Use of a priori Information in Image Reconstruction:
An Occam’s Razor Estimate

V. V. Biryukov and V. Yu. Terebizh
Sternberg Astronomical Institute, Universitetskii pr. 13, Moscow, 119899 Russia
Received January 27, 1997

Abstract—An Occam’s Razor approach for reconstructing images of an object is considered, in which supplementary information about the structure of the object is used together with the non-negativity condition. The object is assumed to include a comparatively smooth component and a set of features of arbitrary specified shape and unknown flux. Taking a priori information into account can substantially enhance the quality of the estimate of the object’s structure. The calculation algorithm and model examples are presented. A procedure for the conditional solution of inverse problems based on this method is proposed.

1. INTRODUCTION

Usually, the image formation process smooths sharp features of the object imaged so much that reconstruction of regions of high spatial frequencies is difficult—such regions “drown” in the background noise of the total image. In addition, insufficient a priori information about the high-frequency regions leads to the appearance of false oscillations in the brightness distribution of the object. Under these conditions, additional information about the properties of the object are important for the quality of the image reconstruction; this information is often at the researcher’s disposal, but is rather difficult to formalize.

Here, we consider the situation when the intensity distribution of the object \( x_0 \) can be divided into two components: a comparatively “smooth” component \( \hat{x} \) and another component that includes a set of features \( x^{(l)} \) \( (l = 1, \ldots, L) \), each of which has some specified brightness distribution but unknown flux \( f_l \) (for example, stars against a background nebula). In the general case, the second component could include both high-frequency and low-frequency features in an object whose nature is known a priori. Assuming without loss of generality that \( x_0 \) can be described by a vector of length \( n \), we can represent the object in the form

\[
x_0 = \hat{x} + \sum_{l=1}^{L} f_l x^{(l)},
\]

(1)

where the features \( x^{(l)} \) satisfy the conditional normalization

\[
\sum_{k=1}^{n} x^{(l)}_k = 1, \quad l = 1, \ldots, L.
\]

(2)

All the quantities introduced above are taken to be non-negative. We are considering here column vectors: the symbol \( T \) denotes transposition. The basis features \( x^{(l)} \) and the number of features \( L \) are specified; we must estimate the vectors \( \hat{x} \) and \( f = [f_1, \ldots, f_L]^T \).

In this model, the number of parameters sought for is less than the extent of \( x_0 \), which is equal to \( n \); this is ensured by the fact that to describe \( \hat{x} \), it is sufficient to reconstruct only a comparatively small number \( \nu \) of the low spatial frequencies. Most often,

\[
\nu + L \ll n.
\]

(3)

Situations are possible in which either the smooth component \( (\nu = 0) \) or the set of sharper features \( (L = 0) \) are absent.

It is obvious that this approach can be applied in its original form not only to the reconstruction of images, but also to the solution of arbitrary linear inverse problems. Astrophysical applications for the reconstruction of images in the presence of a priori information include spatial photometry of stars against a non-uniform background, the Rayleigh resolution limit problem, studies of the central regions of globular clusters and the nuclei of galaxies, the separation of quasar images from the images of galaxies located along the line of sight in gravitational lenses, and the imaging of stellar surfaces using spectral and photometric data. We will examine some of these applications in future papers.

2. FORMULATION OF THE PROBLEM

We will denote the extent of the image by \( m \). Let the \( m \)-dimensional vector \( a \) and the \((m \times n)\) matrix \( H \) with rank \( n \leq m \) represent, respectively, the mean background brightness within the limits of the image and the point spread function for the given observing system. We will assume that both of these quantities are known. Because of the unavoidable random element in the...
image formation process, the observed image \( y_o \) is one of the many possible random realizations of the true image that make up the ensemble \( Y_o \). The mean brightness of the image is given by the \( m \)-dimensional vector

\[
q_0 = \langle Y_o \rangle = Hx_0 + a.
\]  

(4)

For a photon counting system, it is a very good approximation to treat \( Y_o \) as a Poisson ensemble, so that the probability of obtaining a particular image \( y \) of the object \( x_0 \) is

\[
Pr(Y_o = y|x_0) = f(y|x_0) = \prod_{j=1}^{m} \exp(-q_{0j})(q_{0j})^{y_j}/y_j!.
\]  

(5)

Assuming that the object is not very weak, we can substitute \( f(y|x_0) \) with the corresponding Gaussian distribution (in practice, it requires that \( q_0 > 5 \)). It is easy to see that, in this case, the Poisson image formation model reduces to the linear model

\[
Y_o = Hx_0 + \xi, \quad x_0 \geq 0,
\]  

(6)

where the random "background" of Gaussians \( \xi \) is characterized by the mean value \( \langle \xi \rangle = a \) and the diagonal covariance matrix

\[
C(x_0) = \langle (\xi - a)(\xi - a)^T \rangle = \text{diag}(q_0).
\]  

(7)

The dependence of \( C \) on the object \( x_0 \) takes into account the presence of background noise in model (6), thereby making it possible to reconstruct the bright parts of the object with relatively high accuracy, as expected from general considerations. In addition, the dependence of the probability properties of the background on the object make the problem described by equations (4), (6), and (7) substantially different from the usual linear regression problem [1]. Fortunately, the unknown object appears in (7) only in integrated form, via the mean image \( q_0 \), which makes it possible to replace \( q_0 \) with the observed image \( y_o \) without introducing appreciable errors. A series of model calculations carried out by us fully confirms the validity of this approximation. Thus, assuming

\[
C = \text{diag}(q_0),
\]  

(8)

we arrive at a linear model for the data with fully known probability properties for the background \( \xi \).

The next step is to transform the general model (6) to a model with a Gaussian additive background \( \eta \), whose mean is zero and whose covariance matrix is a unit matrix \( E_m \) with dimensions \( m \times m \) (we will call this model the standard linear model). It is known [2] that by introducing the variables

\[
\begin{align*}
Z_o &= C^{-1/2}(Y_o-a) \\
A &= C^{-1/2}H \\
\eta &= C^{-1/2}(\xi-a),
\end{align*}
\]  

(9)

the standard model takes on the form

\[
\begin{align*}
\langle Z_o \rangle &= Ax_0 + \eta, \quad x_0 \geq 0 \\
\langle \eta \rangle &= 0, \quad \text{cov}(\eta) = E_m.
\end{align*}
\]  

(10)

Of most interest in practice is the case when we record a single reduced image \( z_o = C^{-1/2}(y_o-a) \), which is a Gaussian representation of the random quantity \( Z_o \); the inverse problem consists of obtaining a statistical estimate of the object \( x_0 \) using relations (1), (2), and (10). Note that because of the central limit theorem, our assumption of the Gaussian appearance of the distribution \( \eta \) is not important.

3. SOLUTION USING AN OCCAM'S RAZOR APPROACH

To solve the problem formulated above, we turn to the approach considered here; a detailed description of the basis for this approach can be found in [3, 4].

Let \( \{ v_k \}, k = 1, \ldots, n \) be some complete orthonormal system of \( n \)-dimensional vectors. In accordance with our discussion in the Introduction, we expand the smooth component \( \bar{x} \) in the first \( n \leq n \) unit vectors in this basis

\[
\bar{x} = \sum_{k=1}^{n} p_k v_k = V_n p,
\]  

(11)

where \( p = [p_1, \ldots, p_n]^T \) is an unknown \( n \)-dimensional vector and \( V_n = [v_1, \ldots, v_n] \) is an \( n \times n \) matrix whose columns are the unit vectors \( \{ v_k \} \). When \( n = n \), we have the square orthogonal matrix \( V = V_n^T \). As shown in [3, 4], we obtain a natural ordering of the components of the vector \( p \) with accuracy to within our estimates of them when we adopt for \( V \) a system of eigenvectors of the Fisher matrix \( I \), which in our case is equal to \( I = A^T A \). If the eigenvalues \( \lambda = [\lambda_1, \ldots, \lambda_n] \) \( \geq 0 \) of the symmetrical matrix \( I \) are arranged in decreasing order, then \( p_1, \ldots, p_n \) are, in a statistical context, the first principal components of our estimate of the least-squares vector \( \bar{x} \). Increasing the number of principal components corresponds to increasing the spatial frequency. The Occam's Razor approach keeps only the minimum number of principal components required for satisfactory agreement with the observational data.

An efficient means for constructing the matrix \( V \) is provided by singular value decomposition (see, for example, [6]) of the principal matrix \( A \), which plays the role of the point spread function

\[
A = USV^T,
\]  

(12)

where \( U = (m \times n) \) matrix with orthogonal columns and \( S \) is a an \( (n \times n) \) diagonal matrix of the singular values of \( A \). Since

\[
U^T U = E_n, \quad V^T V = I,
\]  

(13)

ASTRONOMY REPORTS  Vol. 41  No. 6  1997
we have $I = VSV^T$ or $IV = VS$. This latter relation reflects the fact that the columns of $V$ are the eigenvectors of $I$ corresponding to the eigenvalues $\lambda = \text{diag}(S^2)$.

One special problem is formulation of the conditions that must be satisfied in order for some estimate of the object $x$ to be considered a feasible solution of the inverse problem (10). We will adopt a discrepancy condition in the form

$$t_{1-\alpha} \leq \frac{\|Z_0 - Ax\|^2 - m}{\sqrt{2m}} \leq t_{1-\alpha},$$

(14)

where $\|c\|$ denotes the Euclidean length of the vector $c$, the probabilities $\alpha_1 \leq \alpha_2$ are the adopted significance levels for the reconstruction, and $t_q$ is the quantile of the normal distribution of order $q$. Thus, the inequality (14) and the condition $x \geq 0$ determine the feasible region in the space of possible objects. Let us present a short elucidation of this determination.

As can be seen from (10), we have for the true object

$$\|Z_0 - Ax_0\|^2 = \|\eta\|^2 = \sum_{j=1}^{m} \eta_j^2,$$

(15)

Since the random quantities $\eta_j$ are independent and obey a standard normal distribution with zero mean and unit variance, the quantity on the left-hand side of (15) obeys a $\chi^2$ distribution with $m$ degrees of freedom. The mean value $\langle \chi^2_m \rangle = m$, the variance is $\text{var}(\chi^2_m) = 2m$, and for $m \gg 1$, this random quantity must be normally distributed with the corresponding parameter values. If we have some acceptable estimate $\hat{x}$ of the unknown object $x_0$, we can require that its discrepancy for the given ensemble $Z_0$, which is determined by the formula

$$\theta(Z_0|\hat{x}) = \frac{\|Z_0 - A\hat{x}\|^2 - m}{\sqrt{2m}},$$

(16)

take on values that are typical for a Gaussian random quantity with zero mean and unit variance. It is obvious that large values of $\theta(Z_0|\hat{x})$ indicate that $\hat{x}$ is not sufficiently close to the object $x_0$; at the same time, we cannot demand an extremely small value for the discrepancy for a single experiment (the well-known maximum likelihood estimate is unstable for precisely this reason). Adopting some reasonable values for the probabilities $\alpha_1$ and $\alpha_2$ (see below) and specifying, as usual, the quantile of the standard normal distribution $\Phi(t)$ by the relation

$$\Phi(t_q) = q, \quad 0 \leq q \leq 1,$$

(17)

we arrive at a definition of the feasible region in the form (14). Generally speaking, there is no special need to use a normal approximation for the exact distribution of the random quantity $\|Z_0 - Ax\|^2$; the exact distribution is somewhat less convenient only because the corresponding quantile depends on the extent of the image.

Experience with statistical estimation shows that it is reasonable to choose significance levels of the order of $\alpha_1 = 0.10$ and $\alpha_2 = 0.50$. It is important to note that a statistical treatment of inverse problems makes it possible to take into account in a natural way the particularities of object reconstruction when the image contains significant fluctuations.

We can use a quadratic measure of the closeness of two images when evaluating the acceptability of a given estimate $x$, i.e., we can use (14) for a given feasible region, when the dimensions of the image studied are comparatively small. In the general case, an image randomness test must be used; a discussion of this test can be found in [3-5].

Let us collect together the conditions that define the formal formulation of the problem in Occam's Razor approach:

$$t_{1-\alpha} \leq \frac{\|Z_0 - Ax\|^2 - m}{\sqrt{2m}} \leq t_{1-\alpha},$$

$$x = \hat{x} + \sum_{i=1}^{L} f_i x^i,$$

$$\hat{x} = V_p p, \quad v \to \min$$

(18)

Recall that here, we seek the $v$-dimensional vector $p$ and the $L$-dimensional vector $f$. Applying straightforward but cumbersome transformations, which we describe in the Appendix, we can reduce (18) to a standard non-negative least-squares problem. Reliable algorithms to solve such problems are known (see, for example, [7]), so that we can consider (18) to be solved.

4. MODEL EXAMPLES

The most rigorous verification of any transformation method is constructing various model examples for which the true objects are completely known. Our set of test objects included about twenty one-dimensional and two-dimensional objects with various smooth and sharp features.

Figure 1 shows our reconstructions of a one-dimensional object that is a superposition of a Gaussian component with total brightness $10^6$ events and standard deviation $\sigma_0 = 5$ pixels and two point sources with equal brightness $1.9 \times 10^4$ events. Thus, the total brightness of this test object was $1.38 \times 10^5$ events. The point spread function was a one-dimensional analog of an Airy diffraction function with the distance from the maximum to the first zero equal to $\Delta = 7$ pixels. We
took the intensity of the background to be constant, \( q_i = 100 \) events/pixel, and adopted significance levels for the reconstruction \( \alpha_1 = 0.12 \) and \( \alpha_2 = 0.50 \). In accordance with (5), we obtained a smeared image of the object (Fig. 1b) via Poisson “randomization” of the smoothed mean image.

In this test, the available a priori information supposed the presence of two independent point sources located in the 7th and 25th pixels. We obtained the estimate of the object \( x \_k \) with the required significance level when we included \( \nu = 7 \) principal components (Fig. 1d). If we characterize the quality of the reconstruction by the quantity

\[
Q = \frac{\|x - x_{\text{d}}\|}{\sqrt{f_0}} = \frac{1}{\sqrt{f_0}} \sqrt{\sum_{k=1}^{n} (x_{vk} - x_{0k})^2},
\]

we achieved \( Q \) values of the order of 1 for this test example, which are typical for natural Poisson brightness fluctuations of the object—the photometric accuracy of the reconstruction procedure is impeccable.

Note that the introduction of additional point components did not change the quality of the reconstructions. The role of a priori information about the presence of point components in the object is clearly visible in comparisons with the results obtained using only information about the form of the point spread function and the mean background level (Fig. 1b).

Figure 2 gives an idea of the effectiveness of the method in reconstructing two-dimensional objects. Here, the test object is the superposition of a symmetrical Gaussian component with total brightness \( 5 \times 10^5 \) events and standard deviation \( \sigma_0 = 7 \) pixels and five point sources with brightnesses 500, 1000, 1500, 2000, and 3000 events. The dimensions of the image are \( m = n = 25 \times 25 = 625 \) pixels. We adopted a symmetrical two-dimensional Gaussian with standard deviation \( \sigma = 2 \) pixels for the point spread function. We took the background intensity to be \( q_i = 100 \) events/pixel. As above, the a priori information was the coordinates for a set of point sources of unknown brightness. To achieve the required significance level for the reconstruction, we required \( \nu = 25 \) principal components in the expansion of the smooth component, which is substantially fewer than the total number of pixels in the image \( n \).

5. CONCLUDING REMARKS

This method for image reconstruction can be considered the first step of a more precise procedure. The next steps are calculating the image \( \bar{q} \) of the smooth component \( \bar{x} \) of the initial estimate of the object, i.e., \( \bar{q} = H\bar{x} + a \), and then calculating the corresponding covariant matrix \( \bar{C} = \text{diag}(\bar{q}) \) and the effective point spread function \( \bar{A} = \bar{C}^{-1/2} H \). The singular value
decomposition of the matrix $\tilde{A}$ specifies a new, refined system of eigenvectors $\tilde{V}$. Further, the calculations are carried out in the same way described above.

It has been pointed out on multiple occasions that a *priori* information can be extremely diverse. An important example that is comparatively easily formulated is the Rayleigh resolution limit problem. In this case, we assume that the object consists of two point sources whose coordinates and brightness must be estimated. The object must be considered single if the distance between its components is small compared to the associated rms errors. Applying pattern recognition theory to the Rayleigh problem makes it possible to find an optimum criterion in explicit form [3, 4]; the results of numerical modelling for components of equal brightness are presented in [8]. In the future, we will extend these calculations to the case of components with arbitrary brightness.

We have supposed that a researcher sometimes has unambiguous *a priori* information about the nature of certain features in an observed image. Such cases are very possible in practice; however, most often, the researcher is not able to classify the image features completely. For example, without a special analysis, it is impossible to resolve doubts about whether an observed stellar-like image is due to a single star or to a close binary system. In this situation, the image should be reconstructed several times; with and without taking into account information about the possible nature of features in the image. Then, the parts of the different images that substantially differ indicate the role that has been played by the information incorporated, and variation of the assumptions used makes it possible to estimate the accuracy of the results. For example, Figs. 1c and 1d outline the extreme cases when there is no information about the coordinates of point components and when these coordinates are fully specified. The conditional image reconstruction we have described (and, more generally, the conditional solution of inverse problems) is undoubtedly of practical interest.
APPENDIX. REDUCTION OF THE SYSTEM (18)
TO A NON-NEGATIVE LEAST SQUARES
PROBLEM
Let \( X \) denote the \((n \times L)\) matrix whose columns are
the given vectors \( x^k, l = 1, \ldots, L \). We can then write
the estimate for the object in (18) in the form
\[
x = V_x p + X f = G r,
\]
where we have introduced the \( n \times (v + L) \) matrix \( G \) and
the \((v + L)\)-dimensional vector \( r \):
\[
G = [V_x X]. \quad r = \begin{bmatrix} p \\ f \end{bmatrix}.
\]
It can easily be seen that the non-negativity conditions for
the components of the vectors \( x \) and \( f \) can be written as
the single inequality \( Wr \geq 0 \), where the \((n + L) \times (v + L)\)
matrix \( W \) is
\[
W = \begin{bmatrix} V_x & Z_1 \\ Z_2 & E_L \end{bmatrix}.
\]
\( Z_1 \) and \( Z_2 \) are the zero matrices of dimensions \( n \times L \) and
\( L \times v \), respectively, and \( E_L \) is the \( L \times L \) unit matrix.
Finally, assuming that \( B = AG \), we reduce (18) to the
system of conditions relative to the unknown vector \( r \)
\[
\begin{cases}
\|z_0 - Br\|_2^2 - m \\
\sqrt{2m} \leq t_{1-a_2} \leq t_{1-a_1} \\
W r \geq 0, \\
v \rightarrow \min.
\end{cases}
\]
The minimum value of \( v \) can be determined if we
consider conditions (A4) consecutively for \( v = 1, v = 2, \ldots, v = n \)
and stop after the first \( v \) for which the inequalities (A4) are satisfied. Since the discrepancy
monotonically decreases with \( v \), we can in succession
solve the problems
\[
\begin{cases}
\|z_0 - Br\|_2^2 \\
W r \geq 0 \\
v = 1, \ldots, n.
\end{cases}
\]
and as soon as the discrepancy becomes acceptable, we
have obtained a solution \( r_v \) for the system (A4).

System (A5) represents the least squares problem
with inequalities, which can be solved either directly, or
after reduction to the simpler least distance programing
problem or the non-negative least squares problem. We chose this second path.

Let us carry out a preliminary singular value decom-
position of the matrix \( B \) in the same way as was done in
(12) for \( A \):
\[
B = U_B S_B V_B^T.
\]
Using the orthogonality properties of the matrices \( U_B \)
and \( V_B \), in analogy to (13), it is not difficult to verify the
validity of the equality
\[
\|z_0 - Br\|_2^2 = \|U_B^T z_0 - S_B V_B^T r\|^2 + \|z_0\|^2 - \|U_B^T z_0\|^2.
\]
The last two terms in this expression do not depend on
\( r \) and, therefore, the minimization condition (A5) is
equivalent to minimization of the first term in (A7).
Denoting
\[
\omega = S_B V_B^T r - U_B^T z_0,
\]
we reduce (A5) to a least distance programming problem:
\[
\begin{cases}
\|\omega\|_2^2 \rightarrow \min \\
K \omega \geq d.
\end{cases}
\]
where the \((n + L) \times (v + L)\) matrix \( K \) and the \((n + L)\)-
dimensional vector \( d \) are equal to
\[
K = WV_B S_B^{-1}, \quad d = -K V_B^T z_0.
\]
According to [7, Sec. 23.4], in order to solve a least
distance programming problem, we must find a solution \( u_v \) of the non-negative least squares problem
\[
\begin{cases}
\|Tu - \gamma\|_2^2 \rightarrow \min \\
u \geq 0,
\end{cases}
\]
where the \((v + L + 1) \times (n + L)\) matrix \( T \) and the \((v +
L + 1)\)-dimensional vector \( \gamma \) are equal to
\[
T = \begin{bmatrix} K^T \\ d^T \end{bmatrix}, \quad \gamma = [0, \ldots, 0, 1]^T.
\]
Note that the function \( u_v = \text{nnls}(T, \gamma) \) is included in
some applied program packages, in particular, in the
widely-used MatLab package. We then calculate the
discrepancy of (A11)
\[
\rho = Tu_v - \gamma
\]
and the \((v + L)\)-dimensional vector \( \omega_v \) with components
\[
\omega_v = -\rho / \rho_{v+L+1}, \quad k = 1, \ldots, v + L.
\]
which is the solution of the corresponding least distance programming problem. After this, the solution of (A5) follows from (A8):

$$r_\ast = V_B S_B^{-1} (U_B^T z_0 + \omega_\ast), \quad (A15)$$

and substituting $r_\ast$ into (A1) yields the solution of (18), i.e., the desired estimate of the object $x_\ast = Gr_\ast$.

The limiting case $\nu = 0$, when there is no smooth component $\check{x}$, is also of interest. In this case, $B = AX$, and the vector $r_\ast$ is the solution of the non-negative least squares problem with the matrix $B$ and the free parameter $z_0$, i.e., $r_\ast = nnls(B, z_0)$, and the estimate of the object is $x_\ast = Xr_\ast$.

The simplification to the case $L = 0$, when information about fixed features is not used, is evident, and we will not consider it further here.

ACKNOWLEDGMENTS
This work was supported by the Russian Foundation for Basic Research (project no. 96-02-07296a).

REFERENCES


Translated by D. Gabuzda