JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 57, NUMBER 3 MARCH 1967

Image Restoration by the Method of Least Squares

CARL W. HELSTROM*
Westinghouse Research Laboratories, Pittsburgh, Pa. 15235

(Received 15 August 1966)

The restoration of optical images, as well as the unfolding of spectroscopic and other data that have been convolved with a window function or an instrumental impulse response, can be viewed as the solution of an integral equation. Solution of such an integral equation when the data are corrupted by noise or experimental error is treated as the problem of finding an estimate that is a linear functional of the data and minimizes the mean squared error between the true solution and itself. The estimate depends on assumptions about the spectral densities of the images and the noise, the choice of which is discussed. Coherent optical processing and digital processing are described.

INDEX HEADINGS: Image restoration; Fourier transform.

1. IMAGE RESTORATION AS THE SOLUTION OF AN INTEGRAL EQUATION

THE degradation suffered by images in optical systems can often be described as a convolution of the true or geometrical image with a spread function.[1] Let $J_0(\mathbf{r})$ be the illuminance at a point $\mathbf{r}=(x,y)$ of the image plane if there were no degradation, and $J(\mathbf{r})$ be the actual illuminance. Then at least over a limited region, or "isoplanatism-patch,"[2]

$$J(\mathbf{r})=\int S(\mathbf{r}')J_0(\mathbf{r}-\mathbf{r}')d^2\mathbf{r}', \tag{1.1}$$

where $S(\mathbf{r})$ is the spread function and $d^2\mathbf{r}'=dx'dy'$ the differential of area. (All integrals can be taken over the infinite plane if the integrands are properly defined.)

* Present address: Department of Applied Electrophysics, University of California, San Diego, La Jolla, California 92037.

[1] E. L. O'Neill, *Introduction to Statistical Optics* (Addison-Wesley Publishing Co., Reading, Mass., 1963), p. 20.

[2] P. B. Fellgett and E. H. Linfoot, Phil. Trans. Roy. Soc. **A247**, 369 (1955).

It is convenient to normalize the spread function so that

$$\int S(\mathbf{r})d^2\mathbf{r}=1. \tag{1.2}$$

If the system were perfect, the spread function would be a two-dimensional delta-function, $S(\mathbf{r})=\delta(\mathbf{r})$, and $J(\mathbf{r})$ would equal $J_0(\mathbf{r})$.

Other impairments of an image can also be expressed as convolutions. If an image moving with velocity v in the x direction is recorded photographically in an exposure of duration T, the transmittance of the developed film is a function of an integrated illuminance given in Eq. (1.1), where the spread function can be taken as

$$\begin{aligned} S(\mathbf{r})=S(x,y)&=(vT)^{-1}\delta(y), \quad -\tfrac{1}{2}vT\leq x\leq\tfrac{1}{2}vT, \\ &=0, \quad |x|>\tfrac{1}{2}vT. \end{aligned} \tag{1.3}$$

The restoration of the true image $J_0(\mathbf{r})$ can be considered as the solution of the integral Eq. (1.1), whose kernel is the spread function $S(\mathbf{r})$, assumed known. The method that first presents itself is to apply the two-dimensional Fourier transform. Putting[3]

$$\begin{aligned} j(\boldsymbol{\omega})&=\int J(\mathbf{r})\exp(-i\boldsymbol{\omega}\cdot\mathbf{r})d^2\mathbf{r}, \\ s(\boldsymbol{\omega})&=\int S(\mathbf{r})\exp(-i\boldsymbol{\omega}\cdot\mathbf{r})d^2\mathbf{r}, \end{aligned} \tag{1.4}$$

we can write the solution of the integral equation as

$$\begin{aligned} J_0(\mathbf{r})&=\int j_0(\boldsymbol{\omega})\exp(i\boldsymbol{\omega}\cdot\mathbf{r})d^2\boldsymbol{\omega}/(2\pi)^2, \\ j_0(\boldsymbol{\omega})&=j(\boldsymbol{\omega})/s(\boldsymbol{\omega}). \end{aligned} \tag{1.5}$$

This method often does not work because the illuminance $J(\mathbf{r})$ cannot be measured precisely enough. Indeed, the illuminance $J(\mathbf{r})$ actually observed is given not by Eq. (1.1), but by

$$J(\mathbf{r})=\int S(\mathbf{r}')J_0(\mathbf{r}-\mathbf{r}')d^2\mathbf{r}'+N(\mathbf{r}), \tag{1.6}$$

where $N(\mathbf{r})$ represents experimental error, background radiation, and, in photography, the effect of granularity of the emulsion. We can encompass all these in the term spatial noise. The noise $N(\mathbf{r})$ cannot be known in advance and differs from experiment to experiment.

Most optical spread functions attenuate high spatial frequencies; $|s(\boldsymbol{\omega})|\rightarrow 0$ as $|\boldsymbol{\omega}|\rightarrow\infty$. An instructive analogy is the passage of a signal through a low-pass filter. To restore the original signal, the output must be passed through the inverse filter, which amplifies the high frequencies. By this time, however, noise has inevitably been added, and it contains frequencies far beyond those of the signal. These noise frequencies are so strengthened by the inverse filter as to blot out the desired restoration.

An attempt to treat the optical problem by Fourier transforms as in Eq. (1.5) is frustrated in the same way. A further difficulty arises when, as with the kernel in Eq. (1.3), the Fourier transform $s(\boldsymbol{\omega})$ of the spread function contains zeros at finite values of $\boldsymbol{\omega}$. Because of the noise, the Fourier transform $j(\boldsymbol{\omega})$ does not vanish at the same values of $\boldsymbol{\omega}$, and the quotient $j(\boldsymbol{\omega})/s(\boldsymbol{\omega})$ acquires troublesome infinities.

The solution of such integral equations involving experimental data is a common problem in physics. It arises in nuclear and optical spectroscopy, in photometry, in astronomy, and indeed wherever a finite instrumental response or entrance window convolves with the quantity to be measured. A method of avoiding certain of the difficulties in optics has recently been demonstrated by Harris.[4] Other approaches have been given by Trumpler and Weaver,[5] Kahn,[6] Phillips,[7] and Twomey.[8,9] Here we propose a method that makes use of the statistical properties of the noise. It is based on the theory of minimum-mean-square prediction and estimation.

2. AN ESTIMATED SOLUTION OF THE INTEGRAL EQUATION

With the spatial noise $N(\mathbf{r})$ unknown, Eq. (1.6) cannot be solved directly; the most that can be done is to produce an estimate $\hat{J}_0(\mathbf{r})$ of the solution. If both the image illuminance $J_0(\mathbf{r})$ and the noise $N(\mathbf{r})$ are viewed as spatial stochastic processes, the best estimate is naturally defined as the one that maximizes the posterior probability density of $J_0(\mathbf{r})$, given $J(\mathbf{r})$, as determined by Bayes's rule.[10]

The processes $J_0(\mathbf{r})$ and $N(\mathbf{r})$ are independent. If they were also gaussian, with zero means and known covariances, the posterior probability density would be greatest for the linear estimate

$$\hat{J}_0(\mathbf{r})=\int M(\mathbf{r}')J(\mathbf{r}-\mathbf{r}')d^2\mathbf{r}', \tag{2.1}$$

in which the estimating kernel $M(\mathbf{r})$ is chosen so as to

[3] In general we shall use a lower-case letter to denote the Fourier transform of a spatial function designated by the corresponding upper-case letter.

[4] J. L. Harris, Sr., J. Opt. Soc. Am. 56, 569 (1966).

[5] R. J. Trumpler and H. F. Weaver, *Statistical Astronomy* (University of California Press, Berkeley, Calif., 1953), Ch. 1.4, p. 95 ff.

[6] F. D. Kahn, Proc. Cambridge Phil. Soc. 51, 519 (1955).

[7] D. L. Phillips, J. Assoc. Comp. Mach. 9, 84 (1962).

[8] S. Twomey, J. Assoc. Comp. Mach. 10, 97 (1963).

[9] S. Twomey, J. Franklin Inst. 279, 95 (1965).

[10] For continuous processes such as these, a probability density must be treated by a limiting procedure as samples of the process are taken closer and closer together in the plane. The assertions of the text remain valid.

minimize the mean squared error

$$E=\mathbf{E}[\hat{J}_0(\mathbf{r})-J_0(\mathbf{r})]^2 \tag{2.2}$$

between the estimate and the true illuminance, $\mathbf{E}$ denoting an expected value.[11]

Although the noise may often be at least approximately treated as a gaussian process, the images to be restored rarely fall into that category. Nevertheless, if an estimator is optimum for a class of gaussian processes whose structure resembles that of the finest images to be encountered, it can be expected to be effective when applied to these images as well. It is admitted that a detailed statistical specification of the class of images anticipated would lead, if the optimization problem could be solved, to a superior estimator; but it would almost certainly be more complicated than the linear estimator of Eq. (2.1), and of restricted applicability besides. The linear estimator that minimizes the mean-squared error, Eq. (2.2), therefore recommends itself as an initial step in the search for good image restoration.

The integral in Eq. (2.1) may be said to represent the action of a restoring filter, and such a linear estimate may be indeed be realizable by coherent optical processing.[12,13] If the illuminances are sampled at discrete points, a discrete counterpart obtained by approximating the integrals in Eqs. (1.6) and (2.1) by summation formulas can be executed by a digital computer. The estimating kernel $M(\mathbf{r})$ depends on the spread function $S(\mathbf{r})$ and on the properties of the images and the noise. We turn to the problem of calculating it.

Of the limited region in which the degradation of the image can be described by Eq. (1.1) there is no reason to favor one part over another, and the stochastic processes representing the images $J_0(\mathbf{r})$ and the noise $N(\mathbf{r})$ may as well be taken as spatially stationary. Since it is chiefly the deviation

$$\Delta J_0(\mathbf{r})=J_0(\mathbf{r})-\mathbf{E}J_0(\mathbf{r})$$

from an over-all mean illuminance that is of interest, the representative processes can without loss be taken to have zero mean values.

The mean value $\mathbf{E}J_0(\mathbf{r})$, which for a spatially ergodic process is equal to the limit as A goes to infinity of the spatial average $A^{-1}\int_A J_0(\mathbf{r})d^2\mathbf{r}$, over a region of area A, does not usually contribute to the identification of an image, and we need not be concerned with its estimation. With the mean value of the noise $N(\mathbf{r})$ taken as 0, however, a good estimate of the mean of $J_0(\mathbf{r})$, in view of Eq. (1.2), is simply the integral of the observed illuminance $J(\mathbf{r})$ over a large region of the image plane, divided by the area of that region.

The independent, zero-mean gaussian processes $\Delta J_0(\mathbf{r})$ and $N(\mathbf{r})$ are completely characterized by their spatial covariance functions

$$\begin{aligned}\Phi_0(\mathbf{r})&=\mathbf{E}\Delta J_0(\mathbf{r}')\Delta J_0(\mathbf{r}'+\mathbf{r})\\ \Phi_n(\mathbf{r})&=\mathbf{E}N(\mathbf{r}')N(\mathbf{r}'+\mathbf{r})\end{aligned} \tag{2.3}$$

or by the Fourier transforms of these, $\varphi_0(\omega)$ and $\varphi_n(\omega)$, which are the spatial spectral densities of the image and noise processes, respectively. When Eqs. (1.6) and (2.1) are substituted into the error expression of Eq. (2.2) and expected values are formed, the mean-squared error becomes

$$\begin{aligned}E=\Phi_0(0)-2\int\int M(\boldsymbol{\tau})S(\mathbf{r})\Phi_0(\boldsymbol{\tau}+\mathbf{r})d^2\boldsymbol{\tau}d^2\mathbf{r}+\int\int M(\boldsymbol{\tau}_1)M(\boldsymbol{\tau}_2)\\ \left\{\int\int S(\mathbf{r}_1)S(\mathbf{r}_2)\Phi_0(\boldsymbol{\tau}_1+\mathbf{r}_1-\boldsymbol{\tau}_2-\mathbf{r}_2)d^2\mathbf{r}_1d^2\mathbf{r}_2+\Phi_n(\boldsymbol{\tau}_1-\boldsymbol{\tau}_2)\right\}d^2\boldsymbol{\tau}_1d^2\boldsymbol{\tau}_2\\ =\int\{\varphi_0(\omega)-[m(\omega)s(\omega)+m^*(\omega)s^*(\omega)]\varphi_0(\omega)+|m(\omega)|^2[|s(\omega)|^2\varphi_0(\omega)+\varphi_n(\omega)]\}d^2\omega/(2\pi)^2.\end{aligned} \tag{2.4}$$

This can be written, as in the usual treatment of Wiener filtering,[14,15]

$$E=\int\left\{\left|(m\omega)-\frac{s^*(\omega)\varphi_0(\omega)}{\gamma(\omega)}\right|^2\gamma(\omega)+\varphi_0(\omega)\varphi_n(\omega)/\gamma(\omega)\right\}d^2\omega/(2\pi)^2,$$

with

$$\gamma(\omega)=|s(\omega)|^2\varphi_0(\omega)+\varphi_n(\omega),$$

from which it is evident that the Fourier transform $m(\omega)$ of the estimating kernel that minimizes the mean-squared error E is

$$m(\omega)=[|s(\omega)|^2\varphi_0(\omega)+\varphi_n(\omega)]^{-1}s^*(\omega)\varphi_0(\omega). \tag{2.5}$$

The minimum mean-squared error attained by this estimator $M(\mathbf{r})$ is then

$$E_{\min}=\int[|s(\omega)|^2\varphi_0(\omega)+\varphi_n(\omega)]^{-1}\times\varphi_0(\omega)\varphi_n(\omega)d^2\omega/(2\pi)^2. \tag{2.6}$$

[11] D. Middleton, *An Introduction to Statistical Communication Theory* (McGraw-Hill Book Co., New York, N. Y., 1960), Section 21.4, p. 994 ff.

[12] L. J. Cutrona, E. N. Leith, C. J. Palermo, and L. J. Porcello, Trans. IRE **IT-6**, 386 (1960).

[13] A. Vander Lugt, Trans. IEEE **IT-10**, 139 (1964).

[14] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series* (John Wiley & Sons, Inc., New York, N. Y., 1949), p. 84.

[15] H. W. Bode and C. E. Shannon, Proc. IRE 38, 417 (1950).

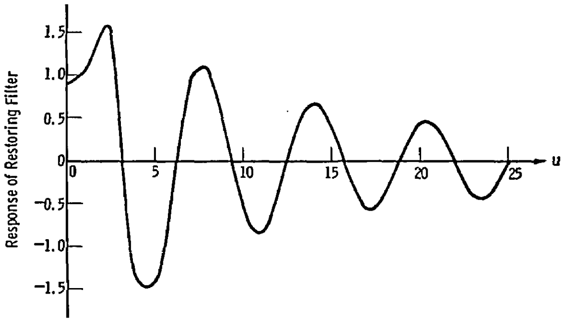

FIG. 1. Restoring filter for a moving image. $u=\frac{1}{2}vT\omega_x$.

There remains the question of what covariance functions or spectral densities to adopt for the images $J_0(\mathbf{r})$ and the spatial noise $N(\mathbf{r})$. For the images there is usually a certain distance δ representing the size of the finest details that should be resolved. The smaller this dimension δ, the larger the minimum error $E_{\min}$. Nothing is to be gained by adopting a value of δ smaller than the distances over which the features of the images $J_0(\mathbf{r})$ change significantly. Thus δ corresponds to a correlation distance for the stochastic processes representing the images. Their covariance function $\Phi_0(\mathbf{r})$ has widths in x and y of the order of δ, and their spectral density $\varphi_0(\omega)$ widths in each direction of the order of $2\pi/\delta$. Since little more than this can usually be said about the images, a plausible choice for the spectral density $\varphi_0(\omega)$ is simply

$$\begin{aligned}\varphi_0(\omega)&=4\pi\sigma^2/W^2, \quad |\omega|\leq W=2\pi/\delta\\ &=0, \quad |\omega|>W,\end{aligned} \tag{2.7}$$

where $\sigma^2=\mathrm{Var}\,J_0(\mathbf{r})$ is the mean-squared deviation of the image illuminances from their average value.

The noise $N(\mathbf{r})$ usually has finer structure than the images to be restored. Its correlation distance is much smaller than δ and its spectral density $\varphi_n(\omega)$ much broader than $\varphi_0(\omega)$. In most cases there is little reason to take $\varphi_n(\omega)$ as anything but constant, $\varphi_n(\omega)\equiv\varphi_n$. This choice corresponds to assuming that the variations $N(\mathbf{r})$ are like what the communications engineer calls white noise. Measurements of the illuminance $J(\mathbf{r})$ however close together have components $N(\mathbf{r})$ that are statistically independent, as is to be expected when experimental error and stray incoherent background radiation are taken into account.

The spatial frequency response $m(\omega)$ of the restoring filter depends on the ratio φ_n/φ_0, which is now a constant, but a rough estimate of this ratio usually suffices. The spectral density φ_n of the noise can be estimated after substituting for the images $J_0(\mathbf{r})$ a completely blank source producing the same average illuminance. A large number of measurements are then made of the integrated illuminance

$$J_T=\int_A J(\mathbf{r})d^2\mathbf{r}$$

over separate regions of area A whose diameters are much greater than the correlation distance of the spatial noise. The sample variance of the values of J_T so obtained provides an estimate of the product $A\varphi_n$.

As an example we show in Fig. 1 the spatial frequency response $m(\omega)$ of the restoring filter for an image moving with velocity v in the x direction. The kernel of the integral equation is given in Eq. (1.3); its Fourier transform is

$$s(\omega)=(\sin u)/u, \quad \omega=(\omega_x,\omega_y), \quad u=\tfrac{1}{2}vT\omega_x. \tag{2.8}$$

Since here the spreading occurs only in the x direction, it is reasonable to cut off the spectral density $\varphi_0(\omega)$ of the representative image processes only in the ω_x direction and to put

$$\begin{aligned}\varphi_0(\omega)&=\pi\sigma^2/W, \quad -W\leq\omega_x\leq W,\\ &=0, \quad |\omega_x|>W.\end{aligned} \tag{2.9}$$

Any distortion in the y direction due to the finite aperture of the imaging system is left unamended. The restoring filter now has the spatial-frequency response

$$\begin{aligned}m(\omega)&=u(\sin u)/(\sin^2 u+Bu^2), \quad |\omega_x|\leq W,\\ &=0, \qquad \omega_x>W,\\ u&=\tfrac{1}{2}vT\omega_x, \quad B=W\varphi_n/\pi\sigma^2,\end{aligned} \tag{2.10}$$

which is plotted vs u in Fig. 1 for a ratio $B=0.1$. It is

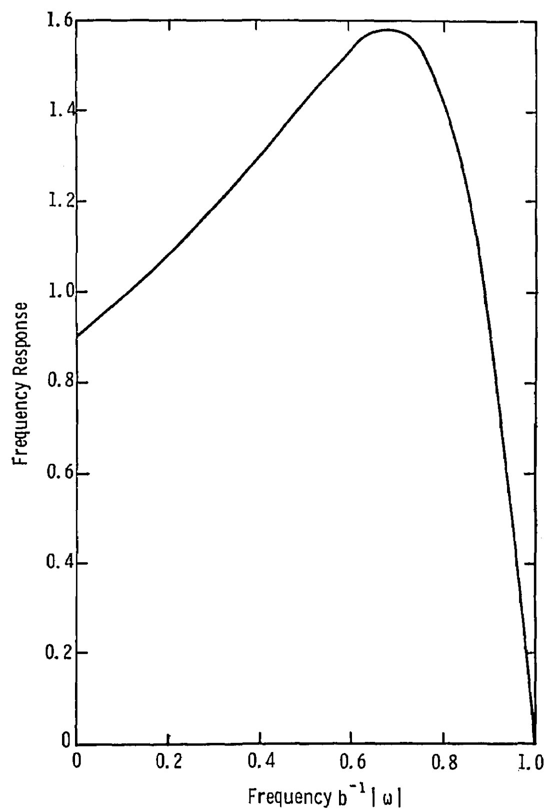

FIG. 2. Restoring filter for a slit-diffracted image.

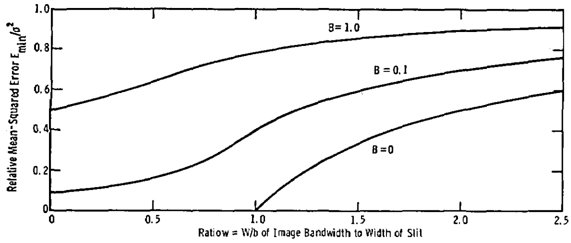

FIG. 3. Minimum mean squared error in restoration: slit diffraction.

symmetrical about $u=0$ and is to be cut off at a value of u equal to $\frac{1}{2}vTW$.

The restoration of an image that has suffered diffraction at a slit can also be viewed from the present standpoint. Let incoherent light be passing through an infinite slit of width a from an object that can be supposed at infinity. The image is then focused on a plane at a distance F from a cylindrical lens. Then if $b=2\pi a/\lambda F$, where λ is the wavelength of the light, the spread function is

$$S(\mathbf{r})=2(\pi b x^2)^{-1}\sin^2(bx/2)\delta(y). \tag{2.11}$$

Its Fourier transform is

$$\begin{aligned} s(\omega)&=1-b^{-1}|\omega_x|, \quad |\omega_x|\leqslant b,\\ &=0, \qquad\qquad |\omega_x|>b, \end{aligned} \tag{2.12}$$

which is the convolution with itself of a rectangular function of width b. Thus b can be identified with the optical bandwidth of the system.

If we take the noise as having a uniform spectral density φ_n and attribute to the geometrical image the same spectral density as in Eq. (2.10), the frequency response of the restoring filter is, by Eq. (2.5),

$$\begin{aligned} m(\omega)&=[(1-b^{-1}|\omega_x|)^2+B]^{-1}(1-b^{-1}|\omega_x|),\\ &\qquad\qquad |\omega_x|<\min(W,b),\\ &=0, \qquad\qquad |\omega_x|>\min(W,b)\\ B&=W\varphi_n/\pi\sigma^2. \end{aligned} \tag{2.13}$$

Figure 2 shows this response function for a value of $B=0.1$.

The minimum mean-squared error $E_{\min}$ is obtained by substituting $s(\omega)$ into the one-dimensional form of Eq. (2.6), which after integration yields

$$\begin{aligned} E_{\min}/\sigma^2&=(\beta/w)\{\tan^{-1}(\beta^{-1})-\tan^{-1}[\beta^{-1}(1-w)]\},\\ &\qquad\qquad 0<w\leqslant 1,\\ &=(\beta/w)[\tan^{-1}(\beta^{-1})+\beta^{-1}(w-1)], \quad w>1\\ w&=W/b, \quad \beta=B^{\frac{1}{2}}. \end{aligned} \tag{2.14}$$

This is plotted in Fig. 3 vs $w=W/b$ for $B=\beta^2=0$, 0.1, 1.0. For $W/b=0$, $E_{\min}/\sigma^2=B/(1+B)$. In the limit of no noise at all, $B\rightarrow 0$,

$$\begin{aligned} E_{\min}/\sigma^2&\rightarrow 0, \quad 0<W<b,\\ E_{\min}/\sigma^2&\rightarrow 1-b/W, \quad W>b. \end{aligned}$$

Thus if the width a of the slit is greater than $\lambda FW/2\pi$, the representative image processes can be restored perfectly in the absence of noise, but not if $a<\lambda FW/2\pi$.

The spatial response $m(\omega)$ of the restoring filter for circular diffraction looks much like that shown in Fig. 2, but it appears difficult to evaluate the minimum mean-squared error analytically.

3. THE DATA PROCESSING

Optical Filtering

The image might be restored optically by use of coherent light, as described by Cutrona *et al.*[12] and Vander Lugt.[13] A transparency must be prepared whose amplitude transmittance deviates from a reference value by an amount proportional to the illuminance $J(\mathbf{r})$. Plane parallel coherent light passes through the transparency and thence to a convergent lens. In the focal plane beyond the lens is a second transparency whose amplitude transmittance is proportional to the function $m(\omega)$ of Eq. (2.5). The coordinates $\mathbf{x}$ in the focal plane are related to ω by

$$\mathbf{x}=\lambda F\omega/2\pi,$$

where λ is the wavelength of the coherent light and F is the focal length of the lens.

The field amplitude of the light falling on this transparency is the Fourier transform $j(\omega)$ of $J(\mathbf{r})$, and the transmitted light has a field amplitude $m(\omega)j(\omega)$. This light then passes through a second lens whose focal plane coincides with that of the first. The emergent light has a field amplitude proportional to the estimate $\hat{J}_0(\mathbf{r})$, which can be obtained by allowing the light to fall on a photographic plate of known characteristics. The limitation of the spectral density $\varphi_0(\omega)$ of the images to a spatial band of radius $|\omega|=W$, as in Eq. (2.7), corresponds to placing an aperture stop in the common focal plane of the two lenses.

Unless $m(\omega)$ is real and positive, preparation of the filter transparency may be difficult. Holographic techniques[16] might be applied, however, by making the amplitude transmittance proportional to

$$\mu+m(\omega)\exp i\omega\cdot\mathbf{a}+m^*(\omega)\exp(-i\omega\cdot\mathbf{a}),$$

where μ is a constant large enough to prevent this expression from becoming negative, and where $\mathbf{a}$ is a suitable vector. The field amplitude of the transmitted light is then proportional to

$$j(\omega)[\mu+m(\omega)\exp i\omega\cdot\mathbf{a}+m^*(\omega)\exp(-i\omega\cdot\mathbf{a})];$$

after Fourier transformation by the second lens, the restored image is displaced by the vector $\mathbf{a}$ from a replica of the original image.

[16] G. W. Stroke, *An Introduction to Coherent Optics and Holography* (Academic Press, New York, London, 1966), p. 79ff.

Digital Processing

If the values of the illuminance $J(\mathbf{r})$ are sampled at a rectangular grid of points, and if the illuminance $J_0(\mathbf{r})$ of the restored image is desired only at the intersections of a similar rectangular grid, the integrations in Eqs. (1.6) and (2.1) are replaced by finite summations obtained by applying a quadrature formula such as Simpson's or the trapezoidal rule. Taking the points in each grid in a certain fixed order we can write the resulting equations in matrix form,

$$\mathbf{J}=\mathbf{S}\mathbf{J}_0+\mathbf{N},\quad \hat{\mathbf{J}}_0=\mathbf{M}\mathbf{J}, \tag{3.1}$$

where $\mathbf{J}$, $\mathbf{J}_0$, and $\mathbf{N}$ are column vectors of sample values, of the functions $J(\mathbf{r})$, $J_0(\mathbf{r})$, and $N(\mathbf{r})$, respectively. $\mathbf{S}$ and $\mathbf{M}$ are matrices and $\hat{\mathbf{J}}_0$ is a column vector of estimated illuminances at the grid points on the restored image. The number n of points at which the illuminance $J(\mathbf{r})$ is measured is generally larger than the number m of points at which the illuminance $J_0(\mathbf{r})$ is to be estimated. The matrix $\mathbf{S}$ is then $n\times m$, and the matrix $\mathbf{M}$ is $m\times n$.

The elements of the estimating matrix $\mathbf{M}$ are to be chosen to minimize an average error that can be written

$$E=\mathrm{Tr}\mathbf{E}\mathbf{G}(\hat{\mathbf{J}}_0-\mathbf{J}_0)(\tilde{\hat{\mathbf{J}}}_0-\tilde{\mathbf{J}}_0), \tag{3.2}$$

where again $\mathbf{E}$ denotes the expected value. Here $\tilde{}$ indicates the transpose of a matrix, and $\mathbf{G}$ is a positive-definite, symmetric error matrix. The simplest kind of error matrix is the identity matrix $\mathbf{I}$, which weights errors at all points equally.

The matrix $\mathbf{M}$ that minimizes the error E turns out to be independent of the precise form of $\mathbf{G}$. Under assumptions akin to those made in Sec. 2, it is given by[17]

$$\mathbf{M}=\varphi_0\tilde{\mathbf{S}}(\mathbf{S}\varphi_0\tilde{\mathbf{S}}+\varphi_n)^{-1}, \tag{3.3}$$

and the minimum error is

$$E_{\min}=\mathrm{Tr}\mathbf{G}(\varphi_0-\mathbf{M}\mathbf{S}\varphi_0). \tag{3.4}$$

Here

$$\varphi_0=\mathbf{E}\Delta\mathbf{J}_0\Delta\tilde{\mathbf{J}}_0,\quad \Delta\mathbf{J}_0=\mathbf{J}_0-\mathbf{E}\mathbf{J}_0,\qquad \varphi_n=\mathbf{E}\mathbf{N}\tilde{\mathbf{N}}, \tag{3.5}$$

are covariance matrices of the sampled image and spatial noise processes. The derivation of Eq. (3.3) is a straightforward exercise in differential calculus.

Equivalent to taking the spatial noise $N(\mathbf{r})$ to be white is the assumption that the covariance matrix φ_n is diagonal: $\varphi_n=\bar{\varphi}_n\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. If the spacings of the grid of points at which the restored image $\hat{J}_0(\mathbf{r})$ are evaluated are of the order of the desired resolution distance δ, the matrix φ_0 can also be taken as diagonal: $\varphi_0=\bar{\varphi}_0\mathbf{I}$; for if structural variations in the image $J_0(\mathbf{r})$ take place in distances of the order of δ, values of the representative stochastic processes at points separated by δ must be nearly uncorrelated.

Under these assumptions the column vector $\hat{\mathbf{J}}_0$ of estimated illuminances and the minimum error E are given by

$$\hat{\mathbf{J}}_0=\tilde{\mathbf{S}}(\mathbf{S}\tilde{\mathbf{S}}+\alpha\mathbf{I})^{-1}\mathbf{J}=(\tilde{\mathbf{S}}\mathbf{S}+\alpha\mathbf{I})^{-1}\tilde{\mathbf{S}}\mathbf{J},\qquad \alpha=\varphi_n/\varphi_0, \tag{3.6}$$

$$E_{\min}=\bar{\varphi}_n\mathrm{Tr}\mathbf{G}(\tilde{\mathbf{S}}\mathbf{S}+\alpha\mathbf{I})^{-1}. \tag{3.7}$$

In the limit $\bar{\varphi}_0\gg\bar{\varphi}_n$ the *a priori* density of the image illuminance is so broad as to be nearly uniform, and in effect no prior knowledge of the strength of the image relative to that of the noise is being assumed. We then find that the best estimate of the illuminances is given by the column vector

$$\tilde{\mathbf{J}}_0=(\tilde{\mathbf{S}}\mathbf{S})^{-1}\tilde{\mathbf{S}}\mathbf{J},\quad \alpha\ll 1, \tag{3.8}$$

and the minimum error is

$$E_{\min}=\bar{\varphi}_n\mathrm{Tr}\mathbf{G}(\tilde{\mathbf{S}}\mathbf{S})^{-1}. \tag{3.9}$$

This estimate represents a solution of the linear equations $\mathbf{J}=\mathbf{S}\mathbf{J}_0$ by the method of least squares.[18] The continuous counterpart of this procedure is to take the response $m(\omega)$ of the restoring filter equal to $1/s(\omega)$ for $|\omega|\leq W$, and to 0 for $|\omega|>W$. It may be unsafe if the Fourier transform $s(\omega)$ of the spread function has any zeros in the region $|\omega|\leq W$.

The $m\times n$ matrix $(\tilde{\mathbf{S}}\mathbf{S}+\alpha\mathbf{I})^{-1}\tilde{\mathbf{S}}$ yielding the estimate $\tilde{\mathbf{J}}_0$ as in Eq. (3.6) has elements significantly different from zero only in places that connect points in the image plane separated by distances of the order of the width of the spread function $S(\mathbf{r})$. Hence although the images contain such a large number of grid points that they tax the storage capacity of a digital computer, the images to be processed can be broken down into sections of a size governed by the dimensions of the spread function, which is usually much smaller.

If it is desired to smooth the solution, as when it is being evaluated at points spaced by somewhat less than the "correlation distance" δ of the true image, a covariance matrix φ_0 that is nondiagonal can be used without greatly increasing the amount of computation. A possible form might be

$$\varphi_0=||\varphi_{0,rs}||=||\bar{\varphi}_0\exp(-|\mathbf{r}_r-\mathbf{r}_s|/\delta||).$$

The noise is still assumed white, $\varphi_n=\bar{\varphi}_n\mathbf{I}$. The estimates are then given by vector

$$\hat{\mathbf{J}}_0=(\tilde{\mathbf{S}}\mathbf{S}+\bar{\varphi}_n\varphi_0^{-1})^{-1}\tilde{\mathbf{S}}\mathbf{J},$$

and the minimum-mean-squared error by

$$E_{\min}=\bar{\varphi}_n\mathrm{Tr}\mathbf{G}(\tilde{\mathbf{S}}\mathbf{S}+\bar{\varphi}_n\varphi_0^{-1})^{-1}.$$

[17] R. Deutsch, *Estimation Theory* (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1965), p. 66.

[18] A. Reiz, Arkiv Mat. Astron. Fysik **29A**, #29, 1 (1943).

An estimate of this form was proposed by Twomey on somewhat different grounds.[8,9]

One-dimensional convolution-integral equations, such as arise in spectroscopy and photometry, can be treated in the same digital fashion. The linear equations that replace them after sampling are of the same form as those in Eq. (3.1), and what we have said about the nature of the covariance functions can usually be applied to those other domains as well.

As a one-dimensional example, we present the convolution of a gaussian function

$$J_0(x)=(2\pi)^{-\frac{1}{2}}\exp(-x^2/2) \qquad (3.10)$$

with a rectangular kernel of width $S=1.5$ and height 1. What corresponds to the geometrical image is now the function

$$J_g(x)=\tfrac{1}{2}[I(x+\tfrac{1}{2}S)-I(x-\tfrac{1}{2}S)],$$
$$I(x)=(2/\pi)^{\frac{1}{2}}\int_0^x \exp(-t^2/2)dt, \qquad (3.11)$$

which was calculated at the 71 points $x_k=k\eta$, $-35\leq k\leq 35$, $\eta=0.125$, to yield a set of pristine data $z_k=J_g(x_k)$. The data vector $\mathbf{J}$ consisted of the 71 numbers

$$J_k=[(z_k+e_k')^2+e_k''^2]^{\frac{1}{2}}, \quad -35\leq k\leq 35, \qquad (3.12)$$

where the e_k', e_k'' were 142 pseudorandom numbers generated by a computer in such a way as to be approximately normally distributed with mean 0 and standard deviation 0.001. The data J_k were formed as in Eq. (3.12) to ensure their being positive, as are the data in any experiments in which an intensity of some kind is being measured.

The values of the estimate $\hat{J}_0(x)$ were sought at 29 points $x_j=jh$, $-14\leq j\leq 14$, $h=0.25$. The solution was assumed equal to 0 outside the interval $-3.5<x<3.5$; the true solution, Eq. (3.10), is less than 10^{-3} for $|x|>3.5$. To put the integral equation into the form of Eq. (3.1), the integral was approximated by Simpson's rule, which involves replacing the integrand over each interval of length $2h$ by a parabola passing through the values of the integrand at the two end-points and at the midpoint. The results of applying the method of Eq. (3.6) to these data are plotted in Fig. 4 along the curve of the true solution $J_0(x)$. Two values of α, 0 and $6.25\cdot 10^{-3}$, were used, with only slightly different results.

The results obtained above may be compared with those from the direct method of solving an integral equation with such a rectangular kernel. It employs the relation

$$dJ_g/dx=J_0(x+\tfrac{1}{2}S)-J_0(x-\tfrac{1}{2}S), \qquad (3.13)$$

which is obtained by differentiating the integral equation. If we denote by J_k' the numerical derivative of the data at the kth point, we can write Eq. (3.13) in the form

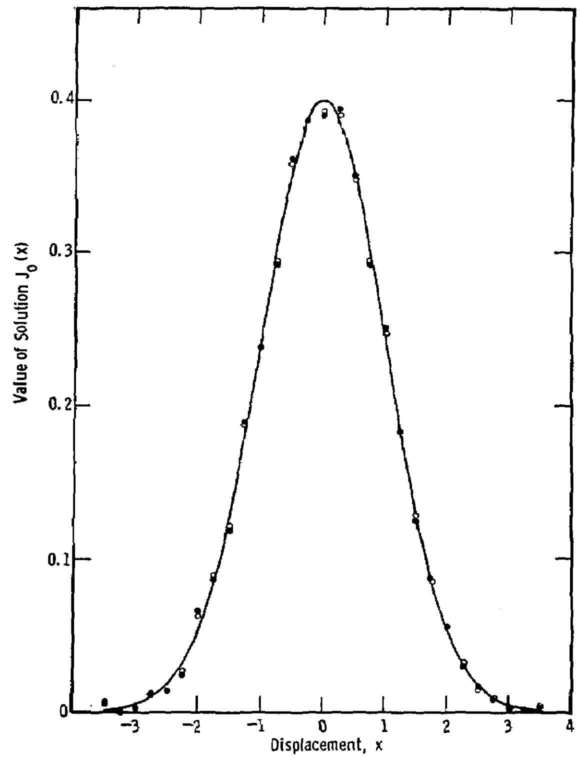

FIG. 4. Restoration of a gaussian function. ○: $\alpha=0$, ●: $\alpha=6.25\cdot 10^{-3}$.

$$J_{0k}=J_{k-6}'+J_{0,k-12}, \quad -29\leq k\leq 0,$$
$$J_{0k}=J_{0,k+12}-J_{k+6}', \quad 0\leq k\leq 29, \qquad (3.14)$$

in which the spacing of the data is $\eta=0.125=S/12$. By assuming $J_{0k}=0$, $|k|>35$, we should be able to derive values of the solution for $|k|\leq 29$ by numerically differentiating the data J_k, starting from each end of the set, and substituting into Eq. (3.14). The numerical differentiation formula was obtained by differentiating Newton's forward interpolation formula, ten terms of which were used.[19] Applied to the pristine data $z_k=J_g(x_k)$, this method produced the correct solution with three-figure accuracy over most of the range. Applied to the very same data J_k as were used in calculating the points on Fig. 4, it yielded a set of positive and negative numbers varying over a range from -1.4 to $+1.5$ and bearing no resemblance whatever to the true solution. It is necessary to smooth such fluctuations in order to obtain a useful solution, although how best to do so in this case is not apparent.

ACKNOWLEDGMENT

I wish to thank Robert Fardo for programming and executing the numerical example on the digital computer.

[19] J. R. Scarborough, *Numerical Mathematical Analysis* (The Johns Hopkins Press, Baltimore, Md., 1930), pp. 48, 114.