USCIPI Report 520

ENGINEERING

UNIVERSITY OF SOUTHERN CALIFORNIA

DIGITAL IMAGE RESTORATION UNDER A REGRESSION MODEL-THE UNCONSTRAINED, LINEAR EQUALITY AND INEQUALITY CONSTRAINED APPROACHES

January 1974

Nelson Delfino d'Avila Mascarenhas

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

Sponsored by
Advanced Research Projects Agency
Contract No. F08606-72-C-0008
ARPA Order No. 1706

IPI

IMAGE PROCESSING INSTITUTE

USCIPI Report 520

DIGITAL IMAGE RESTORATION UNDER A REGRESSION MODEL - THE UNCONSTRAINED, LINEAR EQUALITY AND INEQUALITY CONSTRAINED APPROACHES

January 1974

Nelson Delfino d'Avila Mascarenhas

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Eastern Test Range under Contract No. F08606-72-C-0008, ARPA Order No. 1706.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Image Processing Institute, Electronic Sciences Laboratory, University of Southern California, University Park, Los Angeles, California, 90007

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
6
DIGITAL IMAGE RESTORATION UNDER A REGRESSION MODEL - THE UNCONSTRAINED, LINEAR EQUALITY AND INEQUALITY CONSTRAINED APPROACHES.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
9
Technical Report,

5. AUTHOR(S) (First name, middle initial, last name)
10
Nelson Delfino d'Avila/Mascarenhas

6. REPORT DATE
11
Jan[redacted]74

7a. TOTAL NO. OF PAGES
246
12 249p.

7b. NO. OF REFS

8a. CONTRACT OR GRANT NO.
15
F08606-72-C-0008

b. PROJECT NO.
ARPA Order No. 1706

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
14 USC IPI [redacted] 520

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
None

10. DISTRIBUTION STATEMENT
Approved for release; distribution unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

13. ABSTRACT

The problem of restoring images degraded by blur and corrupted by noise is considered in this report. A discretization of the Fredholm integral equation of the first kind in a two dimensional form is performed. The overdetermined and underdetermined regression models are examined in detail, with particular attention to the problem of ill conditioning. The results of the restoration of simulated pictures under atmospheric turbulence and diffraction limited point spread functions are presented.

A priori information in the form of deterministic constraints is proposed as a means to solve the ill conditioning problem. With linear equality constraints, a combination of estimation and hypothesis testing is used to decide if a reduction of the mean square error occurs upon the imposition of the restrictions. Experimental results show that more acceptable results are obtained in the restoration. Linear inequality constraints are incorporated by a quadratic programming formulation. The use of lower (nonnegativeness) and upper bounds indicate a substantial improvement in the restoration, even for the ill conditioned situation.

****

14. Key Words: Digital Image Restoration, Regression Techniques, Ill Conditioning, Linear Constraints, Quadratic Programming.

DD FORM 1 NOV 65 1473

391141

UNCLASSIFIED
Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

Security Classification

ACKNOWLEDGEMENT

The author is deeply indebted to the Chairman of his Dissertation Committee, Dr. William K. Pratt. His advice and encouragement during the course of this research were truly invaluable. The interest of Professors N. Nahi and D. Ferguson, as the other members of the committee, is also acknowledged.

This research was partially supported by the Advanced Research Project Agency of the Department of Defense and was monitored by the Air Force Test Range under Contract No. F08606-72-C-0008. In addition, the work was also supported by a Brazilian scholarship given by the National Research Council and the Foundation for Support of Research in the State of Sao Paulo, when the author was on leave of absence from the Technological Institute of Aeronautics, Sao Jose dos Campos, Brazil.

ABSTRACT

The problem of restoring images degraded by blur and corrupted by noise is considered in this dissertation.

The Fredholm integral equation of the first kind in a two-dimensional form adequately describes the linear model. A discretization is performed by using quadrature methods. By transforming the two-dimensional array into vector format a regression model results. The overdetermined and underdetermined cases are considered in detail, with the derivation of the estimators, their covariance matrices, confidence intervals and hypothesis testing involving parametric functions of pixel values. The problem of ill conditioning is examined for atmospheric turbulence and diffraction limited spread functions. The results of the restoration of simulated pictures under separable spread functions are presented.

In order to solve the ill conditioning of the restoration problem, a priori information in the form of deterministic constraints is proposed. A comparison with existing methods like Wiener filter, smoothing and regularizing techniques is made. Linear equality constraints reduce the variance of the estimators, but some bias may be introduced if the constraints are not valid.

A combination of estimation and hypothesis testing is proposed to decide if a reduction of the mean square error (taking into account both bias and variance) occurs. Experimental results show that more acceptable restored pictures are obtained in the restoration.

Linear inequality constraints are incorporated by means of a quadratic programming formulation. The natural constraint of nonnegativeness of pixel values is handled in a formal way, as well as other types of restrictions that can be described by linear inequalities. Experimental results indicate a substantial improvement in the restoration even for the ill conditioned situation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The subject of image restoration, encompassing attempts to remove different types of degradations in imaging systems, dates back to the fifties [1-1]. However, it was the space program of the sixties, with its need for high quality imagery, that provided the necessary motivation for the development of the field. In particular, the work developed at the Jet Propulsion Laboratory [1-2] demonstrated the feasibility of using the digital computer to deal with the large quantities of pictorial data involved. The success of the effort opened the path for new applications that now range the large spectrum of biological [1-3] and geological sciences [1-4], high energy physics [1-5], etc.

Image restoration or spatial filtering can be divided into two main classes: optical and digital processing. The former has the advantages of larger storage capacity and faster processing, but does not achieve the precision and flexibility of the latter. This dissertation will be concerned with digital methods for image restoration, with emphasis on a firm theoretical basis in their derivation.

The degradations that an imaging system imposes over a picture can often be roughly described as composed by a smoothing operation due to the finite resolution of the sensor and the addition of disturbances, known only in a statistical sense. The earlier methods of restoration, mostly optically oriented, attempted to undo the first degradation by inverse filtering [1-6]. These techniques used the Fourier transforming properties of lenses, by simply multiplying the Fourier transform of the object by the inverse of the Fourier transform of the blurring function. The presence of statistical noise corrupting the image was disregarded and this fact often limited the effectiveness of these methods. A nonoptimal procedure [1-7] consisted of replacing the inverse Fourier transform of the blur function by zero in the spatial frequencies where the noise is larger than the signal.

Perhaps the first attempt to consider a formal way to deal with the presence of noise in an image is due to Helstrom [1-8]. The image and noise were regarded as uncorrelated random processes with a known blur function. Slepian [1-9] considered the lack of knowledge of the blur function, and also modeled it as a random process. Experiments [1-10 and 1-11] indicated that formal approaches using the mean square error criterion gave better results than ad hoc schemes.

Digital methods for image restoration have had to face the problems of storage and computational time in dealing with large scale sampled images. Some of the methods developed have utilized simple, ad hoc operations while others [1-12, 1-13 and 1-14] have attempted more formal approaches based on mean square error. In these cases the Bayesian approach has predominated, with the modeling of the object as a two-dimensional random process.

In this dissertation a different direction is taken. In many situations the experimenter faces the restoration task with very little or even no a priori knowledge about the object to be restored. In such cases the use of the Bayesian approach does not seem to offer the best alternative. When no a priori knowledge about the image is assumed, a regression model adequately describes the blurring and addition of noise processes. The original object is simply considered as a set of parameters to be determined, given the knowledge of the blurred and noisy image, the blurring function and the statistics of the noise. The necessity of digital processing requires a discrete modeling of both the object and the image.

The use of the least squares criterion leads to a very tractable and general mathematical structure, allowing the image restoration process to be cast in a technique analogous to those used in the field of econometrics, for example. However, the lack of use of any a priori knowledge limits the effectiveness of the restoration process. It will be shown that for certain amounts of blur the estimators have very large variance, masking completely the real content of the image.

The model used is flexible enough to accommodate some a priori information, including the Bayesian approach. Since this path has been considerably explored in the past, a new approach was pursued, namely, the use of deterministic constraints.

Linear equality constraints allow a reduction in variance, as a result of a reduction in the dimensionality of the problem. The detection of any bias imposed as the result of incorrectly formulated constraints is also discussed.

The problem of taking into account some physical inequality constraints that should be satisfied by estimators has been the object of discussion by several authors. The most obvious restriction to be satisfied in image restoration is nonnegativeness. It comes from the basic physical laws governing the process of image formation. Some results [1-15] concerning the properties of Fourier transforms of nonnegative functions were used by Lukosz [1-16] to give bounds on the transfer function of a physical system. Similarly, Cleveland and Schell [1-17] extrapolated the spectrum so that it would become an autocorrelation function, imposing that its Fourier transform pair be nonnegative. Phillip [1-18] considered the problem of finding the maximum likelihood estimator of a continuous function assumed to be nonnegative and upper bounded, under gaussian noise. A quadratic expression has to be minimized under these constraints. Necessary and sufficient conditions for uniqueness of the solution were derived and the problem was explicitly solved in some special cases. Some estimation procedures can give only nonnegative results as a result of an exponentiation, for example. This is the case with the technique of homomorphic filtering [1-19], that assumes the image to be the result of the product of an illumination and a reflected component. The assumption that the image is described by an array of cells whose content is given by the Maxwell Boltzmann distribution also leads to estimators given by exponentials. This has been explored by Frieden [1-20] and Hershel [1-21 and 1-22]. Ad hoc procedures have also been tried, as the control of the relaxation factor in an iterative method to solve a linear system of equations [1-23]. Further details on these proposed methods are given in reference [1-24].

This dissertation will develop the inequality constrained least squares approach to image restoration. The proposed method follows a philosophy similar to the one described by Phillip [1-18] for the case of the discrete model. The optimal solution is given by a quadratic programming procedure. Any kind of linear inequality constraint can easily be incorporated and, as a result, requirements like monotonicity and convexity of the solutions can be satisfied. The statistical analysis of the estimators is considerably more complex than the previous cases, but some approximate confidence intervals for functional values of the original image can be obtained. Besides improving the quality of the restoration by the use of additional a priori information in the statistical procedure, the use of linear inequality constraints in the form of lower (nonnegativeness) and upper bounds facilitates the display of the pictorial information.

A word about notation is necessary. An attempt has been made to maintain coherence by expressing matrices by underlined capital letters, vectors by underlined small letters and scalars by small or capital nonunderlined letters.

2. THE RESTORATION PROBLEM

This chapter presents the mathematical framework in which the restoration problem can be cast. In section 2.1 the modeling of the blurring and addition of noise processes is discussed. Section 2.2 contains a brief discussion of the properties of the Fredholm integral equation of the first kind. Its discretization is examined in section 2.3 and, finally, section 2.4 presents the several numerical methods that have been proposed to solve this equation.

2.1 The Model

Figure (2.1-1) contains the block diagram of an incoherent imaging system. The first source of degradation is represented by the point spread function $h(\alpha, \xi, \beta, \eta)$. It is assumed that this blurring operation is linear so that it can be represented by a linear filtering operation. The second source of degradation represents the addition of noise. Due to the randomness inherent in this process, it can only be characterized in statistical terms. Consequently, due to the lack of complete knowledge of the degradation, the restoration cannot be perfect in the sense of restoring the image to the original value.

Assuming that all the processes involved are available continuously and unboundedly, the following equation characterizes this

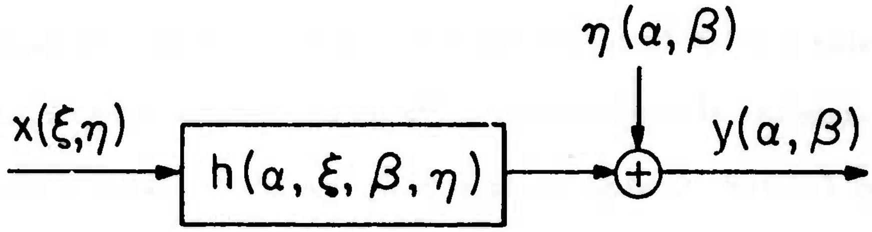


Figure (2.1-1) The Restoration Model

two dimensional model

$$y(\alpha, \beta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x(\xi, \eta) h(\alpha, \xi; \beta, \eta)\, d\xi\, d\eta + \eta(\alpha, \beta)$$

$$-\infty < \alpha, \beta < \infty \qquad (2.1\text{-}1)$$

In many situations, the input image is available only over a finite extent and the previous equation reduces to

$$y(\alpha, \beta) = \int_{a}^{b}\int_{a}^{b} x(\xi, \eta) h(\alpha, \xi; \beta, \eta)\, d\xi\, d\eta + n(\alpha, \beta)$$

$$-\infty < \alpha, \beta < \infty \qquad (2.1\text{-}2)$$

In the particular situation where the blur function is isoplanatic, the point spread is a function of only two variables and the previous equation takes the form

$$y(\alpha, \beta) = \int_{a}^{b}\int_{a}^{b} x(\xi, \eta) h(\alpha - \xi; \beta - \eta)\, d\xi\, d\eta + n(\alpha, \beta)$$

$$-\infty < \alpha, \beta < \infty \qquad (2.1\text{-}3)$$

This model is general enough to include many situations that occur in optical systems. The hypothesis of linear and spatially invariant blur is valid in situations like limitation due to diffraction, for example. In this case, the blur function in a rectangular system assumes the general form [2-1].

$$h(\alpha, \beta) = \left(\frac{\sin \alpha}{\alpha}\right)^2 \left(\frac{\sin \beta}{\beta}\right)^2 \qquad (2.1\text{-}4)$$

A nice feature of this rectangular system is the fact that the degradation is separable. In other words, this function of two variables can be cast into the product of two functions of one variable each. Another example is the blurring due to atmospheric turbulence for long photographic exposure, in which the point spread function is of the form [2-2].

$$h(\alpha, \beta) = \exp\left[-(\alpha^2 + \beta^2)^{5/6}\right] \qquad (2.1\text{-}5)$$

Several other examples could be mentioned. The defocussing [2.1-1] that the optical system may impose over the image is one of them. Other examples could include certain types of optical imperfections and motion blur [2-3].

The assumption of space invariance of blur cannot be validated under certain circumstances. Examples of this are motion blur where objects at different distances from the camera move by different amounts [2-3] or certain optical aberrations like coma, pincushion and barrel distortion. Although most of the experimental work in this dissertation will concentrate on the removing of spatially invariant blur, the regression model will not be restricted to this class of degradation.

The assumptions of additive noise are broad enough to encompass different situations in which the limitations of the optical and/or electrical system impose perturbations known only in a statistical sense. Stray illumination, circuit noise, or round off due to digital processing could be mentioned.

Nevertheless, as it should be expected, there are restrictions in the use of the present model. The assumption of linearity, for example, is subject to criticism, since ultimately the image is recorded on a photographic medium whose characteristic is severely nonlinear [2-1]. Even though this nonlinear function is known, its effect might be such that the addition of noise could occur before and after the nonlinearity. Such would be the case with stray illumination in exposure, followed by the nonlinearity of the H-D curve, followed by roundoff error in digital processing of the picture. In some circumstances, however, the effect of the nonlinearity can be lumped in one block after the addition of noise. Therefore, its effect can be undone by an inverse operation prior to any other operation.

The assumption of additive noise can also be criticized. In particular, the effect of graininess in photographic materials is far from being additive. Huang [2-4] has shown that it could be modeled by a multiplicative process.

Once the limitations of the model are specified, the next step is to clarify the use of a priori information in it. First, the model assumes that the analyst has complete knowledge of the blur function. This hypothesis presupposes that the experimenter has some way of measuring the modulation transfer function. This could be done by measuring the system itself, [2-5], by theoretical analysis [2-1] or by measuring the response of a sharp point or edge in the picture [2-6, 2-7, and 2-8].

With respect to the function $x(\alpha, \beta)$, unless explicitly stated, it will be assumed throughout this dissertation, that it is a fixed but unknown function to be determined, given the values of the output function $y(\xi, \eta)$. This implies that, although the observed values $y(\xi, \eta)$ are random, the desired function $x(\alpha, \beta)$ is not a random process. This approach of parameter identification is in contrast with the Bayesian approach that assumes an a priori statistical distribution on $x(\alpha, \beta)$, characterizing it as a random process. The first method leads itself to the use of other types of a priori information, namely, linear relationships involving values of $x(\alpha, \beta)$ and bounds on their values. These methods will be extensively explored in the present work.

As far as the noise is concerned, all the methods used will assume knowledge of the second order statistical properties. It will not be necessarily white although this assumption will often be made. If additional hypotheses are assumed, further inferences will be drawn.

In order to perform some meaningful restoration, it is necessary to define a goal to the estimation process. The purpose, of course, is to estimate the unknown function $x(\alpha, \beta)$, given the observed values $y(\xi, \eta)$, for some criterion of goodness of the restored image. Assuming that the picture is to be viewed by a human observer, the criterion should take into account the psychophysical properties of human vision. Much research is needed in this field so that reasonable criteria, both from the point of view of realism and mathematical tractability, could be obtained. In the lack of a better one, a squared error criterion will be adopted, namely, minimizing the covariance between the estimated values and true values. Although it is known that the human observer does not judge images according to this criterion [2-9], it has been found (and our experimental work tends to confirm this) that reasonable results can be obtained by its use. Furthermore, and here is its main advantage, the use of a squared error leads to a very tractable mathematical structure, the regression model, that has been considerably explored in mathematical statistics and econometrics.

2.2 The Fredholm Integral Equation of the First Kind

The problem of restoration, as stated in equations (2.1-2) or (2.1-3) consists in solving a two dimensional version of the Fredholm equation of the first kind. The same type of integral equation occurs in different physical problems as radioastronomy [2-10 and 2-11], spectroscopy [2-12], applied optics [2-13], communication theory [2-14] and nuclear engineering [2-15].

The ideal kernel would be

$$h(\alpha, \xi ; \beta, \eta) = \delta(\alpha - \xi, \beta - \eta) \tag{2.2-1}$$

since in this case, with no noise

$$y(\alpha, \beta) = \int_a^b \int_a^b x(\xi, \eta)\, \delta(\alpha - \xi, \beta - \eta)\, d\xi\, d\eta = x(\alpha, \beta) \tag{2.2-2}$$

When the kernel is not the δ function, there is a loss of resolution and the problem that is posed is the one of recovering values of $x(\xi, \eta)$ given the values of $y(\alpha, \beta)$.

In order to keep the equations in their simplest form, only the one dimensional blur will be considered in the following paragraphs. The extension to planar equations is straightforward. Under this condition, under no noise, equation (2.1-2) assumes the form

$$y(\alpha) = \int_a^b x(\xi) h(\alpha, \xi)\, d\xi \tag{2.2-3}$$

where the function $h(\alpha, \xi)$ is the so-called kernel of the integral equation. Associated with this kernel there is an eigenvalue -

eigenfunction problem defined by the equation

$$\int_a^b h(\alpha, \xi)\, \Phi(\xi)\, d\xi = \lambda\, \Phi(\alpha) \tag{2.2-4}$$

The so-called spectrum of the kernel, i.e., the distribution of its eigenvalues, determines the most important properties of the solution $x(\xi)$ for a given observed value $y(\alpha)$. For example, the existence of zero eigenvalues expressed by the equation

$$\int_a^b h(\alpha, \xi)\, \Phi(\xi)\, d\xi = 0 \tag{2.2-5}$$

implies that the solution to equation (2.2-3) will not be unique because a linear combination of eigenfunctions corresponding to zero eigenvalues can always be added to the solution and the result will still be a solution.

A real kernel $h(\alpha, \xi)$ is symmetric if $h(\alpha, \xi) = h(\xi, \alpha)$. The eigenvalues of a symmetric kernel are real and eigenfunctions corresponding to different eigenvalues are orthogonal, that is

$$\int_a^b \phi_i(\xi)\, \Phi_j(\xi)\, d\xi = 0 \tag{2.2-6}$$

$$\lambda_i \neq \lambda_j$$

Furthermore, the eigenfunctions corresponding to the same eigenvalue span a linear subspace. In this subspace an orthogonal basis can be selected (say, by using a Gram-Schmidt procedure) so that it is always possible to have an orthonormal set (as a result of further normalization) of eigenfunctions for a symmetric kernel.

The kernel is defined to be closed if it does not have any zero eigenvalues. As a result, the solution to equation (2.2-3) will be unique. The kernel is said to be separable if it can be expressed as the sum

$$h(\alpha, \xi) = \sum_{n=1}^{N} f_n(\alpha) g_n(\xi) \qquad (2.2\text{-}7)$$

where N is finite and the functions $f_1(\alpha)$, $f_2(\alpha)$, $\cdots$, $f_N(\alpha)$ are linearly independent in $[a, b]$. If the kernel is separable, equation (2.2-3) will have a solution only if $y(\alpha)$ is a linear combination of $f_1(\alpha)$, $f_2(\alpha)$, $\cdots f_N(\alpha)$.

Let λ_1, λ_2, $\cdots$, λn, $\cdots$, in order of decreasing absolute value, be the eigenvalues of the real symmetric kernel and let $\Phi_1(\alpha)$, $\Phi_2(\alpha)$, $\cdots$, $\Phi_n(\alpha)$, $\cdots$ be the corresponding eigenfunctions that are assumed to form an orthonormal set. It can be shown that this kernel can be expressed as

$$h(\alpha, \xi) = \sum_{n=1}^{\infty} \lambda_n \Phi_n(\alpha) \phi_n(\xi) \qquad (2.2\text{-}8)$$

if the series converges uniformly.

Under the conditions that a kernel be symmetric and closed, the set of orthonormal eigenfunctions forms a complete set, i.e., any function in the space can be expressed as a linear combination of the elements of this set. Consequently, the observed value $y(\alpha)$ can also be expressed in this way

$$y(\alpha) = \sum_{n=1}^{\infty} \alpha_n \Phi_n(\alpha) \tag{2.2-9}$$

where the coefficients α_n are given by

$$\alpha_n = \int_a^b y(\alpha) \Phi_n(\alpha)\, d\alpha \tag{2.2-10}$$

In this case a necessary and sufficient condition for equation (2.2-3) to have a solution is that the series

$$\sum_{n=1}^{\infty} \frac{|\alpha_n|^2}{|\lambda_n|^2} \tag{2.2-11}$$

converges. In case of convergence, the solution is given by

$$x(\alpha) = \sum_{n=1}^{\infty} \frac{\alpha_n}{\lambda_n} \Phi_n(\alpha) \tag{2.2-12}$$

2.3 The Discretization of the Integral Equation

When the Fredholm integral equation is to be solved in the digital computer, a discretization has to be performed. This takes into consideration the fact that when images are processed digitally, the information is necessarily finite and discrete. Therefore, suppose that $y(\alpha, \beta)$, the observed function, is sampled at a finite set of points

$$y(\alpha_i, \beta_j) \qquad \begin{array}{l} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J \end{array} \tag{2.3-1}$$

This implies that

$$y(\alpha_i, \beta_j) = \int_a^b \int_a^b x(\xi, \eta)\, h(\alpha_i, \xi, \beta_j, \eta)\, d\xi\, d\eta + n(\alpha_i, \beta_j) \tag{2.3-2}$$

In order to reduce the problem to a complete discrete form, numerical quadrature expressions must be used, replacing the integral by a weighted sum of the values of the integrand at points

$$x(\xi_k, \eta_\ell) \qquad \begin{array}{l} k = 1, 2, \ldots, K \\ \ell = 1, 2, \ldots, L \end{array} \tag{2.3-3}$$

Under these conditions, one obtains the following expression

$$y(\alpha_i, \beta_j) = \sum_{k=1}^{K} \sum_{\ell=1}^{L} W_{k\ell}\, h(\alpha_i, \xi_k ; \beta_j, \eta_\ell)\, x(\xi_k, \eta_\ell) + n(\alpha_i, \beta_j)$$

$$\begin{array}{l} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J \\ k = 1, 2, \ldots, K \\ \ell = 1, 2, \ldots, L \end{array} \tag{2.3-4}$$

Using a lexicographic notation [2-16], it is possible to reduce this two dimensional problem into a one dimensional model. Define the square data arrays of the original and observed images, by the $(K \times L)$ matrix $\underline{X}$ and by the $(I \times J)$ matrix $\underline{Y}$ respectively.

$$\underline{X} = [x(\xi_k, \xi_\ell)] \tag{2.3-5}$$

$$\underline{Y} = [y(\alpha_i, \beta_j)] \tag{2.3-6}$$

Also define a $(L \times 1)$ vector $\underline{v}_\ell$ and a $(K.L \times K)$ matrix $\underline{N}_\ell$ as well as a $(J \times 1)$ vector $\underline{v}_j$, and a $(I.J \times I)$ matrix $\underline{M}_j$

$$\underline{v}_\ell = \begin{array}{cl} \left[\begin{matrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \end{matrix}\right] & \begin{matrix} 1 \\ \\ \\ \ell \\ \ell+1 \\ \\ L \end{matrix} \end{array} \qquad \underline{v}_j = \begin{array}{cl} \left[\begin{matrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \end{matrix}\right] & \begin{matrix} 1 \\ \\ j-1 \\ j \\ j+1 \\ \\ J \end{matrix} \end{array} \tag{2.3-7}$$

$$\underline{N}_\ell = \begin{array}{cl} \left[\begin{matrix} \underline{0}_\ell \\ \vdots \\ \underline{0}_\ell \\ \underline{I}_\ell \\ \underline{0}_\ell \\ \vdots \\ \underline{0}_\ell \end{matrix}\right] & \begin{matrix} 1 \\ \\ \ell-1 \\ \ell \\ \ell+1 \\ \\ L \end{matrix} \end{array} \qquad \underline{M}_j = \begin{array}{cl} \left[\begin{matrix} \underline{0}_j \\ \vdots \\ \underline{0}_j \\ \underline{I}_j \\ \underline{0}_j \\ \vdots \\ \underline{0}_j \end{matrix}\right] & \begin{matrix} 1 \\ \\ j-1 \\ j \\ j+1 \\ \\ J \end{matrix} \end{array} \tag{2.3-8}$$

where $\underline{0}_\ell$ ($\underline{0}_j$) and $\underline{I}_\ell$ ($\underline{I}_j$) represent, respectively, the (KxK) ((IxI)) matrix with all zero elements and the (KxK) ((IxI)) identity matrix.

Using this notation, the vector representations of the matrices $\underline{X}$ and $\underline{Y}$ are given by

$$\underline{x} = \sum_{\ell=1}^{L} \underline{N}_\ell \underline{X} \underline{v}_\ell \qquad (2.3\text{-}9)$$

and

$$\underline{y} = \sum_{j=1}^{J} \underline{M}_j \underline{Y} \underline{v}_j \qquad (2.3\text{-}10)$$

where $\underline{x}$ and $\underline{y}$ are (K.Lx 1) and (I.Jx 1) vectors, respectively. The purpose of the vector $\underline{v}_\ell$ is to extract the ℓ^{th} column from $\underline{X}$. The matrix $\underline{N}_\ell$ has the role of placing this column into the ℓ^{th} segment of the (K.Lx 1) vector $\underline{x}$. As a result, $\underline{x}$ contains the elements of $\underline{X}$ scanned column-wise. Analogous considerations can be made for the vector $\underline{y}$ and the matrix $\underline{Y}$.

At this point, it is also convenient to refer to the inverse relation, that allows the transformation from the vector form back into the two-dimensional format. This manipulation will be useful in transforming blurred and restored images into two-dimensional form for display purposes.

$$\underline{X} = \sum_{\ell=1}^{L} \underline{N}_\ell^T \underline{x} \underline{v}_\ell^T \qquad (2.3\text{-}11)$$

$$\underline{Y} = \sum_{j=1}^{J} \underline{M}_j^T \, \underline{y} \, \underline{v}_j^T \qquad (2.3\text{-}12)$$

With this column scanning of the two dimensional data arrays, the system of equations assumes the form

$$\underline{y} = \underline{H}\,\underline{x} + \underline{n} \qquad (2.3\text{-}13)$$

$\underline{y}$ = (I. J x 1) vector

$\underline{H}$ = (I. J x K. L) matrix

$\underline{x}$ = (K. L x 1) vector

$\underline{n}$ = (I. J x 1) vector

where

$$\underline{y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_J \end{bmatrix}, \quad \underline{y}_j = \begin{bmatrix} y_{1,j} \\ y_{2,j} \\ \vdots \\ y_{I,j} \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_L \end{bmatrix}, \quad \underline{x}_\ell = \begin{bmatrix} x_{1,\ell} \\ x_{2,\ell} \\ \vdots \\ x_{K,\ell} \end{bmatrix}$$

$$\underline{n} = \begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \\ \vdots \\ \underline{n}_J \end{bmatrix}, \quad \underline{n}_j = \begin{bmatrix} n_{1,j} \\ n_{2,j} \\ \vdots \\ n_{I,j} \end{bmatrix}$$

and the matrix $\underline{H}$ is given by

$$\underline{H} = \begin{bmatrix} \underline{H}_{1,1} & \cdots & \underline{H}_{1,L} \\ \vdots & & \\ \underline{H}_{J,1} & \cdots & \underline{H}_{J,L} \end{bmatrix}$$

where the submatrices $H_{j,\ell}$ have the form

$$\underline{H}_{j,\ell} = \begin{bmatrix} W_{1\ell} h(\alpha_1, \xi_1; \beta_j, \eta_j) & \cdots & W_{K\ell} h(\alpha_1, \xi_K; \beta_j, \eta_\ell) \\ \vdots & & \vdots \\ W_{1\ell} h(\alpha_I, \xi_1; \beta_j, \eta_\ell) & \cdots & W_{K\ell} h(\alpha_I, \xi_K; \beta_j, \eta_\ell) \end{bmatrix}$$

The problem of image restoration has now been reduced to a regression framework, that can be stated as follows: given the observed vector $\underline{y}$, the blur matrix $\underline{H}$ and the second order statistics of the noise vector $\underline{n}$, estimate, according to some suitable criterion, the vector of parameters $\underline{x}$. In the next chapter, this regression problem will be treated extensively, as well as the specific questions arising in its solution in the context of image processing.

Furthermore, by the use of additional a priori information, expressed by equality or inequality constraints on the restoration, the problem of ill conditioning will be improved. This will be the object of discussion in chapter 4 of this dissertation.

The problem of selecting the number and the location of the nodes of the quadrature approximation, as well as the observed values is a very complex one, since it involves several different sources of error. The first error comes from the approximation of the integral by the summation and it will be named quadrature error. It tends to decrease as the number of nodes increases. The best location of the nodes is not given, in general, by the equally spaced distribution. In one dimension, if the nodes are located on the zeros of the set of orthogonal polynomials on the interval [a, b], the so called gaussian quadrature is obtained [2-17, pages 392-395]. It provides the optimum precision in the sense of maximizing the degree of the polynomial for which the quadrature is strictly correct. In two dimensions, the technique of gaussian quadrature cannot be easily generalized [2-17, page 419] since the zeros of the orthogonal polynomials may be complex or lie outside the region of integration.

Another source of error may appear when the continuous estimator $\hat{x}(\xi, \eta)$ is obtained from the discrete vector $\underline{\hat{x}}$. Assume that the nonrandom function $x(\xi, \eta)$ is band limited in the frequency plane within, for example, the rectangular region given by the coordinates $-B_u$, $+B_u$ and $-B_v$, $+B_v$, where u and v represent the coordinates of the frequency domain. If the sampling grid is coarser than $\frac{1}{2B_u}$,

$\frac{1}{2B_v}$ (Nyquist rate), then an aliasing error will occur when the continuous function is obtained from the interpolated values. For a given interval, this requires that the number of samples be above a certain threshold if an equally spaced distribution of quadrature nodes is employed. The determination of the threshold will, of course, depend on the a priori knowledge of the frequencies B_u and B_v.

The third source of error comes from the noise inherent in the observations of the blurred picture. While the quadrature error affects the process of passing from the continuous to the discrete description and the aliasing error intervenes in the inverse process, the effect of the noise is over the estimation of the discrete values. It becomes worse as the number of nodes in the quadrature formula increases. It can be measured by the increased condition number or, through a complementary point of view, by the increased variance of the estimators, as will be discussed in chapter 3 of this dissertation. The type of quadrature, the location of the nodes and of the observation values affect the blur matrix $\underline{H}$ and, by consequence, the quality of the estimators.

If the quadrature error can be disregarded with respect to the two other sources of error, a trade-off can be characterized between aliasing and the effect of noise. A small number of nodes implies small variances of the estimators but possibly an aliasing error in the reconstruction. Increasing the number of nodes tends to eliminate the aliasing at the price of increased variances of the discrete estimators.

The number of observation values $y(\alpha_i, \beta_j)$ should be kept at least equal to the number of nodes in the quadrature if no other a priori information is to be incorporated. Otherwise this lack of information will be reflected in infinite variances of the discrete estimators. In the case of use of a priori information, a trade off can be characterized between this information and the one coming from the sample.

This dissertation will be concerned mainly with the third type of error, namely, the one due to noise. It will be implicity assumed that the sampling is enough to avoid aliasing errors and that quadrature errors are negligible compared to noise errors.

2.4 The Existing Methods of Solution

Except for a few cases, the solution of the Fredholm equation of the first kind is far from trivial. Usually numerical techniques are used for its solution. All methods of solution have to face the obstacle of the ill conditioning of the problem. This means that small perturbations on the observed values result in very large changes in the solution. A large research effort has been underway during the last two decades attempting to develop feasible computational methods to deal with this problem.

In general, these methods try to circumvent the problem of ill conditioning by imposing side constraints on the solutions. An example is the method by Phillips (2-18), who imposed the constraint that the solution be smooth by minimizing the criterion

$$\min_{x} \int_a^b [x''(\alpha)]^2 d\alpha \tag{2.4-1}$$

where $x''(\alpha)$ denotes the second spatial derivative of $x(\alpha)$. If a discretization is performed, a linear system of equations is obtained

$$\underline{y} = \underline{H}\,\underline{x} + \underline{\epsilon} \tag{2.4-2}$$

where $\underline{H}$ is a square matrix. In Phillips' method, an estimator $\hat{\underline{x}}$ is forced to satisfy a quadratic equality constraint related to the noise level involved (ℓ^2)

$$(\underline{y} - \underline{H}\,\hat{\underline{x}})^T(\underline{y} - \underline{H}\,\hat{\underline{x}}) = \ell^2 \tag{2.4-3}$$

and the solution is obtained by minimizing a quadratic form measuring the smoothness constraint

$$\min_{\underline{x}} \underline{x}^T \underline{S}\,\underline{x} \tag{2.4-4}$$

The result of the equality constrained optimization problem is given by

$$\underline{\hat{x}} = \left[\underline{H} + \gamma (\underline{H}^{-1})^T \underline{S}\right]^{-1} \underline{y} \tag{2.4-5}$$

where γ is a Lagrange Multiplier that specifies the amount of smoothing imposed by the constraints. It was also shown that an estimator of the perturbation vector $\underline{\epsilon}$ relates to the Lagrange Multiplier γ and the optimal solution through the expression

$$\underline{\hat{\epsilon}} = -\gamma (\underline{H}^{-1})^T \underline{S}\underline{\hat{x}} \tag{2.4-6}$$

A trial and error method was used giving the largest value of the smoothing coefficient γ compatible with the constraint expressed by equation (2.4-3).

A generalization of the method of Phillips as well as a simplification was performed by Twomey [2-19 and 2-20]. The solution expressed by

$$\underline{\hat{x}} = (\underline{H}^T \underline{H} + \gamma \underline{S})^{-1} \underline{H}^T \underline{y} \tag{2.4-7}$$

involves only one matrix inversion instead of two in the method by Phillips. In addition, the method allows solutions in the case where the matrix $\underline{H}$ is not square. It should be observed at this point that for $\underline{S}$ equal to the identity matrix, Twomey's method reduces to the so-called method of ridge regression [2-21, 2-22, and 2-23] that attempts to trade a small amount of bias in the statistical procedures in order to achieve a major reduction in the variance of the estimator.

Among the other methods that yield numerically stable solutions to equation (2.2-3) is the regularization method of the Russian mathematician A.N. Tichonov [2-24, 2-25, and 2-26]. Likewise, the method imposes the constraint that the solution be a piecewise smooth function. It is based on the minimization of a functional which, after discretization, assumes the form

$$M^{\gamma}[\underline{x}, \underline{y}] = \Delta\alpha(\underline{H}\underline{x} - \underline{y})^{T}(\underline{H}\underline{x} - \underline{y}) + \gamma\underline{x}^{T}\left(\frac{1}{\Delta\xi}\underline{S} + \Delta\xi\underline{P}\right)\underline{x} \tag{2.4-8}$$

where $\underline{S}$ and $\underline{P}$ are appropriate positive definite matrices that define the smoothness constraint and $\Delta\alpha$ and $\Delta\xi$ are discrete increments on equation (2.2-3). The solution, for any $\gamma > 0$, was shown to be given by

$$\underline{x}^{\gamma} = \Delta\alpha\left[\Delta\alpha\underline{H}^{T}\underline{H} + \gamma\left(\frac{1}{\Delta\xi}\underline{S} + \Delta\xi\underline{P}\right)\right]^{-1}\underline{H}^{T}\underline{y} \tag{2.4-9}$$

and, again, a trial and error process is involved in the determination of the optimal value of the coefficient γ. The method can also be generalized so that the functional would involve higher order differences on the solution vector.

In the context of image processing, the solution of the planar integral equations involves additional difficulties due to the large dimensionality required when a discretization of the equations is made. Under the conditions of the separability of the matrix $\underline{H}$ as a

Kronecker product, Ekstrom [2-27] restructured the calculations using a singular value decomposition of the same matrix.

Other methods [2-28, 2-29, 2-30, 2-31, 2-32, 2-33, 2-34, and 2-35] have also been suggested to solve the deconvolution problem. In general, these methods were proposed to solve one dimensional, small dimensionality problems and, as pointed out by Ekstrom [2-36], some sort of reformulation of the problem is often needed in order to adapt these procedures to the large dimensionalities that occur in two-dimensional problems.

A significant development is possible in the solution involving large amounts of data, when the kernel $h(\alpha, \xi, \beta, \eta)$ is shift invariant, that is,

$$h(\alpha, \xi, \beta, \eta) = h(\alpha - \beta, \xi - \eta) \qquad (2.4\text{-}10)$$

and the functional that expresses the smoothness constraint is given by a convolution expression. In this case $(\underline{H}^T\underline{H})$ and $\underline{S}$ are Toeplitz matrices in the one dimensional case and block Toeplitz in the two dimensional case. By extending the domain of the convolutions and transforming them into circular operations, Hunt [2-37, 2-38, 2-39, and 2-40] used Fast Fourier Transform techniques to solve Twomey's method.

3. REGRESSION TECHNIQUES

In the previous chapter a blurred digital picture corrupted by noise was modeled by the expression

$$\underline{y} = \underline{H}\,\underline{x} + \underline{n} \tag{3-1}$$

where

$\underline{y}$ = (I x J) x 1 vector

$\underline{H}$ = (I x J) x (K x L) matrix

$\underline{x}$ = (K x L) x 1 vector

$\underline{n}$ = (I x J) x 1 vector

In this discrete form, the problem consists of performing an estimation of the parameter vector $\underline{x}$, given the observed vector $\underline{y}$, the knowledge of the matrix $\underline{H}$ and the statistical distribution of the noise vector $\underline{n}$.

In order to proceed with the derivation of the solution and its properties, it is necessary to consider the possible dimensions involved in the model. For the sake of simplification,

$$I \times J = M \tag{3-2}$$

$$K \times L = N \tag{3-3}$$

Two cases are possible: $M \geq N$ and $M < N$. In the first case, which would occur if, for example, $I \geq K$ and $J \geq L$, the number of nodes in the quadrature expression is less than or equal to the number of samples of the observed image. In the second case the opposite situation occurs. The latter model would tend to occur when the experimenter increases the number of nodes in order to improve the discrete approximation of the integral equation that represents the blurring process. In the case for which $M \geq N$, depending on the values of the $\underline{H}$ matrix, its rank may or may not be given by the number of columns, while in the case for which $M < N$, the rank is necessarily less than the number of columns of $\underline{H}$. As a matter of notation, the model of full column rank is called overdetermined. If the matrix $\underline{H}$ is not of full column rank, the model is said to be underdetermined. The overdetermined model leads to the use of classical regression techniques for its solution, while the underdetermined scheme will require the concept of pseudoinverse and extensions of the previous case.

3.1 The Overdetermined Model

Consider the overdetermined model, i.e., under the conditions of rank of the matrix $\underline{H}$ being determined by the number of columns. Suppose, furthermore, that the noise has zero mean and covariance matrix $\underline{V}$, assumed to be positive definite. The vector $\underline{x}$ is fixed but unknown and the task is to obtain an estimator $\hat{\underline{x}}$ of $\underline{x}$ according to some criterion. The chosen estimator is the best linear unbiased estimator (B.L.U.E.) of $\underline{x}$. This means that one is searching for an estimate

$$\hat{\underline{x}} = \underline{G}\,\underline{y} \tag{3.1-1}$$

such that

$$E(\underline{G}\,\underline{y}) = \underline{x} \tag{3.1-2}$$

Let $\underline{V}_{\hat{\underline{x}}}$ denote the covariance matrix of the optimal estimator vector $\hat{\underline{x}}$, and $\underline{V}_{\bar{\underline{x}}}$ the covariance matrix of any other linear estimator that satisfies (3.1-2). It is noted that $(\underline{V}_{\bar{\underline{x}}} - \underline{V}_{\hat{\underline{x}}})$ is a positive semi-definite matrix. The optimal solution is given by the Gauss-Markov Theorem [3-1, page 52]

$$\hat{\underline{x}} = (\underline{H}^T\,\underline{V}^{-1}\,\underline{H})^{-1}\,\underline{H}^T\,\underline{V}^{-1}\,\underline{y} \tag{3.1-3}$$

and its covariance matrix is

$$\underline{V}_{\hat{\underline{x}}} = (\underline{H}^T\,\underline{V}^{-1}\,\underline{H})^{-1} \tag{3.1-4}$$

Suppose now that, instead of trying to estimate the set of pixel values x_i, i = 1, 2... N, one is interested in estimating a linear functional of the x_i's. An example could be the estimation of the integral of the original picture that would be observed by the output of a photocell. The linear functional Φ can be represented by the inner product

$$\underline{\Phi} = \underline{c}^T \underline{x} \tag{3.1-5}$$

The task then is to obtain a B.L.U.E. estimator for Φ. Using Lagrangean methods [3-2, page 33] it is possible to show that

$$\hat{\Phi} = \underline{c}^T (\underline{H}^T \underline{V}^{-1} \underline{H})^{-1} \underline{H}^T \underline{V}^{-1} \underline{y} \tag{3.1-6}$$

This means that

$$\hat{\Phi} = \underline{c}^T \hat{\underline{x}} \tag{3.1-7}$$

The optimal estimator of $\underline{x}$ could also be derived by considering parametric functions

$$\Phi_i = \underline{e}_i^T \underline{x} \qquad i = 1, 2 \cdots, N \tag{3.1-8}$$

where $\underline{e}_i$ is the $i^{\underline{th}}$ column of the identity matrix. In this case $\hat{\underline{x}}$ will be formed by the set of B.L.U.E. estimators for each one of its components.

The same result could have been obtained by another method, namely, the one that minimizes the weighted sum of squares of the residuals. This is the method of least squares, which was first developed by Gauss. In this case one seeks for the vector $\hat{\underline{x}}$ that minimizes the quadratic expression

$$\theta(\hat{\underline{x}}) = (\underline{y} - \underline{H}\hat{\underline{x}})^T \underline{V}^{-1} (\underline{y} - \underline{H}\hat{\underline{x}}) \tag{3.1-9}$$

Taking derivatives and equating to zero, one obtains

$$-2\underline{H}^T \underline{V}^{-1} \underline{y} + 2\underline{H}^T \underline{V}^{-1} \underline{H}\,\hat{\underline{x}} = 0 \qquad (3.1\text{-}10)$$

or

$$\underline{H}^T \underline{V}^{-1} \underline{H}\hat{\underline{x}} = \underline{H}^T \underline{V}^{-1} \underline{y} \qquad (3.1\text{-}11)$$

This is the set of normal equations of the least squares problem. Under the hypothesis of full column rank of the blur matrix $\underline{H}$ and positive definiteness of the covariance matrix $\underline{V}$, the matrix $(\underline{H}^T \underline{V}^{-1} \underline{H})$ is invertible and the set of normal equations has a unique solution given by

$$\hat{\underline{x}} = (\underline{H}^T \underline{V}^{-1} \underline{H})^{-1} \underline{H}^T \underline{V}^{-1} \underline{y} \qquad (3.1\text{-}12)$$

A comparison of equations (3.1-3) and (3.1-12) will confirm the assertion that the B.L.U.E. and least squares estimators of $\underline{x}$ are identical.

When the noise is white, $\underline{V}$ becomes an identity matrix and expression (3.1-12) reduces to

$$\hat{\underline{x}} = (\underline{H}^T\underline{H})^{-1} \underline{H}^T \underline{y} \qquad (3.1\text{-}13)$$

or

$$\hat{\underline{x}} = \underline{H}^+\underline{y} \qquad (3.1\text{-}14)$$

Let

$$\underline{H}^\dagger = (\underline{H}^T\underline{H})^{-1} \underline{H}^T \qquad (3.1\text{-}15)$$

The matrix $\underline{H}^+$ is called the pseudoinverse or Moore-Penrose generalized inverse of $\underline{H}$ [3-3]. A more complete discussion of its properties will be given in connection with the discussion of the underdetermined model.

Assume for the moment that the noise is white. Therefore, the least squares problem reduces to the minimization of the square of the norm of the residual vector

$$\underset{\underline{x}}{\text{Min}} \; \| \underline{y} - \underline{H}\underline{x} \|^2 \tag{3.1-16}$$

In order to obtain greater understanding over the question of existence and uniqueness of the restoration problem, some heuristic arguments will be presented. Consider Figure (3.1-1) where the decomposition of a finite dimensional linear space into the direct sum of two linear subspaces is represented [3-3], namely,

$$E^N = R_{\underline{H}^T} + N_{\underline{H}} \tag{3.1-17}$$

and

$$E^M = R_{\underline{H}} + N_{\underline{H}^T} \tag{3.1-18}$$

As $\underline{x}$ varies over E^N, the vector $\underline{y} = \underline{H}\,\underline{x}$ varies over $R(\underline{H})$. Therefore, the problem of minimizing $\| \underline{y} - \underline{H}\,\underline{x} \|^2$ over $\underline{x}$ can be reduced to the one of minimizing $\| \underline{y} - \underline{y}_1 \|^2$ where $\underline{y}_1$ is in $R_{\underline{H}}$.

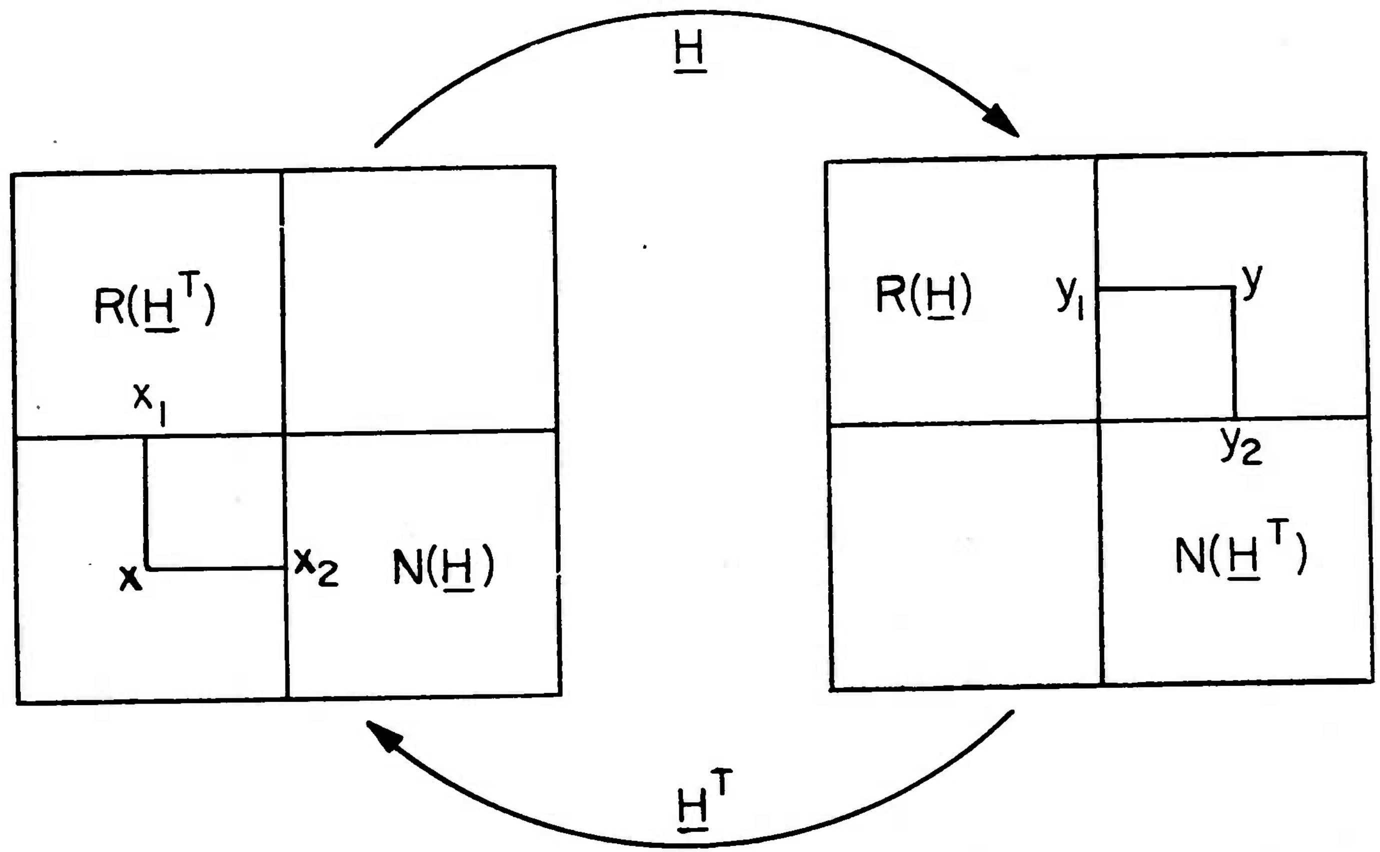


Figure (3.i-1) Decomposition of Linear Spaces

From geometrical considerations it is clear that this is solved for $\underline{y}_1$ given by the projection of $\underline{y}$ onto $R_{\underline{H}}$. Since $\underline{y}_1$ is in $R_{\underline{H}}$, there always exists a solution to the problem

$$\underline{H}\,\underline{x} = \underline{y}_1 \tag{3.1-19}$$

which implies that a solution to the least squares problem always exists. Now, the solution will be unique if and only if the null space of $\underline{H}$, $N_{\underline{H}}$, is composed only of the zero vector. Indeed, assume the solution is unique. Therefore, the null space of $\underline{H}$ has to contain only the zero vector because otherwise a nonzero vector in $N_{\underline{H}}$ could always be added to $\underline{x}$ without affecting $\underline{y}$. On the other hand, assume that $N_{\underline{H}}$ comprises only the zero vector. If the solution is not unique, say $\underline{x}'$ and $\underline{x}''$ being two distinct solutions, then $\underline{x}' - \underline{x}''$ would be in $N_{\underline{H}}$, which is a contradiction.

In the overdetermined case the columns of the blur matrix $\underline{H}$ are assumed to be linearly independent, which implies that the null space of $\underline{H}$ contains only the zero vector, otherwise there would be a nontrivial linear combination of these column vectors resulting in the zero vector. This explains the unique solution that was obtained for the normal equations. In the underdetermined case this will not happen and there will be many solutions to the least squares problem.

3.2 The Hypothesis of Normality and Interval Estimation

It will be assumed in this section that the noise is Gaussian. Besides the fact that it occurs often in practice, this assumption also has the advantage that it will allow the derivation of further properties of the estimators.

Accordingly, let the components of the noise vector $\underline{n}$, n_1, n_2, ..., n_M, be jointly distributed with a multivariate normal distribution

$$\underline{n} \sim N(\underline{0}\,,\,\underline{V}) \tag{3.2-1}$$

denoting that the mean is the zero vector and the covariance matrix is $\underline{V}$. Therefore, given the parameter vector $\underline{x}$, the probability density function of the observed vector $\underline{y}$ is given by

$$p(y|x) = \frac{1}{(2\pi)^{M/2}\,|V|^{\frac{1}{2}}} \exp\left\{-\tfrac{1}{2}(y-Hx)^T V^{-1}(y-Hx)\right\} \tag{3.2-2}$$

Consider now the maximum likelihood estimator of the vector of the original pixel values $\underline{x}$. By definition [3-4, page 193] this estimator is obtained by maximizing over $\underline{x}$ the expression of $p(y|x)$. One may take log before maximizing since it is a monotonically nondecreasing function. In doing this one observes that the maximum likelihood estimator $\underline{\hat{x}}$ minimizes the quadratic expression

$$(\underline{y} - \underline{H}\,\underline{x})^T \underline{V}^{-1} (\underline{y} - \underline{H}\,\underline{x}) \tag{3.2-3}$$

Under these conditions, given the observed values of the blurred and noisy picture, the maximum likelihood estimator of the original pixel values is the least squares estimator (and the B.L.U.E. estimator for the overdetermined model), if the hypothesis of normality is assumed. Since the maximum likelihood estimator has the desirable properties of consistency and asymptotic efficiency, the Gaussian hypothesis allows the extension of these properties to the estimators derived under the two other criteria.

In the following discussion the assumption of white noise will be made. The purpose will be to derive estimators for the variance of the noise that corrupts the image. Under the white noise hypothesis, (3.2-2) assumes the form

$$p(\underline{y}|\underline{x}) = \frac{1}{(2\pi)^{M/2}\,\sigma^M} \cdot \exp\left(\frac{-1}{2\sigma^2} (\underline{y} - \underline{H}\,\underline{x})^T (\underline{y} - \underline{H}\,\underline{x})\right) \tag{3.2-4}$$

If the log likelihood function is maximized by setting the derivative with respect to σ equal to zero, one obtains

$$\frac{1}{2\sigma^4} (\underline{y} - \underline{H}\,\underline{x})^T (\underline{y} - \underline{H}\,\underline{x}) - \frac{M}{2\sigma^2} = 0 \tag{3.2-5}$$

The expression for the maximum likelihood estimator of the coefficient vector $\underline{x}$ has already been derived. By substituting this value, the maximum likelihood estimator for σ^2, $\hat{\sigma}^2$, is obtained

$$\hat{\sigma}^2 = \frac{1}{M}(\underline{y} - \underline{H}\hat{\underline{x}})^T(\underline{y} - \underline{H}\hat{\underline{x}}) \tag{3.2-6}$$

Now, consider the quantity

$$\begin{aligned}\hat{\underline{v}} = \underline{y} - \hat{\underline{y}} &= \underline{y} - \underline{H}\hat{\underline{x}} = \\ &= \underline{y} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{y} = \\ &= [\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T]\underline{y} \\ &= \underline{L}\underline{y}\end{aligned} \tag{3.2-7}$$

Since $\underline{L}$ is the difference of two symmetric matrices, it follows that $\underline{L}$ is also a symmetric matrix. Also $\underline{L}$ is idempotent as it can be shown by the following derivation

$$\begin{aligned}\underline{L}^2 = [\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T]^2 &= \underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T \\ &+ \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T \\ &= \underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T = \underline{L}\end{aligned} \tag{3.2-8}$$

Furthermore, the trace of $\underline{L}$ can be obtained as follows

$$\mathrm{tr}\underline{L} = \mathrm{tr}\underline{I} - \mathrm{tr}\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T = M - \mathrm{tr}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{H} = M - N \tag{3.2-9}$$

From the fact that the rank of an idempotent matrix is equal to its trace, it follows that the rank of $\underline{L}$ is (M-N). Also, observe that

$$\underline{L}\,\underline{H} = [\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T]\underline{H} = \underline{H} - \underline{H} = \underline{0} \qquad (3.2\text{-}10)$$

Now, consider another possible estimator foı the parameter σ^2, namely

$$\begin{aligned} s^2 &= \frac{1}{M-N}(\underline{y} - \underline{H}\,\hat{\underline{x}})^T(\underline{y} - \underline{H}\,\hat{\underline{x}}) \\ &= \frac{1}{M-N}\hat{\underline{v}}^T\hat{\underline{v}} \end{aligned} \qquad (3.2\text{-}11)$$

The following relationship is valid

$$\hat{\underline{v}} = \underline{L}\,\underline{v} = \underline{L}(\underline{H}\,\underline{x} + \underline{n}) = \underline{L}\,\underline{n} \qquad (3.2\text{-}12)$$

where the fact that $\underline{L}\,\underline{H} = \underline{0}$ was used. Therefore,

$$\hat{\underline{v}}^T\hat{\underline{v}} = \underline{n}^T\underline{L}^T\underline{L}\,\underline{n} = \underline{n}^T\underline{L}\,\underline{n} = \mathrm{tr}\,\underline{L}\,\underline{n}\,\underline{n}^T \qquad (3.2\text{-}13)$$

The second equality comes from the idempotency of $\underline{L}$ and the third is based on the fact that $\underline{n}^T\underline{L}\,\underline{n}$ is a scalar and therefore equal to its own trace.

By taking expectations one obtains

$$E(\underline{\hat{v}}^T\underline{\hat{v}}) = E(\mathrm{tr}\underline{L}\,\underline{n}\,\underline{n}^T) = \mathrm{tr}\ \underline{L}E(\underline{n}\,\underline{n}^T) = \sigma^2\ \mathrm{tr}\ \underline{L}\,\underline{I} = \sigma^2(M-N) \qquad (3.2\text{-}14)$$

As a result of the last expression, s^2 is an unbiased estimator for the variance of the noise. Observe that, although $\frac{1}{M}\underline{\hat{v}}^T\underline{\hat{v}}$ is the maximum likelihood estimator of σ^2, it is not an unbiased estimator.

It could be also of interest to determine an estimator for the covariance matrix of the estimator $\underline{\hat{x}}$, namely, $\sigma^2(\underline{H}^T\underline{H})^{-1}$. Since s^2 is an unbiased estimator for σ^2, it follows that $s^2(\underline{H}^T\underline{H})^{-1}$ is an unbiased estimator of $\sigma^2(\underline{H}^T\underline{H})^{-1}$.

It has already been observed that, under the normality assumption, the vector of estimated pixel values, $\underline{\hat{x}}$, is distributed according to a multivariate normal distribution. Observe, furthermore, that $(M-N)s^2 = \underline{y}^T\underline{L}\underline{y} = \underline{n}^T\underline{L}\underline{n}$ and that $\underline{L}$ is an idempotent matrix of rank (M-N). This fact implies that the quadratic form $\frac{(M-N)s^2}{\sigma^2}$ has a χ^2 distribution with (M-N) degrees of freedom [3-5, page 91].

Now, observe that the matrix $\underline{L}$ and the matrix $(\underline{H}^T\underline{H})^{-1}\underline{H}^T$ of $\underline{\hat{x}} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{y}$ satisfy

$$(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{L} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T[\underline{I}-\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T] = (\underline{H}^T\underline{H})^{-1}\underline{H}^T - (\underline{H}^T\underline{H})^{-1}\underline{H}^T = \underline{0} \qquad (3.2\text{-}15)$$

This implies [3-6, page 89] that $\hat{\underline{x}}$ and s^2 are independently distributed.

Let us reconsider now, still under the Gaussian hypothesis, the problem of estimating a linear functional of the pixel values of a picture, like the sum of the pixels or a single pixel value, for example. Expression (3.1-6) gives the value of the B.L.U.E. estimator $\hat{\Phi}$. That expression can be put into the form

$$\hat{\Phi} = \underline{u}^T \underline{y} \tag{3.2-16}$$

where

$$\underline{u}^T = \underline{c}^T (\underline{H}^T \underline{V}^{-1} \underline{H})^{-1} \underline{H}^T \underline{V}^{-1} \tag{3.2-17}$$

Since $\underline{n}$ is normally distributed with zero mean and covariance matrix $\underline{V}$, it follows that $\underline{y}$ is also normally distributed with mean $\underline{H}\,\underline{x}$ and covariance $\underline{V}$. On the other hand, $\hat{\Phi}$, being a linear combination of Gaussian random variables, is also a Gaussian random variable and its variance is expressed by

$$\mathrm{var}(\hat{\Phi}) = \underline{u}^T \underline{V}\, \underline{u} \tag{3.2-18}$$

Since $\hat{\Phi}$ is an unbiased estimator of Φ, the random variable

$$\hat{\eta} = \frac{\hat{\Phi} - \hat{\Phi}}{\sqrt{\mathrm{var}(\hat{\Phi})}} \tag{3.2-19}$$

is zero mean, unit variance and Gaussian. As a consequence, the probability that $\hat{\eta}$ falls in the interval $\lceil -K, K \rceil$ is given by

$$\Pr\{-K \leq \hat{\eta} \leq K\} = \int_{-K}^{K} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \qquad (3.2\text{-}20)$$

or

$$\Pr\left\{-K \leq \frac{\hat{\Phi} - \Phi}{\sqrt{\operatorname{var}(\hat{\Phi})}} \leq K\right\} = \alpha \qquad (3.2\text{-}21)$$

where

$$\alpha = \int_{-K}^{K} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx \qquad (3.2\text{-}22)$$

It is our interest to derive the confidence interval at a given level α for the parametric function Φ. In view of (3.2-21) this can be given by

$$I_K(\hat{\Phi}) = \left\{\hat{\Phi} - K(\operatorname{var}(\hat{\Phi}))^{\frac{1}{2}}, \ \hat{\Phi} + K(\operatorname{var}(\hat{\Phi}))^{\frac{1}{2}}\right\} \qquad (3.2\text{-}23)$$

or

$$I_K(\hat{\Phi}) = \left\{\underline{u}^T \underline{y} - K(\underline{u}^T \underline{V}\,\underline{u})^{\frac{1}{2}}, \ \underline{u}^T \underline{y} + K(\underline{u}^T \underline{V}\,\underline{u})^{\frac{1}{2}}\right\} \qquad (3.2\text{-}24)$$

For each value of K, the corresponding confidence level is tabulated below [3-2, page 38].

K	α
1.0	0.6827
2.0	0.9545
3.0	0.9973
4.0	0.9999

It is possible to give a geometrical interpretation of the confidence ellipsoid. This interpretation will provide considerable insight into the properties of the estimators, and will open the path toward the discussion of the influence of the perturbations in the solution of the linear equations involved. Excellent discussions of this interpretation are found in references [3-2, pages 40-58 and 3-7, pages 406-411].

Consider the expression given by equation (3.2-3). For a given observed value $\underline{y}$, that expression represents a quadratic function in N-dimensional space of the x_i variables. Under the overdetermined model the solution of the normal equations is unique. Therefore, the minimum of the quadratic form is obtained at a unique point $\hat{\underline{x}}$. For other values of $\underline{x}$ the residual surface assumes the shape of a parabolloid. Let r_o denote the minimum value of the quadratic expression. Consider the expression

$$(\underline{y}-\underline{\underline{H}}\underline{x})^T \underline{V}^{-1}(\underline{y}-\underline{\underline{H}}\underline{x}) = r_o + K^2 \qquad (3.2\text{-}25)$$

Upon the substitution of the value of $\hat{\underline{x}}$ given by the solution of the normal equations, it is easy to verify that

$$(\underline{x} - \hat{\underline{x}})^T \underline{H}^T \underline{V}^{-1} \underline{H}(\underline{x} - \hat{\underline{x}}) = K^2 \tag{3.2-26}$$

This is the expression of an ellipsoid with center at the point $\hat{\underline{x}}$ in the N-dimensional $\underline{x}$ - space. Following reference [3-2] this ellipsoid will be called the K - ellipsoid. Figure (3.2-1), obtained from reference [3-2, page 42] shows the residual surface and the K - ellipsoid for N = 2.

Consider now a vector $\underline{h}$ in the N-dimensional space. For a nondegenerate ellipsoid (this is the case with the overdetermined model), there will be two (N-1) dimensional planes orthogonal to $\underline{h}$ and tangent to the ellipsoid. These are planes such that the ellipsoid lies entirely on one side of and has at least one point in common with them. Following Scheffé [3-7] this planes will be called planes of support of the K - ellipsoid.

On the other hand, equation (3.1-7) gives the value of the estimator of the parametric function Φ as expressed by the inner product of the vector $\underline{c}$ and the estimator $\hat{\underline{x}}$. If one considers the planes of support of the K - ellipsoid perpendicular to $\underline{c}$, their analytical expressions will be given by [3-2, page 41]

$$\underline{c}^T \underline{x} = \underline{c}^T \hat{\underline{x}} + K[\underline{c}^T (\underline{H}^T \underline{V}^{-1} \underline{H})^{-1} \underline{c}]^{\frac{1}{2}} \tag{3.2-27}$$

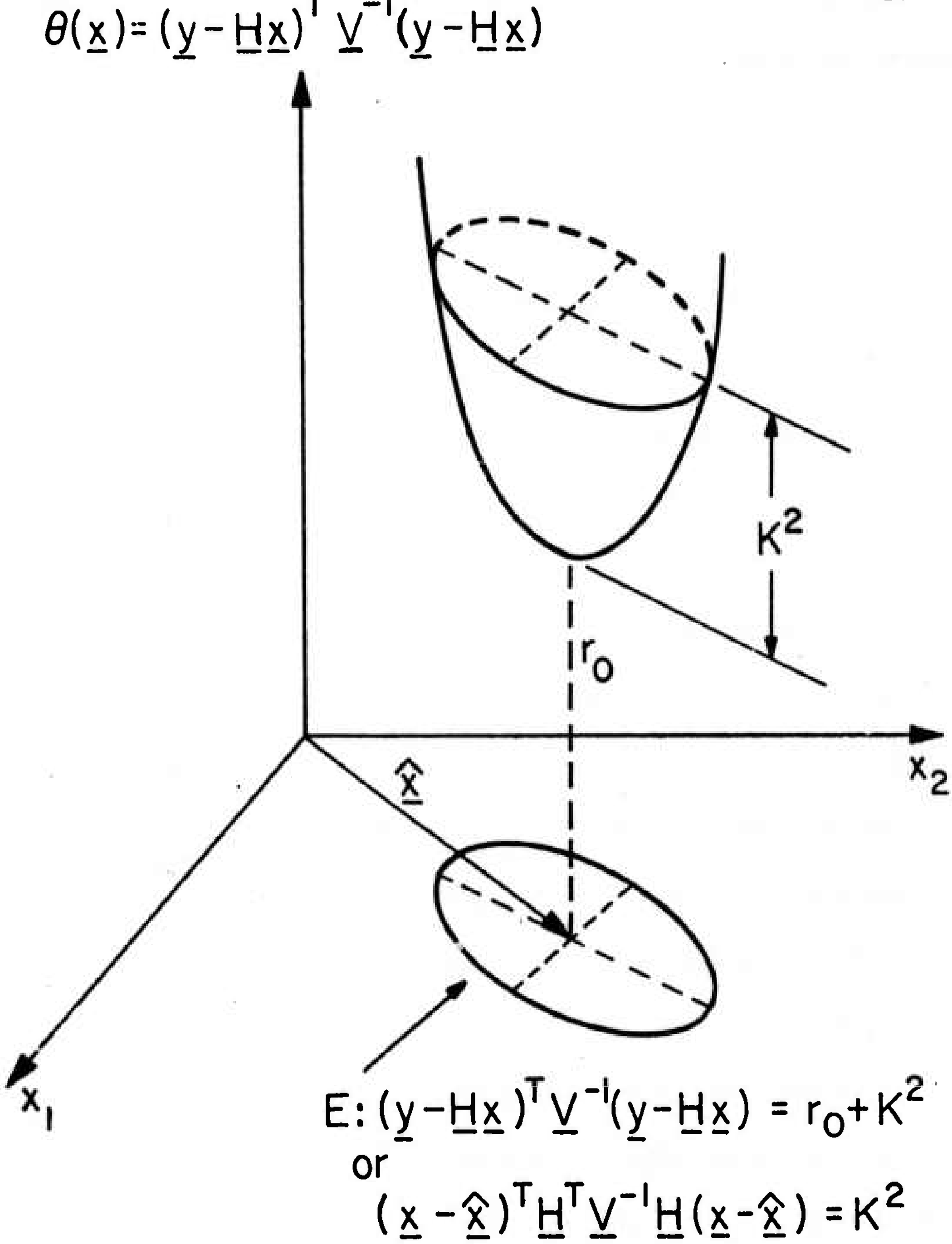


Figure (3.2-1) The Residual Surface and the K-Ellipsoid for N=2

Since

$$\text{var}(\hat{\Phi}) = \underline{u}^T \underline{V}\underline{u} = \underline{c}^T(\underline{H}^T \underline{V}^{-1} \underline{H})^{-1} \underline{c} \qquad (3.2\text{-}28)$$

it follows, comparing (3.1-7), (3.2-24) and (3.2-27), that the confidence interval $I_K(\Phi)$ can be given by the distance between the two points where the planes of support touch the K-ellipsoid. Figure (3.2-2), obtained from reference [3-2, page 43] illustrates the previous assertion. The same figure also shows that the width of the confidence interval is proportional to the distance between the two support planes.

Since the width of the confidence intervals of parametric functions of pixel values is proportional to the distance between the planes of support, and since this distance will vary depending on the direction of the vector $\underline{c}$ with respect to the axes of the ellipsoid, it is important to characterize the directions of these axes in terms of measurable quantities.

Equation (3.2-26) gives the analytical expression of the K-ellipsoid. If a translation of origin in the E^N space is made through the equation

$$\underline{v} = \underline{x} - \underline{\hat{x}} \qquad (3.2\text{-}29)$$

equation (3.2-26) assumes the form

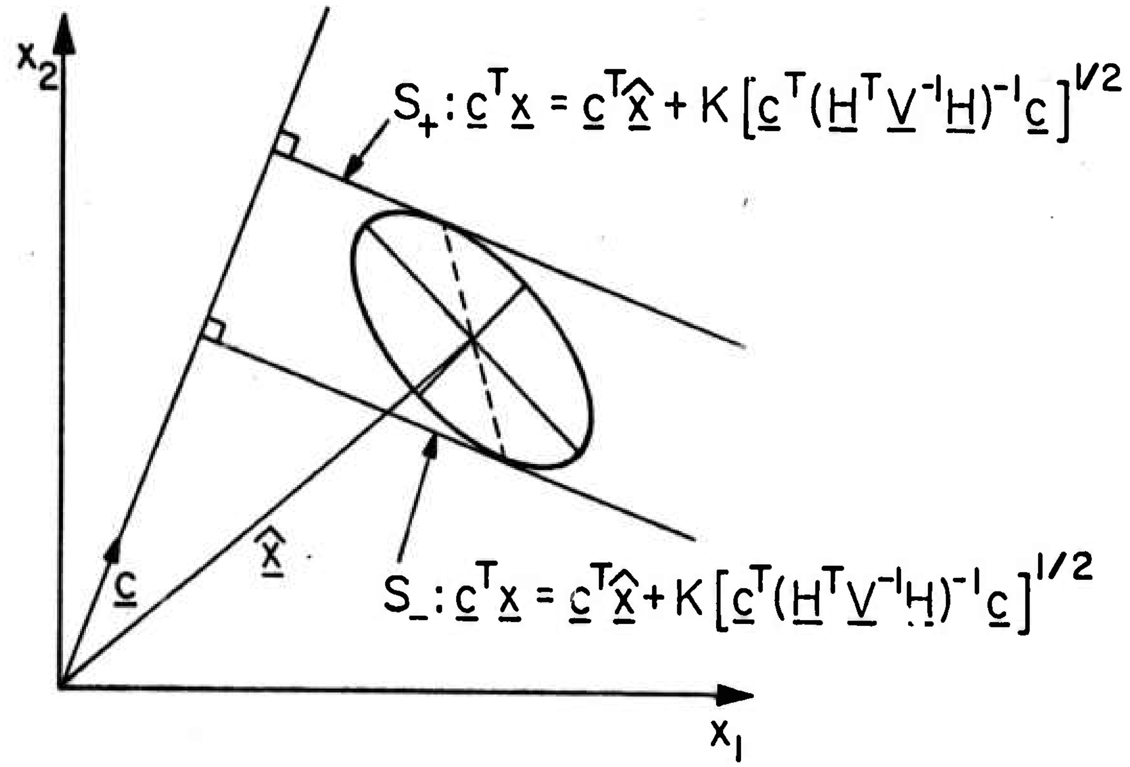


Figure (3.2-2) The Determination of the Confidence Interval for Parametric Functions

$$\underline{v}^T \underline{H}^T \underline{V}^{-1} \underline{H}\underline{v} = K^2 \qquad (3.2\text{-}30)$$

In order to find the directions of the principal axes of the K - ellipsoid, first observe that a radius vector from the origin to any point on the surface of the ellipsoid will be colinear with one of the principal axes if and only if it will be in the direction of the normal to the surface at that point. On the other hand, the ellipsoid can be considered as an equipotential surface [3-2, page 45] of the scalar field

$$\beta(\underline{v}) = \underline{v}^T \underline{H}^T \underline{V}^{-1} \underline{H}\underline{v} \qquad (3.2\text{-}31)$$

so that the normal to the surface can be obtained by the direction of the gradient vector

$$\underline{\nabla}(\beta) = 2\underline{H}^T \underline{V}^{-1} \underline{H}\underline{v} \qquad (3.2\text{-}32)$$

Consequently, the problem of finding the principal axes of the ellipsoid reduces to the one of finding axes that are colinear with the gradient vector. This is expressed by the following equation in $\underline{p}$, for some constant λ

$$\underline{H}^T \underline{V}^{-1} \underline{H}\underline{p} = \lambda \underline{p} \qquad (3.2\text{-}33)$$

The previous equation represents an eigenvector-eigenvalue problem and the fact that $\underline{H}^T \underline{V}^{-1} \underline{H}$ is symmetric and positive definite (since $\underline{V}$ is positive definite and $\underline{H}$ has full column rank in the overdetermined model) guarantees that its eigenvalues are all real and positive. The eigenvectors can always be chosen to be mutually orthogonal and these will be the directions of the principal axes.

Consider now the diagonal matrix $\underline{\Omega}^2$ containing the eigenvalues of $\underline{H}^T \underline{V}^{-1} \underline{H}$ in decreasing order. Consider also the unitary matrix $\underline{P}$ such that its columns are the normalized corresponding eigenvectors. The matrix $\underline{\Omega}^2$ is obtained by the following transformation

$$\underline{P}^T \underline{H}^T \underline{V}^{-1} \underline{H}\underline{P} = \underline{\Omega}^2 \tag{3.2-34}$$

In order to obtain the axes of the ellipsoid a rotation of coordinates is performed

$$\underline{r} = \underline{P}^T \underline{v} \tag{3.2-35}$$

This will align the axes of the ellipsoid with the axes of the coordinate system. By solving for $\underline{v}$ in the previous equation and substituting in (3.2-30), the following expression is obtained

$$\underline{r}^T \underline{P}^T \underline{H}^T \underline{V}^{-1} \underline{H}\underline{P}\underline{r} = K^2 \tag{3.2-36}$$

and, using (3.2-34), this reduces to

$$\underline{r}^T \underline{\Omega}^2 \underline{r} = K^2 \tag{3.2-37}$$

Taking into consideration that $\underline{\Omega}^2$ is diagonal, with entries $w_1, w_2, \cdots, w_n$ in the diagonal, the previous expression can be rewritten as

$$\sum_{i=1}^{N} \frac{r_i^2}{K^2/w_i^2} = 1 \tag{3.2-38}$$

This is the canonical form of the equation of an ellipsoid when the axes are colinear with the coordinate axes. The lengths of these axes are given by

$$l_i = \frac{2K}{w_i}, \quad i = 1, 2, \cdots, N \tag{3.2-39}$$

It also follows that the principal axes of the ellipsoid have lengths inversely proportional to the square root of the corresponding eigenvalues.

Recall that the width of the confidence interval for parametric functions of pixel values is proportional to the distance between the planes of support. Now, if the vector $\underline{c}$ that specifies the parametric function is parallel to an eigenvector that corresponds to a small eigenvalue, the distance between the planes of support will be larger than in the situation where $\underline{c}$ is parallel to an eigenvector corresponding to a large eigenvalue of $\underline{H}^T \underline{V}^{-1} \underline{H}$.

So far the confidence interval for parametric functions of pixel values has been derived under the assumption that the variance of the noise is known to the experimenter. When this is not the case the confidence interval can be determined as follows. First observe that, under white noise conditions, $\underline{c}^T \hat{\underline{x}} - \underline{c}^T \underline{x}$ is normally distributed with zero mean and variance $\sigma^2 \underline{c}^T (\underline{H}^T \underline{H})^{-1} \underline{c}$. Therefore, the ratio

$$\frac{\underline{c}^T (\hat{\underline{x}} - \underline{x})}{\sigma \sqrt{\underline{c}^T (\underline{H}^T \underline{H})^{-1} \underline{c}}} \tag{3.2-40}$$

should be a standardized Gaussian random variable. The parameter σ is not known but it has already been derived that $\hat{\underline{x}}$ and $s^2 = \frac{1}{M - N} (\underline{y} - \underline{H}\,\underline{x})^T (\underline{y} - \underline{H}\,\underline{x})$ are independently distributed. Therefore, the ratio given by (3.2-40) and $\frac{(M-N)s^2}{\sigma^2}$ are also independent. Since $\frac{(M-N)s^2}{\sigma^2}$ is χ^2 distributed with M-N degrees of freedom, it follows that

$$\frac{\dfrac{\underline{c}^T (\hat{\underline{x}} - \underline{x}) \cdot \sqrt{M-N}}{\sigma \sqrt{\underline{c}^T (\underline{H}^T \underline{H})^{-1} \underline{c}}}}{\sqrt{\dfrac{(M-N)s^2}{\sigma^2}}} = \frac{\underline{c}^T (\hat{\underline{x}} - \underline{x})}{s \sqrt{\underline{c}^T (\underline{H}^T \underline{H})^{-1} \underline{c}}} \tag{3.2-41}$$

is distributed according to a Student distribution with (M-N) degrees of freedom. The determination of the confidence interval is then easily done by using tables of this distribution. The determination of the confidence interval for the unknown variance can be done by observing that $\frac{(M-N)s^2}{\sigma^2}$ has a chi-square distribution with (M-N) degrees of freedom.

3.3 Analytic Study of the Condition Number

This section considers the effects that perturbations on the observed blurred pixels have on the estimated original pixel values, from the complementary point of view of the numerical analyst. Under this perspective,the estimation of pixel values would consist in the problem of solving a system of linear equations

$$\underline{H}\,\underline{x} = \underline{y} \tag{3.3-1}$$

such that the right hand side is subject to perturbations. These errors represent the role of the noise in the system. Consider the effect of the perturbation vector $\underline{n}$ on the solution of the system of linear equations. Call

$$\underline{\hat{x}} = \underline{x} + \underline{\Delta x} \tag{3.3-2}$$

the solution, $\underline{x}$ being the true vector. The set of normal equations

$$(\underline{H}^T \underline{V}^{-1} \underline{H})(\underline{x} + \underline{\Delta x}) = \underline{H}^T \underline{V}^{-1} \underline{y} = \underline{H}^T \underline{V}^{-1}(\underline{H}\,\underline{x} + \underline{n}) \qquad (3.3\text{-}3)$$

gives the solution for the perturbed system. Reduction of the previous equation gives

$$\underline{H}^T \underline{V}^{-1} \underline{H}\,\underline{\Delta x} = \underline{H}^T \underline{V}^{-1} \underline{n} \qquad (3.3\text{-}4)$$

At this point, assuming the overdetermined model, one could simply invert $(\underline{H}^T \underline{V}^{-1} \underline{H})$ in order to obtain the change $\underline{\Delta x}$ in the solution of the linear system due to the perturbation $\underline{n}$. A decomposition of the matrices involved will be performed, however, giving more insight into the problem [3-2, pages 47-58]. The assumption that $\underline{V}^{-1}$ is positive definite leads to the possibility of a decomposition of the form

$$\underline{V}^{-1} = \underline{C}^T \underline{C} \qquad (3.3\text{-}5)$$

so that equation (3.4-4) can be written as

$$(\underline{H}^T \underline{C}^T \underline{C}\,\underline{H})\underline{\Delta x} = \underline{H}^T \underline{C}^T \underline{C}\,\underline{n} \qquad (3.3\text{-}6)$$

A factorization of the matrix $\underline{C}\,\underline{H}$ will now be performed. This is the so-called singular value decomposition of a rectangular matrix [3-3, page 38]

$$\underline{C}\,\underline{H} = \underline{P}\,\underline{L}^{\frac{1}{2}}\underline{Q} \qquad (3.3\text{-}7)$$

where $\underline{L}$ is the (N x N) diagonal matrix of the eigenvalues of $(\underline{H}^T\ \underline{C}^T\ \underline{C}\,\underline{H})$, $\underline{P}$ is a unitary matrix whose columns are the eigenvectors of $(\underline{C}\,\underline{H}\,\underline{H}^T\ \underline{C}^T)$ and $\underline{Q}$ is also a unitary matrix whose rows are the eigenvectors of $(\underline{H}^T\ \underline{C}^T\ \underline{C}\,\underline{H})$. As a result, equation (3.3-6) can be rewritten

$$(\underline{Q}^T\ \underline{L}\,\underline{Q})\underline{\Delta x} = \underline{Q}^T\ \underline{L}^{\frac{1}{2}}\underline{P}^T\ \underline{C}\,\underline{n} \tag{3.3-8}$$

or

$$\underline{L}\,\underline{Q}\,\underline{\Delta x} = \underline{L}^{\frac{1}{2}}\underline{P}^T\ \underline{C}\,\underline{n} \tag{3.3-9}$$

Since $(\underline{H}^T\ \underline{C}^T\ \underline{C}\,\underline{H})$ is nonsingular, $\underline{L}$ is also nonsingular and one obtains (by multiplying both sides of the equation by $\underline{Q}^T\ \underline{L}^{-1}$)

$$\underline{\Delta x} = \underline{Q}^T\ \underline{L}^{-\frac{1}{2}}\underline{P}^T\ \underline{C}\,\underline{n} \tag{3.3-10}$$

The previous equation can also be written

$$\underline{\Delta x} = \sum_{i=1}^{N} \frac{(\underline{P}^T\ \underline{C}\,\underline{n})_i}{w_i} \cdot \underline{q}_i \tag{3.3-11}$$

where $\underline{q}_i$ are the rows of $\underline{Q}$ (eigenvectors of $\underline{H}^T\ \underline{C}^T\ \underline{C}\,\underline{H}$) and w_i, $i = 1, 2, \ldots N$ are the square root of the eigenvalues of $(\underline{H}^T\ \underline{C}^T\ \underline{C}\,\underline{H})$. The last quantities are called the singular values of the matrix $\underline{C}\,\underline{H}$.

Equation (3.3-11) shows that the component of the error along each of the eigenvectors of $(\underline{H}^T\ \underline{V}^{-1}\ \underline{H})$ is inversely proportional to the corresponding singular value of $\underline{C}\,\underline{H}$. Assuming that the components of $(\underline{P}^T \underline{C}\,\underline{n})$ do not vary much in magnitude, the components of $\underline{\Delta x}$ will tend to be larger in the direction of eigenvectors corresponding to smaller singular values.

So far the analysis of perturbations has been restricted to absolute changes in the least squares solution due to errors in the observed values. The next step consists in analyzing relative changes in the solution due to perturbations in the data as well as in the matrix $\underline{H}$.

Assume for the sake of simplicity, that the noise is white. This implies that the solution to the normal equations is given by

$$\underline{\hat{x}} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{y} \qquad (3.3\text{-}12)$$

As pointed out before, the previous expression can be put into the form

$$\underline{\hat{x}} = \underline{H}^+ \underline{y} \qquad (3.3\text{-}13)$$

where $\underline{H}^+$ is the pseudoinverse of $\underline{H}$.

Call

$$\underline{\bar{y}} = \underline{y} + \underline{n} \qquad (3.3\text{-}14)$$

and let $\underline{y}_1$ and $\underline{\bar{y}}_1$ be the projections of $\underline{y}$ and $\underline{\bar{y}}$ respectively, onto the range of the transformation $\underline{H}$, denoted $R(\underline{H})$. Under these conditions, the following bound is valid for the relative changes in the solution to the least squares problems [3-8, page 221]

$$\frac{\| \underline{H}^{+}\underline{y} - \underline{H}^{+}\overline{\underline{y}} \|}{\| \underline{H}^{+}\underline{y} \|} \leq c(\underline{H}) \cdot \frac{\| \underline{y}_1 - \overline{\underline{y}}_1 \|}{\| \overline{\underline{y}}_1 \|} \tag{3.3-15}$$

where $c(\underline{H}) = \| \underline{H} \| . \| \underline{H}^{+} \|$. (3.3-16)

The quantity $c(\underline{H})$ is called the condition number of the blur matrix $\underline{H}$. It plays an extremely important role in explaining the effect of perturbations on the accuracy of the computations involved.

Equation (3.3-15) can be obtained by the following reasoning. Decompose $\underline{y}$ into $\underline{y}_1$ and $\underline{y}_2$ where $\underline{y}_1$ belongs to $R(\underline{H})$ and $\underline{y}_2$ is in its orthogonal complement, which is the null space of $\underline{H}^T$, denoted by $N(\underline{H}^T)$. Therefore,

$$\underline{H}^{+}\underline{y} = \underline{H}^{+}\underline{y}_1 + \underline{H}^{+}\underline{y}_2 = \underline{H}^{+}\underline{y}_1 + (\underline{H}^T\underline{H})\underline{H}^T\underline{y}_2 = \underline{H}^{+}\underline{y}_1 \tag{3.3-17}$$

Analogously,

$$\underline{H}^{+}\overline{\underline{y}} = \underline{H}^{+}\overline{\underline{y}}_1 \tag{3.3-18}$$

where $\overline{\underline{y}}_1$ is the projection of $\overline{\underline{y}}$ onto $R(\underline{H})$. Hence

$$\| \underline{H}^{+}\underline{y} - \underline{H}^{+}\overline{\underline{y}} \| = \| \underline{H}^{+}(\underline{y}_1 - \overline{\underline{y}}_1) \| \leq \| \underline{H}^{+} \| . \| \underline{y}_1 - \overline{\underline{y}}_1 \| \tag{3.3-19}$$

On the other hand, it is easily shown [3-8, page 220] that if $\underline{H}^+\underline{y}$ is the solution to the least squares problem, then

$$\underline{H}\,\underline{H}^+\underline{y} = \underline{y}_1 \tag{3.3-20}$$

and

$$\|\underline{y}_1\| = \|\underline{H}\,\underline{H}^+\underline{y}\| \leq \|\underline{H}\| \cdot \|\underline{H}^+\underline{y}\| \tag{3.3-21}$$

or

$$\|\underline{H}^+\underline{y}\| \geq \frac{\|\underline{y}_1\|}{\|\underline{H}\|} \tag{3.3-22}$$

By dividing (3.3-19) by (3.3-22) one obtains the desired result

Observe that it is only the component of the relative error in the observed vector of pixel values lying on the range of the blur matrix that contributes to the relative error in the estimated pixel values.

The condition number will determine the effect of the noise in the restoration process. If its value is small a little relative perturbation on the observed blurred picture will not produce large relative changes in the restored picture. In this case, the normal equations are said to be well conditioned. If, on the other hand, the condition number has a large value, small relative changes in the observed values may greatly affect the estimated pixel values and the normal equations are said to be ill conditioned.

The matrix norm used in the expression of the condition number $\|\underline{H}\| \cdot \|\underline{H}^{+}\|$ can be any one of the matrix norms that are consistent with the vector norm used. In particular, one may select the spectral norm, which is equal to the largest singular value of the matrix.

In order to find the largest singular value of the pseudoinverse $\underline{H}^{+}$, (or $(\underline{C}\,\underline{H})^{+}$ if a colored noise corrupts the image) the factorization expressed by equation (3.3-7) will be used. By doing this and also using the expression for the pseudoinverse one obtains

$$(\underline{C}\,\underline{H})^{+} = [(\underline{C}\,\underline{H})^{T}\,\underline{C}\,\underline{H}]^{-1}\,(\underline{C}\,\underline{H})^{T} = [\underline{Q}^{T}\underline{L}^{\frac{1}{2}}\underline{P}^{T}\underline{P}\,\underline{L}^{\frac{1}{2}}\,\underline{Q}]^{-1}\,\underline{Q}^{T}\underline{L}^{\frac{1}{2}}\,\underline{P}^{T} \tag{3.3-23}$$

But since

$$\underline{P}^{T}\underline{P} = \underline{Q}\,\underline{Q}^{T} = \underline{I} \tag{3.3-24}$$

and

$$(\underline{Q}^{T}\underline{L}^{1}\underline{Q})^{-1} = \underline{Q}^{T}\underline{L}^{-1}\underline{Q} \tag{3.3-25}$$

it follows that

$$(\underline{C}\,\underline{H})^{+} = \underline{Q}^{T}\,\underline{L}^{-\frac{1}{2}}\,\underline{P}^{T} \tag{3.3-26}$$

The matrix $(\underline{C}\,\underline{H})^{+}$ is an $N \times M (M \geq N)$ matrix, so its singular values are calculated by the positive square roots of the matrix

$$(\underline{C}\,\underline{H})^{+}\,[(\underline{C}\,\underline{H})^{+}]^{T} = \underline{Q}^{T}\underline{L}^{-1}\underline{Q} \tag{3.3-27}$$

The eigenvalues of this matrix in factorized form are given by the diagonal elements of L^{-1}, namely, $\frac{1}{w_1^2}, \frac{1}{w_2^2}, \ldots, \frac{1}{w_n^2}$. As a consequence, the singular values of $(\underline{C}\,\underline{H})^{+}$ are the reciprocal of the singular values of $\underline{C}\,\underline{H}$. The largest singular of $(\underline{C}\,\underline{H})^{+}$ is $\frac{1}{w_n}$. Therefore, the condition number is given by

$$\|(\underline{C}\,\underline{H})\|\,\|(\underline{C}\,\underline{H})^{+}\| = \frac{w_1}{w_n} \tag{3.3-28}$$

that means that this number is the ratio of the largest to the smallest singular value of the matrix $(\underline{C}\,\underline{H})$.

A further insight can be obtained by considering the relationship between the condition number and the K-ellipsoid [3-2, page 54]. Equation (3.2-39) expresses the relationship between the length of an axis of the ellipsoid and the corresponding singular value. Using that expression one may immediately conclude that the ratio of the largest to the smallest singular value of $(\underline{C}\,\underline{H})$ is also the ratio of the largest to the smallest axis of the ellipsoid. This means that the more the ellipsoid departs from the shape of a sphere, the more ill conditioned the restoration problem will be. Figure (3.3-1) obtained from reference [3-2, page 54] shows the

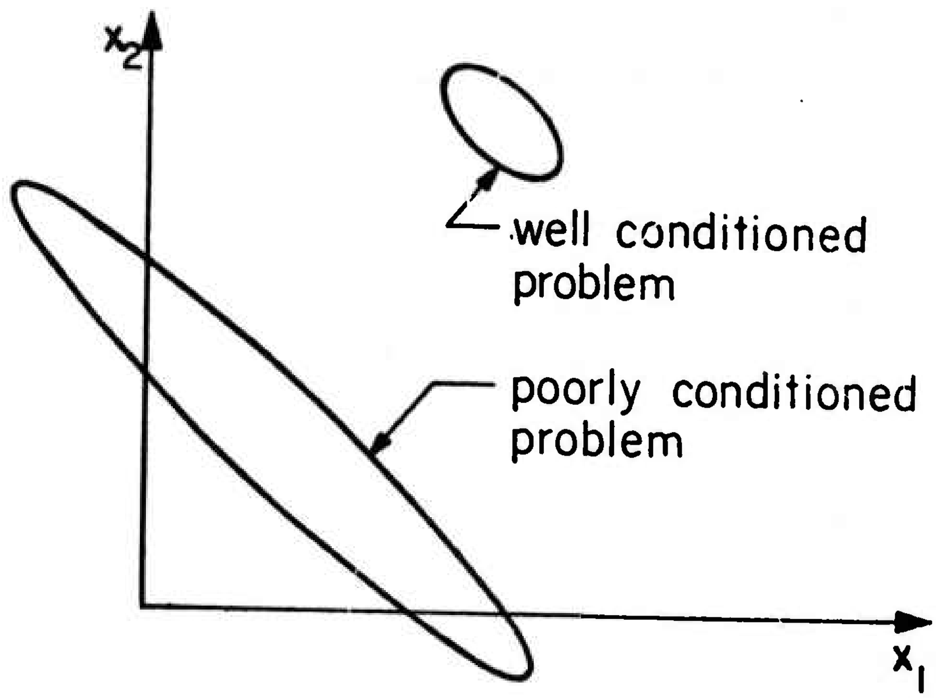


Figure (3.3-1) K-Ellipsoids for a Well Conditioned and a Poorly Conditioned Problem comparison between the K-ellipsoids for the well and ill conditioned system, for the case where N = 2.

When the noise is Gaussian, the observed blurred and noisy pixel values are also normally distributed, under the linear model. In this case one can define ellipsoids centered at the mean value of the observed pixel values (it may be assumed to be the origin for convenience) and containing a given percentage of this multivariate distribution.

Since the estimated pixel values are obtained by combining linearly the observed values, it follows that the estimators are also Gaussian distributed.

It is possible to show [reference 3-2, pages 55-58] that if the ellipsoid for the observed values in a regression model has the expression

$$(\underline{y} - \underline{H}\underline{x})^T \underline{V}^{-1} (\underline{y} - \underline{H}\underline{x}) \leq \rho^2 \qquad (3.3\text{-}29)$$

then the corresponding ellipsoids of the estimators are given by

$$(\underline{x} - \underline{\hat{x}})^T \underline{H}^T \underline{V}^{-1} \underline{H}(\underline{x} - \underline{\hat{x}}) \leq \rho^2 \qquad (3.3\text{-}30)$$

This ellipsoid essentially gives the multidimensional confidence interval for the pixel values under the normality assumption. The eigenvectors and eigenvalues ($\underline{H}^T \underline{V}^{-1} \underline{H}$) will determine the size and shape of the ellipsoid.

The attention will be devoted now to the problem of considering the effects of perturbations of the blur matrix $\underline{H}$ on the restoration problem. This question is of extreme importance since the experimenter rarely knows with great precision the spread function. This is particularly true when that function is derived from measurements that inevitably involve errors.

The analysis of the effect of the perturbation on the blur matrix is quite involved. In order to do this a new terminology is introduced by Stewart [3-9]. Let $\underline{E}$ be a perturbation matrix on the blur matrix $\underline{H}$ and S be a subspace of R^M. Each column of $\underline{E}$ is an M-vector that can be projected onto S. Call $\underline{E}_1(\underline{y}_1)$ and $\underline{E}_2(\underline{y}_2)$ the projections of $\underline{E}(\underline{y})$ onto the range of the blur matrix, denoted by $R(\underline{H})$ and its orthogonal complement $(N(\underline{H}^T))$, respectively.

Assuming the overdetermined model and if

$$\|\underline{H}^+\| \cdot \|\underline{E}_1\| < \tfrac{1}{2} \tag{3.3-31}$$

then the columns of $(\underline{A} + \underline{E})$ are linearly independent. Also, assuming that

$$\hat{\underline{x}} = \underline{H}^+ \underline{y} \tag{3.3-32}$$

and

$$\bar{\hat{\underline{x}}} = (\underline{H} + \underline{E})^{+} \underline{y} \qquad (3.3\text{-}33)$$

then

$$\frac{\|\hat{\underline{x}} - \bar{\hat{\underline{x}}}\|}{\|\hat{\underline{x}}\|} \leq 2c(\underline{H}) \cdot \frac{\|\underline{E}_1\|}{\|\underline{H}\|} + 4c^2(\underline{H}) \cdot \frac{\|\underline{E}_2\|}{\|\underline{H}\|} \cdot \frac{\|\underline{y}_2\|}{\|\underline{y}_1\|} + 8c^3(\underline{H}) \frac{\|\underline{E}_2\|^2}{\|\underline{H}\|^2}$$

(3.3-34)

$$\text{where } c(\underline{H}) = \|\underline{H}\| \cdot \|\underline{H}^{+}\| \qquad (3.3\text{-}35)$$

is the condition number of $\underline{H}$ and the consistent Euclidean norm for vector and Frobenius norm for matrices is used. The third term in the bound depends on the square of $\|\underline{E}_2\|^2$ and usually can be disregarded when compared to the first two terms. The first term is similar to the bound that can be derived for perturbations in non-singular linear systems.

The second term states that the relative perturbation in $N(\underline{H}^T)$ is amplified by $c(\underline{H}) \cdot \frac{\|\underline{y}_2\|}{\|\underline{y}_1\|}$. Since $\underline{y}_1$ is the projection of $\underline{y}$ onto $R(\underline{H})$ and $\underline{y}_2$ is the projection on its orthogonal complement, namely $N(\underline{H}^T)$, it follows that the ratio $\frac{\|\underline{y}_2\|}{\|\underline{y}_1\|}$ measures how nearly $\underline{y}$ lies with respect to $R(\underline{H})$. If $\underline{y}$ is close to $R(\underline{H})$ this ratio will be small. If $\|\underline{E}_1\|$ and $\|\underline{E}_2\|$ are of the same order of magnitude then the first term tends to dominate when $\underline{y}_2$ is small. If, on the other hand, $\underline{y}_2$ is large, the second term is prevalent. Stewart states loosely that "if $\underline{y}$ very nearly lies in $R(\underline{H})$, then $c(\underline{H})$ is the condition number for the least squares problem, otherwise $c^2(\underline{H})$ is the condition number."

An important conclusion in the case of image restoration can be drawn. If there is a small amount of noise present in the observed pixel values (high signal to noise ratio), the $\underline{y}$ will tend to be near $R(\underline{H})$, which implies that $c(\underline{H})$ will be the condition number. If, on the contrary, the signal to noise ratio is low, the component of the noise will tend to place $\underline{y}$ farther away from $R(\underline{H})$ and in this case $c^2(\underline{H})$ will be the condition number. Since $c(\underline{H})$ is always greater or equal to one, the latter case is certainly a worse situation. Incidentally, it should be remarked that $c^2(\underline{H})$ is the condition number of the matrix $(\underline{H}^T\underline{H})$.

3.4 The Underdetermined Model

So far the study of the image restoration problem has been essentially restricted to the overdetermined model. This means that the $(M \times N)$ blur matrix $\underline{H}$ is assumed to have rank N. In other words, the columns of this matrix are supposed to be linearly independent.

On the other hand, if, in the discretization method of the continuous planar equation that describes the blurring process, a number of nodes for the quadrature formula is selected exceeding the number of observed values (i.e., $M < N$), this condition is violated and the rank of the blur matrix $\underline{H}$ is necessarily less than $\underline{N}$.

Under the overdetermined model there was a unique solution for the restoration problem given by the set of normal equations. Furthermore, estimators and finite confidence intervals were obtained for every parametric function of the pixel values. Also, a finite condition number was obtained by considering the ratio of the largest and the smallest singular values of the matrix $\underline{C}\underline{H}$.

If $\underline{H}$ has not full column rank several important consequences are immediately derived. First, the uniqueness of the set of normal equations cannot be guaranteed any more, since the matrix $(\underline{H}^T\underline{V}^{-1}\underline{H})$ is singular and therefore cannot be inverted. Second, the smallest singular value of the matrix $\underline{C}\underline{H}$ is zero, resulting in a condition number with an infinite value. As a consequence of this fact, many linear combinations of pixel values have an infinite confidence interval, which is equivalent to say that these functions are not estimable

There is a concept that not only is necessary for the study of underdetermined systems but also broadens the view over the overdetermined systems, unifying the whole study of the linear model in regression analysis. It is the concept of the generalized inverse of a matrix, which was mentioned briefly in connection with the treatment of the overdetermined model and now is more fully treated.

Initially, a brief survey of generalized inverse concepts will be presented. There are several ways of presenting these concepts. The presentation contained in reference [3-3] will be followed.

Given an M x N matrix $\underline{H}$, the matrix $\underline{H}^+$ obtained by the following limiting operations

$$\underline{H}^+ = \lim_{\alpha \to 0} (\underline{H}^T\underline{H} + \alpha^2 \underline{I})^{-1}\underline{H}^T$$

$$= \lim_{\alpha \to 0} \underline{H}^T(\underline{H}\underline{H}^T + \alpha^2 \underline{I})^{-1} \tag{3.4-1}$$

always exists. Also, for any (M x 1) vector $\underline{y}$, the vector

$$\hat{\underline{x}} = \underline{H}^+\underline{y} \tag{3.4-2}$$

is the vector of minimum norm among those that minimize

$$\|\underline{y} - \underline{H}\underline{x}\|^2 \tag{3.4-3}$$

It can be shown that $\hat{\underline{x}}$ is the unique vector in $R(\underline{H}^T)$ satisfying the equation

$$\underline{H}\underline{x} = \hat{\underline{y}} \tag{3.4-4}$$

where $\hat{\underline{y}}$ is the projection of $\underline{y}$ on $R(\underline{H})$. This vector $\hat{\underline{x}}$ satisfies the set of normal equations

$$\underline{H}^T\underline{H}\underline{x} = \underline{H}^T\underline{y} \tag{3.4-5}$$

The unique matrix $\underline{H}^+$ is called the generalized inverse or the pseudoinverse of the matrix $\underline{H}$.

As a corollary of expressions (3.4-1) and (3.4-2), it follows that for any vector $\underline{y}$, $\underline{H}\underline{H}^+\underline{y}$ is the projection of $\underline{y}$ on $R(\underline{H})$, $(\underline{I}-\underline{H}\underline{H}^+)\underline{y}$ is the projection of $\underline{y}$ on $N(\underline{H}^T)$. Furthermore, for any vector $\underline{x}$, $\underline{H}^+\underline{H}\underline{x}$ is the projection of $\underline{x}$ on $R(\underline{H}^T)$ and $(\underline{I}-\underline{H}^+\underline{H})\underline{x}$ is the projection of $\underline{x}$ on $N(\underline{H})$. It should be observed at this point that a projection matrix $\underline{P}$ is idempotent, i.e., $\underline{P}^2 = \underline{P}$.

If $\underline{H}$ is square and nonsingular, $\underline{H}^+$ is the inverse of $\underline{H}$, $\underline{H}^{-1}$. If the columns of $\underline{H}$ are linearly independent, like in the overdetermined model, the pseudoinverse is given by

$$\underline{H}^+ = (\underline{H}^T\underline{H})^{-1}\underline{H}^T \tag{3.4-6}$$

If, on the other hand, the rows of $\underline{H}$ are linearly independent, the pseudoinverse will be represented as

$$\underline{H}^+ = \underline{H}^T(\underline{H}\underline{H}^T)^{-1} \tag{3.4-7}$$

A better perspective over the pseudoinverse can be obtained by considering some specific cases. Take, for example, the (1 x 1) matrix $\underline{H}$, represented by the value h. In this case,

$$\begin{aligned} \underline{H}^+ &= 0 \quad \text{if } h = 0 \\ &= \frac{1}{h} \quad \text{if } h \neq 0 \end{aligned} \tag{3.4-8}$$

If $\underline{H}$ is diagonal,

$$\underline{H} = \text{diag}\,(h_1, h_2, \cdots h_M) \tag{3.4-9}$$

then

$$\underline{H}^+ = \text{diag}\,(h_1^+, h_2^+, \cdots h_M^+) \tag{3.4-10}$$

where

$$\begin{aligned} h_i^+ &= 0 \quad \text{if } h_i = 0 \\ &= \frac{1}{h_i} \quad \text{if } h_i \neq 0 \end{aligned} \tag{3.4-11}$$

If $\underline{H}$ is a symmetric (M x M) matrix, it is possible to represent it in the following form

$$\underline{H} = \underline{T}\,\underline{D}\,\underline{T}^T \tag{3.4-12}$$

where $\underline{T}$ is an orthogonal matrix and $\underline{D}$ is diagonal. Using (3.4-1), $\underline{H}^+$ can be expressed as

$$\begin{aligned} \underline{H}^+ &= \lim_{\alpha \to 0} \underline{T}(\underline{D}^2 + \alpha^2\underline{I})^{-1}\underline{D}\,\underline{T}^T \\ &= \underline{T}\lim_{\alpha \to 0}(\underline{D}^2 + \alpha^2\underline{I})^{-1}\underline{D}\,\underline{T}^T \\ &= \underline{T}\,\underline{D}^+\,\underline{T}^T \end{aligned} \tag{3.4-13}$$

As a result, the pseudoinverse of a symmetric matrix can be obtained by pseudoinverting the diagonal matrix that consists of its eigenvalues. If $\underline{H}$ is nonsingular, all the eigenvalues are nonzero and $\underline{D}^{+} = \underline{D}^{-1}$, so that $\underline{H}^{+} = \underline{H}^{-1}$.

This result on symmetric matrices leads to spectral representations for the pseudoinverse matrix. If the columns of $\underline{T}$ are denoted by $\underline{t}_1, \underline{t}_2, \ldots, \underline{t}_M$ and the eigenvalues of $\underline{H}$ by $\lambda_1, \lambda_2, \cdots, \lambda_M$ the matrix $\underline{H}$ can be represented as

$$\underline{H} = \sum_{i=1}^{M} \lambda_i \underline{t}_i \underline{t}_j^T \qquad (3.4\text{-}15)$$

and the pseudoinverse $\underline{H}^{+}$ by

$$\underline{H}^{+} = \sum_{i=1}^{M} \lambda_i^{+} \underline{t}_i \underline{t}_j^T \qquad (3.4\text{-}15)$$

where λ_i^{+} has the same meaning as in (3.4-11).

Two results that will be useful in the analysis of the underdetermined model of the restoration process are now stated. For any matrix $\underline{H}$, $\underline{x}$ belongs to the null space of $\underline{H}$ if and only if

$$\underline{x} = (\underline{I} - \underline{H}^{+}\underline{H})\underline{y} \qquad (3.4\text{-}16)$$

for some vector $\underline{y}$. For any matrix $\underline{H}$, $\underline{z}$ belongs to the range of $\underline{H}$ if and only if

$$\underline{z} = \underline{H}\underline{H}^{+}\underline{u} \tag{3.4-17}$$

for some vector $\underline{u}$.

For any rectangular matrix $\underline{H}$, the pseudoinverse can be expressed in terms of pseudoinverses of symmetric square matrices as follows

$$\underline{H}^{+} = (\underline{H}^{T}\underline{H})^{+}\underline{H}^{T} = \underline{H}^{T}(\underline{H}\underline{H}^{T})^{+} \tag{3.4-18}$$

It can be shown that an entirely equivalent way of introducing the pseudoinverse exists. This is the so called Penrose characterization [3-3, page 28]. A matrix $\underline{H}^{+}$ is said to be the pseudoinverse of a matrix $\underline{H}$ if and only if the four conditions are satisfied

$\underline{H}\underline{H}^{+}$ and $\underline{H}^{+}\underline{H}$ are symmetric

$$\underline{H}\underline{H}^{+}\underline{H} = \underline{H}$$

and

$$\underline{H}^{+}\underline{H}\underline{H}^{+} = \underline{H}^{+}$$

These results can now be applied to the restoration problem. Consider first the no noise case

$$\underline{y} = \underline{H}\underline{x} \tag{3.4-20}$$

where $\underline{x}$ represents the vector of pixel values, $\underline{H}$ is the blur matrix and $\underline{y}$ is the vector of observed values. No restriction is placed on the dimension of the blur matrix, either the overdetermined or the underdetermined models could be involved.

Consider first the problem of existence of solution. In order for a solution to exist $\underline{y}$ must be on $R(\underline{H})$. By (3.4-17) this occurs if and only if $\underline{y} = \underline{H}\underline{H}^+\underline{u}$ for some $\underline{u}$. On the other hand, since $\underline{H}\underline{H}^+$ is a projection on $R(\underline{H})$, it follows that

$$\underline{H}\underline{H}^+\underline{y} = (\underline{H}\underline{H}^+)^2\underline{u} = \underline{H}\underline{H}^+\underline{u} = \underline{y} \qquad (3.4\text{-}21)$$

The condition expressed by the previous equation is the so called consistency condition for the solution of a linear system.

At this point it is perhaps useful to point out that, under no noise, for real situations, $\underline{y}$ will always be in $R(\underline{H})$ since it was obtained by blurring an existing picture. This is why the restoration problem is then formulated as searching for the solution of the linear system (3.4-20) instead of directly solving for the least squares problem.

Turning now to the problem of uniqueness of the solution, the homogeneous system $\underline{H}\underline{x} = \underline{0}$ has to be investigated. Observe that, for any vector $\underline{v}$

$$\underline{x} = (\underline{I} - \underline{H}^+\underline{H})\underline{v} \qquad (3.4\text{-}22)$$

is a solution to the homogeneous system since

$$\underline{H}\underline{x} = \underline{H}(\underline{I}-\underline{H}^{+}\underline{H})\underline{v} = (\underline{H}-\underline{H}\underline{H}^{+}\underline{H})\underline{v} = \underline{0} \tag{3.4-23}$$

where in the last equality one of the conditions expressed by (3.4-19) was used. Therefore, if $\underline{H}^{+}\underline{H} \neq \underline{I}$, a nonzero vector can always be added to the solution, without changing the left hand side of system (3.4-20). A necessary and sufficient condition can be expressed by

$$\underline{H}^{+}\underline{H} = \underline{I} \tag{3.4-24}$$

this condition being equivalent to the statement that $N(\underline{H})$ consists only of the zero vector.

A general solution to the linear system (3.4-20) can be expressed as

$$\underline{x} = \underline{H}^{+}\underline{y} + (\underline{I}-\underline{H}^{+}\underline{H})\underline{v} \tag{3.4-25}$$

where $\underline{v}$ is an arbitrary vector.

In the case of the overdetermined system, the blur matrix $\underline{H}$ has linearly independent columns, $\underline{H}^{+}$ is expressed by (3.4-6) and the condition (3.4-24) is satisfied so that the unique solution is given by

$$\underline{x} = \underline{H}^{+}\underline{y} \tag{3.4-26}$$

For the underdetermined model, condition (3.4-24) is not satisfied and there will not be a unique solution to the system of linear equations.

The vector $\underline{H}^+\underline{y}$ is the minimum norm solution. This can be verified by noticing that the set of vectors given by $(\underline{I}-\underline{H}^+\underline{H})\underline{v}$ is orthogonal to $\underline{H}^+\underline{y}$ as shown below

$$\begin{aligned} [(\underline{I}-\underline{H}^+\underline{H})\underline{v}]^T \cdot \underline{H}^+\underline{y} &= \underline{v}^T(\underline{I}-\underline{H}^+\underline{H})^T\underline{H}^+\underline{y} \\ &= \underline{v}^T(\underline{I}-\underline{H}^+\underline{H})\underline{H}^+\underline{y} = \underline{v}^T(\underline{H}^+-\underline{H}^+\underline{H}\,\underline{H}^+)\underline{y} = \underline{0} \end{aligned} \qquad (3.4\text{-}27)$$

where the second equality used the fact that $(\underline{I}-\underline{H}^+\underline{H})$ is a symmetric matrix and the fourth equality was based on one of the relations in (3.4-19). Figure (3.4-1) taken from reference [3-2, page 63], shows the geometry of the solutions to the linear system in the underdetermined case, when N = 2 and the dimension of $N(\underline{H})$ is 1.

Now, suppose that noise is added to the system so that

$$\underline{y} = \underline{H}\underline{x} + \underline{n} \qquad (3.4\text{-}28)$$

In this case one would search for an estimator $\hat{\underline{x}}$ of $\underline{x}$ under some meaningful statistical criterion. In the overdetermined case the best linear unbiased estimator (B.L.U.E.) has already been obtained and it was shown to be unique. Suppose, therefore, that one is looking for a B.L.U.E. estimator in the underdetermined model, where rank $(\underline{H}) < N$. Assume that a linear estimator of the form

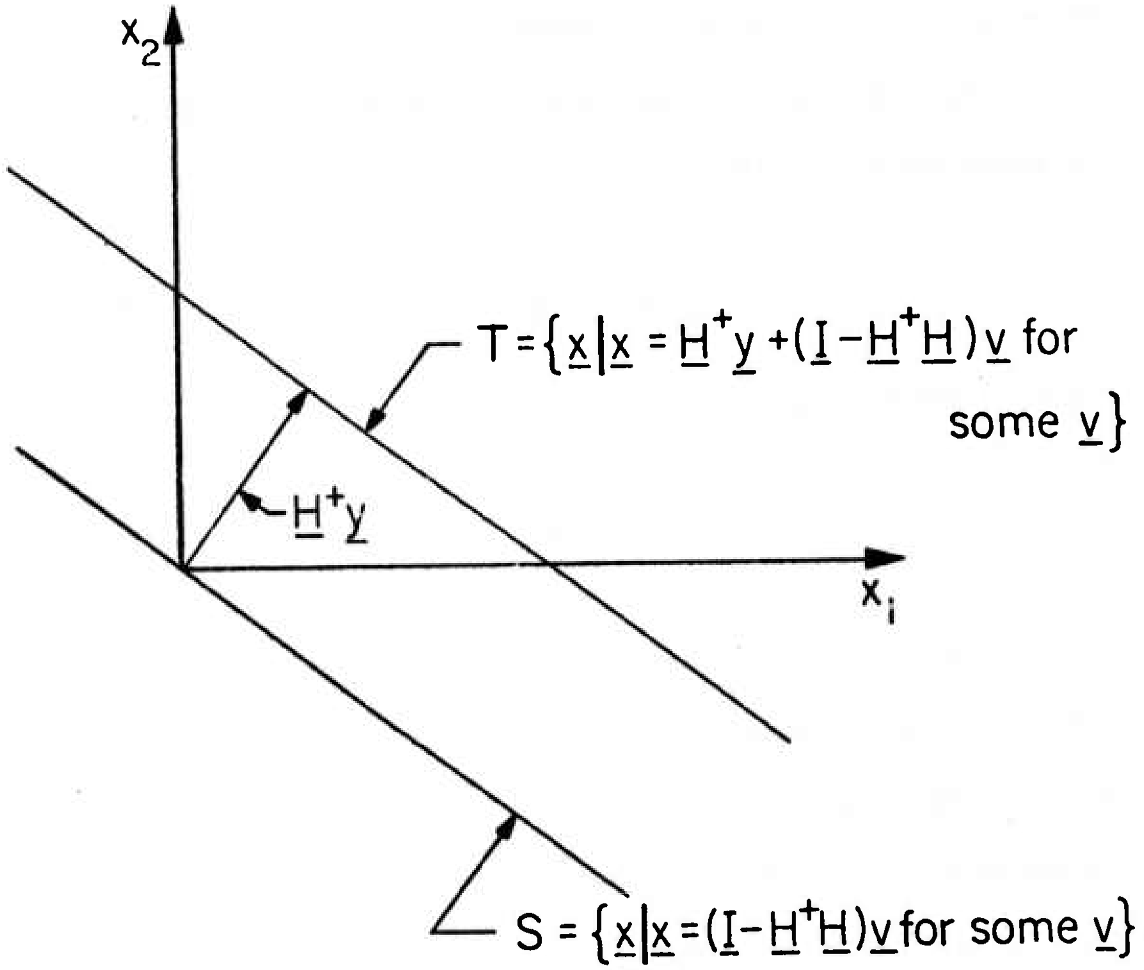


Figure (3.4-1) The Geometry of the Solutions of the Underdetermined System

$$\underline{x} = \underline{U}^T \underline{y} \qquad (3.4\text{-}29)$$

is used, where $\underline{U}$ is an (M x N) matrix.

By imposing the unbiasedness condition, the following expression has to be valid, for any value of $\underline{x}$

$$\underline{U}^T \underline{H} \underline{x} = \underline{x} \qquad (3.4\text{-}30)$$

Thus, it follows that

$$\underline{U}^T \underline{H} = \underline{I}_N \qquad (3.4\text{-}31)$$

But if this is true, the rank of $\underline{I}_N$, N would be larger than the rank of one of its factors ($\underline{H}$, with rank < N) which contradicts the Sylvester Inequality for the product of matrices. Therefore, there is no unbiased linear estimator for the vector of pixel values $\underline{x}$.

This fact greatly limits the usefulness of the underdetermined model. This can be viewed from the perspective of being the price paid for increasing the number of quadrature nodes above the number of observed values. Because of lack of information an unbiased estimator for the pixel values cannot be obtained.

However, restoration can be attempted according to another criterion, namely, the minimization of the least squares quadratic form

$$\theta(x) = (\underline{y} - \underline{H}\,\underline{x})\underline{V}^{-1}(\underline{y} - \underline{H}\,\underline{x}) \tag{3.4-32}$$

where $\underline{V}$ is the covariance matrix of the noise that corrupts the picture.

The set of normal equations is represented by

$$(\underline{H}^T\,\underline{V}^{-1}\,\underline{H})\hat{\underline{x}} = \underline{H}^T\,\underline{V}^{-1}\,\underline{y} \tag{3.4-33}$$

By performing the factorization

$$\underline{V}^{-1} = \underline{C}^T\underline{C} \tag{3.4-34}$$

(3.4-33) can be expressed as

$$(\underline{H}^T\,\underline{C}^T\,\underline{C}\,\underline{H})\hat{\underline{x}} = \underline{H}^T\,\underline{C}^T\,\underline{C}\,\underline{y} \tag{3.4-35}$$

In order to check whether the system is solvable, the consistency condition of equation (3.4-21) would have to be checked. Instead of doing this, a simpler way would be to observe that $R((\underline{C}\,\underline{H})^T) = R((\underline{C}\,\underline{H})^T\,\underline{C}\,\underline{H})$. Since $\underline{H}^T\,\underline{C}^T\,\underline{C}\,\underline{y}$ is in the range of $\underline{H}^T\,\underline{C}^T = (\underline{C}\,\underline{H})^T$, it must be in the range of $\underline{H}^T\,\underline{C}^T\,\underline{C}\,\underline{H} = (\underline{C}\,\underline{H})^T\underline{C}\,\underline{H}$. As a result, it must be the image of some $\hat{\underline{x}}$ under the transformation $(\underline{C}\,\underline{H})^T\,\underline{C}\,\underline{H}$. In other words, the set of normal equations is always consistent.

Using (3.4-25) the general solution of this system of linear equations is given by

$$\underline{\hat{x}} = (\underline{H}^T\underline{C}^T\underline{CH})^+\underline{H}^T\underline{C}^T\underline{Cy} + [\underline{I} - (\underline{H}^T\underline{C}^T\underline{CH})^+(\underline{H}^T\underline{C}^T\underline{CH})]\underline{v} \quad (3.4\text{-}36)$$

for an arbitrary vector $\underline{v}$.

Taking into consideration (3.4-18), the previous expression can be reduced to

$$\underline{\hat{x}} = (\underline{CH})^+\underline{Cy} + [\underline{I} - (\underline{CH})^+\underline{CH}]\underline{v} \quad (3.4\text{-}37)$$

The solution of the least squares problem is, therefore, not unique and any vector in the set expressed by the previous equation minimizes the quadratic form. The vector $(\underline{CH})^+\underline{Cy}$ is now merely the smallest norm solution that gives this minimum value.

In Figure (3.4-2), taken from reference [3-2, page 65], the geometry of the least squares problem for the underdetermined model is shown. The surface of the quadratic form (3.4-32) is infinitely long in the directions of the eigenvectors of $\underline{H}^T\underline{V}^{-1}\underline{H} = \underline{H}^T\underline{C}^T\underline{CH}$ corresponding to the zero singular values of $(\underline{CH})$. The set of solutions given by (3.4-37) is the projection on the $\underline{x}$ space of the bottom of the infinitely long quadratic through. The K-ellipsoids are degenerate, being infinitely long in the directions of the mentioned eigenvectors. One of the principal axes of these ellipsoids is given by the solution set of the least squares problem. The number of dimensions where the ellipsoid is infinite is the dimension of $N(\underline{H})$,

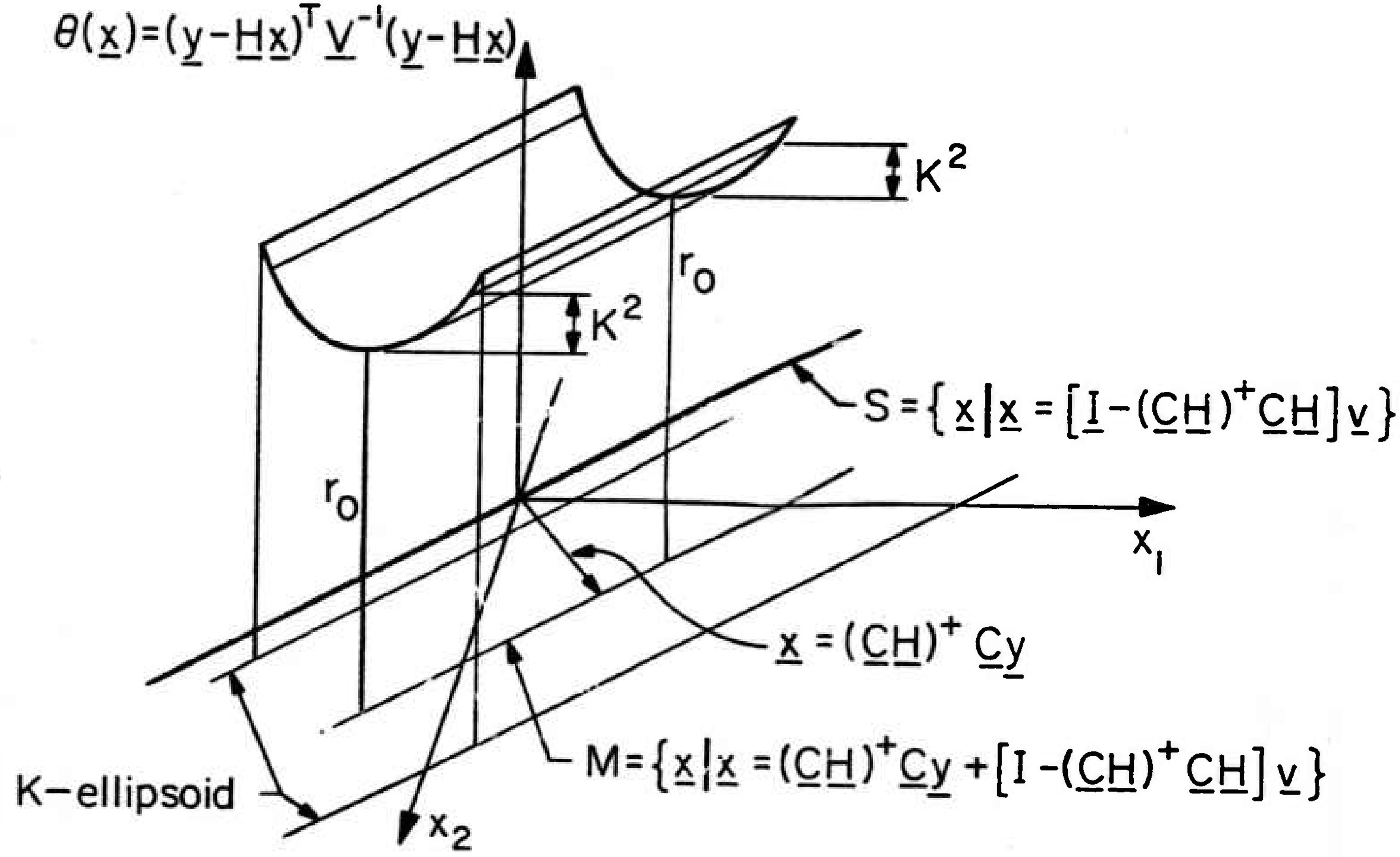


Figure (3.4-2) The Geometry of the Underdetermined Model assuming that $\underline{V}$ is positive definite.

Suppose now that the experimenter is interested in estimating parametric functions of the pixel values, represented by the inner product

$$\Phi = \underline{c}^T \underline{x} \qquad (3.4\text{-}38)$$

A linear combination of the observed pixel values is used to perform this estimation

$$\hat{\Phi} = \underline{u}^T \underline{y} \qquad (3.4\text{-}39)$$

The requirement of unbiasedness implies that

$$\underline{E}(\underline{u}^T \underline{y}) = \underline{c}^T \underline{x} \qquad (3.4\text{-}40)$$

which in turn leads to

$$\underline{u}^T \underline{H}\underline{x} = \underline{c}^T \underline{x} \qquad (3.4\text{-}41)$$

for any value of $\underline{x}$. Therefore, the following equality must be valid

$$\underline{u}^T \underline{H} = \underline{c}^T \qquad (3.4\text{-}42)$$

At this point the analogy between (3.4-31) and (3.4-42) is clear. The previous equation can be also expressed by stating that there must exist a vector $\underline{u}$ such that

$$\underline{H}^T \underline{u} = \underline{c} \tag{3.4-43}$$

In other words, the vector $\underline{c}$ has to be a linear combination of the columns of $\underline{H}^T$, a condition that can also be expressed by saying that $\underline{c}$ belongs to $R(\underline{H}^T)$ or still that

$$\underline{H}^+ \underline{H} \underline{c} = \underline{c} \tag{3.4-44}$$

since $\underline{H}^+\underline{H}$ is a projection matrix onto $R(\underline{H}^T)$. The term $R(\underline{H}^T)$ has dimension, say $K < N$, which will determine the number of linearly independent parametric functions of pixel values that are estimable. The functions $\underline{c}_i^T \underline{x}$, given by

$$\underline{c}_i^T = (0, \cdots, 1, \cdots, 0) \quad i = 1, 2, \cdots N \tag{3.4-45}$$

form a set of N linearly independent parametric functions that cannot all be estimated by an unbiased estimator. This confirms the result that the whole vector of pixel values is inestimable.

If Φ is an estimable parametric function, it can be proved that the estimator $\hat{\Phi}$ is given by

$$\hat{\Phi} = \underline{c}^T (\underline{C}\underline{H})^+ \underline{C}\underline{y} = \underline{c}^T \hat{\underline{x}} \tag{3.4-46}$$

where

$$\hat{\underline{x}} = (\underline{C}\underline{H})^{+}\underline{C}\underline{y} \tag{3.4-47}$$

is the minimum norm solution to the least squares problem. The variance of this estimator is given by

$$\sigma^2(\hat{\Phi}) = \underline{u}^T \underline{V} \underline{u} \tag{3.4-48}$$

where

$$\underline{u}^T = \underline{c}^T (\underline{C}\underline{H})^{+} \underline{c} \tag{3.4-49}$$

As far as confidence intervals and hypothesis testing for parametric functions of pixel values, the analysis can be carried out in a manner analogous to the overdetermined model. The confidence interval will be finite for estimable functions and infinite for inestimable functions. Similarly, hypothesis involving estimable functions will be testable, while those involving nonestimable functions will be nontestable. (See Appendix A.)

4. CONSTRAINED RESTORATION

In the previous chapter the restoration problem was solved by application of regression techniques. However, the experimental results will show that the problem can often be ill conditioned. This fact can be expressed by the large variances of the estimators of the individual pixel values or linear combinations of them. On the other hand, these techniques make use of the minimum possible amount of a priori information about the image to be restored. Pixel values are simply regarded as parameters to be determined in a N^2 dimensional space. Through the use of some additional a priori information it is possible to reduce considerably the uncertainty about the estimators. This can be done in several forms.

4.1 Analysis of Established Techniques

The classical Bayesian approach consists of assuming an a priori joint probability density on the pixel values to be estimated. The problem of estimating these values under a meaningful criterion like the mean square error, for any kind of a priori densities can be very complex, involving nonlinear filters. If only linear operations are allowed or if only second moments are known or if both noise and pixel vectors are gaussian distributed and any operation is allowed, the optimum procedure is the well known Wiener filter.

The problem can be formulated by the linear model

$$\underline{y} = \underline{H}\underline{x} + \underline{n} \tag{4.1-1}$$

where $\underline{H}$ is an $(M^2 \times N^2)$ matrix, $\underline{y}$ is the $(M^2 \times 1)$ vector of observed pixel values, $\underline{x}$ is the $(N^2 \times 1)$ vector of pixels to be estimated and $\underline{n}$ is the $(M^2 \times 1)$ noise vector. The covariance matrices of both $\underline{n}$ and $\underline{x}$ are known to be respectively, $\underline{V}$ and $\underline{C}_{xx}$, assumed to be positive definite. Zero means and uncorrelatedness of signal and noise are assumed for simplicity. It is desired to estimate the vector of random variables $\underline{x}$ by means of a linear operation

$$\hat{\underline{x}} = \underline{G}\underline{y} \tag{4.1-2}$$

in such a way that the covariance matrix $E\{(\hat{\underline{x}}-\underline{x})^T(\hat{\underline{x}}-\underline{x})\}$ is minimized in the sense described in section 3.1 of chapter 3. It is well known that $\underline{G}$ can be obtained by imposing the orthogonality principle, that leads to the result

$$\underline{G} = \underline{C}_{xx}\underline{H}^T(\underline{H}\,\underline{C}_{xx}\underline{H}^T + \underline{V})^{-1} = (\underline{C}_{xx}^{-1} + \underline{H}^T\underline{V}^{-1}\underline{H})^{-1}\underline{H}^T\underline{C}^{-1} \tag{4.1-3}$$

and the covariance matrix of the estimator is given by

$$\underline{C}_{xx}\underline{H}^T(\underline{H}\,\underline{C}_{xx}\underline{H}^T + \underline{V})^{-1}\underline{H}\,\underline{C}_{xx} \tag{4.1-4}$$

A connection between the Wiener filtering technique and the regression techniques can be drawn. Assume that both signal and noise are uncorrelated, that is

$$\underline{V} = \sigma^2 \underline{I} \tag{4.1-5}$$

and

$$\underline{C}_{xx} = \delta^2 \underline{I} \tag{4.1-6}$$

Now suppose that the value of δ^2 is fixed at unity and that σ^2 assumes progressively lower values approaching zero. Under these conditions, the first expression for $\underline{G}$ assumes the form

$$\underline{G} = \lim_{\sigma \to 0} \underline{H}^T (\underline{H}\underline{H}^T + \sigma^2 \underline{I})^{-1} = \underline{H}^+ \tag{4.1-7}$$

where equation (3.4-1) has been used. The same result could be obtained by fixing the noise level $\sigma^2 = 1$ and letting the variance of the signal, δ^2, go to infinity. In this case the second expression for $\underline{G}$ would be

$$\underline{G} = \lim_{\delta \to \infty} \left(\frac{1}{\delta^2} \underline{I} + \underline{H}^T \underline{H} \right)^{-1} \underline{H}^T = \underline{H}^+ \tag{4.1-8}$$

In either case, the pseudoinverse is obtained when the variance of the a priori distribution on the pixel values is much greater than the noise variance. This corresponds to assuming essentially no a priori knowledge about the parameters to be estimated, which is exactly what the regression techniques do.

It is interesting at this point to relate expressions (4.1-7) or (4.1-8) to the method of ridge regression [2-21, 2-22, and 2-23] or Twomey's method [2-19 and 2-20] for the case where the matrix $\underline{V}$ is the identity. The fact that there is some probabilistic prior information about the vector of pixel values plays the same role in the computational procedure as the damping factor γ in (2.4-7), for example.

Instead of a probabilistic a priori information one could also incorporate deterministic a priori constraints. These could be derived, for example, from the knowledge of some physical restrictions that the solution must satisfy. The methods described in chapter 2, Phillips', Twomey's and Tichonov's, can be regarded as equality constrained methods where the constraint expresses some degree of smoothness that the solution of the least squares problem must possess.

In the following discussion a framework to understand these equality constrained methods will be formulated. This will not only help to get a better understanding of these methods but also will make the connection between them and the linear equality and inequality methods to be described later.

Phillips', Twomey's or Tichonov's method can be described as searching for the minimum of a quadratic form $\underline{x}^T \underline{\underline{C}}\underline{x}$ that expresses the smoothness requirement, subject to an equality constraint on the residual vector

$$(\underline{y} - \underline{\underline{H}}\underline{x})^T (\underline{y} - \underline{\underline{H}}\underline{x}) = e \tag{4.1-9}$$

By imposing this restriction, the stationary point of the Lagrangean expression

$$F(\underline{x}, \lambda) = \underline{x}^T \underline{\underline{C}}\underline{x} + \lambda[(\underline{y} - \underline{\underline{H}}\underline{x})^T (\underline{y} - \underline{\underline{H}}\underline{x}) - e] \tag{4.1-10}$$

is searched for. Taking the derivatives with respect to $\underline{x}$ and λ and setting them to zero one obtains

$$\frac{\partial F}{\partial \underline{x}} = 2\underline{\underline{C}}\underline{x} + \lambda[-2\underline{\underline{H}}^T \underline{y} + 2\underline{\underline{H}}^T \underline{\underline{H}}\underline{x}] = 0 \tag{4.1-11}$$

$$\frac{\partial F}{\partial \lambda} = (\underline{y} - \underline{\underline{H}}\underline{x})^T (\underline{y} - \underline{\underline{H}}\underline{x}) - e = 0 \tag{4.1-12}$$

From (4.1-11) it follows that

$$(\underline{\underline{C}} + \lambda \underline{\underline{H}}^T \underline{\underline{H}})\underline{x} = \lambda \underline{\underline{H}}^T \underline{y} \tag{4.1-13}$$

or, dividing by λ, and solving for $\underline{x}$

$$\underline{x} = \left(\underline{H}^T\underline{H} + \frac{1}{\lambda}\underline{C}\right)^{-1}\underline{H}^T\underline{y} \qquad (4.1\text{-}14)$$

where λ is chosen such that (4.1-9) is satisfied.

Now consider another problem related to the previous one. It consists of solving the least squares problem, that is, minimizing the norm of the residual $(\underline{y} - \underline{H}\,\underline{x})$, but with the additional restriction of satisfying an equality constraint expressed by

$$\underline{x}^T\underline{C}\underline{x} = d \qquad (4.1\text{-}15)$$

In this case the Lagrangean expression is

$$G(\underline{x}, \lambda) = (\underline{y} - \underline{H}\underline{x})^T(\underline{y} - \underline{H}\underline{x}) + \gamma[\underline{x}^T\underline{C}\underline{x} - d] \qquad (4.1\text{-}16)$$

Setting the derivatives equal to zero, it follows that

$$\frac{\partial G}{\partial \underline{x}} = 2\underline{H}^T\underline{y} + 2\underline{H}^T\underline{H}\underline{x} + 2\gamma\underline{C}\underline{x} = 0 \qquad (4.1\text{-}17)$$

$$\frac{\partial G}{\partial \gamma} = \underline{x}^T\underline{C}\underline{x} - d = 0 \qquad (4.1\text{-}18)$$

Solving for $\underline{x}$ from the first equation one obtains

$$\underline{x} = (\underline{H}^T\underline{H} + \gamma\underline{C})^{-1}\underline{H}^T\underline{y} \qquad (4.1\text{-}19)$$

A comparison of (4.1-14) and (4.1-19) reveals that the two somehow inverse problems are solved by the same expression, the only difference being the Lagrange multipliers which are inverse of one another. This can be viewed from a geometrical point of view, expressed by Figure (4.1-1). In two-dimensional space, the contours of constant value of both quadratic forms, $(\underline{y}-\underline{H}\underline{x})^T(\underline{y}-\underline{H}\underline{x})$ and $\underline{x}^T\underline{C}\underline{x}$ are represented. The same solution is obtained if $\underline{x}^T\underline{C}\underline{x}$ is minimized subject to $(\underline{y}-\underline{H}\underline{x})^T(\underline{y}-\underline{H}\underline{x}) = e$ or if $(\underline{y}-\underline{H}\underline{x})^T(\underline{y}-\underline{H}\underline{x})$ is minimized subject to $\underline{x}^T\underline{C}\underline{x} = d$.

Phillips', Twomey's and Tichonov's methods can therefore be regarded as iterative methods to solve the equality constrained quadratic minimization problem, with the equality being also expressed by a quadratic expression. For $\gamma = 0$, the solution is equivalent to unconstrained estimation; it will exhibit no bias and the covariance matrix of the estimator is $(\underline{H}^T\underline{H})^{-1}$. For a value of $\gamma \neq 0$ this will correspond to imposing a constraint expressed by some quadratic form. The variance is reduced because the solution is now restricted to the contour, but bias is introduced. When γ tends to infinity, the estimator will be given by the origin of the $\underline{x}$-space. This corresponds to imposing the quadratic constraint $\underline{x}^T\underline{0}\underline{x}$, the variance will be reduced to zero and the bias will be finite and given by the

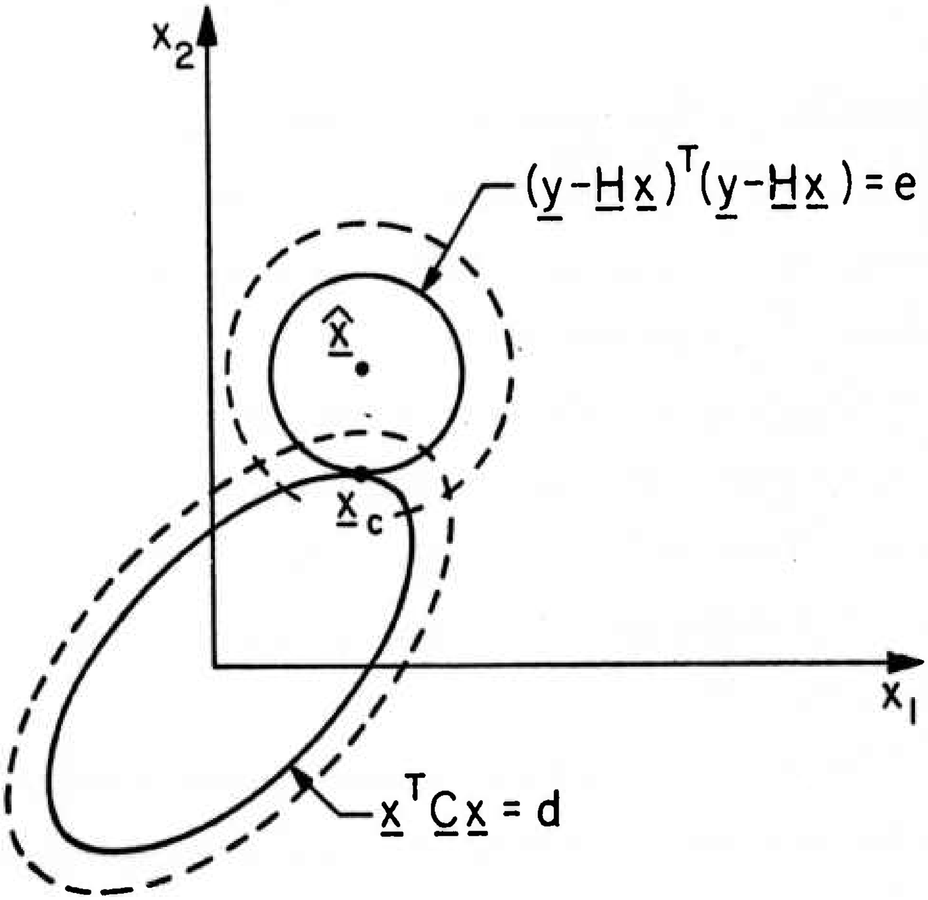


Figure (4.1-1) Geometry of the Smoothing and Regularization Methods

difference between the true solution $\underline{x}$ and the zero vector.

A problem that occurs with the smoothing and regularizing techniques is that, even though the variance of the solution can be calculated, the bias is unknown. It is true, however, that, in practice, it has been observed that there is considerable reduction in variance for a small amount of bias. A possible measure of the quality of the estimator would be the mean square error, computed by the square of the norm of the bias plus the sum of the variances of the individual components of the estimator. The fact that the constraint is quadratic makes it difficult to develop any testing procedure to verify whether the mean square error is reduced or not by the imposition of the constraint.

4.2 Linear Equality Constraints

Another possible type of equality constraint that can be imposed over the solution of the restoration problem is the linear equality constraint. This could be derived from some a priori knowledge that the analyst has about relations involving linear combination of pixel values. Examples could be the specification of individual pixel values, of ratios of the values of some pixels, or the sum of part or all of the pixels, representing the integral in the discrete form of the image as measured by a photocell. Another alternative would be presented when more than one image of the same object is available. In this case, if the blur functions are supposed to be known, the specification of blurred pixel values on additional pictures would represent linear constraints to be met. The value of these constraints will depend on the amount of uncertainty represented by noise in the additional images.

Suppose that the usual overdetermined linear model for restoration is adopted. The covariance matrix of the noise is $\underline{V}$, assumed to be positive definite. The set of linear constraints

$$\underline{A}\tilde{\underline{x}} = \underline{t} \qquad (4.2\text{-}1)$$

is imposed, where $\underline{A}$ is a $(P \times N^2)$ matrix of rank $P < N^2$, $\tilde{\underline{x}}$ is the constrained estimator, and $\underline{t}$ is a $(P \times 1)$ known vector.

The minimization of $(\underline{y} - \underline{Hx})^T \underline{V}^{-1} (\underline{y} - \underline{Hx})$ can be carried out using standard Lagrangean techniques. The optimal estimator will be given by [3-1, page 100]

$$\tilde{\underline{x}} = \hat{\underline{x}} + (\underline{H}^T \underline{V}^{-1} \underline{H})^{-1} \underline{A}^T [\underline{A}(\underline{H}^T \underline{V}^{-1} \underline{H})^{-1} \underline{A}^T]^{-1} (\underline{t} - \underline{A}\hat{\underline{x}}) \qquad (4.2\text{-}2)$$

where $\hat{\underline{x}}$ is the unconstrained solution, expressed by

$$\hat{\underline{x}} = (\underline{H}^T \underline{V}^{-1} \underline{H})^{-1} \underline{H}^T \underline{V}^{-1} \underline{y} \qquad (4.2\text{-}3)$$

and the covariance matrix of the optimal solution is

$$\underline{V}_{\tilde{\underline{x}}} = (\underline{H}^T\underline{V}^{-1}\underline{H})^{-1} - (\underline{H}^T\underline{V}^{-1}\underline{H})^{-1}\underline{A}^T[\underline{A}(\underline{H}^T\underline{V}^{-1}\underline{H})\underline{A}^T]^{-1}\underline{A}(\underline{H}^T\underline{V}^{-1}\underline{H})^{-1} \qquad (4.2\text{-}4)$$

$\underline{V}_{\hat{\underline{x}}} = (\underline{H}^T\underline{V}^{-1}\underline{H})^{-1}$ is the (positive definite) covariance matrix of the unrestricted estimator. The second matrix on the expression of $\underline{V}_{\tilde{\underline{x}}}$ can be shown to be nonnegative definite with rank P. Therefore, $\underline{V}_{\tilde{\underline{x}}}$ is equal to $\underline{V}_{\hat{\underline{x}}}$ minus a nonnegative definite matrix. As a consequence, each diagonal element of $\underline{V}_{\hat{\underline{x}}}$ is less than or equal to the corresponding element of $\underline{V}_{\hat{\underline{x}}}$. Thus, there is a reduction in the variance of each component of the constrained estimator vector as compared to the unconstrained one. However, this should not imply that the former is necessarily better than the latter. In fact, the constrained estimator may present bias, as opposed to the unbiasedness of the unconstrained estimator. The bias of the constrained solution is given by

$$\underline{x} - (\underline{H}^T\underline{V}^{-1}\underline{H})^{-1}\underline{A}^T[\underline{A}(\underline{H}^T\underline{V}^{-1}\underline{H})^{-1}\underline{A}^T][\underline{t} - \underline{A}\underline{x}] \qquad (4.2\text{-}5)$$

This bias will be zero if and only if the specifications are satisfied by the true vector $\underline{x}$, i.e., if $\underline{A}\underline{x} = t$. In this case, the set of constraints could be regarded as additional noise free observations. The following result obtained by Theil [4-1, pages 536-538] follows naturally. It states that, if the restrictions given by (4.2-1) are satisfied by the true parameter vector $\underline{x}$, then the equality constrained estimator $\underline{\tilde{x}}$ is B.L.U.E. of $\underline{x}$ in the sense of giving the minimum variance among the class of all unbiased estimators that are linear in $\underline{y}$ and $\underline{t}$.

The reduction of variance comes from the fact that the solution $\underline{\tilde{x}}$ should lie in a smaller dimensional space. For example, as an extreme case, if the linear system (4.2-1) has a unique solution, the variance of the solution will be zero. However, there will be bias if the true picture is not this solution vector.

Like in the quadratic equality constrained methods, the amount of bias is unknown, because the true vector of pixel values is not accessible. A measure of the quality of the constrained estimator should take into consideration both bias and variance. A possible measure could be given by the mean square error matrix, defined by $E(\underline{\tilde{x}} - \underline{x})(\underline{\tilde{x}} - \underline{x})^T$. The circumstance that the constraints are linear opens up the possibility of developing a statistical test procedure to verify whether or not there is a reduction in the mean square error by the imposition of the constraints. This is done by testing whether or not the hypothesis that the matrix $E(\underline{\hat{x}} - \underline{x})(\underline{\hat{x}} - \underline{x})^T - E(\underline{\tilde{x}} - \underline{x})(\underline{\tilde{x}} - \underline{x})^T$ is positive semidefinite is true. This procedure is due to Toro Vizcarrondo and Wallace [4-2]. It makes use of the F-statistic described in chapter 3. The improvement in the mean square error by the introduction of the equality constraint can be expressed by the fact that the noncentrality parameter λ of the test statistic is less than or equal to one-half. Observe that the test to verify whether or not the linear restrictions are true checks whether or not this noncentrality parameter is zero. In the MSE test one is not concerned whether or not the linear restrictions on the pixel values are true, but whether or not the imposition of these constraints represents an improvement in MSE. One problem that occurs in the application of the Toro Vizcarrondo-Wallace test is that the decision regions have been tabulated so far for only one linear constraint. In this case the experimenter can always perform the F-test for linear hypothesis, which has been tabulated for all degrees of freedom and uses the same statistics as the Toro Vizcarrondo-Wallace test. This will not tell whether there is an improvement in mean square error, but whether or not the imposed linear restrictions are satisfied by the true parameter vector.

The choice of the linear constraint to be imposed should be judged by two factors. The first one is the knowledge, coming from a priori considerations, that the relationship is true or at least approximately true so that an excessive bias will not be introduced in the answers. This is important when the linear relationships are subject to random error, as is the case when they come from the observation of pixel values in another blurred and noisy image of the same object. The second factor that the experimenter should have in mind is that a linear relationship tends to be more effective in reducing variance if it is in the direction of smaller axes of the K-ellipsoid rather than the larger ones. This can also be expressed by the fact that the vector $\underline{a}_p$ in the linear restriction $\underline{a}_p^T \underline{x} = t_p$, $p = 1, 2, \ldots, P$ should be in the direction of the eigenvectors of $\underline{H}^T \underline{C}^{-1} \underline{H}$ corresponding to the smallest eigenvalues. These requirements may not be very easy to conciliate in practice. Nevertheless, the MSE test provides a tool to verify whether or not linear equality restrictions should be used in the restoration.

Nonexact linear constraints involving pixel values can also be incorporated in another way, if the uncertainty can be modeled by a random process, with a known covariance matrix. An example would be the use of an additional blurred and noisy image of the same object. Suppose that the uncertain linear constraints are expressed by

$$\underline{t} = \underline{A}\underline{x} + \underline{v} \tag{4.2-6}$$

where the covariance matrix of $\underline{v}$ is $\underline{T}$ and $\underline{v}$ is assumed to be independent of the $\underline{n}$, for simplification. The combination of the sample information expressed by (4.1-1) and a priori information given by (4.2-5) can be accomplished through the model

$$\begin{bmatrix} \underline{y} \\ \underline{t} \end{bmatrix} = \begin{bmatrix} \underline{H} \\ \underline{A} \end{bmatrix} \underline{x} + \begin{bmatrix} \underline{n} \\ \underline{v} \end{bmatrix}$$

Under this model, the estimator $\hat{\underline{x}}$ is given by (using (5.2-3))

$$\hat{\underline{x}} = \left[(\underline{H}^T \underline{A}^T) \begin{pmatrix} \underline{V} & \underline{0} \\ \underline{0} & \underline{T} \end{pmatrix}^{-1} \begin{pmatrix} \underline{H} \\ \underline{A} \end{pmatrix} \right]^{-1} (\underline{H}^T \underline{A}^T) \begin{pmatrix} \underline{V} & \underline{0} \\ \underline{0} & \underline{T} \end{pmatrix}^{-1} \begin{pmatrix} \underline{y} \\ \underline{t} \end{pmatrix}$$

$$= \left[\underline{H}^T \underline{V}^{-1} \underline{H} + \underline{A}^T \underline{T}^{-1} \underline{A} \right]^{-1} \left[\underline{H}^T \underline{V}^{-1} \underline{y} + \underline{A}^T \underline{T}^{-1} \underline{t} \right] \quad (4.2\text{-}7)$$

The covariance matrix of the new estimator is easily obtained as

$$\underline{C}_{\hat{\underline{x}}} = (\underline{H}^T \underline{V}^{-1} \underline{H} + \underline{A}^T \underline{T}^{-1} \underline{A})^{-1} \quad (4.2\text{-}8)$$

This technique of incorporating random linear constraints in regression is known in econometrics as mixed estimation [4-3 and 4-4]. The connection between this procedure and the Bayesian approach is strong. In fact, suppose that $\underline{t} = \underline{0}$ and $\underline{A} = \underline{I}$. This is equivalent to stating that the vector of pixel values has an priori distribution with zero mean and covariance matrix $\underline{V}$. Under these conditions, expression (4.2-7) reduces to

$$\hat{\underline{x}} = (\underline{H}^T\underline{V}^{-1}\underline{H} + \underline{T}^{-1})^{-1}\underline{H}^T\underline{V}^{-1}\underline{y} \tag{4.2-9}$$

which is the familiar result of the Wiener filter.

Observe at this point that the linear constrained or mixed estimator procedures can make use of Fourier techniques if the matrices $\underline{V}$ and $\underline{T}$ are of the form $\sigma^2\underline{I}$ and $\underline{H}$ and $\underline{A}$ represent convolution operations. This would be the case when the constraints are represented by an additional picture and the blur in the observed picture as well as in the additional picture is space invariant and in both pictures the noise is white.

The extension of the mixed estimation technique to multi-exposure of the same object for more than two pictures is straightforward. Given K independent observations

$$\underline{y}_k = \underline{H}_k\underline{x} + \underline{n}_k \tag{4.2-10}$$

with the covariance matrix of $\underline{n}_k$ being $\underline{V}_k$, the B.L.U.E. estimator $\hat{\underline{x}}$ is obtained as

$$\hat{\underline{x}} = \left(\sum_{k=1}^{K} \underline{H}_k^T \underline{V}_k^{-1} \underline{H}_k\right)^{-1} \left(\sum_{k=1}^{K} \underline{H}_k^T \underline{V}_k^{-1} \underline{y}_k\right) \tag{4.2-11}$$

and the covariance matrix of the estimator is

$$\underline{C} = \left(\sum_{k=1}^{K} \underline{H}_k^T \, \underline{V}_k^{-1} \, \underline{H}_k \right)^{-1} \tag{4.2-12}$$

So far the discussion on linear constraints has been restricted to the overdetermined model for image restoration. When the blur matrix $\underline{H}$ has any rank, the analysis of the equality constrained least squares problem can be carried out by the following procedure. Assuming that the vector $\underline{t}$ in (4.2-1) is in the range of $\underline{A}$, the general solution of this equation is expressed by

$$\hat{\underline{x}}_o = \underline{A}^+ \underline{t} + (\underline{I} - \underline{A}^+ \underline{A}) \underline{u} \tag{4.2-13}$$

for some vector $\underline{u}$. Observe that the system (4.2-1) should be under-determined if the equality constraints do not involve any randomness, otherwise these restrictions would determine the solution by themselves, irrespective of what the observed blurred values are. Under white noise, the solution minimizes the norm of the vector $(\underline{y} - \underline{H}\underline{x})$ over the set expressed by (4.2-13). Therefore, the problem has been transferred to

$$\min_{\underline{u}} \| \overline{\underline{y}} - \underline{H}\underline{u} \| \tag{4.2-14}$$

where

$$\underline{\bar{y}} = \underline{y} - \underline{H}\underline{A}^+\underline{t} \tag{4.2-15}$$

and

$$\underline{\bar{H}} = \underline{H}(\underline{I} - \underline{A}^+\underline{A}) \tag{4.2-16}$$

The general solution of (4.2-14) is given by

$$\underline{u}_o = \underline{\bar{H}}^+\underline{\bar{y}} + (\underline{I} - \underline{\bar{H}}\underline{\bar{H}}^+)\underline{z} \tag{4.2-17}$$

for some vector $\underline{z}$. Substituting this value of $\underline{u}_o$ into expression (4.2-13), the solution for $\underline{x}$ is obtained as

$$\hat{\underline{x}}_o = \underline{A}^+\underline{t} + (\underline{I} - \underline{A}^+\underline{A})\left[\underline{\bar{H}}^+\underline{\bar{y}} + (\underline{I} - \underline{\bar{H}}\underline{\bar{H}}^+)\underline{z}\right] \tag{4.2-18}$$

By observing that

$$\underline{\bar{H}}^+ = \underline{\bar{H}}^T(\underline{\bar{H}}\underline{\bar{H}}^T)^+ \tag{4.2-19}$$

and also that

$$(\underline{I} - \underline{A}^+\underline{A})^2 = (\underline{I} - \underline{A}^+\underline{A}) = (\underline{I} - \underline{A}^+\underline{A})^T \tag{4.2-20}$$

which comes from the fact that $(\underline{I} - \underline{A}^+\underline{A})$ is a projection matrix (onto $N(\underline{A})$), equality (4.2-18) can be expressed as

$$\hat{\underline{x}}_o = \underline{A}^+\underline{t} + \underline{\bar{H}}^+\underline{\bar{y}} + (\underline{I} - \underline{A}^+\underline{A})(\underline{I} - \underline{\bar{H}}^+\underline{\bar{H}})\underline{z} \tag{4.2-21}$$

for some vector $\underline{z}$. In analogy with the unconstrained case, it can be shown [3-3, page 32] that

$$\hat{\underline{x}} = \underline{A}^{+}\underline{t} + \overline{\underline{H}}^{+}\overline{\underline{y}} \qquad (4.2\text{-}22)$$

is the minimum norm solution.

Expression (4.2-18) (or (4.2-21)) gives the general form of any solution to the linear constrained restoration problem. This solution will be unique if and only if $(\underline{I} - \overline{\underline{H}}^{+}\overline{\underline{H}})$ is the zero matrix or, equivalently, if and only if the null space of $\overline{\underline{H}}$, $N(\overline{\underline{H}})$, is the zero vector. The constraints substitute the matrix $\underline{H}$ for the matrix $\overline{\underline{H}}$ for the determination of the uniqueness of the solution. By the definition (4.2-16) and by the observation that $(\underline{I} - \underline{A}^{+}\underline{A})$ projects any vector onto $N(\underline{A})$, this necessary and sufficient condition can also be expressed by the condition

$$N(\underline{H}) \cap N(\underline{A}) = \underline{0} \qquad (4.2\text{-}23)$$

where $\underline{0}$ is the zero vector. This can also be viewed from another perspective: one can always add a vector lying anywhere on $N(\underline{H})$ to the solution of the unconstrained restoration problem and a vector lying anywhere on $N(\underline{A})$ to the solution of the linear system of equality constraints. When the two systems are solved together, the solution will clearly be unique if their intersection contains a single vector, which is necessarily the zero vector. Therefore, the use of linear equality constraints can transform the nonunique solution of the underdetermined restoration problem into a unique solution if the constraints compensate for the lack of information of the sample. This transforms the nonestimable vector $\underline{x}$ into an estimable one.

The imposition of linear equality constraints may still not guarantee the uniqueness and estimability of the solution vector if condition (4.2-23) is not satisfied. Nevertheless, the constraints may transform previously inestimable parametric functions of pixel values into estimable functions. In fact, in the unconstrained case, the only estimable functions $\underline{c}^T\underline{x}$ were those such that the vector $\underline{c}$ was orthogonal to $N(\underline{H})$. With the equality constraints the set of estimable functions comprises all those such that $\underline{c}$ is orthogonal to the intersection of $N(\underline{H})$ and $N(\underline{A})$. The latter set clearly contains the former one.

4.3 Inequality Constraints

In the last section the a priori knowledge involving linear combinations of pixel values was analyzed and used in the restoration of blurred images corrupted by noise. However, quite often the experimenter has a priori information in the form of inequality constraints involving the pixel values. This is particularly true in image processing. In fact, the physics of image formation determine that pixel values should be nonnegative quantities. Furthermore, an upper bound on these values is often known, as is the case when the images are digitized and a finite number of bits is assigned to each pixel. The analyst may also want to combine equality and inequality constraints in the restoration model. It will be shown in the following that if these constraints are linear and if a squared error is used as a criterion, a tractable mathematical model is developed, leading to considerably improved restoration results.

Suppose that the linear model

$$\underline{y} = \underline{\underline{H}}\underline{x} + \underline{n} \tag{4.3-1}$$

is adopted for the blurring process and the corruption by noise. As before, $\underline{\underline{H}}$ is an $(M^2 \times N^2)$ matrix and $\underline{y}$, $\underline{x}$ and $\underline{n}$ are vectors with compatible dimensions. The rank of the $\underline{\underline{H}}$ matrix is R. If $R = N^2$ or $R < N^2$, the model will be overdetermined or underdetermined, respectively. The covariance matrix of the noise will be assumed to be $\underline{\underline{V}}$. The constraints will be expressed by

$$\underline{\underline{A}}\underline{x} = \underline{t} \tag{4.3-2}$$

and

$$\underline{0} < \underline{x} < \underline{u} \tag{4.3-3}$$

where $\underline{A}$ is an $(S \times N^2)$ matrix of rank $S < N^2$.

Under the least squares criterion the objective function to be minimized is given by

$$(\underline{y} - \underline{H}\underline{x})^T \underline{V}^{-1} (\underline{y} - \underline{H}\underline{x}) \tag{4.3-4}$$

The minimization of (4. 3-4) subject to the constraints (4. 3-2) and (4. 3-3) may also be obtained if the vector $\underline{x}$ is supposed to be random, with a uniform distribution in the region defined by the constraints, and gaussian noise corrupts the image. Under the maximum a posteriori (MAP) estimation criterion (or maximum likelihood, since $p(\underline{x})$ is a constant in the interval), one looks for the vector $\hat{\underline{x}}$ such that $p(\hat{\underline{x}}|\underline{y}) \geq p(\underline{x}|\underline{y})$ for any $\underline{x}$. Using the fact that the logarithm is a monotonic increasing function and also the gaussian assumption on the noise it is equivalent to maximize the function

$$\log p(\underline{x}|\underline{y}) = \log p(\underline{x}) - \frac{1}{2} (\underline{y} - \underline{H}\underline{x})\underline{V}^{-1}(\underline{y} - \underline{H}\underline{x}) - \log p(\underline{y})$$

Since $p(\underline{x})$ is constant within the constraints, the maximization of $\log p(\underline{x}|\underline{y})$ leads to the minimization of the quadratic form (4. 3-4) subject to the linear constraints (4. 3-2) and (4. 3-3).

In order to obtain the quadratic programming problem in standard form, some manipulation is necessary. Define a slack vector $\underline{s}$ with nonnegative entries such that

$$(\underline{I}\ \ \underline{I})\begin{pmatrix}\underline{x}\\ \underline{s}\end{pmatrix} = \underline{u} \tag{4.3-5}$$

$$\underline{x}^* = \begin{pmatrix}\underline{x}\\ \underline{s}\end{pmatrix}$$

Also, introduce a new matrix $\underline{B}$, defined by the expression

$$\begin{pmatrix}\underline{A} & \underline{0}\\ \underline{I} & \underline{I}\end{pmatrix}\begin{pmatrix}\underline{x}\\ \underline{s}\end{pmatrix} = \begin{pmatrix}\underline{t}\\ \underline{u}\end{pmatrix} \tag{4.3-7}$$

or

$$\underline{Bx}^* = \underline{v} \tag{4.3-8}$$

where

$$\underline{v} = \begin{pmatrix}\underline{t}\\ \underline{u}\end{pmatrix}$$

If a matrix $\underline{H}^*$ is defined as

$$\underline{H}^* = (\underline{H}\ \ \underline{0}) \tag{4.3-9}$$

The linear model for restoration can then be expressed as

$$\underline{y} = \underline{H}^* \underline{x}^* + \underline{n} \tag{4.3-10}$$

And the objective function of the least squares is

$$(\underline{y} - \underline{H}^* \underline{x}^*)^T \underline{V}^{-1} (\underline{y} - \underline{H}^* \underline{x}^*) \tag{4.3-11}$$

subject to the constraints

$$\underline{B}\underline{x}^* = \underline{v} \tag{4.3-12}$$

and

$$\underline{x}^* \geq \underline{0} \tag{4.3-13}$$

The previous equations express the standard form of a quadratic programming problem, namely, the minimization of a quadratic form subject to linear constraints on the variables.

The necessary and sufficient condition that the solution of such a problem satisfies is expressed by the Kuhn Tucker Theorem [4-5, page 233]. In the particular situation of a quadratic objective function, this theorem can be expressed by the following conditions

a) $\underline{x}^*$ is feasible, that is equations (4.3-12) and (4.3-13) hold,

b) there exist vectors $\underline{u} > \underline{0}$ and $\underline{w}$ such that

$$2\underline{H}^{*T}\underline{V}^{-1}\underline{H}^*\underline{x}^* - \underline{u} + \underline{B}^{*T}\underline{w} - 2\underline{H}^{*T}\underline{V}^{-1}\underline{y} = 0 \tag{4.3-14}$$

c) the vector $\underline{u}$ and the optimal solution are such that

$$\underline{u}^T \underline{x}^* = 0 \qquad (4.3\text{-}15)$$

The Kuhn Tucker conditions for quadratic programming require the existence of vectors $\underline{x}^*$, $\underline{u}$, and $\underline{w}$ satisfying linear equalities and inequalities plus the additional condition that $\underline{u}^T \underline{x}^* = 0$. This suggests the use of a method very similar to the one that verifies the existence of a feasible solution to initialize the simplex method in linear programming. The condition $\underline{u}^T \underline{x}^* = 0$ is nonlinear, however, and requires a modification of the usual procedure.

This is the basis of the algorithms for quadratic programming that rely on the simplex method for linear programming. Two of the main procedures are Wolfe's [4-6] and Dantzig's [4-7] algorithms. The former has two versions, a short form and a long form. The first one is capable of handling only positive definite quadratic forms (which occurs when the model is overdetermined), but the second one can also deal with positive definite forms (for the underdetermined case). There are several other algorithms available for quadratic programming problems [4-8]. In recent years a research effort has been under way towards developing numerically stable methods [4-9, 4-10, 4-11, and 4-12] to solve this important problem.

The method developed so far to solve the linear inequality constrained restoration problem was directed toward the estimation of the image itself. Quite often one would be interested in estimating parametric functions of pixel values. If this is the case, a reformulation of the problem is convenient. This will lead to the calculation of confidence intervals for the constrained problem.

In chapter 3, the computation of the confidence intervals was made by considering the K-ellipsoids and the support planes orthogonal to the vector $\underline{c}$ that was used to calculate the parametric function $\underline{c}\underline{x}$. In the overdetermined model, if the vector $\underline{c}$ is parallel to some axis of the ellipsoid corresponding to a large eigenvalue of $\underline{H}^T\underline{V}^{-1}\underline{H}$, the confidence interval would be small compared to the case when $\underline{c}$ is parallel to an axis corresponding to a small eigenvalue. In the underdetermined model some eigenvalues of $\underline{H}^T\underline{V}^{-1}\underline{H}$ are zero and the confidence interval for some parametric functions is infinite, which corresponds to the fact that these functions are inestimable.

The a priori knowledge involving inequality constraints may change this situation considerably. The restrictions $\underline{0} < \underline{x} < \underline{u}$ may bound the elongated (in the ill conditioned case) or degenerate (in the underdetermined case) ellipsoids, reducing the confidence interval in the former case and transforming inestimable functions into estimable functions in the latter one. It is also clear that the estimation of the vector $\underline{x}$ of pixel values may be improved, with the bounding of the ellipsoids by the hyperplanes. The ill conditioning and the underdeterminacy problems may be solved, at least from the statistical point of view, with the use of linear inequality constraints.

In order to compute the confidence intervals for the inequality constrained case, it is necessary to introduce the idea of a confidence ellipsoid for a multidimensional distribution. The assumption of gaussian noise will be made throughout the discussion.

For a given estimator $\underline{\hat{x}}$ of $\underline{x}$, the 100.α% confidence interval of $\underline{x}$ is the ellipsoid in the $\underline{x}$-space with center in $\underline{\hat{x}}$ and given by the expression

$$(\underline{x} - \underline{\hat{x}})^T \underline{H}^T \underline{V}^{-1} \underline{H}(\underline{x} - \underline{\hat{x}}) = \gamma^2 \tag{4.3-16}$$

where γ^2 is selected such that

$$\Pr\left\{(\underline{x} - \underline{\hat{x}})^T \underline{H}^T \underline{V}^{-1} \underline{H}(\underline{x} - \underline{\hat{x}}) \leq \gamma^2\right\} = \alpha \tag{4.3-17}$$

The value of γ^2 can be computed by observing that the quadratic form in (4.3-17) is distributed according to a χ^2-distribution with r degrees of freedom, where r is the rank of the blur matrix $\underline{H}$. This means that [3-2]

$$\alpha = \int_0^{\gamma^2} \left[\Gamma(\rho/2)2^{r/2}\right]^{-1} \rho^{r/2-1} \cdot \exp(-\rho/2)d\rho \tag{4.3-18}$$

where $\Gamma(\cdot)$ is the gamma function defined by

$$\Gamma(\rho) = \int_0^{\infty} y^{\rho-1} e^{-y} dy \tag{4.3-19}$$

Now, suppose that the constraint $\hat{\underline{x}} \geq 0$ is imposed on the estimator. This means that, with probability one, $\hat{\underline{x}}$ will be in the positive quadrant. Without this constraint, $\hat{\underline{x}}$ is guaranteed to be in the γ - ellipsoid α% of the time. Therefore, with the imposition of the constraint, the estimator $\hat{\underline{x}}$ will be in the intersection of the γ - ellipsoid and the positive quadrant $100.\alpha$% of the time.

Consider now the parametric function $\underline{c}^T\underline{x}$ and the planes obtained by setting this function equal to a given constant. There are two support planes of the region mentioned in the previous paragraph. Figure (4.3-1) taken from [3-2, page 203], illustrates the assertion for the case where the vector $\underline{x}$ has two components. The two support planes will be denoted by

$$S_- = \left\{ \underline{x} \,|\, \underline{c}^T \underline{x} = \phi_E^{\ell} \right\} \qquad (4.3\text{-}20)$$

$$S_+ = \left\{ \underline{x} \,|\, \underline{c}^T \underline{x} = \phi_E^{u} \right\} \qquad (4.3\text{-}21)$$

$\hat{\underline{x}}$ is guaranteed to lie between S_+ and S_- with a probability of 100α%. The interval $I_\gamma(\phi) = \left[\phi_E^{\ell}, \phi_E^{u} \right]$ is a 100α% confidence interval for ϕ.

The interval $I_\gamma(\phi)$ can also be expressed as

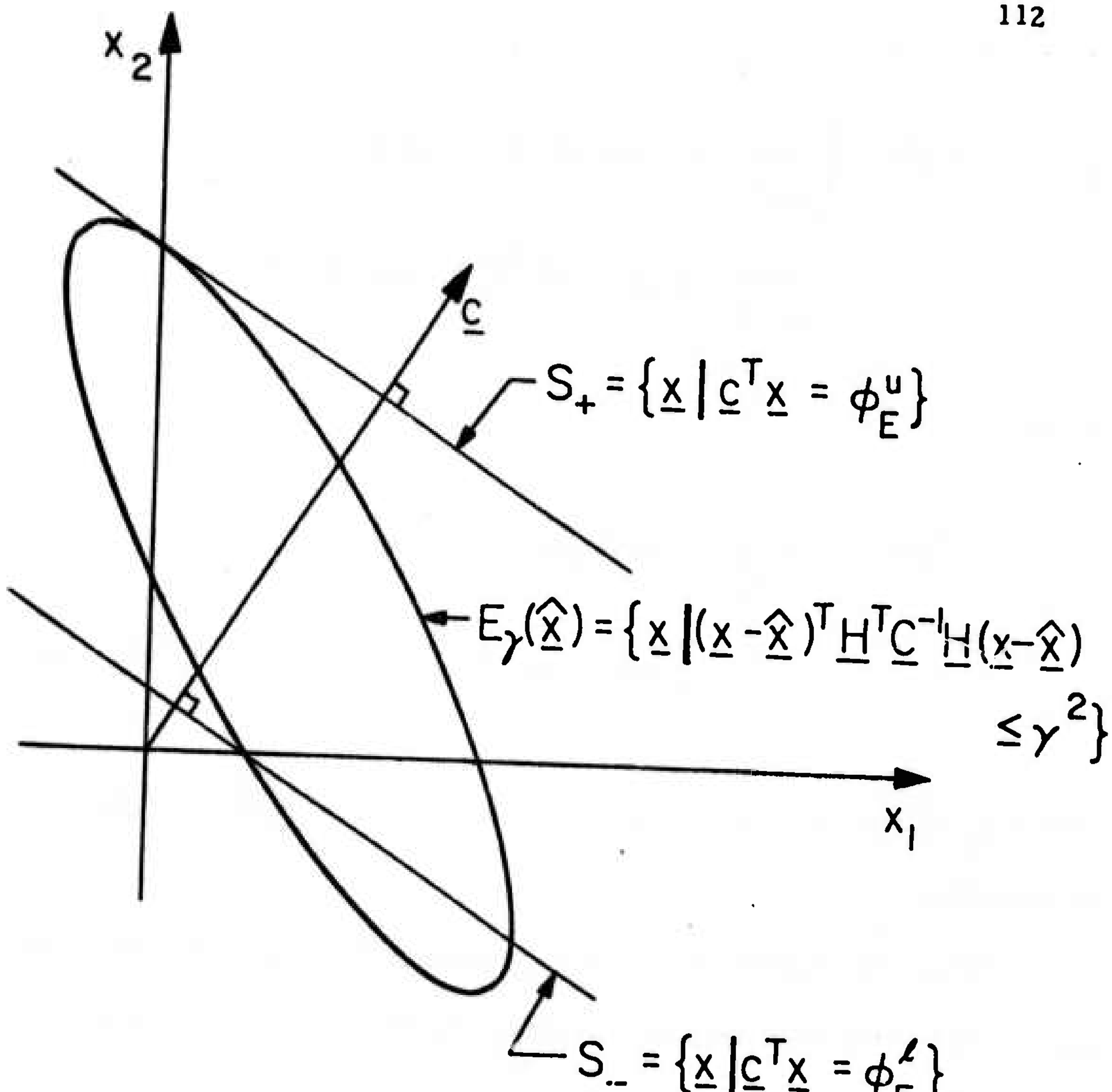


Figure (4.3-1) Support Planes for the Inequality Constrained Restoration

$$I_\gamma(\phi) = \left[\min_{\underline{x} \geq \underline{0}} \left\{ \underline{c}^T \underline{x} \mid (\underline{x}-\hat{\underline{x}})^T \underline{H}^T \underline{V}^{-1} \underline{H}(\underline{x}-\hat{\underline{x}}) \leq \gamma^2 \right\}, \right.$$

$$\left. \max_{\underline{x} \geq \underline{0}} \left\{ \underline{c}^T \underline{x} \mid (\underline{x}-\hat{\underline{x}})^T \underline{H}^T \underline{V}^{-1} \underline{H}(\underline{x}-\hat{\underline{x}}) \leq \gamma^2 \right\} \right] \quad (4.3\text{-}22)$$

or as

$$I_\gamma(\phi) = \left[\min_{\underline{x} > \underline{0}} \left\{ \underline{c}^T \underline{x} \mid (\underline{y}-\underline{H}\underline{x})^T \underline{V}^{-1} (\underline{y}-\underline{H}\underline{x}) \leq r_o + \gamma^2 \right\}, \right.$$

$$\left. \max_{\underline{x} > \underline{0}} \left\{ \underline{c}^T \underline{x} \mid (\underline{y}-\underline{H}\underline{x})^T \underline{V}^{-1} (\underline{y}-\underline{H}\underline{x}) \leq r_o + \gamma^2 \right\} \right] \quad (4.3\text{-}23)$$

where r_o is the minimum value assumed by the quadratic residual expression.

Since the minimum or the maximum of the linear function can occur only at the boundary of the ellipsoid, the expression for the confidence interval can also be written as

$$I_\gamma(\phi) = \left[\min_{\underline{x} \geq \underline{0}} \left\{ \underline{c}^T \underline{x} \mid (\underline{y}-\underline{H}\underline{x})^T \underline{V}^{-1} (\underline{y}-\underline{H}\underline{x}) = r_o + \gamma^2 \right\}, \right.$$

$$\left. \max_{\underline{x} > \underline{0}} \left\{ \underline{c}^T \underline{x} \mid (\underline{y}-\underline{H}\underline{x})^T \underline{V}^{-1} (\underline{y}-\underline{H}\underline{x}) = r_o + \gamma^2 \right\} \right] \quad (4.3\text{-}24)$$

As pointed out by Rust and Burrus [3-2, page 168], the two optimization problems that define the confidence interval can be formulated in a somehow dual way by making the following observation: the end points of the confidence interval are the points at which the two support planes make contact with the lowest level contours in the positive orthant. Figure (4.3-2) illustrates this point.

The inverted problems have the form

$$r_o + \gamma^2 = \min_{\underline{x} \geq 0} \left\{ (\underline{y} - \underline{Hx})^T \underline{V}^{-1} (\underline{y} - \underline{Hx}) \mid \underline{c}^T \underline{x} = \phi^{\ell} \right\} \tag{4.3-25}$$

$$r_o + \gamma^2 = \min_{\underline{x} \geq 0} \left\{ (\underline{y} - \underline{Hx})^T \underline{V}^{-1} (\underline{y} - \underline{Hx}) \mid \underline{c}^T \underline{x} = \phi^{u} \right\} \tag{4.3-26}$$

This formulation now leads to quadratic programming problems, with the role of the constraints and the objective function reversed.

The calculation of the confidence interval is facilitated by the construction of a curve expressed by

$$\Gamma(\phi) = \min_{\underline{x} \geq 0} \left\{ (\underline{y} - \underline{Hx})^T \underline{V}^{-1} (\underline{y} - \underline{Hx}) \mid \underline{c}^T \underline{x} = \phi \right\} \tag{4.3-27}$$

Figure (4.3-3) illustrates the curve given by (4.3-27), for the case of Figure (4.3-2). The bottom of the curve is in parabolic form, showing that for small values of γ^2 the inequality constraints are not being inforced. When γ^2 increases, however, these constraints start becoming effective and the curve rises steeper than the para-

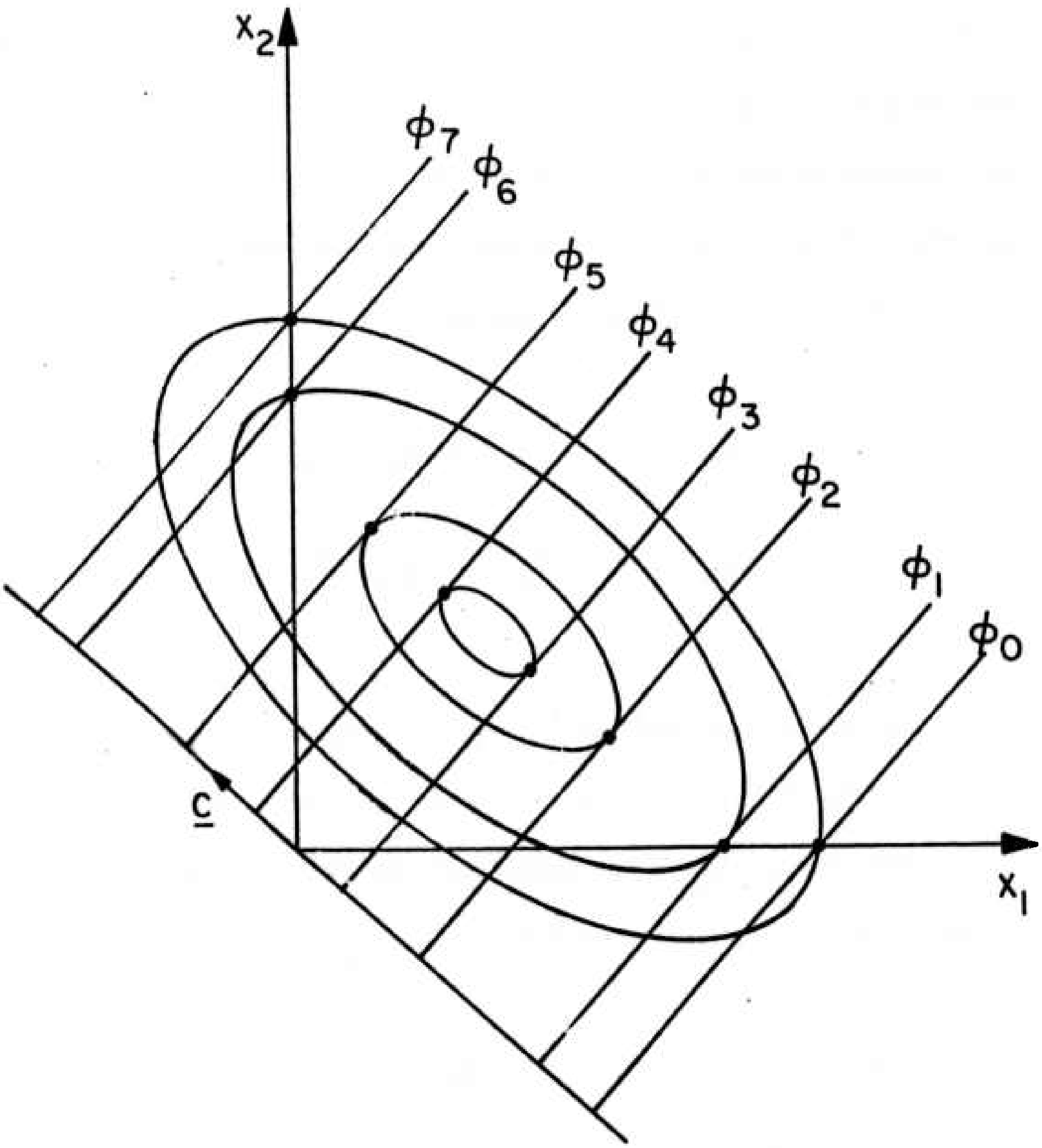


Figure (4.3-2) The Determination of Contact Points for Confidence Intervals

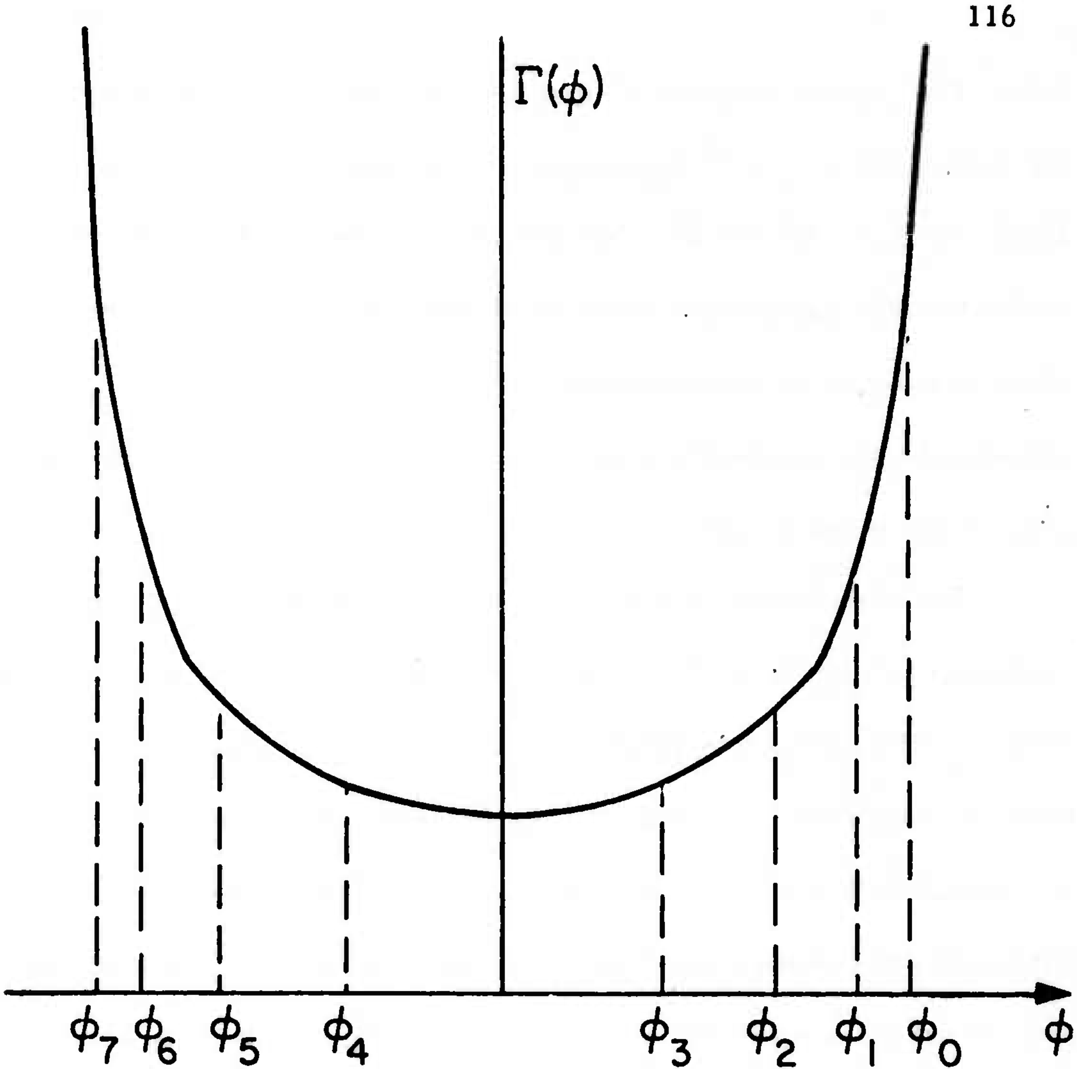


Figure (4.3-3) The Curve for Determination of the Confidence Intervals for Linear Inequality Constrained Restoration

bola. For a given value of γ^2 given by the tables of χ^2 distribution, the value $\Gamma(\emptyset) = r_o + \gamma^2$ determines two values on the abscissa of Figure (4.3-3), $\emptyset^{\ell}$ and $\emptyset^{u}$, that are the extremes the confidence interval for this parametric function of pixel values. The construction of the curve can be accomplished by the procedure of imposing simultaneously the inequality constraint $\underline{x} > \underline{0}$ and the equality constraint $\underline{c}^T\underline{x} = \emptyset$ for several values of $\emptyset$.

An observation should be made at this point. The unconstrained confidence intervals, developed in chapter 3, and making use of the K-ellipsoids, are optimum in the sense of giving the shortest possible confidence interval for a given confidence level. No claim of optimality is made for the constrained confidence intervals using quadratic programming and based on the γ-ellipsoids. In fact, they may be very pessimistic, particularly for large rank of the blur matrix $\underline{H}$, when the ratio γ/K becomes large. Further research is needed in order to obtain tighter intervals.

It should be remarked that if the γ-ellipsoid is centered outside the inequality constraints, the computation of the confidence intervals for small levels of confidence may not be possible. In this case one possible solution consists of replacing the constraint

$$(\underline{y} - \underline{H}\underline{x})^T\underline{V}^{-1}(\underline{y} - \underline{H}\underline{x}) = r_o + \gamma^2 \qquad (4.3\text{-}28)$$

by

$$(\underline{y} - \underline{\underline{H}}\underline{x})^T \underline{V}^{-1} (\underline{y} - \underline{\underline{H}}\underline{x}) = r_o' + \gamma^2 \tag{4.3-29}$$

where

$$r_o' = \min_{\underline{x} \geq \underline{0}} (\underline{y} - \underline{\underline{H}}\underline{x})^T \underline{V}^{-1} (\underline{y} - \underline{\underline{H}}\underline{x}) > r_o \tag{4.3-30}$$

In the preceding discussion of confidence intervals the assumption of gaussian noise has been made. As a consequence, the γ-ellipsoids contain a given percentage of this multidimensional distribution. An entirely similar analysis can be carried out when these ellipsoids contain the whole set of possible values for the estimator. This would be the case when the noise is distributed in some fashion, uniformly or not, within the bounds of an ellipsoid. In this case, the computed intervals give the minimum and maximum values that a parametric function can assume, with probability one. This bounded distribution does not have to be restricted to the ellipsoid shape. Suppose that the noise components satisfy the linear constraint

$$\sum_{i=1}^{M^2} \frac{|n_i|}{s_i} \leq u^2 \tag{4.3-31}$$

where s_i are positive quantities. This equation defines the shape of a diamond in a higher dimensional space. In analogy to expression (4.3-24), the lower and upper bounds of the inequality restricted confidence interval are given by

$$\phi^{\ell} = \min_{\underline{x} \geq 0} \left\{ \underline{c}^T \underline{x} \;\middle|\; \sum_{i=1}^{M^2} \frac{|(\underline{y} - \underline{H}\underline{x})_i|}{s_i} \leq \mu^2 \right\} \qquad (4.3\text{-}32)$$

and

$$\phi^{u} = \max_{\underline{x} \geq 0} \left\{ \underline{c}^T \underline{x} \;\middle|\; \sum_{i=1}^{M^2} \frac{|(\underline{y} - \underline{H}\underline{x})_i|}{s_i} \leq \mu^2 \right\} \qquad (4.3\text{-}33)$$

The two previous equations clearly define linear programming problems, namely, the minimization (or maximization) of a linear function subject to linear constraints.

Another type of bounded distribution that leads to linear programming is the rectangular distribution, defined by

$$\max_{1 \leq i \leq M^2} \left(\frac{|n_i|}{s_i} \right) \leq \mu^2 \qquad (4.3\text{-}34)$$

The confidence interval will be given by

$$\phi^{\ell} = \min_{\underline{x}>\underline{0}} \left\{ c^T x \;\middle|\; \max_{1 \leq j \leq M^2} \frac{|(\underline{y} - \underline{H}\underline{x})_j|}{s_j} \leq u^2 \right\} \tag{4.3-36}$$

With some manipulation it is possible to make it clear [3-5, page 98] that the problems defined above are linear programming problems.

In fact, equation (4.3-34) is equivalent to the statement that

$$\frac{|n_i|}{s_i} \leq u^2 \qquad i = 1, 2, \ldots, M^2 \tag{4.3-37}$$

or, by making u^2 equal to unity

$$-s_i \leq (\underline{y} - \underline{H}\underline{x})_i \leq s_i \qquad i = 1, 2, \ldots, M^2 \tag{4.3-38}$$

This can be expressed as

$$(\underline{H}\underline{x})_i \leq y_i + s_i \tag{4.3-39}$$

and

$$-(\underline{H}\underline{x})_i \leq -y_i + s_i \tag{4.3-40}$$

or, in matrix form

$$\underline{H}\underline{x} \leq \underline{y} + \underline{s} \tag{4.3-41}$$

and

$$-\underline{H}\underline{x} \leq -\underline{y} + \underline{s} \tag{4.3-42}$$

where $\underline{s}^t = (s_1, s_2, \cdots, s_{M^2})^T$.

Defining $\underline{A}$ to be a $2M^2 \times N^2$ matrix expressed by

$$\underline{A} = \begin{pmatrix} \underline{H} \\ -\underline{H} \end{pmatrix} \tag{4.3-43}$$

and $\underline{\ell}$ the $2M^2 \times 1$ vector

$$\underline{\ell} = \begin{pmatrix} \underline{y} + \underline{s} \\ -\underline{y} + \underline{s} \end{pmatrix} \tag{4.3-44}$$

The previous inequalities are expressed as

$$\underline{Ax} \leq \underline{\ell} \tag{4.3-45}$$

so that the confidence interval can be given by

$$\phi^{\ell} = \min_{\underline{x} > \underline{0}} \left\{ \underline{c}^T \underline{x} \mid \underline{Ax} \leq \underline{\ell} \right\} \tag{4.3-46}$$

and

$$\phi^{u} = \max_{\underline{x} \geq \underline{0}} \left\{ \underline{c}^T \underline{x} \mid \underline{Ax} \leq \underline{\ell} \right\} \tag{4.3-47}$$

The last two expressions are clearly linear programming formulations.

Linear programming is also used in two other possible formulations of the restoration problem, for the calculation of the optimal estimator $\hat{\underline{x}}$, instead of linear functionals, as in the previous derivation. As the first example, suppose that the noise components are independent and identically distributed according to an exponential distribution

$$p(n_i) = \frac{1}{2} \exp \left\{ - |n_i| \right\} \tag{4.3-48}$$

Assuming the vector $\underline{x}$ to be random and uniformly distributed within the constraints expressed by (4.3-2) and (4.3-3) and adopting either the criterion of MAP estimation or maximum likelihood, the following expression is to be minimized

$$\underline{Q}(\underline{x}) = \sum_{i=1}^{M^2} |y_i - (\underline{H}\underline{x})_i| \tag{4.3-49}$$

Observe that the same objective function would have been obtained through the criterion of estimation by least sum of absolute deviations under linear equality and inequality constraints.

This problem can be formulated in terms of linear programming by observing [4-13] that the objective function can be expressed as

$$\min \sum_{i=1}^{M^2} \left(\epsilon_{i_1} + \epsilon_{i_2}\right) \tag{4.3-50}$$

such that

$$(\underline{H}\underline{x})_i - y_i = \epsilon_{i_2} - \epsilon_{i_1} \qquad i = 1, 2, \ldots, M^2 \tag{4.3-51}$$

and

$$\epsilon_{i_1}, \epsilon_{i_2} \geq 0 \qquad i = 1, 2, \ldots, M^2 \tag{4.3-52}$$

As a second example of the use of linear programming in regression analysis, suppose that the objective is to minimize the maximum absolute derivation in the regression model (Chebyschev criterion). Then one seeks

$$\min_{\underline{x}} \left(\max_{1 \leq i \leq M^2} | y_i - (\underline{H}\underline{x})_i |\right) \tag{4.3-53}$$

The reduction to a linear programming structure [4-13] can be done by expressing the previous objective function as

$$\min \epsilon \tag{4.3-54}$$

such that

$$-\epsilon \le y_i - (\underline{\underline{H}}\underline{x})_i \le \epsilon \qquad i = 1, 2, \ldots, M^2 \tag{4.3-55}$$

The use of inequality constraints in image restoration also allows for the incorporation of a priori knowledge concerning the variation of the function that defines the original picture. This can be done by a change of variables [3-2, pages 112-115 and 190-193]. Consider the one-dimensional case first. If the solution is known to be monotonically increasing and positive, this can be imposed by expressing it in terms of an integral of a positive function with a positive initial condition. In discrete form, this corresponds to the change of variables

$$\underline{x} = \underline{\underline{R}}\underline{q}, \quad \underline{q} \ge \underline{0} \tag{4.3-56}$$

and $\underline{\underline{R}}$ in lower triangular form

$$\underline{\underline{R}} = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix} \tag{4.3-57}$$

If the solution is increasing, but not necessarily positive, the initial condition can assume any value, so it can be expressed by a difference of two positive quantities. In this case the matrix $\underline{\underline{R}}$ takes the form

$$\underline{R} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & -1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ 1 & -1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix} \qquad (4.3\text{-}58)$$

If the function is known to be positive and convex, it is enough to express it as a double integral of a positive function with a positive initial condition. In matrix form this is done by expressing

$$\underline{x} = \underline{R}\underline{q} \qquad \underline{q} \geq \underline{0} \qquad (4.3\text{-}60)$$

where

$$\underline{R} = \underline{S}\underline{T} \qquad (4.3\text{-}61)$$

where $\underline{S}$ and $\underline{T}$ are matrices with the form expressed by (4.3-57).

If the solution is positive and known to possess some degree of smoothness, this can be subjectively incorporated by expressing $\underline{x}$ as a positive linear combination of N^2 vectors that have positive components, are linearly independent and have some smooth variation in their components. Reference [3-2, page 114] suggests the use of vectors whose components form a triangular function and are shifted from vector to vector.

The application of these techniques to the two-dimensional problems that occur in image processing is easily accomplished by the use of vector notation. If the experimenter feels that the variation of some row or some column in the picture has one of the properties described in the preceding paragraphs, this can be done by substituting the variables corresponding to the row or column by the appropriate transformation. The matrix $\underline{R}$ will be diagonal with elements equal to one everywhere in the diagonal except in the elements that are transformed. Once this is done, both the mathematical programming problems of restoring the image itself or calculating approximate confidence intervals for parametric functions can be solved in terms of the new vector $\underline{q}$, with a new blur matrix $\underline{H}' = \underline{H} \cdot \underline{R}$. When the problem is solved in terms of $\underline{q}$, the transformation $\underline{x} = \underline{Rq}$ is used to obtain the desired solution.

The discussion on inequality constrained restoration will now be concentrated on the problem of the calculation of the sampling distribution of the estimators. The imposition of inequality constraints affects considerably that distribution. While in the unconstrained or linear equality constrained restoration, under the gaussian hypothesis, the estimators would still be normally distributed, in the case of inequality constraints, the distribution is of the mixed type, partly continuous, inside the permissible region and partly discrete, at the boundary of this area. The situation is exemplified by Figure (4.3-4), obtained from reference [3-6, page 354], for the one-dimensional case. In case a) the true parameter x_o satisfies the constraint $x \geq 0$. There is a positive bias since the distribution of the estimator is moved to the right by the constraint. However, if the mean square error is computed, taking into account both bias and variance, there is an improvement, because that distribution tends to be more concentrated around the true value. Case b) shows the opposite situation. The constraint $x \geq 0$ is invalid, that is, the true (and unknown) parameter x_o is negative. The bias is still positive but there will be an improvement in the mean square error only if x_o is not too far from 0. If this is the situation, the probability mass concentrated at 0 will contribute less for the mean square error than the part of the distribution for negative values of x. Zellner [4-14 and 4-15] calculated the moments of this mixed distribution in the one-dimensional, gaussian case. His conclusions can be summarized in the table below that gives the mean square error of the constrained estimator expressed as a fraction of the mean square error of the unconstrained estimator. The entry $\frac{x_o - \alpha}{\sigma}$ measures the distance between the true parameter x_o and the value α that determines that constraint $x \geq \alpha$ as a fraction of the standard deviation σ of the unconstrained estimator.

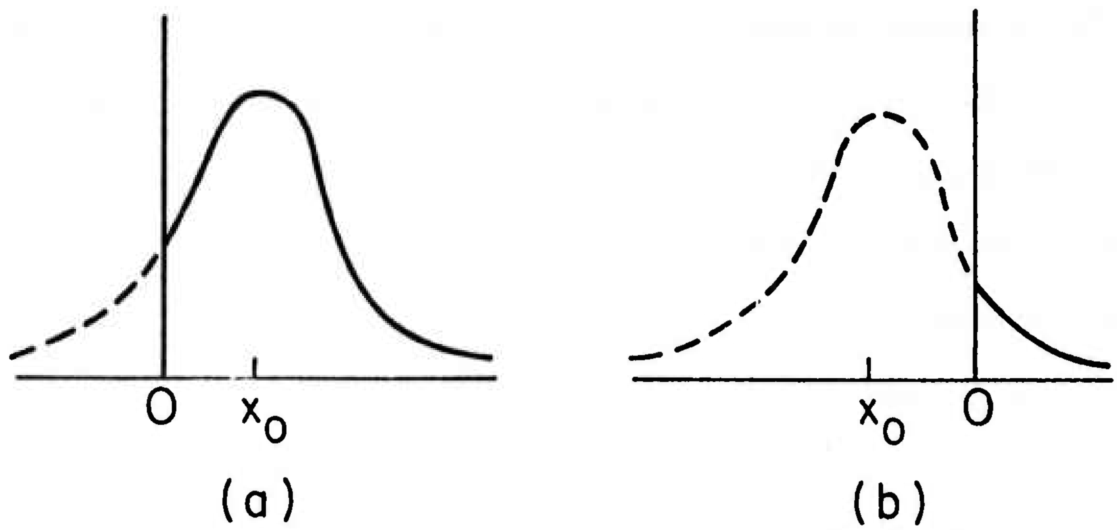


Figure (4.3-4) Distribution of the Inequality Constrained Estimators for the One-Dimensional Case

$\frac{x_o-\alpha}{\sigma}$ 3	2.5	2	1.5	1	0.5	0	-0.5	-1.0	-1.5	-2.0	-2.5	-3.0
1.0	0.99	0.96	0.89	0.76	0.59	0.50	0.66	1.24	2.36	4.04	6.26	9.0

The constrained estimator is seen to be superior to the unconstrained even if the true parameter slightly violates the constraints. In the referred papers, Zellner also studied the distribution of a flexible bound procedure in which the estimator is given by a linear convex combination of the unconstrained and constrained estimators, reflecting the lack of absolute confidence of the analyst in the imposed bounds.

When the dimensionality of the problem is greater than one, the difficulties in computing the sampling distribution or even the first two moments increase considerably. Hocking [4-16] has obtained closed form solutions for the case where the constraints are described by a single, smooth, convex surface. Some Monte Carlo experiments involving small dimensionality have been reported by Lee, Judge and Zellner [4-17] in the estimation of transition probabilities of a Markov probability model.

In the context of image processing, the large dimensionality of the problem makes any attempt to calculate the sampling distribution of the constrained estimator extremely difficult. However, the confidence of the experimenter in the imposition of the constraints is very high, at least for the lower bound $\underline{x} \geq \underline{0}$, due to the fact that they come from physical laws governing the process of image formation. As a result, by extending Zellner's conclusions to a higher dimensional space, there should be considerable reduction in mean square error through the use of these restrictions.

5. EXPERIMENTAL RESULTS WITH UNCONSTRAINED RESTORATION

In this chapter the experimental results obtained with computer simulation studies of digital image restoration are described. In order to expose the most salient features of the statistical and numerical problems found in the restoration under a regression model, a simplified artificial picture has been generated. The picture consists of an 8 x 8 pixel image, containing a bright square of value 245 on a constant background of value 10 over a 0 to 255 scale. Figure (5-1) displays this picture used as the object of blurring and addition of noise in the simulation experiments.

For the purpose of displaying pictorial information, two operations had to be performed in the pictures obtained in these experiments. First, the images were blown up to the size of 256 x 256. This explains the checkerboard pattern that results. Second, a redistribution of the pixel values is necessary. This redistribution consists of clipping the results to the 0-255 scale. In some instances the restored values will be far from this interval and in order to make a meaningful judgment, a display of the actual numerical results will be made. The objective of the experiments was to investigate, in the context of image restoration, the usefulness of the regression model

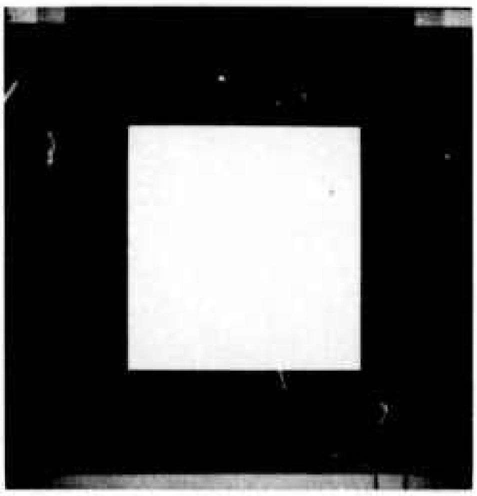


Figure (5-1) Original Picture Used in the Computer Simulation Studies developed in chapter 3 when no additional constraint is imposed in the restored values.

The overdetermined model is studied first, followed by the underdetermined one. The computation of confidence intervals of some parametric functions is performed and, finally, a few results concerning the testing of hypothesis is presented.

5.1 Restoration for the Overdetermined Model

Experiments with the overdetermined model simulated the following real problem. The blurred and noisy image of an object of finite extent (e.g., the moon on a dark background) is available. As pointed out before, the use of digital processing requires that this blurred image be sampled at a finite number of points. Furthermore, the original image has to be estimated based also on a finite number of points that are the nodes of the quadrature integration. The experiments represent the situation in which the analyst decides to place those nodes at equally spaced points on a rectangular grid over the finite object, taking into account the several factors discussed in section 2.3. The 8 x 8 original picture of Figure (5-1) represents the original object as if it were available at these quadrature nodes. The blurred image is assumed to be sampled at points separated by the same distance as the nodes of quadrature. There are more sampling points than nodes, covering a blurred picture that is larger in extent than the original. Some of the sampling points coincide with the quadrature nodes. The blurred picture is still of finite extent, which implies a point spread function of finite support. This represents an approximation to the real case, obtained by truncation of the kernel. In the experiment the support was taken as a multiple of the sampling distance and the kernel was assumed to be zero beyond two times this distance. The full extent of the blurred picture was assumed available. Under these conditions, the original 8 x 8 object was blurred into a 12 x 12 image. Using the notation of equation (2.3-4) the following conditions describe the experiments

$$\begin{aligned} I = J = N = 8 \\ K = L = M = 12 \end{aligned} \qquad (5.1\text{-}1)$$

Figure (5.1-1) describes the data arrays involved. In that figure a translation of the enumeration of the original picture was done. The values of ξ_k and η_ℓ were made to run from $\left(\frac{M-N}{2} + 1\right)$ to $\left(\frac{M+N}{2}\right)$. Under these conditions, the truncation of the impulse response is described by

$$h(\alpha_i, \xi_k ; \beta_j, \eta_\ell) = 0 \qquad (5.1\text{-}2)$$

or

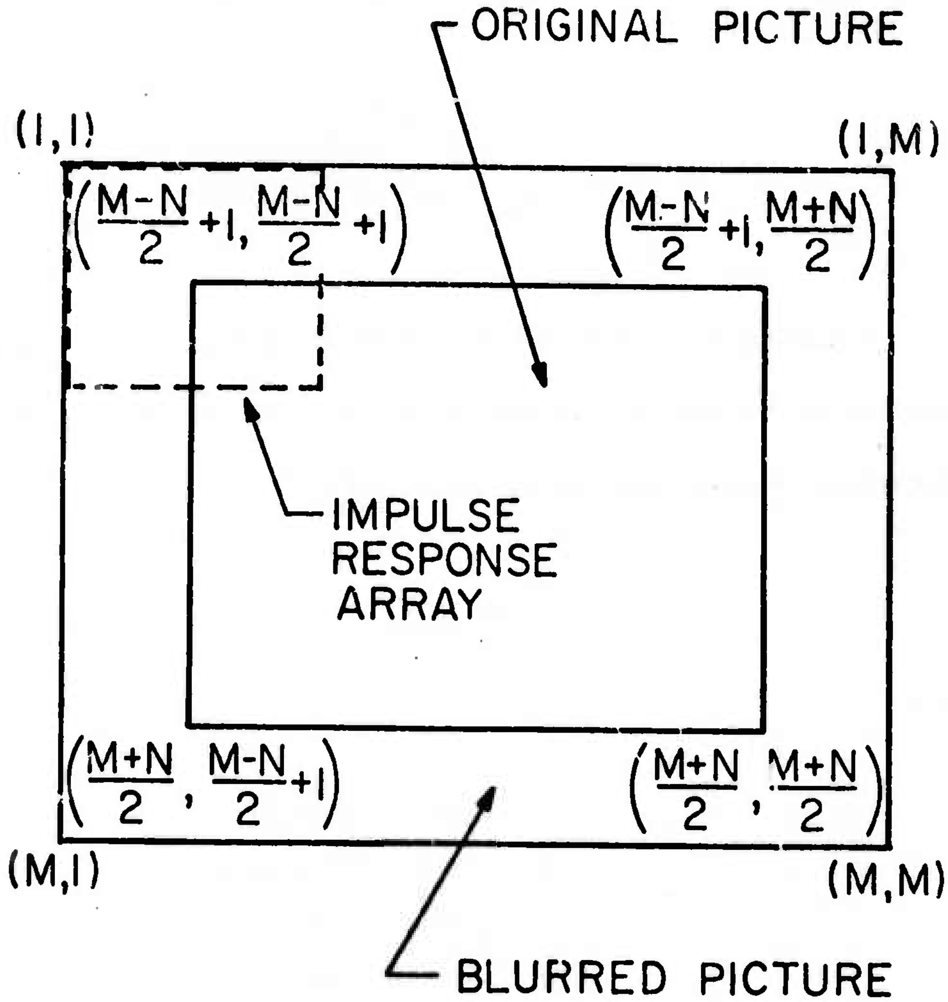


Figure (5.1-1) The Data Arrays in the Overdetermined Model

$$|\alpha_i - \beta_j| > \frac{M-N}{2} = \frac{L-1}{2}$$

or

$$|\xi_k - \eta_\ell| > \frac{M-N}{2} = \frac{L-1}{2}$$

By using the techniques described in section 2.3, a transformation of the two dimensional arrays into vector form is done. It results a model described by the equation

$$\underline{y} = \underline{B}\underline{x} + \underline{n} \qquad (5.1\text{-}3)$$

where

$$\begin{aligned}\underline{y} &= (M^2 \times 1) \text{ vector}\\ \underline{B} &= (M^2 \times N^2) \text{ matrix}\\ \underline{x} &= (N^2 \times 1) \text{ vector}\\ \underline{n} &= (M^2 \times 1) \text{ vector}\end{aligned}$$

The description of the simulation experiments leads to the following structure for the matrix $\underline{B}$. First partition $\underline{B}$ in submatrices $\underline{B}_{i,j}$ of size (M x N), as shown in Figure (5.1-2). Then each matrix $\underline{B}_{i,j}$ is composed by a similar structure, as described by Figure (5.1-3). Observe that the matrix $\underline{B}$ has considerable structure being very sparse. It is composed of a nonzero diagonal band of submatrices which, in turn, contain a nonzero diagonal band of elements.

$$\underline{B} = \begin{bmatrix} \underline{B}_{1,1} & \underline{0} & \underline{0} & \cdots & \underline{0} \\ \underline{B}_{2,1} & \underline{B}_{2,2} & \underline{0} & \cdots & \underline{0} \\ \vdots & & & & \vdots \\ \underline{B}_{L,1} & \underline{B}_{L,2} & & & \\ \underline{0} & \underline{B}_{L+1,2} & & & \underline{0} \\ \underline{0} & \underline{0} & & & \underline{B}_{N,N} \\ \vdots & \vdots & & & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{B}_{M,N} \end{bmatrix}$$

Figure (5.1-2) Partition of the Blur Matrix $\underline{B}$
in the Overdetermined Model

$$
\underline{B}_{m_2, n_2} = \begin{bmatrix}
b_{1,1} & 0 & 0 & \cdots & 0 \\
b_{2,1} & b_{2,2} & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & & \vdots \\
b_{L,1} & b_{L,2} & & \ddots & 0 \\
0 & b_{L+1,2} & & & \\
0 & 0 & \ddots & & b_{N,N} \\
\vdots & \vdots & & \ddots & \vdots \\
0 & 0 & \cdots & 0 & b_{M,N}
\end{bmatrix}
$$

Figure (5.1-3) Composition of the Submatrices $\underline{B}_{m_2, n_2}$ in the Overdetermined Model

In a real experiment the matrix $\underline{B}$ should involve both the quadrature weights and the kernel of the point spread function. The weights depend, however, on the type of quadrature expression that is used. In order to circumvent this problem a simplification was made: all weights were assumed to be equal to one. Therefore, the entries of the blur matrix depend only on the point spread function.

If the impulse response is space invariant, then

$$h(\alpha_i, \xi_k ; \beta_j, \eta_\ell) = h(\alpha_i - \beta_j ; \xi_k - \eta_\ell) \tag{5.1-4}$$

Consequently, the columns of the submatrices of $\underline{B}$ are shifted versions of the first column and the same pattern occurs for each submatrix. Suppose, furthermore, that the impulse response matrix $\underline{H}$ is in separable form, that is, $h(\alpha_i, \xi_k ; \beta_j, \eta_\ell)$ is expressed by the product

$$h(\alpha_i, \xi_k ; \beta_j, \eta_\ell) = h_1(\alpha_i ; \beta_j) \cdot h_2(\xi_k ; \eta_\ell) \tag{5.1-5}$$

Then it can be shown that the blur matrix $\underline{B}$ is given by

$$\underline{B} = \underline{B}_c \otimes \underline{B}_R \tag{5.1-6}$$

where $\underline{B}_c$ and $\underline{B}_R$ are given by matrices of the type described by Figure (5.1-4) and $\otimes$ denotes the Kronecker product.

Two expressions for the blur have been used. The first one

$$\underline{B}_C = \begin{bmatrix} b_V(1) & 0 & 0 & \cdots & 0 \\ b_V(2) & b_V(1) & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ b_V(L) & b_V(L-1) & & & \\ 0 & b_V(L) & & & 0 \\ & & \ddots & & b_V(1) \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 & b_V(L) \end{bmatrix}$$

Figure (5.1-4) Unidimensional Blur Matrix in the Separable, Space Invariant Case, Overdetermined Model simulates the effect of atmospheric turbulence over a long exposure. The blur function is given by

$$h(\alpha_i, \xi_k ; \beta_j, \eta_\ell) = \exp. \left\{ - \frac{(\alpha_i - \beta_j)^2}{b_V} + \frac{(\xi_k - \eta_\ell)^2}{b_H} \right\} \tag{5.1-7}$$

where the coefficients b_V and b_H control the amount of blur imposed on the vertical and horizontal directions, respectively. The exponent 5/6 of expression (2.1-5) was approximated by unity. The second blur function, also space invariant and in separable form, simulates the effect of a diffraction limited optical system as given by

$$h(\alpha_i, \xi_k ; \beta_j, \eta_\ell) = \left[\frac{\sin \left(\frac{\alpha_i - \beta_j}{b_V} \right)}{\left(\frac{\alpha_i - \beta_j}{b_V} \right)} \right]^2 \cdot \left[\frac{\sin \left(\frac{\xi_k - \eta_\ell}{b_H} \right)}{\left(\frac{\xi_k - \eta_\ell}{b_H} \right)} \right]^2 \tag{5.1-8}$$

Both blur functions are truncated to dimension L x L. Once the picture is blurred by multiplying the vector $\underline{x}$ by the blur matrix $\underline{B}$, gaussian noise from a random variable generator is added to the components of the vector $\underline{y}$. The noise is uncorrelated, with a covariance matrix $\underline{V}$ given by

$$\underline{V} = \sigma^2 \underline{I} \tag{5.1-9}$$

Recall that under the gaussian assumption the B.L.U.E. estimator is also the maximum likelihood estimator, besides being the least squares solution. Furthermore, since the overdetermined model is assumed, the solution is unique.

Under white noise, the estimator $\hat{\underline{x}}$ is obtained through the expression

$$\hat{\underline{x}} = (\underline{B}^T \underline{B})^{-1} \underline{B}^T \underline{y} = \underline{B}^+ \underline{y} \tag{5.1-10}$$

The condition that the spread function should be space invariant may be explored for computational purposes. In fact, under this hypothesis, $(\underline{B}^T \underline{B})$ is a block Toeplitz matrix and the inversion of such a matrix can be performed quite efficiently through Fast Fourier Transform techniques [5-1], even for large dimensions. Furthermore, $\underline{B}^T \underline{y}$ represents, in this case, a discrete convolution operation that can also be performed by the use of two-dimensional FFT. The procedure is equivalent to the one used by Hunt [2-40] by making the coefficient γ equal to zero in Twomey's method. If the blur matrix is separable, considerable simplification in the computation of the pseudoinverse can be achieved. Using (5.1-6) it follows that [5-2]

$$
\begin{aligned}
\underline{B}^{+} &= [(\underline{B}_C \otimes \underline{B}_R)^T \; (\underline{B}_C \otimes \underline{B}_R)]^{-1} \; (\underline{B}_C \otimes \underline{B}_R)^T \\
&= [(\underline{B}_C^T \otimes \underline{B}_R^T)\,(\underline{B}_C \otimes \underline{B}_R)]^{-1} \; (\underline{B}_C^T \otimes \underline{B}_R^T) \\
&= (\underline{B}_C^T\underline{B}_C \otimes \underline{B}_R^T\underline{B}_R)^{-1} \; (\underline{B}_C^T \otimes \underline{B}_R^T) \\
&= [(\underline{B}_C^T\underline{B}_C)^{-1} \otimes (\underline{B}_R^T\underline{B}_R)^{-1}] \; (\underline{B}_C^T \otimes \underline{B}_R^T) \\
&= (\underline{B}_C^T\underline{B}_C)^{-1}\underline{B}_C^T \otimes (\underline{B}_R^T\underline{B}_R)^{-1}\,\underline{B}_R^T \\
&= \underline{B}_C^{+} \otimes \underline{B}_R^{+}
\end{aligned}
\tag{5.1-11}
$$

In the previous derivation the following identities were used [5-3]

$$(\underline{A} \otimes \underline{B})^T = \underline{A}^T \otimes \underline{B}^T \tag{5.1-12}$$

$$(\underline{A} \otimes \underline{B})(\underline{F} \otimes \underline{G}) = (\underline{A}\underline{F}) \otimes (\underline{B}\underline{G}) \tag{5.1-13}$$

$$(\underline{A} \otimes \underline{B})^{-1} = \underline{A}^{-1} \otimes \underline{B}^{-1} \tag{5.1-14}$$

Equation (5.1-11) allows the computation of $\underline{B}^T$ through the Kronecker product of pseudoinverses of much smaller dimensions. Since the computational methods for the pseudoinverse [3-3, chapter V] are involved and particularly sensitive to numerical errors due to round-off, the importance of (5.1-11) becomes evident.

The method used for the computation of the pseudoinverses $\underline{B}_C^T$ and $\underline{B}_R^T$ was the gradient projection by Pyle [3-3, pages 69-74]. It is based on an application of the Gram-Schmidt orthonormalization process. In the course of the method some care must be exercised in order to decide, within the precision of the machine, whether a vector can be given by a linear combination of the previous ones or not.

The restored image has been computed through the expression

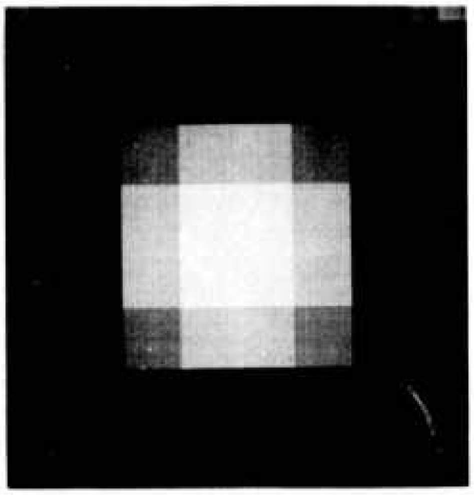
$$\hat{\underline{x}} = (\underline{B}_C^+ \otimes \underline{B}_R^+)\underline{y} \tag{5.1-15}$$

for several values of blur, under noisy or noise-free conditions, blurred by gaussian or sinc^2 functions. In the case of no noise, the least squares problem reduces to the solution of the system of linear equations

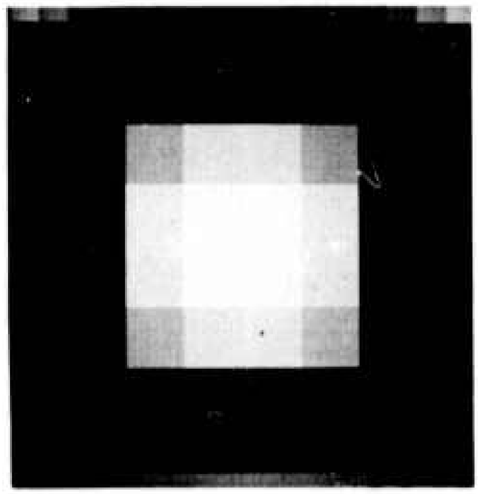
$$\underline{y} = \underline{B}\,\underline{x} \tag{5.1-16}$$

and, under the assumption of full column rank of the matrix $\underline{B}$ (over-determined model), the solution exists and it is unique, given by the same expression as the estimator for the white noise (equation (5.1-10)).

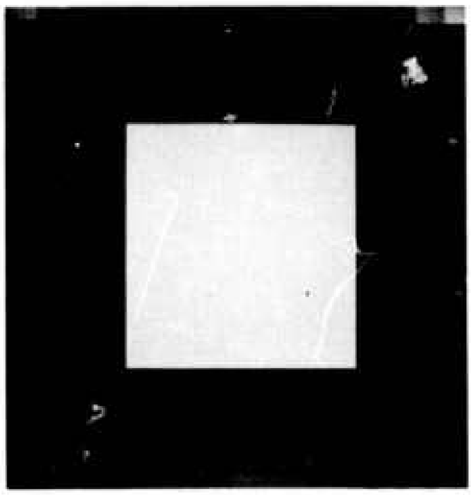
Figure (5.1-5) shows the blurred and restored images under



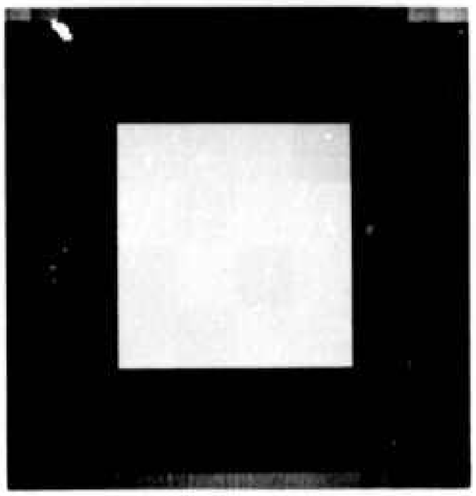
Blurred $b_V = b_H = .70$, Var = 0
Gaussian Blur, Overdetermined



Blurred $b_V = b_H = .70$, Var = 10
Gaussian Blur, Overdetermined



Restored



Restored

Figure (5.1-5) Examples of Restoration with the Overdetermined Model and Gaussian Blur-I gaussian blur and $b_V = b_H = 0.70$. The first column of pictures refers to the no noise case. It can be seen that the restored image coincides with the original picture. This is what should be expected with the overdetermined model: since the solution to the linear system (5.1-16) is unique, the original picture is the only possible solution. In the second column of pictures white gaussian noise has been added to the blurred image. Now the restored picture is not equal to the original, but the difference is relatively small, with the light square in the middle being clearly distinguishable.

A remark should be made at this point: even though all the blurred pictures with the overdetermined model are 12 x 12, only the center 8 x 8 parts are displayed, blown up to 256 x 256.

Figure (5.1-6) presents the same results as Figure (5.1-5) for different values of the blur coefficients, $b_V = b_H = 2.5$. The result obtained with no noise shows that the original picture is still obtained, although a closer inspection of the numerical values will evidence some round-off in the computation. The noisy restoration differs substantially from the previous result. Even though the noisy and blurred picture is visually barely different from the noise free case, the restored image differs considerably, with the bright square in the middle not even visible.

Figure (5.1-7) presents the results for the case $b_V = b_H = 5000$. For the noisy restoration it is observed that the result is

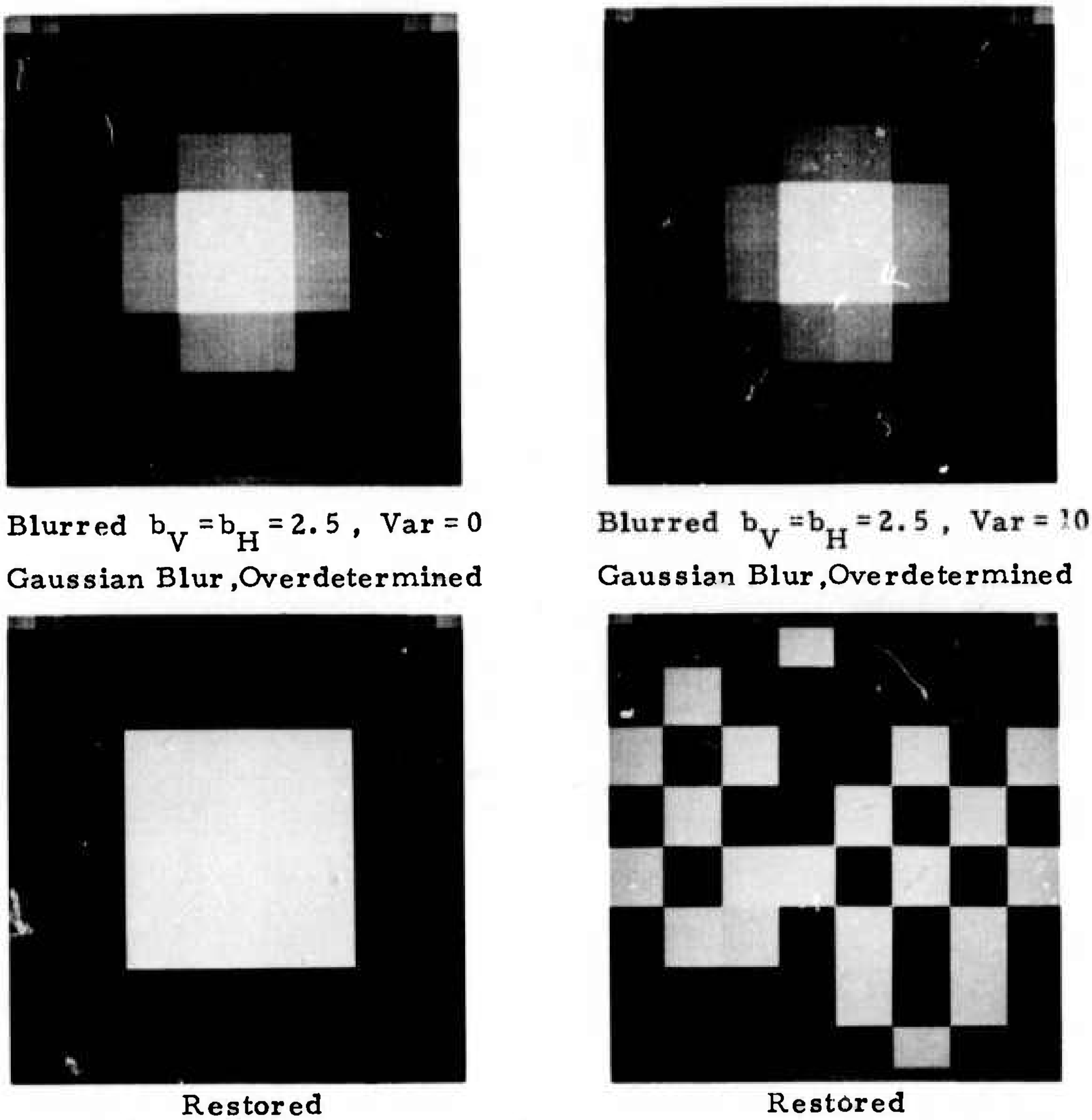
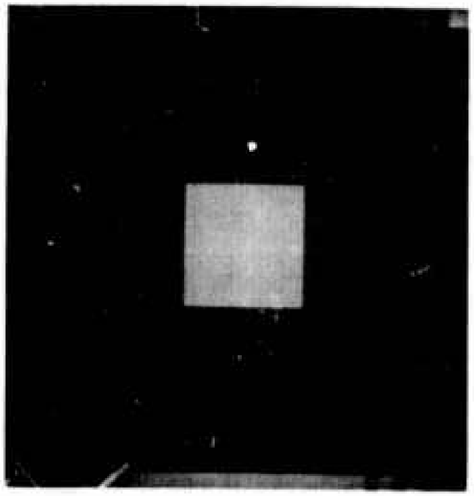
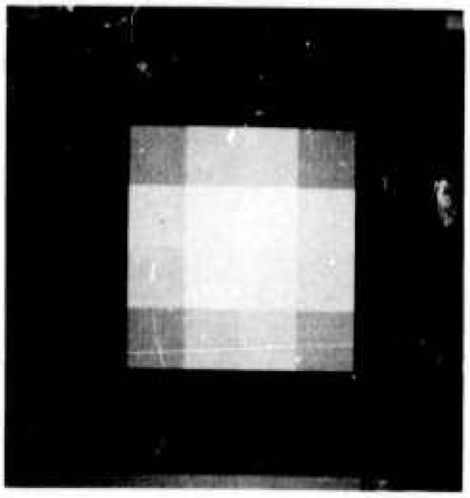


Figure (5. 1-6) Examples of Restoration with the Overdetermined Model and Gaussian Blur -II

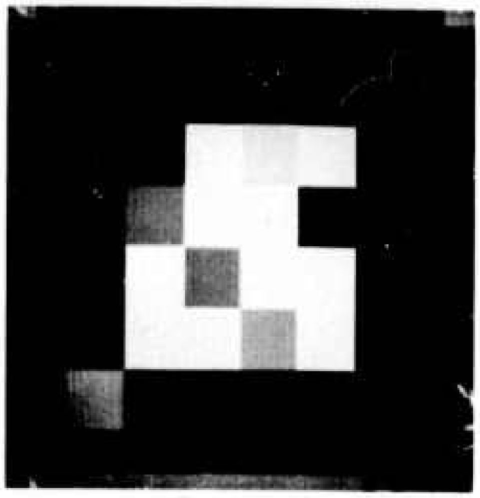


Blurred $b_V = b_H = 5000$, $Var = 0$
Gaussian Blur, Overdetermined



Blurred $b_V = b_H = 5000$, $Var = 10$
Gaussian Blur, Overdetermined

Restored



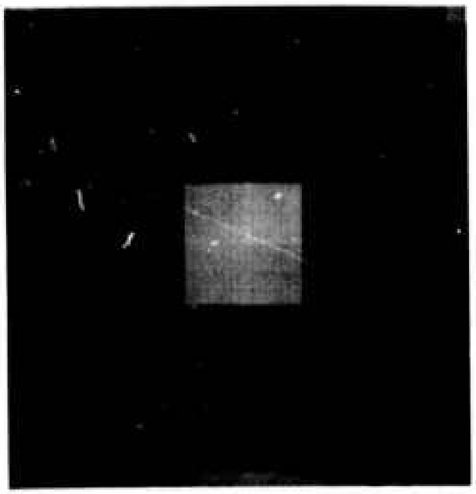
Restored

Figure (5.1-7) Examples of Restoration with the Overdetermined Model and Gaussian Blur-III somewhat in between the two previous results as far as fidelity of the restored picture to the original is concerned.

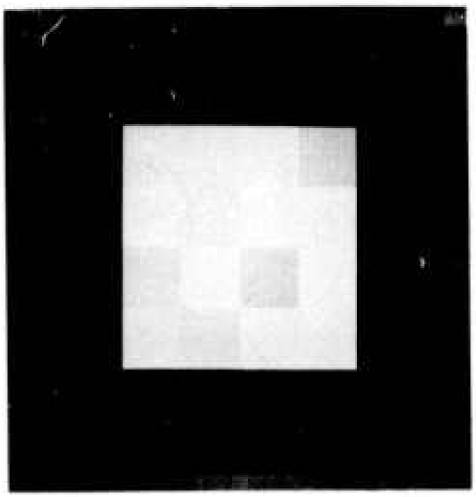
Figure (5.1-8) shows the result for $b_V = b_H = 0.70$ under more severe noise conditions. It evidences the little sensitivity of the solution to the increase in noise level for these values of blur coefficients.

These experimental results suggest that the blur coefficient influences considerably the amount of perturbation on the solution with respect to the corrupting noise. The concept of condition number developed in chapter 3 offers an adequate framework to explain and predict the behavior of the solution with respect to the perturbation represented by the noise.

Therefore, the condition number of the blur matrix $\underline{B} = \underline{B}_C \otimes \underline{B}_R$ for $\underline{B}_C = \underline{B}_R$ was computed as a function of the blur coefficient. This was done by the following procedure. Since the $(M^2 \times N^2)$ matrix $\underline{B}$ is given by the Kronecker product of two $(M \times N)$ matrices $\underline{B}_C$, its N^2 singular values are obtained by all the possible combinations of products of the N singular values of $\underline{B}_C$ [5-4]. The condition number $c(\underline{B})$ is the ratio of the largest to the smallest singular values of $\underline{B}$, which is equal to the square of the ratio of the largest to the smallest singular value of $\underline{B}_C$, this being the condition number of $\underline{B}_C$. Now, the square of the condition number of $\underline{B}_C$ is the condition number of $(\underline{B}_C^T \underline{B}_C)$ [3-8, page 223] which can be calculated



Blurred $b_V = b_H = .70$, Var = 30
Gaussian Blur, Overdetermined



Restored

Figure (5.1-8) Example of Restoration with the Overdetermined Model and Gaussian Blur-IV by multiplying the norm of $(\underline{B}_C^T \underline{B}_C)$ by the norm of $(\underline{B}_C^T \underline{B}_C)^{-1}$. This has been done by using the Froebenius norm, for several values of the blur coefficient. Figure (5.1-9) shows the result of this experiment. The condition number is maximum for moderate values of the blur coefficient. The curve explains the results obtained with the restoration of the noisy images. In fact, $b_V = b_H = 0.70$ is in the initial part, with low values of $c(\underline{B})$, and thus implies little effect of the noise on the restored image. $b_V = b_H = 2.5$, on the other hand, is on the peak of the curve with maximum perturbation and, finally, $b_V = b_H = 5000$ gives moderate values for the condition number and effect of the noise. The same type of curve for the condition number was observed for the sinc^2 spread function. Figure (5.1-10) displays the results. The matrix $(\underline{B}_C^T \underline{B}_C)$ becomes nearly singular for moderate values of the blur coefficient. An even greater variation of the condition number was observed as compared to the gaussian spread function.

The restoration experiments were repeated with the sinc^2 blur. Figure (5.1-11) shows the results for $b_V = b_H = 0.25$. The noise-free restoration reproduces the original picture while the noisy image is restored to values that are very close to the original ones. This corresponds to the very small condition number on this part of the curve. Figure (5.1-12) presents the experimental results for $b_V = b_H = 1.0$. In the noise free case, even though the

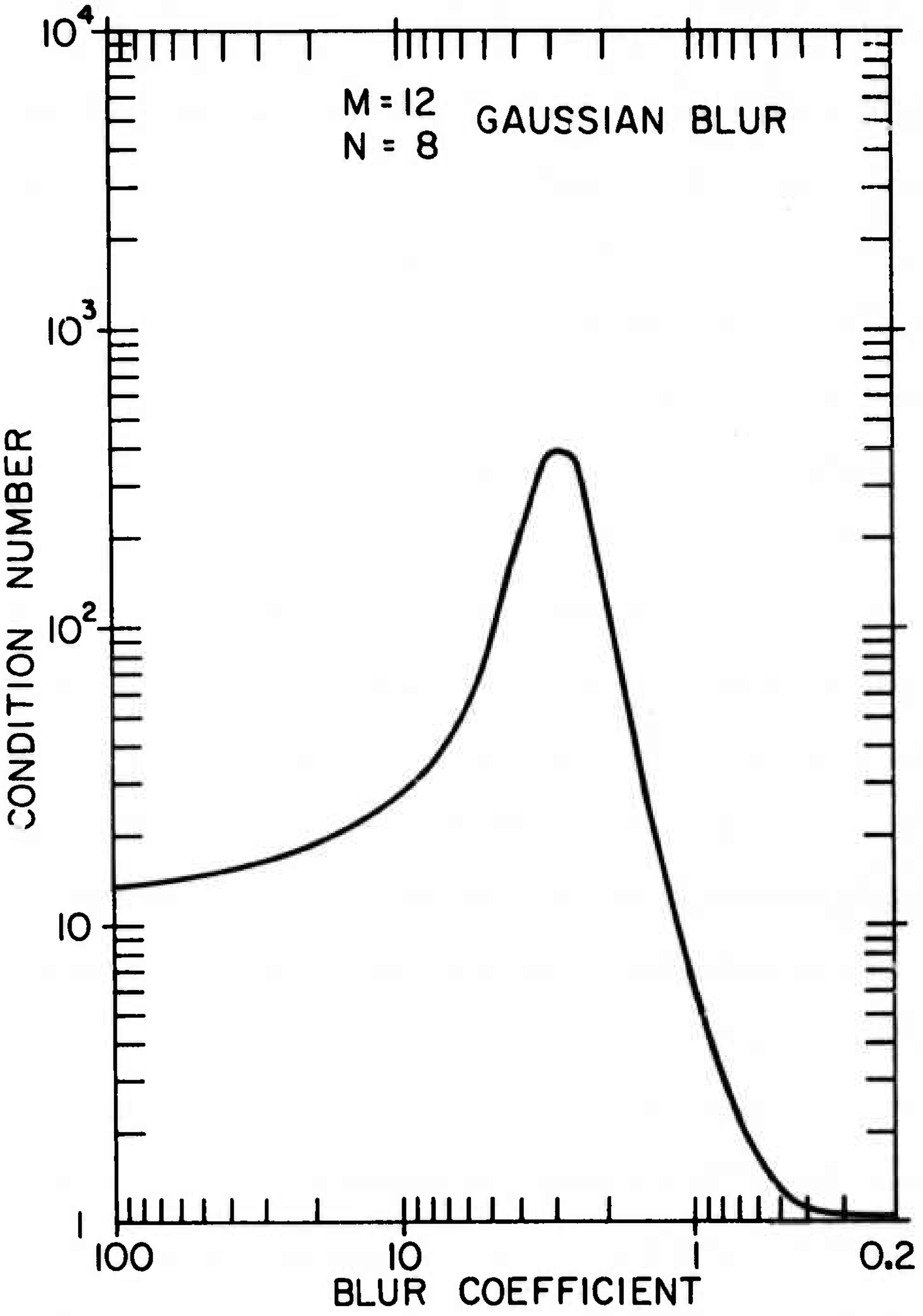


Figure (5.1-9) Blur Coefficient, Condition Number Curve for Gaussian Blur

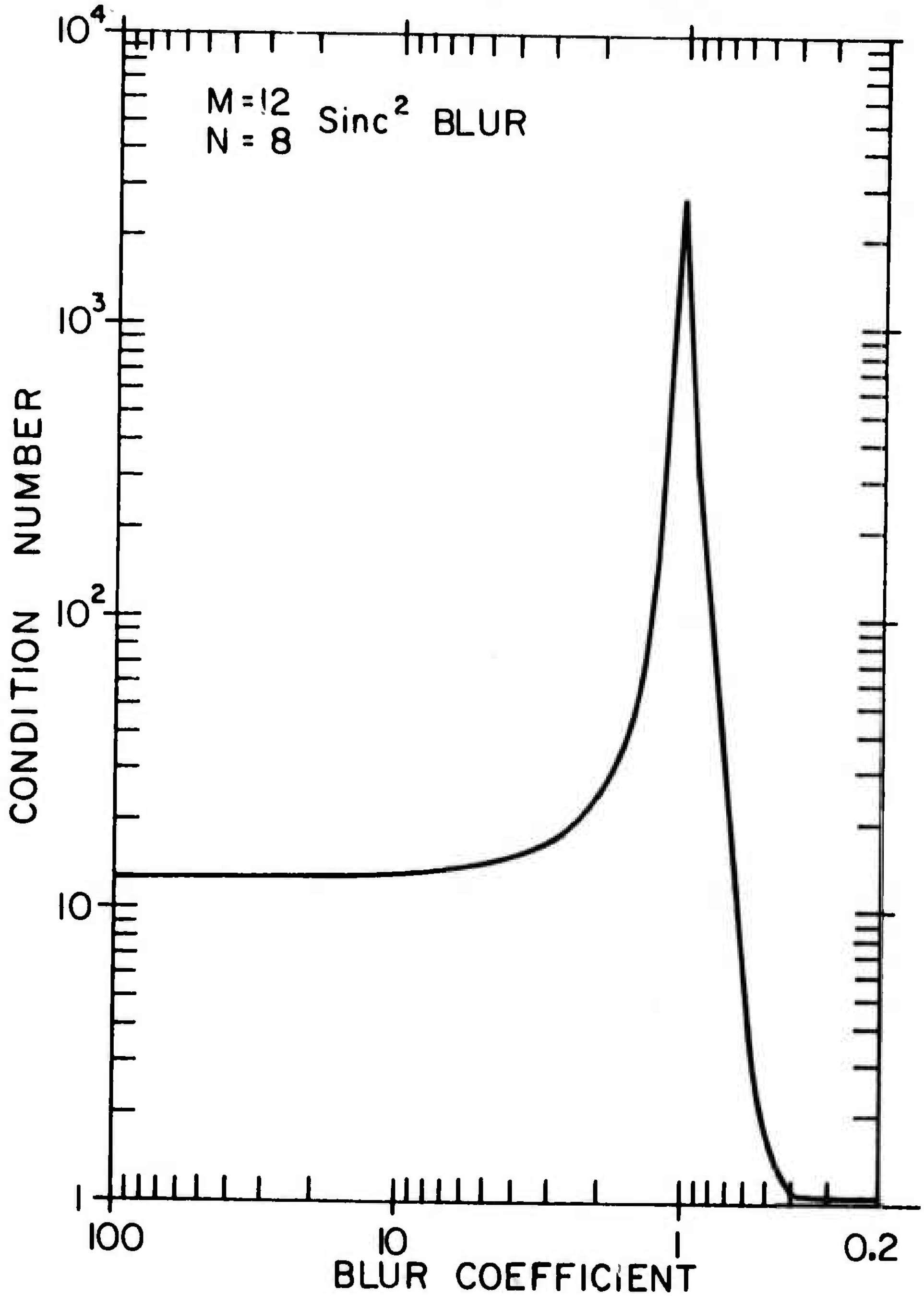
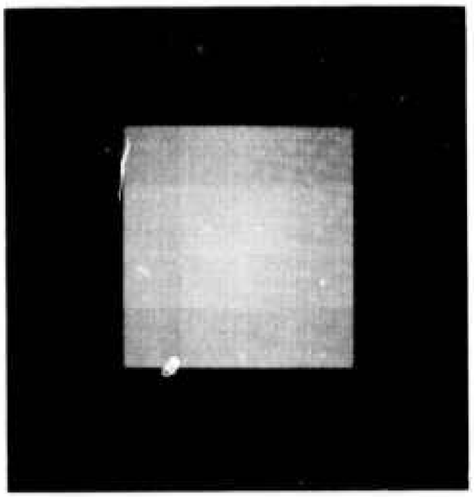
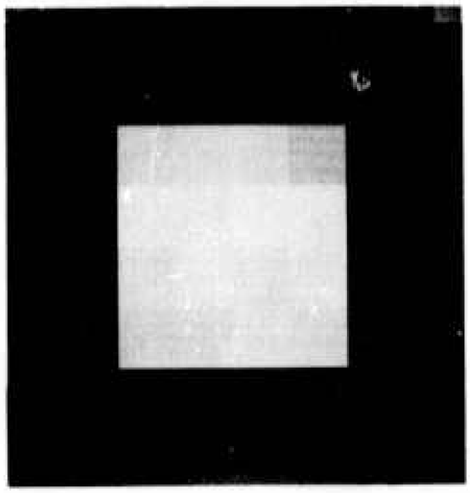


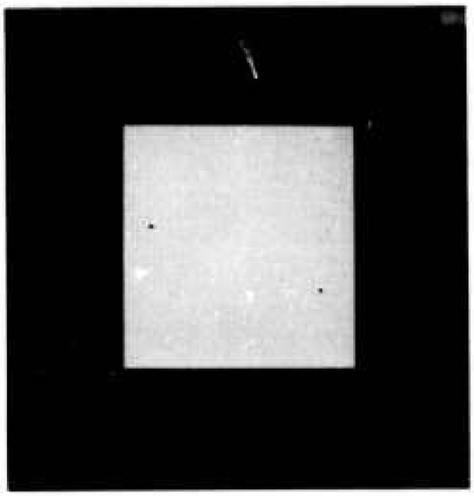
Figure (5.1-10) Blur Coefficient,
Condition Number Curve for $Sinc^2$ Blur



Blurred $b_V = b_H = .25$, Var = 0
$Sinc^2$ Blur, Overdetermined



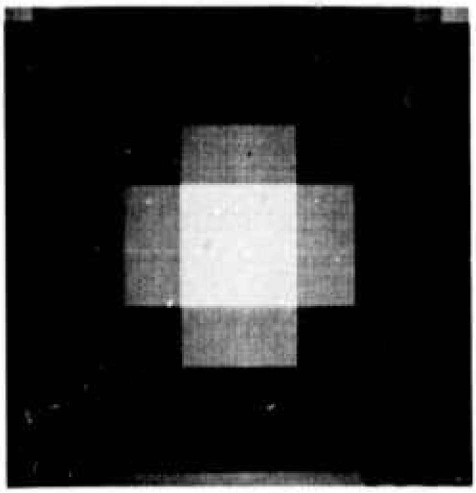
Blurred $b_V = b_H = .25$, Var = 40
$Sinc^2$ Blur, Overdetermined



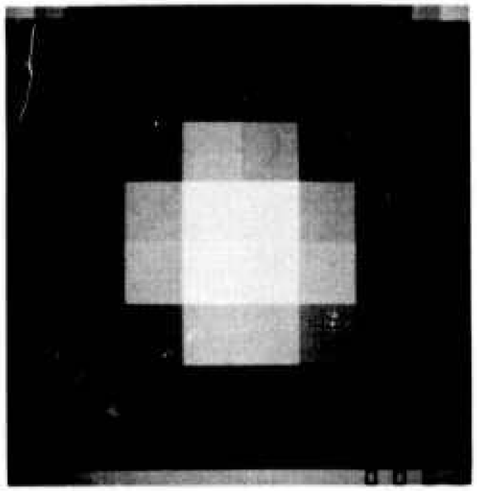
Restored

Restored

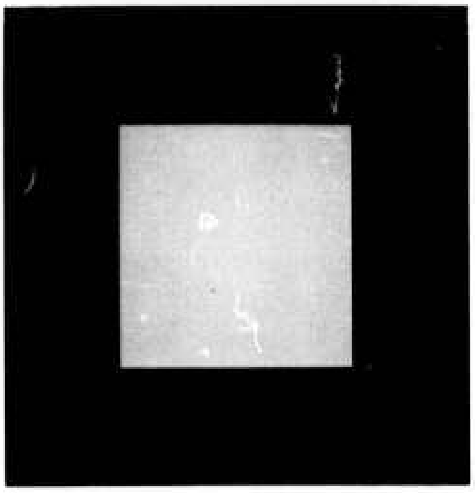
Figure (5.1-11) Examples of Restoration with the Overdetermined Model and $Sinc^2$ Blur-I



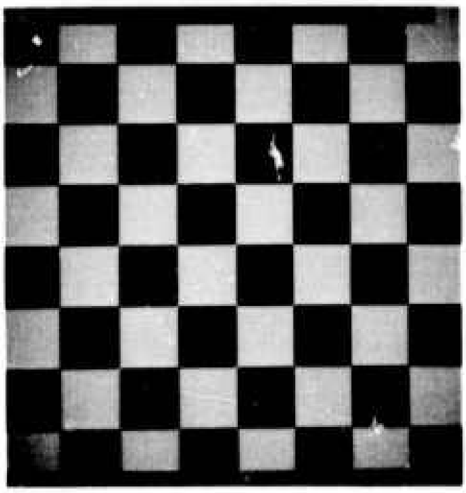
Blurred $b_V = b_H = 1.0$, Var = 0
$Sinc^2$ Blur, Overdetermined



Blurred $b_V = b_H = 1.0$, Var = 40
$Sinc^2$ Blur, Overdetermined



Restored



Restored

Figure (5.1-12) Examples of Restoration with the Overdetermined Model and $Sinc^2$ Blur -II restored image visually reproduces the original, the effect of the high condition number can be observed in the actual numerical results presented in Figure (5.1-13). The perturbation due to round-off is noticeable (single precision has been used in the calculations). In the noisy case, the displayed restored image shows the enormous effect of the perturbation, but in order to give an idea of the order of magnitude of the error in the estimation, the non-clipped numerical results are presented in Figure (5.1-14). The very large perturbations imposed by a high condition number are clearly displayed. The restoration under a moderate value of condition number, under sinc^2 blur, is displayed in Figure (5.1-15).

A complementary point of view to explain the effect of the noise on the restoration can be given. Equation (3.1-4) gives the expression of the covariance matrix of the estimator. Under white noise conditions, that expression reduces to

$$\underline{V}_{\hat{x}} = \sigma^2 (\underline{B}^T \underline{B})^{-1} \qquad (5.1\text{-}17)$$

In the condition number is high the matrix $(\underline{B}^T \underline{B})$ is nearly singular and large variances of the estimated values are expected. On the other hand, if the condition number is low, the opposite situation occurs and the variances of the estimated values are reduced.

Figure (5.1-16) and (5.1-17) display the condition number

ORIGINAL PICTURE

10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.00	10.00	245.00	245.00	245.00	245.00	10.00	10.00
10.00	10.00	245.00	245.00	245.00	245.00	10.00	10.00
10.00	10.00	245.00	245.00	245.00	245.00	10.00	10.00
10.00	10.00	245.00	245.00	245.00	245.00	10.00	10.00
10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00

RESTORED PICTURE

9.00	10.02	9.84	9.91	9.81	9.91	9.97	10.00
10.02	10.06	10.45	10.45	10.42	10.54	10.00	10.04
9.84	10.45	243.40	244.63	242.74	244.82	9.48	10.08
9.91	10.53	244.62	245.89	243.96	246.07	9.55	10.15
9.80	10.42	242.71	243.97	242.05	244.14	9.44	10.05
9.91	10.54	244.82	246.05	244.14	246.24	9.56	10.16
9.97	10.01	9.48	9.56	9.45	9.56	9.95	9.98
10.00	10.03	10.08	10.15	10.05	10.16	9.98	10.01

Figure (5.1-13) Numerical Results, $b_V = b_H = 1.0$

Var = 0, Sinc^2 Blur, Overdetermined Model

ORIGINAL PICTURE

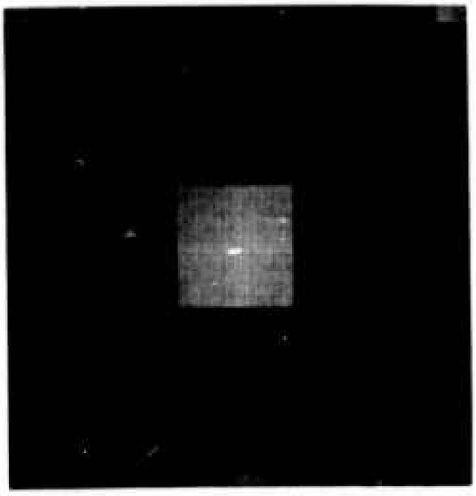
10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.00	10.00	245.00	245.00	245.00	245.00	10.00	10.00
10.00	10.00	245.00	245.00	245.00	245.00	10.00	10.00
10.00	10.00	245.00	245.00	245.00	245.00	10.00	10.00
10.00	10.00	245.00	245.00	245.00	245.00	10.00	10.00
10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00

RESTORED PICTURE

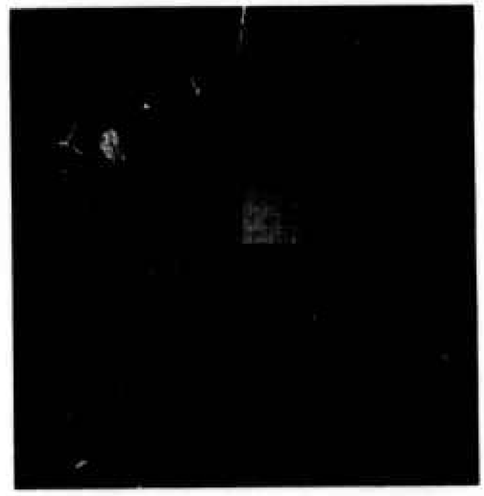
-802.18	3747.56	-7212.29	9442.45	-8991.54	6045.18	-2716.17	570.54
2106.96	-9870.84	19808.70	-27105.73	27210.39	-19439.42	9529.35	-2217.52
-3031.30	15010.06	-30925.02	43557.52	-43755.91	32449.76	-15977.14	3762.48
3636.16	-17655.67	36975.16	-50252.27	51292.08	-37259.20	18301.10	-4134.15
-3750.62	17550.89	-35592.72	48599.66	-47897.68	35320.97	-16198.66	3409.58
3651.48	-15718.55	31029.78	-39922.70	38959.14	-27329.24	11969.57	-2320.60
-2600.10	10499.66	-19248.51	23635.48	-21405.17	14413.81	-5539.95	867.91
992.43	-3670.29	6290.77	-7109.83	5971.90	-3729.23	1275.87	-159.55

Figure (5.1-14) Numerical Results, $b_V = b_H = 1.0$

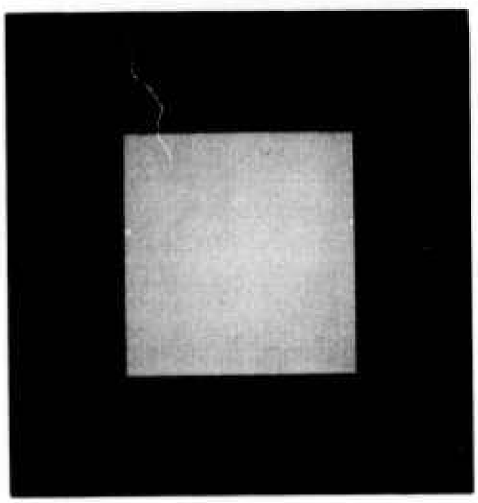
Var = 40, Sinc^2 Blur, Overdetermined Model



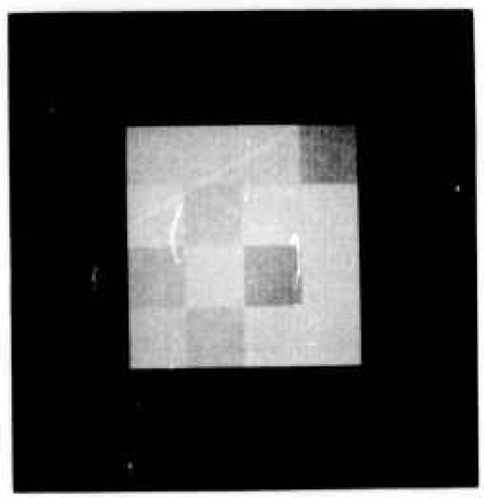
Blurred $b_V = b_H = 500$, Var = 0
Sinc^2 Blur, Overdetermined



Blurred $b_V = b_H = 500$, Var = 40
Sinc^2 Blur, Overdetermined



Restored



Restored

Figure (5.1-15) Examples of Restoration with the Overdetermined Model and Sinc^2 Blur-III

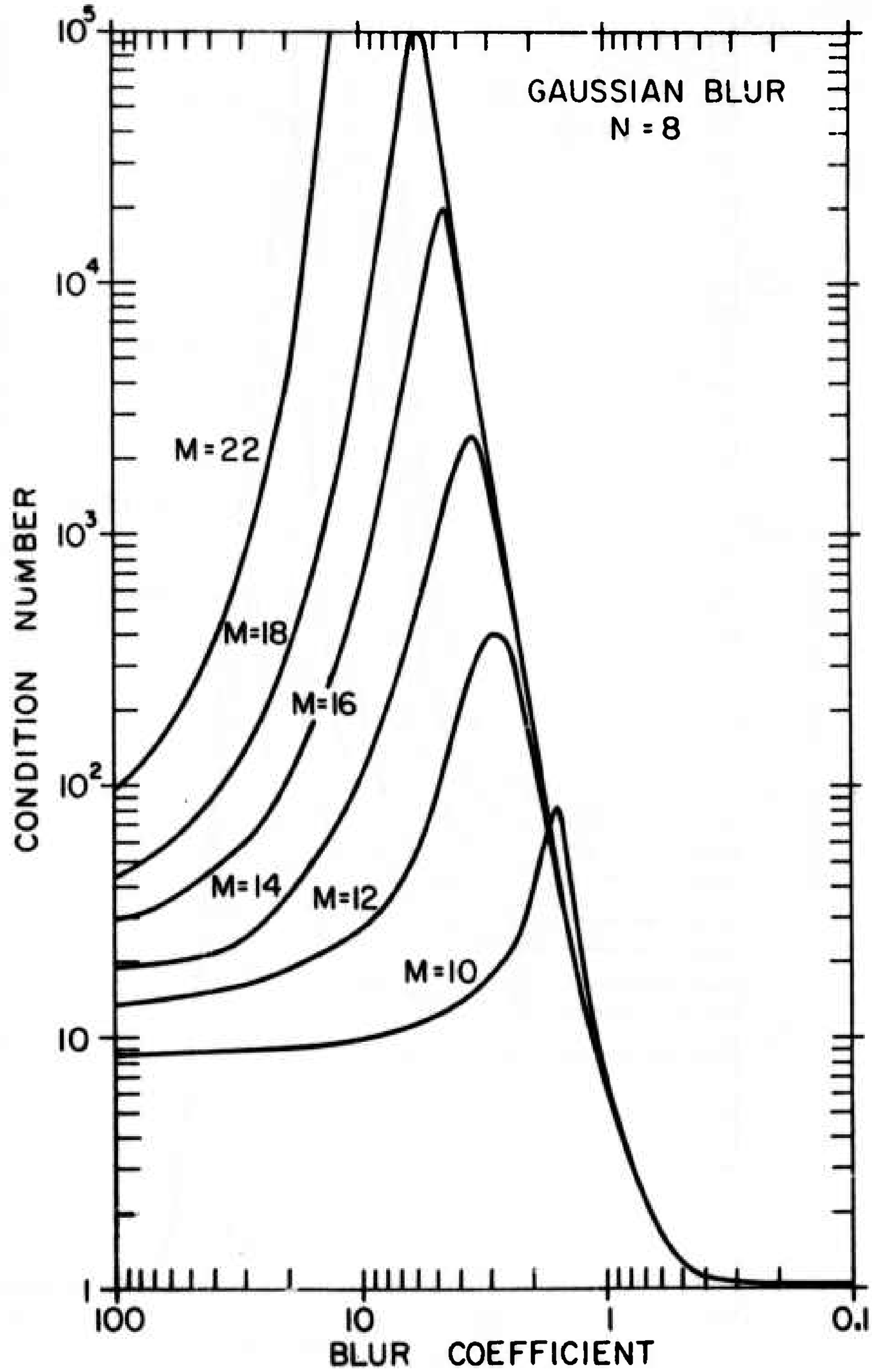


Figure (5.1-16) Blur Coefficient, Condition Number Curves for Gaussian Blur and Different Number of Sampled Values

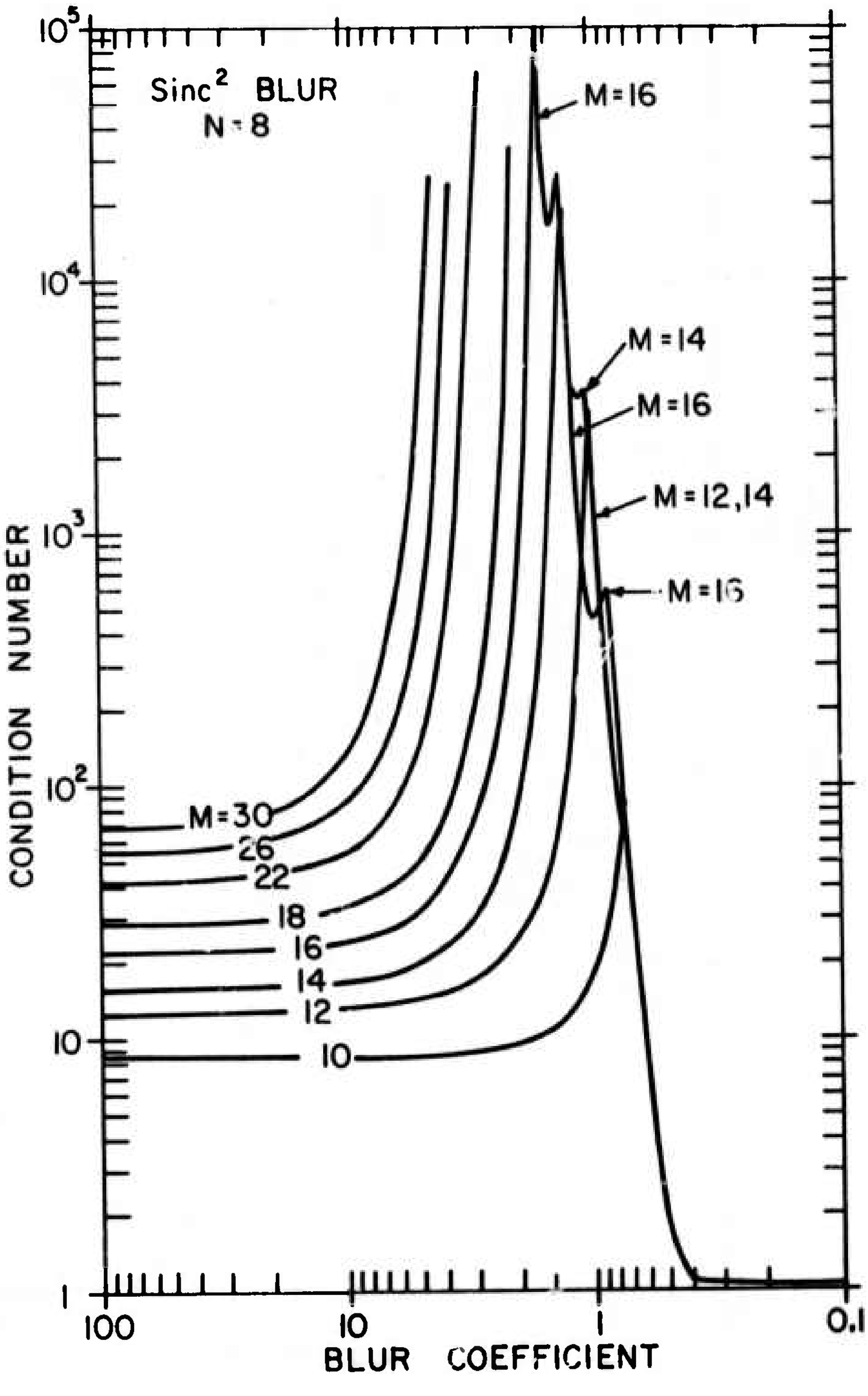
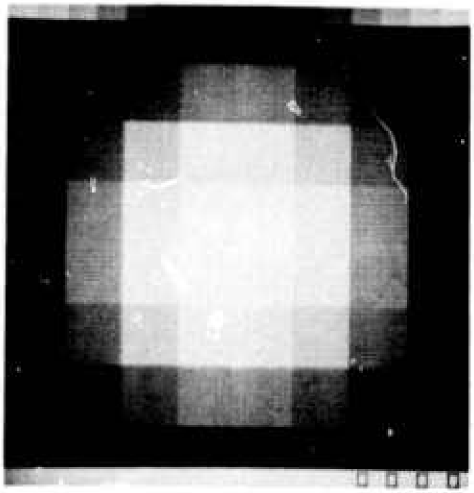


Figure (5.1-17) Blur Coefficient, Condition Number Curves for Sinc^2 Blur and Different Number of Sampled Values curves, for a given number of quadrature nodes ($N = 8$), varying the number of sampled values (M), while maintaining the structure of the blur matrix given in Figure (5.1-2). These results give more insight into the reason for the existence of a maximum of the condition number curves. This is due to the truncation of the point spread function. In fact, for increasing M, the number of points where this function can be nonzero is increased and the effect of the truncation starts only for higher blur coefficients. Consequently, the curves for different values of M have essentially a common ascending branch and the descending part starts at varying points for different values of blur coefficients. If there were no truncation, the curve would approach infinity very fast, the asymptotic value being obtained for the smoothest possible kernel, with constant value one, implying a blur matrix with rank one. With the truncation, the curves show a descending branch that begins at the point where the increasingly wider kernel starts to be cut down substantially. Now, for increasing value of blur coefficient, the curves tend to a finite value.

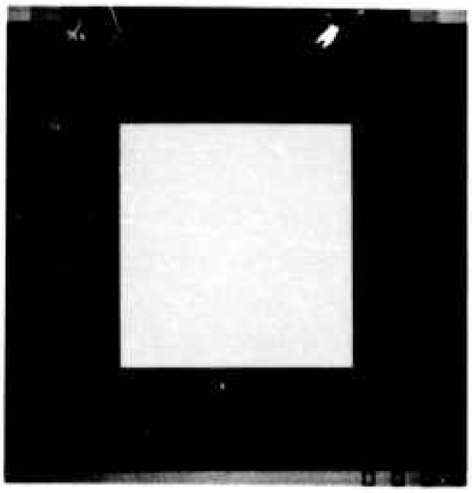
These curves can be used as a guide for the choice of the number of sampling points, once the number of quadrature nodes is fixed. For a very small amount of blur all curves coincide so that the designer may choose $M = N$ with almost no error. Blur in this case plays no role, only noise will affect the restoration. With increasing amount of blur, different numbers of sampling points will give different values of condition number. If a curve on an ascending branch is chosen, truncation will have no effect on the kernel but a high condition number will impose high variances on the estimators. If a curve on a descending branch is selected, lower variances of the estimators will be obtained, at the price of error on the estimation of the continuous function due to the truncation error in the discrete model. Therefore, a trade-off between the variance of the estimators and the truncation error of the discrete model can be characterized.

Although these conclusions are drawn based on the particular model discussed in this section, they are more general. This comes from the fact that the inverse of the integral operator that describes the blur is unbounded. Therefore, the closer the discrete model follows the continuous one, the more ill conditioned the former model tends to be. A move in the opposite direction reduces singularity but imposes modeling errors. This inevitable dilemma can only be broken with the intervention of correct a priori knowledge about the solution.

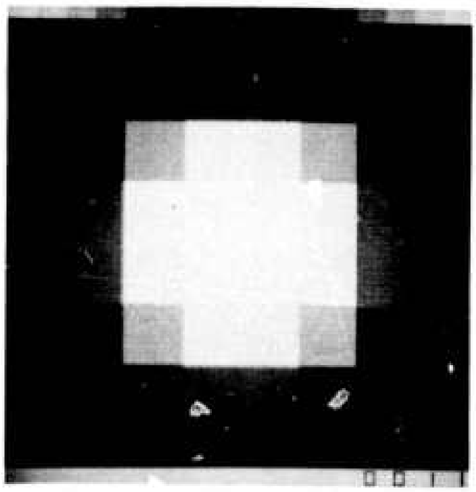
The effect of changes in the blur matrix was also experimentally confirmed. For this purpose an image was restored using a value of blur coefficient different from the one that was used for its blurring. Figure (5.1-18) shows the result for noise-free and



Blurred $b_V = b_H = 1.0$, Var = 0
Sinc^2 Blur, Overdetermined



Restored with $b_V = b_H = 1.0$



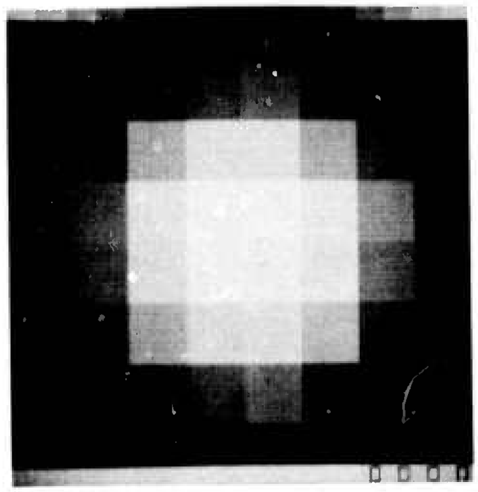
Restored with $b_V = b_H = 1.2$

Figure (5.1-18) Effect of Changes of the
Blur Matrix on the Restoration-Noise Free Case

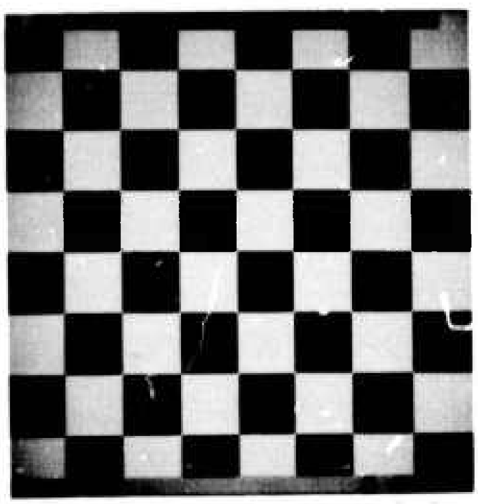
Figure (5.1-19) for noisy observations. The perturbation is much higher in the noisy case, with many reversals of signs in the solution. This in accordance with the conclusions of equation (3.3-34), which predicted that in the noise-free case, the condition number would matter, while in the presence of noise the square of this quantity would determine the amount of perturbation.

5.2 Restoration for the Underdetermined Model

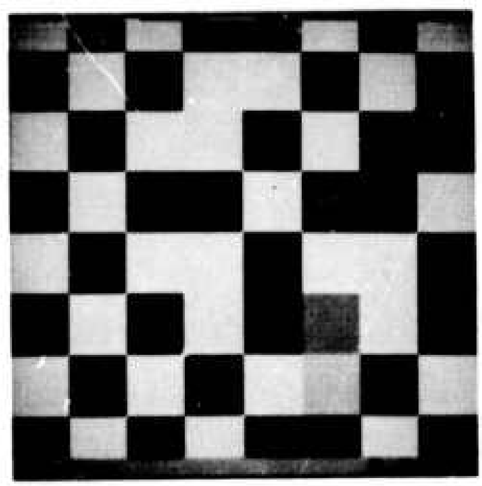
Another set of experiments has been performed for the underdetermined model, i.e., when the number of quadrature nodes exceeds the number of observed values. The following real situation is simulated by these experiments: the image of part of an object is taken (e.g., the photograph of a certain region by an earth resources satellite); as in the overdetermined model, a decision is made to place the nodes of the quadrature integration at equally spaced points on a rectangular grid. The sampling points are separated by the same distance as the quadrature nodes and they coincide with some of the quadrature nodes. The number of sampling points in this case will be determined by the size of the image. The point spread function is assumed to be truncated to twice the sampling distance, like in the overdetermined model. This determines the extent of the original picture that contributes to the blurred picture and only the quadrature nodes that make a nonzero contribution with this trun-



Blurred $b_V = b_H = 1.0$, Var = 40
Sinc^2 Blur, Overdetermined



Restored with $b_V = b_H = 1.0$



Restored with $b_V = b_H = 1.2$

Figure (5.1-19) Effect of Changes of the Blur Matrix on the Restoration - Noisy Case cated kernel are retained.

Figure (5.2-1) describes the data arrays. The vector representation for the original and blurred arrays is still valid, but the matrix $\underline{B}$ is the transpose of the matrix of the overdetermined model. Figure (5.2-2) shows the partition of $\underline{B}$ in this case. The structure of the submatrices $\underline{B}_{i,j}$ is described in Figure (5.2-3). In the case of separable, space invariant blur, the unidimensional blur matrix has the form expressed by Figure (5.2-4).

The experiments have been performed with the gaussian shaped blur and white gaussian noise. The original picture is composed of (12 x 12) pixels, coinciding with the (8 x 8) picture depicted in Figure (5.1) on the center part.

As pointed out in the previous chapter, there is no unbiased estimator in this case and the solution of the least squares problem is not unique. The minimum norm solution is given by

$$\hat{\underline{x}} = \underline{B}^{+} \underline{y} \qquad (5.2\text{-}1)$$

Since the blur matrix is the transpose of the corresponding matrix for the overdetermined model, it follows that $\underline{B}$ has full row rank and $\underline{B}^{+}$ can be given by

$$\underline{B}^{+} = \underline{B}^{T}(\underline{B}\underline{B}^{T}) \qquad (5.2\text{-}2)$$

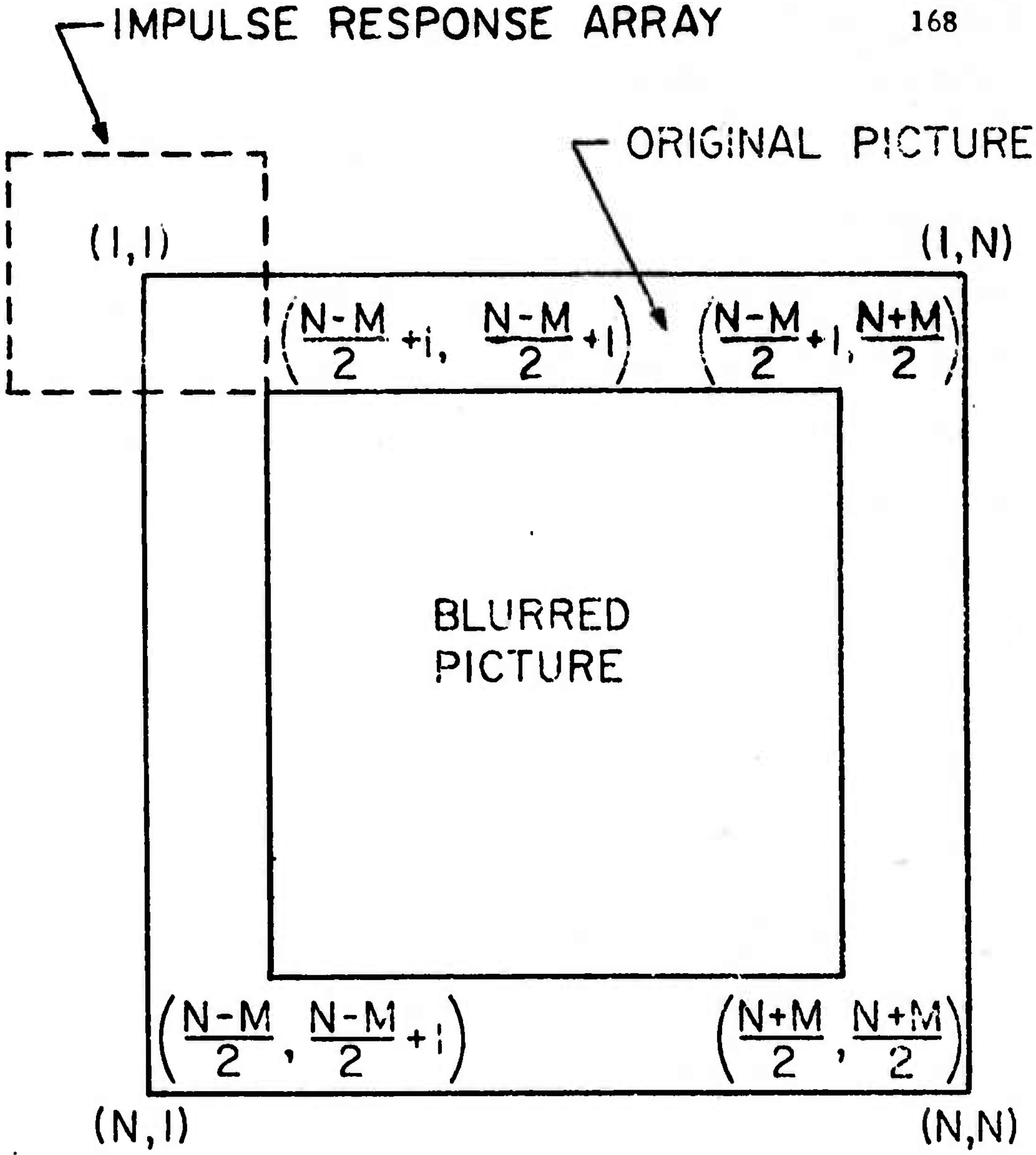


Figure (5.2-1) The Data Arrays in the Underdetermined Model

$$
\underline{B} = \begin{bmatrix}
B_{1,1} & B_{1,2} & \cdots & B_{1,L} & 0 & 0 & \cdots & 0 \\
0 & B_{2,2} & \cdots & B_{2,L} & B_{2,L+1} & 0 & \cdots & 0 \\
0 & 0 & & & & & & \vdots \\
\vdots & \vdots & & & & & & 0 \\
0 & 0 & \cdots & & & 0 \;\; B_{M,M} & \cdots & B_{M,N}
\end{bmatrix}
$$

Figure (5.2-2) Partition of the Blur Matrix $\underline{B}$ in the Underdetermined Model

$$\underline{B}_{m_2, n_2} = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,L} & 0 & 0 & \cdots & 0 \\ 0 & b_{2,2} & \cdots & b_{2,L} & b_{2,L+1} & 0 & \cdots & 0 \\ 0 & 0 & & & & & & \vdots \\ \vdots & \vdots & & & & & & 0 \\ 0 & 0 & \cdots & & 0 & b_{M,M} & \cdots & b_{M,N} \end{bmatrix}$$

Figure (5.2-3) Composition of the Submatrices $\underline{B}_{m_2, n_2}$

in the Underdetermined Model

$$B_C = \begin{bmatrix} b_V(1) & b_V(2) & \cdots & b_V(L) & 0 & 0 & \cdots & & 0 \\ 0 & b_V(1) & \cdots & b_V(L-1) & b_V(L) & 0 & \cdots & & 0 \\ 0 & 0 & & & & & & & \vdots \\ \vdots & \vdots & & & & & & & 0 \\ 0 & 0 & \cdots & & 0 & b_V(1) & b_V(2) & \cdots & b_V(L) \end{bmatrix}$$

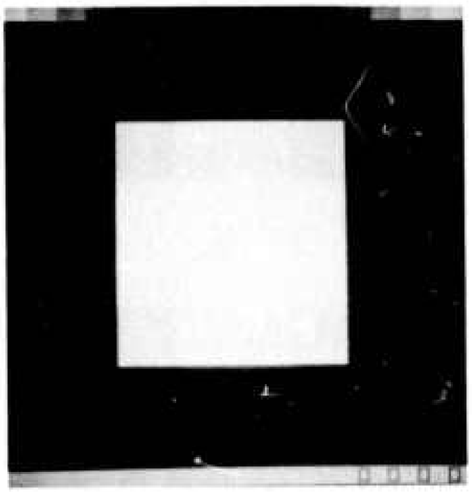
Figure (5.2-4) Unidimensional Blur Matrix in the Separable, Space Invariant Case Underdetermined Model

The assumption of shift invariance allows the computation of (5.2-1) to be done very efficiently using Fast Fourier Transform algorithms. In the separable case, using a derivation entirely analogous to the one that led to equation (5.1-11), it is possible to conclude that

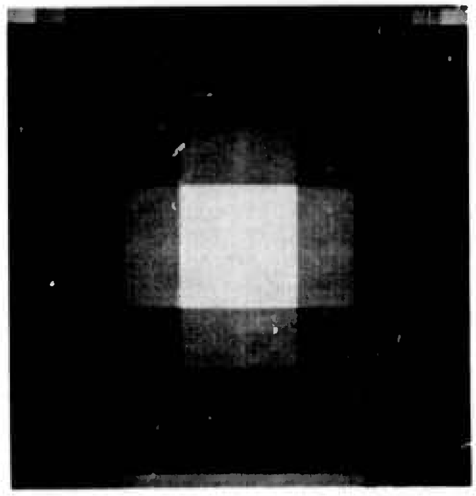
$$\underline{B}^{\dagger} = \underline{B}_c^{\dagger} \otimes \underline{B}_R^{\dagger} \qquad (5.2\text{-}3)$$

The noise-free minimum norm solution is not necessarily the original picture and this is clearly shown in Figures (5.2-5) and (5.2-6), for blur coefficients set at .5, 5 and 500. Note that only the center (8 x 8) part of the restored (12 x 12) picture is shown, blown up to (256 x 256).

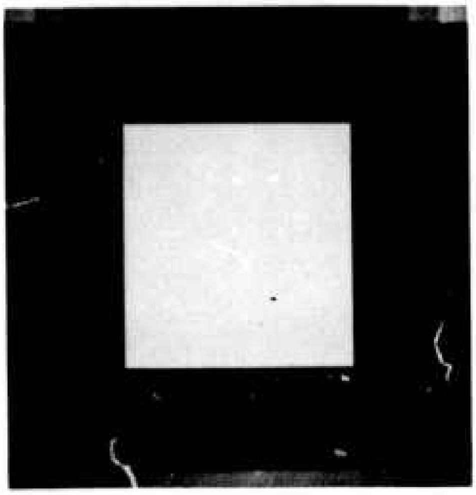
The noisy restorations, displayed in Figure (5.2-7) and (5.2-8) show the same pattern of the overdetermined model, namely, small perturbation in the solution due to noise for small blur, followed by large and moderate perturbations for increasing values of the blur coefficient. This fact cannot be explained by the condition number since it is infinite in this case. However, since $\underline{B}$ is the transpose of the matrix in the overdetermined case, and considering the fact that the nonzero eigenvalues of $\underline{B}^T\underline{B}$ and $\underline{B}\underline{B}^T$ are the same [3.5, page 41] it turns out that the ratio of the largest to the smallest nonzero eigenvalue of $\underline{B}^T\underline{B}$ follows the curve given by



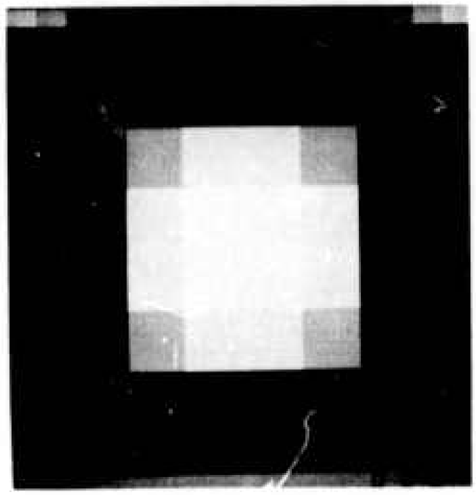
Blurred $b_V = b_H = .50$, Var = 0
Gaussian Blur, Underdetermined



Blurred $b_V = b_H = 5.0$, Var = 0
Gaussian Blur, Underdetermined

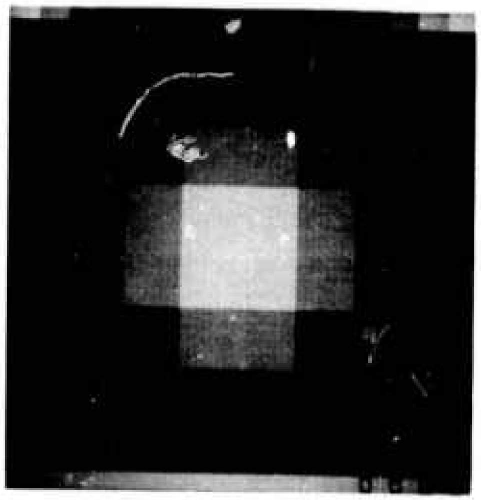


Restored

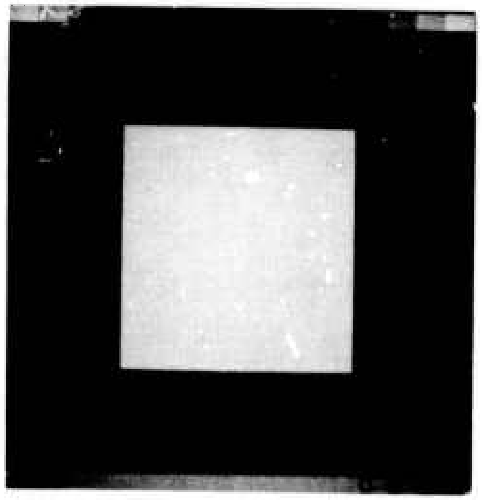


Restored

Figure (5.2-5) Restoration for the
Underdetermined Model - Noise Free Case I

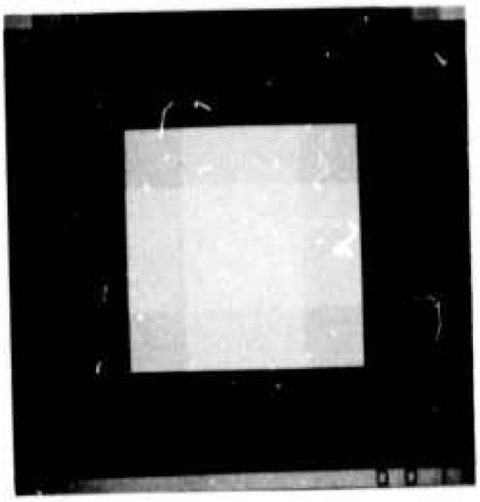


Blurred $b_V = b_H = 500$, Var = 0
Gaussian Blur, Underdetermined

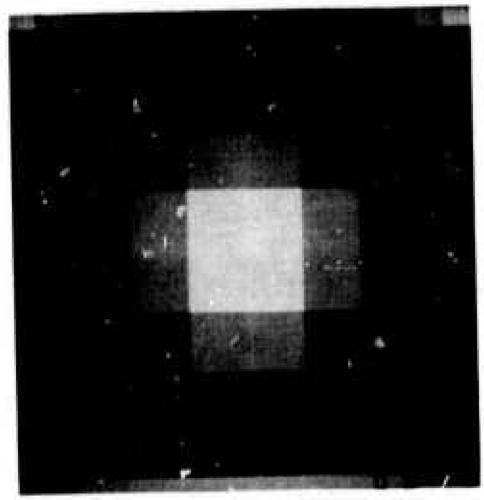


Restored

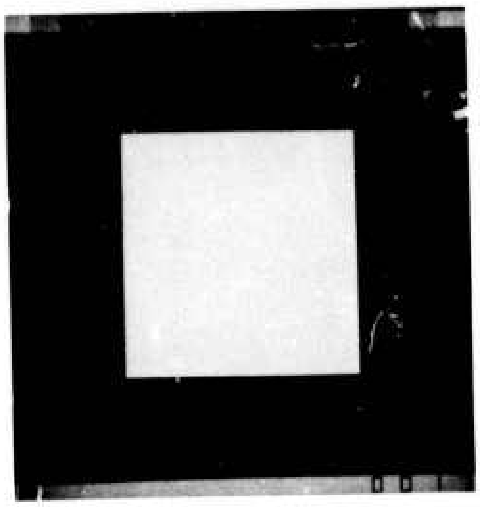
Figure (5.2-6) Restoration for the Underdetermined Model - Noise Free Case II



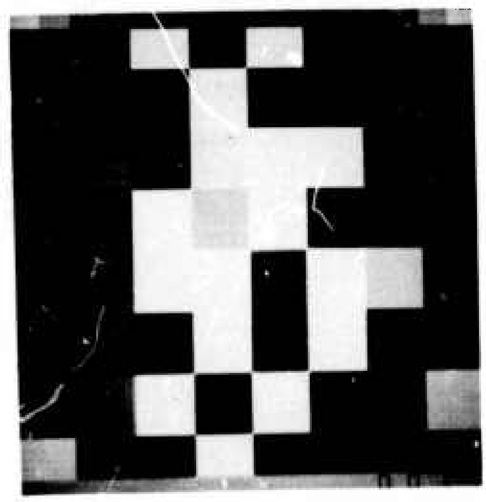
Blurred $b_V = b_H = .50$, Var = 10
Gaussian Blur,Underdetermined



Blurred $b_V = b_H = 5.0$, Var = 10
Gaussian Blur,Underdetermined

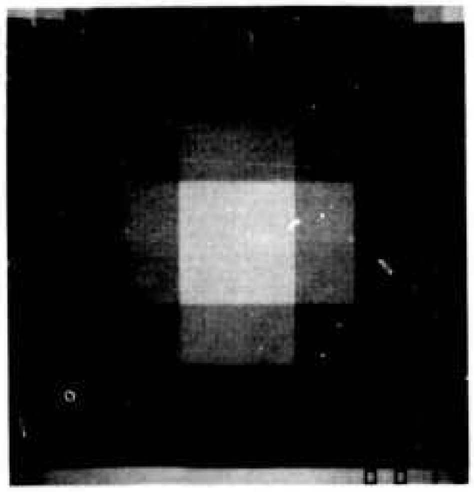


Restored

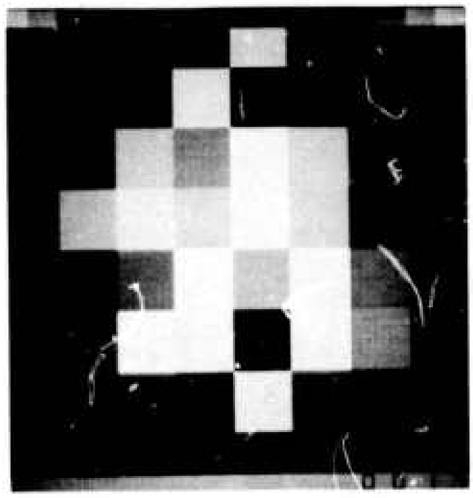


Restored

Figure (5.2-7) Restoration for the Underdetermined Model - Noisy Case I



Blurred $b_V = b_H = 500$, Var = 10
Gaussian Blur, Underdetermined



Restored

Figure (5.2-8) Restoration for the Underdetermined Model
Noisy Case II

Figure (5.1-9). Therefore, the ratio of the largest to the smallest finite principal axes of the degenerate K-ellipsoid follows the same curve.

On the other hand, the minimum norm solution is obtained by projecting the origin of the x-space orthogonally onto the subspace which consists of null space of $\underline{B}$, $N(\underline{B})$, added to $\underline{B}^{+}\underline{y}$. Therefore, when the minimum norm solution ($\underline{B}^{+}\underline{y}$) is taken from this subspace, no variation in this solution due to noise is allowed in the direction of the eigenvectors that span $N(\underline{B})$. These are precisely the eigenvectors corresponding to the zero eigenvalues of $\underline{B}^{T}\underline{B}$. Only variations along the eigenvectors corresponding to nonzero eigenvalues are allowed. These variations are in the (nondegenerate) ellipsoid that consists of the intersection of the original (degenerate) ellipsoid and the hyperplance that passes through the origin and is orthogonal to $N(\underline{B})$, that is, the range of $\underline{B}^{T}$, $R(\underline{B}^{T})$. The shape of this ellipsoid is the same as the shape of the ellipsoid of the dual, overdetermined model because the eigenvalues are identical. Therefore, the variations of the solution of the underdetermined model in this subspace of restricted dimensions should be of the same type as in the corresponding overdetermined model. Viewed from another perspective, this situation can be described as follows: by projecting the origin onto $N(\underline{B})$ added to $\underline{B}^{+}\underline{y}$, to obtain the minimum norm solution, $\underline{B}^{+}\underline{y}$, the infinite variances of the underdetermined model are avoided, leaving only the finite variances in the directions of the nondegenerate axes of the ellipsoid. This is done at the price of imposing bias, since the lack of information in the sample is compensated, not by a correct a priori information about the original picture, but by merely imposing a minimum norm solution. This trade-off between bias and variance is somehow analogous to the one between modeling error and high condition number in the choice of the size of the point spread function. A similar situation will also occur with the use of linear equality and inequality constraints in the restoration.

5.3 The Computation of Confidence Intervals and Hypothesis Testing in the Overdetermined Model

Some computational work has been performed with the objective of determining both confidence intervals and results of hypothesis testing in the linear model for restoration. In order to simplify the calculations, the unidimensional regression model has been employed.

Figures (5.3-1) and (5.3-2) present the results of the computation of the 68% confidence interval for individual pixel values, under gaussian and sinc^2 blur, respectively. The correlation of these curves with those of the condition number is clear. The higher this quantity the greater the confidence interval for the same pixel value,

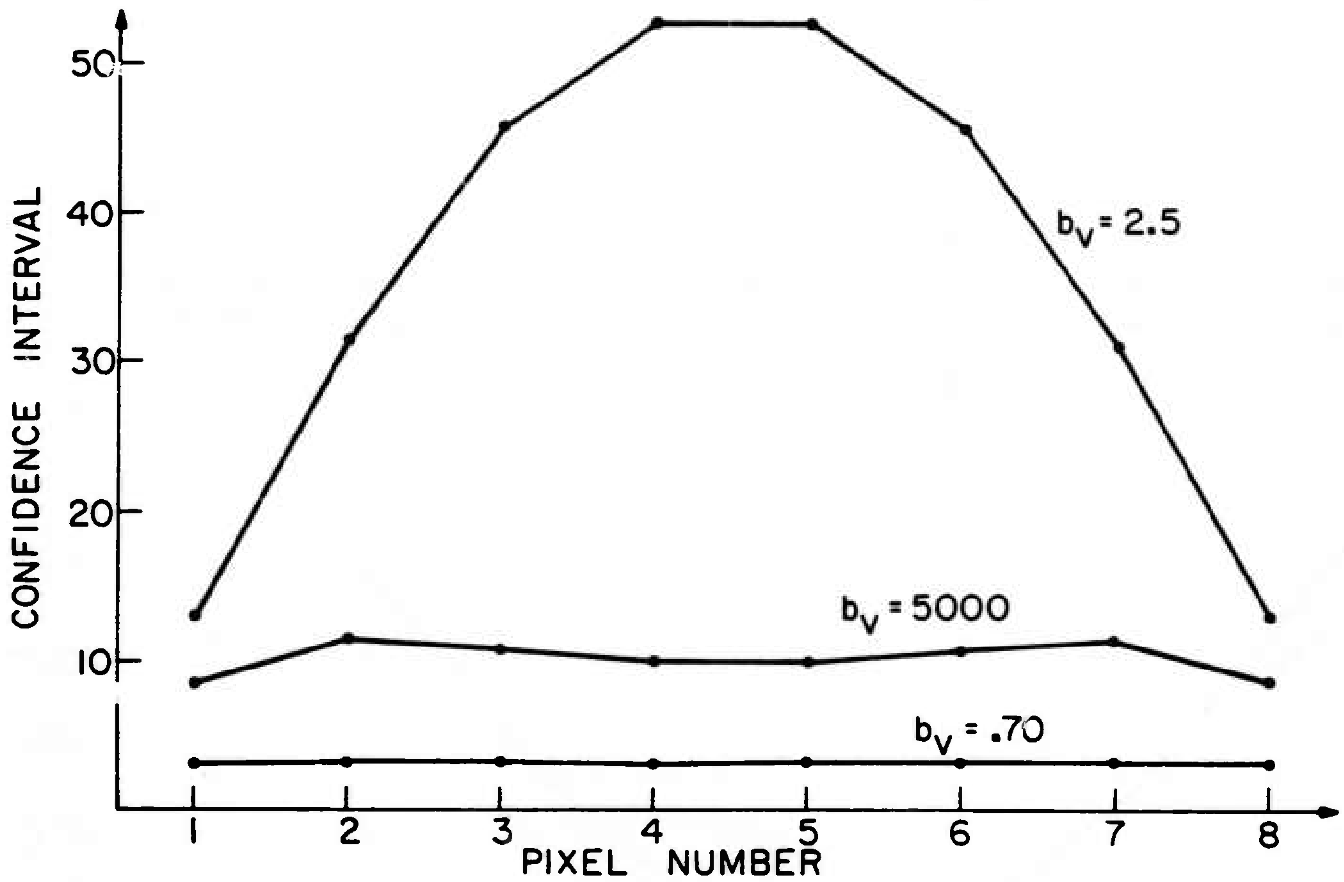


Figure (5.3-1) 68% Confidence Intervals for the Unidimensional, Overdetermined Model, Gaussian Blur, Unit Variance

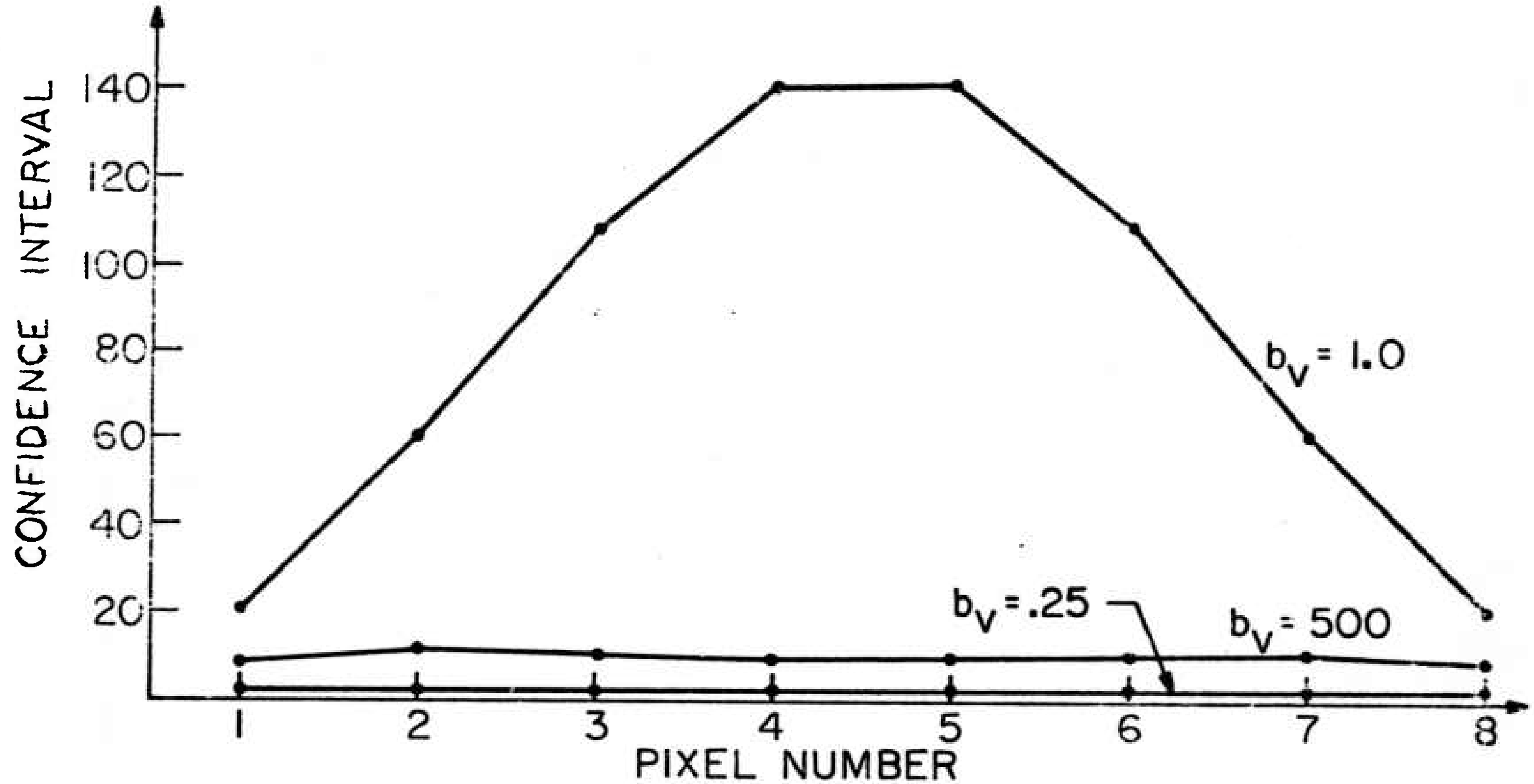


Figure (5.3-2) 68% Confidence Intervals for the Unidimensional, Overdetermined Model, Sinc^2 Blur, Unit Variance due to the larger variances involved. Furthermore, for large values of blur coefficient, the curves for gaussian and sinc^2 blur tend to coincide. This is due to the fact that the truncated spread functions assume the constant value unity when the blur coefficient tends to infinity.

The hypothesis testing experiments involve one pixel value in the unidimensional model, with sinc^2 blur and white gaussian noise. Two distinct sets of tests have been performed, the first with known variance, using the normal distribution, and the second under unknown variance, making use of the Student's distribution. The reader is referred to Appendix A for the theoretical material concerning hypothesis testing.

In both tests the level of confidence is set at 10%; the tests are two-sided, testing the fourth pixel value, with the null hypothesis H_o being $x_4 = A$ against the alternative hypothesis $x_4 \neq A$, for different values of A; the true value of x_4 is set at 245, the variance of the gaussian noise is 50 in both tests. Tables 5.3-1 and 5.3-2 present the results for the normal and Student's distribution tests, respectively.

Again, the correlation of the testing results with the condition number is evident through the inspection of the tables. A higher condition number is associated with a higher variance of the statistics used in the test. For a given size of the test (or probability of false alarm), fixed by the Neyman Pearson criterion, the power (or

Table 5.3-1 Hypothesis Testing
For Pixel Values, Normal Distribution

a) Variance known (normal distribution)

Level of confidence: 10% (two-sided)

True value: 245

Variance: 50

Blur: sinc^2 (diffraction limited)

H_o: $x_4 = A$ $\quad$ H_1: $x_4 \neq A$

Blur Coefficient	Condition Number	A: 245	A: 145	A: 45
0.25	1.0	-0.156 (accept)	12.62 (reject)	12.40 (reject)
1.0	2500	0.999 (accept)	1.20 (accept)	1.40 (accept)
500	13.0	-1.42 (accept)	1.40 (accept)	4.23 (reject)

Decision Regions

Reject H_o | Accept H_o | Reject H_o

-1.645 | +1.645

Table 5.3-2 Hypothesis Testing
For Pixel Values, Student's Distribution

b) Variance unknown (Student's distribution)

Level of confidence : 10% (two-sided)

True value: 245

Variance: 50

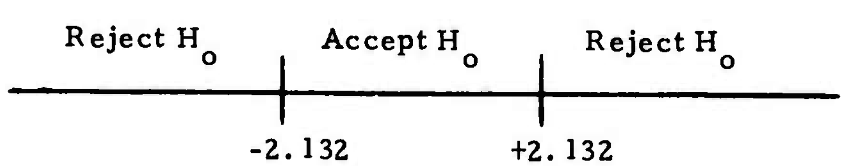
Blur: sinc^2 (diffraction limited)

Degrees of freedom: 12-8 = 4

$H_o: x_4 = A \quad H_1: x_4 \neq A$

Blur Coefficient	Condition Number	A: 245	A: 145	A: 45
0.25	1.0	-0.140 (accept)	11.34 (reject)	22.83 (accept)
1.0	2500	0.751 (accept)	0.902 (accept)	1.05 (accept)
500	13.0	-1.49 (accept)	1.46 (accept)	4.41 (reject)

Decision Regions



probability of detection) decreases with the variance. This explains the smaller probability of rejecting the null hypothesis when it is false, obtained for higher condition numbers.

6. EXPERIMENTAL RESULTS WITH LINEAR CONSTRAINED RESTORATION

In this chapter the results obtained with linear constrained restoration will be discussed. The first section presents experiments with linear equality constraints while the second section focuses on linear inequality restrictions.

6.1 Equality Constraints

As a first attempt to overcome the instabilities found in the use of unconstrained restoration with regression techniques, a single equality constraint has been imposed on the overdetermined model. The constraint consists of restricting the sum of the restored values to be equal to the sum of the original pixels. Since the analyst would not have direct access to this value in a real world experiment, the sum has been varied. The application of the Toro Vizcarrondo-Wallace test showed that there was an improvement in the mean square error for considerable variation of the constrained value, under a given confidence level. However, the variation of the numerical answer was minimal. This can be explained in view of the fact that for small condition number the unconstrained solution nearly satisfies the constraint and with moderate or large condition number the instabilities are in the form of oscillations from pixel to pixel.

The imposition of a given value for the sum of the restored pixels simply changes the D.C. component of the waveform without affecting the oscillations. Adopting the point of view of ellipsoids, this means that this elimination of one dimension in a $8 \times 8 = 64$-dimensional ellipsoid does not seem to be done in the direction of the eigenvectors corresponding to the smallest singular values.

In order to obtain a reasonable decrease in the variance, a higher dimensional and more appropriate restriction should be necessary. Considering both the nature of the image and the characteristic of the oscillations, it was felt that the restriction that pairs of adjacent pixels should be equal would tend to damp out the oscillations. This is a 32-dimensional restriction in a 64-dimensional space. In the particular case of the image used in these experiments this restriction is satisfied by the original image. In other cases, the imposition of these constraints will represent a smoothing of the solution in relation to the original. Observe that the rows of the matrix $\underline{A}$ of equation (4.2-1) have in this case the property of being shifted versions of the first row which opens the possibility of computation of (4.2-2) by Fourier Transform methods in the case of space invariant blur and white noise.

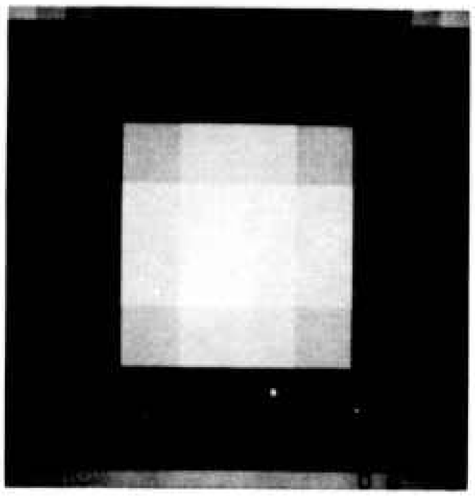
Equation (4.2-2) has been solved using the simplification that comes from the assumption of separable blur functions and white noise. This implies that the matrix $(\underline{H}^T\underline{H})$ can be inverted by taking the Kronecker product of smaller matrices.

Figures (6.1-1), (6.1-2), and (6.1-3) compare the results of unconstrained and linear equality constrained restorations in the overdetermined model for gaussian shaped blur, under the situations of small, large and moderate condition number. With small condition number the constrained restoration differs very little from the unconstrained case and both are very close to the true value. It is with larger values of condition number that the effect of the constraints, blocking the oscillatory nature of the unconstrained estimator, can be observed.

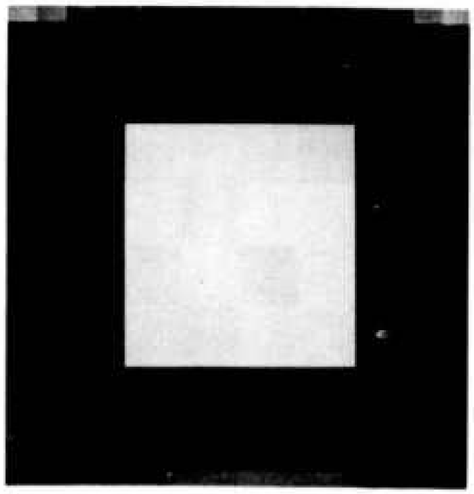
Similar results have been obtained with diffraction limited blur, shown in Figures (6.1-4), (6.1-5), and (6.1-6). In this case the problem can become extremely ill conditioned, and for blur coefficients equal to 1.0 the round-off error in evaluating the inverse of the matrix $[\underline{A}^T(\underline{H}^T\underline{H})^{-1}\underline{A}]$ prevents a meaningful result to be obtained, so that the high condition number situation is exemplified by the somehow better conditioned case of $b_V = b_H = 1.3$.

In both types of blur the statistics for the F-test have been computed and, under any reasonable confidence level, the hypothesis specifying that the linear relationships are true is accepted, confirming their validity.

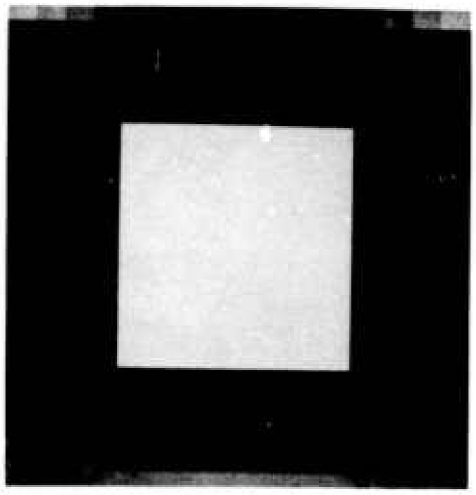
The results obtained with both types of blur seem to be indicative of the degree of damping of the oscillations that can be achieved



Blurred $b_V = b_H = .70$, Var = 10
Gaussian Blur, Overdetermined

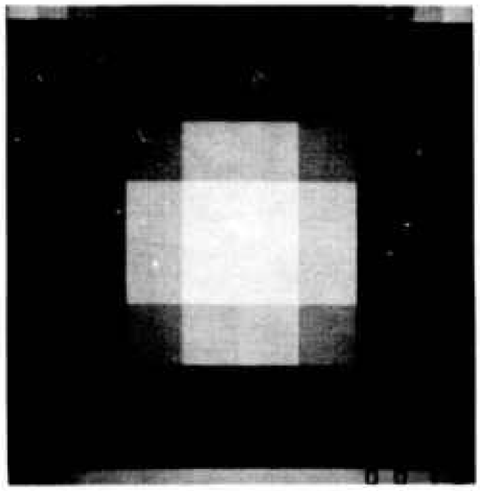


Unconstrained Restoration

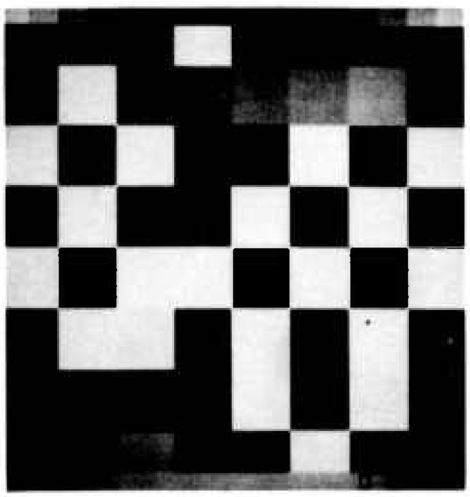


Equality Constrained Restoration

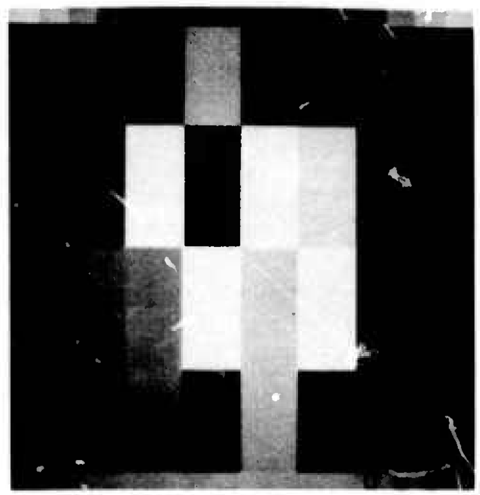
Figure (6.1-1) Comparison of Unconstrained and Equality Constrained Restorations
Gaussian Blur, $b_V = b_H = .70$



Blurred $b_V = b_H = 2.5$, Var = 10
Gaussian Blur, Overdetermined

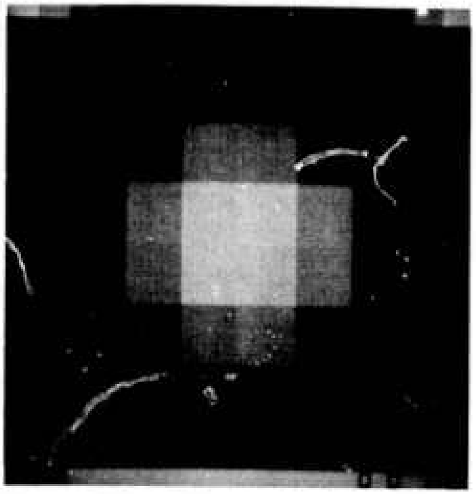


Unconstrained Restoration

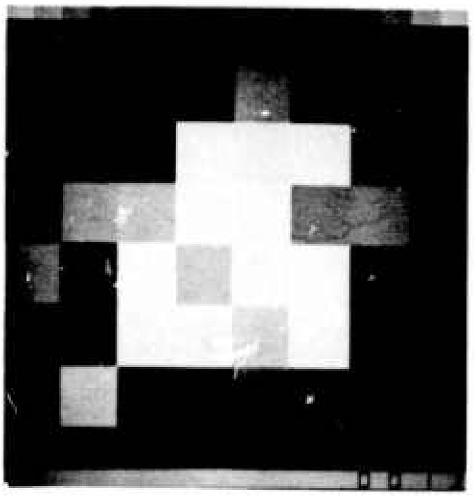


Equality Constrained Restoration

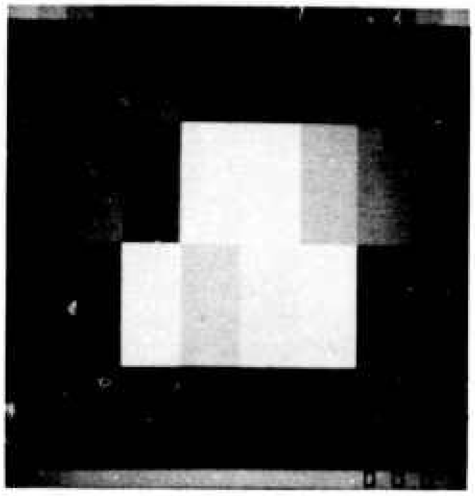
Figure (6.1-2) Comparison of Unconstrained and Equality Constrained Restorations
Gaussian Blur, $b_V = b_H = 2.5$



Blurred $b_V = b_H = 5000$, Var = 10
Gaussian Blur, Overdetermined

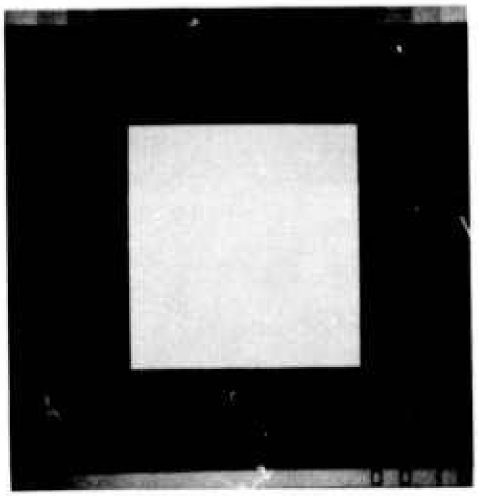


Unconstrained Restoration

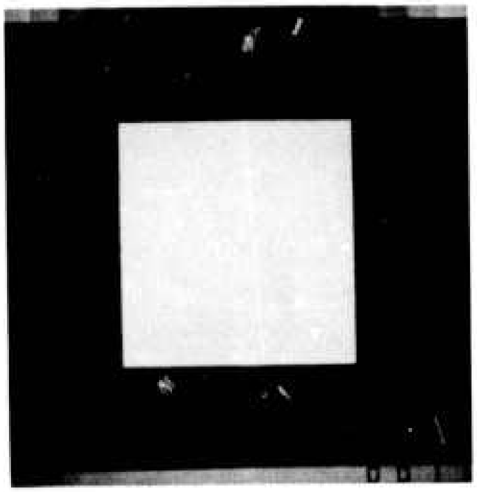


Equality Constrained Restoration

Figure (6. 1-3) Comparison of Unconstrained and Equality Constrained Restorations Gaussian Blur, $b_V = b_H = 5000$



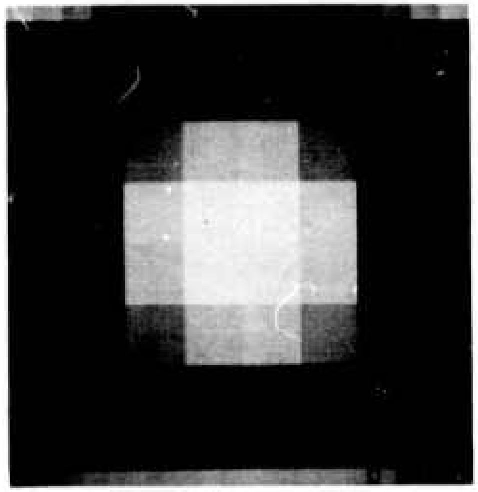
Blurred $b_V = b_H = .25$, Var = 40
Sinc^2 Blur, Overdetermined



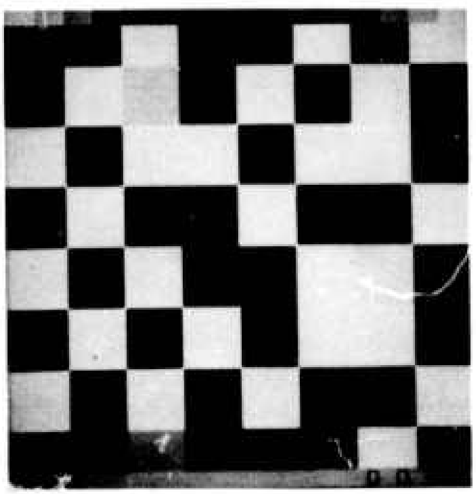
Unconstrained Restoration

Equality Constrained Restoration

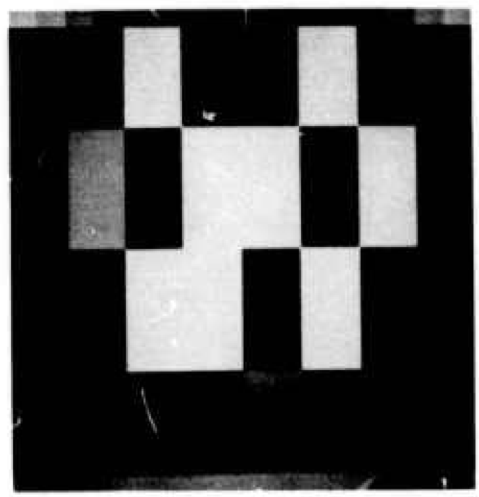
Figure (6.1-4) Comparison of Unconstrained and Equality Constrained Restorations Sinc^2 Blur, $b_V = b_H = .25$



Blurred $b_V = b_H = 1.3$, Var = 40
Sinc^2 Blur, Overdetermined

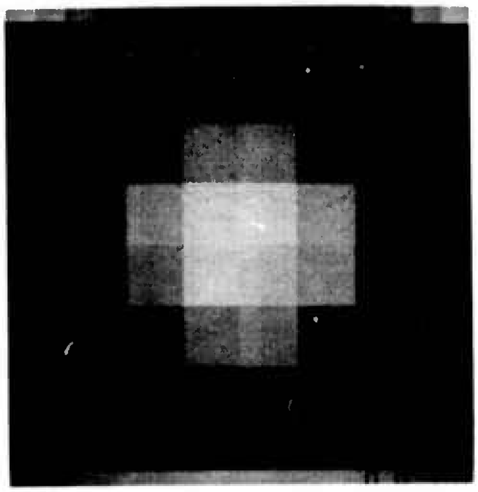


Unconstrained Restoration

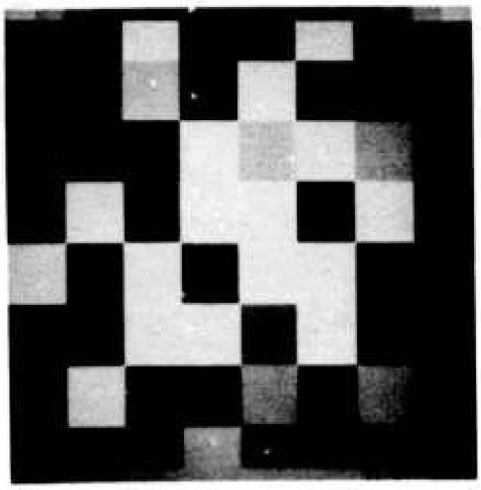


Equality Constrained Restoration

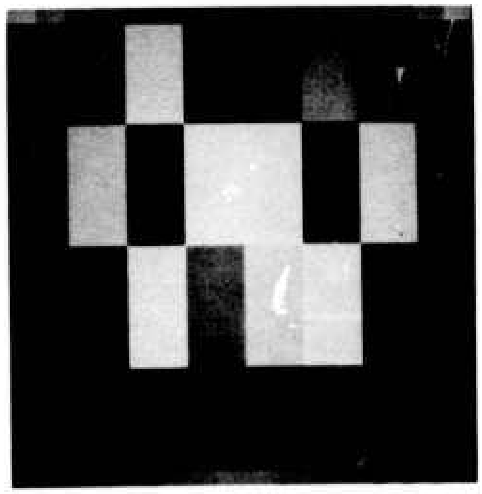
Figure (6. 1-5) Comparison of Unconstrained and Equality Constrained Restorations
Sinc^2 Blur, $b_V = b_H = 1.3$



Blurred $b_V = b_H = 500$, Var = 40
Sinc^2 Blur, Overdetermined



Unconstrained Restoration



Equality Constrained Restoration

Figure (6.1-6) Comparison of Unconstrained and Equality Constrained Restorations
Sinc^2 Blur, $b_V = b_H = 500$

by the use of linear equality constraints involving only a reasonable a priori knowledge of the smoothness of the solution. There is a trade-off between the elimination of the oscillations and the achieved resolution of the picture.

The use of the linear equality constrained method leads to the determination of confidence intervals for parametric functions of pixel values. The unidimensional, overdetermined regression model has been set up, with gaussian shaped blur. The dimensions of the vector of observations and the vector of original pixel values are (12 x 1) and (8 x 1) respectively. The fourth pixel value has been constrained to vary from 0 to 500 and the norm of the residual vector, $\|\underline{y} - \underline{H}\hat{\underline{x}}\|$ is computed. Figure (6.1-7) shows the result for different values of the variance. These curves give examples of the type of result represented in Figure (4.3-3) for the case when no inequality constraint is imposed upon the solution.

Observe that the minimum value of each curve is obtained for the pixel value x_4 that corresponds to the unconstrained solution. It is only the true value (245) for the no noise case. In this case the K-ellipsoids are degenerate so that the parabola also reduces to straight lines.

The confidence interval for the 95% confidence level has been computed using the K value given by the table of normal distribution, as described in chapter 3. It should be remarked that the parabola

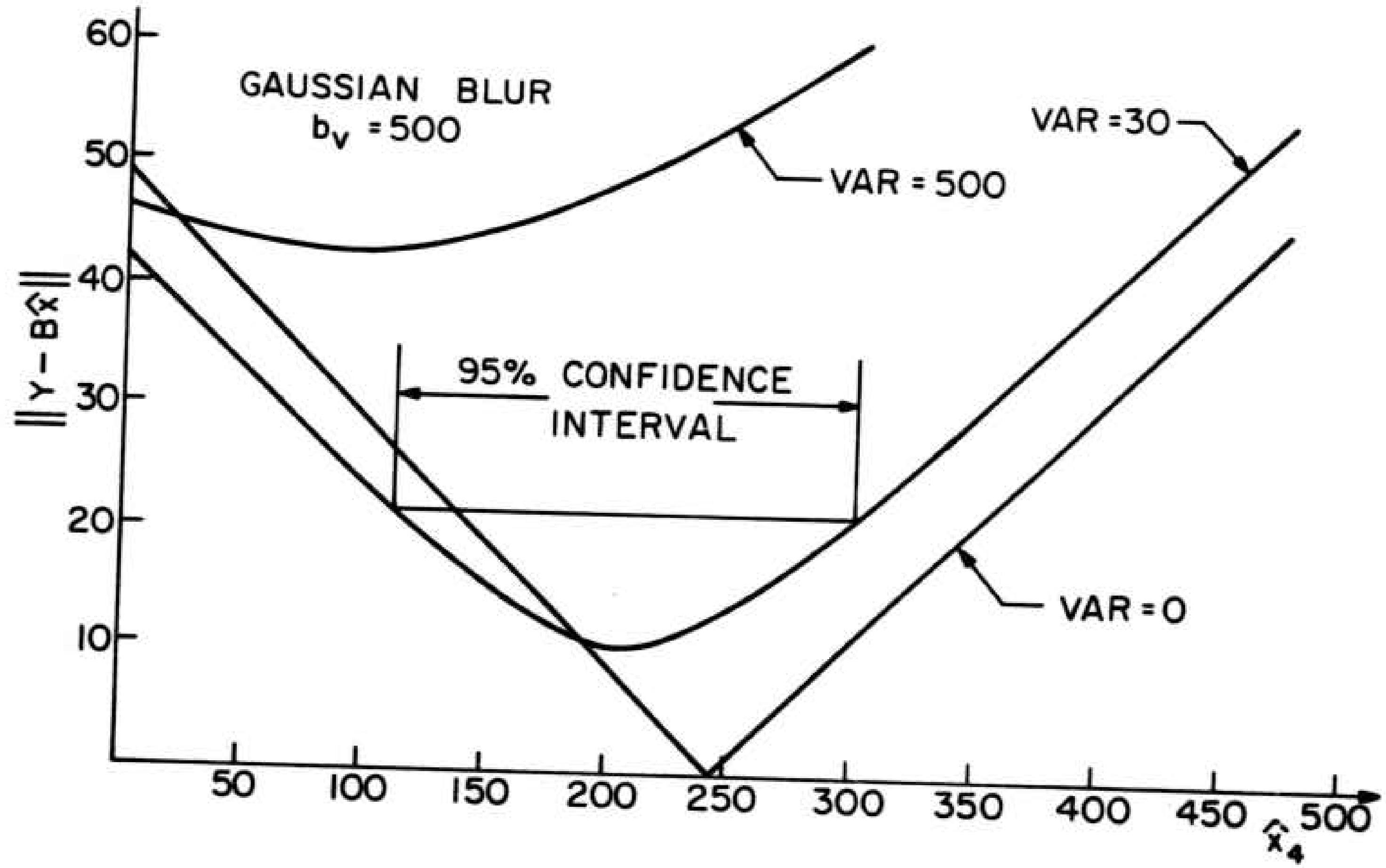
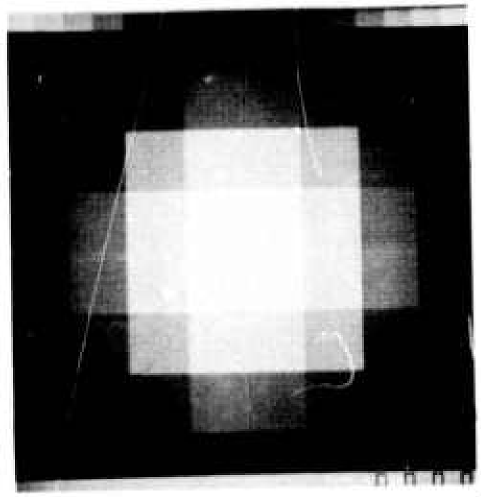


Figure (6.1-7) Computation of Confidence Interval for a Pixel Value for Var = 500 is more open than for Var = 30, indicating that larger confidence intervals are obtained in the former case.

6.2 Inequality Constraints

The experiments with linear inequality constrained restoration involve the solution of the quadratic programming problems discussed in chapter 4. The problems have been solved through the use of the Dantzig's algorithm [6-1]. The lower and upper bounds were set at 0 and 255, respectively. An attempt to solve the two-dimensional problem directly has to face the serious storage requirements of the matrix ($\underline{H}^T\underline{H}$) of the quadratic expression, involving approximately $\frac{1}{2} \times N^4$ elements, where N is the size of one dimension in the original image. Moreover, the attempt revealed numerical convergence problems due to the large amount of computations involved, ev n though double precision was used. Therefore, the two-dimensional problem was solved by a sequence of solutions involving first restorations on the rows and then restorations on the columns. This is clearly an approximate method. The approximation tends to be better when there is little blur. This occurs because on one hand a moderate amount of blur implies large condition number and the input to the column restoration will be clipped. This tends to invalidate the model of additive noise that is the basis of the quadratic programming algorithm. On the other hand, for large blur, with the truncation of the point spread function, the condition number is moderate, but now the random variables in the same row will tend to be correlated and this is not taken into account in the column restoration.

The linear inequality constrained restorations have been performed under gaussian or sinc^2 blur, for different values of condition number. Figures (6.2-1), (6.2-2), and (6.2-3) illustrate the results for gaussian shaped blur, while Figures (6.2-4), (6.2-5), and (6.2-6) refer to the diffraction limited case. The improvement over the unconstrained restoration is clear, particularly with high condition number. Figure (6.2-5) illustrates an example of a completely unfeasible restoration using straightforward regression techniques becoming feasible by the addition of inequality constraints. The improvement by the use of this type of restriction is greater than with the equality constraints, although a much higher computational task has to be performed. The solution of each eight variable quadratic programming problem took between 6 and 7 seconds, completing 10 or 12 iterations of the algorithm. However, most of this time was spent writing and reading from the disk where the data is stored. It is felt that a substantial reduction in time is possible by all-in-core programming. It should be noted that the upper and lower bounds of the quadratic programming, respectively 255 and 0, determine exactly the range of values that are displayed. Therefore, while the uncon-

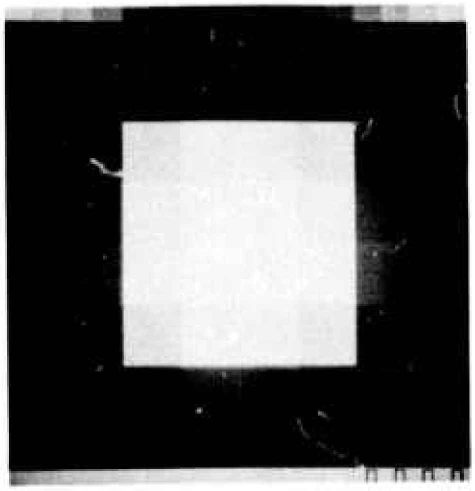


Blurred $b_V = b_H = .70$, Var = 10
Gaussian Blur, Overdetermined

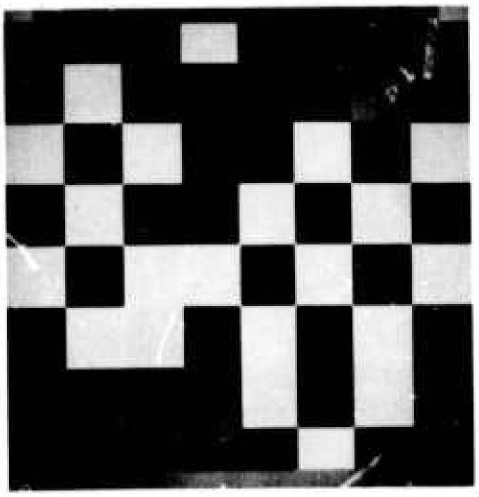
Unconstrained Restoration

Inequality Constrained
Restoration
$0 \leq x \leq 255$

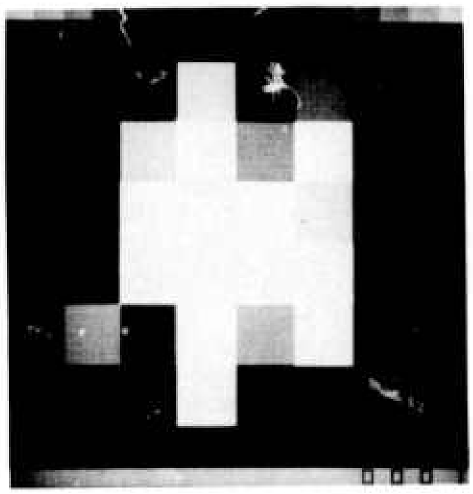
Figure (6.2-1) Comparison of Unconstrained and Inequality Constrained Restorations Gaussian Blur, $b_V = b_H = .70$



Blurred $b_V = b_H = 2.5$, Var = 10
Gaussian Blur, Overdetermined

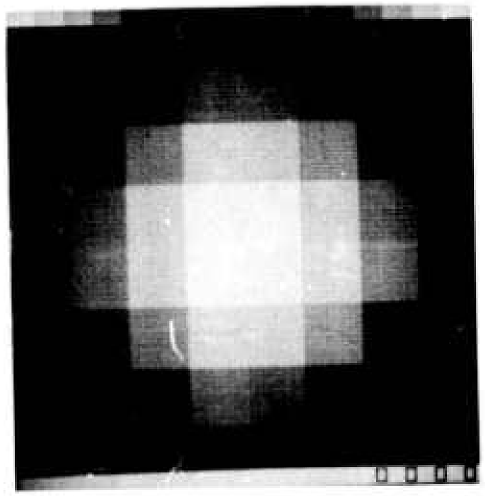


Unconstrained Restoration

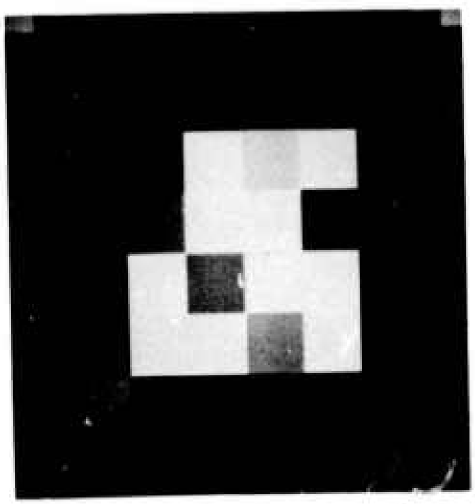


Inequality Constrained
Restoration
$0 \leq x \leq 255$

Figure (6.2-2) Comparison of Unconstrained and Inequality Constrained Restorations Gaussian Blur, $b_V = b_H = 2.5$



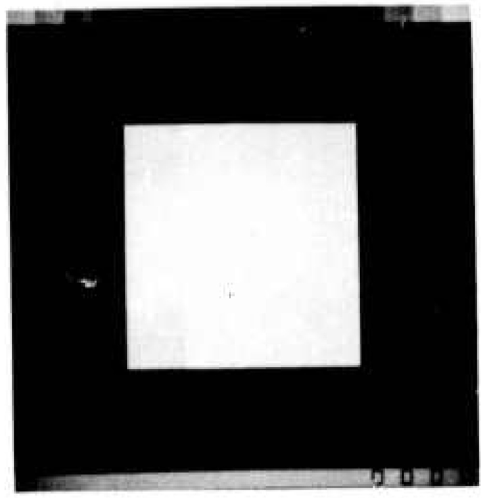
Blurred $b_V = b_H = 5000$, Var = 10
Gaussian Blur, Overdetermined



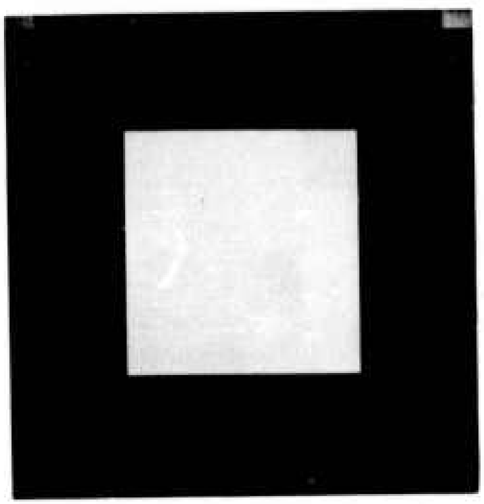
Unconstrained Restoration

Inequality Constrained
Restoration
$0 \leq x \leq 255$

Figure (6.2-3) Comparison of Unconstrained and Inequality Constrained Restorations Gaussian Blur, $b_V = b_H = 5000$



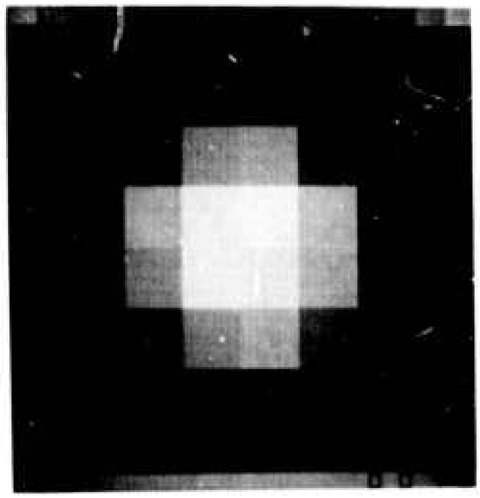
Blurred $b_V = b_H = .25$, Var = 40
Sinc^2 Blur, Overdetermined



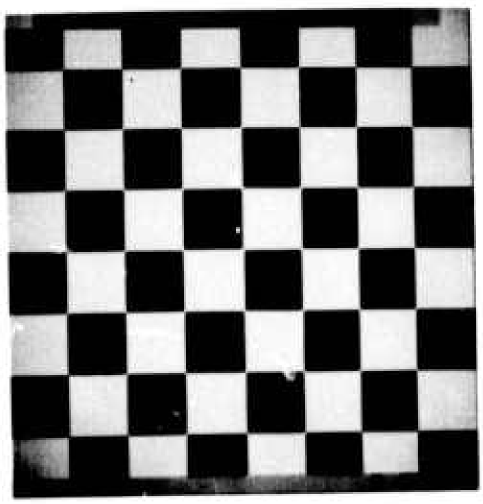
Unconstrained Restoration

Inequality Constrained
Restoration
$0 \leq x \leq 255$

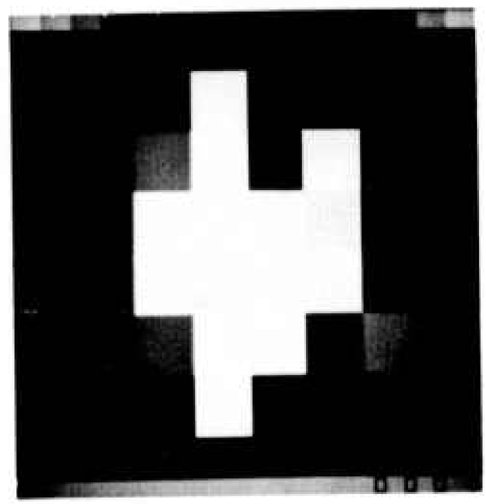
Figure (6.2-4) Comparison of Unconstrained
and Inequality Constrained Restorations
Sinc^2 Blur, $b_V = b_H = .25$



Blurred $b_V = b_H = 1.0$, Var = 40
$Sinc^2$ Blur, Overdetermined

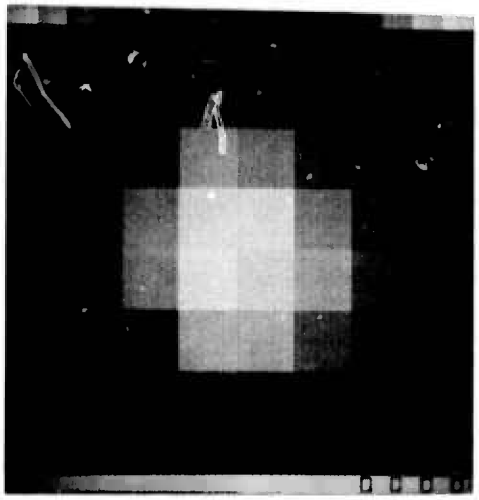


Unconstrained Restoration

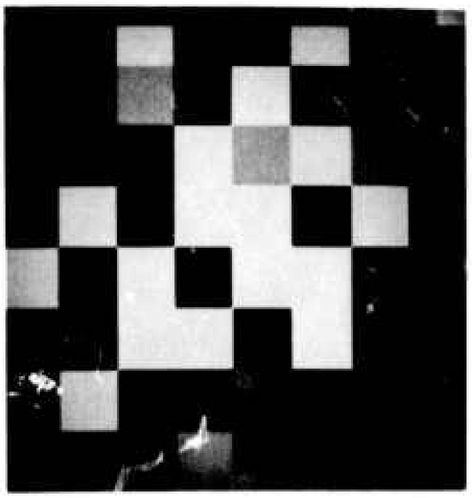


Inequality Constrained
Restoration
$0 \leq x \leq 255$

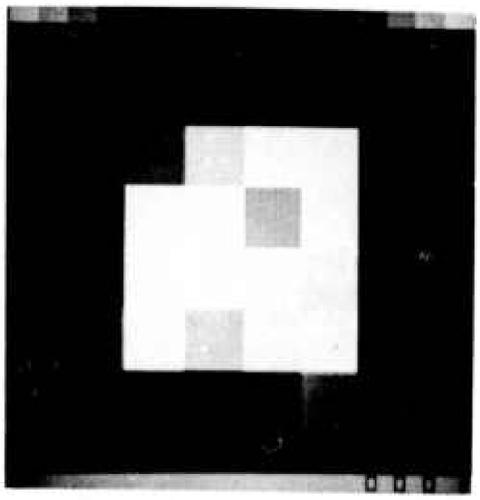
Figure (6.2-5) Comparison of Unconstrained
and Inequality Constrained Restorations
$Sinc^2$ Blur, $b_V = b_H = 1.0$



Blurred $b_V = b_H = 500$, Var = 40
Sinc^2 Blur, Overdetermined



Unconstrained Restoration



Inequality Constrained
Restoration
$0 \le x \le 255$

Figure (6.2-6) Comparison of Unconstrained and Inequality Constrained Restorations
Sinc^2 Blur, $b_V = b_H = 500$

strained or equality constrained results are clipped to these bounds, sometimes from much higher or lower values, the displayed inequality constrained restored pictures reflect precisely the numerical results.

The determination of approximate confidence intervals for parametric functions of pixel values is also done. Figure (6.2-7) illustrates the comparison with the unconstrained restoration. Observe that both confidence intervals, unconstrained and constrained are pessimistic. In fact, these intervals were computed using the γ value which is approximately 1.89 times greater than the K-value used in Figure (6.1-7) for the unconstrained case. This makes the unconstrained confidence interval larger than in that figure, in order to be compared with the constrained interval. The last interval is also pessimistic for the reasons stated in chapter 4.

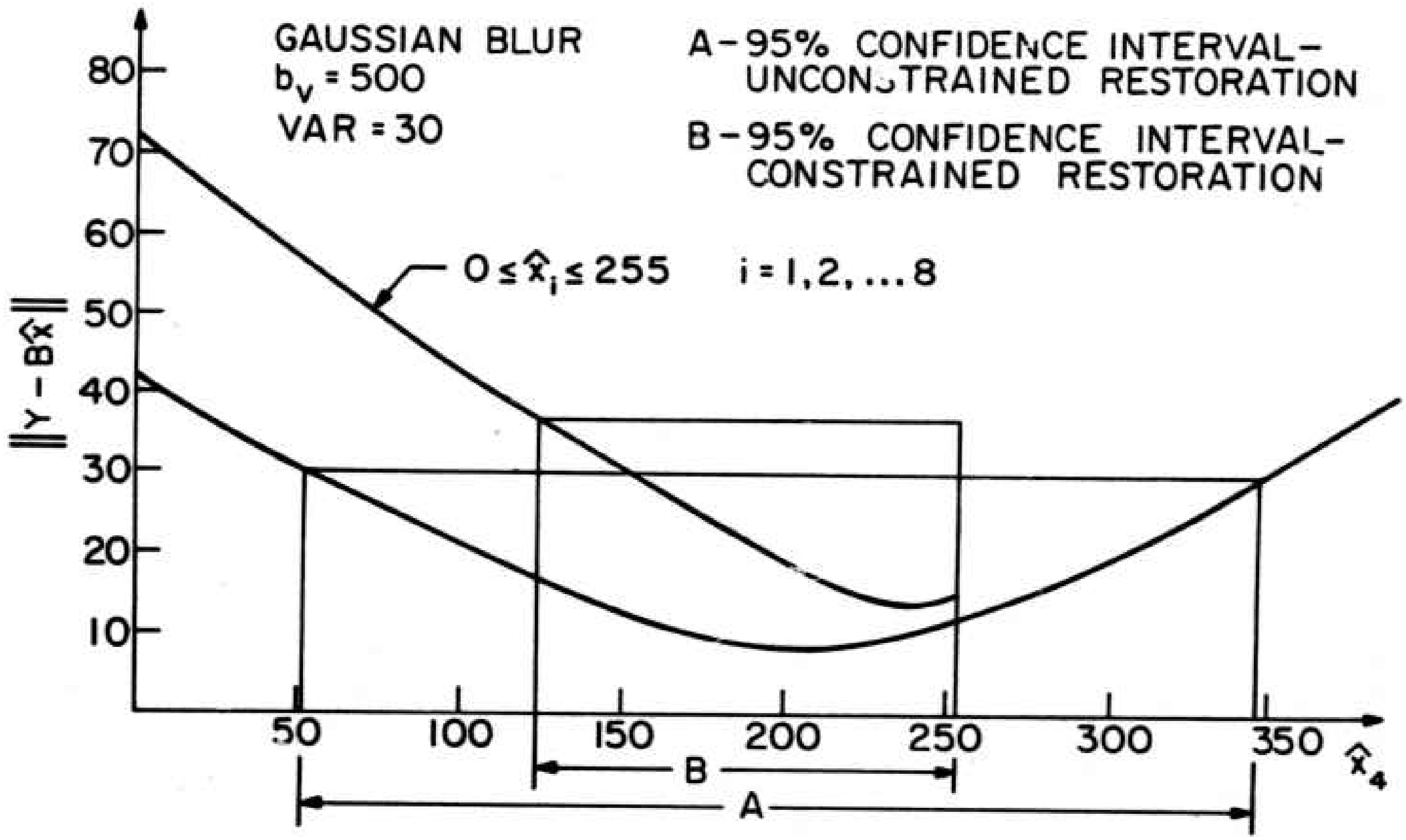


Figure (6.2-7) Comparison Between Unconstrained and Inequality Constrained Approximate Confidence Intervals

7. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

In this dissertation an attempt has been made to put the problem of image restoration on a firm theoretical basis. The linear integral equation that describes the imaging formation process has been discretized in order to adapt the method to the requirements of digital computation. This leads to a regression model whose numerical and statistical properties have been extensively studied in this work. Through the use of this regression model it has been possible to incorporate the large body of knowledge developed in mathematical statistics, econometrics, optimization theory and numerical analysis into the field of image restoration.

The first developed method consists of the solution of the least squares problem by the set of normal equations. This solution assumes the minimum possible amount of a priori information about the image to be restored. In the regression model the vector of pixel values representing the image is simply a set of parameters to be determined. A price is paid, of course, for this lack of knowledge. In the case of the overdetermined model the restoraion problem can become extremely ill conditioned and the solution may exhibit wildly oscillating behavior. The study of the variation of the condition number as a function of the blur coefficient has been made, for two types of kernel, namely, atmospheric turbulence and diffraction limitation. The effect of truncation of the kernel has been shown to be one of reduction of condition number. A trade-off is developed between the uncertainty in the estimators, when the discrete model approximates the continuous one, and the error in modeling when the opposite approach is taken. The use of the underdetermined model, representing an attempt by the experimenter to estimate more points than observed values implies infinite variances of the estimator. When a pseudoinverse solution is obtained this unbounded uncertainty is traded for an unknown bias in the estimators.

It becomes clear that, in order to obtain meaningful solutions, some sort of a priori knowledge has to be introduced. Most of the prior work in image restoration has employed probabilistic a priori information, as a Bayesian type of approach to the problem. This dissertation has explored deterministic a priori constraints, of two types: linear equality and linear inequality restrictions. A comparison with existing methods of solution has been made. In particular, it has been shown that the smoothing and regularization techniques for solving the discrete version of the Fredholm integral equation of the first kind can be interpreted as least squares with quadratic equality constraints.

The use of linear equality constraints has shown the advantage of being able to formulate the restrictions in terms of a hypothesis that could be tested, revealing to the experimenter the validity or not of these restrictions. The test could be performed by verifying the restrictions themselves (F-test) or testing whether there is an improvement in mean square error by the use of these (perhaps incorrect) constraints (Toro Vizcarrondo-Wallace test). A moderate visual improvement has been observed by the use of linear equality constraints imposing that pairs of adjacent pixels should be equal. Nevertheless, for high condition number, large oscillations are still present in the solution.

When inequality constraints are applied by solving the quadratic programming problem, better results are obtained, damping the oscillations even in the high condition number situation. This is done at the price of increased computational requirements, in terms of both storage and time. It is felt that restoration of large images by quadratic programming could only be done in blocks, using an all-in-core algorithm. Its use could be justified when large condition numbers are involved in the restoration of valuable imagery.

In the analysis of the effect of perturbations on the solution of the least squares problem it has been pointed out that the condition number for changes in the blur matrix tends to increase as the square power when noise is present. As a consequence, extreme care must be taken in the measurement of the blur function, particularly when a high condition number is involved.

The research pursued in this dissertation may be extended in several directions. A more detailed study of the interplay of the different sources of error in the estimation of the continuous image, when a discrete model is used, would be of considerable interest. The choice of the number and location of nodes of quadrature and sampling points affects the error of numerical quadrature, the modeling error due to the truncation of the point spread function, the variance of the estimators and the aliasing error. The problem is of considerable complexity and one should be probably satisfied with a suboptimal solution. The study of the condition number for different types of quadrature formulas and different kernels would provide valuable information.

Another area that deserves some exploration consists of modeling the lack of knowledge of the blur matrix by a probabilistic description, leading to the use of stochastic regressors. This will be particularly valuable in the restoration of images taken through turbulent atmosphere, when the duration of the exposure is not long enough so that the blur function stabilizes. This work would complement the study by Slepian [1-9], by considering for example, statistical dependence between noise and the point spread function.

The use of recursive computational schemes of the regression estimators would be particularly useful in real time applications, when the stream of data comes from a row or column scanning of the image. Another possibility would be the exploration of iterative methods to solve the least squares problem, in the unconstrained or equality constrained constrained cases. This method is equivalent to solving the linear system of normal equations in an iterative manner. Finally, the area of nonlinear regression could provide an adequate framework to deal with a situation in which the blur cannot be modeled by a linear operation.

APPENDIX A

HYPOTHESIS TESTING IN THE OVERDETERMINED MODEL

One of the advantages of the model developed for the image restoration problem is its feature of being mathematically very tractable. In particular, it offers the framework of extensive hypothesis testing that seems well suited to solve several image detection problems. In chapter 4 the linear equality constrained restoration combines estimation and hypothesis testing in a common framework of image restoration.

The assumption of an overdetermined model ($M > N$) will be made, together with the hypothesis of white Gaussian noise corrupting the image. Nevertheless, some recent work in the statistical literature seems to indicate that some of these tests are robust, in the sense that they seem to perform well even when the Gaussian assumption is violated [3-6, page 615].

Suppose first that a linear functional of the pixel values is to be tested. In this case the experimenter could be interested, for example, in testing whether a single pixel is equal to a certain value or not or whether the sum of all pixel values, representing an integral over the entire picture, is also equal to a prespecified value or not equal.

Assume that the linear functional is represented by the inner product $\underline{c}^T \underline{x}$ and that the hypothesized value (represented by the hypothesis H_o) is c_o. Under the hypothesis H_o, the random variable

$$\underline{c}^T \underline{x} - c_o \tag{A-1}$$

is zero mean and Gaussian. Its variance is given by

$$\sigma^2(\underline{c}^T(\underline{H}^T\underline{H})^{-1}\underline{c}) \tag{A-2}$$

Thus, under H_o,

$$\frac{\underline{c}^T \underline{x} - c_o}{\sigma\sqrt{\underline{c}^T(\underline{H}^T\underline{H})^{-1}\underline{c}}} \tag{A-3}$$

is a zero mean, unit variance Gaussian random variable. In order to perform the test under a given level of significance, a table of standardized normal distribution should be used. The threshold $a_\alpha > 0$ is chosen such that

$$P\{N(0, 1) > a_\alpha\} = P\{-N(0, 1) < -a_\alpha\} = \frac{\alpha}{2} \tag{A-4}$$

where α is the level of significance of the test also known as the probability of false alarm, using detection theory terminology. If the random variable given by (A-3) falls between $-a_\alpha$ and a_α, Ho is accepted. Otherwise H_o is rejected.

It should be noted that H_o is a simple hypothesis and the alternative H_1 is a composite hypothesis. It is well known that there is no uniformly most powerful test involving a Gaussian random variable in this situation. The test specified in the previous paragraph can be shown to be the uniformly most powerful unbiased test for the simple hypothesis H_o against the alternative composite hypothesis H_1.

Assume now that the variance of the noise is not known. The procedure in this case would be equivalent to performing an estimation of the variance prior to testing. In the section on confidence intervals it was shown that the variance of the parametric function could be estimated by

$$s^2(\underline{c}^T(\underline{H}^T\underline{H})^{-1}\underline{c}) \qquad \text{(A-5)}$$

where

$$s^2 = \frac{1}{M-N}(\underline{y}-\underline{H}\hat{\underline{x}})^T(\underline{y}-\underline{H}\hat{\underline{x}})$$

As a result, the statistic

$$\frac{\underline{c}^T \underline{x} - c_o}{s\sqrt{\underline{c}^T(\underline{H}^T\underline{H})^{-1}\underline{c}}} \qquad \text{(A-7)}$$

is Student distributed with (M-N) degrees of freedom. The test $H_o: \underline{c}^T\underline{x} = c_o$ against $H_1: \underline{c}^T\underline{x} \neq c_o$ can then be performed by the use of a Student distribution table. It should be remarked that for a sample size larger than 40, the test can be performed by using tables of normal distribution. The procedure is entirely analogous to the case of known variance.

Suppose now that the problem consists in testing the unknown variance of the noise. Let $\sigma^2 = \sigma_o^2$ be the H_o hypothesis and $\sigma^2 \neq \sigma_o^2$ be the alternative hypothesis. In the section concerning confidence intervals it was pointed out that the statistics $\frac{(M-N)s^2}{\sigma^2}$ is chi-square distributed with (M-N) degrees of freedom. Under the H_o hypothesis this quantity becomes $\frac{(M-N)s^2}{\sigma_o^2}$ and the test, for a given level of significance, can be performed using tables of the chi-square distribution. More specifically, using a 100 α percent significance level, the thresholds a_α and b_α of this two-sided test are chosen satisfying the relationships

$$P[\chi^2(M-N) < a_\alpha] = P[\chi^2(M-N) > b_\alpha] = \frac{\alpha}{2} \qquad \text{(A-8)}$$

The hypothesis H_o is accepted if the value of $\frac{(M-N)s^2}{\sigma_o^2}$ falls between a_α and b_α. Otherwise the hypothesis is rejected.

Consider now the problem of testing the whole vector $\underline{x}$ of pixel values. In order to simplify the discussion, it will be assumed initially that the H_o hypothesis specifies the zero vector, denoted by $\underline{0}$. The purpose of the test is to verify whether the whole picture is zero against the alternative that it is nonzero.

The following discussion will make use of three quantities, denoted by Q_1, Q_2, and Q_3, that will be defined now. Q_1 will measure the variation of the observed data around the hypothesized regression line, specified by the hypothesis H_o

$$Q_1 = [\underline{y} - \underline{H}\,\underline{0}]^T [\underline{y} - \underline{H}\,\underline{0}] = \underline{n}^T \underline{n} \qquad \text{(A-9)}$$

Q_2 will measure the variation of the observed data around the regression line obtained by estimating the regression coefficients irrespective of whether or not H_o is true.

$$Q_2 = [\underline{y} - \underline{H}\,\hat{\underline{x}}]^T [\underline{y} - \underline{H}\,\hat{\underline{x}}] \qquad \text{(A-10)}$$

Using the notation defined previously, it follows that

$$Q_2 = \hat{\underline{u}}^T \hat{\underline{u}} = \underline{yL}^T \underline{Ly} = \underline{yLy} = \underline{n}^T \underline{Ln} \qquad \text{(A-11)}$$

Q_3 will denote the variation of the estimated regression line around the hypothesized regression line, that means

$$\begin{aligned} Q_3 &= [\underline{H}\hat{\underline{x}} - \underline{H}\,\underline{o}]^T[\underline{H}\hat{\underline{x}} - \underline{H}\,\underline{o}] \\ &= (\hat{\underline{x}} - \underline{o})^T(\underline{H}^T\underline{H})(\hat{\underline{x}} - \underline{o}) \qquad \text{(A-12)} \\ &= \hat{\underline{x}}^T(\underline{H}^T\underline{H})\hat{\underline{x}} \end{aligned}$$

Now since

$$\begin{aligned} \hat{\underline{x}} &= (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{y} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T(\underline{H}\,\underline{x} + \underline{n}) = \\ &\quad \underline{x} + (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{n} \qquad \text{(A-13)} \end{aligned}$$

for $\underline{x} = \underline{0}$ it follows that

$$\hat{\underline{x}} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{n} \qquad \text{(A-14)}$$

So that (A-12) can be put into the form

$$\begin{aligned} Q_3 &= \underline{n}^T\underline{H}(\underline{H}^T\underline{H})^{-1}(\underline{H}^T\underline{H})(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{n} \\ &= \underline{n}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{n} \qquad \text{(A-15)} \\ &= \underline{n}^T(\underline{I} - \underline{L})\underline{n} \end{aligned}$$

Now, by observing expressions (A-9), (A-11) and (A-15) together with the observations that

$$\text{rank } \underline{I} = M \tag{A-16}$$

$$\text{rank } \underline{L} = M-N \tag{A-17}$$

$$\text{rank}(\underline{I}-\underline{L}) = N \tag{A-18}$$

it follows that $\frac{Q_1}{\sigma^2}$ is χ^2 (M), $\frac{Q_2}{\sigma^2}$ is χ^2 (M-N) and $\frac{Q_3}{\sigma^2}$ is χ^2 (N) distributed. Also, it is easy to verify that

$$Q_1 = Q_2 + Q_3 \tag{A-19}$$

Furthermore, by Cochran's Theorem [A-1, pages 212-214], Q_2 and Q_3 are independent random variables.

From the definitions of Q_2 and Q_3 it should be clear that if Q_3 is large compared to Q_2, this would lead to the belief that the hypothesis H_o is not true. In fact, this would mean that the hypothesized regression line is far from the estimated regression line as compared to the variation of the data around the estimated regression line. Under this perspective, the statistics

$$F = \frac{Q_3}{Q_2} \frac{M-N}{N} \tag{A-20}$$

is appropriate for the testing procedure [3-5, page 96]. This ratio is distributed according to the Fischer distribution, with N and M-N degrees of freedom. The null hypothesis H_o is rejected when this ratio exceeds the significance limit.

So far the test has been developed assuming that the null hypothesis specifies the zero vector. When, in general, a nonzero pixel vector $\underline{x}_o$ is specified, Q_3 is given by

$$Q_3 = [\underline{H}\hat{\underline{x}} - \underline{H}\underline{x}_o]^T [\underline{H}\hat{\underline{x}} - \underline{H}\underline{x}_o] \tag{A-21}$$

and still, after some manipulation, the previous expression assumes the form of equation (A-15). Since Q_2 is not affected by the form of the null hypothesis, it remains the same. Therefore, expression (A-20) can still be applied.

In some circumstances the experimenter might be interested in testing only part of the total number of pixels in a picture. This would happen when it is known that a certain object to be detected could occur only at a certain location on the overall picture. In this case, the previous test can be modified to take into account this special feature. By performing, if necessary, a reordering of the pixel values, the vector $\underline{x}$ can be subdivided into two subvectors.

$$\underline{x} = \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} \tag{A-22}$$

Suppose that the subvector of interest is $\underline{x}_2$ and that the H_o hypothesis consists in testing $\underline{x}_2 = \underline{x}_{2o}$ against $\underline{x}_2 \neq \underline{x}_{20}$. Under this partition the linear model is given by

$$\underline{y} = (\underline{H}_1 \underline{H}_2)\begin{pmatrix}\underline{x}_1 \\ \underline{x}_2\end{pmatrix} + \underline{n} \tag{A-23}$$

where $\underline{x}_1$ is an $(N_1 \times 1)$ vector, $\underline{x}_2$ an $(N_2 \times 1)$ vector such that $N_1 + N_2 = N$ and $\underline{H}_1$ and $\underline{H}_2$ are, respectively $(M \times N_1)$ and $(M \times N_2)$ matrices. Under this partition, the estimator will have the form [3-5, page 100]

$$\hat{\underline{x}} = \begin{pmatrix}\hat{\underline{x}}_1 \\ \hat{\underline{x}}_2\end{pmatrix} = [(\underline{H}_1 \underline{H}_2)^T (\underline{H}_1 \underline{H}_2)]^{-1} (\underline{H}_1 \underline{H}_2)^T \underline{y}$$

$$= \begin{pmatrix}\underline{H}_1^T \underline{H}_1 & \underline{H}_1^T \underline{H}_2 \\ \underline{H}_2^T \underline{H}_1 & \underline{H}_2^T \underline{H}_2\end{pmatrix}^{-1} \begin{pmatrix}\underline{H}_1^T \underline{y} \\ \underline{H}_2^T \underline{y}\end{pmatrix}$$

$$= \begin{pmatrix}(\underline{H}_1^T \underline{H}_1)^{-1} \underline{H}_1^T \underline{y} - (\underline{H}_1^T \underline{H}_1)^{-1} \underline{H}_1^T \underline{H}_2 \underline{E}^{-1} \underline{H}_2^T \underline{L}_1 \underline{y} \\ \underline{E}^{-1} \underline{H}_2^T \underline{L}_1 \underline{y}\end{pmatrix} \tag{A-24}$$

where

$$\underline{L}_1 = \underline{I} - \underline{H}_1 (\underline{H}_1^T \underline{H}_1)^{-1} \underline{H}_1^T \tag{A-25}$$

and

$$\underline{E} = \underline{H}_2^T \underline{H}_2 - \underline{H}_2^T \underline{H}_1 (\underline{H}_1^T \underline{H}_1)^{-1} \underline{H}^T \underline{H}_2 = \underline{H}_1^T \underline{L}_1 \underline{H}_2 \tag{A-26}$$

Also, $\hat{\underline{x}}_2$ can be expressed as

$$\underline{\hat{x}}_2 = \underline{x}_2 + \underline{E}^{-1}\underline{H}_2^T\underline{L}_1\underline{n} \tag{A-27}$$

Comparing (A-27) with (A-13) and taking into consideration (A-26), it follows that in this case the quantity Q_3 of interest for testing has the form

$$Q_3 = (\underline{x}_2 - \underline{x}_2^*)\underline{E}(\underline{x}_2 - \underline{x}_2^*) \tag{A-28}$$

where $\underline{x}_2 = \underline{x}_2^*$ is the null hypothesis. Since $\underline{E}$ has the rank N_2, $\frac{Q_3}{\sigma^2}$ has a χ^2 distribution with N_2 degrees of freedom and the ratio

$$F = \frac{Q_3}{Q_2} \cdot \frac{M-N}{N_2} \tag{A-29}$$

will be distributed according to the Fisher distribution with N_2 and (M-N) degrees of freedom. A large value of this ratio leads to the rejection of the null hypothesis.

Several other tests could be devised concerning the pixel values of an image or their linear combinations. For example, suppose that the problem consists in detecting whether there is a difference between the gray level of two parts of a picture. This is the so called edge detection problem. Two versions of this question can be formulated. In the first one, the null hypothesis specifies $\underline{x}_1 = \underline{x}_2$

where $\underline{x}_1$ and $\underline{x}_2$ are partitions of $\underline{x}$ but the common value is unknown. In the second version the common value is known and represented by a vector, say $\underline{x}^*$. In the field of econometrics tests of this type have been called tests for structural stability [3-5, page 103-116].

Before closing this appendix some observations should be made. First, even though no justification was given for the statistics used in the tests, it can be shown [3-6, page 141] that these tests can be rigorously derived through the use of likelihood ratios, using some suitable criterion, like the Neyman-Pearson for example.

Second, the examples of tests of pixel values or their linear combinations can be put into the framework of a general linear hypothesis represented by the expressions

$$H_o : \underline{R}\underline{x} = \underline{r} , H_1 : \underline{R}\underline{x} \neq \underline{r} \quad \text{(A-30)}$$

where $\underline{R}$ and $\underline{r}$ are matrices of dimensions Q x M and Q x 1, respectively, the rank of $\underline{R}$ being Q. A test procedure [3-6, page 143] can be developed for this general case, leading to the previous cases for specific choices of the matrices $\underline{R}$ and $\underline{r}$. Reference [3-3, pages 100-104] presents broader results, encompassing the cases where the rank of $\underline{R}$ may be not Q and the underdetermined model, where the rank of $\underline{H}$ may not be N.

Further tests could be devised, using the large body of material developed so far in regression analysis. As examples, could be mentioned a test to verify whether the white noise corrupting the image has constant variance or the test to verify if any degree of correlation exists between the noise elements corrupting different pixel values. This last procedure would be based on the so called Durbin-Watson test developed in econometrics [3-6, pages 199-201]. Also, tests to verify the hypothesis of linearity of the blurring process could be used, being based on the von Neumann ratio [3-6, pages 222-224].

APPENDIX B

COMPUTATION OF THE NUMBER OF OPERATIONS FOR DIFFERENT METHODS OF RESTORATION

In this appendix, the number of operations involved in the computation of the estimators in the restoration problem is presented. The unconstrained, linear equality and linear inequality constrained methods are considered. It is assumed that the first two methods use nonrecursive forms of the estimators and the overdetermined model is involved.

The unconstrained solution is obtained by a simple matrix-vector multiplication. If the observed vector $\underline{y}$ is $(M^2 \times 1)$ and the vector of pixel values $\underline{x}$ is $(N^2 \times 1)$, the solution would involve $M^2 \times N^2$ multiply and add operations.

Expression (4.2-2) gives the solution for the linear equality constrained restoration. Assuming that P linear constraints are incorporated, the number of operations involved in obtaining the solution can be computed as follows:

$M^2 \times N^2$ mult. and add for $\underline{\hat{x}}$

$N^2 \times P$ mult. and add for $\underline{A}\underline{\hat{x}}$

P add for $\underline{t} - \underline{A}\underline{\hat{x}}$

$N^2 \times P$ mult. and add for $\underline{K}(\underline{t} - \underline{A}\hat{\underline{x}}) = \overline{\underline{x}}$

where $\underline{K} = (\underline{H}^T\underline{C}^{-1}\underline{H})^{-1}\underline{A}^T[\underline{A}(\underline{H}^T\underline{C}^{-1}\underline{H})^{-1}\underline{A}^T]$

N^2 add for $\hat{\underline{x}} + \overline{\underline{x}}$

Total: $N^2(2P + M^2)$ mult. and add + $(P + N^2)$ add

The computation of the number of operations involved in the solution of the quadratic programming problems for inequality constrained restoration is more complex and only an estimated value can be given. Suppose that the problem consists of minimizing the quadratic form given by (4.3-4) subject to the constraints expressed by (4.3-2) and (4.3-3). Wolfe's procedure [4-6] consists essentially in solving the equations that express the Kuhn Tucker conditions.

The procedure consists of two steps: in the first one, three nonnegative slack vectors, $\underline{m}$, $\underline{t}_1$ and $\underline{t}_2$ are introduced and the step consists in finding a set of vectors $\underline{x} \geq \underline{0}$, $\underline{z} \geq \underline{0}$ and $\underline{w}$ and the slack vectors that solve the problem of minimizing the objective function

$$\sum_{i=1}^{S+N^2} m_i \tag{B-1}$$

subject to the constraints

$$\underline{B}\underline{x}^* + \underline{m} = \underline{v} \tag{B-2}$$

$$2\underline{H}^{*T}\underline{V}^{-1}\underline{H}^{*}\underline{x}^{*} - \underline{z} + \underline{B}^{T}\underline{w} - 2\underline{H}^{*T}\underline{V}^{-1}\underline{y} + \underline{t}_1 - \underline{t}_2 = \underline{0} \qquad \text{(B-3)}$$

$$\underline{z}^{T}\underline{x}^{*} = 0 \qquad \text{(B-4)}$$

where the notation defined by equations (4.3-5) to (4.3-13) is used. Except for the last restriction, this is a linear programming problem. The procedure is very much similar to the simplex method, with some modification to take into consideration the nonlinear restriction.

The minimum of (B-1) is attained for $\underline{m} = \underline{0}$ and the second step consists of minimizing the objective function

$$\sum_{i=1}^{2N^2} t_i \qquad \text{(B-5)}$$

subject to the constraints

$$\underline{Bx}^{*} = \underline{v} \qquad \text{(B-6)}$$

$$2\underline{H}^{*T}\underline{V}^{-1}\underline{H}^{*}\underline{x}^{*} - \underline{z} + \underline{B}^{*T}\underline{W} - 2\underline{H}^{*T}\underline{V}^{-1}\underline{y} + \underline{Et} = \underline{0} \qquad \text{(B-7)}$$

$$\underline{z}^{T}\underline{x}^{*} = 0 \qquad \text{(B-8)}$$

where $\underline{t}$ contains the remaining nonzero components of $\underline{t}_1$ and $\underline{t}_2$ and E is the corresponding coefficient matrix.

The number of operations of each step will be computed as if they were standard simplex methods. This is evidently an approximation but it should be a reasonable one since a few extra operations are necessary to impose conditions (B-4) or (B-8).

Under this assumption, the number of operations for each step is estimated [4-5, page 55] as 1.5x (number of linear equality restrictions) pivot operations. Since in each step the number of linear equality restrictions is given by $(S + 3N^2)$, it follows that the total number of pivot operations should be about $3x(S + 3N^2)$.

A pivot operation in linear programming comprises three series of suboperations:

a) Determination of the column of the pivot, that is, the nonbasic variable that will enter the basis. This is done by a search of the most negative reduced cost coefficient among the nonbasic variables.

b) Division of all the elements of this column by the corresponding right hand side elements and search for the smallest nonnegative quocient. This operation will determine the pivot.

c) The operation of pivoting, that is, reducing to one the pivot element and to zero all other elements of the pivot column by elementary row operations and updating the entire simplex tableau.

REFERENCES

[1-1] A.Marechal, P. Croce, and K. Dietzel, "Amelioration du Contraste des Details des Images Photographiques par Filtrage des Frequencies Spatiales," Opt. Acta, vol. 5, pp. 256-262, 1958.

[1-2] R. Natham, "Picture Enhancement for the Moon, Mars and Man," Pictorial Pattern Recognition, G.C. Cheng, R.S. Ledley, D.K. Pollock, and A. Rosenfield, Eds., Washington, D.C., Thompson, 1967.

[1-3] R.H. Selzer, "Recent Progress in Computer Processing of X-Ray and Radioisotope Scanner Images," Biomed. Sci. Instrum., vol. 6, pp. 225-234, 1969.

[1-4] F.P. Asterberg, "Computer Techniques in Geology," Earth-Science Review, vol. 3, pp. 47-77, March 1967.

[1-5] B.R. Hunt, D.H. Janney, and R.K. Zeigler, Introduction to the Restoration and Enhancement of Radiographic Images, Los Alamos Scientific Laboratory, Los Alamos, New Mexico, April 1970.

[1-6] J. Tsujinchi, "Correction of Optical Images by Compensation of Aberrations and by Spatial Frequency Filtering," Progress in Optics, E. Wolf, Ed., Amsterdam, North-Holland, 1963.

[1-7] J.L. Harris, "Image Evaluation and Restoration," J. Opt. Soc. Amer., vol. 56, May 1966, pp. 569-579.

[1-8] C.W. Helstrom, "Image Restoration by the Method of Least Squares," J. Opt. Soc. Amer., vol. 57, pp. 297-303, March 1967.

[1-9] D. Slepian, "Linear Least Squares Filtering of Distorted Images," J. Opt. Soc. Amer., vol. 57, July 1967, pp. 918-922.

[1-10] J.L. Horner, "Optical Spatial Filtering with the Least Mean Square Error Filter," J. Opt. Soc. Amer., vol. 59, May 1969, pp. 553-558.

[1-11] J.L. Horner, "Optical Restoration of Images Blurred by Atmospheric Turbulence using Optimum Filter Theory," Appl. Opt., vol. 9, Jan. 1970, pp. 167-171.

[1-12] M.P. Ekstrom, Numerical Restoration of Random Images, Ph.D. Dissertation, University of California, Davis, 1971.

[1-13] W.K. Pratt, "Generalized Wiener Filter Computation Techniques," IEEE Transactions on Computers, vol. C-21, July 1972, pp. 636-641.

[1-14] N.D.A. Mascarenhas, "Causal Generalized Wiener Filters for Image Noise Reduction," presented at the Symposium on Image Processing and Recognition, University of Missouri - Columbia, August 1972.

[1-15] R.P. Boas and M. Kac, "Inequalities for Fourier Transforms of Positive Functions," Duke Math. J., vol. 12, pp. 189-206, 1945.

[1-16] W. Luckosz, "Transfer of Nonnegative Signals through Linear Filters," Optica Acta., vol. 9, pp. 335-369, 1962.

[1-17] F.H. Cleveland and A.C. Schell, "The Enhancement of Antenna Images," AFCRL-69-0345, Phys. and Math. Sc. Res. Paper No. 391, August 1969.

[1-18] J. Philip, "Reconstruction from Measurements of Positive Quantities by the Maximum-Likelihood Method," J. Math. Analysis and Appl., vol. 7, no. 3, pp. 327-327, 1963.

[1-19] A.V. Oppenheim, M.W. Schafer and T.G. Stockham, Jr., "Non-linear Filtering of Multiplied and Convolved Signals," Proc. IEEE, vol. 56, pp. 1264-1291, August 1968.

[1-20] B.R. Frieden, "Restoring with Maximum Likelihood and Maximum Entropy," J. Opt. Soc. Amer., vol. 62, No. 4, pp. 511-518, April 1972.

[1-21] R.S. Hershel, "Unified Approach to Restoring Degraded Images in the Presence of Noise," Optical Sciences Center, University of Arizona Technical Report No. 72, December 1971.

[1-22] R.S. Hershel, "Positive Restoration by Deconvolution," Semiannual Technical Report of the Image Processing Institute, USCEE Report 459, University of Southern California, March - August 1973, pp. 64-66.

[1-23] P.A. Jansson, R.H. Hunt and E.K. Plyler, "Reduction Enhancement of Spectra," J. Opt. Soc. Amer., vol. 60, No. 5, pp. 596-599, May 1970.

[1-24] H.C. Andrews, "Positive Digital Image Restoration Techniques -- A Survey," Report No. ATR-73 (8139) - 2, Aerospace Corp., 1973.

[2-1] J.W. Goodman, Introduction to Fourier Optics, McGraw-Hill, New York, 1968.

[2-2] R.E. Hufnagel and N.R. Stanley, "Modulation Transfer Function Associated with Image Transmission through Turbulent Media," J. Opt. Soc. Amer., vol. 54, pp. 52-61, January 1964.

[2-3] M.M. Sondhi, "Image Restoration: The Removal of Spatially Invariant Degradations," Proceedings of the IEEE, vol. 60, No. 7, July 1972.

[2-4] T.S. Huang, "Some Notes on Film Grain Noise," Woods Hole Summer Study Rep. on Restoration of Atmospherically Degraded Images, vol. 2, Alexandria, Va., Defense Documentation Center, July 1966, pp. 105-109.

[2-5] R. Nathan, "Picture Enhancement for the Moon, Mars, and Man," Pictorial Pattern Recognition, Washington, D.C., Thompson, 1967.

[2-6] R.A. Jones, "An Automated Technique for Deriving MTF's from Edge Traces," Photogr. Sci. Engr., vol. 11, March - April 1967, pp. 102-106.

[2-7] R.A. Jones, "Accuracy Test Procedure for Image Evaluation Techniques," Appl. Opt., vol. 7, January 1968, pp. 133-136.

[2-8] R.A. Jones and E.C. Yeadon, "Determination of the Spread Function from Noisy Edge Scans," Photogr. Sci.

Engr., vol. 13, July - August 1969, pp. 200-204.

[2-9] T.N. Cornsweet, Visual Perception, Academic Press, New York, 1970.

[2-10] R.N. Bracewell and J.A. Roberts, "Aerial Smoothing in Radio Astronomy," Australian J. Phys., 7, 1954, pp. 615-640.

[2-11] F.D. Kahn, "The Correction of Observational Data for Instrumental Bandwidth," Proc. Cambridge Phil. Soc., 51, 1955, pp. 519-525.

[2-12] C. Eckart, "The Correction of Continuous Spectra for the Finite Resolution of the Spectrometer," Phys. Rev., 51, 1937, pp. 735-738.

[2-13] C.W. Barnes, "Object Restoration in a Diffraction-Limited Imaging System," J. Opt. Soc. Amer., 56, 1966, pp. 575-578.

[2-14] D. Slepian and T.T. Kadota, "Four Integral Equations of Detection Theory," Siam. J. Appl. Math, 17, 1969, pp. 1102-1117.

[2-15] R.O. Lane, N.F. Morehouse, and D.L. Phillips, "The Correction of Resonance Beams for Experimental Resolution," Nucl. Instr. Methods, 9, 1960, pp. 87-91.

[2-16] W.K. Pratt, "Vector Space Formulation of Two Dimensional Signal Processing Operations," Internal Report Electrical Engineering Dept., University of Southern California, February 1973.

[2-17] E.K. Blum, Numerical Analysis and Computation: Theory and Practice, Addison Wesley, Reading, Mass., 1972.

[2-18] D.L. Phillips, "A Technique for the Numerical Solution of Certain Integral Equations of the First Kind," J. Assoc. Comput. Mach. 9, 1964, pp. 84-97.

[2-19] S. Twomey, "On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature," J. Assoc. Comput. Mach., vol. 10, 1963, pp. 97-101.

[2-20] S. Twomey, "The Application of Numerical Filtering to the Solution of Integral Equations Encountered in Indirect Sensing Measurements," J. Franklin Inst., vol. 279, pp. 95-109, February 1965.

[2-21] A.E. Hoerl, and R.W. Kennard, "Ridge Regression, Biased Estimation for Nonorthogonal Problems," Technometrics, vol. 12, 1970, pp. 55-67.

[2-22] A.E. Hoerl, and R.W. Kennard, "Ridge Regression Applications to Nonorthogonal Problems," Technometrics, vol. 12. 1970, pp. 69-82.

[2-23] D.W. Marquardt, "Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation," Technometrics, vol. 12, 1970, pp. 591-612.

[2-24] A.N. Tichonov, "Solution of Incorrectly Formulated Problems and the Regularization Method," Soviet Math. Dokl., vol. 4, 1963, pp. 1035-1038.

[2-25] A.N. Tichonov, "Regularization of Incorrectly Posed Problems," Soviet Math. Dolk., vol. 4, 1963, pp. 1624-1627.

[2-26] A.N. Tichonov and V.B. Glasko, "The Approximate Solution of Fredholm Integral Equations of the First Kind," USSR Comput. Math. and Math. Phys., vol. 4(3), 1964, pp. 236-247.

[2-27] M.P. Ekstrom, "Computational Aspects of Solving Planar Integral Equations by Regularization," presented at the Sixth Asilomar Conference on Circuits and Systems, Pacific Grove, California, November 1972.

[2-28] H.J. Longley, "Numerical Solutions and Applications of the Fold Integral," Los Alamos Scientific Laboratory, Rept. LA02729, 1962.

[2-29] R.J. Hanson, "A Numerical Method for Solving Fredhold Integral Equations of the First Kind Using Singular Values," Siam J. Numer. Anal., vol. 8, pp. 616-622, November 1971.

[2-30] B. Backus, and F. Gilbert, "Uniqueness in the Inversion of Inaccurate Gross Earth Data," Phil. Trans. Roy. Soc., London, vol. 266, ser. A, pp. 123-192, March 1970.

[2-31] J.M. Varah, "On the Numerical Solution of Ill Conditioned Linear Systems with Applications to Ill-Posed Problems," Siam J. Numer. Anal., vol. 10, No. 10, April 1973, pp. 257-267.

[2-32] M.P. Ekstrom, and R.L. Rhoads, "On the Application of Eigenvector Expansions to Numerical Deconvolution," Internal Report #74100, Lawrence Livermore Laboratory, September 1972.

[2-33] R. Bellman, R. Kalaba, and J. Lockett, "Dynamic Programming and Ill-Conditioned Linear Systems II," J. Math Anal. Appl., vol. 12, 1965, pp. 393-400.

[2-34] C.T.H. Baker, L. Fox, D.F. Mayers, and K. Wright, "Numerical Solution of Fredholm Integral Equations of First Kind," Computer J., vol. 7, 1965, pp. 141-148.

[2-35] O.N. Strand and E.R. Westwater, "Statistical Estimation of the Numerical Solution of a Fredholm Integral Equation of the First Kind," J. of A.C.M., vol. 15, No. 1, January 1968, pp. 100-114.

[2-36] M.P. Ekstrom, "On the Numerical Feasibility of Digital Image Restoration," Proceedings of the IEEE, 61, 1973, No. 8, pp. 1155-1156.

[2-37] B.R. Hunt, "The Inverse Problem of Radiography," Mathematical Biosciences, vol. 8, 1970, pp. 161-179.

[2-38] B.R. Hunt, "Biased Estimation for Nonparametric Identification of Linear Systems," Mathematical Biosciences, vol. 10, 1971, pp. 215-237.

[2-39] B.R. Hunt, "Deconvolution of Linear Systems by Constrained Regression and its Relationship to the Wiener Theory," IEEE Trans. on Automatic Control, vol. AC-17, No. 5, October 1972, pp. 703-705.

[2-40] B.R. Hunt, "The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer," IEEE Trans. on Computers, vol. C-22, No. 9, September 1973, pp. 805-812.

[3-1] T.O. Lewis, and P.L. Odell, Estimation in Linear Models, Prentice-Hall, New Jersey, 1971.

[3-2] B.W. Rust, and W.R. Burrus, Mathematical Programming and the Numerical Solution of Linear Equations, American Elsevier, New York, 1972.

[3-3] A. Albert, Regression and the Moore-Penrose Pseudoinverse, Academic Press, New York, 1972.

[3-4] N.E. Nahi, Estimation Theory and Applications, Wiley, New York, 1969.

[3-5] D.S. Huang, Regression and Econometric Methods, Wiley, New York, 1970.

[3-6] H. Theil, Principles of Econometrics, Wiley, New York, 1971.

[3-7] H. Scheffe, The Analysis of Variance, Wiley, New York, 1959.

[3-8] G.W. Stewart, Introduction to Matrix Computations, Academic Press, New York, 1973.

[3-9] G.W. Stewart, "On the Continuity of the Generalized Inverse," Siam J. Appl. Math., vol. 17, pp. 33-45.

[4-1] H. Theil, Economic Forecast and Policy, North-Holland, Amsterdan, 1969.

[4-2] C. Toro-Vizcarrando and T.D. Wallace, "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression," J. of the Amer. Stat. Assoc., vol. 63, June 1968, pp. 558-572.

[4-3] H. Theil and A.J. Goldberger, "On Pure and Mixed Statistical Estimation in Economics," International Economic Review, vol. 2, 1961, pp. 65-78.

[4-4] H. Theil, "On the Use of Incomplete Prior Information in Regression Analysis," J. of the Amer. Stat. Assoc., vol. 58, 1963.

[4-5] D.G. Luenberger, Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading, Mass., 1973.

[4-6] P. Wolfe, "The Simplex Method for Quadratic Programming," Econometrica, vol. 27, No. 3, July 1959.

[4-7] G. B. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, New Jersey, 1963.

[4-8] J.C.G. Boot, Quadratic Programming, North Holland, Amsterdam, 1964.

[4-9] G.H. Golub and M.A. Saunders, "Linear Least Squares and Quadratic Programming," Integer and Nonlinear Programming, edited by J. Abadie, North Holland, Amsterdam, 1970, pp. 229-256.

[4-10] J. Stoer, "On the Numerical Solution of Constrained Least Squares Problems," Siam J. of Numer. Anal., vol. 8, No. 2, June 1971, pp. 382-411.

[4-11] R.J. Hanson, "Computing Quadratic Programming Problems: Linear Inequality and Equality Constraints," Jet Propulsion Laboratory Report, 1970.

[4-12] C.L. Lawson and R.J. Hanson, Solving Least Squares Problems, Prentice Hall, to be published.

[4-13] H.M. Wagner, "Linear Programming Techniques for Regression Analysis," J. of the Amer. Stat. Assoc., vol. 54, March 1959.

[4-14] A. Zellner, "Linear Regression with Inequality Constraints on the Coefficients: an Application of Quadratic Programming and Linear Decision Rules," Report 6109 of the International Center for Management Science, Rotterdam, 1961.

[4-15] A. Zellner, "Decision Rules for Economic Forecasting," Econometrica, vol. 31, No. 1-2, Jan. - April, 1963.

[4-16] R.R. Hocking, "The Distribution of a Projected Least Squares Estimator," Annals of the Institute of Statistical Mathematics, vol. 17, 1965, pp. 357-362.

[4-17] T.C. Lee, G.G. Judge and A. Zellner, Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North Holland, Amsterdam, 1970.

[5-1] M.P. Ekstrom, "An Iterative Improvement Approach to the Numerical Solution of Vector Toeplitz Systems," Fifth Asilomar Conference on Circuits and Systems, Pacific Grove, California.

[5-2] M.P. Ekstrom, "A Numerical Algorithm for Identifying Spread Functions of Shift-Invariant Imaging Systems," IEEE Trans. on Computers, vol. C-22, No. 4, April 1973.

[5-3] F.A. Graybill, Introduction to Matrices with Application in Statistics, Wadsworth, Belmont, California, 1969.

[5-4] P. Lancaster, Theory of Matrices, Academic Press, New York, 1969.

[6-1] L. Cutler and D.S. Pass, A Computer Program for Quadratic Mathematical Models to be Used for Aircraft Design and Other Applications Involving Linear Constraints, Rand Corporation, Santa Monica, 1971.

[A-1] S.S. Wilks, Mathematical Statistics, Wiley, New York, 1962.