Restoring with Maximum Likelihood and Maximum Entropy*

B. Roy Frieden
Optical Sciences Center, University of Arizona, Tucson, Arizona 85721
(Received 19 January 1971; revision received 12 October 1971)

Given M sampled image values of an incoherent object, what can be deduced as the most likely object? Using a communication-theory model for the process of image formation, we find that the most likely object has a maximum entropy and is represented by a restoring formula that is positive and not band limited. The derivation is an adaptation to optics of a formulation by Jaynes for unbiased estimates of positive probability functions. The restoring formula is tested, via computer simulation, upon noisy images of objects consisting of random impulses. These are found to be well restored, with resolution often exceeding the Rayleigh limit and with a complete absence of spurious detail. The proviso is that the noise in each image input must not exceed about 40% of the signal image. The restoring method is applied to experimental data consisting of line spectra. Results are consistent with those of the computer simulations.

Index Headings: Computers; Deconvolution; Image formation; Information theory; Resolution; Spectroscopy.

An incoherent object scene $O(x)$ is a spatial radiance distribution; it therefore cannot be negative:

$$O(x) \geq 0, \quad \text{all} \quad x. \tag{1}$$

Let us consider the problem of restoring $O(x)$ from knowledge of its image $I(x)$ and the point-impulse response $S(x)$ characteristic of the imagery. Restoring schemes that are tailored to the positivity constraint have recently been developed,[1-4] and used with encouraging results. One worker reports the attainment of superresolution.[1] A common feature of these methods is their *ad hoc* nature: They do not follow from a prior statistical principle (such as, e.g., minimum mean-square error). Instead, the positivity is built into the restoring scheme in an arbitrary manner, e.g., by Biraud[1] through representing $O(x)$ as the square of a Fourier series.

Here we show that a positive restoring formula logically follows from a statistical, communication-theory model of image formation. By this model the most likely object scene implied by given image data is found to obey a principle of maximum entropy. The benefits that result from use of this principle are theoretically developed and are illustrated by computer simulation and the use of experimental image data. In brief, restorations based on this principle are (a) positive, (b) not, in general, band limited, (c) spatially smooth (as defined below), and (d) not overly sensitive to noise in the image data at object points that are near the background radiance.

A COMMUNICATION-THEORY MODEL FOR RESTORATION OF DEGRADED IMAGES

A. The Object

Let the unknown object be an incoherent, planar radiance distribution. The object plane is imagined to be subdivided into J equal-sized, elemental cells of extent Δx, and x_j locates the center of the jth cell. (We use one-dimensional notation, for simplicity.) It is logical to make Δx just the required limit of resolution desired by the user in the final restoration. It therefore suffices to represent the object as a sequence of average radiance values $O(x_j)=O_j$, $j=1, \ldots, J$, over the elemental cells.

Suppose that, in addition to the input resolution limit Δx, there is an allowed uncertainty $\pm\Delta O$ in each restored, average (see above) radiation value. Then it suffices to restore each O_j as a multiple o_j of ΔO. That is, we represent O_j as

$$O_j = o_j \Delta O, \tag{2}$$

where o_j is a pure number and ΔO contains the unit of radiance.

Assume, in addition, that the total number of object units

$$\sum_{j=1}^{J} o_j = \mathcal{O}, \quad \text{where } \mathcal{O} \gg 1, \tag{3a}$$

is known. An approximate value for $\mathcal{O}$ can be established, e.g., by observation of the total power in the image. For later use, define a total radiant power

$$P_0 = \mathcal{O}\Delta O. \tag{3b}$$

We now seek the most likely object $O_1, \cdots, O_J \equiv \{O_j\}$ consistent with M observed image values. The derivation to follow will closely parallel that of Jaynes[5] for finding the most likely (positive) probability density obeying M specified constraints. Goldman[6] also displays the main steps in the derivation, although in another context.

We denote the probability of any one object unit ΔO being located in cell x_j as p_j. The p_j are assumed to be unknown and in fact will be estimated by the restoring procedure. For a very large number $\mathcal{O}$ of object units [see Eq. (3a)], it is asymptotically true that

$$p_j = O_j/\mathcal{O}, \tag{4}$$

i.e., the $\{O_j\}$ occurring are "typical,"[5] complying with the fixed (but unknown) $\{p_j\}$.

Assuming a complete lack of *a priori* information regarding the statistical makeup of the object, we make the simple assumption that the occurrence of one object unit ΔO in a cell j does not affect the possible location of any other object unit. This is analogous to Goldman's[6] subdivision of messages into "symbol groups" having no mutual intersymbol influence. Thus, the object units are statistically independent within each cell and from cell to cell. Then[5,6] the number of ways W_0 that a typical object $\{o_j\}$ can occur is

$$W_0(o_1,\ldots,o_J)=\frac{\mathcal{O}!}{o_1!o_2!\ldots o_J!}. \tag{5}$$

Permutations within any one cell of its o_j object units are not counted as contributing to W_0; i.e., the individual object units are assumed to be distinguishable. Jaynes[5] and Goldman[6] also make this implicit assumption in their analogous problems. Equation (5) corresponds precisely to Goldman's[6] result, Eq. (4); our cell-occupation numbers $\{o_j\}$ correspond to his occurrence numbers $\{N_j\}$ for symbol groups.

For later use we take the natural logarithm of Eq. (5) and use Eq. (2) and Stirling's approximation[6] to find

$$\ln W_0(O_1,\ldots,O_J)=-\Delta O^{-1}\sum_{j=1}^{J} O_j \ln O_j \tag{6}$$

to a good approximation for $\mathcal{O}$ very large. The sum defines a spatial entropy for the object, related to ordinary (probabilistic) entropy[5] by use of Eq. (4).

B. Image and Noise

The light from the object is assumed to be imperfectly imaged by a linear system such as a lens. Let an image plane intercept the aerial image and be sampled by a light-collecting aperture of some type at points y_m, $m=1,\ldots,M$. Unit magnification is assumed. Points y_m are not necessarily at a constant separation, and the ratio M/J is arbitrary.

If the image is band limited, it obeys a sampling theorem.[7] In this case, it is natural to space points y_m uniformly at the Nyquist interval. In order to permit a higher resolution in the restoration of $\{O_j\}$ than in the image, object-cell size Δx must then be smaller than the image-data interval,

$$\Delta x<y_m-y_{m-1}. \tag{7}$$

In this case $M<J$; that is, the number of image data is less than the number J of object values to be restored. The restoring algorithm described below was tested under these conditions.

The collected image data I_m, $m=1,\ldots,M$ (having units of irradiance) are further degraded by additive noise n_m, $m=1,\ldots,M$. With the assumption of a linear imaging system, the data relate back to the object through[7]

$$I_m=\sum_{j=1}^{J} O_j S(y_m,x_j)+n_m, \quad m=1,\ldots,M \tag{8}$$

where S denotes the (dimensionless) point-spread function for the system (including the detector) normalized to unit area. S is not necessarily isoplanatic.[7]

The noise $\{n_m\}$ can, of course, be negative, but we would like to work with a positive quantity directly related to the noise in order to develop a theory of noise analogous to Eqs. (3)–(6) for the object. For this purpose, we introduce a noise bias B and a positive noise N_m associated with each n_m through

$$N_m=n_m+B, \quad N_m\geq 0, \quad B\geq 0. \tag{9}$$

Bias B is defined as an upper bound n_{neg} to the amplitude of the most negative noise value n over the measured image,

$$B\equiv n_{\text{neg}}. \tag{10}$$

A value n_{neg} can be guessed at, e.g., by observing fluctuations of the image data where the signal image is known, by object context, to be flat. From Eqs. (9) and (10), all

$$N_m\geq 0, \tag{11}$$

as required.

Let us now anticipate that noise values $\{N_m\}$ are themselves to be estimated by the restoring scheme below. Let ΔN denote an acceptable uncertainty in each restored N_m, and define a signal-to-noise uncertainty

$$\rho\equiv\Delta O/\Delta N. \tag{12}$$

This ratio will be seen to control partially the smoothness of restored $\{O_j\}$ relative to restored $\{N_m\}$. We also now have each

$$N_m=\eta_m\Delta N, \tag{13}$$

where η_m is a whole number. Each ΔN therefore represents a quantity of noise radiance which we call a "noise unit" in analogy to the "object units" used above.

We now make the same statistical assumptions for the occurrence, at sampling points, for these noise units as for the object units, with the exception that we do not assume the total positive noise $\sum_m \eta_m\equiv\mathfrak{N}$ to be known. Expressions analogous to Eqs. (3)–(6) may be constructed. The number of ways $W_N(N_1,\ldots,N_M)$ that a typical sequence $\{N_m\}$ of positive noise values can occur thereby obeys, to a good approximation,

$$\ln W_N(N_1,\ldots,N_M)=-\Delta N^{-1}\sum_{m=1}^{M} N_m \ln N_m \tag{14}$$

for $\mathfrak{N}$ very large.

C. Most Likely Object–Noise Combination

Observe that the problem of determining unknowns $\{O_j\}$ and $\{N_m\}$ from constraint equations (3a) and (8) [as supplemented by Eqs. (2) and (9)] alone has an infinite number of possible solutions: There are $J+M$ unknowns but only $M+1$ equations. Out of this infinite ensemble of solutions, we now seek the typical [see Eq. (4)] object $\{O_j\}$ and noise $\{N_m\}$ sets that are "most likely" to have caused the observed data $\{I_m\}$ and P_0. As with Jaynes,[5] by "most likely" sets we mean those that can be formed in the maximum number of ways, consistent with the data.

The requirement of maximum likelihood will also be seen to impose a constraint of smoothness upon the restored $\{O_j\}$. Phillips[8] showed the extreme importance of enforcing smooth restorations.

If we assume that noise values $\{N_m\}$ are independent of object values $\{O_j\}$, the required sets then obey

$$W_0 W_N = \text{maximum}, \tag{15}$$

subject to constraint equations (3a) and (8). This assumption of statistical independence is made mainly because it allows a tractable solution to the estimation problem. However, it does have physical applicability, as, e.g., in the case of a phototube that is thermal-noise limited. Also, a photograph of a sufficiently low-contrast aerial image obeys both a linear theory[9] [Eq. (8)] and statistical independence between object and noise.[10] For this case, photographic density would necessarily replace irradiance as the physical unit for the image. However, the simple ideas of the model equations (9)–(14) would still apply.

Proceeding with the development, we take the natural logarithm of Eq. (15) and use Eqs. (3b), (6), (8), (9), (12), and (14) to produce a variational principle for the solution sets $\{\hat{O}_j\}$ and $\{\hat{N}_m\}$ obeying Eq. (15),

$$-\sum_{j=1}^{J} \hat{O}_j \ln \hat{O}_j - \rho \sum_{m=1}^{M} \hat{N}_m \ln \hat{N}_m - \sum_{m=1}^{M} \lambda_m \Big[\sum_{j=1}^{J} \hat{O}_j S(y_m, x_j) + \hat{N}_m - B - I_m\Big] - \mu\Big(\sum_{j=1}^{J} \hat{O}_j - P_0\Big) = \text{maximum}. \tag{16}$$

In the conventional manner, we have delegated undetermined multipliers $\{\lambda_m\}$ and μ to the input constraints.

The solution to Eq. (16) is found by the usual method of calculus for finding extrema. Differentiation with respect to $\hat{O}_j$ produces the object solution

$$\hat{O}_j = \exp\Big[-1-\mu-\sum_{m=1}^{M} \lambda_m S(y_m, x_j)\Big]. \tag{17a}$$

Differentiation with respect to $\hat{N}_m$ produces the noise solution

$$\hat{N}_m = \exp[-1-\lambda_m/\rho]. \tag{17b}$$

Differentiation with respect to the $\{\lambda_m\}$ and μ produce the constraint equations

$$I_m = \sum_{j=1}^{J} \hat{O}_j S(y_m, x_j) + \hat{N}_m - B, \quad m = 1, \ldots, M \tag{17c}$$

$$P_0 = \sum_{j=1}^{J} \hat{O}_j. \tag{17d}$$

Equations (17a) and (17b) constitute the restoring formulas proposed in this paper. The parameters $\{\lambda_m\}$ and μ, which define restorations $\{\hat{O}_j\}$ and $\{\hat{N}_m\}$, are found by substituting Eqs. (17a) and (17b) into Eqs. (17c) and (17d). The resulting $M+1$ equations in $M+1$ unknowns $\{\lambda_m\}$ and μ are highly nonlinear, but were found to be always solvable by an $(M+1)$-dimensional Newton–Raphson[11] method, provided that B is sufficiently large [see discussion following Eq. (9)]. Note that if B is too small, $\hat{N}_m$ might have to be negative in order to satisfy Eq. (17c), and this is impossible because representation (17b) for $\hat{N}_m$ is always positive.

DISCUSSION

The variational principle, Eq. (16), was found to result from the requirement of a most likely object–noise combination consistent with M image inputs. Note that the first two sums in Eq. (16) define spatial entropies for object and noise. We have thereby proved that the most likely object–noise configurations share a maximum entropy (with relative weight ρ).

That maximum likelihood implies maximum entropy is, we think, a beautiful result, for a restoration having maximum entropy also conveys maximum information, in Shannon's sense. More precisely, Eqs. (4) and (17) establish the set of object-cell probabilities $\{p_j\}$, consistent with the input data, that would cause an ensemble of objects whose average information content is maximal. Restoration $\{\hat{O}_j\}$ is simply the most typical member [because of Eq. (4)] of that ensemble.

On the practical side, maximum entropy is proving to be a useful tool for numerical estimation. For the analogous problem of probability estimation previously discussed,[5] Wragg and Dowson[12] have shown that the principle can be put to practice. Also, maximum entropy has been used to estimate the power spectrum corresponding to a measured autocorrelation function.[13] Again, results were promising.

The principle of maximum entropy therefore seems to be useful for attacking a wide class of estimation problems. Its numerical utility perhaps lies in the fact that a set $\{O_j\}$ with maximum entropy is smooth in a certain sense: Suppose that set $\{O_j\}$ has entropy H, and let two values—say, O_1 and O_2—approach one

another. Then the entropy for the set must increase (see the Appendix). Hence, an object estimate having maximum entropy is maximally smooth in this sense. The requirement of maximum entropy therefore imposes a smoothing operation upon the restoration. This operation is analogous to Phillips's[8] use of second-difference smoothing, with ρ corresponding to Phillips's smoothness parameter γ.

Inputs ρ and B reflect the user's *a priori* knowledge of fluctuations in $\{O_j\}$ and $\{N_j\}$ and must be regarded as experimental inputs that supplement the $\{I_m\}$ and P_0. Thus, $\{\hat{O}_j\}$ actually represents the most likely object subject to all the experimental inputs $\{I_m\}$, P_0, ρ, and B.

The actual value of ρ to be used in an experiment reflects the user's judgment as to which of $\hat{O}$ and $\hat{N}$ is more apt to be in error. Thus, if *a priori* the true O is known to fluctuate wildly with x whereas the noise has small amplitude, estimate $\hat{O}$ has a higher probable error at each x than an estimate $\hat{N}_m$, and ρ should be made large. The numerical effect of ρ upon $\hat{O}$ and $\hat{N}$ is in line with this viewpoint. Observing Eq. (16), we see that a large ρ will give more weight to smoothing the noise estimate $\{\hat{N}_m\}$ over that of the object, and vice versa. This is further borne out by examination of restoring Eqs. (17a) and (17b). For a given set $\{\lambda_m\}$, a large ρ causes a smooth set $\{N_m\}$ relative to the set $\{\hat{O}_m\}$, and vice versa.

Bias B also affects the smoothness of $\hat{O}$, as is best observed for the case where the image values $\{I_m\}$ are normalized to about unit irradiance and the noise obeys $|n_m|/I_m \ll 1$. Then, by Eq. (9), all values $N_m - B$ will be small. Hence, if B is made large relative to the noise—say, e^{-1}—$\{N_m\}$ is also approximately e^{-1}, and by Eq. (17b) the $\{\lambda_m\}$ tend to be small. With small $\{\lambda_m\}$, Eq. (17a) shows that restoration $\{\hat{O}_j\}$ is smooth. In summary, if the noise is small but B is made very large, then the restored object will be too smooth. This effect is observed in the numerical cases described below. In the same way, it is seen that a small B—e.g., the least upper bound [see discussion following Eq. (9)]—causes large $\{\lambda_m\}$, which permits a restored object with steep gradients. In effect, use of a small B means the injection into the restoring scheme of a large amount [considering defining equation (10)] of *a priori* knowledge regarding the noise; thus the restoration is allowed to become highly fluctuating.

Other useful properties of Eqs. (17a) and (17b) are best realized after the continuous limit $x_j - x_{j-1} \to 0$ for object-cell size is taken. Now, in any numerical use of the restoring scheme, a discrete (albeit small) cell size must be used. Therefore, we are actually interested in seeing how the restoring scheme is affected by the limit $x_j - x_{j-1} \to \epsilon$, a small but finite number. Making the usual replacements $x_j \to x$ continuous, $\hat{O}_j \to \hat{O}(x)\Delta x$ in the variational principle, Eq. (16), the result is that all sums over j approach integrals over x, and the discrete object entropy approaches

$$-\int_{\text{scene}} dx \hat{O}(x) \ln \hat{O}(x) - P_0 \ln \epsilon. \tag{18}$$

The last term is but a finite constant and hence does not affect the solution of Eq. (16). The solution $\hat{O}(x)$, $\{\hat{N}_j\}$ to the continuous version of Eq. (16) is once more easily obtained, with the $\hat{O}(x)$ part a simple Euler-Lagrange problem and the $\{\hat{N}_j\}$ part as before. The solution is Eq. (17b) and

$$\hat{O}(x) = \exp[-1-\mu-\sum_{m=1}^{M} \lambda_m S(y_m, x)]. \tag{19}$$

Solution set $\{\lambda_m\}, \mu$ now determines $\hat{O}(x)$ at a continuum of points.

RESOLUTION, SMOOTHNESS, AND STABILITY TO NOISE

Examination of object restoration equation (19) yields the following useful properties:

(1) $\hat{O}(x)$ is intrinsically positive, as required for an incoherent object.

(2) Each image input I_m contributes a degree of freedom to $\hat{O}(x)$. As the number of input image data decreases, $\hat{O}(x)$ approaches an increasingly conservative, i.e., smooth, estimate. With no image inputs, $\hat{O}(x)$ is a constant P_0/L, where L is the object extent (area, in a two-dimensional analysis), found by combining Eqs. (17d) and (19) with all $\lambda_m = 0$. Conversely, the larger M is, the freer is $\hat{O}(x)$ to fluctuate.

(3) $\hat{O}(x)$ is intrinsically not band limited, except in the limit of all $\lambda_m \to 0$. In the latter case $\hat{O}(x)$ approaches

$$\text{const} \times [1 - \sum_{m=1}^{m} \lambda_m S(y_m, x)], \tag{20}$$

which is band limited if the point-spread function S is band limited. By contrast, a solution set $\{\lambda_m\}$ of sufficiently large numbers can yield an $\hat{O}(x)$ with very high resolution, e.g., exceeding the Rayleigh limit for the lens system that forms S. This point will be verified by numerical cases below. Biraud[1] had already exceeded the Rayleigh limit, using experimental data, by means of his positive restoring formalism.

(4) All space derivatives of $\hat{O}(x)$ are proportional to $\hat{O}(x)$,

$$\frac{d^{(n)}}{dx^{(n)}} \hat{O}(x) \propto \hat{O}(x), \quad n = 1, 2, \ldots. \tag{21}$$

This means that, where $\hat{O}(x)$ is small, it will be a smooth curve. Hence, objects that are mainly zero, such as star fields and spectra consisting of isolated lines, should be restored particularly well by Eq. (19), provided

that image noise is not too severe. This will be exemplified below. Regarding smoothness, we previously found (see Discussion and Appendix) that $\hat{O}(x)$ is also intrinsically smooth because it has a maximum entropy.

(5) All mixed partial derivatives with respect to the λ's and μ are proportional to $\hat{O}(x)$,

$$\frac{\partial^{(n)}\hat{O}(x)}{\partial\lambda_{m1}\partial\lambda_{m2}\cdots\partial\lambda_{mn}} \propto \hat{O}(x), \quad n=1,2,\ldots,M. \quad (22)$$

Therefore, where $\hat{O}(x)$ approximates zero it will be nearly invariant with changes of solution set $\{\lambda_m\},\mu$. Hence, if the same object is restored many times, by use of differing sets $\{I_m\}$ of noise-prone data, the solution sets $\{\lambda_m\},\mu$ for these restorations will differ, but the various restorations will all be nearly invariant where $\hat{O}(x)\simeq 0$. This amounts to a condition of stability to inputs where $\hat{O}(x)\simeq 0$, a very valuable property.[8]

Finally, we address ourselves to the question of whether the sets of solutions $\{\hat{O}_j\},\{\hat{N}_m\}$ of Eqs. (17a)–(17d) are necessarily consistent with our assumption that the number of object units and noise units in each elemental cell is large. Suppose, e.g., that the image inputs $\{I_m\}$ are all small. Then solutions $\{\hat{O}_j\},\{\hat{N}_m\}$ will likewise be small. This inconsistency is illusory, however, because if $\{I_m\}$, B, and P_0 are each multiplied by a sufficiently large constant (or, equivalently, if smaller units of radiance are used), the solution sets $\{\hat{O}_j\},\{\hat{N}_m\}$ will likewise be multiplied by that constant, as shown by Eqs. (17c) and (17d).

TEST CASES

We have extensively tested the restoring equations (17a)–(17d) by means of computer simulation of noisy images and with experimental, spectroscopic data kindly furnished by Jansson *et al.*[2] In all cases, the smallest value of B was used that permitted a solution $\{\lambda_m\},\mu$ to Eqs. (17) (see Discussion). Also, image sampling was uniform, at the Nyquist interval, and the object-cell size Δx was $\frac{1}{5}$ of the Nyquist interval.

For the computer simulations, nine object scenes consisting of rectangular pulses of random spacing and intensity were used. Examples of physical objects of this type are line spectra (see below) and star fields. Imagery was as if by diffraction-limited optics, with the signal image denoted as $\{I_m{}^S\}$.

A simple measure of the accuracy of a number is its relative error. We sought to create image noise characterized by a fixed, maximum relative error β over all points of the image. Accordingly, for each m a random number (the noise) was chosen from a zero-mean, rectangular probability distribution of width $2\beta\cdot I_m{}^S$, and added to $I_m{}^S$. Thus, we generated a type of signal-dependent noise. Although this appears to violate premise equation (15) of signal independence, in practice each test image $\{I_m{}^S\}$ was so degraded by diffraction that $I_m{}^S$ was approximately constant with m, thereby causing the generation of nearly signal-independent noise. The small amount of signal dependence in the noise might be thought of as specifically biasing the tests toward low-contrast photographic imagery.

The image data were spaced at the Nyquist interval and used as inputs to restoring equations (17). For comparison, the perfect-band-limited version O^{pbl} of each object was also displayed. This comparison is important because all current methods of restoring real data are band limited to the frequency domain defined by the image data.[14]

Figure 1 shows the computer outputs for the first four object scenes of the series, with noise amplitude

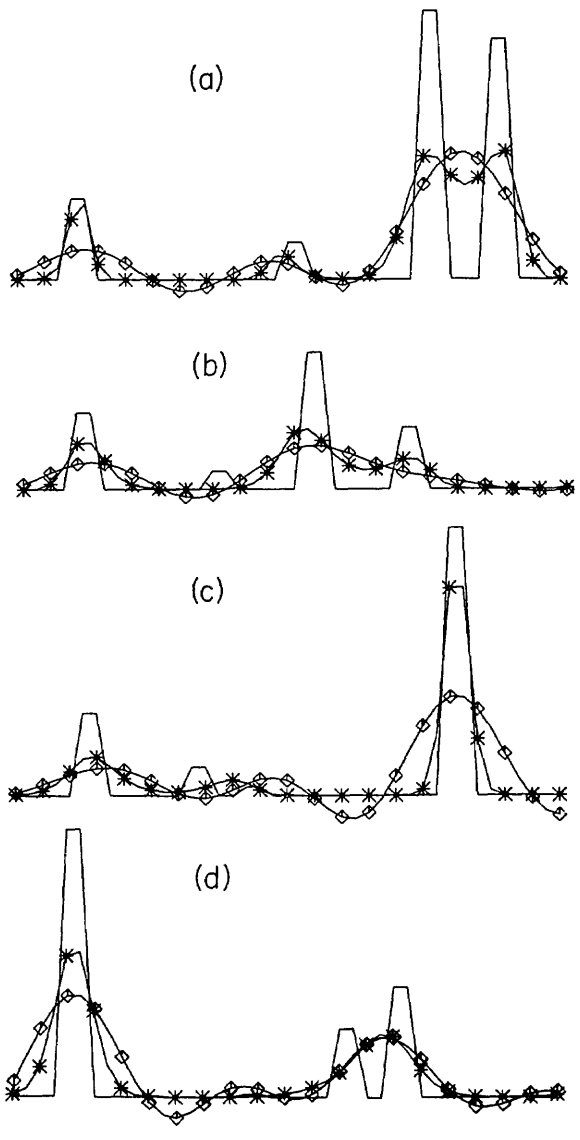

FIG. 1. Restoration by maximum entropy (starred curves) of random-impulse patterns (solid), compared with the perfect-band-limited version (boxed curves) of each object. A value $\rho=1$ was used. The input image points $\{I_m\}$ (not shown) have a maximum relative error of 5% due to uniformly random noise.

$\beta=5\%$ and $\rho=1$. The object (solid curve), its perfect-band-limited version (boxed curve), and the maximum-entropy restoration (starred curve) are shown. The stars and boxes label every other restored point. For reference, the Rayleigh resolution distance spans six consecutive starred points. Visually, the resolution in the maximum-entropy restorations appreciably exceeds that of the perfect-band-limited results; e.g., see the far right-hand impulse pairs of outputs a and b. Further, the maximum-entropy restorations do not oscillate at all when the true object is zero [suggested by Eq. (21)], whereas by comparison the perfect-band-limited curves oscillate and go negative. Finally, there is no false detail (spurious peaks) in the maximum-entropy restorations. It will be seen that these effects hold even when noise amplitude β is as much as 40%.

As a quantitative measure of accuracy, the spatial rms error e from the true object was computed for the two restoration types,

$$e^2=J^{-1}\sum_{j=1}^{J}(O_j-O_j')^2,$$
$$e\equiv\begin{cases}e_{me} & \text{for } O_j'=\hat{O}_j\\ e_{pbl} & \text{for } O_j'=O_j^{pbl}.\end{cases}\tag{23}$$

As a figure of merit for the maximum-entropy results, it is useful to define the average

$$\alpha=9^{-1}\sum_{n=1}^{9}(e_{me}/e_{pbl})_n \tag{24}$$

over the nine test cases generated by the computer. A value $\alpha=1$ would indicate no average improvement over the perfect-band-limited object. For the case $\beta=5\%$, $\rho=1$, we obtained $\alpha=0.79$.

A value $\rho=1$, as used above, is realistic when nothing *a priori* is known about the relative amount of fluctuation across the object and noise distributions (see Discussion). However, if we know that the average fluctuation in the noise is, say, about 1/20 of that in the object, then a value $\rho\simeq 20$ might seem appropriate

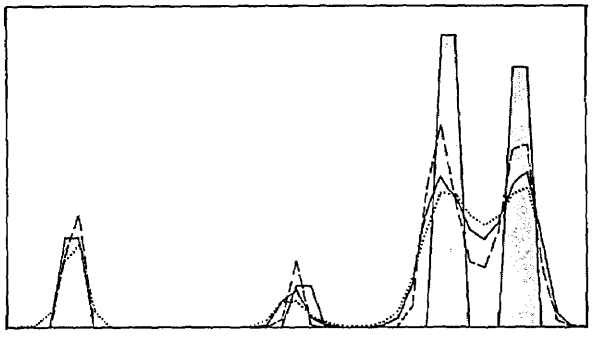

FIG. 2. Maximum-entropy restorations of the shaded object, all based upon the same image data. Dotted curve, $\rho=1$; solid curve, $\rho=20$; dashed curve, $\rho=100$. Where the curves for $\rho=20$ and 100 coincide, only the dashed one is shown. Noise amplitude $\beta=5\%$.

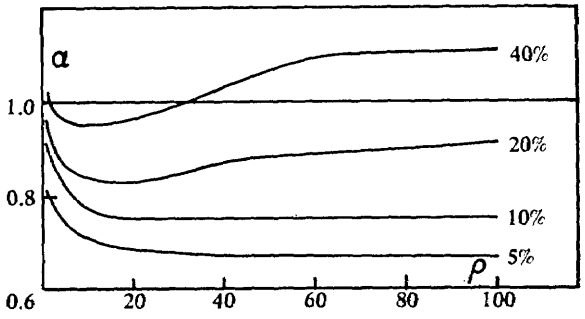

FIG. 3. The effect of ρ upon the relative rms error α over the class of objects, for various levels β of image noise. The values of β label the curves. Level $\alpha=1$ represents egality of rms error with that of the perfect-band-limited versions of the true objects.

[see defining equation (12) of ρ]. Hence, we tested the effect of varying ρ upon restoration of the nine images. Figure 2 is an example, showing the effect of using values $\rho=1$, 20, and 100 for maximum-entropy restorations of object figure 1(a). For this case of $\beta=5\%$, the quality increases with ρ both visually and on an rms-error basis.

By restoring all nine test images (with $\beta=5\%$) at a subdividion of ρ values and computing the relative rms error α over the class for each ρ, we established a continuous curve of α vs ρ, shown as the bottom curve in Fig. 3. From this curve it is apparent that α saturates with ρ, once value $\rho\simeq 20$ is used. Thus, even if the user is grossly wrong in his estimate of relative fluctuation in object and noise, the maximum-entropy restoration will not suffer at all as long as $\rho\gtrsim 20$.

To investigate this saturation effect further, we amplified the noise in the nine images by factors of 2, 4, and 8 (to $\beta=10$, 20, and 40%) and again restored, using various ρ values. The resulting α-vs-ρ curves, Fig. 3, again show approximate saturation. However, perhaps of greater import, from these curves we can establish a single value of ρ to use—regardless of noise level!—for impulse objects. The value $\rho=20$ causes close to the minimum of α for each of the curves.

To clarify the effect of ρ upon the ability to restore when the noise level is high, Fig. 4 shows maximum-entropy restorations of object figure 1(a) using values $\rho=1$, 20, and 100, when $\beta=40\%$. These can be compared with those in Fig. 2, where a noise level $\beta=5\%$ was present. Evidently there is some degradation due to the increased noise, but not an excessive amount for values $\rho=20$ and 100. And, once again, there is no spurious detail. These effects also hold for the other eight objects of the test class.

The image from a conventional spectrometer obeys a linear theory,[2] Eq. (8), but suffers from noise that is not necessarily independent of signal. In view of Eq. (15), our premise of statistical independence, would restoring algorithm equations (17) work well for spectroscopic data? Jansson *et al.*[15] provided us with spectroscopic data for the Q branch of the ν_3 band of

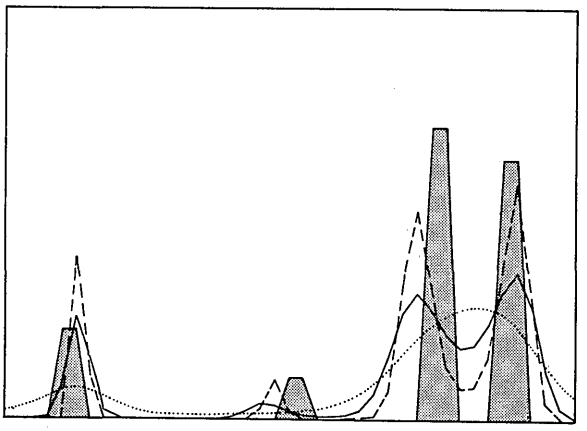

FIG. 4. As in Fig. 2. Here, however, noise amplitude $\beta=40\%$.

CH_4. This was also a critical test in that it permitted comparison to be made with their particular type of positive restoration.[2] Figure 5 shows our maximum-entropy restoration of their data, based on use of a gaussian point-impulse response of the proper half-width. The triangles indicate the experimental inputs $\{I_m\}$. Parameters $\rho=1$ and $B=0.007$ were used. Larger values of B caused restorations with smaller resolution; for example, with $B=0.02$ the right-hand peaks were just barely separated. Smaller values of B did not permit convergence to the required $\{I_m\}$. Again, note the over-all smoothness of the maximum-entropy restoration, especially in regions where $\hat{O}(x)$ is near zero. The restoration by Jansson *et al.*[15] is similar to ours, with differences of relative heights amounting to a few percent.

Computer time for all test cases was about 0.12 s per image-datum input for convergence to a given restoration. Convergence usually required about ten iterations of the Newton–Raphson algorithm.

CONCLUSIONS

Both theory and experiment indicate that the maximum-entropy formalism, Eqs. (17), is well suited to restoring noisy, degraded images of randomly pulsed objects. Of practical interest is the combination of superresolution and lack of spurious oscillation, which ultimately arises from Eqs. (21) and (22). For these reasons, star-field data should be well suited to enhancement by the maximum-entropy technique, provided that photographic nonlinearities and granularity do not contribute more than about 40% error to each collected exposure datum.

A question of basic importance is whether the spectrum of the true object is actually matched by its maximum-entropy restoration outside the optical passband. Certainly superresolution, as in Figs. 1(a) and 1(b), suggests the possibility. We tested the proposition by calculating the spectra of a test object and its maximum-entropy restoration (based on imagery with noise amplitude $\beta=5\%$). These are shown in Fig. 6. We conclude that, despite the presence of significant noise in the image data, it is often possible for the maximum-entropy restoration to accurately match the object at spatial frequencies well beyond the optical cutoff.

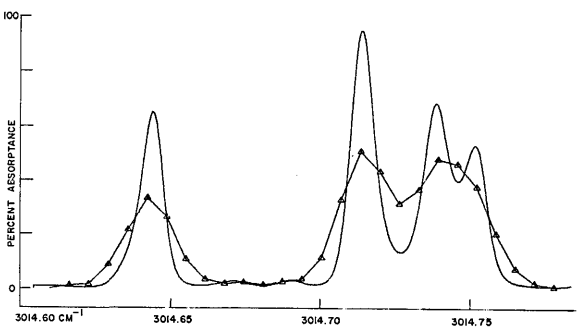

FIG. 5. Restoration by maximum entropy (solid) of a portion of the Q branch of the ν_3 band of the spectrum for CH_4. Triangles indicate the experimental inputs $\{I_m\}$ used in the restoring scheme, some of which lie outside the plotted region and are not shown.

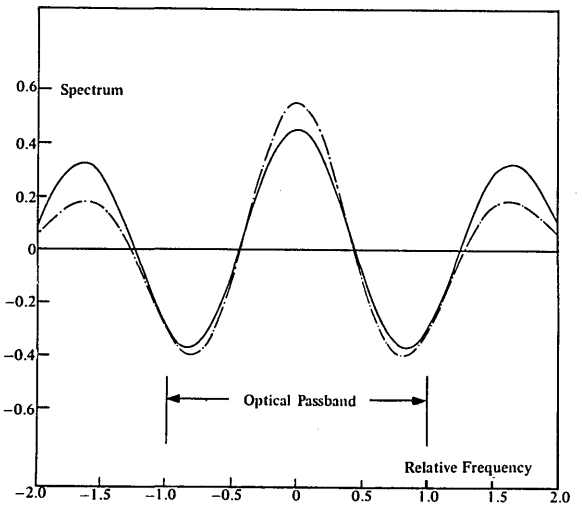

FIG. 6. Comparison of the spectrum (dot-dashed curve) of a maximum-entropy restoration with that (solid curve) of the true object. Image noise amplitude β was 5%. Because the true object was symmetric, only the real part of each spectrum is shown.

APPENDIX

The entropy H for a set $\{O_j\}$ of numbers is defined as

$$H=-\sum_{j=1}^{J} O_j \ln O_j. \tag{A1}$$

We now show that if two numbers O_m and O_n approach each other by differential amounts while their sum remains constant,

$$\sum_{j=1}^{J} O_j=\text{const}, \tag{A2}$$

H must increase. Note that Eq. (A2) is equivalent to constraint equation (17d).

Suppose that $O_m > O_n$. Then for the two to approach equality while obeying Eq. (A2), changes $\Delta O_m = -\epsilon$, $\Delta O_n = +\epsilon$ must be made, $\epsilon > 0$. From Eq. (A1), the change ΔH due to changes ΔO_m and ΔO_n obeys

$$\Delta H = -(1+\ln O_m)\Delta O_m - (1+\ln O_n)\Delta O_n$$
$$= \epsilon \ln(O_m/O_n). \quad \text{(A3)}$$

Because $O_m > O_n$, $\Delta H > 0$. Q.E.D.

This result may also be shown to hold for ϵ finite.

ACKNOWLEDGMENTS

This paper is dedicated to R. S. Hershel, Y. Biraud, P. A. Jansson, and R. C. Jones, whose inestimable friendship and aid made this work possible.

REFERENCES

* Part of this material was presented at the 1971 Spring meeting of the Optical Society, Tucson, Ariz. [J. Opt. Soc. Am. **61**, 657A (1971)].

[1] Y. Biraud, Astron. Astrophys. **1**, 124 (1969).

[2] P. A. Jansson, R. H. Hunt, and E. K. Plyler, J. Opt. Soc. Am. **60**, 596 (1970).

[3] R. S. Hershel, J. Opt. Soc. Am. **60**, 1546A (1970).

[4] D. P. MacAdam, J. Opt. Soc. Am. **60**, 1617 (1970).

[5] E. T. Jaynes, Trans. IEEE **SSC-4**, 227 (1968).

[6] S. Goldman, *Information Theory* (Prentice-Hall, New York, 1955), Appendix II.

[7] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).

[8] B. L. Phillips, J. ACM **9**, 84 (1962).

[9] E. H. Linfoot, *Fourier Methods in Optical Image Evaluation* (Focal, New York, 1964), p. 41.

[10] Reference 9, p. 57.

[11] See, e.g., L. A. Pipes, *Mathematics for Engineers and Physicists* (McGraw-Hill, New York, 1958), p. 116.

[12] A. Wragg and D. C. Dowson, Trans. IEEE **IT-16**, 226 (1970).

[13] A. van den Bos, Trans. IEEE **IT-17**, 494 (1971).

[14] See, e.g., R. Nathan, J. Opt. Soc. Am. **57**, 578A (1967).

[15] The data in question cover the first three maxima of curve A in Fig. 2 of Ref. 2. Jansson's corresponding restoration is Fig. 2(c).