident light, which in turn is related to the radiance distribution of the object plane, our estimandum, and to the background radiance. The object radiance is considered as one of an ensemble of spatial random processes, which form the class of objects to be examined. The squared error of the radiance estimate is averaged over this ensemble of object processes, as well as with respect to the statistical distributions of the light fields and the photoelectrons. The spatial filtering that minimizes the resultant mean square error is then easily determined.

The final expressions for the mean square error will be found to have the same form as those derived from the Wiener filtering theory, and by comparing them we can determine the spatial spectral density that must be assigned to the random noise in utilizing the Wiener technique. This equivalent noise density is the sum of a constant term due to the randomness of the photoelectron emissions ("shot noise") and a spatially variable term arising from the fluctuations of the light fields produced by object and background. Finally, since removing distortion by atmospheric turbulence is a primary goal of image restoration, the effect of a simple kind of turbulence--the random phase screen--will also be analyzed in this way.

1. The Photocurrents

The system to be analyzed is shown in Fig. 1. The object plane O emits incoherent light, which is received at the aperture A of an observing optical instrument. A narrowband filter, not shown, cuts out background light outside the temporal spectrum of the object light. The aperture A contains a lens to focus the light on the image plane I, which consists of a mosaic of isolated, but close photosensitive spots of area s. The light ejects photoelectrons from these spots, and the resulting integrated currents I_m are measured and constitute the data on which estimates of the object radiance are to be based. The spots are so small and so close together that ultimately we shall treat them as infinitesimal.

The object plane is likewise sampled at a dense set of points $\underset{\sim}{u}_k = (u_{kx}, u_{ky})$, and the radiance $B(\underset{\sim}{u}_k)$ is estimated as a linear combination of the integrated currents I_m,

$$\hat{B}(\underset{\sim}{u}_k) = \sum_m L_{km} I_m + \text{const.}, \qquad (1.1)$$

where the weights L_{km} introduced by the processing matrix $\underset{\sim}{L}$ are selected to minimize a mean square error; the constant has a known value.

Dividing time into small intervals Δt, we write the current from the m-th spot as

$$I_m(t) = \sum_i n_m(t_i)\, f(t - t_i), \qquad (1.2)$$

where $n_m(t_i)$ is the number of photoelectrons emitted during the i-th interval, and $f(t)$ is the current pulse resulting from each emission; Δt is much smaller than the duration of $f(t)$. For a given light field at the image plane, the numbers $n_m(t_i)$ are random variables with a Poisson distribution, and they are statistically independent from one spot to another and from one interval to another.

The object, aperture, and image planes are assumed to be so far apart, and the aperture so small, that all light rays can be considered paraxial. If for simplicity we suppose the light linearly polarized, it suffices to use a scalar theory, and we represent the field at point $\underset{\sim}{x}$ of the image plane by the analytic signal $\psi(\underset{\sim}{x}, t)\, e^{-i\Omega t}$, where $\Omega = 2\pi c/\lambda$ is its central angular frequency, λ is its wavelength, and $\psi(\underset{\sim}{x}, t)$ is the complex envelope. The conditional expected value of the number $n_m(t_i)$ is given in terms of this field by

$$\underset{\sim}{E}[n_m(t_i)|\psi] = \frac{1}{2}\, \alpha s |\psi(\underset{\sim}{x}_m, t_i)|^2\, \Delta t, \tag{1.3}$$

where $\alpha = \eta/\hbar\Omega$ is the quantum efficiency η of the surface divided by the quantum $\hbar\Omega$ of energy of the light. The conditional mean of the current I_m from the spot centered at $\underset{\sim}{x}_m$ is then

$$\begin{aligned}\underset{\sim}{E}(I_m|\psi) &= \frac{1}{2}\, \alpha s \sum_i |\psi(\underset{\sim}{x}_m, t_i)|^2\, f(t - t_i)\, \Delta t \\ &\to \frac{1}{2}\, \alpha s \int_{-\infty}^{\infty} |\psi(\underset{\sim}{x}_m, \tau)|^2\, f(t - \tau)\, d\tau \end{aligned} \tag{1.4}$$

when we pass to the limit $\Delta t \to 0$.

The current from each spot will be integrated by a filter with a long time constant T in order to determine the total charge ejected by the light during the observation interval. This filter, acting on each component pulse f(t) of the current, replaces it in Eq. (1.2) by a new pulse whose duration is of the order of the integration time T. We can therefore regard the data as represented by $I_m(t)$ of Eq. (1.2), but with a pulse shape f(t) that is positive and of duration T. If we suppose $f(0) = \max_t f(t)$, a useful definition of the duration T is

$$T = \int_{-\infty}^{\infty} [f(t)]^2 \, dt/[f(0)]^2; \tag{1.5}$$

T is in effect the observation time. Without loss of generality we put $f(0) = 1$.

In order to work out the mean square error of our estimate, we shall need the covariance of the currents from two different spots. It can be evaluated as in the semi-classical analysis of the Hanbury-Brown-Twiss effect,[9,10] which utilizes the equality of the variance of the Poisson distribution of $n_m(t_i)$ with the mean value given in Eq. (1.4), and which also involves taking the expected value of a product of four correlated complex gaussian variables by the formula[11]

$$\langle \psi_1 \psi_2^* \psi_3 \psi_4^* \rangle = \langle \psi_1 \psi_2^* \rangle \langle \psi_3 \psi_4^* \rangle + \langle \psi_1 \psi_4^* \rangle \langle \psi_2^* \psi_3 \rangle .$$

The resulting covariance is

$$\{I_m(t),\ I_p(t)\} = \langle I_m(t)\ I_p(t) \rangle - \langle I_m(t) \rangle \langle I_p(t) \rangle =$$

$$\alpha^2 s^2 \ |\Psi_I(\underset{\sim}{x}_m,\ \underset{\sim}{x}_p)|^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\chi(\tau_1 - \tau_2)|^2 \, f(t - \tau_1)\, f(t - \tau_2)\, d\tau_1 d\tau_2$$

$$+ \alpha s\ \Psi_I(\underset{\sim}{x}_m,\ \underset{\sim}{x}_m)\ \delta_{mp} \int_{-\infty}^{\infty} [f(t - \tau)]^2 \, d\tau, \tag{1.6}$$

where δ_{mp} is the Kronecker delta and $\Psi_I(\underset{\sim}{x}_1,\ \underset{\sim}{x}_2)$ is the spatial part of the mutual coherence function of the light field at the image plane,

$$\tfrac{1}{2}\langle \psi(\underset{\sim}{x}_1,\ t_1)\ \psi^*(\underset{\sim}{x}_2,\ t_2) \rangle = \Psi_I(\underset{\sim}{x}_1,\ \underset{\sim}{x}_2)\ \chi(t_1 - t_2). \tag{1.7}$$

The temporal part $\chi(t_1 - t_2)$ is the Fourier transform of the absolute square $X(\omega)$ of the transfer function of the input filter and is so normalized that $\chi(0) = 1$,

$$\chi(\tau) = \int_{-\infty}^{\infty} X(\omega)\ e^{-i\omega t} \, d\omega/2\pi .$$

Here angular frequency ω is defined with respect to Ω as origin. The angular brackets in Eq. (1.6) refer to averages with respect to the distributions of both the light fields and the photoelectron emissions.

If we define the bandwidth W of the filtered light as

$$W = |\chi(0)|^2 \Big/ \int_{-\infty}^{\infty} |\chi(\tau)|^2 \, d\tau = \left[\int_{-\infty}^{\infty} X(\omega)\, d\omega/2\pi\right]^2 \Big/ \int_{-\infty}^{\infty} |X(\omega)|^2 \, d\omega/2\pi, \tag{1.8}$$

and if we assume $WT >> 1$, as is normal in practice, the double integral in Eq. (1.6) can be simplified, and the covariance of the currents becomes

$$\{I_m(t),\ I_p(t)\} = \alpha^2 s^2 \ |\Psi_I(\underset{\sim}{x}_m,\ \underset{\sim}{x}_p)|^2 \ T/W + \alpha s \ \Psi_I(\underset{\sim}{x}_m,\ \underset{\sim}{x}_m)\ \delta_{mp}\ T. \tag{1.9}$$

In terms of the mutual coherence function, the mean value of the current from the m-th spot is, by Eq. (1.4),

$$\langle I_m(t)\rangle = \alpha s \ \Psi_I(\underset{\sim}{x}_m,\ \underset{\sim}{x}_m)\ \rho T, \tag{1.10}$$

where ρ is a positive constant defined by

$$\rho = T^{-1} \int_{-\infty}^{\infty} f(t)\ dt/f(0). \tag{1.11}$$

For rectangular current pulses $\rho = 1$.

We must now evaluate the spatial coherence function $\Psi_I(\underset{\sim}{x}_1,\ \underset{\sim}{x}_2)$ in terms of the radiance distribution $B(\underset{\sim}{u})$ of the object plane and the spectral density N of the background light.

2. The Mutual Coherence Function

The field $\psi_I(\underset{\sim}{x}, t)$ at point $\underset{\sim}{x}$ of the image plane is given in terms of the aperture field $\psi_a(\underset{\sim}{r}, t)$ by

$$\psi_I(\underset{\sim}{x}, t) = \int_A S_1(\underset{\sim}{x}, \underset{\sim}{r})\, \psi_a(\underset{\sim}{r}, t)\, d^2\underset{\sim}{r}, \tag{2.1}$$

where $S_1(\underset{\sim}{x}, \underset{\sim}{r})$ is the point-spread function (psf) between aperture and image plane. In the paraxial approximation

$$S_1(\underset{\sim}{x}, \underset{\sim}{r}) = i(\lambda R_1)^{-1} \exp\left[ikR_1 + \frac{ik}{2R_1}|\underset{\sim}{r} - \underset{\sim}{x}|^2 - \frac{ik}{2F}\underset{\sim}{r}^2\right],$$

where $k = \Omega/c = 2\pi/\lambda$ is the propagation constant and R_1 is the distance from the aperture to the image plane. The last term in the exponent is the phase shift introduced by the lens, whose focal length is F and which focuses the object plane at distance R onto the image plane,

$$F^{-1} = R^{-1} + R_1^{-1}. \tag{2.2}$$

The aperture field consists of two statistically independent parts, the light from the object and the background light,

$$\psi_a(\underset{\sim}{r}, t) = \int_O S_2(\underset{\sim}{r}, \underset{\sim}{u})\, \psi_o(\underset{\sim}{u}, t)\, d^2\underset{\sim}{u} + \psi_n(\underset{\sim}{r}, t), \tag{2.3}$$

where $\psi_o(\underset{\sim}{u}, t)$ is the field at point $\underset{\sim}{u}$ of the object plane, $S_2(\underset{\sim}{r}, \underset{\sim}{u})$ is the psf between object and aperture, and $\psi_n(\underset{\sim}{r}, t)$ is the background component of the aperture field. In the paraxial approximation

$$S_2(\underset{\sim}{r}, \underset{\sim}{u}) = i(\lambda R)^{-1} \exp\left[ikR + \frac{ik}{2R}|\underset{\sim}{u} - \underset{\sim}{r}|^2\right]. \tag{2.4}$$

As the background light is spatially incoherent, its mutual coherence function after filtering has the form

$$\tfrac{1}{2}\langle\psi_n(\underset{\sim}{r}_1, t_1)\, \psi_n^*(\underset{\sim}{r}_2, t_2)\rangle = N\, \delta(\underset{\sim}{r}_1 - \underset{\sim}{r}_2)\, \chi(t_1 - t_2), \tag{2.5}$$

where $\delta(\underset{\sim}{r})$ is the two-dimensional delta-function. The object light is spatially incoherent at the object plane, where its mutual coherence function is

$$\frac{1}{2}\langle\psi_o(\underset{\sim}{u}_1, t_1)\ \psi_o^*(\underset{\sim}{u}_2, t_2)\rangle = (\lambda^2/4\pi)\ B(\underset{\sim}{u}_1)\ \delta(\underset{\sim}{u}_1 - \underset{\sim}{u}_2)\ \chi(t_1 - t_2) \tag{2.6}$$

if we suppose the temporal filtering already applied. Here $B(\underset{\sim}{u})$ is the radiance of the object plane integrated over the passband of our input filter. It is decomposed as

$$B(\underset{\sim}{u}) = B_0 + b(\underset{\sim}{u}) \tag{2.7}$$

into a known average level B_0 and a deviation $b(\underset{\sim}{u})$. It is $b(\underset{\sim}{u})$ that the system is to estimate; it represents the scene of interest.

When we combine all these equations, we find for the spatial coherence function of the light at the image plane

$$\begin{aligned}\Psi_I(\underset{\sim}{x}_1, \underset{\sim}{x}_2) &= (4\pi R_1{}^2)^{-1} \exp\left[ik(\underset{\sim}{x}_1{}^2 - \underset{\sim}{x}_2{}^2)/2R_1\right] \\ &\times \left\{B_0{}' \int_A \exp\left[ik\underset{\sim}{r}\cdot(\underset{\sim}{x}_2 - \underset{\sim}{x}_1)/R_1\right] d^2\underset{\sim}{r} + \right. \\ &(\lambda R)^{-2} \int_A\int_A\int_O b(\underset{\sim}{u}) \exp\left[\frac{ik}{R_1}(\underset{\sim}{x}_2\cdot\underset{\sim}{r}_2 - \underset{\sim}{x}_1\cdot\underset{\sim}{r}_1) + \frac{ik}{R}\underset{\sim}{u}\cdot(\underset{\sim}{r}_2 - \underset{\sim}{r}_1)\right] \\ &\left.\times\ d^2\underset{\sim}{u}\,d^2\underset{\sim}{r}_1\,d^2\underset{\sim}{r}_2\right\},\end{aligned} \tag{2.8}$$

where

$$B_0{}' = B_0 + 4\pi N/\lambda^2 \tag{2.9}$$

and A and O indicate integrations over the aperture and the object planes, respectively. The uniform radiance level B_0 adds to the background density, and we merge the two into a single uniform radiance level $B_0{}'$. In our discussions $4\pi N/\lambda^2$ will be neglected in comparison with B_0; it can easily be restored by using Eq. (2.9).

It is convenient to rescale the coordinates in the image plane so that the geometrical image of a point $\underset{\sim}{u}$ in the object plane also has coordinates $\underset{\sim}{u}$;

we put

$$\underset{\sim}{v} = -R\,\underset{\sim}{x}/R_1 \tag{2.10}$$

and write the spatial coherence function as

$$\Psi_I(\underset{\sim}{v}_1, \underset{\sim}{v}_2) = A(4\pi R_1{}^2)^{-1} \exp\,[ikR_1(\underset{\sim}{v}_1{}^2 - \underset{\sim}{v}_2{}^2)/2R_0{}^2]$$
$$\times \left\{B_0'\,\mathscr{I}(\underset{\sim}{v}_1 - \underset{\sim}{v}_2) + (\lambda R)^{-2} A \int b(\underset{\sim}{u})\ \mathscr{I}(\underset{\sim}{v}_1 - \underset{\sim}{u})\,\mathscr{I}^*(\underset{\sim}{v}_2 - \underset{\sim}{u})\ d^2\underset{\sim}{u}\right\}, \tag{2.11}$$

where

$$\mathscr{I}(\underset{\sim}{u}) = A^{-1} \int_A I_A(\underset{\sim}{r}) \exp\,(ik\underset{\sim}{u}\cdot\underset{\sim}{r}/R)\ d^2\underset{\sim}{r} \tag{2.12}$$

is the Fourier transform of the indicator function of the aperture, defined as

$$I_A(\underset{\sim}{r}) = 1, \quad \underset{\sim}{r} \in A; \qquad I_A(\underset{\sim}{r}) = 0, \quad \underset{\sim}{r} \notin A. \tag{2.13}$$

3. The Radiance Estimate

If we now use Eqs. (1.10) and (2.11) to evaluate the expected value of the current from the m-th spot on the image plane, we obtain

$$\langle I_m \rangle = I_0 + \rho\, C_1 s \int_0 b(\underset{\sim}{u})\ |\mathscr{I}(\underset{\sim}{v}_m - \underset{\sim}{u})|^2\, d^2\underset{\sim}{u}, \tag{3.1}$$

$$I_0 = \alpha\rho T s\ A\ B_0'(4\pi R_1{}^2)^{-1},$$

$$C_1 = \alpha T\ A^2(\lambda R)^{-2}(4\pi R_1{}^2)^{-1}, \tag{3.2}$$

where $\underset{\sim}{v}_m$ is the scaled coordinate vector of the center of the m-th spot. Except for the known constant I_0, the mean value of I_m is proportional to the weighted integral of the true radiance deviation $b(\underset{\sim}{u})$ over the neighborhood of the object point $\underset{\sim}{v}_m$. The kernel $|\mathscr{I}(\underset{\sim}{v} - \underset{\sim}{u})|^2$ is the incoherent point-spread function (psf); its diameter is of the order of the conventional resolution interval $\delta = \lambda R/a$, where a is the radius of the aperture, taken as circular.

If $b(\underset{\sim}{u})$ varies only slightly over a distance δ, $I_m - I_0$ itself provides a good estimate of $b(\underset{\sim}{v}_m)$; the linear processing of the currents I_m is intended to improve that estimate, and the new estimate of the radiance deviation at point $\underset{\sim}{u}_m$ of the object plane is taken as

$$\hat{b}(\underset{\sim}{u}_m) = \sum_n L_{mn}(I_n - I_0), \tag{3.3}$$

with the coefficients L_{mn} to be determined as shown hereafter.

The expected value of the radiance estimate $\hat{b}(\underset{\sim}{v}_m)$ can be obtained from Eqs. (3.1) and (3.3),

$$\langle \hat{b}(\underset{\sim}{u}_m) \rangle = \rho\, C_1 s \sum_n L_{mn} \int_0 b(\underset{\sim}{v})\ |\mathscr{I}(\underset{\sim}{u}_m - \underset{\sim}{v})|^2\, d^2\underset{\sim}{v}, \tag{3.4}$$

where s is the area of an element of the image plane. We now assume that s is so small, and the spots so close together, that this summation can be approximated by an integral,

$$\langle \hat{b}(\underset{\sim}{u}) \rangle = \rho\ C_1 (R_1/R)^2 \int\int L(\underset{\sim}{u}, \underset{\sim}{w})\ |\mathscr{I}(\underset{\sim}{w} - \underset{\sim}{v})|^2\ b(\underset{\sim}{v})\ d^2\underset{\sim}{w} d^2\underset{\sim}{v}, \qquad (3.5)$$

where we introduce a smooth weighting function $L(\underset{\sim}{u}, \underset{\sim}{w})$ whose values at the sampling points are $L(\underset{\sim}{u}_m, \underset{\sim}{u}_n) = L_{mn}$. The factor $(R_1/R)^2$ arises from the scaling adopted in Eq. (2.10).

4. The Mean Square Error

Any system for image restoration must be evaluated with respect to a definite class of objects. The statistical properties assigned to the class reflect the nature of the objects whose images the system is designed to restore. Here the objects are regarded as having radiance distributions $B(\tilde{u}) = B_0 + b(\tilde{u})$ that are homogeneous two-dimensional random processes of mean value

$$\tilde{E}\, B(\tilde{u}) = B_0 \tag{4.1}$$

and autocovariance function

$$\tilde{E}[b(\tilde{u})\, b(\tilde{v})] = \varphi(\tilde{u} - \tilde{v}), \quad \sigma_B{}^2 = \varphi(\tilde{0}). \tag{4.2}$$

The width of $\varphi(\tilde{u})$ as a function of $|\tilde{u}|$ represents the size of typical details in the object plane; the ratio $\sigma_B{}^2/B_0{}^2$ specifies the mean square contrast. The Fourier transform of $\varphi(\tilde{u})$ provides the distribution of spatial frequencies in objects of the class. Objects in which sharp edges predominate, for example, will have spatial spectra decreasing slowly at high spatial frequencies.

As the actual object radiance is unknown *a priori*, the system can be designed only to attain a certain average performance with respect to the given class of objects. If it does so, it can be expected to restore efficiently most objects of the class. The measure of performance adopted here is the common mean square error

$$\mathscr{E} = \tilde{E}\langle [\hat{b}(\tilde{u}) - b(\tilde{u})]^2 \rangle, \tag{4.3}$$

in which the angular brackets denote, as before, averages over the distributions of the fields and the photoelectron emissions, and the symbol $\tilde{E}$ now denotes an expected value over the class of anticipated objects. The set of coefficients

L_{mn}, or the weighting function $L(\underset{\sim}{u}, \underset{\sim}{v})$ of which they are samples, is to be chosen to minimize $\mathscr{E}$.

The mean square error can be written as

$$\mathscr{E} = \underset{\sim}{E}\langle [\hat{b}(\underset{\sim}{u}) - \langle \hat{b}(\underset{\sim}{u})\rangle + \langle \hat{b}(\underset{\sim}{u})\rangle - b(\underset{\sim}{u})]^2\rangle$$

$$= \underset{\sim}{E} \text{ Var } \hat{b}(\underset{\sim}{u}) + \underset{\sim}{E}[\langle \hat{b}(\underset{\sim}{u})\rangle - b(\underset{\sim}{u})]^2, \tag{4.4}$$

where Var stands for the variance with respect to the field and photoelectron distributions. From Eqs. (3.3) and (1.9)

$$\text{Var } \hat{b}(\underset{\sim}{u}_m) = \sum_n \sum_p L_{mn} L_{mp} \{I_n, I_p\} =$$

$$\sum_n \sum_p L_{mn} L_{mp} \{\alpha^2 s^2 |\Psi_I(\underset{\sim}{x}_n, \underset{\sim}{x}_p)|^2 T/W +$$

$$\sum_n L_{mn}^2 \alpha s \Psi_I(\underset{\sim}{x}_n, \underset{\sim}{x}_n) T \rightarrow$$

$$\alpha^2 (T/W)(R_1/R)^4 \iint L(\underset{\sim}{u}_m, \underset{\sim}{v}_1) L(\underset{\sim}{u}_m, \underset{\sim}{v}_2) |\Psi_I(\underset{\sim}{v}_1, \underset{\sim}{v}_2)|^2 d^2\underset{\sim}{v}_1 d^2\underset{\sim}{v}_2 +$$

$$\alpha T(R_1/R)^2 \int [L(\underset{\sim}{u}_m, \underset{\sim}{v})]^2 \Psi_I(\underset{\sim}{v}, \underset{\sim}{v}) d^2\underset{\sim}{v} \tag{4.5}$$

in the limit $s \rightarrow 0$, where $\Psi_I(\underset{\sim}{v}_1, \underset{\sim}{v}_2)$ is the spatial coherence function given by Eq. (2.11). Substituting from Eqs. (3.5) and (4.5) into Eq. (4.3), using Eq. (2.11), and averaging by means of Eq. (4.2) over the ensemble of radiance deviations $b(\underset{\sim}{u})$, we obtain for the mean square error

$$\mathscr{E} = C_2{}^2 (WT)^{-1} (\lambda R)^4 A^{-2} \iint L(\underset{\sim}{u}, \underset{\sim}{v}_1) L(\underset{\sim}{u}, \underset{\sim}{v}_2) \Big\{ B_0'^2 |\mathscr{I}(\underset{\sim}{v}_1 - \underset{\sim}{v}_2)|^2 +$$

$$A^2 (\lambda R)^{-4} \iint \varphi(\underset{\sim}{u}_1 - \underset{\sim}{u}_2) \mathscr{I}(\underset{\sim}{v}_1 - \underset{\sim}{u}_1) \mathscr{I}(\underset{\sim}{u}_1 - \underset{\sim}{v}_2) \mathscr{I}(\underset{\sim}{v}_2 - \underset{\sim}{u}_2) \mathscr{I}(\underset{\sim}{u}_2 - \underset{\sim}{v}_1)$$

$$\times\, d^2\underset{\sim}{u}_1 d^2\underset{\sim}{u}_2 \Big\} d^2\underset{\sim}{v}_1 d^2\underset{\sim}{v}_2$$

$$+ C_2 B_0'(\lambda R)^2 A^{-1} \int |L(\underset{\sim}{u}, \underset{\sim}{v})|^2 d^2\underset{\sim}{v} +$$

$$\rho^2 C_2{}^2 \iint L(\underset{\sim}{u}, \underset{\sim}{v}_1)\, L(\underset{\sim}{u}, \underset{\sim}{v}_2)\, |\mathscr{I}(\underset{\sim}{v}_1 - \underset{\sim}{w}_1)|^2\, |\mathscr{I}(\underset{\sim}{v}_2 - \underset{\sim}{w}_2)|^2\, \varphi(\underset{\sim}{w}_1 - \underset{\sim}{w}_2) \times d^2\underset{\sim}{v}_1 d^2\underset{\sim}{v}_2 d^2\underset{\sim}{w}_1 d^2\underset{\sim}{w}_2$$

$$- 2\,\rho\, C_2 \int L(\underset{\sim}{u}, \underset{\sim}{v})\, |\mathscr{I}(\underset{\sim}{v} - \underset{\sim}{w})|^2\, \varphi(\underset{\sim}{w} - \underset{\sim}{u})\, d^2\underset{\sim}{v} d^2\underset{\sim}{w} + \varphi(\underset{\sim}{0}), \tag{4.6}$$

with $C_2 = C_1(R_1/R)^2$.

Because of the homogeneity of the object and of the optical imaging in our paraxial approximation, the weighting function $L(\underset{\sim}{u}, \underset{\sim}{v})$ that minimizes $\mathscr{E}$ will depend on $\underset{\sim}{u}$ and $\underset{\sim}{v}$ only through $\underset{\sim}{u} - \underset{\sim}{v}$. We introduce its spatial Fourier transform $\Lambda(\underset{\sim}{r})$, defined by

$$L(\underset{\sim}{u}, \underset{\sim}{v}) = (\lambda R)^{-4} A\, C_2^{-1} \int \Lambda(\underset{\sim}{r}) \exp\,[ik\underset{\sim}{r}\cdot(\underset{\sim}{u} - \underset{\sim}{v})/R]\, d^2\underset{\sim}{r}; \tag{4.7}$$

the transform variable $\underset{\sim}{r}$ is scaled so that it matches the coordinates in the aperture plane. Similarly we define the spatial spectral density of the radiance deviation $b(\underset{\sim}{u})$ through the Fourier integral

$$\varphi(\underset{\sim}{u}) = \int \Phi(\underset{\sim}{r}) \exp\,(ik\underset{\sim}{u}\cdot\underset{\sim}{r}/R)\, d^2\underset{\sim}{r}. \tag{4.8}$$

In terms of these the mean square error can be written as

$$\mathscr{E} = (WT)^{-1} \int |\Lambda(\underset{\sim}{r})|^2 \left[B_0'^2\, I_A^{(2)}(\underset{\sim}{r}) + \int I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s})\, \Phi(\underset{\sim}{s})\, d^2\underset{\sim}{s}\right] d^2\underset{\sim}{r}/A\, +$$

$$C \int |\Lambda(\underset{\sim}{r})|^2\, d^2\underset{\sim}{r}/A + \int |\rho\Lambda(\underset{\sim}{r})\, I_A^{(2)}(\underset{\sim}{r}) - 1|^2\, \Phi(\underset{\sim}{r})\, d^2\underset{\sim}{r}, \tag{4.9}$$

where $C = 4\pi\, B_0'/\lambda^2\alpha T$,

$$I_A^{(2)}(\underset{\sim}{r}) = \int_A I_A(\underset{\sim}{s})\, I_A(\underset{\sim}{s} - \underset{\sim}{r})\, d^2\underset{\sim}{s}/A \tag{4.10}$$

is the self-convolution of the indicator function of the aperture, and

$$I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s}) = \int I_A(\underset{\sim}{t})\, I_A(\underset{\sim}{t} + \underset{\sim}{r})\, I_A(\underset{\sim}{t} + \underset{\sim}{s})\, I_A(\underset{\sim}{t} + \underset{\sim}{r} + \underset{\sim}{s})\, d^2\underset{\sim}{t}/A \tag{4.11}$$

is a quadruple convolution. For a circular aperture of radius a, both $I_A^{(2)}(\underset{\sim}{r})$

and $I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s})$ vanish outside a circle of radius 2a. Their maximum values occur at $\underset{\sim}{r} = \underset{\sim}{s} = \underset{\sim}{0}$ and equal 1. The region where $I_A^{(2)}(\underset{\sim}{r}) \neq 0$ is called the convolved aperture.

The expression in Eq. (4.9) has the same form as the mean square error in Wiener filtering; $\Phi(\underset{\sim}{r})$ is the object spectral density, $\rho\, I_A^{(2)}(\underset{\sim}{r})$ the effective optical transfer function, and

$$\Phi_n(\underset{\sim}{r}) = 4\pi\, B_0'/\lambda^2\alpha TA + A^{-1}(WT)^{-1}\left[B_0'^2\, I_A^{(2)}(\underset{\sim}{r}) + \int I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s})\, \Phi(\underset{\sim}{s})\, d^2\underset{\sim}{s}\right] \qquad (4.12)$$

the equivalent spatial spectral density of the noise. In $\Phi_n(\underset{\sim}{r})$ the first term represents the shot noise due to the photoelectrons, and the second arises from the fluctuations of the light field. Thus the mean square error is

$$\mathscr{E} = \int |\Lambda(\underset{\sim}{r})|^2\, \Phi_n(\underset{\sim}{r})\, d^2\underset{\sim}{r} + \int |\rho\Lambda(\underset{\sim}{r})\, I_A^{(2)}(\underset{\sim}{r}) - 1|^2\, \Phi(\underset{\sim}{r})\, d^2\underset{\sim}{r}, \qquad (4.13)$$

in which the first integral gives the contribution of the noise and the second represents the mean square bias of the estimate, averaged over the ensemble of radiance distributions.

The filter transfer function $\Lambda(\underset{\sim}{r})$ that minimizes the mean square error is determined as previously described.[2] As usual in linear processing, spatial frequencies in the object that when multiplied by λR fall outside the convolved aperture are not restored to the image. When the aperture is so large that diffraction can be neglected, only the shot noise and the fluctuations of the light field prevent perfect image restoration. The equivalent noise density is then constant, and the minimum mean square error is

$$\mathscr{E}_{min} = \int [1 + H\, \Phi(\underset{\sim}{r})/\Phi(\underset{\sim}{0})]^{-1}\, \Phi(\underset{\sim}{r})\, d^2\underset{\sim}{r}, \qquad (4.14)$$

where the effective signal-to-noise ratio (snr) H is defined by

$$H = A\, \rho^2\, \Phi(\underset{\sim}{0})/\Phi_n(\underset{\sim}{0}) =$$

$$\rho^2 \, WT \, (\sigma_B{}^2/B_0'^2) \, (\mathscr{A}_o/\mathscr{A}_r) \left[1 + \frac{\sigma_B{}^2}{B_0'^2} + \frac{4\pi W}{\lambda^2 B_0' \alpha}\right]^{-1} , \tag{4.15}$$

with $\mathscr{A}_r = (\lambda R)^2/A$ the area of a resolution element in the object plane and

$$\mathscr{A}_o = \sigma_B{}^{-2} \int_0 \varphi(\underset{\sim}{u}) \, d^2\underset{\sim}{u}$$

an area of the order of that of typical details in the object; $\sigma_B{}^2/B_0'^2$ is the mean square contrast of the object, with $\sigma_B{}^2 = \varphi(\underset{\sim}{0})$. Table 1 gives the minimum relative mean square error for some simple object spectra. The more high frequencies attributed to the objects by $\Phi(\underset{\sim}{r})$, the more slowly $\mathscr{E}_{min}$ decreases with increasing snr H.

The shot noise predominates over the effect of the field fluctuations when

$$r_B = \lambda^2 B_0'\alpha/4\pi W = \lambda^2 B_0'\eta/4\pi W\hbar\Omega << 1,$$

where η is the quantum efficiency of the photosensitive surface. If we consider a scene illuminated by moonlight on a clear night, we can put for B_0'/W the value 1.6×10^{-3} watt$\cdot$m$^{-2}\mu^{-1}$ at $\lambda = 5150$ Å.U.[12] This ratio r_B then takes the value $0.8 \times 10^{-13}\ \eta$, which must be further diminished by the mean reflectivity of the scene. For illumination by full sunlight at the same wavelength the ratio r_B is larger than this by a factor of about 10^6. The quantum nature of light and its interaction with the recording medium is thus under most circumstances the principal hindrance to perfect image restoration.

5. Phase-Screen Turbulence

When the object light passes through a turbulent medium, the point-spread function (psf) $S_2(\underset{\sim}{r}, \underset{\sim}{u})$ appearing in Eq. (2.3) becomes a randomly time-varying function whose statistical properties derive from those of the turbulence. In order to assess the effect of the resulting distortion on the restorability of the received image, we take a simple model of the medium for which the average psf is readily calculated.

The turbulence is supposed to be confined to a thin, homogeneous planar region between object and aperture, and its only effect is postulated as the introduction of random phase shifts into the rays passing through it. This is called the random phase screen. As a result the spatial coherence function of the object light at the aperture acquires a factor[13]

$$p(\underset{\sim}{r}) = \exp\left[-\frac{1}{2} D(R'\underset{\sim}{r}/R)\right], \tag{5.1}$$

where $D(\underset{\sim}{v})$ is the structure function of the random phase shifts $\theta(\underset{\sim}{u})$ introduced at points $\underset{\sim}{u}$ of the screen,

$$D(\underset{\sim}{v}) = 2[\Theta(\underset{\sim}{0}) - \Theta(\underset{\sim}{v})],$$

$$\Theta(\underset{\sim}{v}) = \underset{\sim}{E}[\theta(\underset{\sim}{u})\ \theta(\underset{\sim}{u} + \underset{\sim}{v})].$$

Here R' is the distance from object to phase screen.

When in our image-processing system the integration time T is much longer than the time constant τ_θ of the phase fluctuations, which in turn is much greater than the reciprocal bandwidth W^{-1} of the received and filtered light, the factor $p(\underset{\sim}{r})$ simply multiplies the optical transfer function $\rho\ I_A^{(2)}(\underset{\sim}{r})$ in the last integral of Eq. (4.9). If there are many independent phase screens lying between object and aperture and separated by many wavelengths of the light, the exponent in Eq. (5.1) contains a sum of structure functions for the several screens; and if each screen introduces only a small, random phase shift, a good approximation to the net effect assigns to the factor $p(\underset{\sim}{r})$ a gaussian form,

$$p(\underset{\sim}{r}) = \exp(-\underset{\sim}{r}^2/4L^2), \tag{5.2}$$

where L is a characteristic length for the turbulence, whose distortions are the more severe, the smaller L is.[13] Alternatively, this gaussian form can be taken as an approximation for a single phase screen. It corresponds to the light from each point of the object arriving with plane wavefronts tilted at random angles, which have a gaussian distribution about perpendicularity to the optical axis.

The turbulence also affects the term in the mean square error, Eq. (4.9), containing the quadruple convolution $I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s})$. This term arises from the product of four psf's

$$S_2(\underset{\sim}{r}_1, \underset{\sim}{v}_1)\ S_2^*(\underset{\sim}{r}_2, \underset{\sim}{v}_1)\ S_2^*(\underset{\sim}{r}_3, \underset{\sim}{v}_2)\ S_2(\underset{\sim}{r}_4, \underset{\sim}{v}_2)$$

occurring in the variance Var $\hat{b}(\underset{\sim}{u})$ of the estimate. It must now be averaged over the ensemble of randomly fluctuating psf's S_2 resulting from the turbulence. A tedious calculation making use of the inequalities $W^{-1} \ll \tau_\theta \ll T$ replaces $I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s})$ by a new function $I_A^{(\theta)}(\underset{\sim}{r}, \underset{\sim}{s})$ involving the structure function of the random phase shifts,

$$I_A^{(\theta)}(\underset{\sim}{r}, \underset{\sim}{s}) = (\lambda R)^{-2} A^{-1} \int \cdots \int \mathscr{F}\left(\frac{R'}{R}\underset{\sim}{s}, \frac{R''}{R}\underset{\sim}{t}\right)$$
$$\times \exp\left[\frac{ik}{R}\left(\frac{R'}{R''}\underset{\sim}{r} - \underset{\sim}{t}\right)\cdot(\underset{\sim}{r}_1 - \underset{\sim}{r}_2 - \underset{\sim}{s})\right] I_A(\underset{\sim}{r}_1)\ I_A(\underset{\sim}{r}_2)\ I_A(\underset{\sim}{r}_1 + \underset{\sim}{r})\ I_A(\underset{\sim}{r}_2 + \underset{\sim}{r})$$
$$\times\ d^2\underset{\sim}{r}_1 d^2\underset{\sim}{r}_2 d^2\underset{\sim}{t}, \tag{5.3}$$

where R" is the distance from phase screen to aperture, R = R' + R", and

$$\mathscr{F}(\underset{\sim}{p}, \underset{\sim}{q}) = \exp\{-[D(\underset{\sim}{p}) + D(\underset{\sim}{q}) - \tfrac{1}{2}D(\underset{\sim}{q} - \underset{\sim}{p}) - \tfrac{1}{2}D(\underset{\sim}{q} + \underset{\sim}{p})]\}. \tag{5.4}$$

For a quadratic approximation to the structure function as in Eq. (5.2), $D(\underset{\sim}{p}) \propto \underset{\sim}{p}^2$, $\mathscr{F} \equiv 1$ and $I_A^{(\theta)}(\underset{\sim}{r}, \underset{\sim}{s}) = I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s})$ of Eq. (4.11). Furthermore, when the aperture is so large that diffraction imposes no significant limitations,

$$I_A^{(\theta)}(\underset{\sim}{r}, \underset{\sim}{s}) \cong \mathscr{F}\left(\frac{R'}{R}\,\underset{\sim}{r}, \frac{R'}{R}\,\underset{\sim}{s}\right). \tag{5.5}$$

The function $\mathscr{F}$ can be shown to be at most equal to 1. Here we shall for simplicity set $\mathscr{F} = 1$ and $I_A^{(\theta)}(\underset{\sim}{r}, \underset{\sim}{s}) = I_A^{(4)}(\underset{\sim}{r}, \underset{\sim}{s})$, thus replacing the integral involving this kernel with its upper bound; the term involved is usually much smaller than the shot-noise term anyhow.

We now assume that the diameter of the aperture is so much greater than the distortion length L that we can neglect diffraction of the image. The minimum mean square error is then, after suitable modification of Eq. (4.9) and subsequent equations,

$$\mathscr{E}_{\min} = \int [1 + H|p(\underset{\sim}{r})|^2\, \Phi(\underset{\sim}{r})/\Phi(\underset{\sim}{0})]^{-1}\, \Phi(\underset{\sim}{r})\, d^2\underset{\sim}{r}, \tag{5.6}$$

where H is the snr defined in Eq. (4.15). As an example we assign to the spectral density of the radiance distribution the constant bandlimited form

$$\begin{aligned} \Phi(\underset{\sim}{r}) &= \sigma_B{}^2/\pi\varepsilon^2, & |\underset{\sim}{r}| &< \varepsilon, \\ &= 0, & |\underset{\sim}{r}| &> \varepsilon, \end{aligned} \tag{5.7}$$

where with Δ the diameter of a typical detail in the object, $\varepsilon = \lambda R/\Delta$. For the optical transfer function $p(\underset{\sim}{r})$ we use the gaussian form in Eq. (4.2). We then find for the minimum relative mean square error

$$\mathscr{E}_{\min}/\sigma_B{}^2 = 2(L/\varepsilon)^2\, \ell n\, \{[H + \exp\,(\varepsilon^2/2L^2)]/(H + 1)\},$$

which has been plotted versus H in Fig. 2 for various values of $\varepsilon/L = \ell/\Delta$, where ℓ is the scale of the phase-screen turbulence projected onto the object plane, $\ell = \lambda R/L$.

For comparison we present in Fig. 3 the minimum relative mean square error for the object spectral density

$$\Phi(\underset{\sim}{r}) = \Phi(\underset{\sim}{0})\ \varepsilon^4\ (\underset{\sim}{r}^2 + \varepsilon^2)^{-2}, \tag{5.8}$$

which extends to high spatial frequencies. The ratio $\mathscr{E}_{min}/\sigma_B{}^2$, which was calculated numerically, decreases with H much more slowly than for the spectral density in Eq. (5.7).

Acknowledgment

I wish to thank Mr. Yie-Ming Hong for carrying out the numerical calculations.

References

1. J. L. Harris, Sr., J. Opt. Soc. Am. **56**, 569 (1966).
2. C. W. Helstrom, J. Opt. Soc. Am. **57**, 297 (1967).
3. D. Slepian, J. Opt. Soc. Am. **57**, 918 (1967)
4. D. Slepian, Bell System Tech. J. **46**, 2353 (1967).
5. B. R. Frieden, J. Opt. Soc. Am. **57**, 1013 (1967).
6. C. K. Rushforth and R. W. Harris, J. Opt. Soc. Am. **58**, 539 (1968).
7. J. E. Falk, SIAM J. Appl. Math. **17**, 582 (1969).
8. C. W. Helstrom, J. Opt. Soc. Am. **60**, 1608 (1970).
9. R. Hanbury Brown and R. Q. Twiss, Proc. Roy. Soc. (London) **A242**, 300 (1957).
10. H. Gamo, J. Opt. Soc. Am. **56**, 441 (1966).
11. I. S. Reed, Trans. IEEE **IT-8**, 194 (1962).
12. A. Boileau, Scripps Institution of Oceanography (private communication).
13. C. W. Helstrom, J. Opt. Soc. Am. **59**, 331 (1969).

Table 1

Minimum Relative Mean Square Error

Object Spectral Density	$\mathscr{E}_{min}/\sigma_B^2$
$\Phi(\underset{\sim}{r}) = \Phi(\underset{\sim}{0}), \quad \lvert\underset{\sim}{r}\rvert < \varepsilon$ $\quad = 0, \quad \lvert\underset{\sim}{r}\rvert > \varepsilon$	$(1 + H)^{-1}$
$\Phi(\underset{\sim}{r}) = \Phi(\underset{\sim}{0}) \exp(-\underset{\sim}{r}^2/2\varepsilon^2)$	$H^{-1} \ln(1 + H)$
$\Phi(\underset{\sim}{r}) = \Phi(\underset{\sim}{0})\, \varepsilon^4 (\underset{\sim}{r}^2 + \varepsilon^2)^{-2}$	$H^{-\frac{1}{2}} \tan^{-1}(H^{\frac{1}{2}})$
$\Phi(\underset{\sim}{r}) = \Phi(\underset{\sim}{0})\, \varepsilon^n (\lvert\underset{\sim}{r}\rvert^n + \varepsilon^n)^{-1}, \; n > 2$	$(1 + H)^{-(n-2)/n}$

Figure Captions

Fig. 1. The object and the image-processing system: O = object plane, A = aperture plane, I = image plane. A narrowband temporal filter for object and background light is not shown.

Fig. 2. Minimum relative mean square error for restoring image distorted by a random phase screen. Object spectral density $\Phi(\underset{\sim}{r}) = \Phi(\underset{\sim}{0})$, $|\underset{\sim}{r}| < \varepsilon$; $\Phi(\underset{\sim}{r}) = 0$, $|\underset{\sim}{r}| > \varepsilon$. Curves are indexed by the value of $\varepsilon/L = \ell/\Delta$.

Fig. 3 Minimum relative mean square error for restoring image distorted by a random phase screen. Object spectral density $\Phi(\underset{\sim}{r}) = \Phi(\underset{\sim}{0})\ \varepsilon^4\ (\underset{\sim}{r}^2 + \varepsilon^2)^{-2}$. Curves are indexed by the value of $\varepsilon/L = \ell/\Delta$.

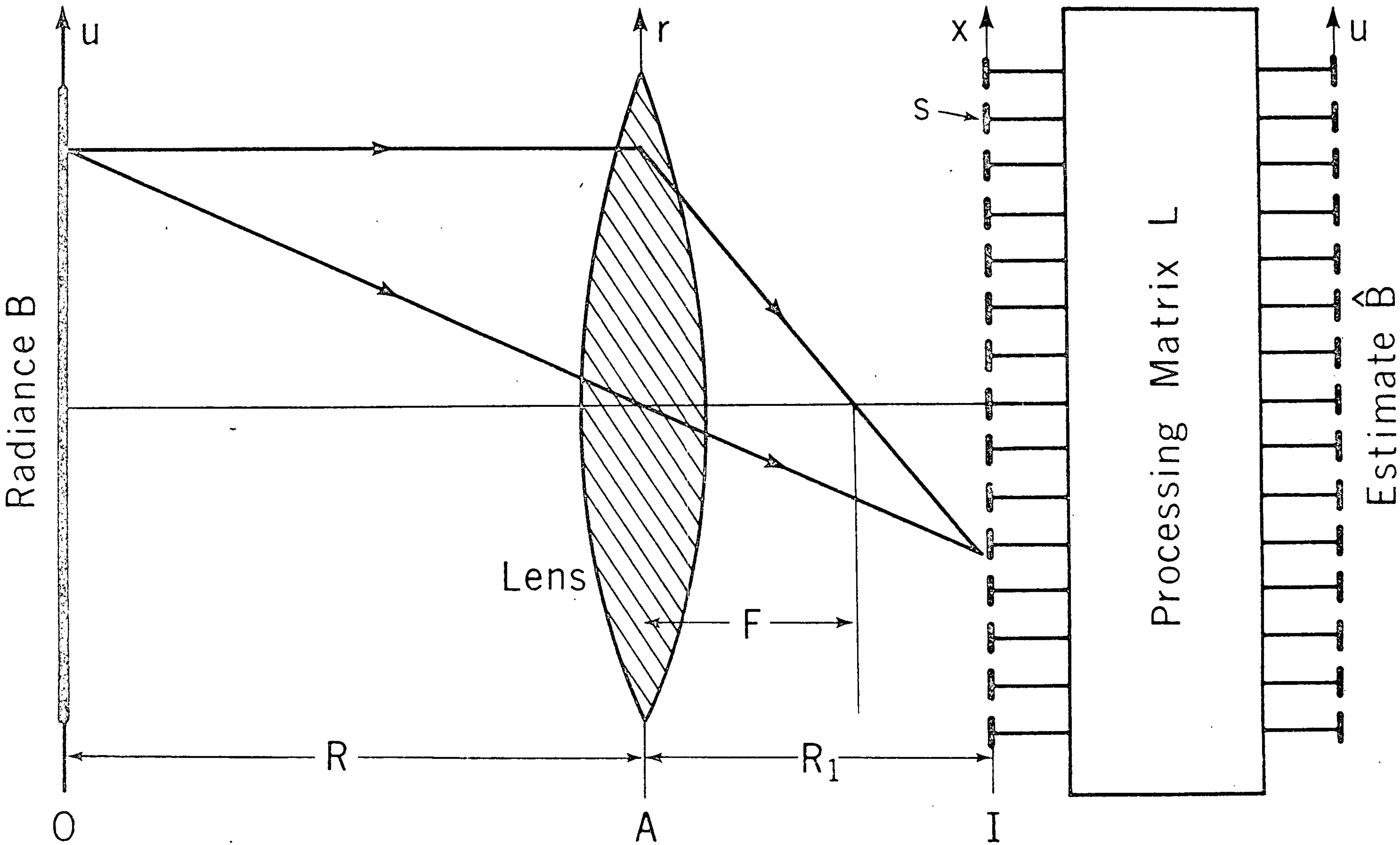
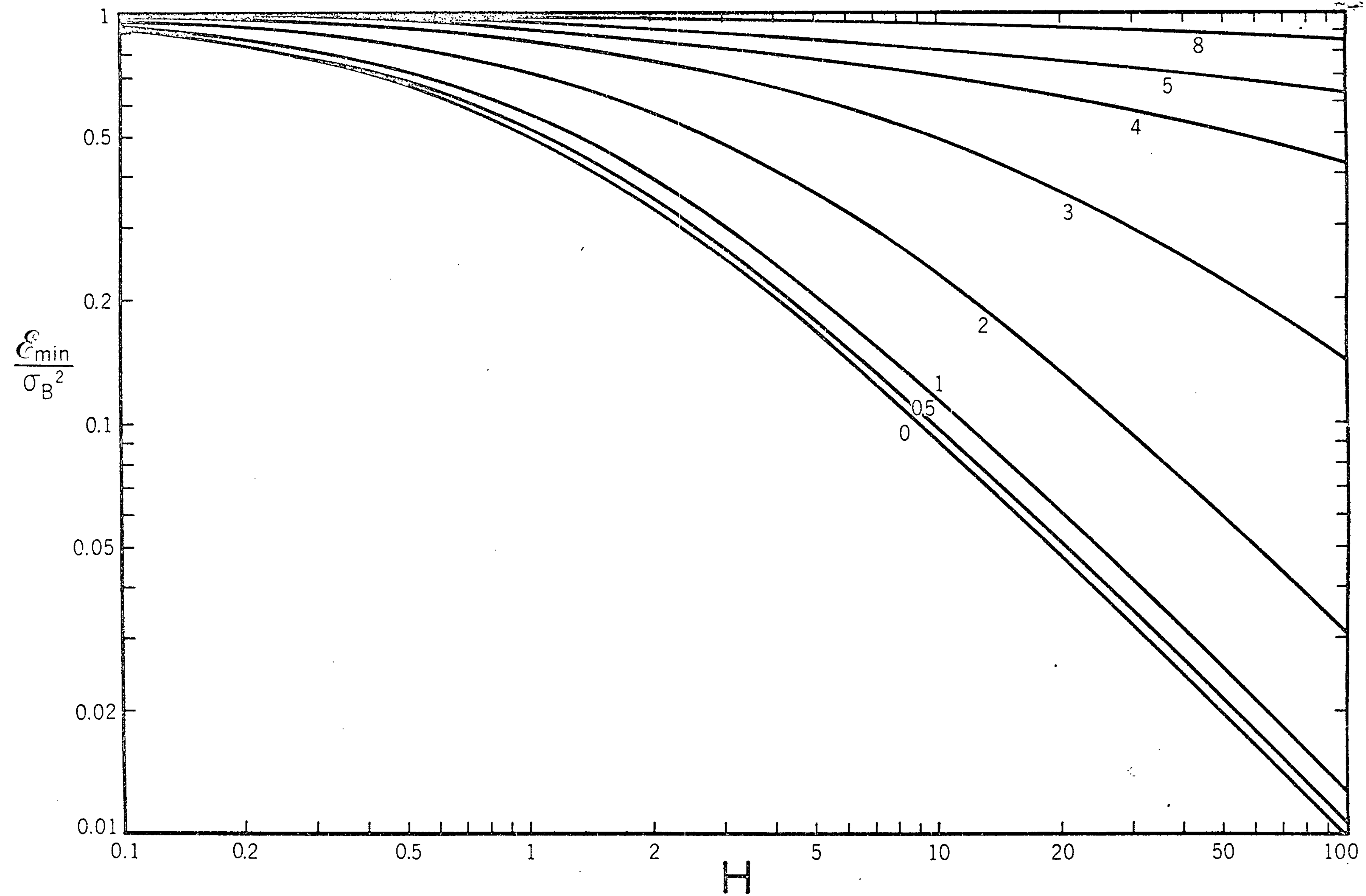
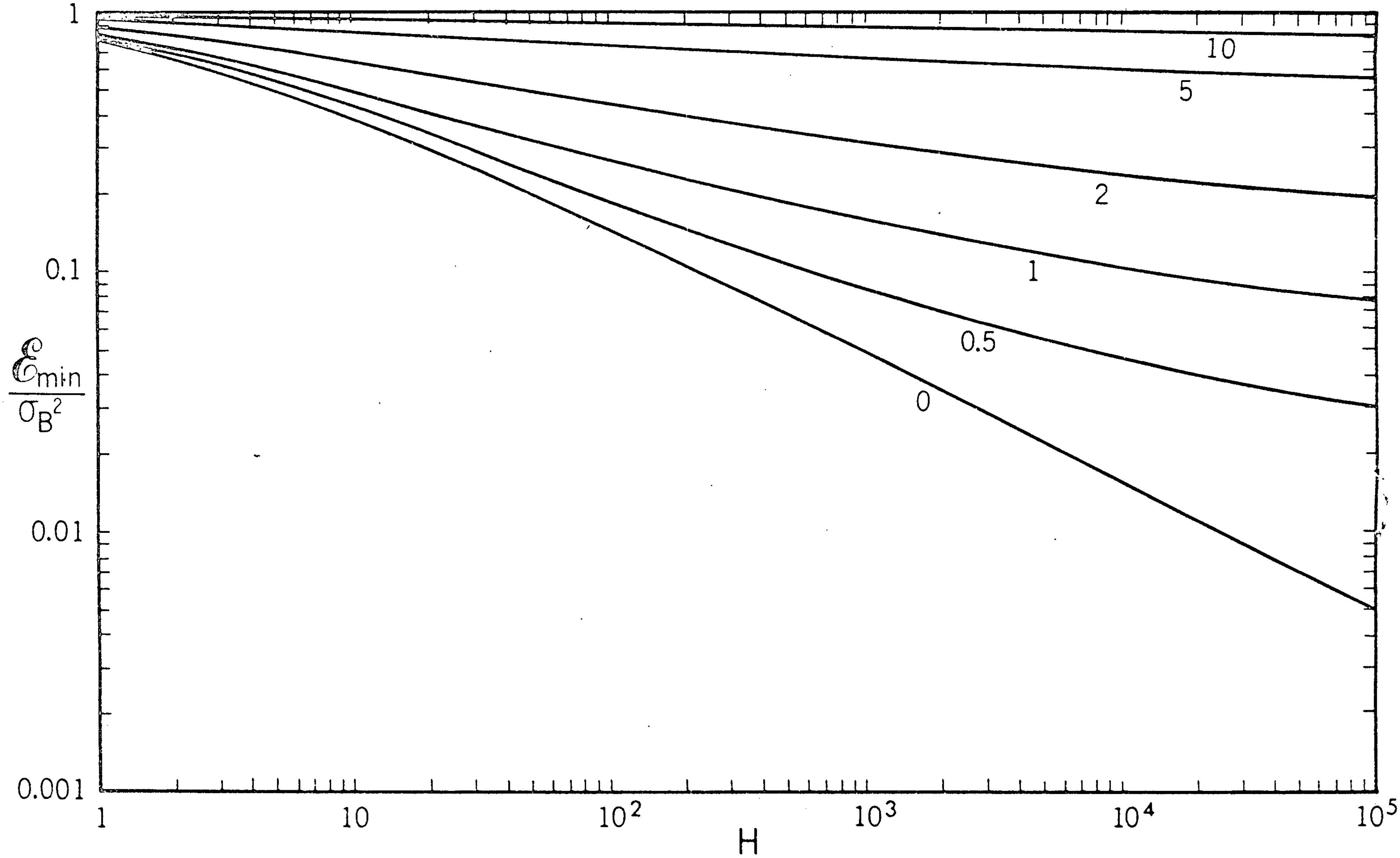

u
r
x
u
S
Radiance B
Lens
F
Processing Matrix L
Estimate $\hat{B}$
R
R_1
O
A
I

Figure 1



1
0.5
0.2
0.1
0.05
0.02
0.01
$\frac{\mathcal{E}_{min}}{\sigma_B^2}$
8
5
4
3
2
1
05
0
0.1
0.2
0.5
1
2
5
10
20
50
100
H

Figure 2



1
0.1
0.01
0.001
$\frac{\mathcal{E}_{min}}{\sigma_B^2}$
10
5
2
1
0.5
0
1
10
10^2
10^3
10^4
10^5
H

Figure 3

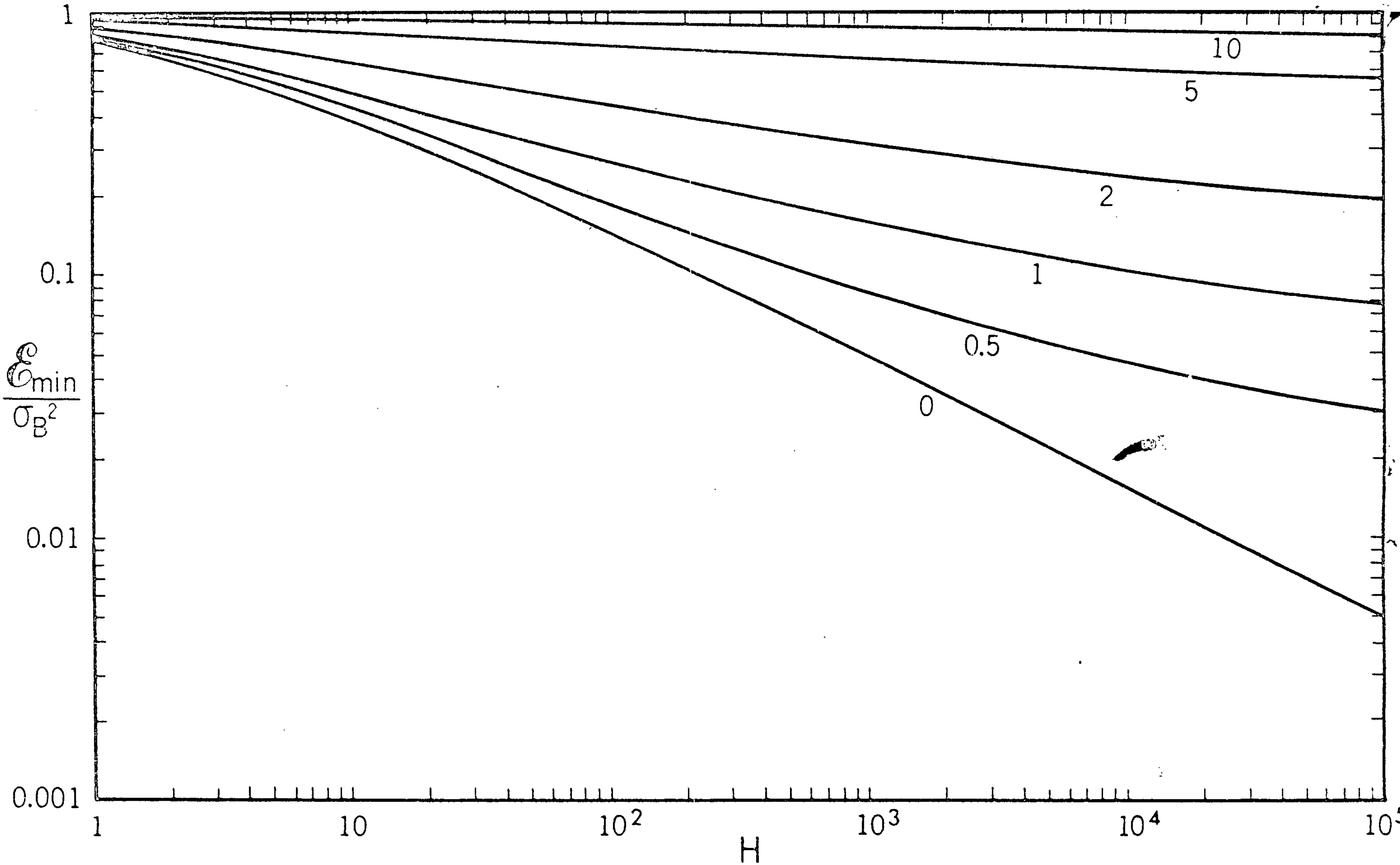

1
0.1
0.01
0.001
$\frac{\mathcal{E}_{min}}{\sigma_B^2}$
1
10
10^2
10^3
10^4
10^5
H
10
5
2
1
0.5
0

Figure 3