INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.

Liang, Zhengrong

BAYESIAN IMAGE PROCESSING OF DATA FROM CONSTRAINED SOURCE DISTRIBUTIONS

City University of New York PH.D. 1987

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) 169 seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received ✓
16. Other _____

_____

_____

University Microfilms International

BAYESIAN IMAGE PROCESSING OF DATA FROM CONSTRAINED SOURCE DISTRIBUTIONS

by

ZHENGRONG LIANG

A dissertation submitted to the Graduate Faculty in Physics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1987

This manuscript has been read and accepted for the Graduate Faculty in Physics in satisfaction of the dissertation requiement for the degree of Doctor of Philosophy.

12/5/86
date

Chairman of Examining Committee

12/8/86
date

Executive Officer

Supervisory Committee

The City University of New York

Abstract

BAYESIAN IMAGE PROCESSING OF DATA
FROM CONSTRAINED SOURCE DISTRIBUTIONS

by

Zhengrong Liang

Adviser: Professor Hiram Hart

A new Bayesian image processing (BIP) formalism which incorporates various categories of a priori source distribution information in treating measured data which obeys poisson or gaussian statistics is introduced and two classes of multiplicative and additive forms of iterative BIP algorithms are formulated. Different categories of a priori source information are described and the resulting probability source distributions developed.

Practical application of this work falls in the areas of biomedical and space related image reconstruction and restoration. Specifically BIP is most effective in those situations in which the a priori source information can be characterized in terms of probability functions of the source distributions and the noise in the measured data would otherwise significantly obscure the resulting images.

The results obtained using the BIP formalism on computer generated and experimental data from phantoms are clearly superior to standard methods wherever the a priori source information is reasonably accurate and sufficiently restrictive.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation and gratitude to Professor Hiram Hart, who, as my advisor, offered invaluable guidance, inspiration, and encouragement through every phase of this investigation and in the preparation of this dissertation.

It is indeed a pleasure to thank Professor D. Blaufox, Professor V. Chung, Professor R. Hatcher, Professor J. Laughlin and Professor M. Lax, for many enlightening discussions and helpful suggestions as well as their critical review of this thesis.

I am deeply indebted to Professor T.D. Lee, for providing the opportunity to come to the United States of America for study at the graduate school of City University of New York via the examination program of CUSPEA.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER I
INTRODUCTION

In many communication applications it is desired to extract information of known frequency, phase, or repetitive characteristics from a noisy signal. Many techniques have been developed to successfully discriminate against essentially random noise and so dramatically improve the signal to noise ratio. In radioisotope measurements and many other applications, however, the signal is itself intrinsically randomly fluctuating, and many of these communications techniques have not been applied. It has not in general been recognized in medical imaging that even a randomly fluctuating signal can reflect a priori known spatial or chromatic information. Furthermore in, for example, radioisotopic organ scans (Kuhl et al 1963 and Brownell 1964) the information desired is typically three-dimensional (four-dimensional if time-dependent dynamic functions are being studied), while the measured outputs are often only two-dimensional scans. Specialized equipment and methods of data analysis have therefore been developed in the field of radioisotopic medical imaging, such as described in "*Radionuclide Section Scanning*" by Ell et al (1982, Grune & Stratton) and "*Principles of Radionucide Emission Imaging*" by Kuhl (1983, Pergamon Press LTD). Attempts have also been made to reconstruct three-dimensional radioisotopic tissue distributions by acquiring two-dimensional signal gamma scan data at different angles of projection and carrying out a three-dimensional reconstruction by projective techniques, such as the X-ray computerized assisted tomographic (CAT) techniques described in "*Image Reconstruction* from *Projections*" by Herman (1980, Academic Press, Inc). Radioisotope tissue distributions have now been determined using positron emission tomographic (PET) coincidence techniques and single photon emission computerized tomographic (SPECT) techniques (Phelps 1977). These techniques are currently limited as to the organ scanned, the extent of the inhomogeneity of the tissues in the field of scanning, the available positron emitting isotopes, the dose delivered to the normal tissues and require specialized equipment, usually with either 2π planar or $\approx 4\pi$ solid angle geometries (Coulam et al 1981, Johns et al 1983 and Curry et al 1984).

Mathematical formalisms used in these techniques to reconstruct the radioisotope tissue distributions (or, more general, source distributions) are generally classified into two categories (Lewitt 1983 and Censor 1983). The transform (Fourier/convolution) methods are based on analytic inversion of formulas which involve the relation of the source distribution and the measured data and finish the source distribution reconstruction (or image reconstruction) in one step, so they need less computation time and are generally preferred in practice. On the other hand, the finite series-expansion (algebraic or iterative) methods are based on the discretization of the source distribution domain priori to any mathematical analysis and approach the final result iteratively; they need more computation time and converge more slowly.

It should be noted that transform methods which suppress high frequency noise necessarily tend to limit the resolution obtainable. Algebraic methods therefore can challenge transform methods in dealing

with stochastic variation in measured data and in suppressing noise propagation in the process of image reconstruction.

In general, the mathematical difficulties associated with a formal linear algebraic approach to the problem of determining the three-dimensional source distributions, Φ, from measured data, **Y**, are very formidable:

$$\mathbf{Y} = R\Phi + \mathbf{N} \tag{1.1}$$

where **N** is the noise vector.

Formal solutions would not only involve inversion of a series of relatively large three-dimensional point source response matrices or of three-dimensional tissue-detector geometric matrices, R, but also the utilization of noisy data appropriately weighted in accordance with overall statistical optimization constraints. In view of these difficulties, it has been customary instead to use various iterative approximation techniques, generally making use of projective (measured) data at multiple angles of orientation. (It should be noted, however, that if the point source response is itself sufficiently depth dependent, there is, in principle, no a priori requirement that multiple projections be taken.)

These iterative approximation techniques have been variably successful in extracting the true source distribution from the projected noisy data.

Some of the iterative algorithms are additive:

a. The Jacobi method and Gauss-seidel method (Dahlquist et al 1974) solve directly for solution Φ of Eq.(1.1) but are only generally effective for dominant diagonal R matrices and essentially noise free measured data **Y**;

b. The ART (algebraic reconstruction techniques) (Herman 1980 and Gordon et al 1970) search for the estimated solution of Eq.(1.1) using iterative orthogonal projections onto hyperplanes of $\{Y_i = (R, \Phi)_i\}$;

c. The ILST (iterative least square techniques) (Budinger et al 1974) and SIRT (simultaneous iterative reconstruction techniques) (Gilbert 1972 and Lakshminarayanan et al 1979) approach minimum measured data noise solutions of Eq.(1.1);

d. The Quadratic optimization methods (Herman 1980, Artzy et al 1979 and Herman et al 1976) are directed at finding minimum measured data noise and smoothing solutions of Eq.(1.1);

e. The Bayesian method (Hunt 1977) selects that solution of Eq.(1.1), which minimizes the noise subject to the a priori condition of smoothing solution.

Other algorithms are multiplicative:

f. The Gold's ratio method (Jansson 1984) uses the ratio of measured data and calculated value to correct the solution $\Phi^{(n)}$ (where n represents the iterative index) at each iteration rather than using additive corrections as in a);

g. The MART (multiplicative ART) (Lent 1976) and MENT (maximum entropy techniques) (Minerbo 1979) serach for the results of maximum entropy with constraints of Eq.(1.1);

h. The NLM (non-local method) (Hart 1983) results in a statistically weighted solution of Eq.(1.1);

i. The maximum likelihood and maximum entropy method (Frieden 1972 and 1980) determines that solution of Eq.(1.1), which is resulted from most likely source-noise combination without any a priori source information available;

j. The Bayesian deconvolution method (Kennett et al 1979) is actually a maximum likelihood method assuming that the measured data obeys poisson statistics.

Recently a maximum likelihood (ML) approach has been developed by Shepp et al (1982, IEEE Trans. Med. Imaging), and Lange et al (1984, J. Comput. Assis. Tomography) which makes use of modern statistical methodology to derive algorithms which under certain conditions serve to formally specify those source distributions which are probabilistically most likely, considering the measured data, the statistical distributions satisfied by the measured data (e.g. Poisson distribution, or Gaussian distribution) and the known point source response function (*PSF*, distortion characteristcs of the equipment and measuring process). Excellent solutions are achieved if the *PSF* are sharp implying relatively little distortion to begin with and in the presence of minimal noise. However, these maximum likelihood expectation and maximization algorithms (MLEM, Shepp et al 1982 and Lange et al 1984) do not in general give accurate results when the *PSF* is broad and the data is noisy. Since a broad *PSF* implies a low resolution detection system and noisy data is necessarily inherently less reliable, the limitations in solution specificity would appear to be inherent and the above approaches therefore still optimal.

Nonetheless in certain situations of very basic importance, a further major improvement in image processing can be achieved.

Consider a situation, for example, in which while the spatial distribution of the source distribution is relatively unknown, the radioisotope concentration of the individual source elements is readily estimated. This occurs, for example, in cardiac imaging immediately following I.V. administration of a

radioisotopic dose. The blood concentration curves can be relatively well measured or otherwise estimated over the next few minutes while the cardiac muscle will be comparatively free of isotope. In brain imaging, hepatic imaging and thyroid imaging studies, probable configurations of cold or hot lesions may be characterizable in terms of estimated amplitudes and spatial relations between source elements. Lesion detection could then be viewed as a search for fuzzy patterns superimposed upon a uniform background.

It follows then that a generally effective method for extracting the most probable source distribution would take into account *all* of the information which may be available.

These observations can be expressed more precisely in terms of the standard statistical relation, Bayes' Law:

$$P(\Phi|\mathbf{Y}) = P(\mathbf{Y}|\Phi)\,P(\Phi)/P(\mathbf{Y}) \qquad (1.2)$$

where:

$P(\Phi|\mathbf{Y})$ is the conditional probability of the source vector Φ subject to the data vector $\mathbf{Y}$;

$P(\mathbf{Y}|\Phi)$ is the conditional probability of the data vector $\mathbf{Y}$ subject to the source vector Φ;

$P(\Phi)$ is the probability distribution of the source vector Φ;

and

$P(\mathbf{Y})$ is the probability distribution of the data vector $\mathbf{Y}$.

When $P(\Phi)$ and $P(\mathbf{Y})$ are assumed constant, $P(\Phi|\mathbf{Y})$ and $P(\mathbf{Y}|\Phi)$ are of course linearly proportional. Maximizing $P(\mathbf{Y}|\Phi)$ (which is the approach previously taken) is then equivalent to maximizing $P(\Phi|\mathbf{Y})$. In the Bayesian formalism, however, any non-trivial a priori source information is first incorporated in $P(\Phi)$ and the more logical direct maximization of $P(\Phi|\mathbf{Y})$ carried out, taking into account, of course, the now non-constant $P(\Phi)$ on the right side of Eq.(1.2).

Hunt (1977, IEEE Trans. Computers) indicated the possibility of utlization of a priori knowledge about the noise and/or the sources in the Bayesian formalism, where he assumed that functions $P(\mathbf{Y}|\Phi)=P(\mathbf{N})$ and $P(\Phi)$ are described by the multivariate normal probability densities.

In this thesis, a new Bayesian image processing (BIP) formalism which incorporates various categories of a priori source distribution information in treating measured data which obeys Poisson or Gaussian statistics is now developed. The source distribution information $P(\Phi)$ in this formalism is viewed as falling into two general categories:

1. generic a priori source information, which results in general global constraints on the source distribution;

2. quantitative a priori source information, which specifies a most probable values of a sub-set of source elements $\{\phi_j\}$, or pattern distribution in the imaging field, and so can result in much more restriction on the source distribution.

Two classes of multiplicative and additive forms of iterative BIP algorithms, which involving various examples of each of these two general types of $P(\Phi)$'s will be illustrated subsequently in detail.

This BIP work can be applied to the areas of biomedical and space related image reconstruction and restoration.

In the biomedical imaging area, this work could apply to clinical studies in nuclear medicine involving quantitative assessment of the distribution of injected radiopharmaceuticals and radionuclides in man. Data arising from basic physiological and biochemical studies in animals can also be more effectively interpreted using the Bayesian considerations.

In satellite and other space related applications this work may also be of value in searching for specific patterns in a noisy data field.

The BIP formalism can also be used to develope a parallel route of "pattern recognition" from processed results comparable to the standard pattern recognition from measured data (Fukunaga 1972).

The organization of this dissertation is as follows. The transform methods and iterative methods of image processing are discussed in chapters II and III respectively. In addition, the ill-posed problem and direct matrix-inversion methods are mentioned. In chapter IV, different categories of a priori source information $P(\Phi)$ are considered. The EM (expectation and maximization) technique (Dempster et al 1977) and its application to radioisotope organ imaging study are described in chapter V. In chapter VI, a BIP formalism for generic source constraints is formulated and the BIP algorithms are derived using the EM technique; while a BIP formalism for quantitative source constraints is formulated in chapter VII. A BIP formalism for fuzzy source constraints is outlined in chapter VIII and an objective spatial-amplitude metric is used to evaluate comparative resuluts. In chapter IX, the BIP algorithms derived in chapters VI, VII and VIII are applied to imaging data from experimental phantoms. Practical considerations relating to the fast Fourier transform, use of reduced data ranges and the relative speed of convergence of the BIP algorithms are indicated. Chapter X is devoted to the formulation of additive BIP algorithms. The concavity of various Bayesian functions is demonstrated in part 1 of the Appendix. The entropy analysis and the random number generators used are detailed in parts 2 and 3. Appendix 4 describes Bayesian contributions in the recent literature. A table of acronyms appears in Appendix 5.

CHAPTER II
IMAGE PROCESSING OF MEASURED DATA

An arbitrary source distribution function, $\Phi(x,y,z)$, can not be uniquely determined from a finite number of measured data elements, $\{Y_i\}=\mathbf{Y}$. However as in most algebraic schemes, the region of source distribution is usually divided into small elements, or voxels,++ $\{\phi_j\}=\Phi$, and the value of ϕ_j associated with the voxel j is the average of $\Phi(x,y,z)$ over the voxel volume. It may then be possible to estimate these voxel values from a finite number of data values, or terminologically, carry out image reconstruction from projection data or image restoration from "convolution" data (Herman 1980 and Ekstrom 1984).

The somewhat arbitrary distinction between image reconstruction and image restoration relates to the function $R_i(x,y,z)$ in the general Eq.(2.1):

$$Y_i = \int R_i(x,y,z)\,\Phi(x,y,z)\,dxdydz + N_i\,,\quad i=1,2,...,I \tag{2.1}$$

reflecting the fundamental assumption of a linear imaging system.

The function $R_i(x,y,z)$ is the probability of including contributions from voxel $\phi_j(x,y,z)$ in the data element Y_i and is in general a function of the voxel-detector geometry, tissue attenuation and scattering. It is determined, in image restoration by scanning a point source (Rosenfeld et al 1982) and, in image reconstruction by calculating the geometric relation of voxel-detector taking the tissue attenuation and scattering into account (Cho et al 1974).

If the region of source distribution is divided into J voxels, the value of R_{ij} can be expressed mathematically as:

$$R_{ij} = (A_{ij} + B_{ij})\,C_{ij}\,,\quad i=1,2,...,I;\ j=1,2,...,J \tag{2.2}$$

where A_{ij} represents the purely geometric factor (or weighting factor) and its value may be determined by the line length of intersection of voxel j with a ray in projection i (Brooks et al 1976); B_{ij} describes the scattering effect (Pang et al 1979) (i.e., R_{ij} can be non-vanishing even if voxel j on a purely geometric basis should not contribute to projection i) and $C_{ij} = \exp[-\int_0^{L_{ij}} u(x)dx\,]$ represents the effect of tissue attenuation (u(x) is the linear attenuation coefficient) for photons from voxel j through line L_{ij} to a detector in projection i.

++ The usual distinction between a voxel and a pixel is illustrated in Fig.(2.1). The voxel is a 3-dimensional volumn element denoted by V in the figure and the pixel is a an areal element indicated by P in the plane $ABCD$ of the figure.

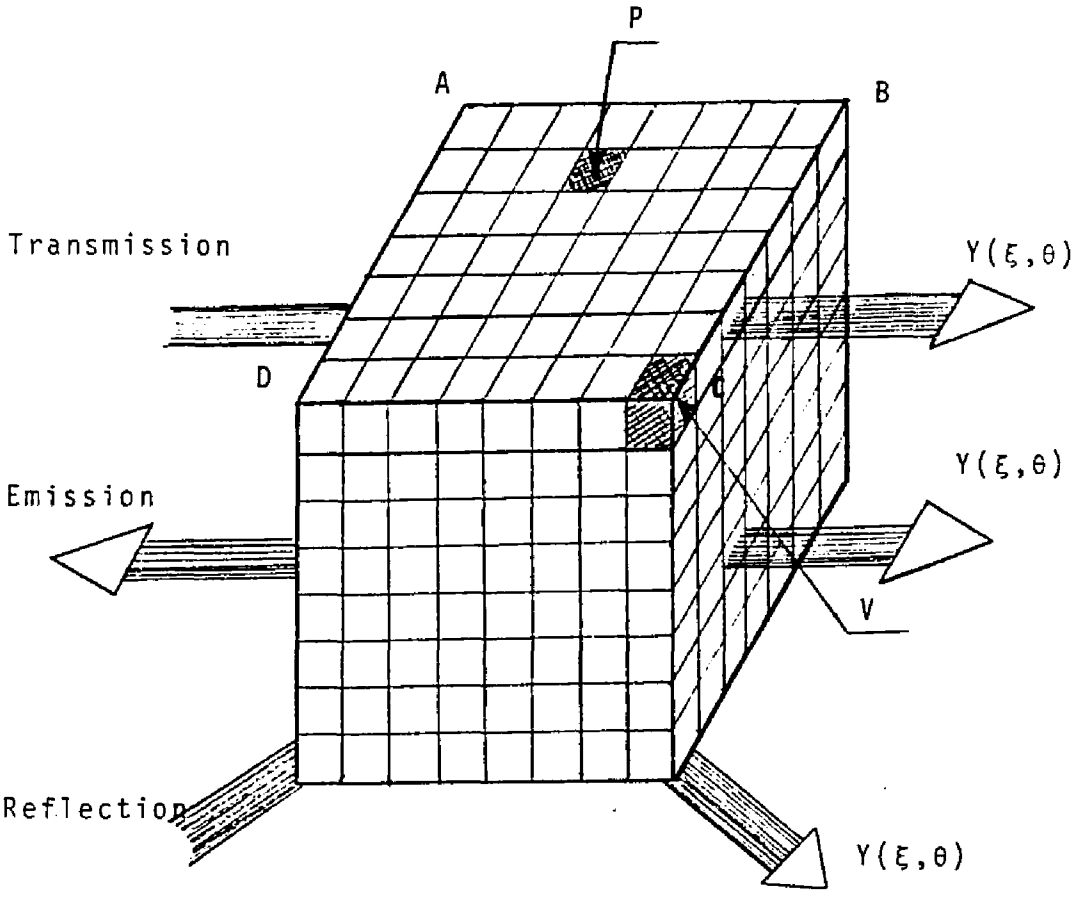


Figure 2.1: Characterization of three dimensional source distributions. The volume element V is called a voxel and the areal element P is called a pixel. Image reconstruction data results from transmission (e.g. X-ray computerized tomography), emission (e.g. single photon emission computerized tomography) and reflection (e.g. Ultrasound tomography).

The effects of radioactive decay of the source and of variable projection time intervals can also be included in R_{ij} by a multiplying factor, $D_i=(e^{-\lambda t_{i1}}-e^{-\lambda t_{i2}})/\lambda$, on the right side of Eq.(2.2); here λ is the decay constant and $t_{i2}-t_{i1}$ the time interval of the i-th measurement (or projection i).

For both image reconstruction and image restoration, image processing comprises the following:

a. first of all, one has to find out to what extent the source is determined by the measured data;

b. secondly, one needs to know to what extent the obtainable source is stable with respect to data errors;

c. thirdly, one has to derive the algorithms which are actually applied;

d. finally, calculation of the algorithms and display of the resulting images.

1) IMAGE RECONSTRUCTION

Image reconstruction from projections is the process of producing an image of a two-dimensional distribution (usually of some physical property, e.g. X-ray attenuation coefficient for CT -- computerized tomography, concentration of γ-ray emitting substances for PET -- positron emission tomography and SPECT -- single photon emission computerized tomography, tissue refractive index for Ultrasound and tissue relaxation time for MRI -- nuclear magnetic resonance imaging) from estimates of its line integrals along a finite number of lines of known location. A three-dimensional volume distribution can then be constructed from a series of parallel two-dimensional sections.

In Fig.(2.1), the volume image is built up from a series of parallel sectional pictures. The sectional image (represented in the area of *ABCD*) is displayed on the picture region *ABCD* by the sum of numerical numbers (or gray levels) at specified pixels (areal elements, one is denoted by P, of *ABCD*) respectively. The numerical numbers are derived from measured data by use of some mathematical algorithms and are related to specified voxels (volume elements, one is indicated by V, of the volume) respectively.

Transmission computerized tomography (TCT, or CT) as shown in the figure is one in which an image is produced that displays the absorption coefficients for X-rays passing through the volume (or body) from an external source (Ledley 1976 and White et al 1981).

Emission computerized tomography (ECT, or PET and SPECT) display the distribution of the concentration of radioisotopes distributing within the body (Phelps et al 1975 and Jaszczak et al 1980).

A third modality (reflection CT, or Ultrasound) produces an image of refractive index distribution or of Ultrasound propagation speed distribution within the body (Greenleaf 1983, Greenleaf et al 1978 and Glover et al 1977).

Finally, MRI produces an image reflecting magnetic resonance tissue relaxation times of body tissues (Mansfield et al 1982, Hinshaw et al 1983 and Cho et al 1982).

1.1) X-ray Computerized Tomography (CT)

The aim of CT is to obtain information regarding the nature of material occupying positions inside the body and to discover and display the precise shape of selected parts of the material (or organs inside the body) --- the anatomic structure of the body.

As shown in Fig.(2.2), the general expression for the detected X-ray intensity transmitted through an attenuation medium is (Macovski 1983):

$$I(\xi,\theta)=\int\int S(\varepsilon,z)\, e^{-\int u(x,y,z,\varepsilon)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy}\,dz\,d\varepsilon \qquad (2.3)$$

where $I(...)$ is the detected X-ray intensity for the slice positioned at z, $S(...)$ the energy spectrum of the incident X-ray beam and $u(...)$ the attenuation distribution within the body. The coordinates (ξ,θ) specify the projection ray which passes through point $(\xi,\eta=0)$ at projection angle θ.

If we assume that the detectors are perfect (i.e., the collimator is ideally small and the electronic process is linear) and ignore scattering effect, then $I(\xi,\theta)$ will be the experimental data.

For a monoenergetic source

$$S(\varepsilon,z)=I_0 S(z)\,\delta(\varepsilon-\varepsilon_0)$$

and slice beam detection, there is:

$$S(\varepsilon,z)=I_0\,\delta(\varepsilon-\varepsilon_0)\delta(z-z_0)$$

then Eq.(2.3) becomes:

$$I(\xi,\theta)=I_0\, e^{-\int u(x,y)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy}$$

or simply

$$ln(\frac{I_0}{I(.)})=\int u(x,y)\,|_\xi\, d\eta\,.$$

Let $\Phi(x,y)$ be the attenuation function $u(x,y)$ and $Y(\xi,\theta)$ the "detected" data $ln(\frac{I_0}{I})$ at projection angle θ, then (Cormack 1973),

$$\begin{aligned} Y(\xi,\theta)&=\int\Phi(x,y)\,d\eta \\ &=\int\Phi(x,y)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy\,. \end{aligned} \qquad (2.4)$$

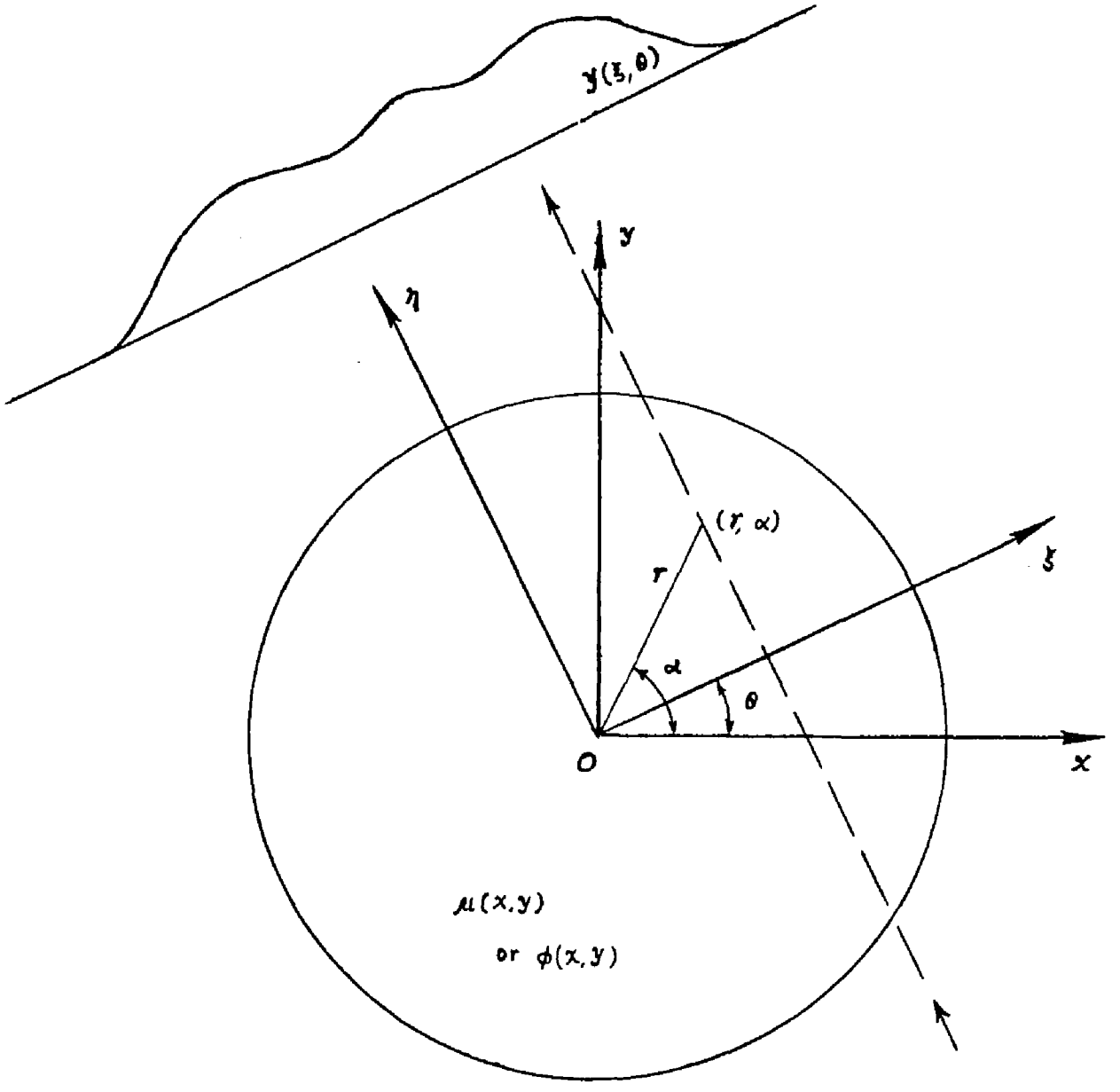


Figure 2.2: The relation of sectional source distribution $\Phi(x,y)$ and projection $Y(\xi,\theta)$:

1) CT:

$$Y(\xi,\theta)=\int_0^\infty \Phi(x,y)\, d\eta\ ,\quad \Phi(x,y)=u(x,y);$$

2) PET:

$$Y(\xi,\theta)=\exp\left[-\int_0^\infty u(x,y)\, d\eta\right] \int_0^\infty \Phi(x,y)\, d\eta;$$

3) SPECT:

$$Y(\xi,\theta)=\int_0^\infty \Phi(x,y)\, \exp\left[-\int_\eta^\infty u(x',y')\, d\eta'\ \right] d\eta.$$

A discrete formulation follows:

Let

$$\Phi(x,y)=\sum_j \phi_j\, b_j(x,y) \tag{2.5}$$

where $b_j(x,y)$ serves as a means for dividing the continuous function $\Phi(x,y)$ into discrete values $\{\phi_j\}$. For simplicity, $b_j(x,y)=1$ when $b_j(x,y)$ is within voxel j and zero when $b_j(x,y)$ outside of voxel j.

Substitute Eq.(2.5) into Eq.(2.4),

$$Y(\xi,\theta)=\sum_j \phi_j \int b_j(x,y)\, d\eta .$$

For the i-th projection ray Y_i at projection angle θ, there results:

$$Y_i=\sum_j \phi_j \int b_j(x,y)\, d\eta_i = \sum_j R_{ij}\phi_j \tag{2.6}$$

where $R_{ij}=\int b_j(x,y)\,d\eta_i$ and index i covers all projection rays at different projection angles.

Consider noise in the data **Y**, then Eq.(2.6) becomes:

$$Y_i=\sum_j R_{ij}\phi_j + N_i . \tag{2.7}$$

The transform solutions of Eq.(2.4) are given below and iterative solutions of Eq.(2.7) will be discussed later.

A). THE RADON TRANSFORM SOLUTION

For any point o in the x–y plane, take the average of the line integrals on lines which are tangent to a circle of radius $r=\sqrt{x^2+y^2}$ with the point o as its center, i.e.,

$$\bar{Y}(\xi)=\frac{1}{2\pi}\int_0^{2\pi} Y(\xi,\theta)\,d\theta$$

where $Y(\xi,\theta)$ is related to $\Phi(x,y)$ by Eq.(2.4), then the value of $\Phi(r,\alpha)$ at the point o is (Radon 1917):

$$\Phi(o)=-\frac{1}{\pi}\int_0^{\infty}\frac{d\bar{Y}(\xi)}{\xi}=-\frac{1}{\pi}\int_0^{\infty}\frac{d\bar{Y}(\xi)}{d\xi}\,\frac{d\xi}{\xi} . \tag{2.8}$$

This pioneering result which serves as the basic theoretic foundation for much of image reconstruction has been reproduced in the current literature (Herman 1980).

Noting that $Y(\xi,\hat{\xi})=Y(-\xi,-\hat{\xi})$, it follows that (John 1955):

$$\Phi(r,\alpha)=\frac{1}{2\pi^2}\nabla^2\left[\int_0^{\pi}d\theta\int_{-\infty}^{\infty}\ln|\xi-r(\hat{x}\cdot\hat{\xi})|\;Y(\xi,\hat{\xi})\,d\xi\right]$$

$$=\frac{1}{2\pi^2}\int_0^{\pi}d\theta\int_{-\infty}^{\infty}\ln|\xi-r(\hat{x}\cdot\hat{\xi})|\;\frac{\partial^2 Y(\xi,\theta)}{\partial\xi^2}\,d\xi$$

$$=-\frac{1}{2\pi^2}\int_0^{\pi}d\theta\int_{-\infty}^{\infty}\frac{\partial Y(\xi,\theta)}{\partial\xi}\;\frac{d\xi}{\xi-r(\hat{x}\cdot\hat{\xi})} \qquad (2.9)$$

where $(\hat{x}\cdot\hat{\xi})=\cos(\alpha-\theta)$ and the Cauchy principle has been used.

B). FOURIER TRANSFORM SOLUTION

The Fourier transform (F.T.) is defined as:

$$\Psi(\xi,\eta)=\frac{1}{2\pi}\int\int\Phi(x,y)\;e^{-i\,(x\,\xi+y\,\eta)}\,dxdy$$

with

$$\Phi(x,y)=\frac{1}{2\pi}\int\int\Psi(\xi,\eta)\;e^{i\,(x\,\xi+y\,\eta)}\,d\xi\,d\eta\,. \qquad (2.10)$$

Rotation by an angle θ, there are relations of

$$x=\xi\cos\theta-\eta\sin\theta\,,\qquad y=\xi\sin\theta+\eta\cos\theta \qquad (2.11)$$
$$dxdy=d\xi\,d\eta\,,\qquad r\,dr\,d\alpha=k\,dk\,d\theta$$

and

$$\xi=k\cos\theta\,,\qquad \eta=k\sin\theta\,.$$

By use of Eqs.(2.10) and (2.11)ff, there results (Mersereau 1976):

$$\Psi(\xi,\eta)=\Psi(k,\theta)=\frac{1}{2\pi}\int\int\Phi(x,y)\;e^{-i\,(x\,k\cos\theta+y\,k\sin\theta)}\,dxdy$$

$$=\frac{1}{2\pi}\int\int\Phi(\xi\cos\theta-\eta\sin\theta,\xi\sin\theta+\eta\cos\theta)\;e^{-i\,k\xi}\,d\xi\,d\eta$$

$$=\frac{1}{2\pi}\int d\xi\;e^{-i\,k\xi}\int\Phi(\xi\cos\theta-\eta\sin\theta,\xi\sin\theta+\eta\cos\theta)\,d\eta$$

$$=\frac{1}{2\pi}\int Y(\xi,\theta)\;e^{-i\,k\xi}\,d\xi \qquad (2.12)$$

so,

$$\Phi(x,y)=\frac{1}{2\pi}\int\int\Psi(\xi,\eta)\,e^{i(x\xi+y\eta)}d\xi\,d\eta$$

$$=(\frac{1}{2\pi})^2\int\int d\xi d\eta\,e^{i(x\xi+y\eta)}\,[\int e^{-ik\xi'}Y(\xi',\theta)\,d\xi'].\qquad(2.13)$$

C). THE FOURIER---BESSEL SOLUTION

On the plane of $x-y$, the function $\Phi(r,\alpha)$ is a periodic function in α with period 2π, so, using Fourier expansion, there is:

$$\Phi(r,\alpha)=\sum_{m=-\infty}^{\infty}\Phi_m(r)\,e^{im\alpha}\qquad(2.14)$$

with

$$\Phi_m(r)=\frac{1}{2\pi}\int_0^{2\pi}\Phi(r,\alpha)\,e^{-im\alpha}\,d\alpha.$$

By use of 2-dimensional polar coordinate F.T. (Bracewell 1978), i.e.,

$$\Psi(k,\theta)=\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty}\Phi(r,\alpha)\,e^{ikr\cos(\theta-\alpha)}r\,dr\,d\alpha\qquad(2.15)$$

with

$$\Phi(r,\alpha)=\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty}\Psi(k,\theta)\,e^{ikr\cos(\alpha-\theta)}k\,dk\,d\theta$$

and noting that

$$e^{iz\cos\theta}=\sum_{l=-\infty}^{\infty}(i)^lJ_l(z)e^{il\theta}$$

where $i=e^{i\frac{\pi}{2}}$ and $J_l(.)$ is the Bessel function, there is:

$$\Phi_m(r)=\frac{1}{2\pi}\int_0^{2\pi}d\alpha e^{-im\alpha}\,[\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty}\Psi(k,\theta)\,e^{ikr\cos(\alpha-\theta)}k\,dk\,d\theta]$$

$$=(\frac{1}{2\pi})^2\int_0^{2\pi}d\alpha\,e^{-im\alpha}\int_0^{2\pi}d\theta\int_0^{\infty}k\,dk\,\Psi(k,\theta)\sum_{l=-\infty}^{\infty}e^{il\frac{\pi}{2}}J_l(kr)\,e^{il(\alpha-\theta)}$$

$$=(\frac{1}{2\pi})^2\int_0^{2\pi}d\theta\int_0^{\infty}k\,dk\sum_l J_l(kr)e^{il(\frac{\pi}{2}-\theta)}\Psi(k,\theta)\int_0^{2\pi}d\alpha e^{i\alpha(l-m)}$$

$$=\frac{1}{2\pi}\int_0^{2\pi}d\theta\int_0^{\infty}k\,dk\sum_l J_l(kr)e^{il(\frac{\pi}{2}-\theta)}\Psi(k,\theta)\,\delta_{lm}$$

$$=\frac{1}{2\pi}\int_0^{2\pi}d\theta\int_0^{\infty}k\,dk\,J_m(kr)e^{im(\frac{\pi}{2}-\theta)}\Psi(k,\theta)\qquad(2.16)$$

where the delta function

$$\delta_{lm} = \frac{1}{2\pi}\int_0^{2\pi} d\alpha\, e^{i\alpha(l-m)}$$

has been used.

Since $\Psi(k,\theta)$ is also a periodic function in θ with period 2π, Fourier expansion gives:

$$\Psi(k,\theta) = \sum_{m=-\infty}^{\infty} \Psi_m(k)\, e^{i m(\theta-\frac{\pi}{2})} = \frac{1}{2\pi}\int Y(\xi,\theta)\, e^{i k\xi}\, d\xi \qquad (2.17)$$

with

$$\Psi_l(k) = \frac{1}{2\pi}\int_0^{2\pi} \Psi(k,\theta)\, e^{-i l(\theta-\frac{\pi}{2})}\, d\theta$$

then,

$$\begin{aligned}
\Phi_m(r) &= \frac{1}{2\pi}\int_0^{\infty} k\, dk\, J_m(kr) \int_0^{2\pi} d\theta e^{i m(\frac{\pi}{2}-\theta)} \sum_{l=-\infty}^{\infty} \Psi_l(k) e^{-i l(\theta-\frac{\pi}{2})} \\
&= \int_0^{\infty} k\, dk\, J_m(kr) \sum_l \Psi_l(k) \frac{1}{2\pi}\int_0^{2\pi} d\theta\, e^{i\theta(l-m)}\, e^{-i\frac{\pi}{2}(l-m)} \\
&= \int_0^{\infty} k\, dk\, J_m(kr) \sum_l \Psi_l(k)\, \delta_{lm} \\
&= \int_0^{\infty} k\, dk\, J_m(kr)\, \Psi_m(k). \qquad (2.18)
\end{aligned}$$

So starting from Eq.(2.17) and then using Eq.(2.18), $\Phi_m(r)$ is obtained. Finally, $\Phi(r,\alpha)$ is given by Eq.(2.14). All together

$$\begin{aligned}
\Phi(r,\alpha) &= \sum_m e^{i m\alpha} \int_0^{\infty} k\, dk\, J_m(kr) \frac{1}{2\pi}\int_0^{2\pi} d\theta\, e^{-i m(\theta-\frac{\pi}{2})} \frac{1}{2\pi}\int d\xi\, e^{i k\xi}\, Y(\xi,\theta) \\
&= (\frac{1}{2\pi})^2 \int_0^{\infty} k\, dk \int_0^{2\pi} d\theta \int d\xi\, Y(\xi,\theta)\, e^{i k\xi} \sum_m J_m(kr)\, e^{i m(\alpha-\theta+\frac{\pi}{2})}. \qquad (2.19)
\end{aligned}$$

D). CONVOLUTION BACKPROJECTION SOLUTION

Defining the operators below:

a). Backprojector B_x:

$$[B_y\Phi](x,y) = \int_0^{\pi} \Phi(x\cos(\theta-y),\theta)\, d\theta$$

b). Differentiator D_x:

$$[D_x\Phi](x,y)=\lim_{\Delta x\to 0}\frac{\Phi(x+\Delta x,y)-\Phi(x,y)}{\Delta x}$$

c). Hilbert transformer H_x:

$$[H_x\Phi](x,y)=-\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\Phi(x',y)}{x-x'}dx'$$

then there is:

$$\Phi(r,\alpha)=-\frac{1}{2\pi^2}\int_0^{\pi}d\theta\int_{-\infty}^{\infty}d\xi\,\frac{\partial Y(\xi,\theta)}{\partial\xi}\frac{d\xi}{\xi-r\cos(\alpha-\theta)}$$

$$=-\frac{1}{2\pi}[B_\alpha H_\xi D_\xi Y](\xi,\theta). \qquad (2.20)$$

The regularizing family approach can give the result of (Chang et al 1978)

$$\lim_{\mu\to\infty}[\Phi * W_\mu](x,y)=[H_x\Phi](x,y)$$

with the set $\{W_\mu|\mu>0\}$ of functions on the real numbers μ and the star * the convoluting symbol.

If W_μ satisfies:

$$[(D_\xi Y(\xi,\theta))*W_\mu](\xi,\theta)=\int_{-\infty}^{\infty}\frac{\partial Y(\xi',\theta)}{\partial\xi'}W_\mu(\xi-\xi')\,d\xi'$$

$$=\int_{-\infty}^{\infty}Y(\xi',\theta)\,W_\mu(\xi-\xi')\,d\xi=[Y_\theta * W_\mu]_\xi(\xi) \qquad (2.21)$$

then,

$$\Phi(r,\alpha)=-\frac{1}{2\pi}\{B_\alpha[Y_\theta * W_\mu]_\xi(\xi)\}(\xi,\theta). \qquad (2.22)$$

So, the objective is to chose the function W_μ, so that it satisfies:

$$W_\mu(x)=2\int_0^{\mu/2}z\,F_\mu(z)\cos(2\pi xz)\,dz$$

with

(i) $0\le F_\mu(z)\le 1$, and $F_\mu(z)=0$ if $z\ge\frac{\mu}{2}$;

(ii) $F_\mu(z)\le c$ for z increasing, where c is a constant;

(iii) $\lim_{\mu\to\infty} F_\mu(z)=1.$

Some other transform methods were indicated by Lewitt (1983, Proc. IEEE) and Louis et al (1983, Proc. IEEE).

E). FILTERED BACKPROJECTION SOLUTION

By use of the F.T., there is Eq.(2.12) of (Brooks et al 1976)

$$\Psi(k,\theta)=\frac{1}{2\pi}\int_{-\infty}^{\infty} Y(\xi,\theta)\,e^{-ik\xi}\,d\xi.$$

The backprojection is expressed, for

$$Y(\xi,\theta)=\int_{-\infty}^{\infty}\Phi(x,y)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy$$

as:

$$\Phi(x,y)=\int_0^{\pi} Y(\xi,\theta)\,d\theta=\int_0^{\pi} Y(x\cos\theta+y\sin\theta,\theta)\,d\theta \qquad (2.23)$$

$$=\frac{1}{2\pi}\int_0^{\pi}d\theta\int_{-\infty}^{\infty}\Psi(k,\theta)\,e^{ik(x\cos\theta+y\sin\theta)}\,dk$$

$$=\frac{1}{2\pi}\int_0^{\pi}\int_{-\infty}^{\infty}\frac{\Psi(k,\theta)}{|k|}|k|\,e^{ik(x\cos\theta+y\sin\theta)}\,dk\,d\theta$$

$$=\frac{1}{2\pi}\int\int\frac{\Psi(k,\theta)}{|k|}\,e^{i(x\xi+y\eta)}\,d\xi\,d\eta \qquad (2.24)$$

where the rotation transform relations Eq.(2.11) have been used.

Note that Eq.(2.24) is just the inverse F.T. of Eq.(2.10) with filter $\frac{1}{|k|}$.

a). Fourier Filtering Backprojection

From the 2-dimensional polar coordinate F.T. of Eq.(2.15), or

$$\Psi(k,\theta)=\frac{1}{2\pi}\int_0^{\pi}d\alpha\int_{-\infty}^{\infty}|r|\,dr\,\Phi(r,\alpha)\,e^{ikr\cos(\theta-\alpha)}$$

there is:

$$\Phi(r,\alpha)=\frac{1}{2\pi}\int_0^{\pi}d\theta\int_{-\infty}^{\infty}|k|\,dk\,\Psi(k,\theta)\,e^{ikr\cos(\alpha-\theta)}. \qquad (2.25)$$

Comparing with Eq.(2.23), there is:

$$Y(x\cos\theta+y\sin\theta,\theta)=\frac{1}{2\pi}\int_{-\infty}^{\infty}|k|\,dk\,\Psi(k,\theta)\,e^{i\,k\,r\cos(\alpha-\theta)}$$

$$=\frac{1}{2\pi}\int|k|\,dk\,\Psi(k,\theta)\,e^{i\,k\,(x\cos\theta+y\sin\theta)} \qquad (2.26)$$

where

$$k\,r\cos(\alpha-\theta)=k\,r\,(\cos\alpha\cos\theta+\sin\alpha\sin\theta)=k\cos\theta\,r\cos\alpha+k\sin\theta\,r\sin\alpha$$
$$=k\,(x\cos\theta+y\sin\theta)=x\xi+y\eta$$

has been used.

The steps of the reconstructing procedure are:

(i) compute 1-dimensional F.T. of Eq.(2.12); $\Psi(k,\theta)$;

(ii) introduce the filtering function $|k|$;

(iii) compute filtered 1-dimensional inverse F.T.,

$$\hat{Y}(\xi,\theta)=\frac{1}{2\pi}\int|k|\,dk\,\Psi(k,\theta)\,e^{i\,k\xi};$$

(iv) note that

$$\hat{Y}(x\cos\theta+y\sin\theta)=\hat{Y}(\xi,\theta)$$

and particularize the continuous variables x, y to the discrete voxel coordinate sets and evaluate $\hat{Y}(x\cos\theta+y\sin\theta)$ by interpolation (Smith et al 1973, Mersereau 1976 and Stark et al 1981);

(v) calculate backprojection

$$\Phi(x,y)=\frac{\pi}{m_\theta}\sum_{l=1}^{m_\theta}\hat{Y}(x\cos\theta_l+y\sin\theta_l,\theta_l)$$

where m_θ is the number of projection angles.

b). Radon Filtering Backprojection

From Eq.(2.26) and let $v=x\cos\theta+y\sin\theta$, there is:

$$Y^*(\nu,\theta)=\frac{1}{2\pi}\int|k|dk\,\Psi(k,\theta)\,e^{ik\nu}=\frac{1}{\sqrt{2\pi}}F\{|k|\Psi(k,\theta)\}$$
$$=\frac{1}{\sqrt{2\pi}}F\{\Psi(k,\theta)\}*F\{|k|\}=\frac{1}{\sqrt{2\pi}}Y(\nu,\theta)*(\frac{-1}{2\pi^2\nu^2}) \qquad (2.27)$$
$$=-(\frac{1}{2\pi})^{3/2}\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{Y(\nu',\theta)}{(\nu-\nu')^2}d\nu'$$

where

$$F\{|k|\}=-\frac{1}{2\pi^2\nu^2}$$

and Eq.(2.12)ff have been used.

The steps of computation are:

(i) make interpolation $Y(\nu,\theta)=Y(\xi,\theta)$;

(ii) compute convolution

$$Y^*(\nu,\theta)=-\frac{1}{\pi}(\frac{1}{2\pi})^{3/2}\int_{-\infty}^{\infty}\frac{Y(\nu',\theta)}{(\nu-\nu')^2}d\nu';$$

(iii) make interpolation $\hat{Y}(x\cos\theta+y\sin\theta,\theta)=Y^*(\nu,\theta)$;

(iv) calculate backprojection

$$\Phi(x,y)=\frac{\pi}{m_\theta}\sum_{l=1}^{m_\theta}\hat{Y}(x\cos\theta_l+y\sin\theta_l,\theta_l);$$

c). Convolution Filtering Backprojection

In order to avoid the divergence of $F\{|k|\}$ in Eq.(2.27), let

$$F\{|k|\}=\begin{cases}F\{|k|\} & |k|<k_m\\ F\{|0|\} & \text{otherwise}\end{cases}$$

there is:

$$F\{|k|\}=\frac{1}{\sqrt{2\pi}}\int_{-k_m}^{k_m}|k|\,e^{ik\nu}dk=\frac{1}{\sqrt{2\pi}}[\frac{k_m}{\pi}\frac{\sin(k_m\nu)}{\nu}-\frac{1}{\pi^2}\frac{\sin^2(k_m\nu/2)}{\nu^2}]$$

then,

$$Y^*(\nu,\theta)=Y(\nu,\theta)*F\{|k|\}$$
$$=\frac{1}{\sqrt{2\pi}}k_mY(\nu,\theta)-\int_{-\infty}^{\infty}Y(\nu',\theta)\frac{\sin^2[k_m(\nu-\nu')/2]}{\pi^2(\nu-\nu')^2}d\nu'.$$

Summation with

$$k_m=\frac{\pi}{2\nu}\ ,\ \text{and}\ \ \sin^2(k_m\frac{\nu-\nu'}{2})=\begin{cases}0 & (4n)\nu\\ 1 & (4n+2)\nu\end{cases}$$

$$Y^*(\nu_i,\theta)=\frac{1}{\sqrt{2\pi}}[-\frac{1}{4\pi}Y(\nu_i,\theta)-\frac{1}{\pi^2\nu}\sum_{j=1}^{N}\frac{Y(\nu_j,\theta)}{(j-i)^2}]\ ,\ \ j-i=(2n+1)\ ,\ \text{add}.$$

The steps of the program are:

(i) interpolation $Y(\nu,\theta)=Y(\xi,\theta)$;

(ii) summation

$$Y^*(\nu_i,\theta)=-\frac{1}{\sqrt{2\pi}}[\frac{1}{4\nu}Y(\nu_i,\theta-\cdots];$$

(iii) interpolation $\hat{Y}(x\cos\theta+y\sin\theta,\theta)=Y^*(\nu,\theta)$;

(iv) backprojection

$$\Phi(x,y)=\frac{\pi}{m_\theta}\sum_{l=1}^{m_\theta}\hat{Y}(x\cos\theta_l+y\sin\theta_l,\theta_l).$$

The reference (Hanson 1980) discusses the optimality of the filtered backprojection algorithm.

Additional developments in CT include:

i) understanding the details of the interaction of the incident beam with the biological tissues of the body (e.g., scattering, absorption, etc for X-ray, and other actions for Ultrasound and MRI);

ii) improving algorithms for handling the various "artifacts" (e.g., hardening, scattering, partial volume effect, etc) powerfully;

iii) improving hardware design to decrease the "artifacts" which have something to do with detection system;

iv) optimizing for low dosage and easy operation clinically.

1.2) Positron Emission Tomography (PET)

The general formula of PET analogous to Eqs.(2.3) and (2.4) of X-ray tomography is expressed as:

$$\tilde{Y}(\xi,\theta)=e^{-\int u(x,y)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy}\int\Phi(x,y)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy \qquad (2.28)$$

where $u(...)$ is the attenuation function and $\Phi(...)$ the radioisotope concentration distribution within the scanning medium.

If $u(x,y)$ is known and letting

$$Y(\xi,\theta)=\tilde{Y}(\xi,\theta)\, e^{\int u(x,y)\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy}$$

then there is (Huang et al 1979):

$$Y(\xi,\theta)=\int\Phi(x,y)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy \qquad (2.29)$$

which is identical to Eq.(2.4). So, all the transform methods mentioned before are valid for searching the solution of Eq.(2.29).

The discrete form is given below:

$$\tilde{Y}(\xi,\theta)=\sum_j \phi_j e^{-\int u(x,y)\delta(\xi-x\cos\theta-y\sin\theta)dxdy}\int b_j(x,y)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy$$

$$=\sum_j \phi_j\, e^{-\int u(x,y)d\eta}\int b_j(x,y)\,d\eta\,.$$

where Eq.(2.5) has been used.

For the i-th projection ray $\tilde{Y}_i$ at projection angle θ, there results in an identical equation as Eq.(2.6) of

$$\tilde{Y}_i=\sum_j \phi_j[e^{-\int u(x,y)\,d\eta_i}\int b_j(x,y)\,d\eta_i]=\sum_j R_{ij}\phi_j \qquad (2.30)$$

where

$$R_{ij}=e^{-\int u(x,y)d\eta_i}\int b_j(x,y)\,d\eta_i\,.$$

More discussions about PET are in references (Hoffman et al 1979, Huang et al 1980, Hoffman et al 1981 and Mazziotta et al 1981).

1.3) Single Photon Emission Computerized Tomograohy (SPECT)

The general expression for the projected data from an emitting medium within the body is:

$$Y(\xi,\theta)=\int\int\Phi(x,y)\,a(x,y,\xi,\theta)\,\delta(\xi-x\cos\theta-y\sin\theta)\,dxdy \qquad (2.31)$$

$$=\int d\eta\,\Phi(x,y)\, e^{-\int_{\omega_{xy}} u(\eta')d\eta'} \qquad (2.32)$$

where $u(...)$ and $\Phi(...)$ were defined in Eq.(2.28) and

$$a(x,y,\xi,\theta)=e^{-\int\int_{\omega_{xy}} u(x',y')\delta(\xi-x'\cos\theta-y'\sin\theta)\,dx'dy'}\,.$$

Note that Eq.(2.32) is more complicated than Eq.(2.4) due to the non-linear attenuation factor.

Using the transform relations Eq.(2.11), Eq.(2.31) is written as:

$$Y(\xi,\theta)=\int\Phi(\xi\cos\theta-\eta\sin\theta,\xi\sin\theta+\eta\cos\theta)\,a(\xi,\eta,\theta)\,d\eta \tag{2.33}$$

where

$$a(\xi,\eta,\theta)=e^{-\int_{\xi}u(\xi\cos\theta-\eta'\sin\theta,\xi\sin\theta+\eta'\cos\theta)\,d\eta'}.$$

The discrete form is:

$$Y_i=\sum_j\phi_j\int d\eta_i\,b_j(x,y)\exp[-\int_{(x,y)}u(\eta')\,d\eta_i'\,]=\sum_j R_{ij}\phi_j \tag{2.34}$$

where

$$R_{ij}=\int d\eta_i\,b_j(x,y)\exp[-\int_{(x,y)}u(\eta')\,d\eta_i'\,].$$

A). Backprojection

$$[B_v Y\,](x,y)=\int_0^{2\pi}Y(x\cos\theta+y\sin\theta,\theta)\frac{d\theta}{a(x,y,x\cos\theta+y\sin\theta,\theta)}. \tag{2.35}$$

B). The adjoint attenuated Radon transform (Gullberg 1979)

$$[A_v^{*}\,Y](x,y)=\frac{1}{W_a(x,y)}\int_0^{2\pi}[\,d\theta\,Y(x\cos\theta+y\sin\theta)\,a(x,y,x\cos\theta+y\sin\theta,\theta) \tag{2.36}$$

$$\times W_a(x\cos\theta+y\sin\theta,\theta)\,]$$

where $W_a(...)$ is the weight function satisfying:

$$\int U_i(x,y)\,U_j(x,y)\,W_a(x,y)\,dxdy=\delta_{ij}$$

and $U_i(...)$ is an eigenfunction of operator $A_v^{*}A_v$ corresponding to an eigenvalue λ_i.

For u constant function,

$$a(x,y,\xi,\theta)=e^{(-x\sin\theta+y\cos\theta)u}.$$

a). The convolution solution

$$\Phi(x,y)=\int_0^{2\pi}d\theta\int_{-\infty}^{\infty}dt\;C_v(x\cos\theta+y\sin\theta-t)\,Y(t,\theta)\,e^{(x\sin\theta-y\cos\theta)u} \tag{2.37}$$

where $C_v(...)$ is a function which is convoluted with the data before backprojection.

b). Fourier transform solution

$$\Phi(r,\alpha)=\int_0^{\pi} d\theta \int_{-\infty}^{\infty} |k|\, dk\, Y\left(\sqrt{k^2+\left(\frac{\mu}{2\pi}\right)^2}, \theta + i \sinh^{-1}\left(\frac{\mu}{2\pi k}\right)\right) e^{i\, 2\pi\mu\, k\, r\, (\alpha-\theta)} \qquad (2.38)$$

where

$$Y(k,\theta)=F\{Y(\xi,\theta)\}$$

Some other methods are indicated below:

(i) the reprojection method (Walters et al 1981);

(ii) the simultaneous iterative algorithm (Censor et al 1979);

(iii) the two-step procedure method (Chang 1978 and 1979);

(iv) the PSF convolution method (Hsieh et al 1976);

(v) the geometric method (Keyes 1976);

(iv) the least-square iterative technique (Budinger et al 1974).

The fundamental differences between the TCT and ECT (or PET, SPECT) is that TCT searches for the attenuation distribution within the body and displays the anatomic structure of the body, in which source position and strength are known, while ECT looks for the distribution of the emission and then can infer the status of physiological functions in the body. In another words, it reveals the distribution of radioactive tracers that can often indicate various aspects of physiological functions (Knoll 1983).

Comparing to SPECT, PET has the advantages of more accurate photon attenuation correction, more restrictive collimation and more uniform sensitivity and resolution with depth (Phelps et al 1975 and Snyder et al 1981). SPECT has an advantage in the availability of radionuclides and higher detection efficiency (Budinger 1980).

The assumptions of a linear response of the detection system over dynamic count rate range and of uniform detection response at the same depth are indicated in the both cases of TCT and ECT.

Further developments in ECT might be anticipated in (Phelps 1977):

i) expand technical effort;

ii) clear definition of purpose;

iii) clinic results that provide information not obtainable by other techniques and clearly demonstrate that its importance exceeds its cost.

2) IMAGE RESTORATION

A commonly used expression of the relation reflecting noise-free output of an apparatus is the Fredholm integral equation of the first kind (Carley et al 1979):

$$Y(\xi,\eta)=\int R(\xi,\eta,x,y)\,\Phi(x,y)\,dxdy \tag{2.39}$$

where $R(...)$ is the point source response function (PSF).

For noise-associated output, Eq(2.39) becomes:

$$Y(\xi,\eta)=\int R(\xi,\eta,x,y)\,\Phi(x,y)\,dxdy+N(\xi,\eta)\,. \tag{2.40}$$

So, in practice, Eq.(2.40) gives the perturbed form of the first-order Fredholm equation which, if the measurement can not be refined so as to make $R(\xi,\eta,x,y)$ approximate a $\delta(\xi-x,\eta-y)$ function, must be solved for $\Phi(x,y)$ with $R(\xi,\eta,x,y)$ and $Y(\xi,\eta)$ known.

If the imaging system is space-invariant, there is:

$$Y(\xi,\eta)=\int R(\xi-x,\eta-y)\,\Phi(x,y)\,dxdy+N(\xi,\eta) \tag{2.41}$$

i.e., searching for the solution $\Phi(x,y)$ becomes a deconvolution problem.

The function $R(\xi-x,\eta-y)$, convolution data $Y(\xi,\eta)$ and source distribution $\Phi(x,y)$ are shown in Fig.(2.3).

In matrix notation, Eqs.(2.40) and (2.41) respectively are expressed as:

$$\mathbf{Y}=R\,\Phi \tag{2.42}$$

and

$$\mathbf{Y}=R\,\Phi+\mathbf{N} \tag{2.43}$$

where the data vector $\mathbf{Y}$ and source vector Φ are formed from the two-dimensional (or 3-d) experimental data array and source distribution array respectively by use of column scanning and R is a block Toeplitz matrix (Hall 1979).

The major problems in obtaining a solution to Eq.(2.40) arise from the perturbing nature of the noise $N(\xi,\eta)$ and the smoothing function of $R(\xi,\eta,x,y)$.

For oscillating $\Phi(x)=\cos(\omega x)$, for example, and smoothing R in one dimensional case, there is:

$$Y_0(\xi)=\int R(\xi,x)\cos(\omega x)\,dx\longrightarrow 0\ ,\quad \text{as}\quad \omega\longrightarrow\infty,$$

in this case, the $Y_0(\xi)$ becomes arbtrarily small as ω becomes arbitrarily large, and the noise $N(\xi)$

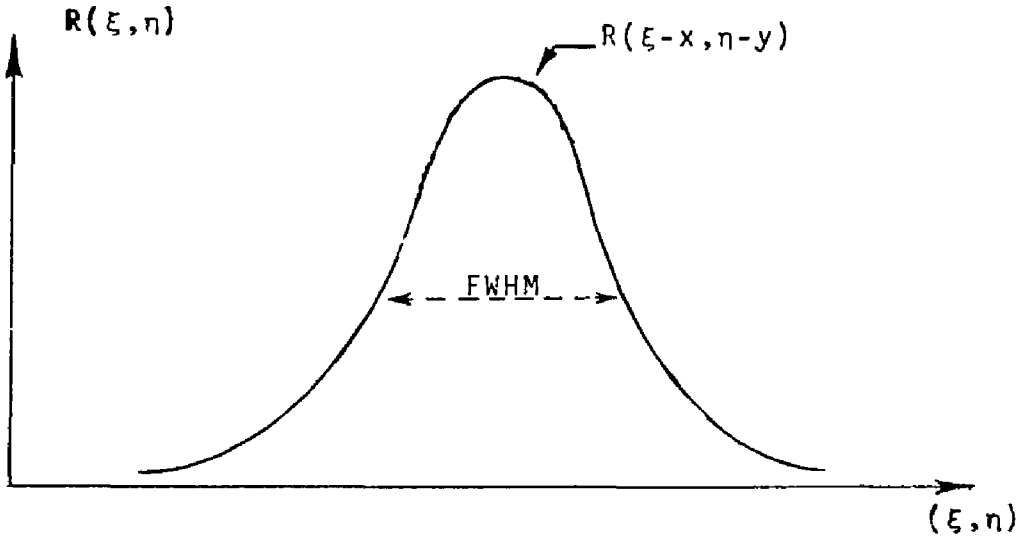
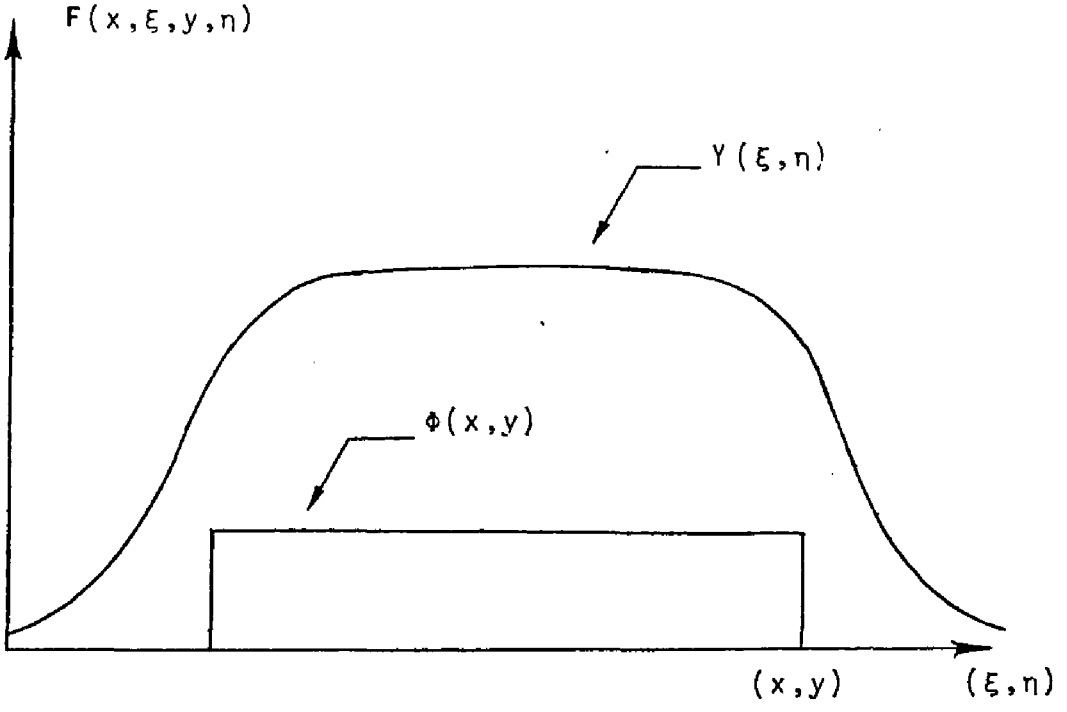


Figure 2.3: (a) A sagittal plane of a point source response function, PSF; or a line source response function, LSF.



(b) One dimensional source distribution, $\Phi(x,y)$, and "convolution" data distribution, $Y(\xi,\eta)$, where

$$Y(\xi,\eta)=\int_0^\infty R(\xi-x,\eta-y)\ \Phi(x,y)\ dx\ dy$$

dominates:

$$Y(\xi)=\int R(\xi,x)\,\Phi(x)\,dx+N(\xi)=Y_0(\xi)+N(\xi).$$

Since the noise is unknown, unique solution may not be possible. Any function $E(x)$ which satisfies $\int R(\xi,x)\,E(x)\,dx=0$ either exactly or approximately may be added to $\Phi(x)$ without greatly affecting its "convoluted" value $Y(\xi)$.

Random high-frequency noise and antisymmetric spikes of period much less than the width of $R(...)$ are included in the class of such functions $E(x)$, and the general effect is to permit the introduction of noise and spurious (unphysical and/or divergent) oscillations into the processed image.

In practical applications of Eq.(2.40) or deconvolution of Eq.(2.41), the data field is finite, and boundary discontinuities can lead to truncation errors in the processed results. This problem may be overcome by analytic continuation of the data near the end points or by use of a suitable background removal procedure which constrains the data points to decrease smoothly to zero at either end of the measured data.

Physically in radioisotope image restoration, the *PSF*, $\{R_{ij}\}$, includes all the factors (e.g. attenuation, scattering, electronics and collimator effects) (Iinuma et al 1967). While in ECT, or Eqs.(2.28) and (2.32), all of the factors except the electronics are considered separately (i.e., considering the attenuation scattering, etc, one by one) and the detectors are assumed as perfect.

The *PSF* function $R(\xi-x)$, as in the one dimensional case, may be decomposed into individual broadening functions, that is,

$$R(\xi-x)=\int U(\xi-x-y)\,f(y)\,dy$$
$$U(\xi-x-y)=\int V(\xi-x-y-z)\,g(z)\,dz$$

and so on.

In radioisotope image restoration, for example, the function $U(...)$ can describe the tissue broadening and the function $V(...)$ may represents the detector electronic broadening, etc.

The transform methods used for the solutions of convolution Eq.(2.41) are given below and the iterative methods for Eq.(2.43) will be discussed later.

2.1). Least-squares Techniques:

It is desired to find that solution Φ^* which minimizes function

$$g(\Phi)=(\mathbf{Y}-R\,\Phi)^{\tau}\Lambda^{-1}(\mathbf{Y}-R\,\Phi)$$

and is given by the normal equation of

$$R^{\tau}\Lambda^{-1}R\,\Phi^{*}=R^{\tau}\Lambda^{-1}\mathbf{Y} \tag{2.44}$$

where Λ is the covariance matrix for $\mathbf{Y}$.

By use of the Fourier transform of

$$\Psi_F(k_x,k_y)=F\{\Phi(x,y)\}\ ,\quad \Phi(x,y)=F^{-1}\{\Psi_F(k_x,k_y)\}$$

to the convolution Eq.(2.41) for $\mathbf{N}=0$ or Eq.(2.44) for space invariant R, there results:

$$\mathbf{Y}_F(k_x,k_y)=R_F(k_x,k_y)\,\Phi_F(k_x,k_y)$$

or (Andrews et al 1977)

$$\Phi_F(k_x,k_y)=\frac{\mathbf{Y}_F(k_x,k_y)}{R_F(k_x,k_y)} \tag{2.45}$$

and the solution is:

$$\Phi^{*}=F^{-1}\{\Phi_F(k_x,k_y)\}\,.$$

2.2). Regularization Techniques:

That solution Φ^{*} is sought which minimizes function of (Ernest 1979)

$$g(\Phi)=(\mathbf{Y}-R\,\Phi)^{\tau}(\mathbf{Y}-R\,\Phi)+a\,(D\,\Phi)^{\tau}(D\,\Phi) \tag{2.46}$$

and is determined by equations of

$$(R^{\tau}R+a\,D^{\tau}D)\,\Phi^{*}=R^{\tau}\mathbf{Y} \tag{2.47}$$

where D is a matrix within the space of $\{R\ \varepsilon IR^{n}\}$ and a is a numerical parameter.

Using the F.T. and noting that R is space invariant, then there is:

$$|R_F(k_x,k_y)|^2\,\mathbf{Y}_F(k_x,k_y)=(|R_F(k_x,k_y)|^2+a\,|D_F(k_x,k_y)|^2)R_F(k_x,k_y)\,\Phi_F(k_x,k_y)$$

or (Price 1982)

$$\Phi_F=\frac{|R_F|^2}{|R_F|^2+a\,|D_F|^2}\,\frac{\mathbf{Y}_F}{R_F}\,. \tag{2.48}$$

and the solution Φ^{*} is given by inverse F.T. of $\Phi_F(k_x,k_y)$.

2.3). Quadratic Optimization Techniques:

It is expressed as (Herman et al 1976):

$$g(\Phi)=a\,(\mathbf{Y}-R\,\Phi)^{\tau}A\,(\mathbf{Y}-R\,\Phi)+(\Phi-\bar{\Phi})^{\tau}(b\,B+c\,C^{-1})\,(\Phi-\bar{\Phi}) \qquad (2.49)$$
$$=\text{minimum}$$

where $\bar{\Phi}$ is the average of Φ and A, B and C are three symmetric positive definite matrices, and there results in:

$$a\,(R^{\tau}A\,R\,\Phi^{*}-R^{\tau}A\,\mathbf{Y})+(b\,B+c\,C^{-1})\,(\Phi^{*}-\bar{\Phi})=0$$

or

$$(a\,R^{\tau}A\,R+b\,B+c\,C^{-1})\,\Phi^{*}=a\,R^{\tau}A\,\mathbf{Y}+(b\,B+c\,C^{-1})\,\bar{\Phi}.$$

The F.T. generates:

$$(a\,|R_F|^2A_F+b\,B_F+c\,C_F^{-1})\,\Phi_F=a\,R_F^{*}A_F\,\mathbf{Y}_F+(b\,B_F+c\,C_F^{-1})\,\bar{\Phi}_F$$

or

$$\Phi_F=\frac{a\,R_F^{*}A_F\,\mathbf{Y}_F}{a\,|R_F|^2A_F+b\,B_F+c\,C_F^{-1}}+\frac{(b\,B_F+c\,C_F^{-1})\,\bar{\Phi}_F}{a\,|R_F|^2A_F+b\,B_F+c\,C_F^{-1}}$$
$$=\frac{|R_F|^2}{|R_F|^2+(b\,B_F+c\,C_F^{-1})/(aA_F)}\,\frac{\mathbf{Y}_F}{R_F}+W_F(\ldots). \qquad (2.50)$$

and the inverse F.T. of $\Phi_F(k_x,k_y)$ determines the solution Φ^{*}.

2.4). Minimum Mean-square-error Techniques

The error of estimated $\hat{\Phi}$ and actual Φ is:

$$\varepsilon=\Phi-\hat{\Phi}$$

This technique is to minimize (Andrews et al 1977)

$$<\varepsilon^{\tau}\varepsilon>=<(\Phi-\hat{\Phi})^{\tau}(\Phi-\hat{\Phi})>$$

with respect to the L, subject to

$$\hat{\Phi}=L\,\mathbf{Y} \qquad (2.51)$$

where L stands for the "pseudo-inverse" matrix of R and

$$\mathbf{Y}=R\,\Phi+\mathbf{N}.$$

The derivation steps are:

$$
\begin{aligned}
<\varepsilon^{\tau}\varepsilon> &= <(\Phi - L\,Y)^{\tau}(\Phi - L\,Y)> = <\Phi^{\tau}\Phi - \Phi^{\tau}L\,Y - Y^{\tau}L^{\tau}\Phi + Y^{\tau}L^{\tau}L\,Y> \\
&= <\Phi^{\tau}\Phi - \Phi^{\tau}L\,(R\,\Phi + N) - (R\,\Phi + N)^{\tau}L^{\tau}\Phi + (R\,\Phi + N)^{\tau}L^{\tau}L\,(R\,\Phi + N)> \\
&= <\Phi^{\tau}\Phi> - <\Phi^{\tau}LR\,\Phi> - <\Phi^{\tau}L\,N> - <\Phi^{\tau}R^{\tau}L^{\tau}\Phi> - <N^{\tau}L^{\tau}\Phi> \\
&\quad + <\Phi^{\tau}R^{\tau}L^{\tau}LR\,\Phi> + <\Phi^{\tau}R^{\tau}L^{\tau}L\,N> + <N^{\tau}L^{\tau}LR\,\Phi> + <N^{\tau}L^{\tau}L\,N> \\
&= <\Phi^{\tau}\Phi> - <N^{\tau}L^{\tau}L\,N> - 2<\Phi^{\tau}LR\,\Phi> - 2<\Phi^{\tau}L\,N> \\
&\quad + 2<N^{\tau}L^{\tau}L\,R\,\Phi> + <\Phi^{\tau}R^{\tau}L^{\tau}L\,R\,\Phi>.
\end{aligned}
$$

For N and Φ independent, then,

$$<\Phi^{\tau}L\,N> = 0\,, \qquad <N^{\tau}L^{\tau}L\,R\,\Phi> = 0 \tag{2.52}$$

and

$$
\begin{aligned}
<\varepsilon^{\tau}\varepsilon> &= <\Phi^{\tau}\Phi> + <N^{\tau}L^{\tau}L\,N> - 2<\Phi^{\tau}L\,R\,\Phi> + <\Phi^{\tau}R^{\tau}L^{\tau}L\,R\,\Phi> \\
&= V(\Phi) + L^{\tau}V(N)L - 2L\,R\,V(\Phi) + L\,R\,V(\Phi)R^{\tau}L^{\tau}
\end{aligned} \tag{2.53}
$$

where $V(...)$ is the variance function.

From $\dfrac{\partial <\varepsilon^{\tau}\varepsilon>}{\partial L} = 0$,

$$[V(N) + R\,V(\Phi)R^{\tau}]\,L = R\,V(\Phi)$$

or

$$L = \frac{V(\Phi)R^{\tau}}{V(N) + R\,V(\Phi)R^{\tau}}\,. \tag{2.54}$$

The estimated solution is therefore (Andrews et al 1977):

$$\hat{\Phi} = \frac{|R|^2}{|R|^2 + V(N)/V(\Phi)}\,\frac{Y}{R}\,. \tag{2.55}$$

The F.T. of Eq.(2.55) for space invariant R is:

$$\Phi_F = \frac{|R_F|^2}{|R_F|^2 + A_F^V/B_F^V}\,\frac{Y_F}{R_F} \tag{2.56}$$

and the inverse F.T of $\Phi_F(k_x, k_y)$ results in the solution Φ^*, where A_F^V and B_F^V are the F.T. of $V(N)$ and $V(\Phi)$ respectively.

2.5). Homomorphic Estimation

When the power spectrum of the processed image is constrained to be the same as that of the unprocessed image (Andrew et al 1977), i.e.,

$$B_F^V(k_x,k_y,\hat{\Phi})=B_F^V(k_x,k_y,\Phi)$$

and

$$L_F(k_x,k_y)\,\mathbf{Y}_F=\hat{\Phi}_F \tag{2.57}$$

then,

$$|L_F|^2\bar{A}_F^V=B_F^V$$

i.e.,

$$|L_F|=\sqrt{B_F^V/\bar{A}_F^V}.$$

For a stationary random process, the power spectrum of $\mathbf{Y}_F$, $\bar{A}_F^V$, is:

$$\bar{A}_F^V=<\mathbf{Y}_F(k_x,k_y)\mathbf{Y}_F^*(k_x,k_y>=|R_F|^2B_F^V+A_F^V$$

so (Cannon 1974),

$$\hat{\Phi}_F=\sqrt{B_F^V/(|R_F|^2B_F^V+A_F^V)}\;\mathbf{Y}_F\,. \tag{2.58}$$

2.6) General Filtering Techniques:

One of the important sources of image degradation is noise. One way to suppress the effect of noise is to filter out the high frequency components in the signal (Barrett et al 1981 and Beauchamp et al 1979).

Let the first linear system be R and the filter be F, then the first output is, as in one dimensional case,

$$Y(\xi)=R(\xi)*\Phi(\xi)=\int R(\xi-x)\,\Phi(x)\,dx$$

and think of it as input for F, the second output is:

$$\begin{aligned}\hat{Y}(\xi)&=F(\xi)*[R(\xi)*\Phi(\xi)]=F(\xi)*[\int dx\,R(\xi-x)\,\Phi(x)]\\&=\int dx\,\Phi(x)\,[\int dt\,F(\xi-t)\,R(t-x)]\end{aligned} \tag{2.59}$$

let $z=t-x$, then,

$$\hat{Y}(\xi)=\int dx\,\Phi(x)\,[\int dz\,F((\xi-x)-z)\,R(z)]=\int dx\,\Phi(x)\,F^{(R)}(\xi-x) \tag{2.60}$$

where $F^{(R)}(\xi-x)=F(\xi)*R(\xi)$.

The F.T. of Eq.(2.60) gives the result:

$$\mathbf{Y}_F(k)=F_F^{(R)}(k)\,\Phi_F(k)=\mathbf{F}_F(k)\,R_F(k)\,\Phi_F(k). \tag{2.61}$$

To overcome the degradation, the filter inverse is required, i.e., from Eq.(2.59),

$$\hat{Y}(\xi)=[F(\xi)*R(\xi)]*\Phi(\xi)$$

letting

$$F(\xi)*R(\xi)=\delta(\xi) \tag{2.62}$$

so that

$$\hat{Y}(\xi)=\delta(\xi)*\Phi(\xi)=\Phi(\xi).$$

In the frequency domain, $F_F^{(R)}(k)=\mathbf{F}_F(k)\,R_F(k)=1$, with $R_F(k)$ being known, then,

$$\mathbf{F}_F(k)=R_F^{-1}(k)\ ,\ \text{and}\quad F(\xi)=\int dk\,R_F^{-1}(k)\,e^{i2\pi k\xi}. \tag{2.63}$$

Eq.(2.62) is unrealistic in most practical situations and so $\mathbf{F}_F(k)$ is chosen depending on the individual case. For example, Eq.(2.61) can be written as:

$$\Phi_F(k)=\mathbf{F}_F^{-1}\,\frac{\mathbf{Y}_F(k)}{R_F(k)} \tag{2.64}$$

which has the equivalent form as Eqs.(2.48), (2.50), (2.56) and (2.58).

Another source of image degradation is the low resolution of a detection system which does not adequately respond to the high frequency components in the input. Without improvement in the detection system, noise suppression will cost some sacrifice of resolution. As a compromise, the choice of function $\mathbf{F}_F^{-1}$ is extended to modulate the ratio of $\mathbf{Y}_F(k)/R_F(k)$.

Now let's take a look at Eq.(2.45). It illustrates a situation in which R_F goes to zero faster than $\mathbf{Y}_F$. Since Y_i contains noise and R does not approximate a δ function, $\mathbf{Y}_F$ has a more uniform distribution. Consequently, Φ_F has accentuated high frequency components and if R_F has zeros in the range of interest, then Eq.(2.45) breaks down. To overcome this ill-posed problem, a window function $W(k_x,k_y)$ is introduced as:

$$\Phi_F(k_x,k_y)=W(k_x,k_y)\,\frac{\mathbf{Y}_F(k_x,k_y)}{R_F(k_x,k_y)} \tag{2.65}$$

to suppress the high frequency components due to the noise, i.e., when $R_F(k_x,k_y)$ goes to very small, $W(k_x,k_y)$ (or equivalently, $\mathbf{F}_F^{-1}(k_x,k_y)$) is almost vanishing and modulate the values of ratio $\mathbf{Y}_F/R_F$ in the frequency range of interest.

By use of the window function, Eq.(2.65) becomes well-posed but the solution remains non-unique in the sense that arbitrary parameters appear in the subjective chose window function $W(k_x,k_y)$.

Note that the use of function $W(k_x,k_y)$ may result in a reduction of noise propagation in the restored image but at the expense of some resolution enhancement, that is, the noise is suppressed and the resolution is made poorer as discussed before. So the choice of $W(k_x,k_y)$ is somewhat critical for optimal image restoration.

Although with the use of $W(k_x,k_y)$, there may be two types of oscillation in applying Eq.(2.65). One is base-line oscillation which arises very readily in the case of poor signal to noise ratio (S/N) or inaccurate knowledge of the function PSF. Another one is diffraction-like oscillation which may arise owing to inproper use of the window function $W(k_x,k_y)$.

Various window functions which have been used are listed as following:

a). Rectangular window

$$W(k_x,k_y)=\begin{cases} 1 & 0\le k_x\le k_{xN}\ ,\ \ 0\le k_y\le k_{yN} \\ 0 & \text{otherwise}\ . \end{cases}$$

where subscript N denotes the Nyquist frequency.

b). Hann window (Webb et al 1985)

$$W(k_x,k_y)=\begin{cases} \frac{1}{4}[1+\cos(\frac{\pi k_x}{k_{xN}})][1+\cos(\frac{\pi k_y}{k_{yN}})] & 0\le k_x\le k_{xN}\ ,\cdots \\ 0 & \text{otherwise}\ . \end{cases}$$

c). Lagrange parameter window (Phillips 1962)

$$W(k_x,k_y)=\frac{1}{1+\eta(2\pi)^4(k_x^2+k_y^2)^2/|R_F|^2}$$

where η is an adjustable parameter.

d). Hamming window (Hamming 1977)

$$W(k_x,k_y)=\begin{cases} [0.54+0.46\cos(\frac{\pi k_x}{k_{xN}})][0.54+0.46\cos(\frac{\pi k_y}{k_{yN}})] & 0\le k_x\le k_{xN}\ ,\cdots \\ 0 & \text{otherwise}\ . \end{cases}$$

e). Parzen window

$$W(k_x,k_y)=\begin{cases}[1-6(\frac{k_x}{k_{xN}})^2(1-\frac{k_x}{k_{xN}})][1-6(\frac{k_y}{k_{yN}})^2(1-\frac{k_y}{k_{yN}})] \\ [2(1-\frac{k_x}{k_{xN}})^3][2(1-\frac{k_y}{k_{yN}})^3] \\ 0\end{cases}$$

for

$$\begin{cases}0\le k_x\le k_{xN}/2\ ,\ \ 0\le k_y\le k_{yN}/2 \\ k_{xN}/2<k_x\le k_{xN}\ ,\ \ k_{yN}/2<k_y\le k_{yN} \\ \text{otherwise}\ .\end{cases}$$

f). Shepp and Logan window (Shepp et al 1974)

$$W(k_x,k_y)=\begin{cases}[\frac{sin(\pi k_x/2k_{xN})}{\pi k_x/2k_{xN}}][\frac{sin(\pi k_y/2k_{yN})}{\pi k_y/2k_{yN}}] & 0\le k_x\le k_{xN}\ ,\ \ 0\le k_y\le k_{yN} \\ 0 & \text{otherwise}\ .\end{cases}$$

g). Gaussian window

$$W(k_x,k_y)=\begin{cases}e^{-\pi k_x^2/\delta_x^2-\pi k_y^2/\delta_y^2} & 0\le k_x\le k_{xN}\ ,\ \ 0\le k_y\le k_{yN} \\ 0 & \text{otherwise}\end{cases}$$

where δ_x and δ_y are resolution parameters and

$$\delta_x^2=\frac{4\,ln(2)}{\pi(FWHM)_x^2}\ ,\ \ \delta_y^2=\frac{4\,ln(2)}{\pi(FWHM)_y^2}\ .$$

h). Butterworth window (Hamming 1977)

$$W(k_x,k_y)=\frac{1}{1+(k_x/k_{xN})^{2n}}\cdot\frac{1}{1+(k_y/k_{yN})^{2n}}$$

where $n \approx 5$ to 350 and $k_{xN} \approx 0.25$ to 1.

i). Wiener window (Walkup et al 1974 and Jansson 1984)

$$W(k_x,k_y)=\frac{|R_F|^2}{|R_F|^2+a\,A_F^V/B_F^V}\ .$$

Comparisons of the window functions are shown in Figs.(2.4), (2.5) and (2.6). Since the Lagrange parameter, gaussian and Wiener windows are dependent on the PSF, in drawing the figures, it was assumed that

$$R(\xi - x) = e^{-\frac{\ln(2)}{T^2}(\xi - x)^2} \tag{2.66}$$

where $T = FWHM/2$.

The F.T. of $R(\xi - x)$ is:

$$R_F(k) = e^{-(\pi k T)^2/\ln(2)}. \tag{2.67}$$

For curve Lage1, $T = 0.5$ and $\eta = 0.001$ were chosen; and Lage2 with $T = 1$ and $\eta = 0.001$. For curve Guss1, $T = 0.5$ and Guss2, $T = 1$. For curve Wien1, $T = 0.5$ and $a\, A_F^V/B_F^V = 0.1$ and for Wien2, $T = 1$ and $a\, A_F^V/B_F^V = 0.1$; while Wienx with $T = 0.5$ and $a\, A_F^V/B_F^V = 0.001$, and Wieny with $T = 1$ and $a\, A_F^V/B_F^V = 0.001$.

Fig.(2.5) exhibits the sensitive dependence of the Lagrange parameter, Gaussian and Wiener window functions on the FWHM of the PSF. For lower resolution imaging system, i.e., larger T, these windows suppress more high frequency components.

Fig.(2.6) shows that for larger value of $a\, A_F^V/B_F^V$, i.e., emphasizing the ratio of A_F^V/B_F^V and the higher frequencies of $V(N)$, the dependence of the Wiener window on the FWHM is more sensitive.

Some properties of the windows are (Gullberg et al 1981):

(i). A rectangular window gives the best resolution in the reconstructed image for perfect data but amplifies noise for data with statistical fluctuations (for not very broad PSF).

The sharp cutoff of the rectangular window gives a convolution function, $f_c(x) = F^{-1}\{k\, W(k)\}$, with a narrow central window but side lobes which continue to oscilate even for large x. This gives rise to intensity oscilations in regions of sharp contrast and thus generates artifacts in the reconstructed image.

(ii). For the Hann window, the central lobe of the convolution function is wider than the central lobe of the corresponding convolution function for the rectangular window, but its side-lobes are greatly reduced. Therefore, the reconstructed image has a smoother texture (less artifact) with a loss in resolution.

(iii). For the Hamming window, the central lobe of the convolution is greater than that of Hann window and has smaller extreme values in side-lobes than that of Hann window.

(iv). The central lobe of the Parzen window is about 30 percent wider than either the Hann or Hamming window, so the reconstructed image resolution is less than can be achieved with Hann or Hamming windows.

(v). The Shepp and Butterworth windows exhibit the effects intermediate between the Hamming and rectangular windows.

(vi). For the Gaussian, Lagrange parameter and Wiener windows, increasing the *FWHM* causes the resolution to deteriorate and suppresses statistical fluctuation. As shown in the figures, these windows are very sensitive to the *FWHM*, and drop down rapidly as the *FWHM* increases.

(vii). A window function with narrow transition band ($\Delta k = k_x - k_x$ as shown in Fig.(2.7)) in frequency space is equivalent to having a narrow central lobe for the *PSF*, which means that the reconstruction will have good resolution with high statistics, but concurrently the side-lobes for such a window function are larger, thus amplifying noise for low statistics. On the other hand, a wider transition band gives poorer spatial resolution with reduced noise amplification.

The solutions of Eq.(2.40) for space-variant imaging system were determined by Eqs.(2.44), (2.47), (2.49) and (4.54) respectively for different constraints (Angel et al 1978, Sawchuk 1974 and Robbins et al 1972). Their iterative approach forms are shown in chapters III and X.

Three dimensional image restoration techniques were indicated using focusing collimator coincidence scanning (FCCS) and focusing collimator single gamma scanning (FCSGS) (Hart et al 1977 and Hart 1968).

Comparison of image reconstruction and restoration methods were discussed by Lewitt (1983, IEEE Proc.) and Canon et al (1978, Appl. Optics).

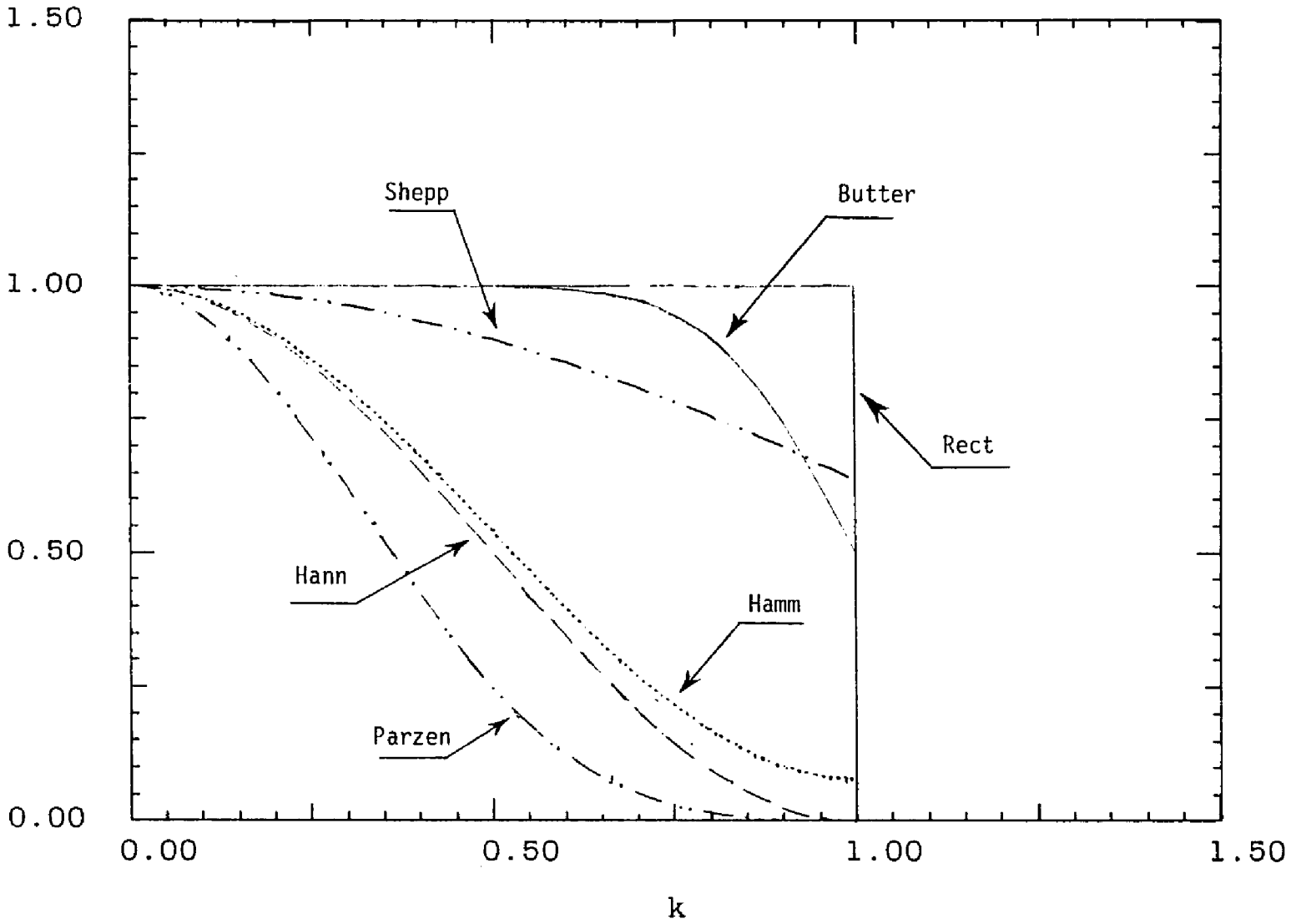


Figure2.4: Comparison of the window functions: Rect (rectangular), Hann, Hamm (hamming), Parzen, Shepp (Shepp and Logan) and Butter (Butterworth).

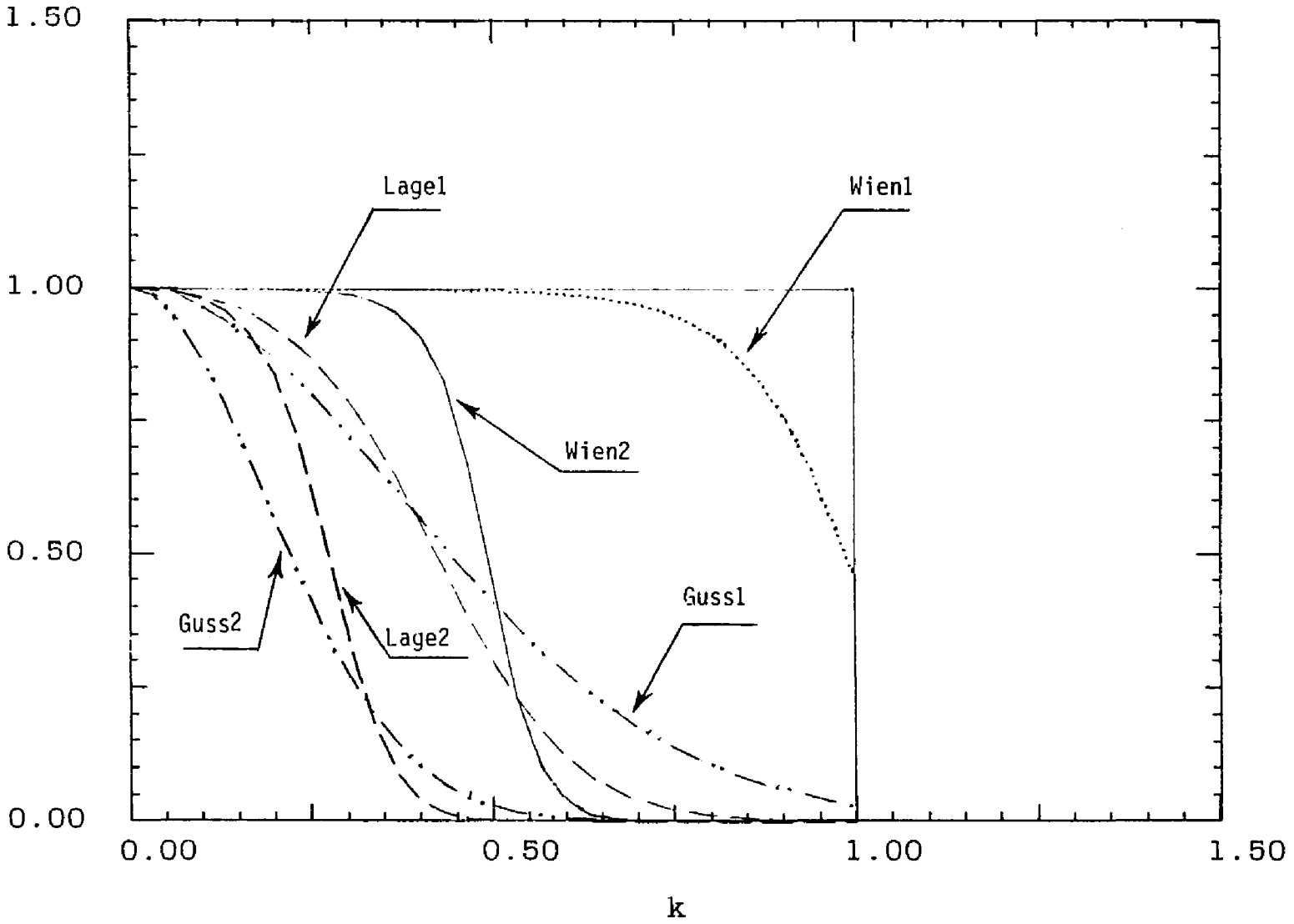


Figure 2.5: Comparison of the Lagel (Lagrange parameter), Guss (Gaussian) and Wien (Wiener) window functions; and their changes as different T of PSF.

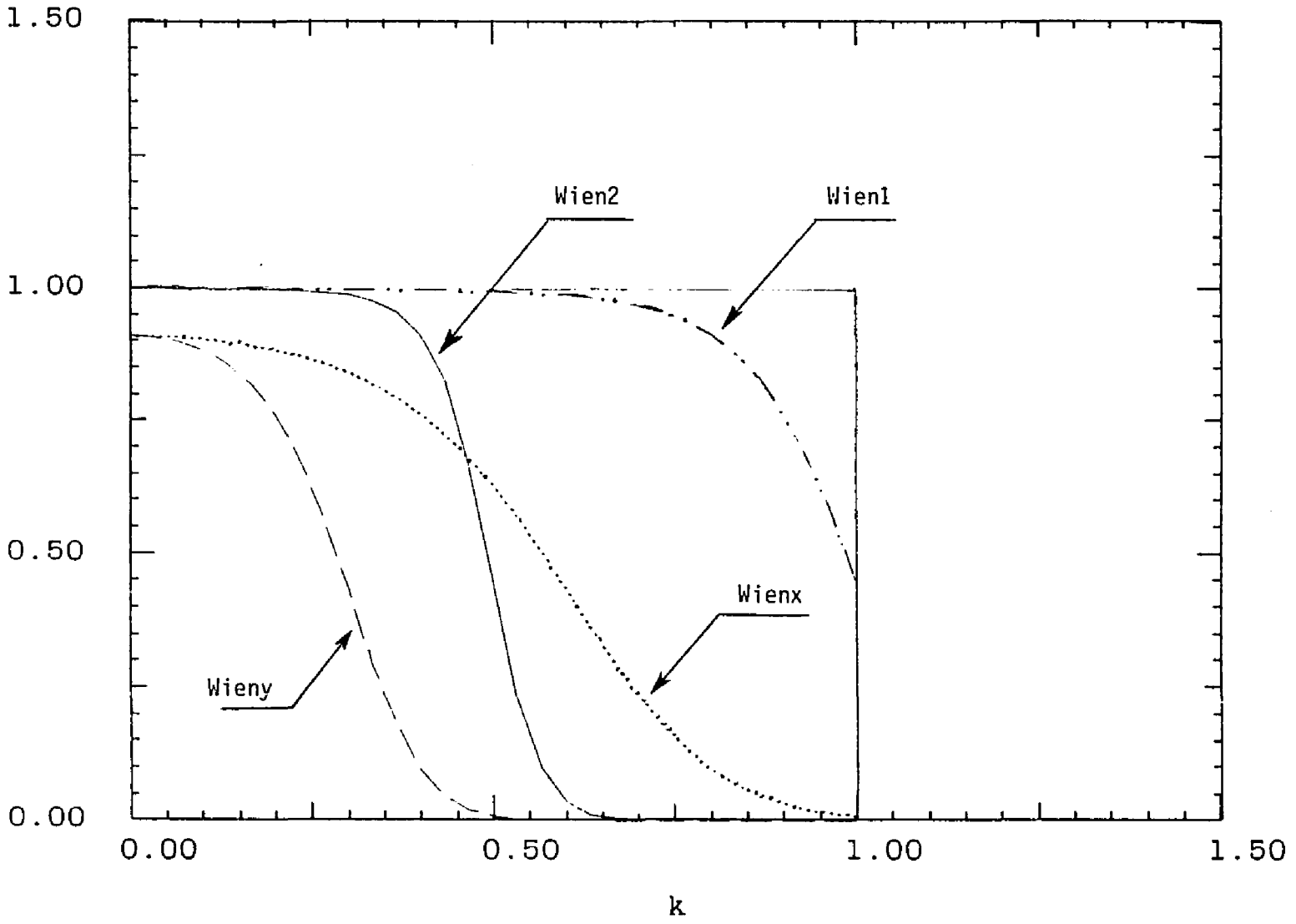


Figure 2.6: Changes of the Wiener window function as different T of PSF.

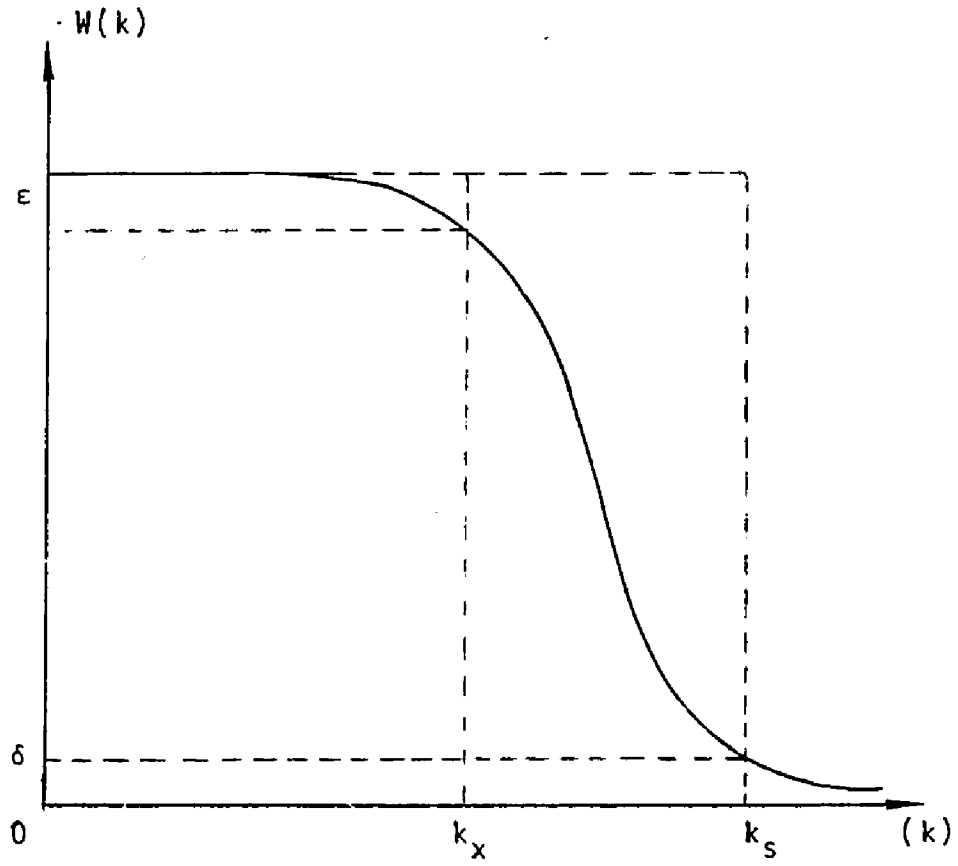


Figure 2.7: The transition band, $k_s - k_x$, of a window function W(k). The parameters ε and δ depend on the change rate of W(k) and may simply be chosen as:

$$\varepsilon = \delta = 10\ \%\ W_{max}$$

CHAPTER III
ITERATIVE TECHNIQUES FOR IMAGE PROCESSING

The system of linear equations of

$$\mathbf{Y}=R\Phi+\mathbf{N} \tag{3.1}$$

was formulated before for image reconstruction (i.e., Eqs.(2.7) --- CT, (2.30) --- PET, and (2.34) --- SPECT) and image restoration (i.e., Eq.(2.43)), where vectors $\mathbf{Y}$ and Φ have dimensions of I and J respectively, $\mathbf{N}$ has the same dimension as $\mathbf{Y}$ and R is a $I \times J$ matrix (i.e., I rows and J columns).

Some comments are given below about the solution of Φ.

(i) If $I \leq J$, that is, there are not enough measured data to provide the J independent equations, there will be many possible solutions.

(ii) due to noise and other artifacts, the measured data may be inconsistent, there is no possible solution or there is more than one solution.

(iii) the off-diagonal terms of the matrix R can be large, in which case even for non-singular matrices there can exist large errors in the calculation and unrealistic solutions.

Different iterative methods are formulated below to approach the solution subject to different constraints respectively. For example, minimizing $|R\Phi-\mathbf{Y}|^2$ results in the solution which minimizes the noise (Budinger et al 1974) and minimizing $|D\Phi|^2$ gives the smoothest solution, where D is a positive definite matrix (Price 1982).

Iterative methods have the advantage that the noise effects tend to build up slowly and linearly with each iteration since there is an inherent smoothing operation in the "convolution" $R\Phi^{(n)}$ for each iteration. Therefore some remedial action to suppress noise building-up can be taken in each iteration (Quittner 1966).

There are two factors to quantify the noise-suppression, one is the quality of the final result and another one is the rate of convergence.

The quality of the final result can be explored by comparing various solutions given by various methods.

The rate of convergence of an iterative method depends exponentially on the signal-to-noise ratio (S/N) of data $\mathbf{Y}$ (Dahlquist 1974), i.e.,

$$ln(\rho)=c\ (S/N)$$

where c is a constant and ρ denotes the rate of convergence and also depends on $FWHM$ of function R.

1). Jacobi Method

It is valid for a consistent system of linear equations of

$$\mathbf{Y} = R\,\Phi \tag{3.2}$$

and for diagonal dominate square matrix R and almost noise-free measured data. It is expressed as follows (Dahlquist et al 1974):

For any initial vector $\Phi^{(0)}$, there is:

$$\phi_k^{(n+1)} = \phi_k^{(n)} + \frac{Y_k - \sum_j R_{kj}\phi_j^{(n)}}{\sum_i R_{ik}} \tag{3.3}$$

or in Gauss-Seidel form:

$$\phi_k^{(n+1)} = \phi_k^{(n)} + \frac{Y_k - \sum_{j=1}^{k-1} R_{kj}\phi_j^{(n+1)} - \sum_{j=k} R_{kj}\phi_j^{(n)}}{\sum_i R_{ik}}.$$

The multiplicative form or the Gold's ratio method (Jansson 1984) is:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{Y_k}{\sum_j R_{kj}\phi_j^{(n)}}. \tag{3.4}$$

2). ART Method (algebraic reconstruction techniques)

For the consistent system of equations (3.2) at the n-th iteration, the solution vector $\Phi^{(n)}$ satisfies the equation of

$$<R_j, \Phi^{(n)}> = Y_j - r_j^{(n)} \tag{3.5}$$

where R_j is the j-th row of the square matrix R and $r_j^{(n)}$ the residual in the j-th projection.

A new solution to Φ is determined by solving the equation

$$<R_j, \Phi^{(n)} + \mathbf{q}^{(n)}> = Y_j \tag{3.6}$$

for $\mathbf{q}^{(n)}$.

From Eqs.(3.5) and (3.6), there is:

$$<R_j, \mathbf{q}^{(n)}> = r_j^{(n)} \tag{3.7}$$

and

$$\mathbf{q}^{(n)}=R_j^G\, r_j^{(n)} \tag{3.8}$$

where R_j^G denotes the generalized inverse of the row matrix R_j and is given by (Isreal et al 1974):

$$R_j^G=\frac{1}{|R_j|^2}R_j^G \tag{3.9}$$

and the new solution is (Herman et al 1976):

$$\begin{aligned}\Phi^{(n+1)}&=\Phi^{(n)}+\mathbf{q}^{(n)}\\ &=\Phi^{(n)}+\frac{R_j^G}{|R_j|^2}(Y_j-<R_j,\Phi^{(n)}>).\end{aligned} \tag{3.10}$$

For the inconsistent system of equations (3.1) due to the statistical fluctuation of Y, with

$$R^{\tau}\mathbf{N}=0 \tag{3.11}$$

which indicates that the back-projection of the errors N are constructed to be zero, there is:

$$\begin{bmatrix} R & I^{(0)} \\ 0 & R^{\tau}\end{bmatrix}\begin{bmatrix}\Phi\\ \mathbf{N}\end{bmatrix}=\begin{bmatrix}\mathbf{Y}\\ 0\end{bmatrix} \tag{3.12}$$

where $I^{(0)}$ is the unit matrix.

The steps of implementing Eq.(3.12) iteratively are:

(i) choose any initial $\Phi^{(0)}$ and $\mathbf{N}^{(0)}$;

(ii) for $j=n\ mod\ (I+J)+1$:

if $1\le j\le I$, then,

$$r_j^{(n)}=Y_j-<R_j,\Phi^{(n)}>-N_j^{(n)}$$

if $I+1<j<I+J$, then,

$$j'=I+J-j,\ \text{and}\ \ r_j^{(n)}=-<R_j^{\tau},\mathbf{N}^{(n)}>;$$

(iii) if $1\le j\le I$, then,

$$\mathbf{q}^{(n)}=r_j^{(n)}R_j^{\tau}/(|R_j|^2+1)$$
$$V_j^{(n)}=r_j^{(n)}/(|R_j|^2+1)$$

and

$$V_i^{(n)}=0,\ \ i=1,2,\ldots,I,\ i\neq j$$

if $I+1\le j\le I+J$, then,

$$j' = I + J - j \text{ , and } \mathbf{q}^{(n)} = 0$$

and

$$V_j^{(n)} = r_j^{(n)} \, (R^\tau)_j^\tau / |R_j^\tau|^2 ;$$

(iv) compute

$$\Phi^{(n+1)} = \Phi^{(n)} + \mathbf{q}^{(n)}$$

and

$$\mathbf{N}^{(n+1)} = \mathbf{N}^{(n)} + \mathbf{V}^{(n)} ; \tag{3.14}$$

(v) junp back to (ii).

More detailed discussion about ART-type algorithms and modified ART methods are given in references (Herman et al 1976, Computer Biol. Med. and Herman 1979).

3). Ratio Relaxation Method

A multiplicative formulation of the ILST method (Budinger et al 1974) is readily derived. For the inconsistent system of Eq.(3.1), the least-square condition on

$$\varepsilon = \sum_i^I \frac{(Y_i - \sum_j R_{ij} \phi_j)^2}{\sigma_i^2} \tag{3.15}$$

results in the normal equations

$$R^\tau \Lambda^{-1} R \, \Phi^* = R^\tau \Lambda^{-1} \mathbf{Y} \tag{3.16}$$

where Λ is the covariance matrix of $\mathbf{Y}$.

The iterative procedure of searching for the least square solution is given below:

For any initial vector $\Phi^{(n)}$, the next iterated result is:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \left[\left(\frac{R^\tau \Lambda^{-1} \mathbf{Y}}{R^\tau \Lambda^{-1} R \, \Phi^{(n)}} - 1 \right) q + 1 \right] \tag{3.17}$$

note that the parameter $q \approx 1$ plays a role in adjusting the speed of convergence and stability of the calculation.

4). NLM Method (non-local method)

Since the multiplicative form of Jacobi method Eq.(3.4) is generally effective for dominant diagonal R matrices and essentially noise free data $\mathbf{Y}$ (Hart 1983), the correlation ratio $Y_k / (\sum_j R_{ij} \phi_j^{(n)})$ is

statistically weighted by

$$(\sum_i F_i^{(1)} \frac{Y_i}{\sum_j R_{ij}\phi_j^{(n)}} F_{ik}^{(2)})/(\sum_i F_i^{(1)} F_{ik}^{(2)}) \tag{3.18}$$

to approach that solution of Eq.(3.1), which is statistically smoothed by appropriately reflecting all the data elements. $F_i^{(1)}$ is a measure of the statistical reliability of the individual data measurements:

$$F_i^{(1)} = \frac{Y_i}{\sigma_i} = Y_i/\sqrt{Y_i + BKGD} \ , \text{ (i.e., signal/noise ratio)}$$

and $F_{ik}^{(2)}$ is the relative extent to which each voxel, on the basis of the last iteration, is calculated to contribute to the data value, i.e.,

$$F_{ik}^{(2)} = R_{ik}\phi_k^{(n)}/\sum_j R_{ij}\phi_j^{(n)}$$

so,

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik} (Y_i/\sum_j R_{ij}\phi_j^{(n)})^2 (Y_i + BKGD)^{-1/2}}{\sum_i R_{ik} (Y_i/\sum_j R_{ij}\phi_j^{(n)}) (Y_i + BKGD)^{-1/2}} . \tag{3.19}$$

5). Maximum Entropy Formalism

For the assumption of $\sum_j \Phi = 1$ and $\Phi \geq 0$, a maximum entropy formalism (Goldman 1955) is directed at finding that Φ^* which maximizes:

$$H = -\sum_j^J \phi_j \ln(\phi_j) \tag{3.20}$$

subject to constraint of

$$Y_i = \sum_j R_{ij}\phi_j . \tag{3.21}$$

The Lagrange function (Bertsekas 1982) is:

$$\Psi(\Phi, \nu) = -\sum_j \phi_j \ln(\phi_j) - a \sum_i \nu_i (Y_i - \sum_j R_{ij}\phi_j) \tag{3.22}$$

where $\{\nu_i\}$ are the Lagrange multipliers and a is an adjustable parameter.

From $\frac{\partial \Psi}{\partial \phi_k}|_{\Phi=\Phi^*} = 0$, the Φ^* is given by:

$$\phi_k^* = \frac{1}{e}\, e^{a\sum_i R_{ik} v_i}. \tag{3.23}$$

Substituting (3.23) to the constraint condition (3.21), there is:

$$Y_i = \frac{1}{e}\sum_j R_{ij}\, e^{a\sum_i R_{ij} v_i}. \tag{3.24}$$

Eqs.(3.23) and (3.24) determine the solutions $\{\phi_j^*\}$ and $\{v_i\}$.

Agmon et al (1979, J. Comput. Physics) discussed a feasible solution of Eqs.(3.23) and (3.24).

a). MART (multiplicative ART)

By iterative approach, there is:

$$\phi_k^{(n)} = \frac{1}{e}\, e^{a\sum_i R_{ik} v_i^{(n)}} \tag{3.25}$$

$$\phi_k^{(n+1)} = \frac{1}{e}\, e^{a\sum_i R_{ik} v_i^{(n+1)}} = \frac{1}{e}\, e^{a(\sum_i R_{ik} v_i^{(n+1)} + R_{kk} v_k^{(n+1)})} \tag{3.26}$$

and

$$\phi_k^{(n+1)} = \phi_k^{(n)}\left[\, e^{(v_k^{(n+1)} - v_k^{(n)})}\right]^{a R_{kk}} \tag{3.27}$$

if

$$R_{kk}(v_k^{(n+1)} - v_k^{(n)}) = R_{ik}\, ln(\frac{Y_i}{\sum_j R_{ij}\phi_j^{(n)}})$$

then the MART algorithm (Lent 1976) results:

$$\phi_k^{(n+1)} = \phi_k^{(n)}(\frac{Y_i}{\sum_j R_{ij}\phi_j^{(n)}})^{a R_{ik}}. \tag{3.28}$$

b). MLEM (maximum likelihood EM method)

If there exists:

$$a\, R_{kk}\,(v_k^{(n+1)} - v_k^{(n)}) = \ln\left[\frac{\sum_i R_{ik} Y_i / \sum_j R_{ij}\phi_j^{(n)}}{\sum_i R_{ik}}\right]$$

then the MLEM algorithm (Lange et al 1984) results:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik} Y_i / \sum_j R_{ij} \phi_j^{(n)}}{\sum_i R_{ik}} . \qquad (3.29)$$

The application of this formalism to image processing was indicated by Minerbo (1979, Comput. Graph. Image Processing) and Wernecke et al (1977, IEEE Trans. Computers).

c). IMENT (Iterative Maximum Entropy Technique)

Here we approximate an iterative algorithm to generate the solutions $\{\phi_j^*\}$ and $\{v_i\}$.

From Eqs.(3.21) and (3.27), there is:

$$Y_i = \frac{1}{e} \sum_j R_{ij} \phi_j^{(n)} e^{a R_{ii} (v_j^{(n+1)} - v_j^{(n)})}$$

$$= \frac{1}{e} \sum_{j \neq i} R_{ij} \phi_j^{(n)} e^{a R_{ii} (v_j^{(n+1)} - v_j^{(n)})} + \frac{1}{e} R_{ii} \phi_i^{(n)} e^{a R_{ii} (v_i^{(n+1)} - v_i^{(n)})}$$

Assume that the off diagonal term $|v_j^{(n+1)} - v_j^{(n)}|$ is much less significant than the diagonal term $|v_i^{(n+1)} - v_i^{(n)}|$, then,

$$e\, Y_i - \sum_{j \neq i} R_{ij} \phi_j^{(n)} = R_{ii} \phi_i^{(n)} e^{a R_{ii} (v_i^{(n+1)} - v_i^{(n)})}$$

and

$$v_i^{(n+1)} - v_i^{(n)} = a\, R_{ii} \ln \left[\frac{e\, Y_i - \sum_{j \neq i} R_{ij} \phi_j^{(n)}}{R_{ii} \phi_i^{(n)}} \right] \qquad (3.30)$$

Eqs.(3.27) and (3.30) are the expression of the algorithm IMENT.

6). Minimum Norm Formalism

It is to minimize the variance of voxel values of (Censor 1983)

$$\sum_j^J (\phi_j - \bar{\phi})^2$$

subject to constraint of

$$Y_i = \sum_j R_{ij} \phi_j \qquad (3.31)$$

where $\bar{\phi}$ stands for the average of $\{\phi_j\}$.

The Lagrange function for this case is:

$$\Psi(\Phi,\nu)=\sum_j(\phi_j-\bar{\phi})^2+a\sum_i\nu_i\,(Y_i-\sum_j R_{ij}\phi_j)\,.$$

The searched solution Φ^* of this formalism is given by:

$$\frac{\partial\Psi}{\partial\phi_k}\Big|_{\Phi=\Phi^*}=0$$

or

$$\phi_k^*=\bar{\phi}+\frac{a}{2}\sum_i\nu_i R_{ik}\,. \tag{3.32}$$

Substituting Eq.(3.32) into the constraint Eq.(3.31), there is:

$$Y_i=\bar{\phi}\sum_j R_{ij}+\frac{a}{2}\sum_{jl}R_{lj}\nu_l\,. \tag{3.33}$$

By iterative approach, there is:

$$\phi_k^{(n+1)}=\bar{\phi}+\frac{a}{2}\sum_i R_{ik}\,\nu_i^{(n+1)}=\phi_k^{(n)}+\frac{a}{2}\sum_i R_{ik}(\nu_i^{(n+1)}-\nu_i^{(n)})\,. \tag{3.34}$$

Based on this formalism, the relaxation methods for image reconstruction were proposed by Herman et al (1978, Comm. ACM).

a). SIRT (simultaneous iterative reconstruction techniques)

If

$$\nu_i^{(n+1)}-\nu_i^{(n)}=\frac{Y_i-\sum_j R_{ij}\phi_j^{(n)}}{\sum_i R_{ik}}$$

then (Gilbert 1972),

$$\phi_k^{(n+1)}=\phi_k^{(n)}+\frac{a}{2}\,\frac{\sum_i R_{ik}(Y_i-\sum_j R_{ij}\phi_j^{(n)})}{\sum_i R_{ik}}\,. \tag{3.35}$$

b). IMINT (iterative minimum norm techniques)

From Eqs.(3.32) and (3.33), we formulate an iterative algorithm for the solutions of $\{\phi_j^*\}$ and $\{\nu_i\}$ as below:

$$Y_i-\bar{\phi}\sum_j R_{ij}=\frac{a}{2}\sum_j(\sum_{l\neq k}R_{lj}\nu_l^{(n)})+\frac{a}{2}\nu_k^{(n+1)}\sum_j R_{kj}$$

or

$$v_k^{(n+1)} = \frac{Y_i - \bar{\phi}\sum_j R_{ij} - (\frac{a}{2})\sum_j (\sum_{l \neq k} R_{ij} v_l^{(n)})}{(a/2)\sum_j R_{kj}} \tag{3.36}$$

and

$$\phi_k^{(n)} = \bar{\phi} + \frac{a}{2}\sum_i R_{ik} v_i^{(n)}. \tag{3.37}$$

Eqs.(3.36) and (3.37) generate the solutions of $\{\phi_j^*\}$ and $\{v_i\}$ iteratively.

7). Maximum Likelihood Approach

This approach has the advantage of handling the fluctuation of measured data in accordance with a statistically valid, analytic formalism.

a). If each data element Y_i obeys Gaussian statistics with mean $\sum_j R_{ij}\phi_j$ and variance σ_i, and all elements $\{Y_i\} = \mathbf{Y}$ are statistically independent, then the probability density function of detecting the set of elements $\{Y_i\}$ is:

$$P(\mathbf{Y}, \sigma, R\Phi) = \prod_i^I (\frac{1}{2\pi\sigma_i^2})^{1/2} e^{-(Y_i - \sum_j R_{ij}\phi_j)^2/(2\sigma_i^2)} \tag{3.38}$$

Maximum likelihood (ML) estimation looks for that solution Φ^* which maximizes the function $P(\mathbf{Y}, \sigma, R\Phi)$.

As usual, the ln likelihood function is:

$$g(\Phi) \equiv \ln P(\mathbf{Y}, \sigma, R\Phi)$$
$$= -\frac{1}{2}\sum_i [(Y_i - \sum_j R_{ij}\phi_j)^2/\sigma_i^2 + ln(2\pi\sigma_i^2)]$$

and the solution Φ^* is then determined by the equations:

$$\frac{\partial g(\Phi)}{\partial \phi_k}\Big|_{\Phi=\Phi^*} = 0. \tag{3.39}$$

If $\{\sigma_i\}$ are assumed independent of Φ, then Eq.(3.39) becomes:

$$\sum_i R_{ik}\frac{Y_i}{\sigma_i^2} = \sum_i R_{ik}\frac{\sum_j R_{ij}\phi_j^*}{\sigma_i^2}, \quad k = 1,2,...,J \tag{3.40}$$

the least square equations.

The iterative solutions for Φ^* are discussed in the chapters VI and X.

b). If each data element Y_i obeys Poisson statistics with mean $\sum_j R_{ij}\phi_j$ and all elements $\{Y_i\} = Y$ are statistically independent, then the probability density function of detecting the set of elements $\{Y_i\}$ is:

$$P(Y, R\Phi) = \prod_i^I (e^{-\sum_j R_{ij}\phi_j}) \frac{(\sum_j R_{ij}\phi_j)^{Y_i}}{Y_i!} \tag{3.41}$$

and the ln likelihood function is:

$$g(\Phi) \equiv \ln P(Y, R\Phi)$$
$$= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j) - \ln(Y_i!)] .$$

The maximum likelihood solution Φ^* of Eq.(3.41) is determined by $\frac{\partial g(\Phi)}{\partial \phi_k}|_{\Phi=\Phi^*} = 0$, which results in:

$$\sum_i R_{ik} = \sum_i R_{ik} \frac{Y_i}{\sum_j R_{ij}\phi_j^*} , \quad k = 1,2,...,J . \tag{3.42}$$

Shepp et al (1982) and Lange et al (1984) derived the iterative maximum likelihood EM algorithms (MLEM) for image reconstruction by employing the EM technique described in detail by Dempster et al (1977, JRSS).

A method combining maximum likelihood and maximum entropy considerations in image restoration was reported by Frieden (1972, J. Opt. Soc. America and 1980, Comput. Graph. Image Processing).

8). Bayesian Deconvolution Method

The Bayesian deconvolution method is directed at solving the equation

$$Y(y) = \int R(y,x)\,\Phi(x)\,dx .$$

If $Y(y)$ specifies the measured data profile $\{y_i\}$, i=1,2,...,I, and $\Phi(x)$ the source distribution $\{x_j\}$, j=1,2,...,J, then the response function $R(y,x)$, if normalized, is the conditional probability $R(y|x)$. The inverse of $R(y,x)$ is the backward probability $R(x|y)$. So there are:

$$Y(y) = \int R(y|x)\,\Phi(x)\,dx \tag{3.43}$$

and

$$\Phi(x) = \int R(x|y)\,Y(y)\,dy .$$

Bayes' Law is expressed as:

$$R(x\,|y)=\frac{R(y\,|x)\Phi(x)}{Y(y)}=\frac{R(y\,|x)\Phi(x)}{\int R(y\,|x')\Phi(x')\,dx'}$$

and substituting it into Eq.(3.43), there is:

$$\Phi(x)=\int\frac{R(y\,|x)\Phi(x)}{\int R(y\,|x')\Phi(x')dx'}Y(y)\,dy \tag{3.44}$$

which reveals that if $\Phi(x)$ is correct, then,

$$\int\frac{R(y\,|x)}{\int R(y\,|x')\Phi(x')\,dx'}Y(y)\,dy\equiv 1$$

otherwise, any non-optimal solution should be modified to approach the correct solution. An iterative form then is (Kennett et al 1979):

$$\Phi^{(n+1)}(x)=\Phi^{(n)}(x)\int\frac{R(y\,|x)Y(y)}{\int R(y\,|x')\Phi^{(n)}(x')\,dx'}dy\,. \tag{3.45}$$

In matrix form:

$$\phi_k^{(n+1)}=\phi_k^{(n)}\sum_i R_{ik}\frac{Y_i}{\sum_j R_{ij}\phi_j^{(n)}}\,. \tag{3.46}$$

If $R(y\,|x)$ is not normalized, then the MLEM algorithm results:

$$\phi_k^{(n+1)}=\phi_k^{(n)}\frac{\sum_i R_{ik}Y_i/\sum_j R_{ij}\phi_j^{(n)}}{\sum_i R_{ik}}\,. \tag{3.47}$$

Note that the above Bayesian deconvolution method (Eq.(3.45)) does not contemplate the use of supplementary information.

9). Bayesian Image Processing Formalism

The Bayesian image processing (BIP) formalism has the advantage, over other algebraic methods, of both handling the statistical fluctuation of measured data and using of any a priori source information which may exist.

Bayes' Law is written in terms of probability density function as Eq.(1.2):

$$P(\Phi\,|\mathbf{Y})=P(\mathbf{Y}\,|\Phi)P(\Phi)/P(\mathbf{Y})$$

The conditional probability density function $P(\mathbf{Y}|\Phi)$ reflects the statistical nature of data measurement processes, for example, in radioisotope organ imaging, each data element Y_i obeys Poisson

statistics and if all elements $\{Y_i\}$ are statistically independent, then,

$$P(\mathbf{Y}|\Phi)=\prod_i^I (e^{-\sum_j R_{ij}\phi_j})(\sum_j R_{ij}\phi_j)^{Y_i}/(Y_i!) \tag{3.48}$$

For independent data elements obeying Gaussian statistics, $P(\mathbf{Y}|\Phi)$ has the same form as Eq.(3.38), i.e.,

$$P(\mathbf{Y}|\Phi)=\prod_i^I (2\pi\sigma_i^2)^{-1/2} exp[-(Y_i-\sum_j R_{ij}\phi_j)^2/(2\sigma_i^2)]. \tag{3.49}$$

The probability density function $P(\Phi)$ in Eq.(1.2) contains the a priori source information and will be discussed later.

For all possible sets $\Phi \varepsilon \Omega$, the probability density function $P(\mathbf{Y})$ in Eq.(1.2) is:

$$P(\mathbf{Y})=\int P(\mathbf{Y}|\Phi)P(\Phi)d\Phi. \tag{3.50}$$

The BIP formalism is directed at finding that solution Φ^* which is most likely to give rise to the measured data, subject to a priori source information. Mathematically, it is expressed as:

$$\begin{aligned} g(\Phi) &\equiv \ln P(\Phi|\mathbf{Y}) \\ &= \ln P(\mathbf{Y}|\Phi)+\ln P(\Phi)-\ln P(\mathbf{Y}) \end{aligned} \tag{3.51}$$

and the Φ^* is determined by equations of

$$\frac{\partial g(\Phi)}{\partial \phi_k}\Big|_{\Phi=\Phi^*}=\frac{\partial \ln P(\mathbf{Y}|\Phi)}{\partial \phi_k}\Big|_{\Phi=\Phi^*}+\frac{\partial \ln P(\Phi)}{\partial \phi_k}\Big|_{\Phi=\Phi^*}=0 \tag{3.52}$$

Hunt (1977) used Bayesian concepts principally to treat non-linearities in image recording by film and the associated behavior of film grain noise in image restoration applications. He assumed that $P(\Phi)$, in our notation, was given by a multivariate normal density distribution with all mean values of $\phi_j > 0$. The work emphasized applications in two dimensional image restoration, although the possible Bayesian utlization of a priori knowledge in general multi-dimensional image analysis was clearly identified, "the medical diagnostician is not looking for airplane shapes in his images, the military photo-interpreter usually is". No development of these possibilities was presented. Other references relating to Bayesian analysis for image processing were similar to Hunt's (i.e., both $P(\mathbf{Y}|\Phi)$ and $P(\Phi)$ were assumed as multivariate normal probability density functions, and the average $\bar{\Phi}=E[\Phi]$; see Appendix 4). To our knowledge, no further reference to Bayesian concepts in complex multi-dimensional image processing applications has appeared other than that of Vardi et al (1985, JASA). In this paper, the possibility of incorporating a priori knowledge of the brain's metabolism into the reconstruction is contemplated. Vardi et al (1985) conclude, however, that for large samples (equivalent to "ideal data") and weak smoothness conditions "F", the Bayesian estimate would be very close to the ML (maximum likelihood) approach. They note "we find this to be a strong argument for preferring the ML rather than a particular

Bayes procedure". Apparently, the possibilities of more aggressive acquisition and utilization of restrictive quantitative a priori source information in $P(\Phi)$ have not been intensively explored. It is this emphasis which perhaps characterizes the current doctoral research.

In chapter IV, two general categories (generic constraints of source elements and specific constraints of source distributions) of a priori source information $P(\Phi)$ are formulated. Various additive algorithms approaching the solution Φ^* for the different a priori source information are represented in chapter X.

The multiplicative iterative forms specifying the solution Φ^*, derived using the EM technique are formulated in chapters VI, VII and VIII.

To complete this compilation of methods used for the solution of Eq.(3.1), a direct inversion method is indicated below.

10). Discussion of Direct Inversion Techniques

a). Invertibility and Well-posed Problem

Let Ω_ϕ and Ω_Y be linear spaces, a linear operator R operates on each function $\Phi \varepsilon \Omega_\phi$, resulting in $R\Phi \varepsilon \Omega_Y$.

(i) R is called continuous if there is a constant k such that

$$| R\Phi | \leq k | \Phi | ;$$

(ii) R is invertible if $R\Phi=0$ implies $\Phi=0$, i.e., for each $\mathbf{Y}$ there is precisely one element Φ denoted by $\Phi=R^{-1}\mathbf{Y}$, such that $R\Phi=\mathbf{Y}$, where R^{-1} is also a linear operator;

(iii) R^{-1} exists and is continuous if there is some constant $k > 0$ such that

$$| R\Phi | \geq k | \Phi | .$$

So if a continuous linear operator R has its inverse operator R^{-1} being continuous and linear there must exists:

$$| R\Phi | = k | \Phi |$$

that means that $R\Phi$ must be finite if Φ is finite.

(iv) well-posed and ill-posed: an equation $R\Phi=\mathbf{Y}$ is said to be well-posed if R is invertible with R^{-1} continuous. Otherwise, $R\Phi=\mathbf{Y}$ is said to be ill-posed.

The well-posed equation can be expressed by:

$$\mathbf{Y}=R\Phi+\mathbf{N}$$

where N can be referred to as a noise vector.

The ill-posed problem arises in the equation

$$\mathbf{Y}=R\Phi$$

that a small perturbation in Y, due to noise and detection errors, causes an arbitrary large perturbation in Φ. The solution is unstable.

The ill-posed sometimes is also called improperly-posed. Note that not all equations $R\Phi=\mathbf{Y}$ are improperly-posed, depending upon the matrix R.

(v) ill-conditioned: in order to actually carry out the image reconstruction, one has to derive or apply an algorithm. An algorithm is ill-conditioned if it is poorly constructed, resulting in output data of poor accuracy. For example, the iterative Jacobi method Eq.(3.3) is ill-conditioned if the matrix R is not diagonal dominant and Y is noisy. A problem is ill-conditioned (the same as ill-posed), differing from an ill-conditioned algorithm, if the output data is very sensitive to disturbances in the input data --- independent of the choice of algorithms.

b). Direct inversion Method

As mentioned before that least-squares solution of

$$\mathbf{Y}=R\Phi+\mathbf{N}$$

is given by the normal equation

$$R^{\tau}R\Phi^{*}=R^{\tau}\mathbf{Y}.$$

Let $\bar{R}=R^{\tau}R$ and $\bar{\mathbf{Y}}=R^{\tau}\mathbf{Y}$, then the normal equation becomes:

$$\bar{\mathbf{Y}}=\bar{R}\Phi^{*} \tag{3.53}$$

and the solution is expressed as (Liacer 1982):

$$\Phi^{*}=\sum_{i}\frac{\omega_i}{\lambda_i}\Theta_i \tag{3.54}$$

where $\{\lambda_i\}$ are the eigenvalues of $\bar{R}$, $\{\Theta_i\}$ are the corresponding eigenvectors and $\omega_i=<\Theta_i,\bar{\mathbf{Y}}>$.

One way of obtaining a "pseudo-inverse" for $\bar{R}$ consists in replacing ω_i and $\bar{\mathbf{Y}}$ by their definitions, arriving at the expression

$$\phi_k^* = \sum_{j=1}^{J} \sum_{l=1}^{J} \sum_{i=1}^{I} Y_i R_{ij} \frac{\Theta_l^k \Theta_l^j}{\lambda_i} . \tag{3.55}$$

Let

$$h_j = \sum_{i=1}^{I} Y_i R_{ij} \quad , \quad b_{jk} = \sum_{i=1}^{J} \frac{\Theta_l^j \Theta_l^k}{\lambda_i}$$

then a system of linear equations is obtained,

$$\phi_k^* = \sum_j b_{jk} h_j \tag{3.56}$$

where Θ_l^j is the j-th element of the l-th eigenvector, $\{R_{ij}\}$ are the elements of the rectangular system R and $\{Y_i\}$ the elements of the data vector.

More details about matrix inversion applications in image restoration can be found in references (Hunt 1973 and Ekstron 1973).

The problem of solving Φ of Eq.(3.2) for known Y and R is essentially ill-posed if R is not approximated diagonal dominant. For square matrices R, stability and/or validity of the solution can be expressed quantitatively in terms of a condition no = $| \lambda_m | / | \lambda_s |$, where λ_m and λ_s are the maximum and minimum eigenvalues of the R matrix. The smaller the condition number, the more likely is the solution valid. When R is not diagonal dominant, it is still possible to improve the solution Φ if some constraints and criteria are considered, for example, the maximum entropy and mimimum norm criteria, etc, are.

Well-posed stochastic extensions of ill-posed linear problems and their application in resolution improvement of an instrument were discussed by Franklin (1970, J. Math. Analy. Appl.) and Shaw (1972, J. Math. Analy. Appl.).

CHAPTER IV
ESTIMATION OF A PRIORI PROBABILITY FUNCTION OF SOURCE DISTRIBUTION

As mentioned in chapters I and III the BIP formalism has the advantage that it considers *both* the statistical character of the measured data (i.e. function $P(\mathbf{Y} \mid \Phi)$ in Eq.(1.2)) and any a priori information about the source distribution (i.e., function $P(\Phi)$ in Eq.(1.2)) which may exist.

The statistical character functions of measured data, $P(\mathbf{Y} \mid \Phi)$, were given by Eqs.(3.48) and (3.49) when each data element obeys Gaussian or Poisson statistics and all of them are statistically independent. For each data element obeying Gaussian statistics and all of them statistically correlated, the form of function $P(\mathbf{Y} \mid \Phi)$ is shown in chapter X.

In this chapter, the a priori source information functions, $P(\Phi)$'s, are formulated in detail for two general categories: generic and quantitative information.

1). Entropy Analysis for Generic Constraints of Source Elements

For continuous random variable ϕ_j, the definition of entropy H of the voxel j is (see Appendix):

$$H = -\int_{-\infty}^{\infty} P_j(\phi_j) \ln P_j(\phi_j)\, d\phi_j \tag{4.1}$$

where $P_j(\phi_j)$ is the a priori probability density function of random variable ϕ_j.

The properties of the probability density function $P_j(\phi_j)$ specifying the random variable (or voxel value) ϕ_j may be expressed as (Lathi 1968):

(i) normalization

$$\int_{-\infty}^{\infty} P_j(\phi_j)\, d\phi_j = 1; \tag{4.2}$$

(ii) various rank averages

$$\int_{-\infty}^{\infty} B_i(\phi_j) P_j(\phi_j)\, d\phi_j = A_i, \quad i=1,2,\ldots,q\,. \tag{4.3}$$

For $B_1(\phi_j)=\phi_j$, $A_1=\bar{\phi}_j$ is the mean of ϕ_j, and for $B_2(\phi_j)=(\phi_j-\bar{\phi}_j)^2$, A_2 is the variance of ϕ_j.

Let $A_0=1$ and $B_0(\phi_j)=1$, Eqs.(4.2) and (4.3) can be combined as:

$$\int_{-\infty}^{\infty} B_i(\phi_j) P_j(\phi_j)\, d\phi_j = A_i, \quad i=0,1,2,\ldots,q\,. \tag{4.4}$$

The maximum entropy technique aims to determine that functional form of $P_j(\phi_j)$ which maximizes the entropy H subject to the "non-valued" constraints of Eq.(4.4) [where only the existence of A_i, $(i>0)$, is assumed].

Following standard Lagrange analysis (Tribus 1971), the optimal function $P_j(\phi_j)$ is determined by maximizing

$$\tilde{H} = -\int_{-\infty}^{\infty} P_j(\phi_j) \ln P_j(\phi_j)\, d\phi_j + \sum_{i=0}^{q} \nu_i \left[\int_{-\infty}^{\infty} B_i(\phi_j) P_j(\phi_j)\, d\phi_j - A_i \right] \tag{4.5}$$

where $(\nu_0, \nu_1, ..., \nu_q)$ are the Lagrange parameters.

That is:

$$\frac{\partial \tilde{H}}{\partial P_j(\phi_j)} = -\int_{-\infty}^{\infty} [\ln P_j(\phi_j) + 1]\, d\phi_j + \sum_{i=0}^{q} \nu_i \int_{-\infty}^{\infty} B_i(\phi_j)\, d\phi_j$$

$$= -\int_{-\infty}^{\infty} d\phi_j \left[\ln P_j(\phi_j) + 1 - \sum_{i=0}^{q} \nu_i B_i(\phi_j) \right] = 0. \tag{4.6}$$

For random variable ϕ_j, Eq.(4.6) gives:

$$\ln P_j(\phi_j) + 1 - \sum_{i=0}^{q} \nu_i B_i(\phi_j) = 0$$

i.e.,

$$P_j(\phi_j) = \exp\left(\sum_{i=0}^{q} \nu_i B_i(\phi_j) - 1 \right) \tag{4.7}$$

where the $k+1$ lagrange parameters are to be determined by the $k+1$ constraints of Eq.(4.4).

For different a priori source information, the probability functions $P_j(\phi_j)$ have the following forms:

a). If the a priori information about the voxel value ϕ_j is simply that ϕ_j has a non-vanishing probability only in a finite interval [a,b] (i.e., $a_j \le \phi_j \le b_j$), then only Eq.(4.2) applies and the a priori probability density function of ϕ_j is:

$$P_j(\phi_j) = exp(\nu_0 - 1) = \begin{cases} 1/(b_j - a_j) & \text{for } a_j \le \phi_j \le b_j \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

i.e.,a uniform distribution;

b). If the a priori information about ϕ_j is $-\infty < \phi_j < \infty$ and ϕ_j has a defined mean $(\bar{\phi}_j)$ and a defined variance (ρ_j^2), then,

$$P_j(\phi_j) = exp(\nu_0 + \nu_1 \phi_j + \nu_2 \phi_j^2 - 1)$$

and by use of the constraints Eq.(4.4), there is:

$$P_j(\phi_j)=\frac{1}{\sqrt{2\pi\rho_j^2}}\,exp\,[-\frac{(\phi_j-\bar{\phi}_j)^2}{2\rho_j^2}] \tag{4.9}$$

i.e..the gaussian distribution;

c). For case b) with ϕ_j further restricted to be $\phi_j \geq 0$, the mathematical calculation of the a priori source probability density function is very complicated but a reasonable approximation is:

$$P_j(\phi_j)=\begin{cases} e^{-\bar{\phi}_j}\dfrac{(\bar{\phi}_j)^{\phi_j}}{\phi_j!} & \text{for } 0\leq\phi_j<\infty \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

the poisson distribution.++

Assuming all the voxels $\{\phi_j\}$ are statistically independent, then the three a priori probability density functions chosen for the source distribution Φ are:

$$P_a(\Phi)=\prod_j^J P_j(\phi_j)=\prod_j 1/(b_j-a_j) \tag{4.11}$$

$$P_b(\Phi)=\prod_j (2\pi\rho_j^2)^{-1/2}exp\,[-\frac{(\phi_j-\bar{\phi}_j)^2}{2\rho_j^2}] \tag{4.12}$$

and

$$P_c(\Phi)=\prod_j e^{-\bar{\phi}_j}\frac{(\bar{\phi}_j)^{\phi_j}}{\phi_j!}. \tag{4.13}$$

If all the voxels are not independent and the neighboring elements of $\{\phi_j\}$ are correlated with correlation parameter $\{\chi_{jk}\}$, then, Eq.(4.12) becomes (Fukunaga 1972):

$$P(\Phi)=\prod_{jk} C(\chi_{jk},\rho_j,\rho_k)\,exp\,[-\frac{1}{2}\frac{\chi_{jk}}{\rho_j\rho_k}(\phi_j-\bar{\phi}_j)(\phi_k-\bar{\phi}_k)] \tag{4.14}$$

where $C(...)$ is the normalization constant.

The correlation parameters $\{\chi_{jk}\}$ reflect the physical property of the sources and may be expressed as:

++ Note that in principle a logical inconsistancy can occur in case c) since while ϕ_j is restricted ≥ 0, if a gaussian data distribution is assumed, the theoretical "probability" of negative data values exists. Negative data values are, of course, excluded physically anyway.

$$\chi_{jk} = a\ exp(-\eta\ |\ j-k\ |^b) \qquad (4.15)$$

where a, b are numerical constants and $\eta = \eta(\rho_j, \rho_k)$.

2). Formulation of Specific (or Quantitative) Constraints of Source Distributions

The source probability function $P(\Phi)$ reflecting "valued" (different from the "non-valued" constraints mentioned in section 1) a priori point source information can be classified into 4 categories:

(i) spatially uncorrelated and distributionally uncorrelated;

(ii) spatially correlated and distributionally uncorrelated;

(iii) spatially uncorrelated and distributionally correlated;

(iv) spatially correlated and distributionally correlated.

Spatial correlation refers to patterns and distributional correlation refers to non-patterned inter-element correlation. The distinction will be clear in considering the separate cases. It will be assumed throughout that the point sources (hot or cold spots) are so sparsely distributed compared to the number of background elements that mutual superposition of the point sources can be neglected.++

a). Spatially uncorrelated and distributionally uncorrelated a priori information

(i) graphical appearance:

As shown in the Fig.(4.1);

(ii) description:

There are an average of J_s point sources of strength $\phi_s^e - \phi_b^e$, $s=1,2,...,\alpha$, randomly superimposed upon a background level ϕ_b^e divided into J voxels;

(iii) characterization of $P(\Phi)$:

Since the value of each voxel is independent of the values of the $(J-1)$ other voxels, the probability of the voxel j having the value ϕ_s^e is J_s/J and the probability of the voxel j having the value ϕ_b^e is $(J - \sum_s J_s)/J$. Mathematically,

++ Source probability functions for densely distributed overlapping patterns may be cumbersome to specify in terms of spatial correlation - particularly for multi-dimensional geometries. In such situations only categories i) and iii) of a priori source information may be useful.

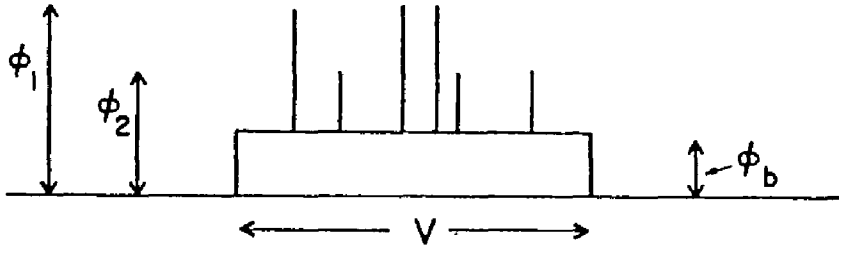


Figure 4.1: The voxel values are restricted to the values ϕ_1^e , ϕ_2^e and ϕ_b^e .

$$P(\Phi)=\prod_{j=1}^{J}P_j(\phi_j)$$

$$=\prod_j\left[\sum_s\frac{J_s}{J}\,\delta(\phi_j-\phi_s^e)+[(J-\sum_s J_s)/J]\,\delta(\phi_j-\phi_b^e)\right]. \qquad (4.16)$$

If the variation of ϕ_j about ϕ_s^e or ϕ_b^e satisfies the gaussian form, then Eq.(4.16) becomes:

$$P(\Phi)=\prod_j\{\sum_s\frac{J_s}{J}C_s\exp(-\frac{(\phi_j-\phi_s^e)^2}{2\rho_s^2})+[(J-\sum_s J_s)/J]\,C_b\exp(-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2})\} \qquad (4.17)$$

where C_b and C_s are normalization factors and ρ_s and ρ_b the variances of variable ϕ_j.

In Fig.(4.2) below, $P(\Phi)$ is sketched for $s=1,2$.

$$P_j(\phi_j)=20\,exp[-\frac{ln(2)}{\phi_b^e}(\phi_j-\phi_b^e)^2]+10\,exp[-\frac{ln(2)}{\phi_1^e}(\phi_j-\phi_1^e)^2]+15\,exp[-\frac{ln(2)}{\phi_2^e}(\phi_j-\phi_2^e)^2]$$

b). Spatially correlated and distributionally uncorrelated a priori information

(i) graphical appearance:

As shown in Fig.(4.3);

(ii) description:

There are an average of J_k patterns of type k, k=1,2,...,α, each consisting of β_k point sources of strength $\phi_{k_s}^e-\phi_b^e$, $k_s=1,2,...,\beta_k$, sparsely superimposed randomly upon a background level ϕ_b^e. Superposition of the point sources of interleaved patterns are excluded and boundary effects are neglected;

(iii) characterization of $P(\Phi)$:

Since each of the k-type patterns consists of β_k point sources, the probability of any voxel being included in the k-type patterns is $(J_k\beta_k)/J$. The probability distribution for a voxel j included in one possible superposition of the β_k point sources is:

$$\prod_k C_{k_s}\exp[-\frac{(\phi_{j+l_{k_s}}-\phi_{k_s}^e)^2}{2\rho_{k_s}^2}] \qquad (4.18)$$

where the index l_{k_s} reflects the appropriate pattern spacings and C_{k_s} is a normalization factor.

If there are q_k possible configurations, then the probability distribution for voxel j being included in a pattern is:

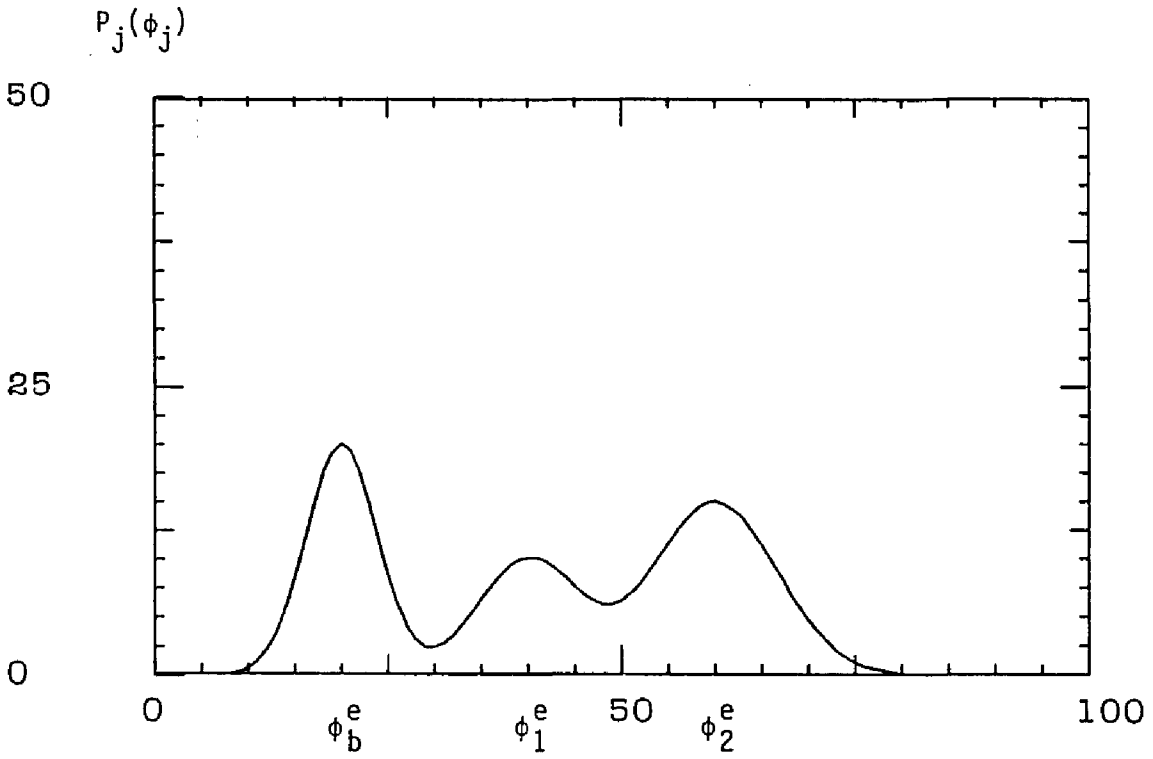


Figure 4.2: Probability distribution of a voxel value, ϕ_j. Note that the different probability amplitudes are emphasized.

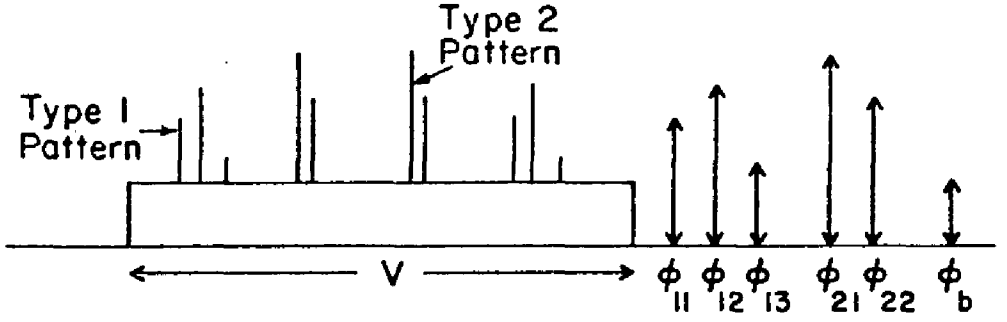


Figure 4.3: The voxel values have either the value ϕ_b^e or they are part of type 1 or type 2 pattern.

$$P_j(\phi_j)=\sum_k\{\frac{J_k\,\beta_k}{J}\,q_k^{-1}\sum_q[\prod_{k_s}C_{k_s}\exp(-\frac{(\phi_{j+l_s}-\phi_{k_s}^e)^2}{2\rho_{k_s}^2})]\}$$
$$+[\frac{(J-\sum_k J_k\,\beta_k)}{J}]\,C_b\exp(-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}) \qquad (4.19)$$

where the index q indicates summation over all possible configurations of the k-th pattern.

Neglecting inter-pattern correlation,

$$P(\Phi)=\prod_j\{\sum_k\frac{J_k\,\beta_k}{J}\,q_k^{-1}\sum_q[\prod_{k_s}C_{k_s}\exp(-\frac{(\phi_{j+l_s}-\phi_{k_s}^e)^2}{2\rho_{k_s}^2})]$$
$$+[\frac{(J-\sum_k J_k\,\beta_k)}{J}]\,C_b\exp(-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2})\}. \qquad (4.20)$$

c). Spatially uncorrelated and distributionally correlated a priori information

(i) graphical appearance:

As shown in the Fig.(4.4);

(ii) description:

There are *exactly* J_s point sources of strength $\phi_s^e-\phi_b^e$, $s=1,2,...,\alpha$, randomly superimposed upon a background level ϕ_b^e;

(iii) characterization of $P(\Phi)$:

Since there are *exactly* J_1 point sources of strength $\phi_1-\phi_b^e$ to be distributed among J voxels, J_2 point sources to be distributed among the remaining $(J-J_1)$ voxels etc, the total number of possible source distributions is:

$$\Omega_1=\prod_{k=1}^{\alpha+1}C_{J_k}^{Jm} \qquad (4.21)$$

where $Jm=J-\sum_{l=0}^{k-1}J_l$, $J_0=0$ and $J_{\alpha+1}=J_b=J-\sum_s J_s$ and C_n^m is the coefficient of binomial expansion.

Since there is a distributional correlation among the J voxels, $P(\Phi)\neq\prod_j P_j(\phi_j)$. The complete $P(\Phi)$ is now the sum of the separate (disjoint) Ω_1 possible source distribution probabilities.

The probability of a single source distribution is:

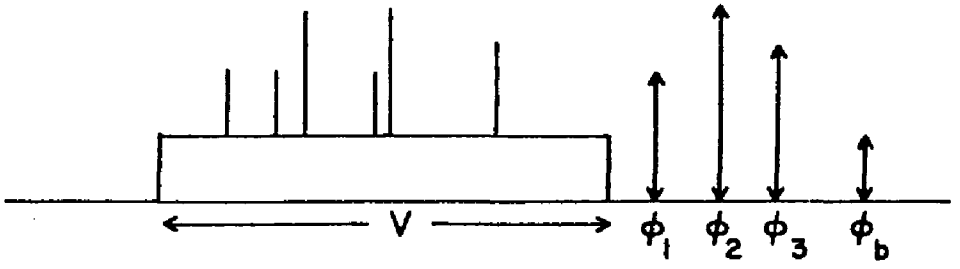


Figure 4.4: There are three voxels of value ϕ_1^e , two voxels of value ϕ_2^e , one voxel of value ϕ_3^e and J-6 voxels of value ϕ_b^e.

$$P_t(\Phi)=\frac{1}{\Omega_1}\prod_{js}^{*} C_{js}\exp[-\frac{(\phi_j-\phi_s^r)^2}{2\rho_s^2}] \tag{4.22}$$

where $\{j=1,2,...,J;\ s=1,2,...,\alpha\}$, C_{js} is a normalization factor and $\prod^{*}$ indicates that there are J product terms reflecting the associated values of j and s corresponding to that particular source distribution.

Finally, the source distribution function $P(\Phi)$ is:

$$P(\Phi)=\sum_{t=1}^{\Omega_1} P_t(\Phi). \tag{4.23}$$

d). Spatially correlated and distributionally correlated a priori information

(i) graphical appearance:

As shown in the Fig.(4.5);

(ii) description:

There are *exactly* J_k patterns of type k, $k=1,2,...,\alpha$, each consisting of β_k point sources of strength $\phi_k^e-\phi_b^e$, $k_s=1,2,...,\beta_k$, sparsely superimposed randomly upon a background level ϕ_b^e;

(iii) characterization of $P(\Phi)$:

Since there are *exactly* J_1 patterns of type 1 to be distributed among the J voxels, it follows, neglecting boundary effects, that there are $C_{J_1}^{J}$ possible source distributions for the patterns of type 1. After a type 1 pattern distribution, the number of possible source distributions of the type 2 pattern is $C_{J_2}^{(J-J_1\beta_1\beta_2)}$, where commensurate pattern spacing is neglected.

The total number of possible source distributions for α different pattern types is thus:

$$\Omega_2=\prod_{m=1}^{\alpha+1} C_{J_m}^{Jn} \tag{4.24}$$

where $Jn=J-\sum_{l=0}^{m-1} J_l\beta_l\beta_m$, $J_0=0$, $\beta_{\alpha+1}=1$ and $J_{\alpha+1}=J_b=J-\sum_k J_k\beta_k$.

The probability distribution for any one complete source distribution (comprising the $\sum_k J_k$ patterns + the remaining $(J-\sum_k J_k\beta_k)$ background elements) is then:

$$P_t(\Phi)=\frac{1}{\Omega_2}\prod_{klk_s} C_{klk_s}\exp[-\frac{(\phi_{\theta_i}-\phi_{k,k}^e)^2}{2\rho_{k,k}^2}]^{*}\prod_{m}^{(J-\sum_k J_k\beta_k)}\exp[-\frac{(\phi_m-\phi_b^e)^2}{2\rho_b^2}] \tag{4.25}$$

where $\{k=1,2,...,\alpha,\ l=1,2,...,J_k$ and $k_s=1,2,...,\beta_k\}$, C_{klk_s} is a normalization factor and for a particular source distribution $\{\phi_k\}$, the subscript θ_i reflects that voxel i corresponding to the k_s-th point source of the k-th

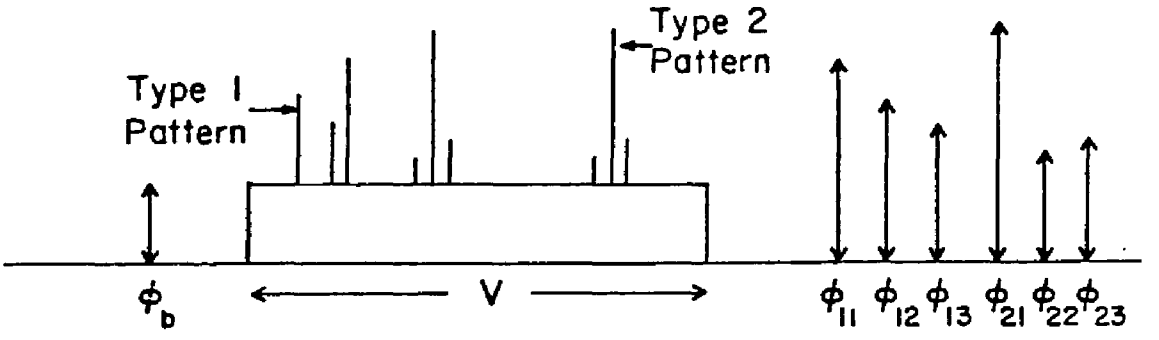


Figure 4.5: There are two type 2 patterns, one type 1 pattern present in the source field and the remaining J-9 voxels have the value ϕ_b^e .

pattern appearing for the l-th time counting from the left and index m comprises the $(J-\sum_k J_k\beta_k)$ background elements.

Finally:

$$P(\Phi)=\sum_{l=1}^{\Omega_s} P_l(\Phi). \tag{4.26}$$

The multiplicative forms of $P(\Phi)$ for cases a) and b) are:

i) for spatially uncorrelated and distributionally uncorrelated a priori source information,

$$P(\Phi)=\prod_j C_j \exp[-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}\prod_s \frac{(\phi_j-\phi_s^e)^2}{2\rho_s^2}]. \tag{4.27}$$

In the Fig.(4.6) below, $P(\Phi)$ is sketched for s=1,2 and it essentially preserves the normal distribution about ϕ_b^e, ϕ_1^e and ϕ_2^e but does not take the different probability amplitudes into account (as shown in Fig.(4.2)).

$$P_j(\phi_j)=20\,exp[-\frac{ln(2)}{\phi_b^e}(\phi_j-\phi_b^e)^2\,\frac{ln(2)}{(\phi_1^e)^2}(\phi_j-\phi_1^e)^2\,\frac{ln(2)}{\phi_2^e}(\phi_j-\phi_2^e)^2]$$

ii) for spatially correlated and distributionally uncorrelated a priori source information,

$$P(\Phi)=\prod_j C_j \exp[-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}\prod_q(\sum_s \frac{(\phi_{j+l_{qs}}-\phi_s^e)^2}{2\rho_{qs}^2})]. \tag{4.28}$$

3). Probability Density Functions of Fuzzy Pattern Sources

In section 1), the generic source constraints $P(\Phi)$ were formulated when each voxel value ϕ_j is assumed as a random variable and its average $\bar{\phi}_j$ and variance ρ_j^2 are assumed to exist. The specific (or quantitative) source constraints $P(\Phi)$, represented in section 2), reflect the a priori information about the probable voxel values $\{\phi_s^e\}$ and the known spatial relations of pattern elements, $\{l_{ks}\}$. If both the voxel values and spatial relations of pattern elements are probabilistic, then fuzzy source constraints result.

Consider first a uniform background of strength ϕ_b^e. The probability density function of this source distribution can be expressed mathematically as:

$$P(\Phi)=\prod_{j=1}^{J}\delta(\phi_j-\phi_b^e) \tag{4.29}$$

where $\delta(...)$ is the delta function and J the number of voxels.

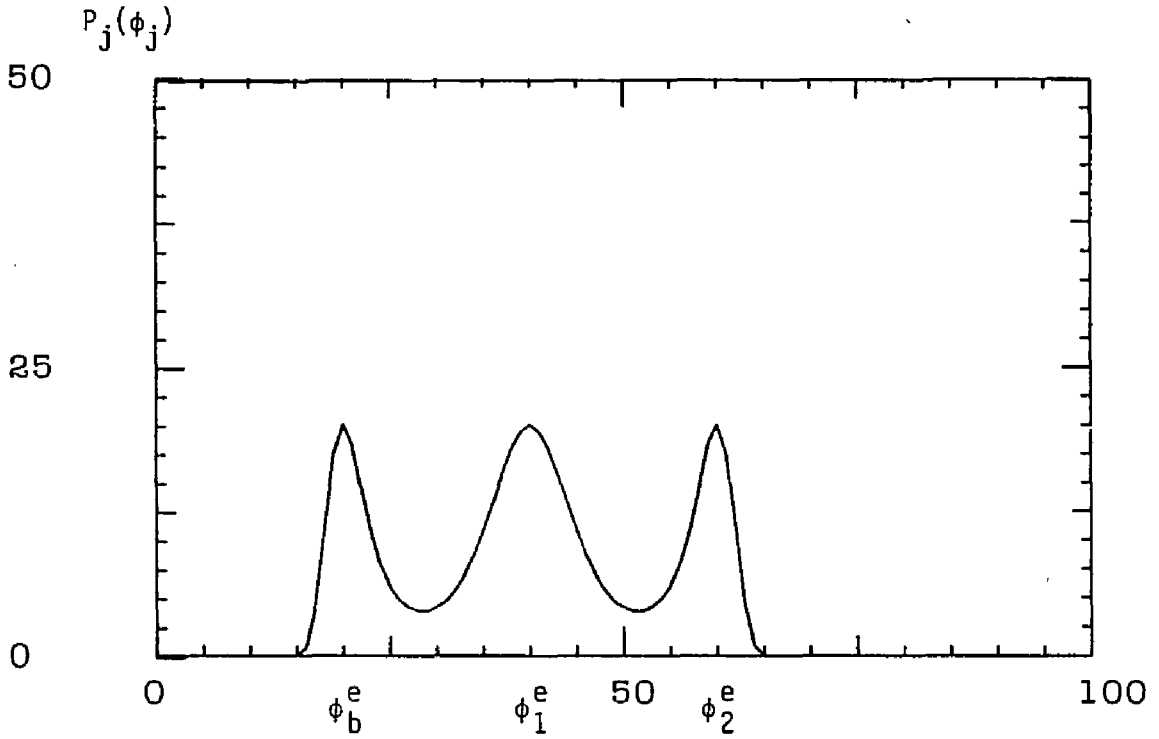


Figure 4.6: Probability distribution of a voxel value ϕ_j. Note that it essentially preserves the normal distribution about ϕ_b^e, ϕ_1^e and ϕ_2^e with the same probability amplitude.

Assuming a fixed normal distribution about the average background value of ϕ_b^e for each voxel, the function (4.29) becomes:

$$P(\Phi)=\prod_j C_b \exp[-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}] \tag{4.30}$$

where C_b is the normalization constant and ρ_b is the standard deviation characterizing the strength variation about the estimated ϕ_b^e (i.e..the uncertainty of the anticipated amplitudes).

If on the average a single spot with estimated strength $\phi_1^e-\phi_b^e$ is superimposed on the background, then, the function (4.30) becomes:

$$P(\Phi)=\prod_j \{\frac{1}{J} C_1 \exp[-\frac{(\phi_j-\phi_1^e)^2}{2\rho_1^2}]+\frac{(J-1)}{J} C_b \exp[-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}]\} \tag{4.31}$$

where C_1 and ρ_1 are analogous to the C_b and ρ_b respectively.

If it is anticipated that there are two spots with estimated strengths of $\phi_1^e-\phi_b^e$ and $\phi_2^e-\phi_b^e$ respectively and separated by l_1 voxels (i.e., they form a simple pattern) superimposed on the background, the probability density function has the form of:++

$$P(\Phi)=\prod_j \{\frac{2}{J}\frac{C_1C_2}{2}[\exp(-\frac{(\phi_{j-l_1}-\phi_1^e)^2}{2\rho_1^2}-\frac{(\phi_j-\phi_2^e)^2}{2\rho_2^2})+\exp(-\frac{(\phi_j-\phi_1^e)^2}{2\rho_1^2}-\frac{(\phi_{j+l_1}-\phi_2^e)^2}{2\rho_2^2})]$$
$$+\frac{J-2}{J} C_b \exp[-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}]\}$$

or

$$P(\Phi)=\prod_j \{\frac{2}{J}\frac{1}{Q}\sum_q [\prod_{s=1}^{2} C_s \exp(-\frac{(\phi_{j+l_{qs}}-\phi_s^e)^2}{2\rho_s^2})]+\frac{(J-2)}{J} C_b \exp(-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2})\} \tag{4.32}$$

where the index q covers the Q possible pattern configurations. Note that for a one-dimensional system the number of possible pattern configurations including voxel j as a component in the pattern is equal to the number of pattern elements. Thus:

$$l_{11}=-l_1,\ l_{12}=0,\ l_{21}=0,\ l_{22}=+l_1$$

Since the estimated value of l_1 may not be exact, a weighting function $W(l)$ is introduced to specify the variation of l around l_1 (i.e..the uncertainty in the anticipated spatial relation). The function

++ For ease of presentation, the source strength elements ϕ_1^e, ϕ_2^e,...., will always be ordered from left to right. In one dimension, pattern reversal actually does not pose any particular difficulty.

(4.32) then can be written as:

$$P(\Phi)=\prod_j\{\frac{2}{J}\frac{1}{Q}\sum_q\sum_l W(l)[\prod_s C_s\exp(-\frac{(\phi_{j+L_{qs}}-\phi_s^e)^2}{2\rho_s^2})]+\frac{(J-2)}{J}C_b\exp[-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}]\}(4.33)$$

where $L_{qs}\ \varepsilon\ l_{qs}\pm\Delta_{qs}$ and Δ_{qs} is the range of the voxels around the voxel $j\pm l_{qs}$ for which $W(l)$ is non-vanishing.

If there are β spots making up the single fuzzy pattern, the probability density function (4.33) becomes:

$$P(\Phi)=\prod_j\{\frac{\beta}{J}\frac{1}{Q}\sum_q\sum_l W(l)[\prod_{s=1}^{\beta} C_s\exp(-\frac{(\phi_{j+L_{qs}}-\phi_s^e)^2}{2\rho_s^2})]+\frac{(J-\beta)}{J}C_b\exp(-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2})\}(4.34)$$

A numerical example of Eq.(4.34) for $\beta=3$ appears as below:

Consider a pattern as shown in the Fig.(4.7), and choose voxel $j=15$ arbitrarily.

Designate configuration $q=1$ as in the Fig.(4.8), and assume:

$$w(\Delta_{qs})=\begin{cases}0.25 & \Delta_{qs}=-1\\ 0.50 & \Delta_{qs}=0\\ 0.25 & \Delta_{qs}=1\end{cases}\qquad(4.35)$$

then,

$$\begin{aligned}P_{q=1}(\phi_{15})\approx\exp(-(\phi_{15}-\phi_3^e)^2)\ [\ &0.125\exp(-(\phi_6-\phi_1^e)^2-(\phi_{10}-\phi_2^e)^2)+0.25\exp(-(\phi_7-\phi_1^e)^2-(\phi_{10}-\phi_2^e)^2)\\ &+0.125\exp(-(\phi_8-\phi_1^e)^2-(\phi_{10}-\phi_2^e)^2)+0.125\exp(-(\phi_7-\phi_1^e)^2-(\phi_9-\phi_2^e)^2)\\ &+0.125\exp(-(\phi_7-\phi_1^e)^2-(\phi_{11}-\phi_2^e)^2)+0.0625\exp(-(\phi_6-\phi_1^e)^2-(\phi_9-\phi_2^e)^2)\\ &+0.0625\exp(-(\phi_6-\phi_1^e)^2-(\phi_{11}-\phi_2^e)^2)+0.0625\exp(-(\phi_8-\phi_1^e)^2-(\phi_9-\phi_2^e)^2)\\ &+0.0625\exp(-(\phi_8-\phi_1^e)^2-(\phi_{11}-\phi_2^e)^2)\ ]\end{aligned}$$

The configurations $q=2$ and $q=3$ are shown in Figs.(4.9) and (4.10).

Similar expressions apply for $P_{q=2}(\phi_{15})$ and $P_{q=3}(\phi_{15})$.

The likelihood of voxel $j=15$ being included in the fuzzy pattern is then:

$$P(\phi_{15})\approx P_{q=1}(\phi_{15})+P_{q=2}(\phi_{15})+P_{q=3}(\phi_{15})\qquad(4.36)$$

If there are an average of J_k fuzzy patterns of type k, $k=1,2,\ldots,\alpha$, each consisting of β_k point sources of strength $\phi_{k_s}^e-\phi_b^e$, $k_s=1,2,\ldots,\beta_k$, sparsely superimposed randomly upon the background level ϕ_b^e, and superposition of the point sources of interleaved patterns are excluded and boundary effects are negleted, then Eq.(4.34) becomes:

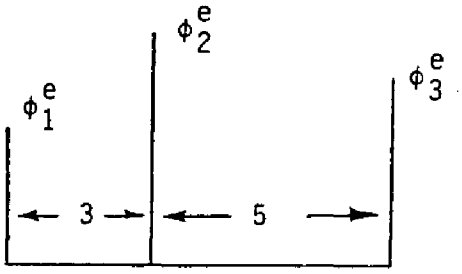


Figure 4.7: A three element pattern. The three elements have the most probable strengths ϕ_1^e , ϕ_2^e and ϕ_3^e respectively and the most probable separations of 3 and 5 voxel elements.

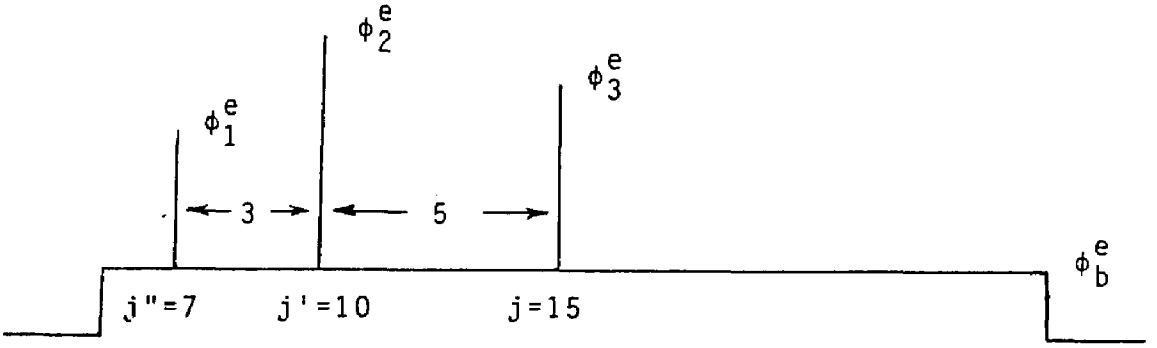


Figure 4.8: One possible configuration, q=1, of the three element pattern, in which an element is located at an arbitary position j=15.

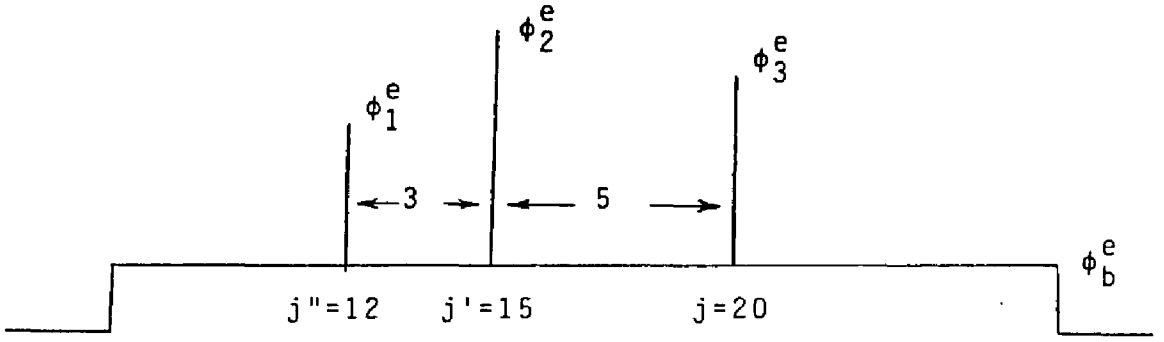


Figure 4.9: One possible configuration, q=2, of the three element pattern, in which an element is located at the position j=15.

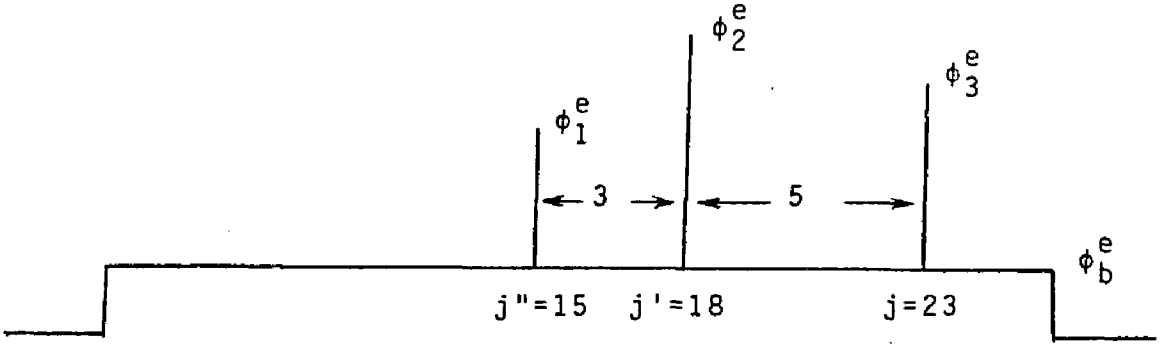


Figure 4.10: One possible configuration, q=3, of the three element pattern, in which an element is located at the position j=15.

$$P(\Phi)=\prod_j \{ \sum_k \frac{J_k \beta_k}{J} q_k^{-1} \sum_q \sum_l W(l) [\prod_k^{\beta_k} C_k \exp(-\frac{(\phi_{j+l_n}-\phi_k^s)^2}{2\rho_k^2})] + \frac{(J-\sum J_k \beta_k)}{J} C_b \exp(-\frac{(\phi_j-\phi_b^s)^2}{2\rho_b^2}) \} \quad (4.37)$$

where q_k indicates the possible configurations of a type k pattern.

The multiplicative form of $P(\Phi)$, corresponding to Eq.(4,34), can be:

$$P(\Phi)=\prod_j \sum_l W(l) \exp[-\frac{(\phi_j-\phi_b^s)^2}{2\rho_b^2}] \prod_q (\sum_{qs} \frac{(\phi_{j+l_n}-\phi_s^s)^2}{2\rho_{qs}^2})]. \quad (4.38)$$

Note that $P(\Phi)$ is only approximately given by $\prod_j P_j(\phi_j)$ since the possible existence of a pattern implies of course a dependence between appropriate voxel values. In one dimensional geometries, the effect of this dependence may not be completely trivial. In two and three dimensions, the inclusion of a single voxel j in a pattern does not specify the pattern orientation and the statistical dependence of regional voxel values arising from the possible presence of a pattern is correspondingly weaker. The probability of a sequence of voxels corresponding in amplitude and spatial orientation to a fuzzy pattern on a random basis is, however, likely to be small and the approximation $P(\Phi)=\prod_j P_j(\phi_j)$ which actually tends to understate the probability of a pattern being present does not seem to compromise the calculated results.

The utilization of these a priori source information functions, $P(\Phi)$'s, are demonstrated in chapters VI, VII and VIII.

CHAPTER V
DESCRIPTION OF THE EM TECHNIQUE

1). Description of the EM (Expectation and Maximization) Algorithm

Assume that there exist two sample spaces Ω_x and Ω_y and a many-one mapping from Ω_x to Ω_y. The measured data $\{Y_i\}$, called incomplete data, are a realization from space Ω_y and the unmeasured data $\{X_j\}$ which are indirected to $\{Y_i\}$, called complete data, are in space Ω_x.

The mapping of from Ω_x to Ω_y is expressed as:

$$\mathbf{Y}=M(\mathbf{X}) \tag{5.1}$$

and the $\mathbf{X}$ is within the subset Ω_x which is determined by the equation of

$$\mathbf{Y}=Y(\mathbf{X})=M(\mathbf{X}).$$

Let the two families of sampling density function of $\mathbf{X}$ and $\mathbf{Y}$ be respectively denoted by $f(\mathbf{X},\Phi)$ and $g(\mathbf{Y},\Phi)$, which both depend on parameter Φ and related by

$$g(\mathbf{Y},\Phi)=\int_{\Omega_x} f(\mathbf{X},\Phi)\, d\mathbf{X}. \tag{5.2}$$

Assume that the conditional average of

$$Q(\tilde{\Phi} \mid \Phi)=E[\ln f(\mathbf{X},\tilde{\Phi}) \mid \mathbf{Y},\Phi] \tag{5.3}$$

exists for all $(\tilde{\Phi},\Phi)$ and $0<f(\mathbf{X},\Phi)<\infty$ almost everywhere in space Ω_x for all $\Phi \varepsilon \Omega$, here Ω is J-dimensional convex set (i.e., $-\infty<\Phi<\infty$), the EM algorithm (Dempster et al 1977) is expressed, after the n-th iteration, as:

a). Expectation

$$Q(\Phi \mid \Phi^{(n)})=E(\ln f(\mathbf{X},\Phi) \mid \mathbf{Y},\Phi^{(n)}) \tag{5.4}$$

gives the conditional average value for the given data $\mathbf{Y}$ and the n-th resulted parameter $\Phi^{(n)}$;

b). Maximization

$$\frac{\partial Q(\Phi \mid \Phi^{(n)})}{\partial \phi_k}\Big|_{\Phi=\Phi^{(n+1)}}=0 \tag{5.5}$$

generates the $(n+1)$th iteration result $\Phi^{(n+1)}$.

If $g(\mathbf{Y},\Phi)$ is strictly concave for all $\Phi \varepsilon \Omega$, then there is an unique solution Φ^* maximizing the function $g(\mathbf{Y},\Phi)$ for a given set of measured data $\mathbf{Y}$, in another words, Φ^* is the maximum likelihood solution of function $g(\mathbf{Y},\Phi)$.

For a sequence $\{\Phi^{(n)}\}$, $n=1,2,...,\infty$, if $\ln g(\mathbf{Y},\Phi^{(n)})$ is bounded and

$$Q(\Phi^{(n+1)} \mid \Phi^{(n)}) - Q(\Phi^{(n)} \mid \Phi^{(n)}) \geq a\,(\Phi^{(n+1)}-\Phi^{(n)})^{T}(\Phi^{(n+1)}-\Phi^{(n)}) \tag{5.6}$$

for some scalar $a > 0$, then $\Phi^{(n)}$ converges to the solution Φ^{*}.

2). Maximum Likelihood Solution via the EM Algorithm

As an example, let's apply the EM technique to obtain the radioisotopic imaging maximum likelihood solution.

Let ϕ_j be the average photon emissions from voxel j at time $t=0$, which is directly proportional to the radioisotope concentration at point j within patient bady, Y_i be the data element measured at scanning position i and R_{ij} the probability of detecting photons from voxel j at scanning position i, its value can be determined by imaging a point source positioned at voxel j while detector is at position i (Andrew et al 1977).

It is well-known that random variable Y_i satisfies Poisson statistics with mean value of $\sum_j R_{ij}\phi_j$. So, the probability distribution of sampling Y_i, for the estimated parameter (or source) vector Φ, is:

$$P_i(Y_i,\Phi) = e^{-\sum_j R_{ij}\phi_j}\,\frac{(\sum_j R_{ij}\phi_j)^{Y_i}}{Y_i!}. \tag{5.7}$$

If all the data elements $\{Y_i\}=\mathbf{Y}$ are statistically independent each other, the probability distribution (or the family of sampling density function) of sampling the incomplete data $\mathbf{Y}$ is then:

$$g(\mathbf{Y},\Phi) = \prod_i^I P_j(Y_i,\Phi) = \prod_i^I e^{-\sum_j R_{ij}\phi_j}\,\frac{(\sum_j R_{ij}\phi_j)^{Y_i}}{Y_i!}. \tag{5.8}$$

Assume that X_{ij} is the number of photons emitted by voxel j and contributing to the data element Y_i, there is:

$$\sum_j X_{ij} = Y_i \quad , \quad \text{(if there is no other error in } Y_i\text{)}$$

and random variable X_{ij} also satisfies Poisson statistics and has mean of $R_{ij}\phi_j$.

If all the complete data elements $\{X_{ij}\}=\mathbf{X}$ are statistically independent each other, the family of sampling density function of $\mathbf{X}$ is, for the estimated parameters Φ, expressed as:

$$f(\mathbf{X},\Phi) = \prod_{ij}^{IJ} e^{-R_{ij}\phi_j}\,\frac{(R_{ij}\phi_j)^{X_{ij}}}{X_{ij}!}. \tag{5.9}$$

The relation of the two families of sampling density function of Eq.(5.2) now is written as:

$$g(\mathbf{Y},\Phi)=\sum_{\{X_{ij}\}} f(\mathbf{X},\Phi) \tag{5.10}$$

where the summation covers all possible sets $\{X_{ij}\}$ with constraint of $\sum_j X_{ij}=Y_i$.

The maximum likelihood solution Φ^* is given by:

$$\frac{\partial g(\mathbf{Y},\Phi)}{\partial \phi_k}\Big|_{\Phi=\Phi^*}=0$$

i.e.,

$$\sum_i R_{ik}=\sum_i R_{ik}\frac{Y_i}{\sum_j R_{ij}\phi_j^*}. \tag{5.11}$$

The EM algorithm approaches the solution Φ^* iteratively in two steps at each iteration.

a). Expectation

$$\begin{aligned} Q(\Phi\mid\Phi^{(n)}) &= E[\ln f(\mathbf{X},\Phi)\mid\mathbf{Y},\Phi^{(n)}] \\ &= -E[\sum_{ij} R_{ij}\phi_j\mid\mathbf{Y},\Phi^{(n)}]+E[\sum_{ij} X_{ij}\ln(R_{ij}\phi_j)\mid\mathbf{Y},\Phi^{(n)}] \\ &\quad -E[\sum_{ij}\ln(X_{ij}!)\mid\mathbf{Y},\Phi^{(n)}] \\ &= -\sum_{ij} R_{ij}\phi_j+\sum_{ij} X_{ij}^{(n)}\ln(R_{ij}\phi_j)-\sum_{ij}\ln(X_{ij}^{(n)}!) \end{aligned} \tag{5.12}$$

where the assumption of $\{X_{ij}\}$ being statistically independent each other has been used and the conditional average value of $X_{ij}^{(n)}$ is (Rao 1952 and Shepp et al 1982):

$$X_{ij}^{(n)}=Y_i\frac{R_{ij}\phi_j^{(n)}}{\sum_l R_{il}\phi_l^{(n)}}. \tag{5.13}$$

b). Maximization

$$\frac{\partial Q(\Phi\mid\Phi^{(n)})}{\partial\phi_k}=-\sum_i R_{ik}+\frac{\sum_i X_{ik}^{(n)}}{\phi_k}.$$

Let it be zero for $\Phi=\Phi^{(n+1)}$, then (MLEM),

$$\phi_k^{(n+1)}=\frac{\sum_i X_{ik}^{(n)}}{\sum_i R_{ik}}=\phi_k^{(n)}\frac{\sum_i R_{ik}Y_i/\sum_j R_{ij}\phi_j^{(n)}}{\sum_i R_{ik}}. \tag{5.14}$$

From Eq.(5.8), the ln likelihood function

$$g(\Phi) \equiv \ln g(\mathbf{Y},\Phi)$$
$$= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j) - \ln(Y_i!)] \tag{5.15}$$

is bounded for non-vanishing $\{\phi_j\}$ and is strictly concave if matrix R has at least as many rows as columns and is of full columns rank, since

$$Z^\tau \nabla^2 g(\Phi) Z = -\sum_i Y_i (\sum_j R_{ij} Z_j / \sum_j R_{ij}\phi_j)^2 < 0 \tag{5.16}$$

for non-vanishing Z (see the Appendix).

So, a unique solution Φ^* exists and is determined by Eqs.(5.11).

Similarly for $Q(\Phi \mid \Phi^{(n)})$ of Eq.(5.12), the $\Phi_k^{(n+1)}$ is uniquely determined by Eqs.(5.14) since $Q(\Phi \mid \Phi^{(n)})$ is a strictly concave function of Φ,

$$Z^\tau \nabla^2 Q(\Phi \mid \Phi^{(n)}) Z = -\sum_j [(\sum_i X_{ij}^{(n)})(Z_j/\phi_j^2)^2] < 0 \tag{5.17}$$

for non-vanishing Z and $X_{ij}^{(n)}$.

Expansion of $Q(\Phi \mid \Phi^{(n)})$ at point $\Phi^{(n+1)}$ is given by:

$$Q(\Phi \mid \Phi^{(n)}) = Q(\Phi^{(n+1)} \mid \Phi^{(n)}) + (\Phi - \Phi^{(n+1)})^\tau \nabla Q(\Phi^{(n+1)} \mid \Phi^{(n)})$$
$$+ \frac{1}{2}(\Phi - \Phi^{(n+1)})^\tau \nabla^2 Q(\bar{\Phi} \mid \Phi^{(n)})(\Phi - \Phi^{(n+1)}) \tag{5.18}$$

where $\bar{\Phi} \in (\Phi^{(n)}, \Phi^{(n+1)})$.

Since $\nabla Q(\Phi^{(n+1)} \mid \Phi^{(n)}) = 0$, by use of Eq.(5.12) and let $\Phi = \Phi^{(n)}$, there is:

$$\nabla^2 Q(\bar{\Phi} \mid \Phi^{(n)})_{jk} = -\sum_i \frac{X_{ij}^{(n)}}{\bar{\phi}_k^2} \delta_{jk} .$$

So,

$$Q(\Phi^{(n+1)} \mid \Phi^{(n)}) - Q(\Phi^{(n)} \mid \Phi^{(n)}) = \frac{1}{2}\sum_j (\phi_j^{(n+1)} - \phi_j^{(n)})^2 \frac{\sum_i X_{ij}^{(n)}}{\bar{\phi}_j^2} \geq 0 \tag{5.19}$$

and according to the meaning of Eq.(5.6), $\Phi^{(n)}$ will converges to the solution Φ^*.

3). Convergence of the Algorithm MLEM

A brief analysis of the convergence of the EM algorithm, Eq.(5.6), is given below:

Let

$$K(\mathbf{X} \mid \mathbf{Y},\Phi)=f(\mathbf{X},\Phi)/g(\mathbf{Y},\Phi)$$

and

$$H(\tilde{\Phi} \mid \Phi)=E[\ln K(\mathbf{X} \mid \mathbf{Y},\tilde{\Phi}) \mid \mathbf{Y},\Phi]=Q(\tilde{\Phi} \mid \Phi)-g(\tilde{\Phi}) \tag{5.20}$$

there is:

$$g(\Phi^{(n+1)})-g(\Phi^{(n)})=[Q(\Phi^{(n+1)} \mid \Phi^{(n)})-Q(\Phi^{(n)} \mid \Phi^{(n)})] + [H(\Phi^{(n)} \mid \Phi^{(n)})-H(\Phi^{(n+1)} \mid \Phi^{(n)})]. \tag{5.21}$$

By use of the Jensen's inequality of (Dempster et al 1977)

$$H(\Phi^{(n+1)} \mid \Phi^{(n)}) \le H(\Phi^{(n)} \mid \Phi^{(n)}) \tag{5.22}$$

we have:

$$Q(\Phi^{(n+1)} \mid \Phi^{(n)})-Q(\Phi^{(n)} \mid \Phi^{(n)}) \le g(\Phi^{(n+1)})-g(\Phi^{(n)}) \tag{5.23}$$

or

$$g(\Phi^{(n+1)})-g(\Phi^{(n)}) \ge \sum_j (\phi_j^{(n+1)}-\phi_j^{(n)})^2 \sum_i X_{ij}^{(n)}/(2\bar{\phi}_j^2) \ge 0. \tag{5.24}$$

So, functions $g(\Phi)$ and $Q(\Phi \mid \Phi^{(n)})$ monotonically increase to the values $g(\Phi^*)$ and $Q(\Phi^* \mid \Phi^*)$ respectively as n goes to infinite and at that point, $\Phi^{(n+1)}=\Phi^{(n)}=\Phi^*$.

From Eqs.(5.14), we know that

(i) if all elements $\{\phi_j^{(0)}\} \ge 0$, then $\{\phi_j^{(n)}\} \ge 0$;

(ii) if there are no other errors in $\{Y_i\}$, $\sum_{ij} X_{ij}^{(n)}=\sum_i Y_i$ is confirmed for all n (or the total photons emitted by all voxels are detected), i.e.,

$$\sum_{ij} X_{ij}^{(n+1)}=\sum_{ij} R_{ij}\phi_j^{(n+1)}=\sum_j R_{ij}\frac{\phi_j^{(n)}}{\sum_l R_{lj}}\sum_l R_{lj}\frac{Y_l}{\sum_k R_{lk}\phi_k^{(n)}}$$

$$=\sum_j \phi_j^{(n)}\sum_l R_{lj}\frac{Y_l}{\sum_k R_{lk}\phi_k^{(n)}}=\sum_i Y_i\frac{\sum_j R_{ij}\phi_j^{(n)}}{\sum_k R_{ik}\phi_k^{(n)}}=\sum_i Y_i;$$

(iii) as n goes to infinite, Eq.(5.14) becomes Eq.(5.11).

The rate of convergence of Eq.(5.14) is roughly disscussed as follows.

a). The Convergence Rate of Φ

To examine the change of $\Phi^{(n)}$ to $\Phi^{(n+1)}$ as n increases, let $\Phi^{(n+1)}=M(\Phi^{(n)})$, the convergence rate is then expressed as $\frac{dM(\Phi)}{d\Phi}$.

Let

$$D^{10}Z(x_1 \mid x_2)=\frac{dZ}{dx_1}\ ,\ D^{11}Z(x_1 \mid x_2)=\frac{d^2Z}{dx_1dx_2}\ \text{and}\ D^{20}Z(x_1 \mid x_2)=\frac{d^2Z}{dx_1^2}$$

the expansion of $D^{10}Q(\Phi^{(n+1)} \mid \Phi^{(n)})$ around point (Φ^*,Φ^*) is of

$$\begin{aligned} D^{10}Q(\Phi^{(n+1)} \mid \Phi^{(n)})=&D^{10}Q(\Phi^* \mid \Phi^*)+(\Phi^{(n+1)}-\Phi^*)^{\tau}D^{20}Q(\Phi^* \mid \Phi^*) \\ &+(\Phi^{(n)}-\Phi^*)^{\tau}D^{11}Q(\Phi^* \mid \Phi^*)+\Delta Q(0). \end{aligned} \tag{5.25}$$

Since $D^{10}Q(\Phi^{(n+1)} \mid \Phi^{(n)})=0$, $D^{10}Q(\Phi^* \mid \Phi^*)=0$ and $\Phi^*=M(\Phi^*)$, so,

$$\frac{(M(\Phi^{(n)})-M(\Phi^*)}{\Phi^{(n)}-\Phi^*}D^{20}Q(\Phi^* \mid \Phi^*)+D^{11}Q(\Phi^* \mid \Phi^*)+\frac{\Delta Q(0)}{\Phi^{(n)}-\Phi^*}=0. \tag{5.26}$$

When $\Phi^{(n)} \to \Phi^*$, there is, for $\Delta Q(0) \to 0$,

$$DM(\Phi^*)D^{20}Q(\Phi^* \mid \Phi^*)+D^{11}Q(\Phi^* \mid \Phi^*)=0.$$

From the definition of H, $Q(\tilde{\Phi} \mid \Phi)=H(\tilde{\Phi} \mid \Phi)+g(\tilde{\Phi})$, there is:

$$D^{11}Q(\tilde{\Phi} \mid \Phi)=\frac{d^2H(\tilde{\Phi} \mid \Phi)}{d\tilde{\Phi}d\Phi}+\frac{d^2g(\tilde{\Phi})}{d\tilde{\Phi}d\Phi}=D^{11}H(\tilde{\Phi} \mid \Phi). \tag{5.27}$$

For all $\Phi \in \Omega$, there are:

$$E\left[\frac{d}{d\Phi}\ln K(\mathbf{X} \mid \mathbf{Y},\Phi) \mid \mathbf{Y},\Phi\right]=D^{10}H(\Phi \mid \Phi)=0$$

and the conditional covariance of

$$V\left[\frac{d}{d\Phi}\ln K(\mathbf{X} \mid \mathbf{Y},\Phi) \mid \mathbf{Y},\Phi\right]=D^{11}H(\Phi \mid \Phi)=-D^{20}H(\Phi \mid \Phi)$$

so,

$$DM(\Phi^*)=D^{20}H(\Phi^* \mid *)/D^{20}Q(\Phi^* \mid \Phi^*). \tag{5.28}$$

The convergence rate is then

$$DM_k(\Phi^*)=1-\phi_k^*(\sum_j G_{kj})/(\sum_i R_{ik}) \tag{5.29}$$

where

$$G_{kj}=\sum_i R_{ik}\frac{Y_i}{(\sum_l R_{il}\phi_l^*)^2}R_{ij}\,.$$

b). The Convergence of Φ

The convergence of Eq.(5.14) may be readily estimated as follows:

Let

$$M_k(\Phi)=\phi_k\sum_i R_{ik}\frac{Y_i}{\sum_j R_{ij}\phi_j}$$

then $\Phi^{(n+1)}-\Phi^*=M(\Phi^{(n)})-M(\Phi^*)$ and

$$\begin{aligned}|\,\Phi^{(n+1)}-\Phi^*\,| &= |\,M(\Phi^{(n)})-M(\Phi^*)\,|\\ &\le |\,(\Phi^{(n)}-\Phi^*)\nabla M(\Phi^*)\,|+\frac{1}{2}|\,(\Phi^{(n)}-\Phi^*)\,|^2\,|\,\nabla^2 M(\bar{\Phi})\,|^2\end{aligned}\qquad(5.30)$$

where $\bar{\Phi}\,\varepsilon\,(\Phi^{(n)},\Phi^*)$.

Since

$$\nabla M_k(\Phi^*)=\sum_i R_{ik}+\phi_k^*\sum_i R_{ik}^2\frac{Y_i}{(\sum_j R_{ij}\phi_j^*)^2}=\omega_k(\Phi^*)\neq 0\qquad(5.31)$$

so, the second term in Eq.(5.30) can be neglected and

$$|\,\Phi^{(n+1)}-\Phi^*\,|\le\omega^0\,|\,\Phi^{(n)}-\Phi^*\,|\qquad(5.32)$$

where $\omega^0=|\,\omega\,|=\sqrt{\sum_i\omega_i^2}$ and therefore it is of order one convergence.

4). Two Modified Versions of the MLEM Algorithm

Two modified versions of Eq.(5.14) are disscussed as follows:

4.1). Relaxation Parameter Method

To speed up the convergence of Eq.(5.14) and adjust the stability of the calculation, a parameter $q>0$ is considered and the Eq.(5.14) is then expressed as:

$$\phi_k^{(n+1)}=\phi_k^{(n)}\{[\frac{\sum_i R_{ik}(Y_i/\sum_j R_{ij}\phi_j)}{\sum_i R_{ik}}-1]q-1\}.\qquad(5.33)$$

4.2). Newton-step Method

The Eq.(5.14) resulted from maximizing $Q(\Phi \mid \Phi^{(n)})$ of Eq.(5.12) directly. If $Q(\Phi \mid \Phi^{(n)})$ is expanded first, its expansion is then maximized:

$$Q(\Phi \mid \Phi^{(n)}) \simeq Q(\Phi^{(n)} \mid \Phi^{(n)}) + (\Phi - \Phi^{(n)})^{\tau} \nabla Q(\Phi^{(n)} \mid \Phi^{(n)}) + \frac{1}{2}(\Phi - \Phi^{(n)})^{\tau} \nabla^2 Q(\Phi^{(n)} \mid \Phi^{(n)})(\Phi - \Phi^{(n)})$$

and

$$\Phi^{(n+1)} = \Phi^{(n)} - (\nabla^2 Q(\Phi^{(n)} \mid \Phi^{(n)}))^{-1} \nabla Q(\Phi^{(n)} \mid \Phi^{(n)}) \tag{5.34}$$

where

$$(\nabla Q(\Phi^{(n)} \mid \Phi^{(n)}))_k = \frac{dQ}{d\phi_k} = \sum_i [R_{ik} (\frac{Y_i}{\sum_j R_{ij} \phi_j^{(n)}} - 1)]$$

and

$$(\nabla^2 Q(\Phi^{(n)} \mid \Phi^{(n)}))_{jk} = \frac{d^2 Q}{d\phi_j d\phi_k} = -(\phi_k^{(n)})^{-2} \sum_i X_{ik}^{(n)} \delta_{jk} = -A_{jk} .$$

Since A_{jk} is diagonal matrix, $(A_{jk})^{-1} = (A_{jk}^{-1})$, so,

$$\phi_k^{(n+1)} = \phi_k^{(n)} [2 - \frac{\sum_i R_{ik}}{\sum_i R_{ik} Y_i / (\sum_j R_{ij} \phi_j^{(n)})}] . \tag{5.35}$$

It is clear that as $n \to \infty$ and $\Phi^{(n+1)} = \Phi^{(n)} = \Phi^*$, then Eq.(5.35) becomes Eq.(5.14).

Since

$$M(\Phi) = \Phi - (\nabla^2 Q(\Phi \mid \Phi^{(n)}))^{-1} \nabla Q(\Phi \mid \Phi^{(n)})$$

and

$$\frac{dM(\Phi^*)}{d\Phi} = 1 + \nabla^3 Q(\Phi^* \mid \Phi^*)(\nabla^2 Q(\Phi^* \mid \Phi^*))^{-1} \nabla Q(\Phi^* \mid \Phi^*) - (\nabla^2 Q(\Phi^* \mid \Phi^*))^{-1} \nabla^2 Q(\Phi^* \mid \Phi^*) = 0$$

the convergence of Eq.(5.35) is of order two, i.e.,

$$\mid \Phi^{(n+1)} - \Phi^* \mid \leq \frac{1}{2} \mid \nabla^2 M(\Phi^*) \mid^2 \mid \Phi^{(n)} - \Phi^* \mid^2 . \tag{5.36}$$

CHAPTER VI
BAYESIAN IMAGE PROCESSING OF DATA FROM GENERIC CONSTRAINED SOURCE DISTRIBUTIONS

The formal mathematical objective in many image processing applications may be viewed as an attempt to find that source or object distribution which is most likely to have given rise to the measured data. To that end a variety of analytical or iterative methods of matrix inversion of the class of Eqs.(1.1) have been employed as shown in chapter I.

In practice, to suppress instabilities arising in application of the additive algorithms, the computer programs often exclude negative values of the elements of $\Phi^{(n)}$ by setting them to zero whenever they occur in the iterations.

Negative elements of $\Phi^{(n)}$ are also excluded automatically in the multiplicative algorithm iterations by assuming the initial estimate vector $\Phi^{(0)}$ is positive definite (i.e., all the elements of $\Phi^{(0)}$ are positive).

It is well known, therefore, that the ad hoc imposition of the above constraint of a positive definite Φ is often necessary for effective image processing and that neither Eq.(1.1) nor algorithms which are equally broad in the range of solutions Φ accepted are always sufficient, by themselves, for optimal results. It follows conversely that neither Eq.(1.1) nor any of the standard algorithms above, *by themselves*, rigorously specify the general maximum likelihood solution $\Phi^{\bullet}$ since there is no explicit formal utilization of all of the a priori source information which may be available.

This a priori source information may simply be the knowledge as above that Φ is positive definite but a wide range of a priori source information might also be available. The exact values of a particular sub-set of the voxels $\{\phi_j\}$ of Φ might be known. Less precisely, the most probable values of a sub-set of voxels $\{\phi_j\}$ might be known together with their associated probability distribution. The a priori source information can also be very general consisting of only global constraints involving all the voxels.

All of these examples of a priori source information can be symbolized as specifying different a priori source probability density functions $P(\Phi)$ in the Bayes' Law. It is convenient to consider two general types of a priori source probability distributions --- generic constrained and quantitative constrained distributions. If no information is available which quantitatively characterizes the values of voxels $\{\phi_j\}$, then obviously only generic "soft" a priori information can exist. The resulting $P(\Phi)$ will not be likely to impose severe restraints on the solutions obtained and only relative improvement in image processing (over prior methods) can be anticipated. If the a priori information characterizes probable values of the voxels $\{\phi_j\}$, however, $P(\Phi)$ is much more restrictive and as will be seen in the next chapter striking improvement in image processing results is then possible.

The generic source information of BIP formalism was discussed in chapter IV and iterative BIP algorithms will be derived as below using the EM technique.

1). Uncorrelated Constraints

It will be assumed throughout this section that all $\{\phi_j\}$ are statistically independent of each other (i.e., uncorrelated).++

1.1). Equations of Bayesian Image Processing

In terms of Bayes' Law Eq.(1.2), the source distribution to be estimated is Φ, the projection data is Y, and the Bayesian image processing objective is to find the source distribution Φ^* which maximizes the conditional probability density function $P(\Phi|\mathbf{Y})$ subject to a priori knowledge of the source and measurement probability characteristics.

For each projection Y_i obeying poisson statistics and all of $\{Y_i\}=\mathbf{Y}$ statistically independent, the conditional probability density function $P(\mathbf{Y}|\Phi)$ for sampling all $\{Y_i\}$ for estimated $\{\phi_j\}$ is given by Eq.(3.48).

The a priori probability density function $P(\Phi)$ are chosen following the analysis of maximum entropy when each of the voxel values $\{\phi_j\}$ is assumed to be a random variable as shown in chapter IV, Eqs.(4.11), (4.12), (4.13) and (4.14).

Since the a priori probability density function $P(\mathbf{Y})$ is independent of Φ, it constitutes a constant multiplicative factor when differentiating $P(\Phi|\mathbf{Y})$ partially with respect to each element of $\{\phi_j\}$ and as such can be neglected as usual.

From Eqs.(1.2) and (3.48), a ln Bayesian function can be defined in the usual way as:

$$g(\Phi)=\ln P(\Phi|\mathbf{Y})=\ln P(\mathbf{Y}|\Phi)+\ln P(\Phi)$$
$$=\sum_i[-\sum_j R_{ij}\phi_j+Y_i\ln(\sum_j R_{ij}\phi_j)-ln(Y_i!)]+\ln P(\Phi). \qquad (6.1)$$

a). For case a) as shown by Eq.(4.11), $P(\Phi)=P_a(\Phi)$ is constant for $a_j\le\phi_j\le b_j$, the equations determining that solution Φ^* which maximizes the function $g(\Phi)$ are given by partial differentiation of $g(\Phi)$ with respect to each of the $\{\phi_k\}$ and equating the results to zero, that is,

++ Note that while statistical independence implies uncorrelated variables, the reverse is in general only true if the variables are normally distributed.

$$\sum_i R_{ik} = \sum_i R_{ik} \left(\frac{Y_i}{\sum_j R_{ij} \phi_j^*} \right) \quad , k = 1,2,...,J \tag{6.2}$$

the maximum likelihood equations resulting from maximizing the function $P(\mathbf{Y} \mid \Phi)$.

b). For $P(\Phi) = P_b(\Phi)$, Eq.(6.1) becomes:

$$g(\Phi) = \sum_i \left[- \sum_j R_{ij} \phi_j + Y_i \ln \left(\sum_j R_{ij} \phi_j \right) - \ln(Y_i!) \right] - \sum_j^J \left[\frac{(\phi_j - \bar{\phi}_j)^2}{2\rho_j^2} + \frac{1}{2} \ln(2\pi\rho_j^2) \right]. \tag{6.3}$$

The equations of Bayesian image processing for the a priori gaussian probability source distribution Eq.(4.12) are:

$$\sum_i R_{ik} + \xi_k \frac{\phi_k^* - \bar{\phi}_k}{\rho_k^2} = \sum_i R_{ik} \left(\frac{Y_i}{\sum_j R_{ij} \phi_j^*} \right) \quad , k = 1,2,...,J \tag{6.4}$$

where ξ_k is an adjustable parameter > 0. The considerations involved in choosing ξ_k and the other adjustable parameters introduced subsequently (η_k, μ_k and ν_k) are indicated in section 3.

c). For $P(\Phi) = P_c(\Phi)$, and following the same steps above:

$$g(\Phi) = \sum_i \left[- \sum_j R_{ij} \phi_j + Y_i \ln \left(\sum_j R_{ij} \phi_j \right) - \ln(Y_i!) \right] - \sum_j \left[\ln(\phi_j!) - \phi_j \ln(\bar{\phi}_j) + \bar{\phi}_j \right]. \tag{6.5}$$

The Bayesian image processing equations for a priori poisson probability source distribution Eq.(4.13) become:

$$\sum_i R_{ik} + \eta_k \left(\ln \phi_k^* - \ln \bar{\phi}_k \right) = \sum_i R_{ik} \left(\frac{Y_i}{\sum_j R_{ij} \phi_j^*} \right) \quad , k = 1,2,...,J \tag{6.6}$$

by applying the Stirling approximation (Feller 1968)

$$\ln(n!) \approx n \ln(n) - n \tag{6.7}$$

and where η_k is an adjustable parameter serving the same purpose as ξ_k.

For comparison, the equations of Bayesian image processing for data obeying gaussian statistics rather than poisson statistics are derived as below:

The conditional probability density function $P(\mathbf{Y} \mid \Phi)$ of Eq.(1.2) now is given by Eq.(3.49).

The ln Bayesian function of Eq.(6.1) then becomes:

$$\bar{g}(\Phi)=-\sum_i[(Y_i-\sum_j R_{ij}\phi_j)^2/(2\sigma_i^2)+\frac{1}{2}ln(2\pi\sigma_i^2)]+\ln P(\Phi). \tag{6.8}$$

a). For $P(\Phi)=P_a(\Phi)$, maximizing $\bar{g}(\Phi)$ is equivalent to minimizing $-\bar{g}(\Phi)$, that is,

$$\sum_i[(Y_i-\sum_j R_{ij}\phi_j)^2/(2\sigma_i^2)]=\text{minimum} \tag{6.9}$$

the least square condition.

Eqs.(6.2), (6.4) and (6.6) were derived assuming poisson statistics obeyed by the data from radioactive measurements. For data obeying gaussian statistics and assuming $\sigma_i^2=Y_i$, the equations corresponding to Eq.(6.2) are derived by maximizing $\bar{g}(\Phi)$ of Eq.(6.8) for $P(\Phi)=P_a(\Phi)$:

$$\sum_i R_{ik}=\sum_i R_{ik}[\frac{\sum_j R_{ij}\phi_j^*}{Y_i}]\ ,k=1,2,...,J\ . \tag{6.10}$$

b). For $P(\Phi)=P_b(\Phi)$, the equations corresponding to Eq.(6.4) are:

$$\sum_i R_{ik}-\mu_k\frac{\phi_k^*-\bar{\phi}_k}{\rho_k^2}=\sum_i R_{ik}[\frac{\sum_j R_{ij}\phi_j^*}{Y_i}]\ ,k=1,2,...,J \tag{6.11}$$

where $\sigma_i^2=Y_i$ and the parameter μ_k has the same purpose as ξ_k.

c). For $P(\Phi)=P_c(\Phi)$, the equations corresponding to Eq.(6.6) are:

$$\sum_i R_{ik}-\nu_k(\ln\phi_k^*-\ln\bar{\phi}_k)=\sum_i R_{ik}[\frac{\sum_j R_{ij}\phi_j^*}{Y_i}]\ ,k=1,2,...,J \tag{6.12}$$

where, as above, $\sigma_i^2=Y_i$ and ν_k is an adjustable parameter.

Since the functions $g(\Phi)$ and $\bar{g}(\Phi)$ for the cases above are strictly concave (see Appendix) whenever the matrix R has at least as many rows (projection rays) as columns (voxels) and is of full column rank, the Bayesian image processing Eqs.(6.2), (6.4), (6.6), (6.10), (6.11) and (6.12) respectively determine unique optimal solutions Φ^*.

In practice, iterative solutions $\Phi^{(n)}$ are indicated whenever the matrix R is not diagonal dominant and the dimensions of Y and Φ are large.

1.2). Derivation of BIP Algorithms via the EM Technique

The EM technique (Dempster et al 1977) as shown in chapter V generates iterative solutions for Eqs.(6.2), (6.4), (6.6), (6.10), (6.11) and (6.12).

By analogy with the Bayes' Law of incomplete data of Eq.(1.2), the expression of complete data Bayes' Law can be written as:

$$P(\Phi \mid \mathbf{X}) = P(\mathbf{X} \mid \Phi) P(\Phi) / P(\mathbf{X}). \tag{6.13}$$

The conditional probability function $P(\mathbf{X} \mid \Phi)$ was given by Eq.(5.9) for independent poisson random variables $\{X_{ij}\}$, i.e.,

$$P(\mathbf{X} \mid \Phi) = \prod_{ij}^{IJ} (e^{-R_{ij}\phi_j}) (R_{ij}\phi_j)^{X_{ij}} / (X_{ij}!). \tag{6.14}$$

The function $P(\Phi)$ is the a priori source information function as discussed before and $P(\mathbf{X})$ can be omitted without affecting the solution Φ^* since it is independent of Φ.

From Eqs.(6.13) and (6.14), we have:

$$\ln f(\mathbf{X},\Phi) = \ln P(\Phi \mid \mathbf{X}) = \sum_{ij} [-R_{ij}\phi_j + X_{ij} \ln(R_{ij}\phi_j) - \ln(X_{ij}!)] + \ln P(\Phi). \tag{6.15}$$

The E-step (expectation) in the EM technique computes the conditional average of

$$\begin{aligned} Q(\Phi \mid \Phi^{(n)}) &= E[\ln f(\mathbf{X},\Phi) \mid \mathbf{Y}, \Phi^{(n)}] \\ &= \sum_{ij} [-R_{ij}\phi_j + X_{ij}^{(n)} \ln(R_{ij}\phi_j) - \ln(X_{ij}^{(n)}!)] + \ln P(\Phi) \end{aligned} \tag{6.16}$$

where $\Phi^{(n)}$ represents the n-th iterated result and where the conditional average value of the random variables $\{X_{ij}\}$ for the given measured data $\{Y_i\}$ in Eq.(6.16) is given by Eq.(5.13).

The M-step (maximization) of the EM technique determines the (n+1)th iterative result, $\Phi^{(n+1)}$, which maximizes the function $Q(\Phi \mid \Phi^{(n)})$ of Eq.(6.16):

$$\frac{\partial Q(\Phi \mid \Phi^{(n)})}{\partial \phi_k} \Big|_{\Phi = \Phi^{(n+1)}} = 0. \tag{6.17}$$

a). For $P(\Phi) = P_a(\Phi)$ constant in the range $a_j \le \phi_j \le b_j$, maximizing $Q(\Phi \mid \Phi^{(n)})$ gives:

$$\sum_i (-R_{ik} + X_{ik}^{(n)} / \phi_k^{(n+1)}) = 0 \tag{6.18}$$

or

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik} (Y_i / \sum_j R_{ij} \phi_j^{(n)})}{\sum_i R_{ik}} \tag{6.19}$$

the maximum likelihood result (MLEM) of Eq.(6) (Lange et al 1984) with, however, the additional constraints of $a_j \le \phi_j^{(n)} \le b_j$. Eq.(6.19) is the desired iterative algorithm for solving Eq.(6.2).

b). For $P(\Phi)=P_b(\Phi)$ and using Eq.(4.12), Eq.(6.16) becomes:

$$Q(\Phi \mid \Phi^{(n)}) = \sum_{ij}[-R_{ij}\phi_j + X_{ij}^{(n)}\ln(R_{ij}\phi_j) - ln(X_{ij}^{(n)}!)] - \sum_j[\frac{(\phi_j-\bar{\phi}_j)^2}{2\rho_j^2} + \frac{1}{2}\ln(2\pi\rho_j^2)] \quad (6.20)$$

by using Eqs.(5.13) and (6.17),

$$\sum_i(-R_{ik} + X_{ik}^{(n)}/\phi_k^{(n+1)}) = \xi_k\,(\phi_k^{(n+1)} - \bar{\phi}_k)/\rho_k^2 \quad (6.21)$$

or

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik}(Y_i/\sum_j R_{ij}\phi_j^{(n)}) + \xi_k\,(\bar{\phi}_k/\phi_k^{(n)})}{\sum_i R_{ik} + \xi_k} \quad (6.22)$$

where we have assumed that $\rho_k^2 = \phi_k^{(n+1)}$ and Eq.(6.22) is the iterative algorithm for solving Eq.(6.4).

c). For $P(\Phi)=P_c(\Phi)$ and using Eq.(4.13), there is:

$$Q(\Phi \mid \Phi^{(n)}) = \sum_{ij}[-R_{ij}\phi_j + X_{ij}^{(n)}\ln(R_{ij}\phi_j) - ln(X_{ij}^{(n)}!)] - \sum_j[ln(\phi_j!) - \phi_j ln(\bar{\phi}_j) + \bar{\phi}_j] \quad (6.23)$$

and by use of Eqs.(6.7) and (6.17),

$$\sum_i(-R_{ik} + X_{ik}^{(n)}/\phi_k^{(n+1)}) = \eta_k\,[ln(\phi_k^{(n+1)}) - ln(\bar{\phi}_k)]. \quad (6.24)$$

The correction term on the right hand side of Eq.(6.24) can be approximated as follows:

Let $\phi_k^{(n+1)} = \phi_k^{(n)} + \tau d_k^{(n)}$ and $\tau \approx 1$, and assuming $d_k^{(n)} \ll \phi_k^{(n)}$, then,

$$ln(\phi_k^{(n+1)}) = ln(\phi_k^{(n)}) + \tau d_k^{(n)}/\phi_k^{(n)}$$

therefore, considering Eq.(5.13),

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik}(Y_i/\sum_j R_{ij}\phi_j^{(n)})}{\sum_i R_{ik} + \eta_k\,[ln(\phi_k^{(n)}/\bar{\phi}_k) + \tau d_k^{(n)}/\phi_k^{(n)}]} \quad (6.25)$$

the iterative algorithm for solving Eq.(6.6) where $d_k^{(n)}$ may be further assumed as $\phi_k^{(n)} - \phi_k^{(n-1)}$.

Similarly for data obeying gaussian statistics:

$$P(\mathbf{X} \mid \Phi) = \prod_{ij}^{IJ} (2\pi\sigma_{ij}^2)^{-1/2} exp\,[-(X_{ij} - R_{ij}\phi_j)^2/(2\sigma_{ij}^2)] \tag{6.26}$$

and

$$Q(\Phi \mid \Phi^{(n)}) = \sum_{ij}[(R_{ij}\phi_j)X_{ij}^{(n)}/\sigma_{ij}^2 - (R_{ij}\phi_j)^2/(2\sigma_{ij}^2)] + \ln P(\Phi) + C(X_{ij}^{(n)}, \sigma_{ij}) \tag{6.27}$$

where $X_{ij}^{(n)}$ is given by Eq.(5.13), σ_{ij}^2 is the variance of the random number X_{ij} and will be assumed to be $X_{ij}^{(n)}$ in later use and $C(...)$ is assumed independent of Φ.

a). For $P(\Phi) = P_a(\Phi)$, the iterative algorithm is:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik}}{\sum_i R_{ik}(\sum_j R_{ij}\phi_j^{(n)}/Y_i)} \tag{6.28}$$

for solving Eq.(6.10).

b). For $P(\Phi) = P_b(\Phi)$, the iterative algorithm solving Eq.(6.11) is:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik} + \mu_k\,\bar{\phi}_k/\rho_k^2}{\sum_i R_{ik}(\sum_j R_{ij}\phi_j^{(n)}/Y_i) + \mu_k\,\phi_k^{(n)}/\rho_k^2} \tag{6.29}$$

where ρ_k^2 may be assumed as $\bar{\phi}_k$.

c). For $P(\Phi) = P_c(\Phi)$, the iterative algorithm solving Eq.(6.12) is:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik} - \nu_k\,[\,ln(\phi_k^{(n)}/\bar{\phi}_k) + \tau d_k^{(n)}/\phi_k^{(n)}]}{\sum_i R_{ik}(\sum_j R_{ij}\phi_j^{(n)}/Y_i)}. \tag{6.30}$$

In the limit as n goes to infinity, $\phi_k^{(n+1)} = \phi_k^{(n)} \rightarrow \phi_k^*$, and the algorithms (6.19), (6.22), (6.25), (6.28), (6.29) and (6.30) converge as expected to the Bayesian image processing Eqs.(6.2), (6.4), (6.6), (6.10), (6.11) and (6.12) respectively.

Although Eq.(6.18) is exactly the previously derived result of Lange et al (1984), in numerical application the iterative values of the source elements $\{\phi_j^{(n)}\}$ are of course now always restricted $a_j \leq \phi_j^{(n)} \leq b_j$ and divergent results are therefore suppressed.

As will be seen below, Eqs.(6.21) and (6.24) are sometimes more effective in image processing than Eq.(6.18) due to the terms on their right sides. If $\phi_k^{(n)}$ converges to $\bar{\phi}_k$ for n large, then Eqs.(6.18), (6.21) and (6.24) result in the same solution. However if the noise is relatively large, $\phi_k^{(n+1)}$ can now be bounded around $\bar{\phi}_k$ by the adjustable parameters. From Eq.(6.22), for example, it is seen that when $\phi_k^{(n)}$ is less then $\bar{\phi}_k$, the ratio of the two summation terms gets larger and so $\phi_k^{(n+1)}$ increases. The variations of

$\phi_k^{(n+1)}$ from $\bar{\phi}_k$ are adjusted by the parameter ξ_k.

Both the choice of $\{\bar{\phi}_k\}$ and the adjustable parameters are quite critical for optimal image processing and will be discussed in the section 3.

2). Correlated Constraints

In section I), it was assumed that all $\{\phi_m\}$ are statistically independent (uncorrelated). However, if all elements $\{\phi_m\}$ are not independent and the neighboring elements of $\{\phi_m\}$ are correlated with the correlation parameter $\{\chi_{mr}\}$, $P(\Phi)=\prod_j P_j(\phi_j)$ no longer applies and the a priori probability source distribution can now be expressed as:

$$P(\Phi)=\prod_{mr} C(\chi_{mr},\rho_m,\rho_r)\exp[-\frac{\chi_{mr}}{2\rho_m\rho_r}(\phi_m-\bar{\phi}_m)(\phi_r-\bar{\phi}_r)] \tag{6.31}$$

where the function $C(...)$ is assumed independent of Φ and χ_{mr} is a function of $\{\rho_m,\rho_r,m-r\}$ as shown in Eq.(4.14).

The ln Bayesian function Eq.(6.1) is:

$$\begin{aligned} g(\Phi)&=\ln P(\Phi|Y)=\ln P(Y|\Phi)+\ln P(\Phi)\\ &=\sum_i[-\sum_j R_{ij}\phi_j+Y_i\ln(\sum_j R_{ij}\phi_j)-\ln(Y_i!)]\\ &\quad-\sum_{mr}[\frac{\chi_{mr}}{2\rho_m\rho_r}(\phi_m-\bar{\phi}_m)(\phi_r-\bar{\phi}_r)-C(...)] \end{aligned} \tag{6.32}$$

and the equations determining the optimal solution Φ^* of BIP with correlated constraints are given by

$$\sum_i R_{ik}+\zeta_k\chi_{kk}\frac{\phi_k^*-\bar{\phi}_k}{\rho_k^2}+\zeta_k\sum_{r\neq k}[\frac{\chi_{kr}}{\rho_k\rho_r}(\phi_r^*-\bar{\phi}_r)]=\sum_i R_{ik}(\frac{Y_i}{\sum_j R_{ij}\phi_j^*}) \tag{6.33}$$

where ζ_k is an adjustable parameter and will be included in parameter $\{\chi_{kr}\}$ later.

For $g(\Phi)$ strictly concave, Eq.(6.33) determines the unique optimal solution Φ^*.

The EM technique results in the iterative algorithm:

$$\sum_i(-R_{ik}+\frac{X_{ik}^{(n)}}{\phi_k^{(n+1)}})-\chi_{kk}\frac{(\phi_k^{(n+1)}-\bar{\phi}_k)}{\rho_k^2}=\sum_{r\neq k}[\frac{\chi_{kr}}{\rho_k\rho_r}(\phi_r^{(n+1)}-\bar{\phi}_r)]$$

assuming that $\rho_k^2=\phi_k^{(n+1)}$,

$$\phi_k^{(n+1)} = \frac{\sum_i X_{ik}^{(n)} + \chi_{kk}\,\bar{\phi}_k}{\sum_i R_{ik} + \chi_{kk} + \sum_{r \neq k} (\chi_{kr} Z_r^{(n)})} \tag{6.34}$$

where

$$Z_r^{(n)} = (1 - \frac{\tau d_k^{(n)}}{2\phi_k^{(n)}})(1 - \frac{\tau d_r^{(n)}}{2\phi_r^{(n)}}) \frac{(\phi_r^{(n)} - \bar{\phi}_r + \tau d_r^{(n)})}{\sqrt{\phi_k^{(n)} \phi_r^{(n)}}}$$

and $X_{ik}^{(n)}$ is given by Eq.(5.13).

Similarly an algorithm for data obeying gaussian statistics can be derived:

$$\phi_k^{(n+1)} = \frac{\sum_i R_{ik} + \chi_{kk}\bar{\phi}_k/\rho_k^2 - \sum_{r \neq k} M_r^{(n)}}{\sum_i R_{ik}^2 / X_{ik}^{(n)} + \chi_{kk}/\rho_k^2} \tag{6.35}$$

where

$$M_r^{(n)} = \chi_{kr}(\phi_r^{(n)} - \bar{\phi}_r + \tau d_r^{(n)})/\rho_r \rho_k$$

and ρ_k^2 may be assumed as $\bar{\phi}_k$.

The results obtained with all of the above algorithms are reported in the next section.

3). Simulation Results

The algorithms (6.19), (6.22), (6.25), (6.28), (6.29), (6.30) and (6.34) were each run for 100 iterations using ideal and poisson or gaussian randomized data for a one-dimensional system in which the elements $\{R_{ij}\}$ of matrix R (PSF) are assumed to exhibit the exponential dependence,

$$R_{ij} = A_0 exp[-\frac{\ln(2)}{T^2} * (i-j)^2] \tag{6.36}$$

where A_0 is a constant and T one half of the FWHM of the PSF.

The actual source $\{O_j\}$ and ideal data distributions are shown in Fig.(6.1). The two point sources of 100 units are separated by 8 voxel elements and are superimposed on an uniform background of 10 units. The background ranges from voxel 3 to voxel 27 and is zero outside the range. The ideal data (accurate projections) used in the test is calculated out by $\sum_j R_{ij} O_j$ with T=4 voxel elements (i.e., the matrix R is not diagonal dominant).

The matrix R is rectangular (35 × 25), i.e., there are 35 projections and 25 voxels (a system of overdetermined equations). In Fig.(6.1) the projection data extends two voxel elements beyond both two sides of the graph (from an abscissa of -2 to +32).

The initial values of $\{\phi_k^{(n)}\}$ were chosen to be proportional to the measured data, i.e.,

$$\phi_k^{(0)} = Y_k / B_0 \ , \ \ B_0 = \max[\sum_i R_{ij}] . \tag{6.37}$$

The choice of $\{\bar{\phi}_k\}$ and the adjustable parameters is based on analysis of the results obtained with Eq.(6.22):

If $\bar{\phi}_k = O_k$ and

$$0 < \xi_k \approx \sum_i R_{ik} \tag{6.38}$$

Eq.(6.22) generates excellent results $\{\phi_k^{(n)} = O_k\}$ after a few iterations.

If $\bar{\phi}_k = \phi_k^{(n+1)}$, Eqs.(6.21) and (6.24) reduce to Eq.(6.18) and the maximum likelihood solution is then obtained.

A reasonable improved choice in the absence of knowledge of $\{O_k\}$ would be:

$$\bar{\phi}_k = \phi_k^{(n+1)} + \tau_0 d_k^{(n)} \tag{6.39}$$

where factor τ_0 is the step length of $\phi_k^{(n+1)}$ converging to $\bar{\phi}_k$ in the direction $d_k^{(n)}$ and has value of order one.

Since, however, $\phi_k^{(n+1)}$ is still not known (only $\phi_k^{(n)}$ is immediately available), the choice of

$$\bar{\phi}_k = \phi_k^{(n)} = \phi_k^{(n+1)} - \tau_0 d_k^{(n)}$$

is equivalent to a "correction" applied in the opposite (wrong) direction and therefore ξ_k is chosen to be *negative*. In summary then from Eqs.(6.21), (6.24)ff and (6.34)ff:

$$\bar{\phi}_k = \phi_k^{(n)} \ \text{ and } \ -\sum_i R_{ik} < \xi_k < 0 \tag{6.40}$$

In practice ξ_k and the other adjustable parameters (η_k, μ_k and ν_k) can be chosen to be monotonic functions of iterative index n, i.e.,

$$\xi_k(n) \approx -\frac{n}{n+L} \sum_i R_{ik} \tag{6.41}$$

where L is the total number of iterations.

For algorithm (6.34), the index r covers the neighboring elements of $\phi_k^{(n)}$ and the correlation parameters

$$\chi_{kr} \approx e^{-\alpha(k-r)^2} . \tag{6.42}$$

In these calculations,

$$\chi_{kk} \simeq \xi_k \ , \ \chi_{kr} = 0.5\chi_{kk} \ , \ (r = k \pm 1)$$

were chosen.

For ideal data, the results of the algorithms derived assuming different a priori source information are shown in Fig.(6.2) (Eqs.(6.19), (6.22), (6.25) and (6.34)) and Fig.(6.3) (Eqs.(6.28), (6.29), (6.30)). The improved performance of the new algorithms is easily seen. Note that better performance is achieved with the correlated source constraints (algorithm (6.34)).

Fig.(6.4) compares the results obtained with the four poisson data algorithms applied to the single set of data shown in Fig.(6.1) for which the median data element was about 30. The poisson randomized data were obtained from the ideal data by use of a standard poisson random number generator (Carnahan et al 1978) as shown in the Appendix.

Since relatively wide variation in comparative results can, of course, be possible for individual data sets, the algorithms were then applied to ten sets of such poisson randomized data. The average of the corresponding ten sets of results are graphed in Fig.(6.5).

The results obtained with the three gaussian data algorithms for a single set of gaussian randomized data (Hamming 1962) are shown in Fig.(6.6). The averaged results for ten sets of gaussian randomized data are shown in Fig.(6.7).

The correlated gaussian data algorithm (6.35) resulted in comparative solutions to those of the correlated poisson data algorithm (6.34) for ideal data, gaussian and poisson randomized data respectively.

The algorithms were also tested for 1000 iterations using ideal data. All of the results continue to converge uniformly toward $\{O_k\}$ with no change in the relative effectiveness of the different algorithms.

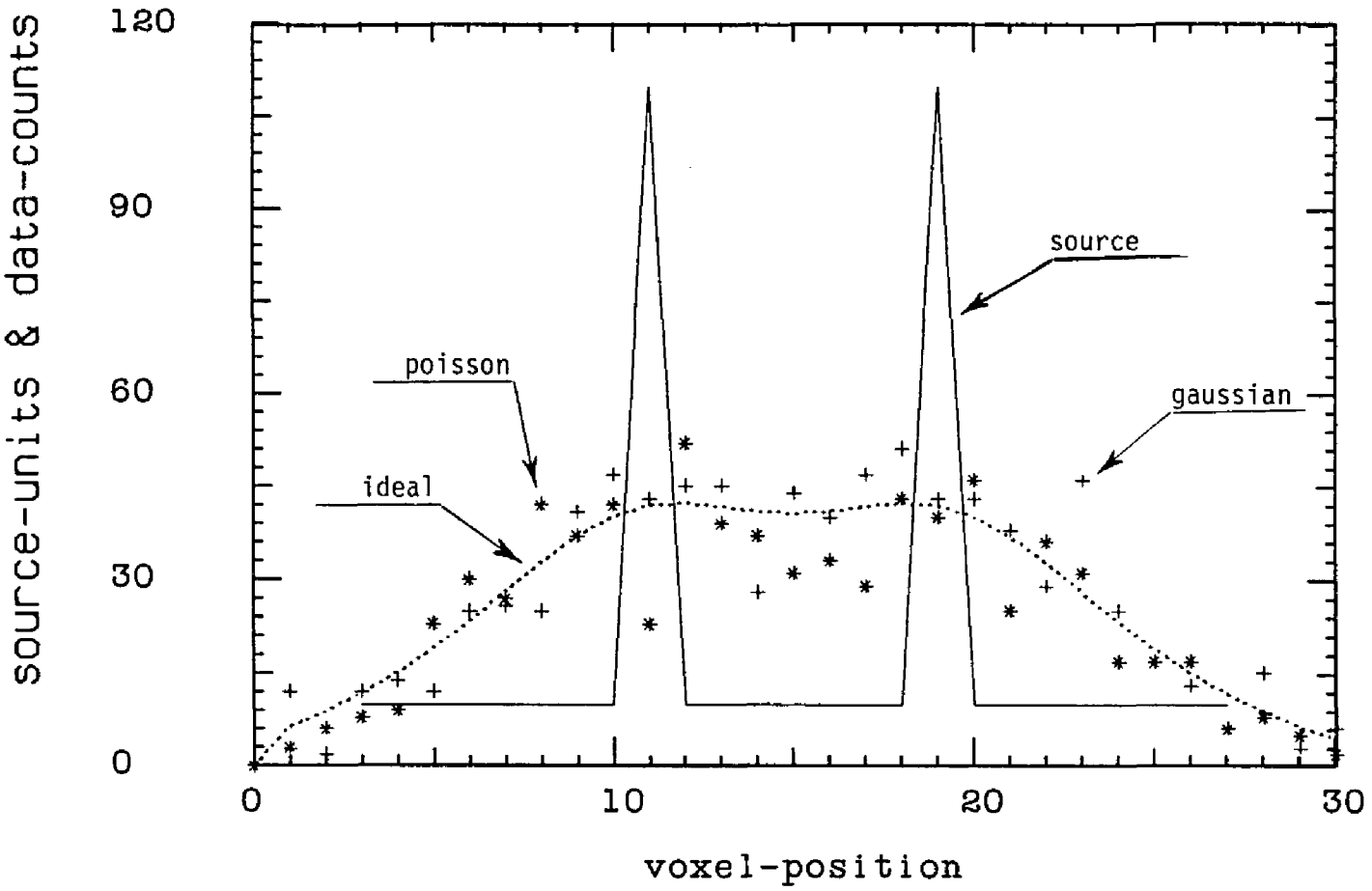


Fig.6.1: One dimensional source and data distributions. The solid line represents the actual source distribution. The dotted line represents the ideal data distribution. The stars represent the computer poisson randomized data. The crosses represent the computer gaussian randomized data.

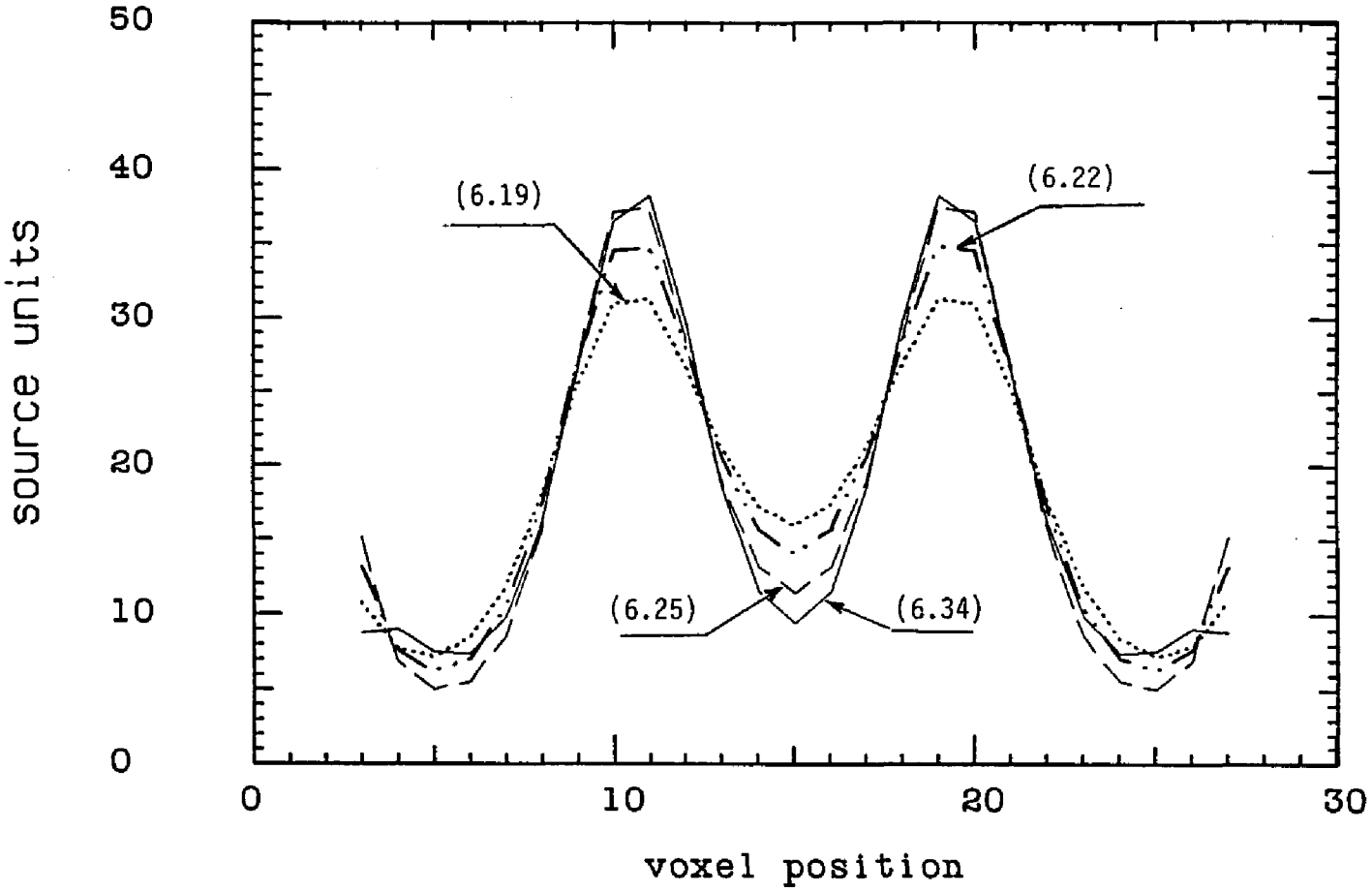


Fig. 6.2: Results using algorithms (6.19), (6.22), (6.25) and (6.34) on ideal data for 100 iterations .

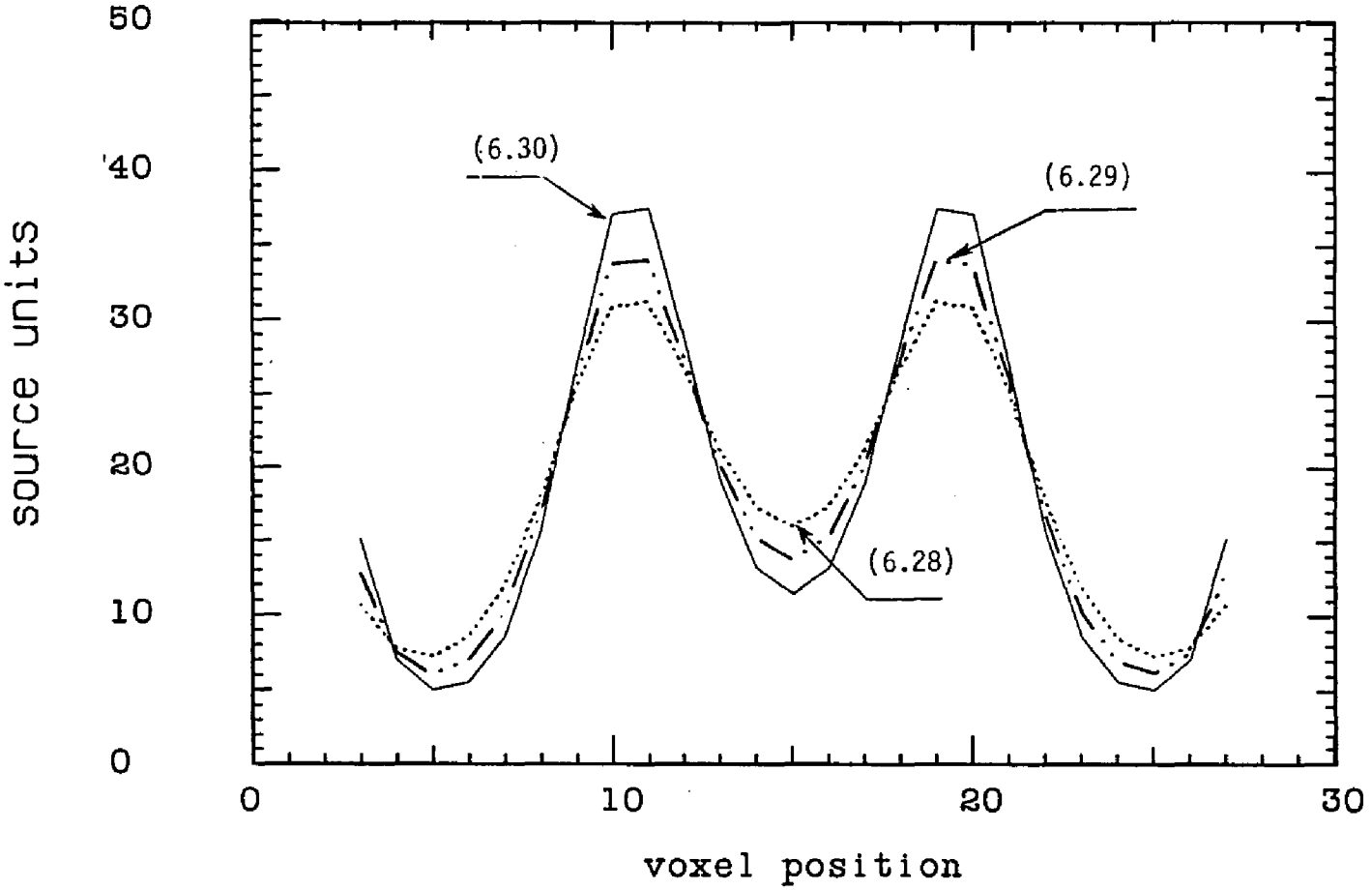


Fig.6.3: Results using algorithms (6.28), (6.29) and (6.30) on ideal data for 100 iterations.

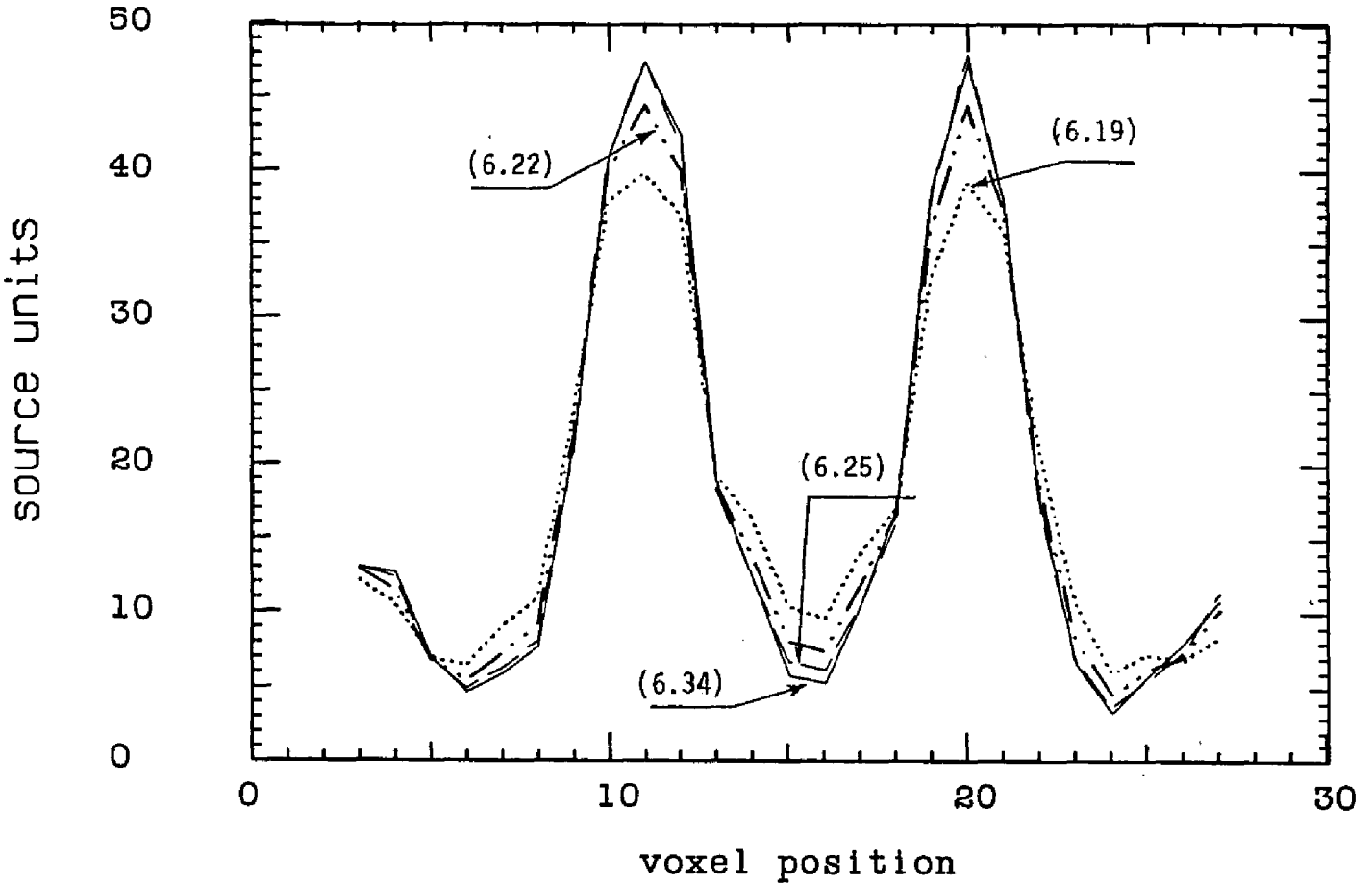


Fig.6.4: Results using algorithms (6.19), (6.22), (6.25) and (6.34) on a single set of poisson randomized data for 100 iterations.

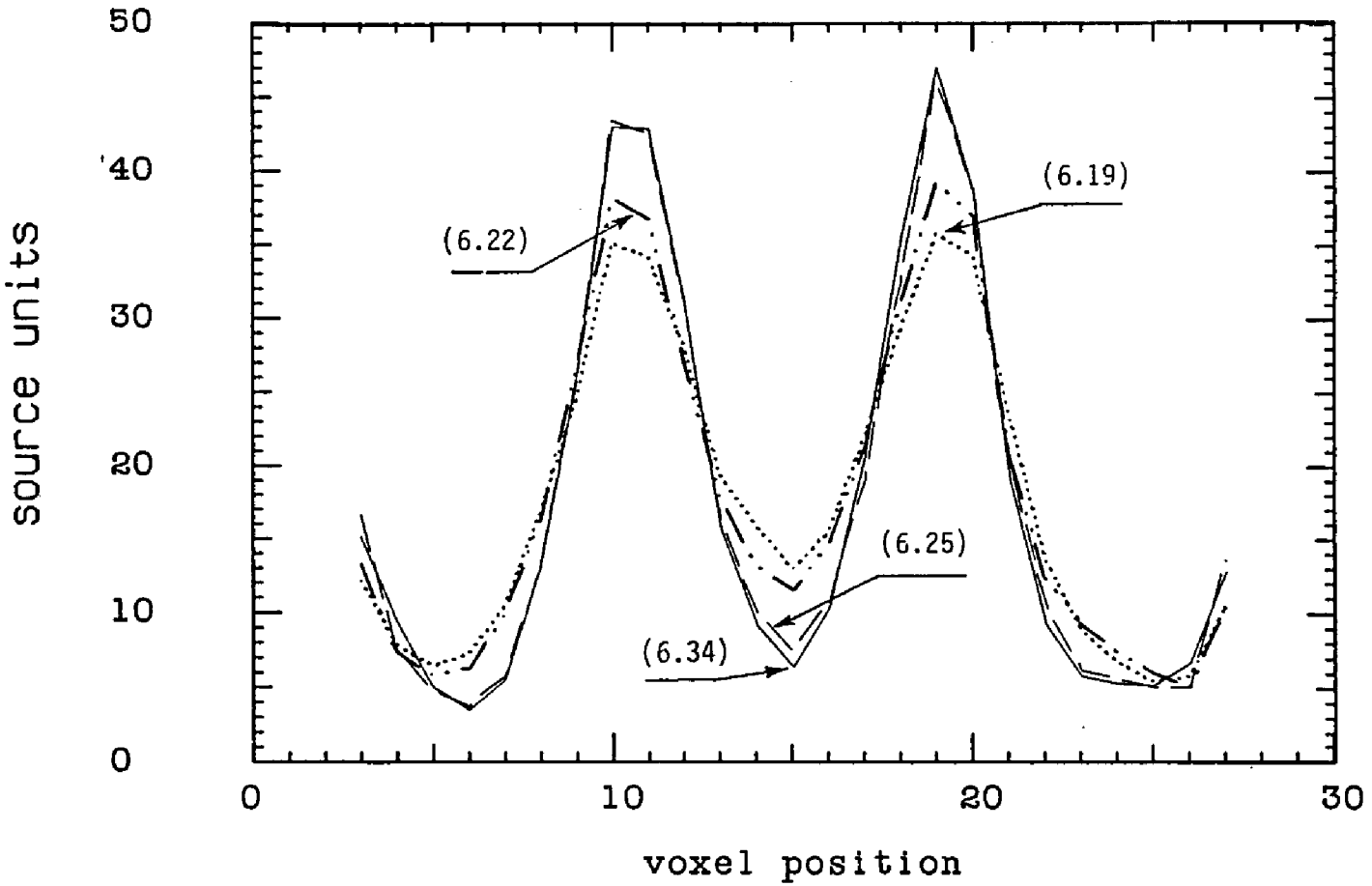


Fig. 6.5: Comparison of (6.19), (6.22), (6.25) and (6.34) for ten sets of poisson random data and 100 iterations.

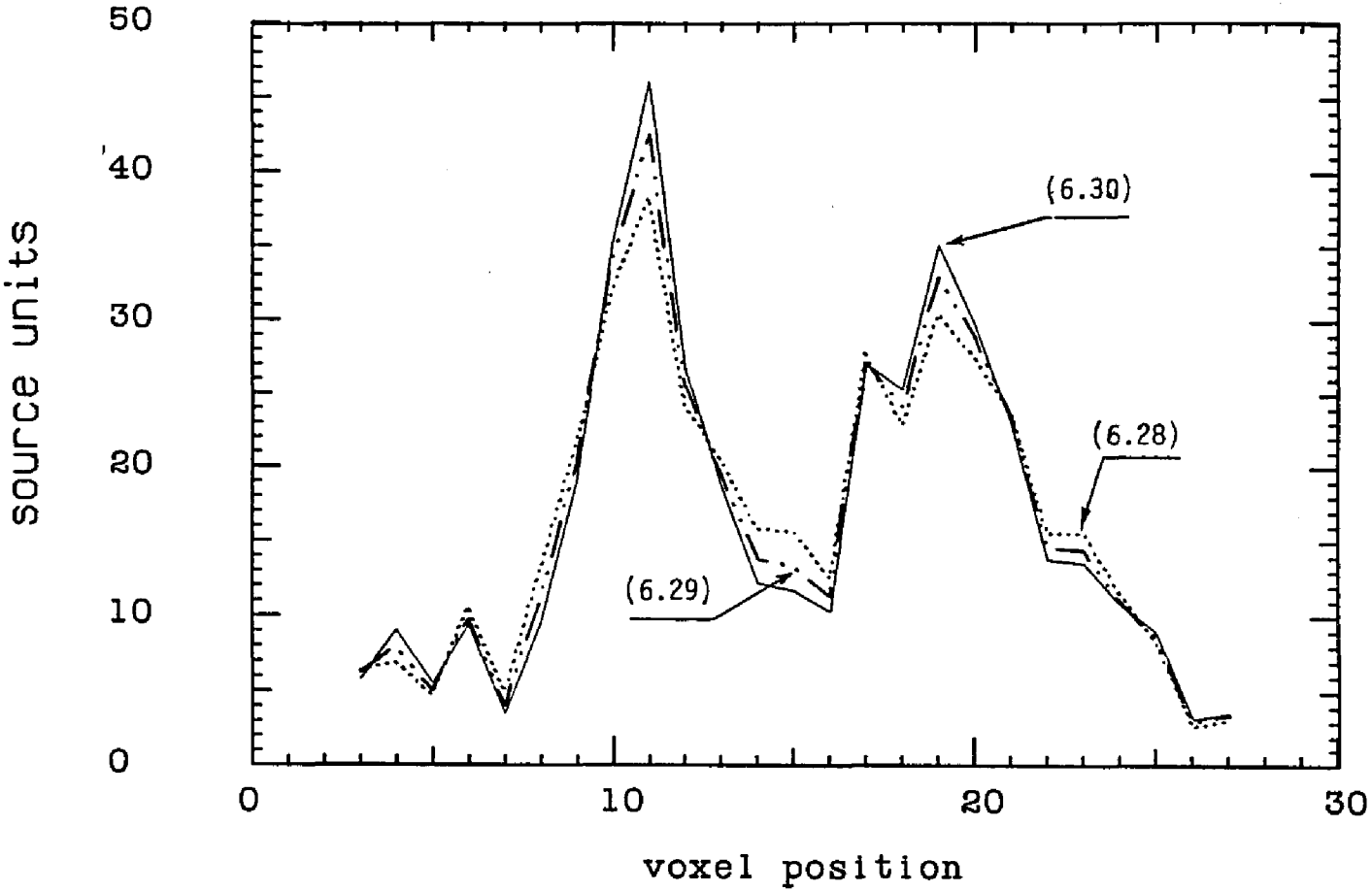


Fig.6.6: Results using algorithms (6.28), (6.29) and (6.30) on a single set of gaussian randomized data for 100 iterations.

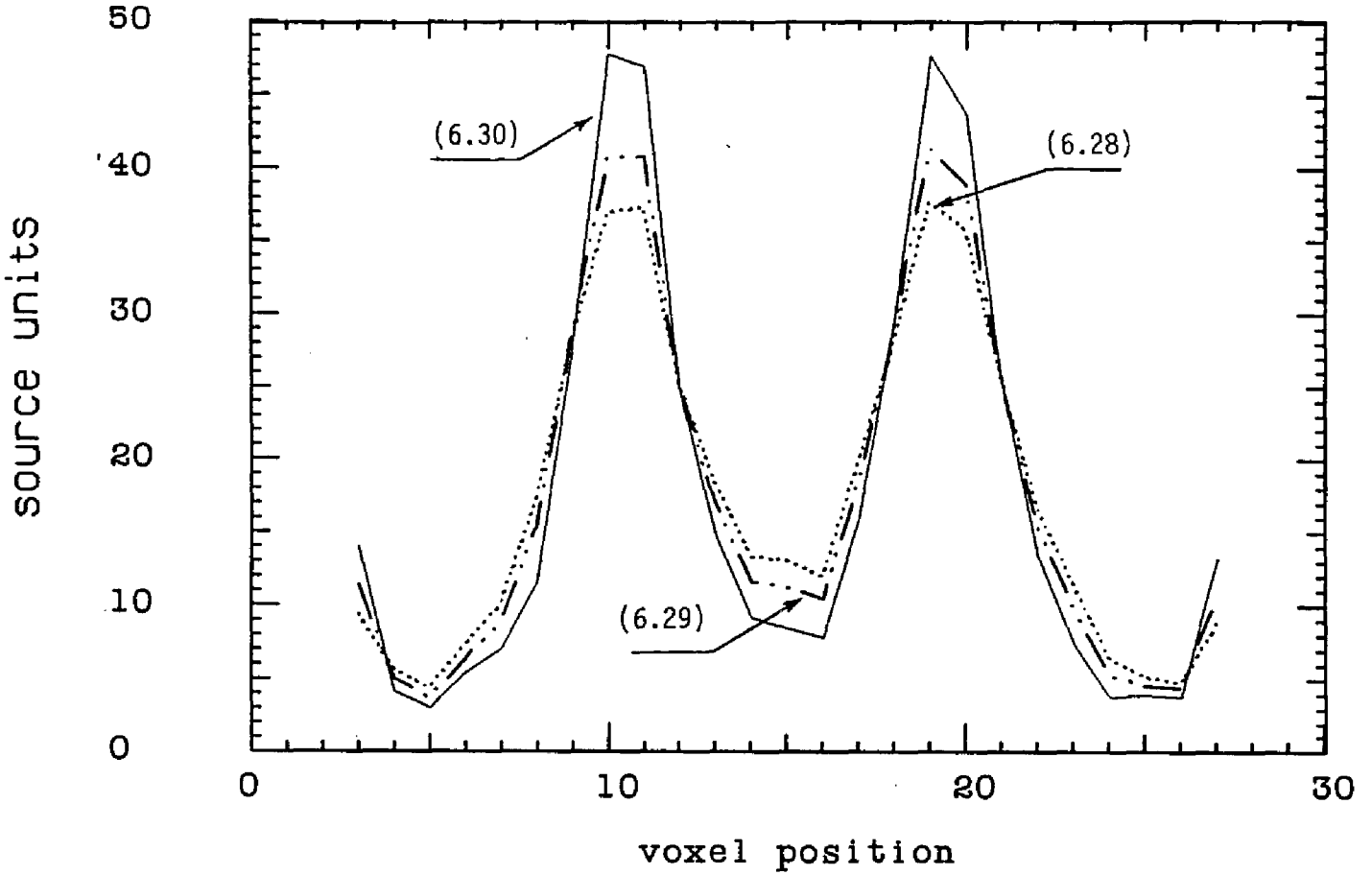


Fig.6.7: Comparison of (6.28), (6.29) and (6.30) for ten sets of gaussian random data and 100 iterations.

CHAPTER VII

BAYESIAN IMAGE PROCESSING OF DATA FROM QUANTITATIVE CONSTRAINED SOURCE DISTRIBUTIONS

Quantitative constrained image processing applications can occur in radioisotopic imaging situations in which it may be possible to estimate the background value and hot or cold lesion values. For example, in cardiac imaging studies it is possible to determine the radioisotope concentration of the blood filling the cardiac chambers as a function of time and to approximate the average radioisotope concentration of surrounding heart tissues. In brain and liver imagings, the spatial relations may be known, again for example, between radioisotopic hot spots or cold spots (i.e., patterned-valued constrained cases). In these situations, the a priori source information function $P(\Phi)$ is much more restrictive and so striking improvement in image processing results is achievable if $P(\Phi)$ is reasonably formulated.

1). Uncorrelated Constraints

In terms of the standard probability notation, Bayes' Law is expressed as Eq.(1.2).

The $P(Y|\Phi)$ in Eq.(1.2) was given by Eq.(3.48) and $P(Y)$ is omitted since is independent of Φ.

The a priori probability function $P(\Phi)$ in Eq.(1.2) can be formulated as follows:

If the permitted hot spot (point source) values are $\{\phi_s^e\}$, (s=1,2,...,β. β is the number of spot values and the $\{\phi_s^e\}$ values are widely spaced), then assuming random superposition on a uniform background level of ϕ_b^e, $P(\Phi)$ can be approximated as (Eq.(4.27)):

$$P(\Phi)=\prod_{j=1}^{J} C_j \exp[-\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}\prod_s(\frac{(\phi_j-\phi_s^e)^2}{2\rho_s^2})] \tag{7.1}$$

where C_j is a normalizing constant and J the total number of voxels, and ϕ_j is most likely to be ϕ_b^e or ϕ_s^e. More exact forms of $P(\Phi)$ appear in chapter IV.

From Eqs.(1.2), (3.48) and (7.1), the ln Bayesian function is:

$$g(\Phi)=\sum_i[-\sum_j R_{ij}\phi_j + Y_i ln(\sum_j R_{ij}\phi_j)] - \sum_j[\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}\prod_s(\frac{(\phi_j-\phi_s^e)^2}{2\rho_s^2})] + C(...) \tag{7.2}$$

where $C(...)=-\sum_i ln(Y_i!)-\sum_j \ln C_j$.

For simplicity, consider the case in which all hot spots have the same value ϕ_1^e; the general Eq.(7.2) reduces to:

$$g(\Phi)=\sum_i[-\sum_j R_{ij}\phi_j+Y_i ln(\sum_j R_{ij}\phi_j)]$$
$$-\sum_j[\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}\frac{(\phi_j-\phi_1^e)^2}{2\rho_1^2}]+C(...). \tag{7.3}$$

Setting $\frac{\partial g(\Phi)}{\partial \phi_k}=0$:

$$\sum_i R_{ik}[\frac{Y_i}{\sum_j R_{ij}\phi_j^*}-1]=\mu_k Z_{kb}^* Z_{k1}^* \frac{(Z_{kb}^*+Z_{k1}^*)}{(2\rho_b^2\rho_1^2)} \tag{7.4}$$

where

$$Z_{kb}^*=\phi_k^*-\phi_b^e\ ,\ Z_{k1}^*=\phi_k^*-\phi_1^e$$

and μ_k is an adjustable parameter > 0.

The first summation in Eq.(7.3) is strictly concave and therefore defines the unique maximum likelihood solution in the absence of a priori source distribution information (Shepp et al 1982 and Lange et al 1984) converging presumably to the neighborhood of the actual source distribution. Once this has been done, the Bayesian contributions arising from the a priori source distribution information represented by the quadratic forms occuring in the second summation (they are then obviously locally concave) should result in still greater negativity and rapid convergence to the generalized Bayesian maximum likelihood solution. The required shift in emphasis is accomplished by the adjustable parameter μ_k. The choice of μ_k and the other adjustable parameters introduced subsequently (ξ_k, ν_k and η_k) will be considered in the section 3 when implementation of the algorithms derived below is described.

The iterative algorithms are derived by using the EM technique.

The ln Bayesian function for the "complete-data" $\{X_{ij}\}=\mathbf{X}$ is:

$$\ln f(\mathbf{X},\Phi)=\sum_{ij}[-R_{ij}\phi_j+X_{ij}ln(R_{ij}\phi_j)]-\sum_j[\frac{(\phi_j-\phi_b^e)^2}{2\rho_b^2}\frac{(\phi_j-\phi_1^e)^2}{2\rho_1^2}]+C_0(...) \tag{7.5}$$

where $C_0(...)=-\sum_{ij} ln(X_{ij}!)-\sum_j ln(C_j)$.

The E-step (expectation) of the EM technique computes the conditional average value of

$$Q(\Phi\mid\Phi^{(n)})=E[\ln f(\mathbf{X},\Phi)\mid\mathbf{Y},\Phi^{(n)}]$$
$$=\sum_{ij}[-R_{ij}\phi_j+X_{ij}^{(n)}ln(R_{ij}\phi_j)]-\sum_j[\frac{(\phi_j-\phi_b^e)^2}{\rho_b^2}\frac{(\phi_j-\phi_1^e)^2}{\rho_1^2}]+C_0(...) \tag{7.6}$$

where $\Phi^{(n)}$ represents the n-th iterated result and $X_{ij}^{(n)}$ is given by Eq.(5.13).

The M-step (maximization) of the EM technique specifies the (n+1)th iterative result:

$$\sum_i [\frac{X_{ik}^{(n)}}{\phi_k^{(n+1)}} - R_{ik}] = \mu_k [\frac{(\phi_k^{(n+1)}-\phi_b^e)}{\rho_b^2}\frac{(\phi_k^{(n+1)}-\phi_1^e)^2}{2\rho_1^2} + \frac{(\phi_k^{(n+1)}-\phi_b^e)^2}{2\rho_b^2}\frac{(\phi_k^{(n+1)}-\phi_1^e)}{\rho_1^2}] . \quad (7.7)$$

The right hand side of Eq.(7.7) is the valued constraint term. Substituting

$$\phi_k^{(n+1)} = \phi_k^{(n)} + \tau d_k^{(n)} \text{ and } \tau = 1$$

where $d_k^{(n)} \ll \phi_k^{(n)}$, on the right hand side of Eq.(7.7):

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik}(Y_i / \sum_j R_{ij}\phi_j^{(n)})}{\sum_i R_{ik} + \mu_k [Z_{kb}^{(n)} Z_{k1}^{(n)} (Z_{kb}^{(n)} + Z_{k1}^{(n)})/(2\rho_b^2\rho_1^2)]} \quad (7.8)$$

where

$$Z_{kb}^{(n)} = \phi_k^{(n)} + \tau d_k^{(n)} - \phi_b^e \ , \quad Z_{k1}^{(n)} = \phi_k^{(n)} + \tau d_k^{(n)} - \phi_1^e$$

and where $d_k^{(n)}$ can be approximated $\simeq \phi_k^{(n)} - \phi_k^{(n-1)}$.

It is clear that in the limit as n goes to infinity, $\phi_k^{(n+1)} = \phi_k^{(n)}$, algorithm (7.8) generates the solution $\{\phi_k^*\}$ of Eq.(7.4).

For the general case of Eq.(7.2), the iterative algorithm corresponding to Eq.(7.8) is:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik}(Y_i / \sum_j R_{ij}\phi_j^{(n)})}{\sum_i R_{ik} + \mu_k Z_{kb}^{(n)} M_k^{(n)}/\rho_b^2} \quad (7.9)$$

where

$$M_k^{(n)} = \prod_s \frac{Z_{ks}^{(n)}}{2\rho_s^2} \{\prod_s Z_{ks}^{(n)} + Z_{kb}^{(n)} [\prod_{s\neq 1} Z_{ks}^{(n)} + \prod_{s\neq 2} Z_{ks}^{(n)} + \cdots + \prod_{s\neq\beta} Z_{ks}^{(n)}]\} . \quad (7.10)$$

If each data element Y_i now obeys gaussian statistics with all $\{Y_i\}$ remaining statistically independent, the Bayesian image equations (Eq.(7.4)) become:

$$\sum_i R_{ik}(1 - \frac{\sum_j R_{ij}\phi_j^*}{Y_i}) = \xi_k Z_{kb}^* M_k^* / \rho_b^2 \quad (7.11)$$

where

$$M_k^* = \prod_s \frac{Z_{ks}^*}{2\rho_s^2} \{\prod_s Z_{ks}^* + Z_{kb}^* [\prod_{s\neq 1} Z_{ks}^* + \prod_{s\neq 2} Z_{ks}^* + \cdots + \prod_{s\neq \beta} Z_{ks}^*]\}$$

and ξ_k is an adjustable parameter > 0.

The iterative algorithm (Eq.(7.9)) becomes:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i R_{ik} - \xi_k Z_{kb}^{(n)} M_k^{(n)} / \rho_b^2}{\sum_i R_{ik} (\sum_j R_{ij} \phi_j^{(n)} / Y_i)}. \tag{7.12}$$

2). Correlated Constraints

In this section, the spatial relations between hot spots (i.e., patterns) are also assumed known. For a single multiple point source pattern of hot spots sparsely superimposed at random upon a uniform background, function $P(\Phi)$ can be chosen as (Eq.(4.28)):

$$P(\Phi) = \prod_j C_j \exp[-\frac{(\phi_j - \phi_b^e)^2}{2\rho_b^2} \prod_q (\sum_s \frac{(\phi_{j+ls} - \phi_s^e)^2}{2\rho_s^2})] \tag{7.13}$$

where the index q covers the possible pattern configurations and the index ls represents the spatial relations.

For two hot spots of value ϕ_1^e and ϕ_2^e respectively, separated by l voxel elements, in the one-dimensional case, Eq(7.13) reduces to:

$$P(\Phi) = \prod_j C_j \exp\{-\frac{(\phi_j - \phi_b^e)^2}{2\rho_b^2} [\frac{(\phi_j - \phi_1^e)^2}{2\rho_1^2} + \frac{(\phi_{j+l} - \phi_2^e)^2}{2\rho_2^2}] * [\frac{(\phi_j - \phi_2^e)^2}{2\rho_2^2} + \frac{(\phi_{j-l} - \phi_1^e)^2}{2\rho_1^2}]\} \tag{7.14}$$

and again assuming poisson data, Eq.(7.4) becomes:

$$\sum_i [\frac{Y_i}{\sum_j R_{ij} \phi_j^*} - 1] = \nu_k \frac{Z_{kb}^*}{4\rho_b^2} \{Z_{k+l}^* Z_{k-l}^* + Z_{kb}^* [Z_{k+l}^* \frac{Z_{k2}^*}{\rho_2^2} + Z_{k-l}^* \frac{Z_{k1}^*}{\rho_1^2}]\} \tag{7.15}$$

where

$$Z_{k+l}^* = (\phi_k^* - \phi_1^e)^2/(\rho_1^2) + (\phi_{k+l}^* - \phi_2^e)^2/(\rho_2^2)$$
$$Z_{k-l}^* = (\phi_k^* - \phi_2^e)^2/(\rho_2^2) + (\phi_{k-l}^* - \phi_1^e)^2/(\rho_1^2)$$

Z_{k1}^* and Z_{k2}^* are defined by Eq.(7.4)ff and parameter ν_k serves the same purpose as μ_k.

An iterative algorithm for solving Eq.(7.15) is again formulated using the EM technique:

$$\phi_k^{(n+1)}=\phi_k^{(n)}\frac{\sum_i R_{ik}\left(Y_i/\sum_j R_{ij}\phi_j^{(n)}\right)}{\sum_i R_{ik}+\nu_k\,\frac{Z_{kb}^{(n)}}{4\rho_b^2}L_k^{(n)}} \tag{7.16}$$

where

$$L_k^{(n)}=Z_{k+l}^{(n)}Z_{k-l}^{(n)}+Z_{kb}^{(n)}\left[Z_{k+l}^{(n)}\frac{Z_{k2}^{(n)}}{\rho_2^2}+Z_{k-l}^{(n)}\frac{Z_{k1}^{(n)}}{\rho_1^2}\right] \tag{7.17}$$

where $Z_{k+l}^{(n)}$ is Z_{k+l}^{*} with replacement of ϕ_k^{*} by $\phi_k^{(n)}+\tau d_k^{(n)}$ and correspondingly for $Z_{k-l}^{(n)}$, $Z_{k1}^{(n)}$ and $Z_{k2}^{(n)}$.

The general iterative algorithm for any number of spatially related hot spots (Eq.(7.13)) is:

$$\phi_k^{(n+1)}=\phi_k^{(n)}\frac{\sum_i R_{ik}\left(Y_i/\sum_j R_{ij}\phi_j^{(n)}\right)}{\sum_i R_{ik}+\nu_k\,\Xi_k^{(n)}} \tag{7.18}$$

where

$$\Xi_k^{(n)}=-\frac{\partial}{\partial\phi_k}\{\ln P(\Phi)\}\,|_{\Phi=\Phi^{(n)}+d^{(n)}}. \tag{7.19}$$

The gaussian data correlated equivalents of Eqs.(7.11) and (7.12) are:

$$\sum_i R_{ik}\left[1-\frac{\sum_j R_{ij}\phi_j^{*}}{Y_i}\right]=\eta_k\Xi_k^{*} \tag{7.20}$$

and

$$\phi_k^{(n+1)}=\phi_k^{(n)}\frac{\sum_i R_{ik}-\eta_k\Xi_k^{(n)}}{\sum_i R_{ik}\left(\sum_j R_{ij}\phi_j^{(n)}/Y_i\right)} \tag{7.21}$$

where $\Xi_k^{*}=\Xi_k^{(n)}$ with $\Phi^{(n)}=\Phi^{*}$ and $\mathbf{d}^{(n)}=\mathbf{d}^{*}=0$ and parameter η_k serves the same purpose as ξ_k.

3). Simulation Results

The algorithms (7.8), (7.12), (7.16) and (7.21) were each run using ideal, and as appropriate, poisson and gaussian randomized data for a one-dimensional system in which the elements $\{R_{ij}\}$ of matrix R (PSF) are assumed to exhibit the exponential dependence of Eq.(6.36).

The actual source $\{O_j\}$ and ideal and randomized data are shown in Fig.(6.1) and correspond to that used in the preceding section.

The initial values $\{\phi_k^{(0)}\}$ were chosen to be proportional to the observed data as shown before in Eq.(6.37).

The exact numerical values or even the functional forms chosen for the adjustable parameters introduced in sections 1 and 2 are in general not critical; a very gradual sigmoidal monotonic dependence on the iterative index n should be exhibited as discussed in section 1, i.e.,

$$D_k(n) \simeq \frac{a\, n^{\kappa}}{b + n^{\kappa}} \sum_i R_{ik} \tag{7.22}$$

where κ, a and b represent numerical values and $D_k(n)$ the parametric dependence in the algorithms.

For example, to obtain smooth convergence, the monotonic dependence Eq.(7.22) is imposed on the factor $\mu_k Z_{k1}^{(n)} (Z_{k1}^{(n)} + Z_{kb}^{(n)})/\rho_b^2$ of Eq.(7.8), where,

$$\mu_k = \frac{a}{b + n^{\kappa}} \sum_i R_{ik} \tag{7.23}$$

and

$$\rho_b^2 = \phi_b^e/\sqrt{2n} \quad , \quad \phi_s^e \rightarrow \phi_b^e + \sqrt{n}\,(\phi_s^e - \phi_b^e)/\sqrt{L} \tag{7.24}$$

here L is the total number of iterations.

The values $\kappa=2$, $a=7.7$ and $b=250$ were chosen for μ_k and ξ_k; and for ν_k and η_k, $\kappa=3$, $a=10$ and $b=3500$.

Fig.(7.1) shows the results obtained using algorithms (7.8), (7.12), (7.16), (7.21) and the maximum likelihood algorithm (6.19) (MLEM, Eq.(6) of Lange et al 1984) on ideal data for 100 iterations.

Fig.(7.2) shows the results of algorithms (7.8) and (7.16) for poisson randomized data, in which the curves labelled (7.8) and (7.16) resulted from a single set of data for which the median data element was $\simeq 30$ while $(7.8)^*$ and $(7.16)^*$ show the average of the corresponding results obtained from ten sets of random data. The results of algorithms (7.12) and (7.21) for gaussian randomized data are shown in Fig.(7.3) in which the labels have the same meaning as those in Fig.(7.2).

In Figs.(7.2) & (7.3), the "hot spots" obtained using algorithms (7.8) and (7.12) for a single set of random data fall on voxels 11 and 19, (i.e., the true positions of the point sources) and the two "hot spots" obtained using algorithms (7.16) and (7.21) shift around the positions (11,19) keeping the spatial separation constant.

Algorithm (7.18) was also tested for the case in which three point sources of value 40, 60 and 60 units are separated by 7 and 8 voxel elements respectively and are superimposed on a one-dimensional uniform background of value 10 units, as shown in Fig.(7.4). The uniform background now extends from -2 to +32 and the observed data extends seven voxel elements beyond both sides of the graph (from -7 to +37), i.e., the matrix R has 45 rows and 35 columns.

The results using algorithm (7.18) for 100 iterations and algorithm (MLEM) for 200 iterations on ideal data are also shown in Fig.(7.4).

Fig.(7.5) compares the results of algorithms of (7.18) and (MLEM) applied to poisson randomized data for which the median data element was ≈ 80. Curves indicated by (7.18) and (MLEM) were obtained using a single set of data while $(7.18)^*$ and $(MLEM)^*$ indicate the average result from the same ten sets of randomized data. Almost identical comparative results were obtained with gaussian randomized data using algorithms (7.21) and Eq.(6.28).

An obviously striking improvement in image processing using the new algorithms is exhibited in Figs.(7.1), (7.4) & (7.5) reflecting the validity of the a priori source distribution information employed. It should be emphasized that the validity of the a priori source distribution information is absolutely crucial. In the absence of a priori source information, algorithm (6.19) or MLEM remains in fact clearly the method of choice. The new algorithms provide maximum likelihood solutions taking the various classes of a priori source distribution information into account. Their solutions differ from the standard maximum likelihood results only because of the additional information assumed to be known.

To determine the results obtained with qualitatively correct, but quantitatively inexact a priori source information, algorithm (7.16) was tested with a double point source having an assumed separation of $l = 8 + \Delta$ and an actual separation 8 voxels.

When $-2 \leq \Delta \leq 4$, the results indicate a double point source of magnitude l voxels apart centered within one voxel of the true position. For $\Delta < -2$ and $\Delta > 4$ the point sources are no longer resolved.

For the same assumed double point source configuration, the actual source distribution was then chosen to be the uniform background with no superimposed any hot (or cold) spots. Algorithm (7.16) gave rise to the correct background distribution with no pattern superposition.

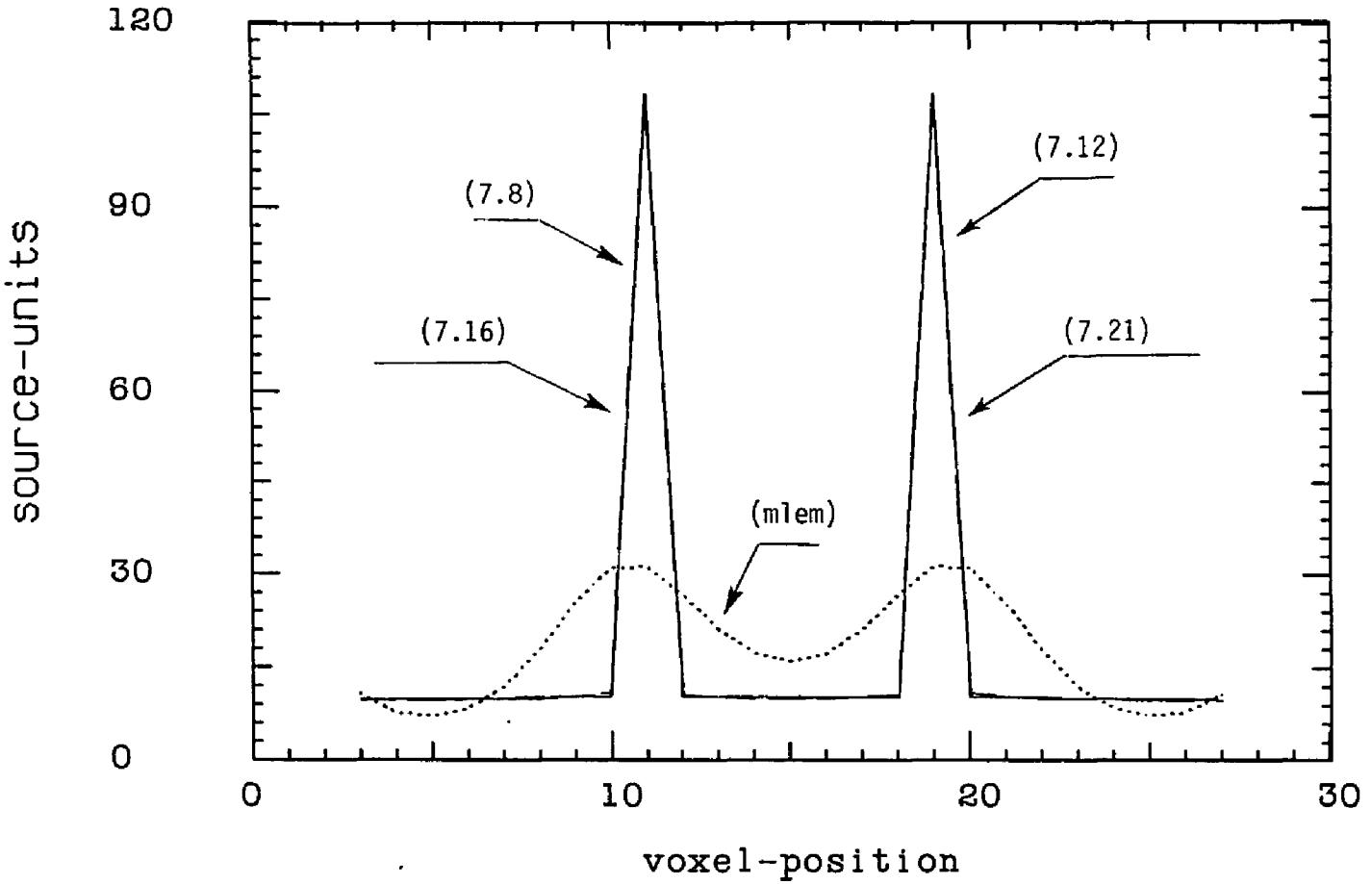


Fig.7.1: Results after 100 iterations using BIP algorithms (7.8), (7.12), (7.16), (7.21) and ML algorithm (mlem) on ideal data.

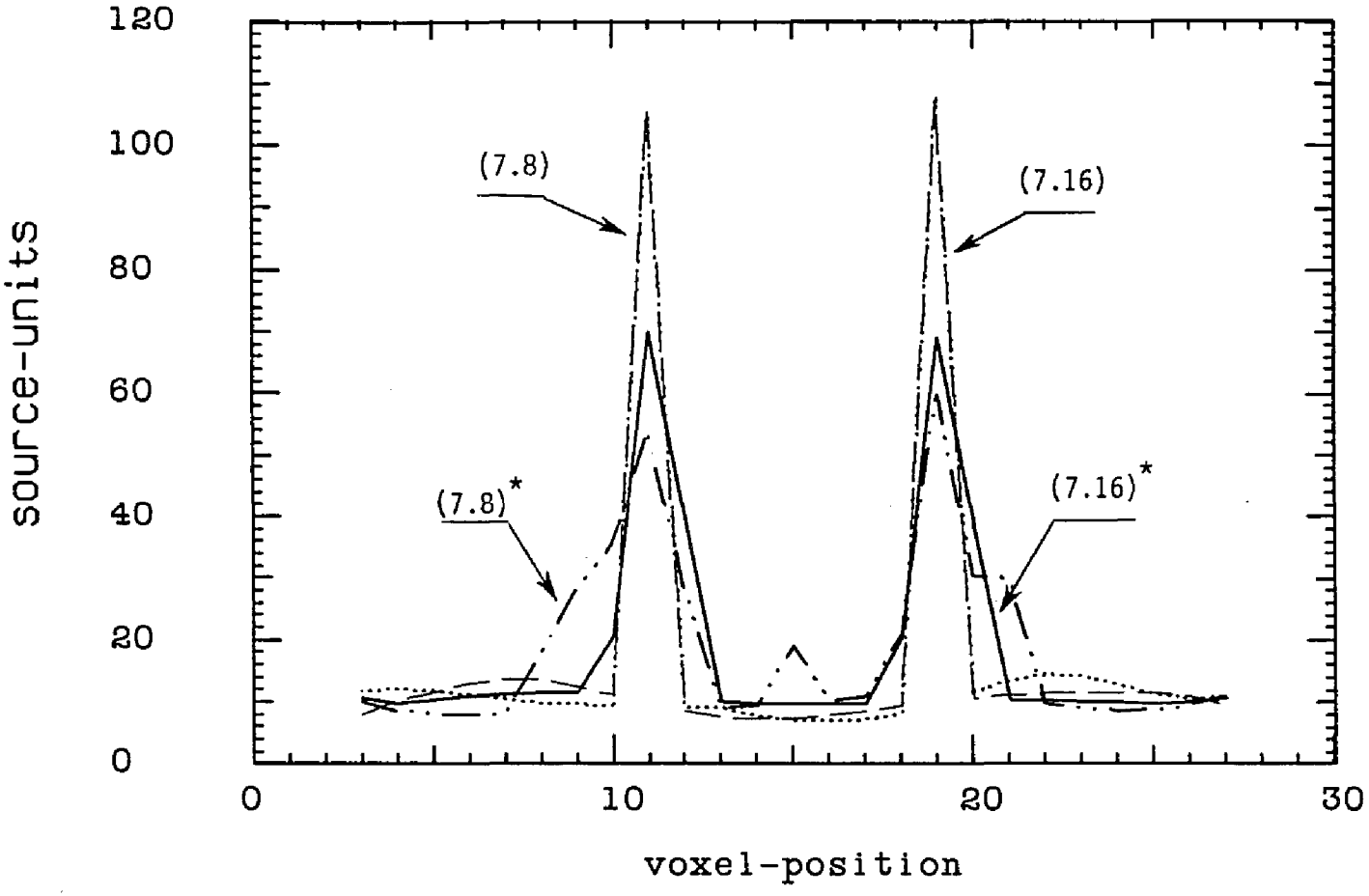


Fig.7.2: Comparison of algorithms (7.8) and (7.16) for a single set (labelled by (7.8) and (7.16) respectively) and ten sets (labelled by (7.8)* and (7.16)* respectively) of poisson randomized data and 100 iterations.

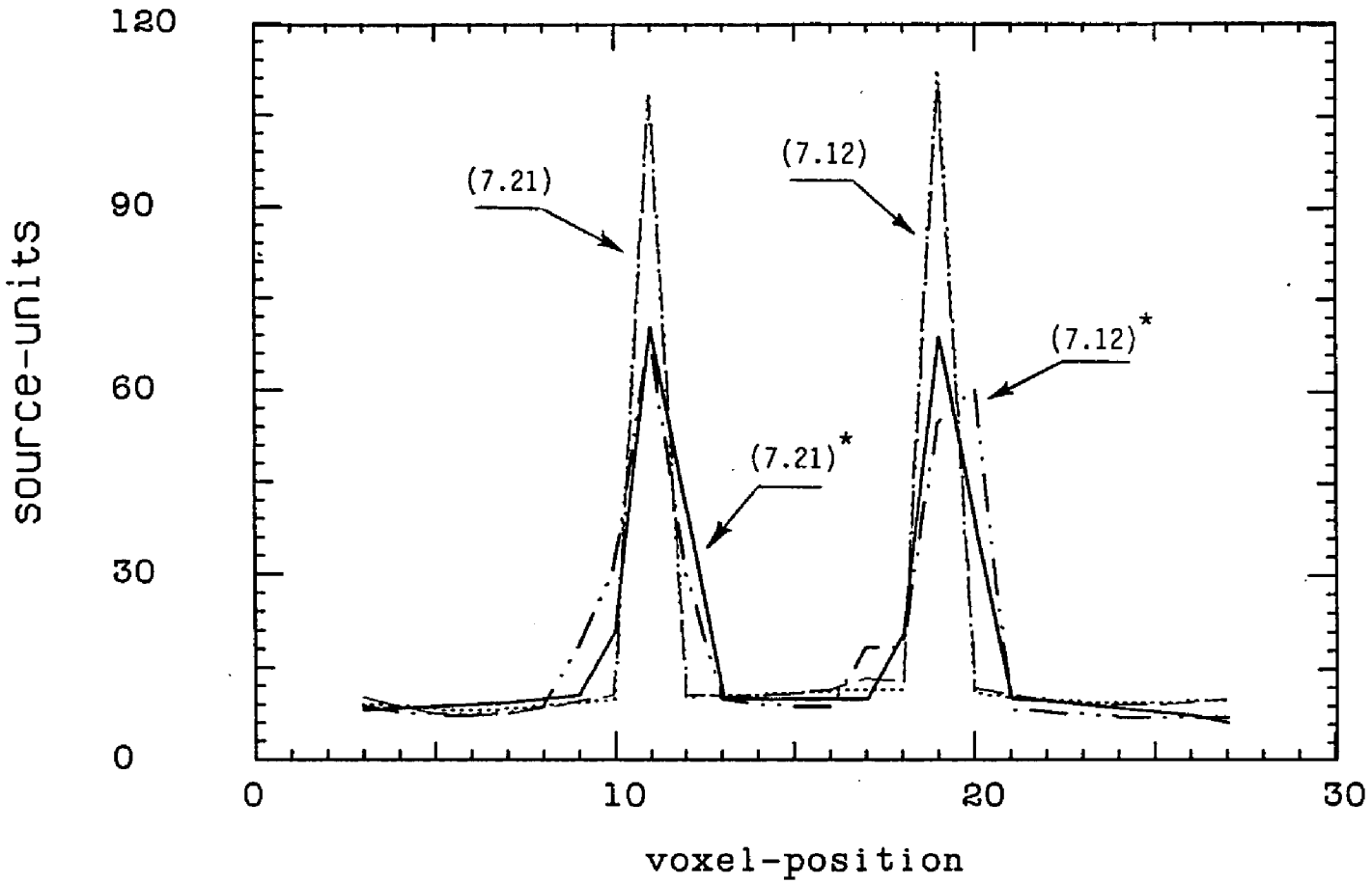


Fig. 7.3: Comparison of algorithms (7.12) and (7.21) for a single set (labelled by (7.12) and (7.21) respectively) and ten sets (labelled by $(7.12)^*$ and $(7.21)^*$ respectively) of gaussian randomized data and 100 iterations.

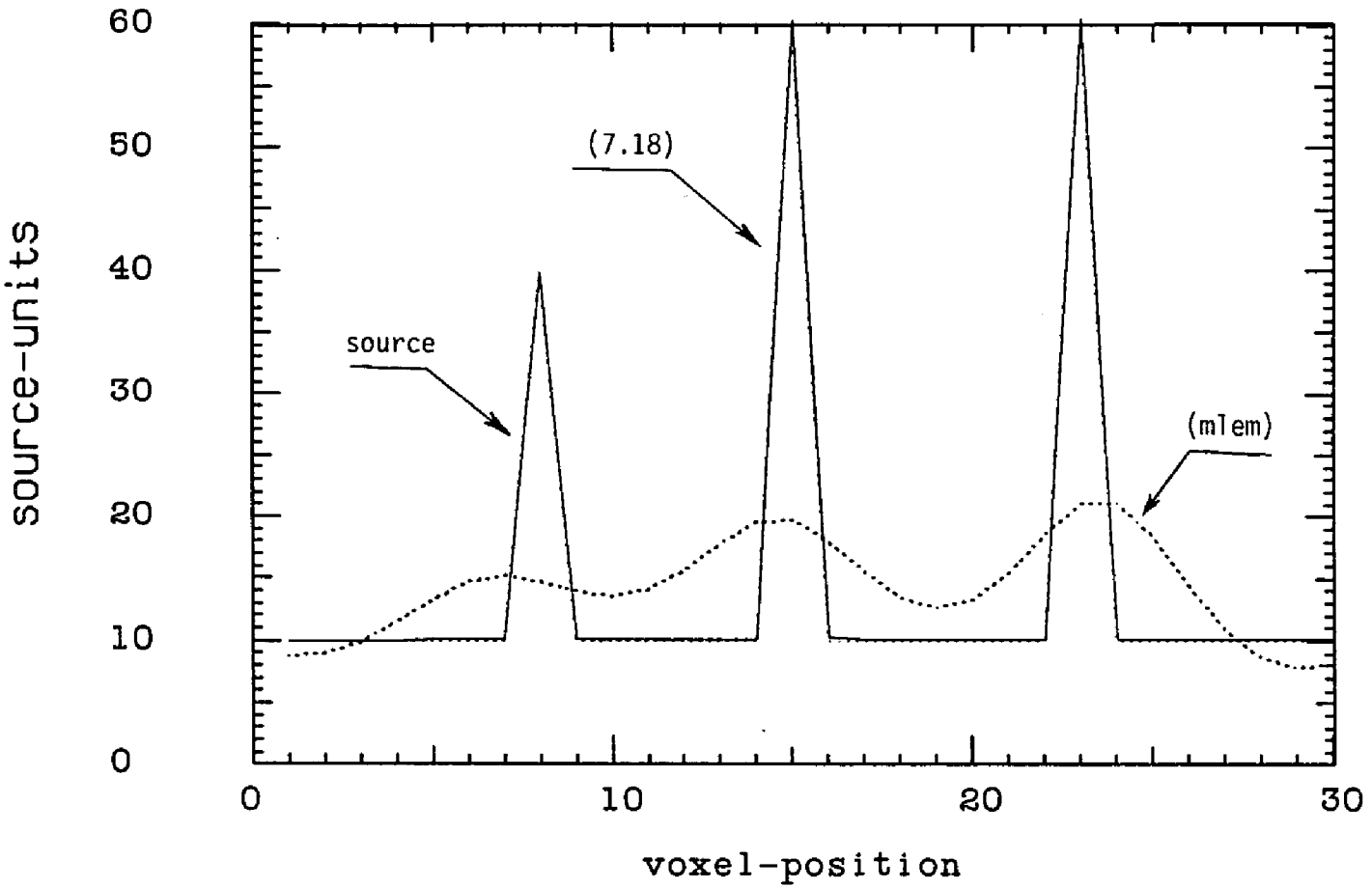


Fig.7.4: One dimensional source distribution (solid line) and results using BIP algorithm (7.18) (broken line) for 100 iterations and ML algorithm (mlem) (dotted line) for 200 iterations on ideal data.

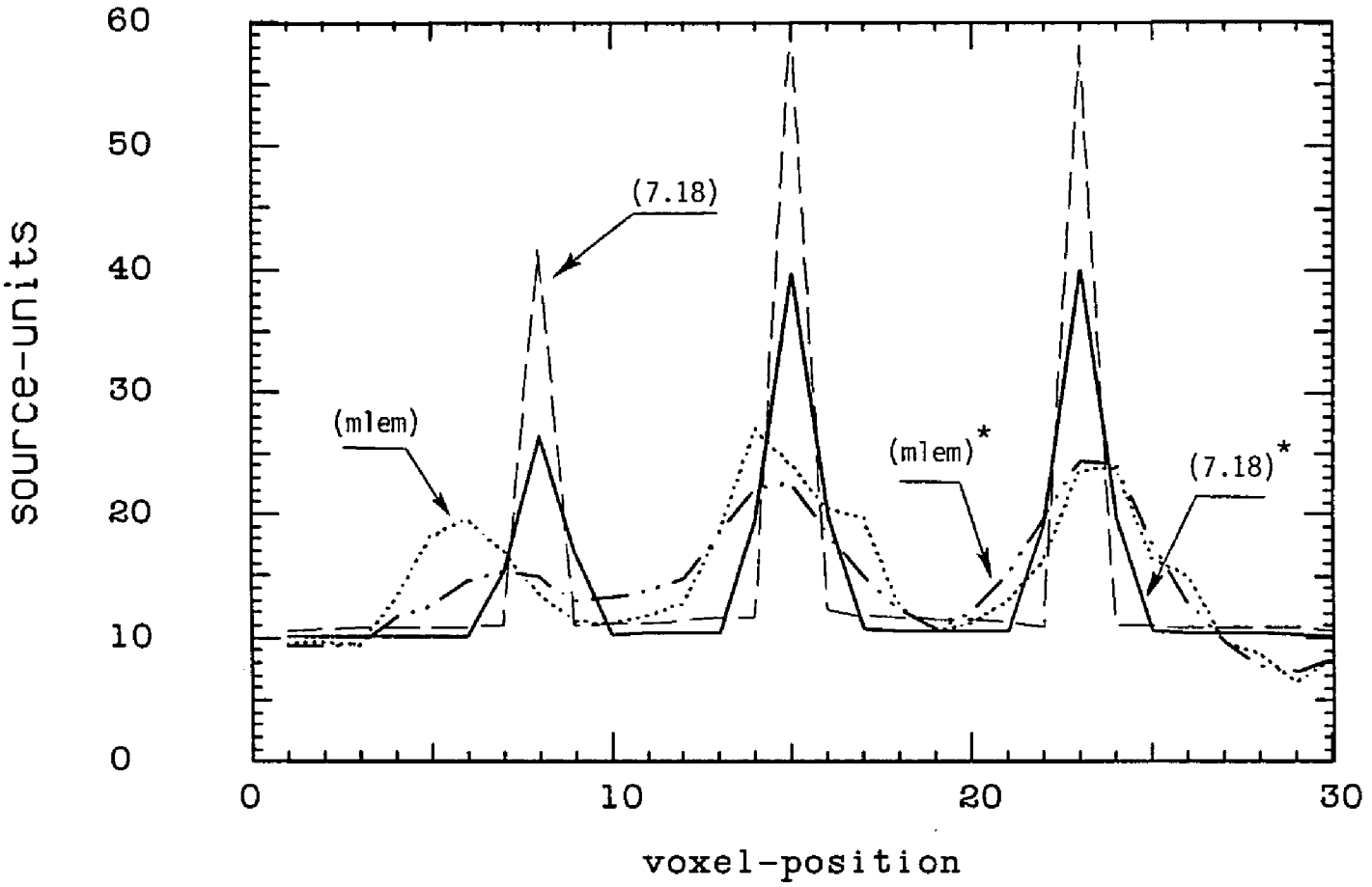


Fig. 7.5: Comparison of BIP algorithm (7.18) and ML algorithm (mlem) for a single set (labelled by (7.18) and (mlem) respectively) and ten sets (labelled by $(7.18)^*$ and $(mlem)^*$ respectively) of poisson randomized data and 100 iterations for BIP and 200 iterations for ML.

CHAPTER VIII

BAYESIAN IMAGE PROCESSING OF DATA FROM FUZZY CONSTRAINED SOURCE DISTRIBUTIONS

In certain imaging situations, a priori quantitative information about a source distribution may exist and can be characterized in terms of a probability density function $P(\Phi)$ of the source elements $\{\phi_j\}=\Phi$ $(j=1,2,...,\alpha)$. In radioisotopic imaging applications, $P_j(\phi_j)$ can be the a priori probability distribution for the photon emission from voxel j per unit time at time $t=0$. For example, in cardiac imaging studies it is possible to determine the radioisotope concentration of the blood filling the cardiac chambers as a function of time by peripheral measurement and to estimate the average radioisotope concentration of surrounding heart tissues. In brain imaging, hepatic imaging and thyroid imaging studies, probable configurations of cold or hot lesions may be characterizible in terms of estimated amplitudes and spatial relations between source elements. Lesion detection can then be viewed as a search for the anticipated fuzzy patterns superimposed upon a relatively uniform background. For such cases, a BIP formalism which incorporates both a priori amplitude and spatial information and the statistical character of measured data (or projections) applies as shown in previous chapters.

1). The BIP Formalism

The BIP formalism is directed at finding that source distribution $\Phi^* \varepsilon \Omega$ which is most likely to have given rise to the measured data, subject to the data vector $\mathbf{Y}$, the matrix R and any a priori source distribution information which may be available (e.g. the $P(\Phi)$ function (4.34)). Mathematically this objective is equivalent to maximizing the conditional probability density function $P(\Phi|\mathbf{Y})$ where from Bayes' Law (Lambert 1968) of Eq.(1.2).

The optimal solution Φ^* is then specified by maximizing $P(\Phi|\mathbf{Y})$ or equivalently $g(\Phi) \equiv ln(P(\Phi|\mathbf{Y}))$ --- algebraically, by solving the system of equations:

$$\frac{\partial g(\Phi)}{\partial \phi_k}\Big|_{\Phi^*}=0 \tag{8.1}$$

Substituting for $P(\mathbf{Y}|\Phi)$ of Eq.(3.48) of independent poisson random data and $P(\Phi)$ of Eq.(4.34), and neglecting $P(\mathbf{Y})$ since it is not a function of Φ, Eq.(8.1) results:

$$\sum_i R_{ik}\left(\frac{Y_i}{\sum_j R_{ij}\phi_j^*}-1\right)=\frac{L_k(\phi_k^*)+M_k(\phi_k^*)}{Z_k(\phi_k^*)} \tag{8.2}$$

where

$$Z_k(\phi_k^*)=\sum_q\sum_l W(l)\,[\prod_s C_s\exp(-\frac{(\phi^*_{k+L_w}-\phi_s^e)^2}{2\rho_s^2})]+C_b\exp[-\frac{(\phi_k^*-\phi_b^e)^2}{2\rho_b^2}]$$

$$L_k(\phi_k^*)=\frac{(\phi_k^*-\phi_b^e)}{\rho_b^2}C_b\exp[-\frac{(\phi_k^*-\phi_b^e)^2}{2\rho_b^2}] \qquad (8.3)$$

$$M_k(\phi_k^*)=\sum_q[\frac{(\phi_k^*-\phi_q^e)}{\rho_q^2}\sum_l W(l)\prod_s C_s\exp(-\frac{(\phi^*_{k+L_w}-\phi_s^e)^2}{2\rho_s^2})]$$

and the α, β and Q in Eq.(4.34) have been incorporated into the C_s and C_b.

It is clear that the function (3.48) is strictly concave if matrix R has at least as many rows as columns and is of full column rank (Lange et al 1984). The function (4.34) is locally concave (i.e., for an iterative solution already in the neighborhood of the maximal vector Φ^*). Therefore, Eq.(8.2) serves to determine that solution Φ^* which would most likely give rise to the measured data, subject to the probabilistic constraint of the estimated strength and spatial separation values of $\{\phi_s^e,\ \phi_b^e\ \&\ l_r\}$. More detail discussion about concavity is given in the Appendix.

For a low resolution imaging system (i.e., a R with an extended FWHM) and large dimension Φ and $\mathbf{Y}$, an iterative approach to the solution Φ^* is preferred and the EM technique can be then employed to derive a BIP algorithm in multiplicative form.

The ln Bayesian function of $P(\Phi|\mathbf{X})$ is:

$$\ln f(\mathbf{X},\Phi)\equiv \ln P(\Phi|\mathbf{X})$$
$$=\sum_{ij}[-R_{ij}\phi_j+X_{ij}\ln(R_{ij}\phi_j)-\ln(X_{ij}!)]+\sum_j \ln Z_j(\phi_j) \qquad (8.4)$$

with $Z_j(\phi_j)$ as defined in Eq.(8.3).

The E-step (expectation) of the EM technique computes the conditional average value of

$$Q(\Phi|\Phi^{(n)})=E[\ln f(\mathbf{X},\Phi)|\mathbf{Y},\Phi^{(n)}]$$
$$=\sum_{ij}[-R_{ij}\phi_j+X_{ij}^{(n)}\ln(R_{ij}\phi_j)-\ln(X_{ij}^{(n)}!)]+\sum_j \ln Z_j(\phi_j) \qquad (8.5)$$

where $\Phi^{(n)}$ represents the n-th iterated result and $X_{ij}^{(n)}$, the conditional average value of random variable X_{ij} for the given projection i (Y_i), was given by Eq.(5.13).

The M-step (maximum) of the EM technique generates the $(n+1)$th iteration result $\Phi^{(n+1)}$ which maximizes the function $Q(\Phi|\Phi^{(n)})$, that is:

$$\frac{\partial Q(\Phi|\Phi^{(n)})}{\partial\phi_k}\Big|_{\Phi=\Phi^{(n+1)}}=0 \qquad (8.6)$$

and

$$\sum_i(-R_{ik}+\frac{X_{ik}^{(n)}}{\phi_k^{(n+1)}})=\frac{L_k(\phi_k^{(n+1)})+M_k(\phi_k^{(n+1)})}{Z_k(\phi_k^{(n+1)})} \tag{8.7}$$

with $Z_k(\phi_k^{(n+1)})$, $L_k(\phi_k^{(n+1)})$ and $M_k(\phi_k^{(n+1)})$ defined as in Eqs.(8.3)ff.

The $\phi_k^{(n+1)}$ on the right hand side of Eq.(8.7) can be approximated as $\phi_k^{(n+1)}=\phi_k^{(n)}+\tau d_k^{(n)}$, the BIP algorithm then becomes:

$$\phi_k^{(n+1)}=\phi_k^{(n)}\frac{\sum_i R_{ik}(Y_i/\sum_j R_{ij}\phi_j^{(n)})}{\sum_i R_{ik}+\xi_k^{(n)}[\frac{L_k^{(n)}(\phi_k^{(n)}+\tau d_k^{(n)})+M_k^{(n)}(\phi_k^{(n)}+\tau d_k^{(n)})}{Z_k^{(n)}(\phi_k^{(n)}+\tau d_k^{(n)})}]} \tag{8.8}$$

where τ and $d_k^{(n)}$ are approximated by:

$$\tau=1 \quad \text{and} \quad d_k^{(n)}=\phi_k^{(n)}-\phi_k^{(n-1)} \tag{8.9}$$

and $\xi_k^{(n)}$ is an adjustable parameter introduced for numerical implementation of Eq.(8.8) and is chosen to be a monotonic sigmoidal function of the iterative index n of the form of Eq.(7.22), which gradually modifies the relative importance of the a priori source information function $P(\Phi)$. The function $Z_k^{(n)}(\phi_k^{(n)}+\tau d_k^{(n)})$ is, again specified by Eq.(8.3), or

$$Z_k^{(n)}(\phi_k^{(n)}+\tau d_k^{(n)})=\sum_q\sum_l W(l)\,[\prod_s C_s\exp(-\frac{(\phi_{k+L_w}^{(n)}+\tau d_{k+L_w}^{(n)}-\phi_s^e)^2}{2\rho_s^2})]$$
$$+C_b\exp[-\frac{((\phi_k^{(n)}+\tau d_k^{(n)}-\phi_b^e))^2}{2\rho_b^2}] \tag{8.10}$$

and similarly for $L_k^{(n)}(\phi_k^{(n)}+\tau d_k^{(n)})$ and $M_k^{(n)}(\phi_k^{(n)}+\tau d_k^{(n)})$.

For each data element obeying gaussian statistics and all of them are statistically independent, substituting for $P(\mathbf{Y}\mid\Phi)$ of Eq.(3.49) and $P(\Phi)$ of Eq.(4.34), Eq.(8.2) becomes:

$$\sum_i R_{ik}\frac{(Y_i-\sum_j R_{ij}\phi_j^*)}{\sigma_i^2}=\frac{L_k(\phi_k^*)+M_k(\phi_k^*)}{Z_k(\phi_k^*)} \tag{8.11}$$

where σ_i^2 may be assumed as Y_i and Eq.(8.8) becomes:

$$\phi_k^{(n+1)}=\frac{\sum_i R_{ik}-\mu_k^{(n)}[(L_k^{(n)}(...)+M_k^{(n)}(...))/Z_k^{(n)}(...)]}{\sum_i R_{ik}{}^2/X_{ik}^{(n)}}. \tag{8.12}$$

where $L_k^{(n)}(...)$, $M_k^{(n)}(...)$ and $Z_k^{(n)}(...)$ are defined by Eq.(8.10)ff and $\mu_k^{(n)}$ serves the same purpose as $\xi_k^{(n)}$.

2). Simulation Results

The actual one dimensional source distribution $\{O_j\}$, as shown in Fig.(8.1) by the solid line, consists of two point sources of 50 unit strengths and of 8 voxels separation ($L=8$), superimposed on a uniform background of 10 unit strengths ($O_{11}=O_{19}=50+10=60$). The one dimensional noise-free or ideal data denoted by the dotted line in Fig.(8.1) was calculated from $\sum_j R_{ij}O_j$. The actual source distribution ranges from voxel-position 3 to 27 and the ideal data distribution extends two voxel units beyond the graph on both sides (i.e. five voxel units beyond the actual source distribution). The R_{ij} was chosen as:

$$R_{ij}=0.5\,e^{-(\ln 2)\left(\frac{i-j}{T}\right)^2},\quad i=-2,-1,0,1,\ldots,32;\quad j=3,4,5,6,\ldots,27;\quad T=4 \tag{8.13}$$

defining a matrix of 35×25 elements.

A single set of poisson randomized data (denoted by stars in Fig(8.1)) was obtained from the ideal data using a standard poisson random number generator (Carnahan et al 1978) as before.

For the two point source pattern considered, the fuzzy pattern function (4.33) applies. The terms defined by Eqs.(8.10)ff in the denominator of the BIP algorithm Eq.(8.8) have the forms:

$$Z_k^{(n)}(\phi_k^{(n)})=\eta_k^{(n)}e^{-U_s(\phi_k^{(n)})}+\Big[\sum_{l\,\varepsilon\, l_1\pm\Delta}W(l)e^{-U_s(\phi_k^{(n)})}+\sum_{l\,\varepsilon\, l_1\pm\Delta}W(l)e^{-V_s(\phi_k^{(n)})}\Big]$$

$$L_k^{(n)}(\phi_k^{(n)})=\eta_k^{(n)}\frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_b^e)}{\rho_b^2}e^{-U_s(\phi_k^{(n)})} \tag{8.14}$$

and

$$M_k^{(n)}(\phi_k^{(n)})=\frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_1^e)}{\rho_1^2}\sum_{l\,\varepsilon\, l_1\pm\Delta}W(l)e^{-V_s(\phi_k^{(n)})}$$
$$+\frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_2^e)}{\rho_2^2}\sum_{l\,\varepsilon\, l_1\pm\Delta}W(l)e^{-U_s(\phi_k^{(n)})}$$

where

$$V_s(\phi_k^{(n)})=\frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_1^e)^2}{2\rho_1^2}+\frac{(\phi_{k+l}^{(n)}+\tau d_{k+l}^{(n)}-\phi_2^e)^2}{2\rho_2^2}$$

$$U_b(\phi_k^{(n)})=\frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_b^e)^2}{2\rho_b^2} \tag{8.15}$$

and

$$U_s(\phi_k^{(n)})=\frac{(\phi_{k-l}^{(n)}+\tau d_{k-l}^{(n)}-\phi_1^e)^2}{2\rho_1^2}+\frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_2^e)^2}{2\rho_2^2}\,.$$

The ϕ_1^e and ϕ_2^e in Eq.(8.14) are the estimated strengths of the assumed two spots respectively and $l\ \varepsilon\, l_1\pm\Delta$ corresponds to L_{qs} defined as in the function (4.33). The adjustable parameter $\eta_k^{(n)}$ is a function

of C_s and C_b and was chosen to have the same functional form as $\xi_k^{(n)}$ of Eq.(7.22), i.e.,

$$\eta_k^{(n)} = \frac{a_0 n^\kappa}{b_0 + n^\kappa} \tag{8.16}$$

The a priori estimates ϕ_s^e and the variance parameter ρ_s are also represented as functions of the iterative index n:

$$\phi_s^e \rightarrow \phi_b^e + \frac{\phi_s^e - \phi_b^e}{\sqrt{L}}\sqrt{n} \quad , \quad \rho_s^2 = \phi_s^e/\sqrt{n} \quad , \quad s=1,2 \tag{8.17}$$

for numerical implementation of Eq.(8.8), where L is the total number of iterations.

The spatial weighting function $W(l)$ was assumed to be of the form:

$$W(l) = \frac{\Gamma^2}{\Gamma^2 + (l - l_1)^2} \tag{8.18}$$

with $l \; \varepsilon \, l_1 \pm \Delta$ as employed in Eqs.(8.14)ff.

In implementing the BIP algorithm expressed by Eqs.(8.8), (8.9) and (8.10) the numerical constants in Eqs.(7.22), (8.16) and (8.18) were chosen as:

$$a_0 \simeq 1, \; a \simeq 0.1, \; b_0 \simeq b \simeq 100, \; \Gamma = 1, \; \kappa = 2$$

The initial values $\{\phi_k^{(0)}\}$ were chosen using Eq.(6.37).

Fig.(8.2) shows the results of the MLEM algorithm (6.19) (i.e., setting $\xi_k^{(n)} = 0$ in Eq.(8.8)) after 100 iterations and of the BIP algorithm (8.8) after only 50 iterations for ideal data.++ In using the BIP algorithm, it was assumed that $\phi_1^e = 55$, $\phi_2^e = 65$ units respectively rather than the actual values of 60 units each. The assumed separation l_1 of the two spots was estimated as 7 voxels rather than the actual separation of 8 voxels. The weighting function $W(l)$ assumes a range of $\Delta = 2$, i.e., $l = 5, 6, 7, 8, 9$, and for this set of calculations $W(l) = 0.2, 0.5, 1, 0.5, 0.2$. Note that in spite of the imperfect a priori assumption of a separation of 7 voxels and the inexact amplitudes 55 and 65 units, the BIP maximum likelihood solution matches the actual source distribution very well. For an assumed separations of $l_1 = 6$ or 10, the BIP solutions remain excellent although the peaks were slightly broader. However, for assumed separations of 5 or 11 corresponding for $\Delta = 2$ to the a priori probabilities $W(l = 5,11) = 0$, the BIP solutions were only slightly superior to the MLEM result.

++ Note that the presence of the a priori probability constraint $P(\Phi)$ tends to suppress the iterative divergences common to most other methods. The results after 50 iterations using the BIP algorithm were generally indistinguishable from those after 100 iterations.

The results of both algorithms MLEM (6.19) and BIP (8.8) applied to the single set of poisson randomized data of Fig.(8.1) after 100 and 50 iterations respectively are shown in Fig.(8.3), where for the BIP algorithm $\phi_1^s=55$, $\phi_2^s=65$ and $l_1=7$ were assumed, and $l \varepsilon l_1 \pm \Delta = 7 \pm 2$ voxels (i.e., the same a priori information as that used in Fig.(8.2). Since the resulting maxima (one set, e.g., as shown in Fig.(8.3) fall in and/or shift around the positions (11,19) of the actual two point sources for different sets of random data, the algorithms were then applied to ten sets of random data respectively and the averages of corresponding results are shown in Fig.(8.4).

It is easily seen that the BIP algorithm (8.8) resulted in substantial improvement over the MLEM algorithm (6.19).

For assumed values of $l_1=6$ or 10 and $\Delta=2$, the BIP solutions remain superior to the MLEM result. For $l_1=5$ or 11, the BIP results on average were comparable to that of the MLEM.

Almost identical comparative solutions were obtained using algorithm (8.12) on ideal data and gaussian randomized data.

Similarly improved results were obtained using the gaussian data BIP algorithm of Eq.(8.12).

In judging the "goodness" of a processed image, one can use objective or subjective criteria.

An objective criterion is provided by a mathematically defined quantitative measure.

An subjective criterion is provided by a visual comparison of the processed image and the test phantom.

To quantitate the deviation of the resulting source distribution $\{\phi_j^{(n)}\}$ from the actual source distribution $\{O_j\}$, an objective test function was introduced,

$$\delta=\sum_j^J(\bar{\phi}_j^{(n)}-\bar{O}_j)^2/\bar{O}_j \tag{8.19}$$

with weighted averages of $\phi_j^{(n)}$ and O_j defined:

$$\bar{O}_j=\sum_r R_{rj}O_r/\sum_r R_{rj} \quad , \quad \bar{\phi}_j^{(n)}=\sum_r R_{rj}\phi_r^{(n)}/\sum_r R_{rj} \tag{8.20}$$

where $r \varepsilon j \pm \Lambda$, Λ stands for the average range.

Function (8.19) gives a more accurate indication than the standard test function (Wine 1964):

$$\delta_0=\sum_j^J(\phi_j^{(n)}-O_j)^2/O_j \tag{8.21}$$

For example, $\delta=11.60$ and 31.14 respectively for the results of the BIP and MLEM algorithms shown in Fig.(8.3), in which $\Lambda=2$ was chosen. Using function (8.21) $\delta_0=211.4$ and 170.2 respectively

actually reversing the subjectively obvious order of agreement.

The inappropriateness of the test function δ_0 for this class of applications arises from its relative inflexibility in evaluating spatial positioning errors. A single voxel shift in the calculated position of a point source compared to its true location results in a very large change in value of δ_0 and implies a grossly incorrect result. By reflecting the extent of spatial shifts quantitatively, the test function δ somewhat more realistically assesses the relative accuracy of image processing results.

The amplitude uncertainties in the a priori fuzzy pattern probability distribution function (4.34), $P(\Phi)$, are reflected by the ρ_s parameters, the spatial uncertainties by the $W(l)$ weighting functions. Numerous calculations using a range of ρ_s values and $W(l)$ functions indicate that:

a). Very accurate results are obtained with BIP when the actual source distribution lies within the central range of the multidimensional $P(\Phi)$ distribution even for extremely noisy data. The results are generally significantly better than those obtained using the MLEM algorithm.

b). If the source distribution lies completely outside the range of the a priori information function $P(\Phi)$, the BIP results are no better and often can be worse than the MLEM results.

The deviation of the resulted source distributions $\{\phi_j^{(50)}\}$ using the BIP algorithm on ideal data after 50 iterations from the actual source distribution $\{O_j\}$ as a function of amplitude variation $\delta\phi = |\phi_s^e \pm 60|, s=1,2$; or spatial variation $\delta l = |l_1 \pm 8|$ of the fuzzy pattern elements respectively were investigated as following, in which function $W(l)$ was first ignored (i.e.,the function (4.32) was used).

When the estimated $l_1=8$ matches the actual separation of the two actual point sources and the estimated strengths of the two assumed spots (plus background) vary in forms of $\phi_1^e=60-\delta\phi$ and $\phi_2^e=60+\delta\phi$, the resulted values $\phi_{11}^{(50)}$ and $\phi_{19}^{(50)}$ are between ϕ_1^e and ϕ_2^e, and $\phi_{11}^{(50)}$ tends to ϕ_1^e and $\phi_{19}^{(50)}$ to ϕ_2^e.

When the estimated $\phi_1^e=\phi_2^e=60$ match the actual strengths of the two point sources and the estimated spatial separation of the assumed spots varies in forms of $l_1=8\pm\delta l$, the resulted values $\phi_{11}^{(50)}$ and $\phi_{19}^{(50)}$ vary in such a way that when the separation of ϕ_1^e and ϕ_2^e matches 8, then $\phi_{11}^{(50)}$ and $\phi_{19}^{(50)}$ almost match value of 60, otherwise they go down from value of 60.

Fig.(8.5) shows the deviations of the resulted $\{\phi_j^{(50)}\}$ from $\{O_j\}$ as a function of $\delta\phi$ and δl respectively using the function (8.19) for $r \varepsilon j \pm 1$. The stars in Fig.(8.5) represents that case in which $l_1=8$ (i.e.,$\delta l=0$) and the estimated amplitudes vary as $\phi_1^e=60-\delta\phi$ and $\phi_2^e=60+\delta\phi$. In the case indicated by symbol triangles, $\phi_1^e=\phi_2^e=60$ (i.e.,$\delta\phi=0$) and the estimated spatial relation varies as $l_1=8\pm\delta l$. The triangles show the average of two results of $l_1=8-\delta l$ and $l_1=8+\delta l$. The lines in Fig.(8.5) reflect the smoothed results of the discrete symbols of star and triangle respectively.

In order to emphasis the effect of the spatial uncertainty function $W(l)$, $\phi_1^e=\phi_2^e=60$ and $l \varepsilon l_1 \pm \Delta = l_1 \pm 8$ were used. Excellent results were obtained for $l_1=2, 3, 4, \ldots, 12, 13, 14$ voxel units

respectively. If weighting range l of $W(l)$ doesn't cover the actual separation of 8 voxels, the results (indicated by triangles in Fig.(8.5)) are obtained.

So far, the BIP algorithms derived in the previous chapters have been tested only on computer simulated ideal data, and poisson and gaussian randomized data for one dimensional cases. In the next chapter, these algorithms are applied to imaging data from experimental phantoms.

Referred to the statements in chapter II (page 8), there are replies as follows:

a. the BIP formalism searches for that source distribution Φ^* which maximizes the conditional probability function $P(\Phi|\mathbf{Y})$ for given a set of measured data $\mathbf{Y}$;

b. since the statistical fluctuation of measured data has been considered accurately in the function $P(\mathbf{Y}|\Phi)$, the solution Φ^* is stable if "artificial" errors in measured data are not dominant;

c. the BIP algorithms, derived using the EM technique, have been tested and are convergent;

d. the acceleration of convergence, using relaxation parameters, has been mentioned in chapters III (Eq.(3.17)) and V (Eq.(5.33)). More discussions will be given in the next chapter.

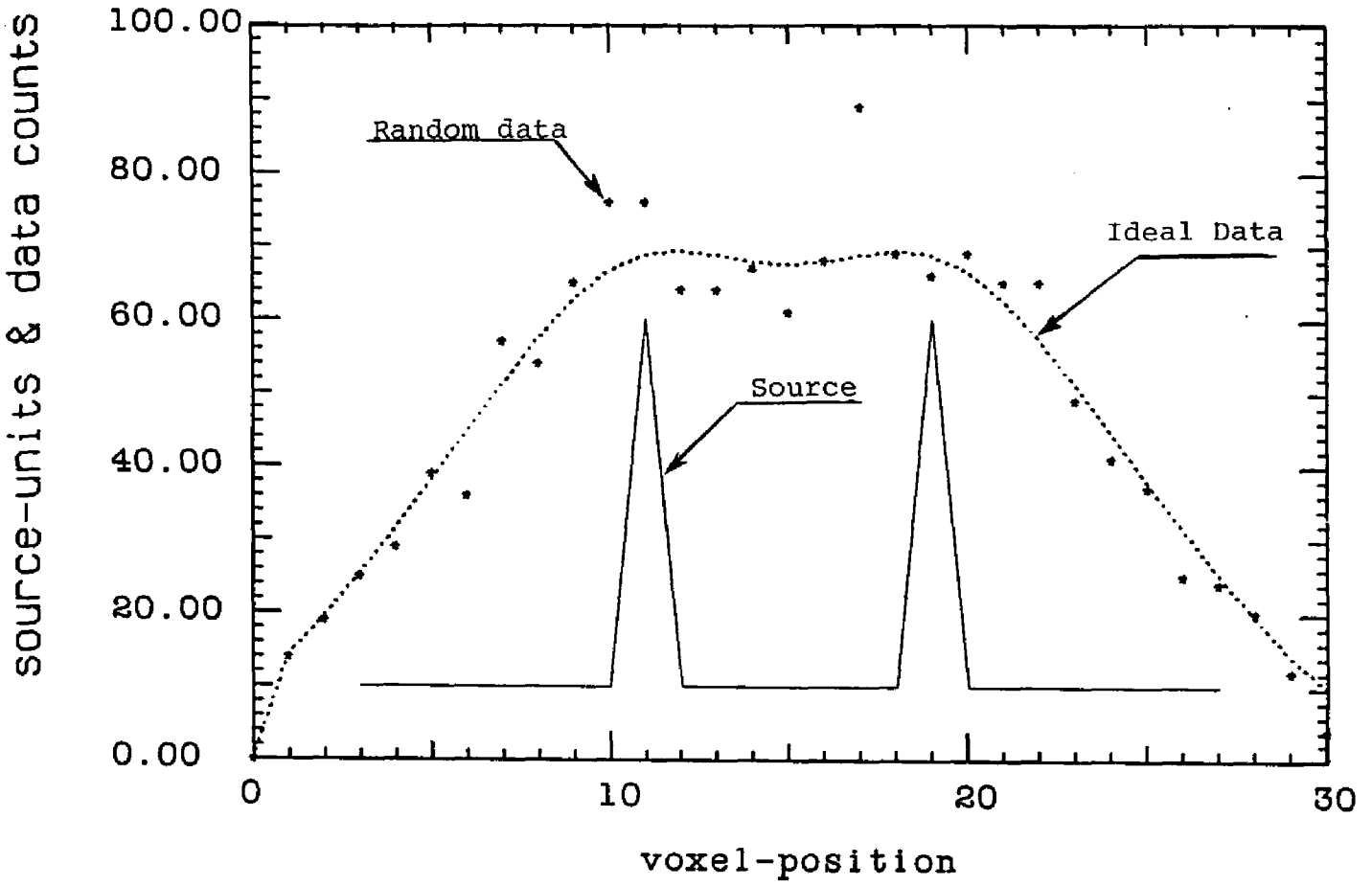


Fig.8.1: One dimensional source and data distributions. The solid line represents actual source distribution. The dotted line represents ideal data distribution. The stars represent computer poisson randomized data.

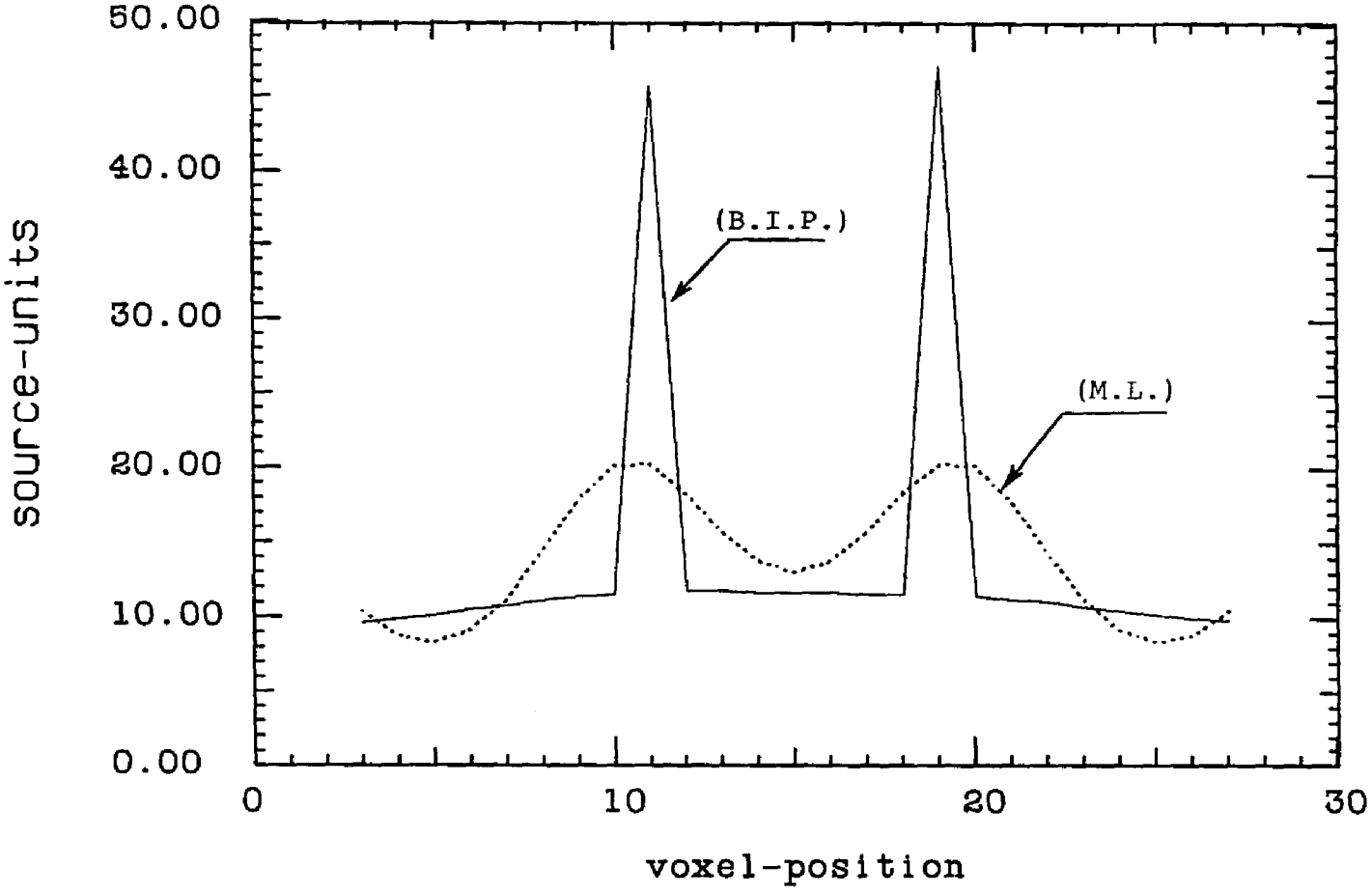


Fig.8.2: Results using algorithm (B.I.P.) at 50 iterations and algorithm (M.L.) at 100 iterations on ideal data.

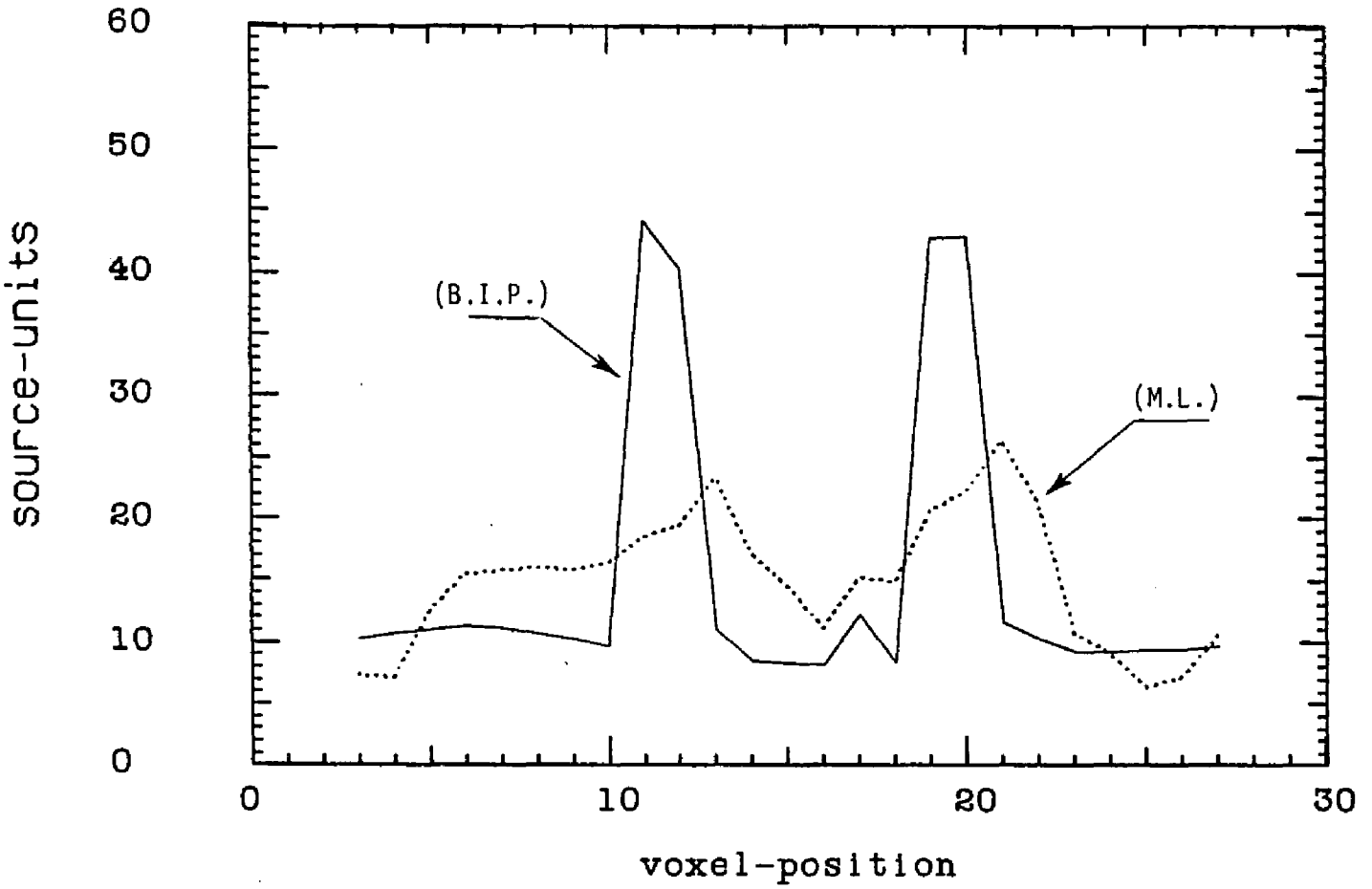


Fig. 8.3: Results using BIP algorithm (B.I.P.) at 50 iterations and ML algorithm (M.L.) at 100 iterations on a single set of computer poisson randomized data.

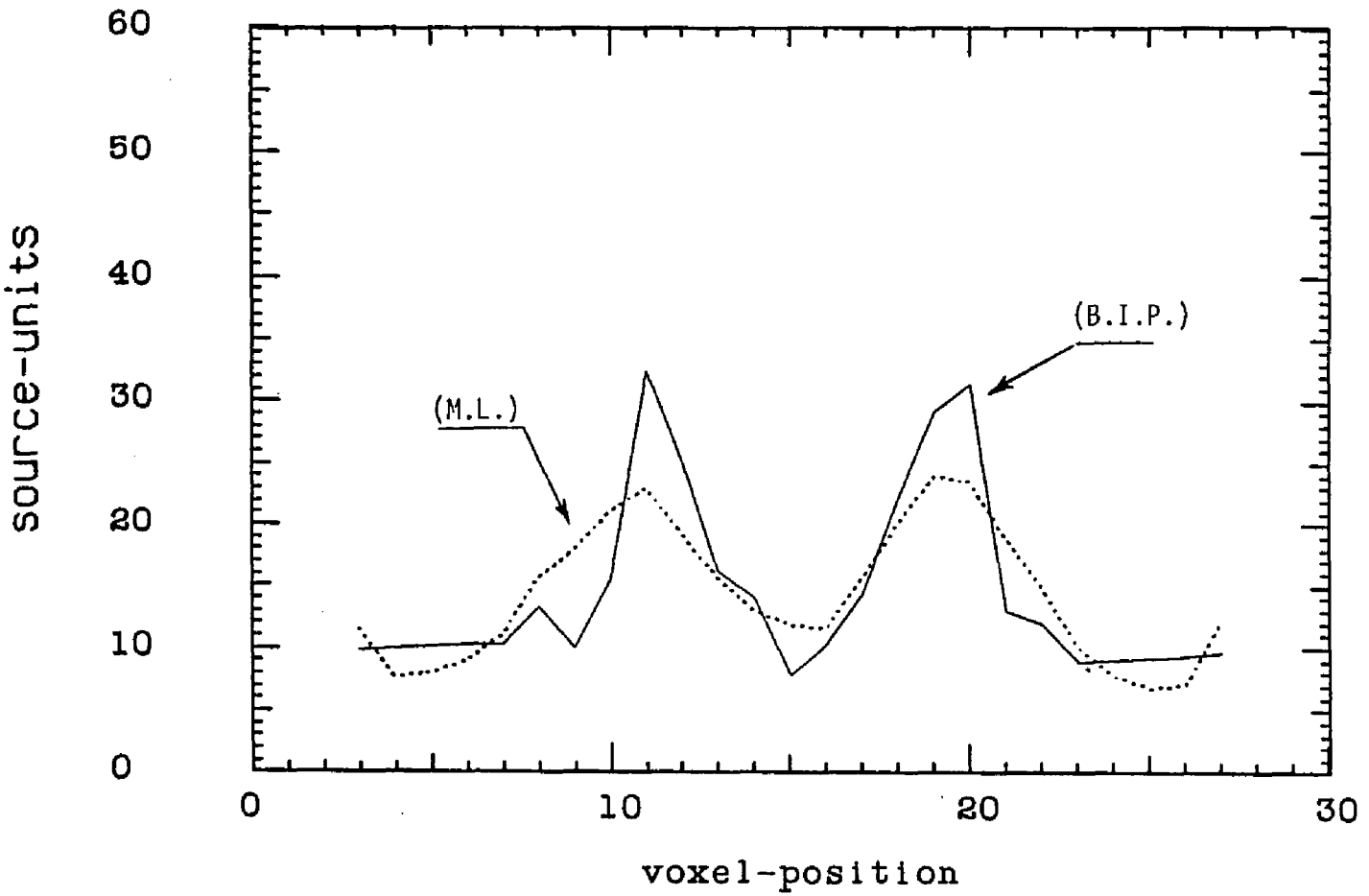


Fig.8.4: Comparison of BIP algorithm (B.I.P.) at 50 iterations and ML algorithm (M.L.) at 100 iterations for ten sets of poisson randomized data.

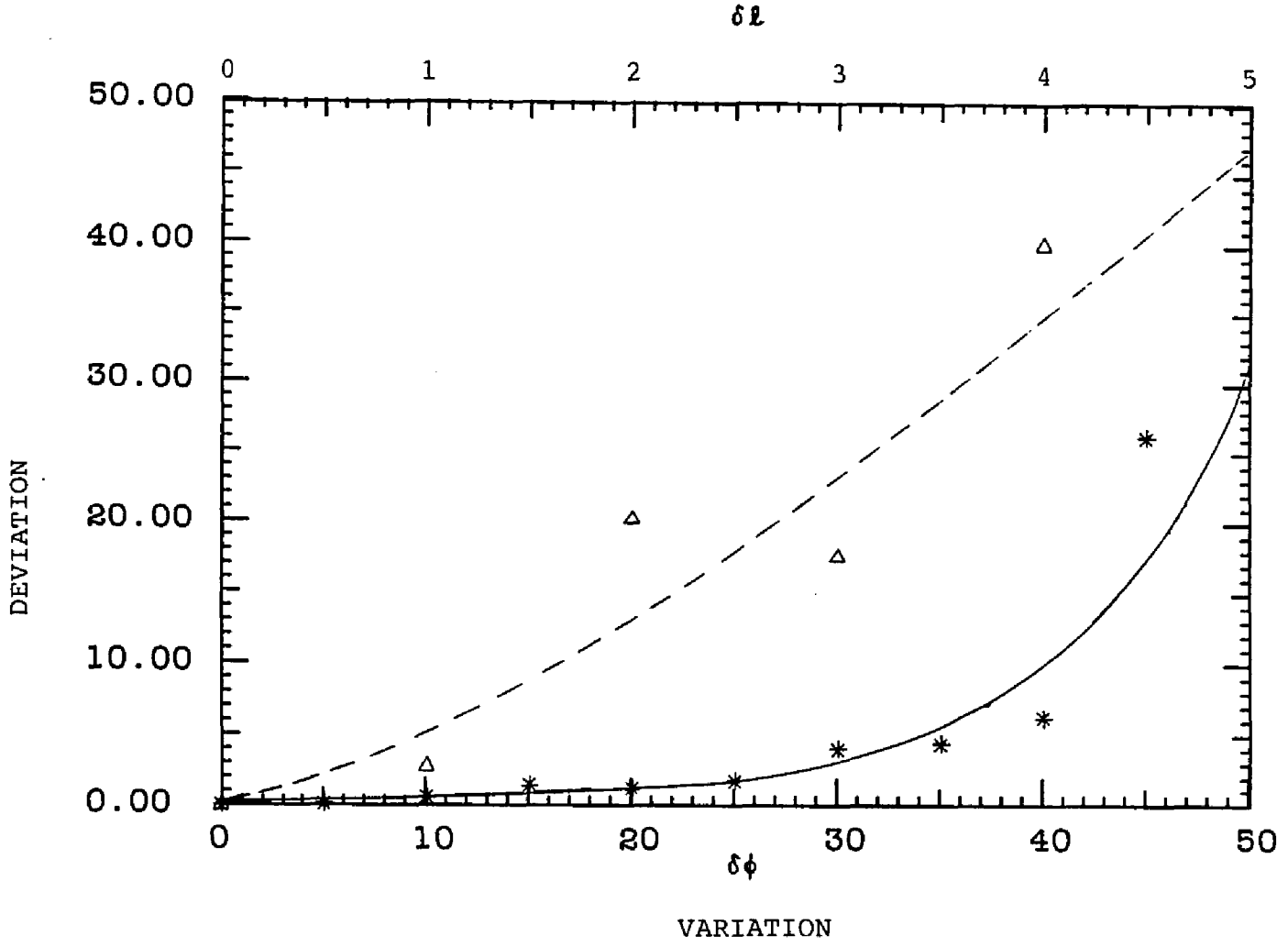


Fig 8.5: Deviations of resulted source distributions from actual source distribution for variations of amplitudes and separations of elements of the fuzzy pattern. The stars represent the deviation as the amplitudes varying. The triangles represent the deviation as the spatial relation varying. The lines reflect the smoothed results of the discrete symbols.

CHAPTER IX
EXPERIMENTAL RESULTS OF PHANTOM IMAGING DATA

1). One Dimensional Results

A point source of Co^{57} was imaged using a Picker Dyna Camera Model No.4 and the output was, for simplicity, arranged as a 32×32 matrix (i.e., a two dimensional point source response function (*PSF*), F_{ij}), from which a line source response function (*LSF*), L_i, could be determined:

$$L_i = \sum_j F_{ij} \tag{9.1}$$

and is shown in Fig.(9.1) by a solid line in which the center value is normalized to 40.

Since the imaging system was approximately spatially invariant, a matrix representation, R_{ij}^L of the one dimensional *LSF*, could be formed from $\{L_i\}$ (Andrews et al 1977, Appendix A).

A simple one dimensional equivalent phantom was prepared by threading two parallel catheters containing a solution of Co^{57} through stainless steel screen. Two-dimensional data was obtained by imaging the phantom at the same depth as the point source and was arranged as a 32×32 matrix, with the two lines of tubing oriented in the column-direction. Row 16 of the data matrix is shown in Fig.(9.1) by stars and was used as the one-dimensional imaging data $\{Y_i\}$. Neglecting the effect of the finite length of the tubing, $\{Y_i\}$ can be viewed as imaging data from a double element source distribution (the projection of the two line sources along the parallel direction superimposed on a uniform background) in a one dimensional geometry.

Algebraically then:

$$Y_i = \sum_i R_{ij}^L \, O_j \tag{9.2}$$

where $\{Y_i\}$ is row 16 of the data matrix, R_{ij}^L the convolution maxtrix generated by the *LSF* and $\{O_j\}$ the equivalent one dimensional phantom.

The a priori source amplitude values of $\phi_b^e = 3$ and $\phi_1^e = \phi_2^e = 60$ were estimated by analysing experimental data with and without presence of the phantom, making use of the factor $\sum_{ij} R_{ij}^L$, i.e.,

$$\sum_i Y_i^0 = \phi_b^e \sum_{ij} R_{ij}^L \tag{9.3}$$

and

$$Y_i = \sum_j R_{ij}^L \phi_j = \phi_b^e \sum_j R_{ij}^L + R_{i1}^L \, \phi_1^e + R_{i2}^L \, \phi_2^e \tag{9.4}$$

where $\mathbf{Y}^0$ is the data vector without presence of the phantom and, to avoid boundary effect, the index i is

chosen at the center of the data vector **Y** as shown in Fig.(9.1).

The numerical constants in Eqs.(7.22) and (8.16) were chosen as:

$$a \approx 5, a_0 \approx 3, b \approx 500, b_0 \approx 100, \kappa = 2$$

and Eq.(8.17) was considered to run the BIP algorithms.

Fig.(9.2) demonstrates the superior result achieved with the uncorrelated constraint BIP algorithm (7.8) (solid line) by 50 iterations compared to the result obtained using the ML algorithm (mlem) (dotted line) after 100 iterations. Similarly improved results were obtained with the BIP algorithm (7.8) for all of the centrally located data rows.

Fig.(9.3) shows the results obtained using the correlated constraint BIP algorithm (7.16) by 50 iterations, in which the a priori spatial relation was assumed as 9 (solid line) and 10 (broken line) (the actual separation of the two lines of tubing was ≈ 7 measured in the phantom frame) respectively, and the ML algorithm (mlem) (dotted line) by 100 iterations.

Note that the blurred image (or measured data) has a flat distribution range of 11 voxel units (or 9.68 cm) with high counts as shown in Fig.(9.1) and the actual separation of the two parallel lines of tubing is about 6.3 cm (or 7 voxel units).

To use the fuzzy constraint BIP algorithm (8.8) and (8.14), the parameters chosen for the spatial weighting function $W(l)$ of Eq.(8.18) are shown below.

In Fig.(9.4), $\Gamma = 1.5$, $l_1 = 10$; $\Delta = 1$ (solid line) and $\Delta = 2$ (broken line) were chosen respectively to run the BIP algorithm at 50 iterations, i.e., the assumed separation is same, the assumed amplitudes are same, but the weight ranges of $W(l)$ are different for the two curves.

In Fig.(9.5), $\Gamma = 1.5$, $\Delta = 3$; $l_1 = 9$ (solid line) and $l_1 = 10$ were chosen respectively to run the BIP algorithm at 50 iterations, i.e., the assumed amplitudes are same, the weight ranges of $W(l)$ are same, but the assumed separation is different for the two curves.

2). Two Dimensional Results

The two dimensional *PSF* is shown in Fig.(9.6) and the data matrix of the phantom is shown in Fig.(9.7). Since the camera crystal is a circular plate, the measured data are zero outside of the plate area as shown in the corners of figures (9.6) and (9.7).

Since the imaging system was approximately spatially invariant, the *PSF* as shown in Fig.(9.6) is almost identical for a point source at any position in the source field. Then the matrix expression of the convolution of the *PSF* and the phantom is:

$$\mathbf{Y} = R\,\mathbf{O} \tag{9.5}$$

where **Y** and **O** are vectors formed from the two dimensional data matrix and phantom respectively, and R is a matrix formed from the two dimensional spatially invariant *PSF* (Andrews et al, Appendix A).

Fig.(9.8) shows the result using the ML algorithm (mlem) at 100 iterations for the phantom imaging data as shown in Fig.(9.7).

The result obtained using the uncorrelated constraint BIP algorithm (7.8) at 50 iterations for the phantom imaging data is shown in Fig.(9.9) and the result of the fuzzy constraint BIP algorithm (8.8) is shown in Fig.(9.10), in which the a priori source amplitude values were estimated following the same analysis of using Eqs.(9.3) and (9.4): $\phi_b^e = 8$ and $\phi_1^e = \phi_2^e = 110$; the numerical constants defined in Eqs.(7.22), (8.16) and (8.18) were: $a = a_0 \simeq 0.08$, $b = b_0 \simeq 0.3$, $\kappa = 2$, $\Gamma = 1.5$, $l_1 = 10$, $\Delta = 3$; the iterative form of ϕ_s^e was given by Eq.(8.17) and $\rho_s^2 = \phi_s^e(6 + 0.4(n-1))/n$; and the orientation of the two parallel catheters was assumed known when using algorithm (8.8).

Although boundary artifacts appear in Figs.(9.8), (9.9) and (9.10), the BIP algorithms exhibit striking improvement in the imaging processing of the phantom imaging data commpared to the ML algorithm. Note that the boundary artifacts were greatly reduced when the a priori spatial relation of the two lines of tubing was considered as shown in Fig.(9.10).

In addition, the ML, and the uncorrelated and fuzzy constraint BIP algorithms were tested in two dimensions for ideal data.

The actual two dimensional source distribution consists of 2 point sources of strength 110 units each and separated by 8 voxel elements, superimposed upon a uniform background of strength 1 unit.

The two dimensional *PSF* has the form of

$$F_{ij} = A \exp[-\frac{ln(2)}{T^2}(i^2 + j^2)]\,,\ i,j = \pm 16, \pm 15, \ldots, 0;\ T = 4, \tag{9.6}$$

from which a spatially invariant R is formed (Andrews et al 1977).

Fig.(9.11) shows the result of the ML algorithm at 100 iterations for the ideal data. The result obtained using the uncorrelated constraint BIP algorithm at 50 iterations is shown in Fig.(9.12). The result of the fuzzy constraint BIP algorithm is almost same as that of Fig.(9.12).

3). Consideration of Computation Time

The results as shown in Figs.(9.8), (9.9) & (9.10) were generated using the VAX/780 computer at CCNY on time sharing basis. It took about two hours to generate the result of Fig.(9.8) using the ML algorithm (6.19) at 100 iterations and about one and half hours to get the result of Fig.(9.9) using the BIP algorithm (7.8) at 50 iterations.

There are two convenient ways to reduce the computation time. One is to use the fast F.T. in the "convolution" calculations. Another is to consider a reduced range of measured data for each voxel in the "deconvolution" calculations. In each method the eventual result can also be used to serve as the initial estimate for the slower, more exact final "deconvolution" calculations.

3.1). Fast F.T. for "Convolution" Calculation

The calculated "measured data" $\{Y_i^{(n)}\}$ at the n-th iteration, for all of the iterative algorithms mentioned in chapters VI, VII and VIII, are expressed as:

$$Y_i^{(n)} = \sum_j R_{ij} \phi_j^{(n)} . \tag{9.7}$$

If the matrix $\{R_{ij}\}$ can be expressed as a function of variable $(i-j)$,

$$R_{ij} = f(i-j) \tag{9.8}$$

i.e., the *PSF* is spatially invariant, then F.T. of Eq.(9.7) is:

$$Y_F^{(n)}(k_x, k_y) = R_F(k_x, k_y)\, \Phi_F^{(n)} \tag{9.9}$$

and, $\{Y_i^{(n)}\}$ are given by inverse F.T. of $Y_F^{(n)}(k_x, k_y)$.

The steps to implement the "convolution" of Eq.(9.7) using fast F.T., instead of calculating the product of matrix and vector directly, are:

(i) at very beginning, calculate the fast F.T. of $\{R_{ij}\}$ (Bracewell 1978) ;

(ii) at n-th iteration, calculate the fast F.T. of $\Phi^{(n)}$;

(iii) compute the $Y_F^{(n)} = R_F\, \Phi_F^{(n)}$;

(iv) compute the $\{Y_i^{(n)}\}$ by applying the inverse fast F.T. to $Y_F^{(n)}$.

3.2). Reduced Data Region for "Deconvolution" Calculation

As shown in chapter VI that the ln Bayesian function of Eq.(6.1) is:

$$g(\Phi) = \sum_i^I [(-\sum_j^J R_{ij}\, \phi_j) + Y_i\, ln(\sum_j^J R_{ij}\, \phi_j) - ln(Y_i!)] + \ln P(\Phi) - lnP(Y) \tag{9.10}$$

where I and J are the total numbers of projection rays and voxels respectively. The BIP solution Φ^* is determined by

$$\frac{\partial g(\Phi)}{\partial \phi_k} \Big|_{\Phi=\Phi^*} = 0 . \tag{9.11}$$

Let's consider a trivial case in which $P(\Phi)$ is uniform distributed (or $P(\Phi)$ is constant, Eq.(4.11)) and $I = 1$, i.e., the sampling density function $P_i(Y_i \mid \Phi)$ of each data element Y_i is considered individually not all of them simultaneously, then Eqs.(9.11) becomes:

$$Y_i = \sum_j^J R_{ij}\, \phi_j^* \;, i = 1,2,...,I \tag{9.12}$$

the noise free linear equations (2.42). As discussed in chapter III, the "deconvolution" of ϕ_j from Eq.(9.12) can be ill-posed if the non-singular matrix R has a rather large ($\gg 1$) condition number.

If all data elements $\{Y_i\}$ are considered simultaneously, for uniform $P(\Phi)$, Eqs.(9.10) and (9.11) result in:

$$\sum_i^I R_{ik} = \sum_i^I R_{ik} \frac{Y_i}{\sum_j R_{ij}\phi_j^*} \;, k = 1,2,...,J \tag{9.13}$$

(i.e., the maximum likelihood Eq.(6.2)).

A optimal solution Φ^* can be obtained if only those measured data are considered, to which the voxel k contributs significantly. The corresponding expression of Eq.(9.13) for this optimal solution Φ^* is then:

$$\sum_i^{I_k} R_{ik} = \sum_i^{I_k} R_{ik} \frac{Y_i}{\sum_j R_{ij}\phi_j^*} \;, k = 1,2,...,J \tag{9.14}$$

and the iterative form:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \frac{\sum_i^{I_k} R_{ik}\,(Y_i / \sum_j R_{ij}\phi_j^{(n)})}{\sum_i^{I_k} R_{ik}} \;, k = 1,2,...,J \tag{9.15}$$

where the set I_k covers those measured data to which the voxel k contributs significantly. If only one data element Y_k is within the set I_k, $i = k$, Eq.(9.15) then reduces to the Jacobi result of Eq.(3.3).

In image restoration applications, the set I_k can be chosen such that only those measured data are used to deconvolute the voxel value ϕ_k, to which voxel k contributs with probability amplitudes defined in the range of *FWHM* of *PSF*.

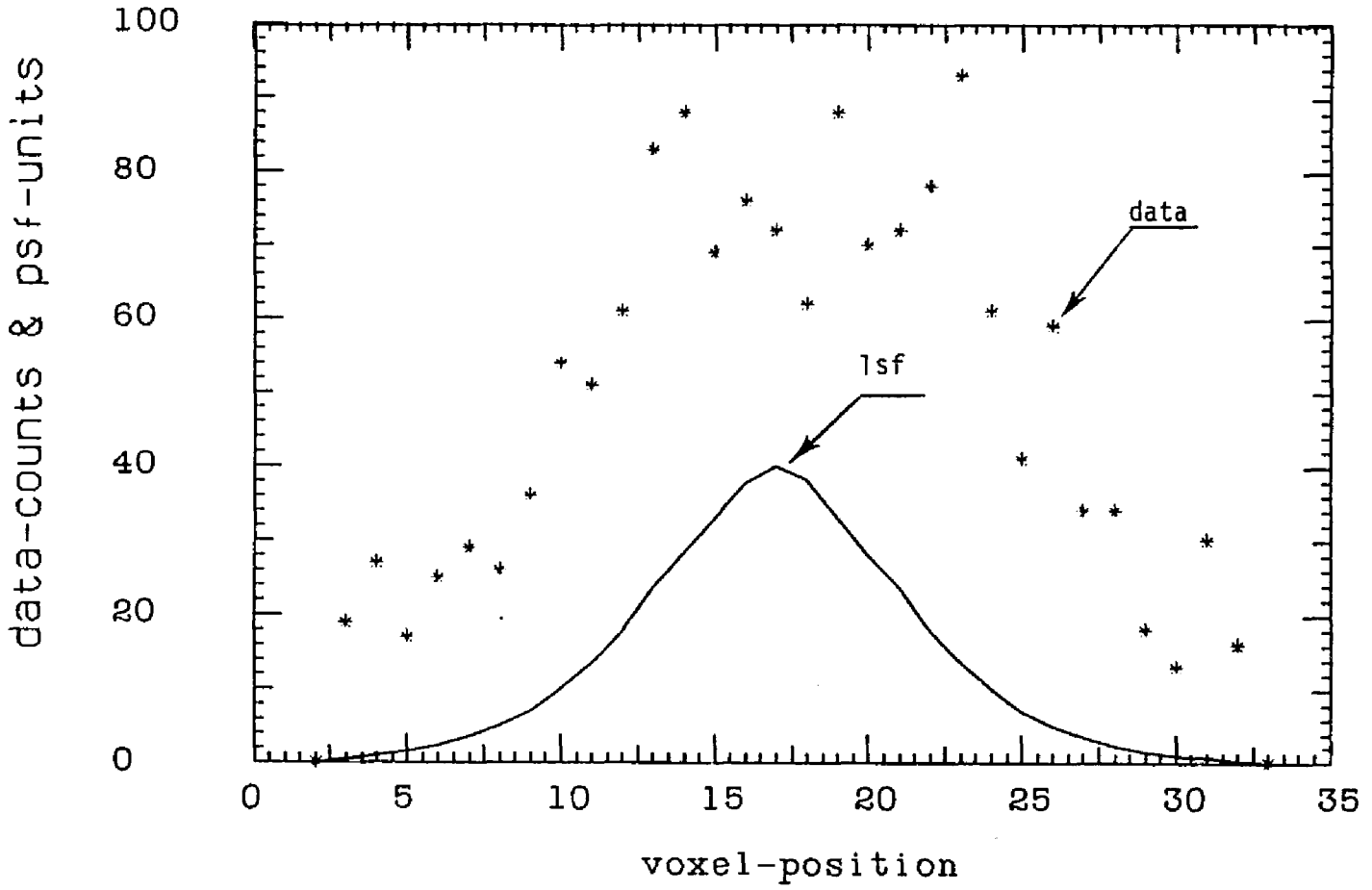


Fig.9.1: The phantom imaging data (stars) from a center column of the two dimensional data matrix and the line source response function for Co^{57} (solid line).

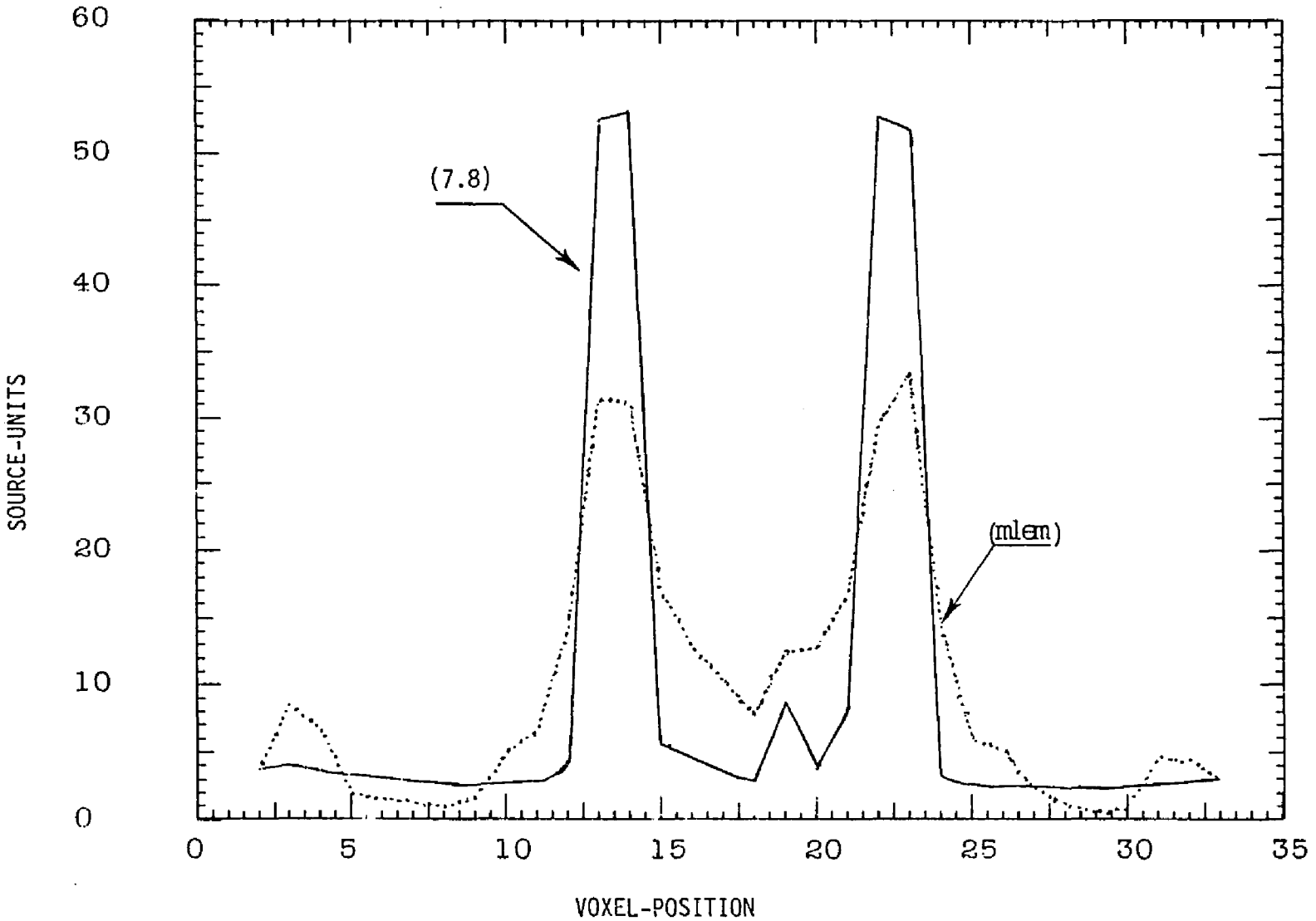


Fig.9.2: Comparison of the uncorrelated constraint BIP algorithm (7.8) at 50 iterations (solid line) and the ML algorithm (mlem) at 100 iterations (dotted line) for the phantom imaging data shown in figure (9.1).

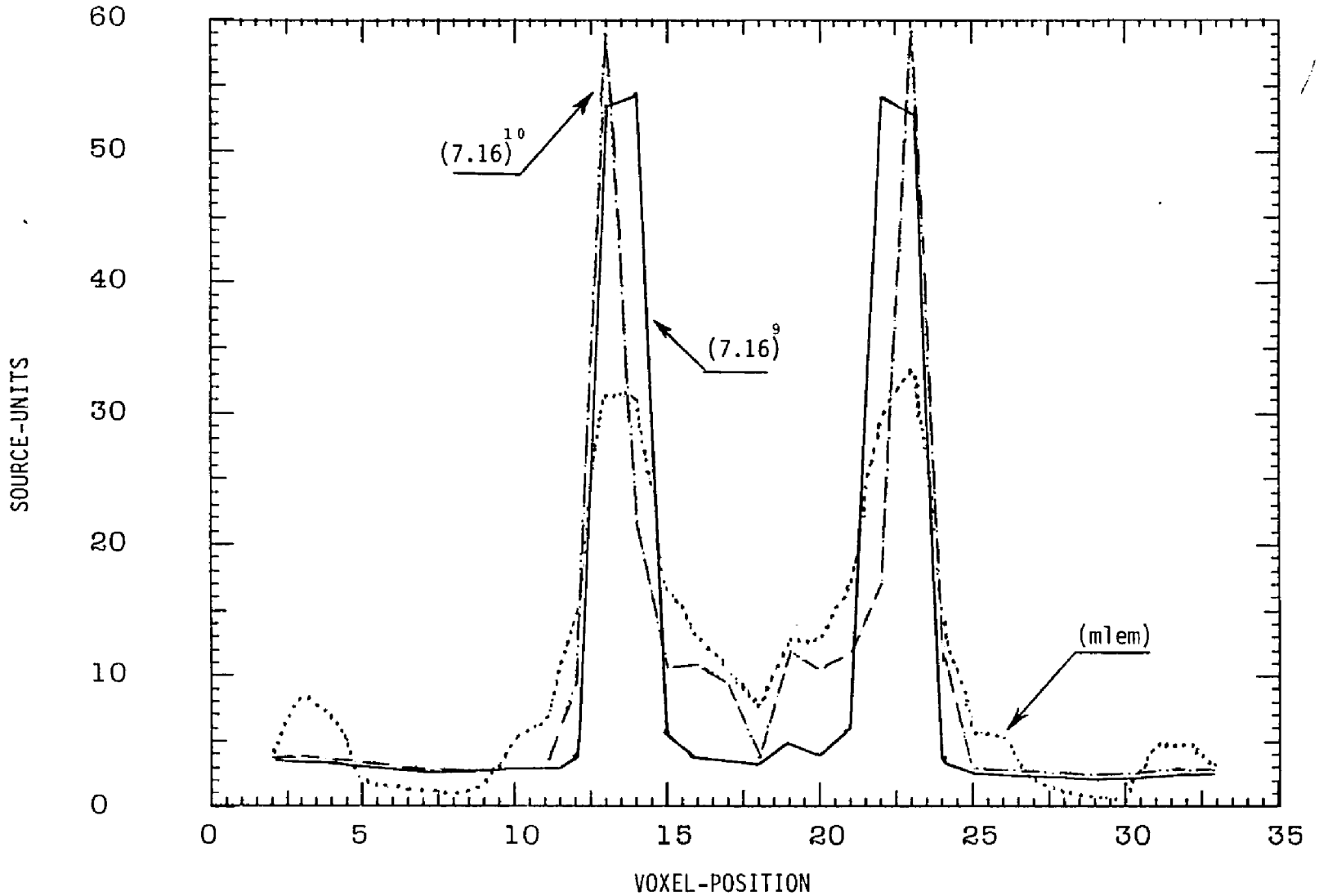


Fig.9.3: Comparison of the ML algorithm (mlem) at 100 iterations (dotted line) and the correlated constraint BIP algorithm (7.16) at 50 iterations for the phantom imaging data; the solid line represents the result with l_1=9 and the broken line with l_1=10.

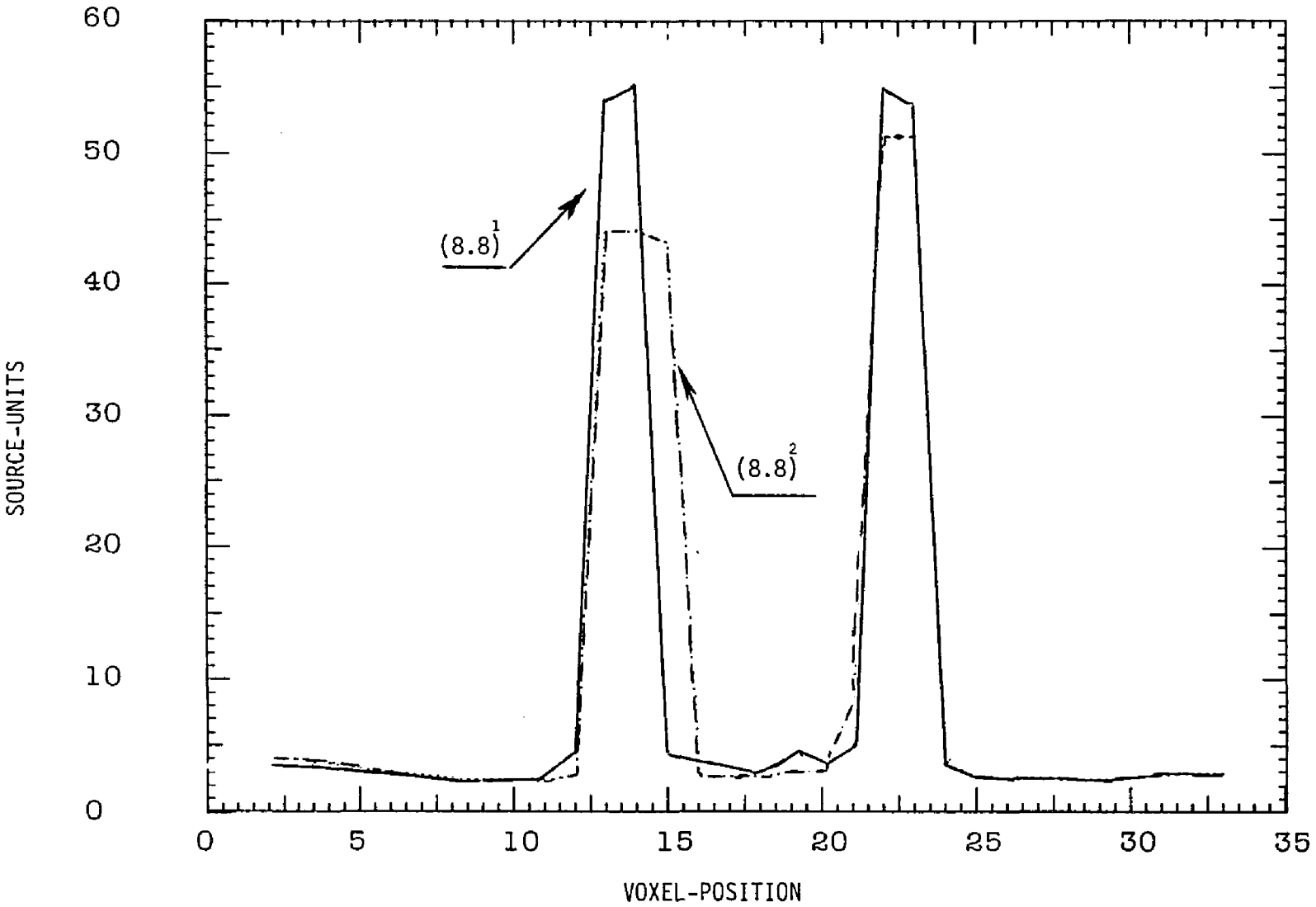


Fig.9.4: Results using the fuzzy source constraint BIP algorithm (8.8) at 50 iterations for the phantom imaging data; the solid line represents the result with l_1=10 and Δ=1, the broken line represents the result with l_1=10 and Δ=2.

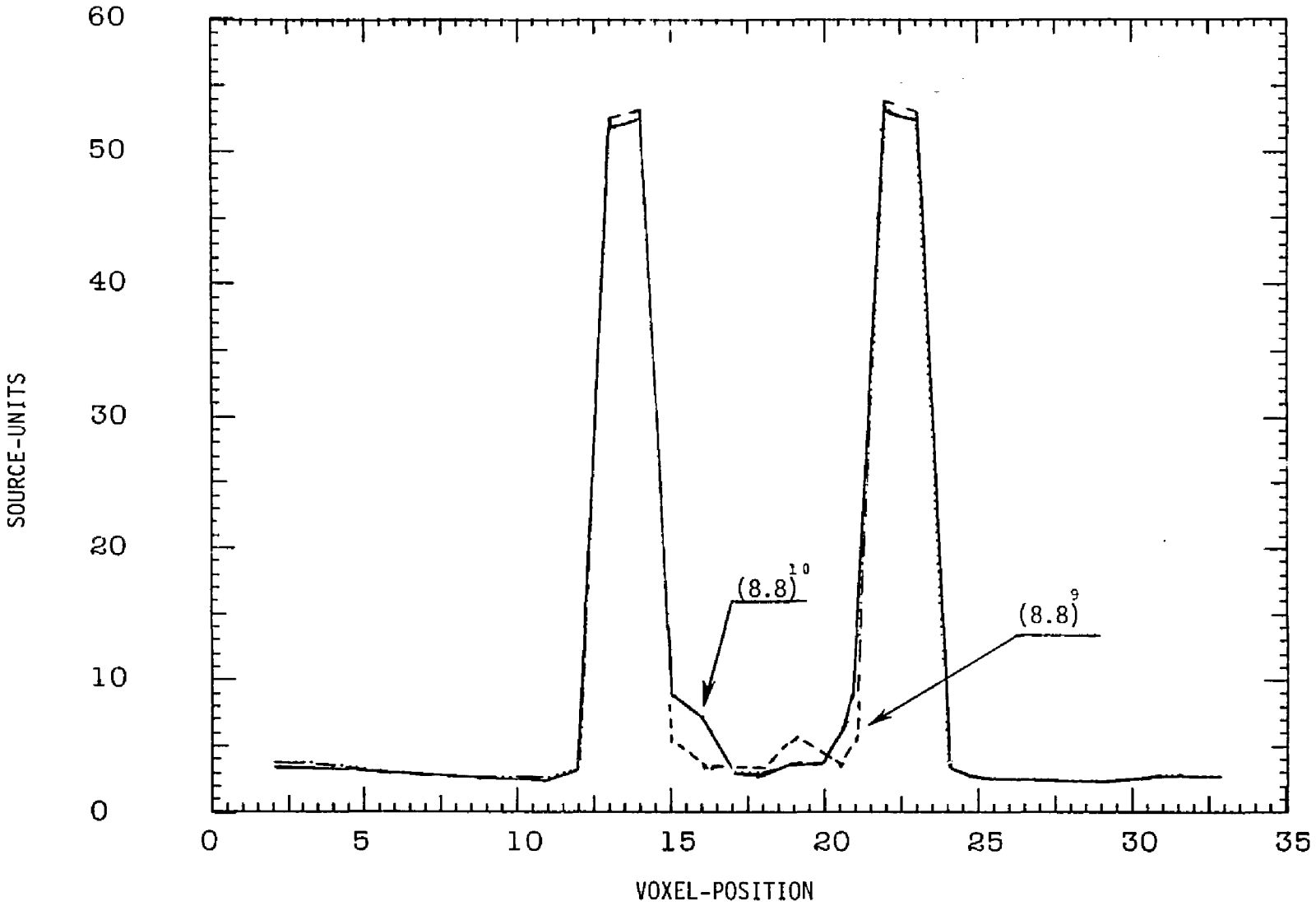


Fig.9.5: Results using the fuzzy source constraint BIP algorithm (8.8) at 50 iterations for the phantom imaging data; the solid line represents the result with l_1=10 and Δ=3, the broken line reprsents the result with l_1=9 and Δ=3.

Fig.9.6: The experimental point source response function for Co^{57}, obtained by using a Picker Dyna Camera Model No.4.

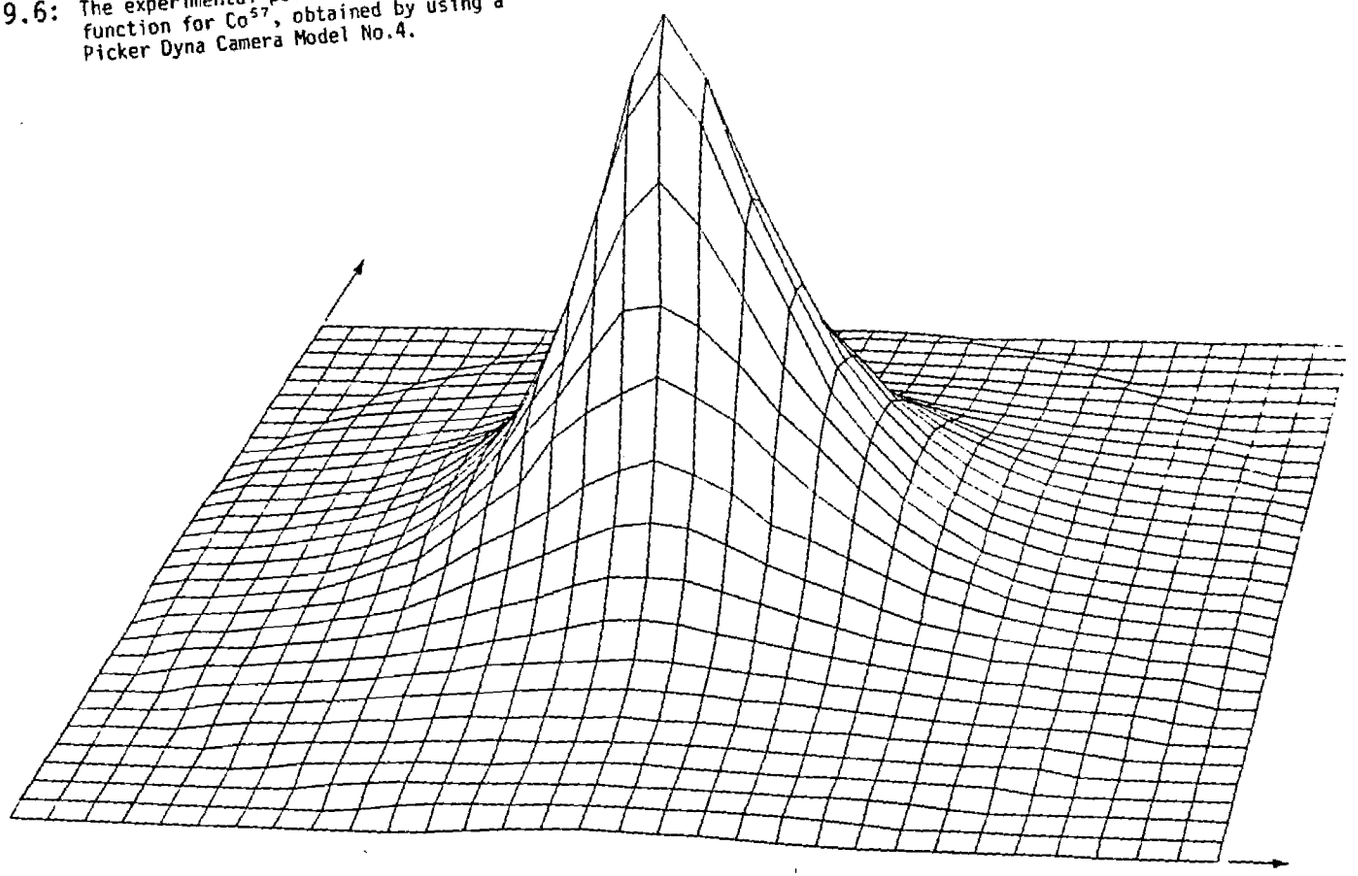


Fig.9.7: Imaging data from a phantom consisting of two parallel lines of tubing, along the collumn direction.

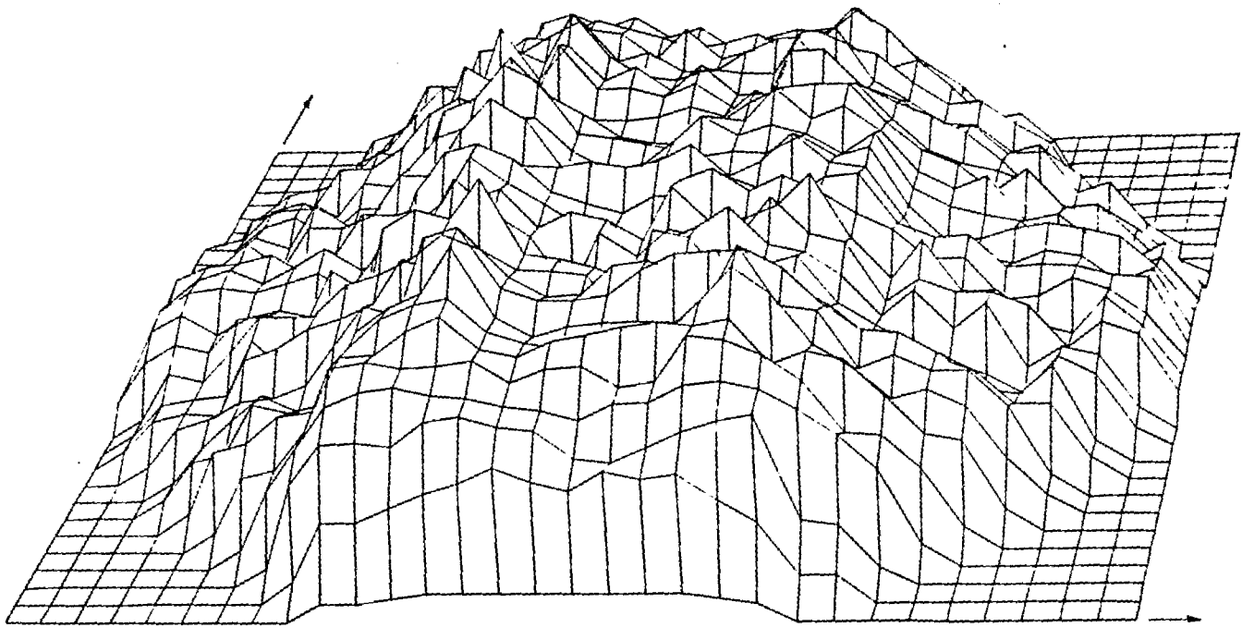


Fig.9.8: Result using ML algorithm for 100 iterations on the phantom imaging data.

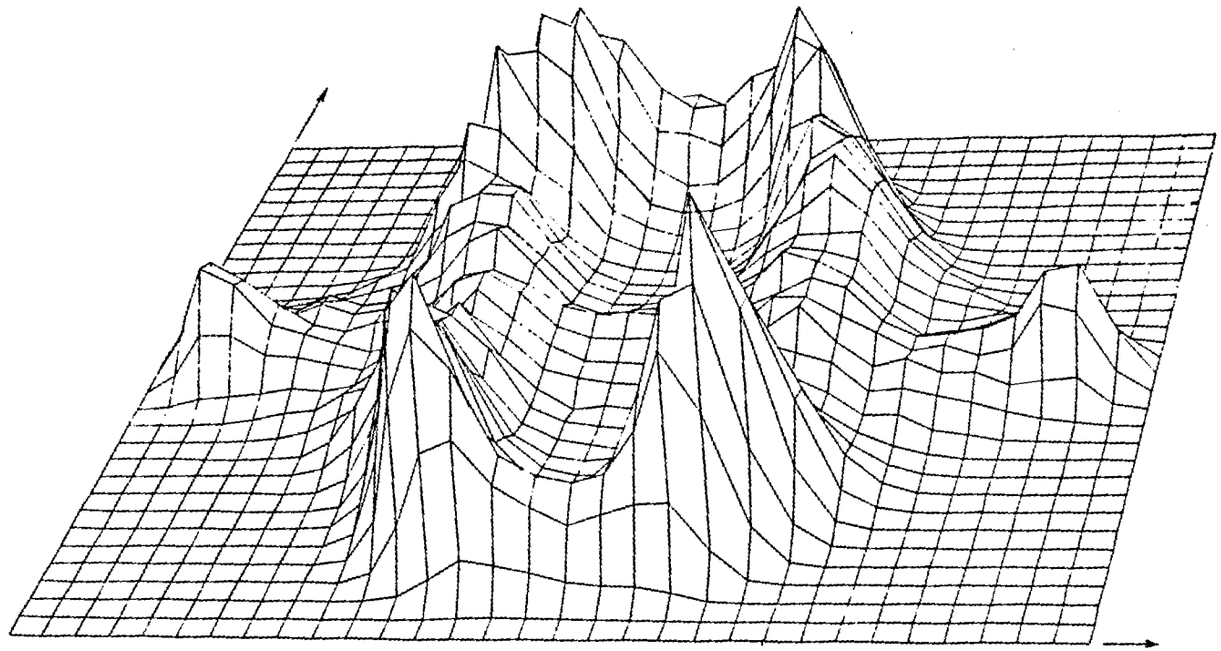


Fig.9.9: Result using the amplitude constraint BIP algorithm for 50 iterations on the phantom imaging data. The amplitudes of the two lines of tubing were assumed as 110 units each and the background of 8 units.

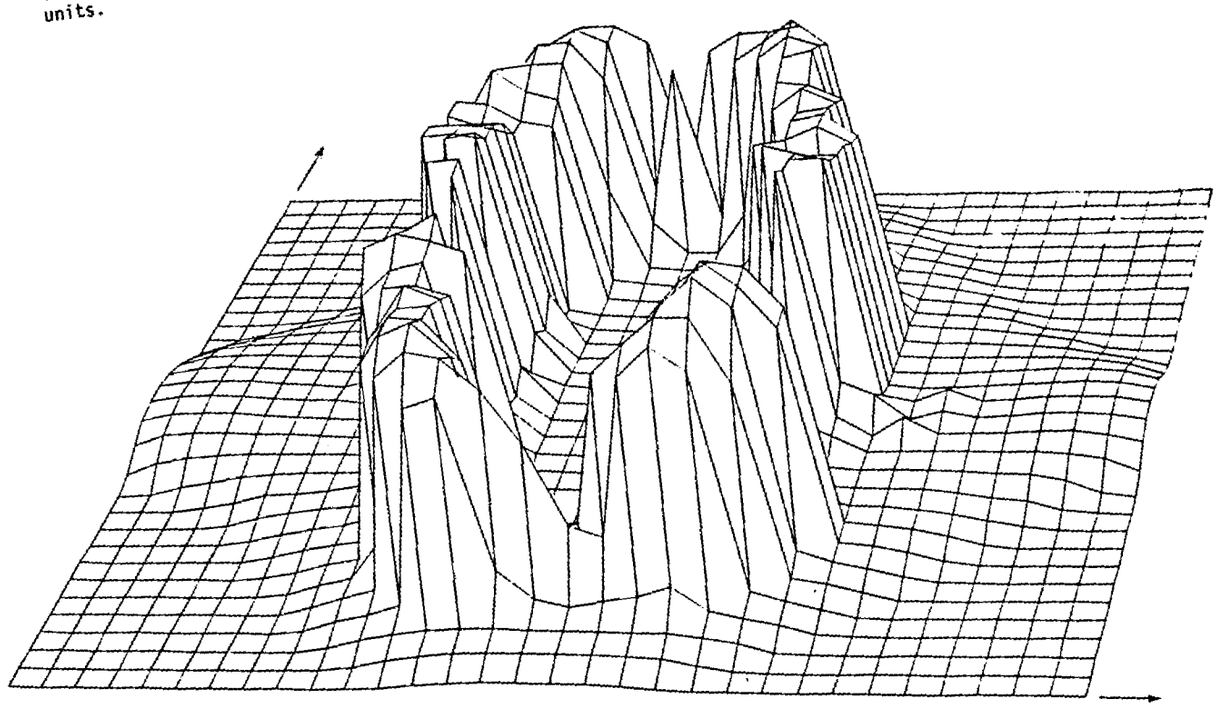


Fig.9.10: Result using the fuzzy pattern source constraint BIP algorithm for 50 iterations on the phantom imaging data. The assumed amplitudes of the two lines of tubing were 110 units each and background of 8 units. The separation of the two line was estimated as 10 voxel elements and the weight range was 7 voxel elements.

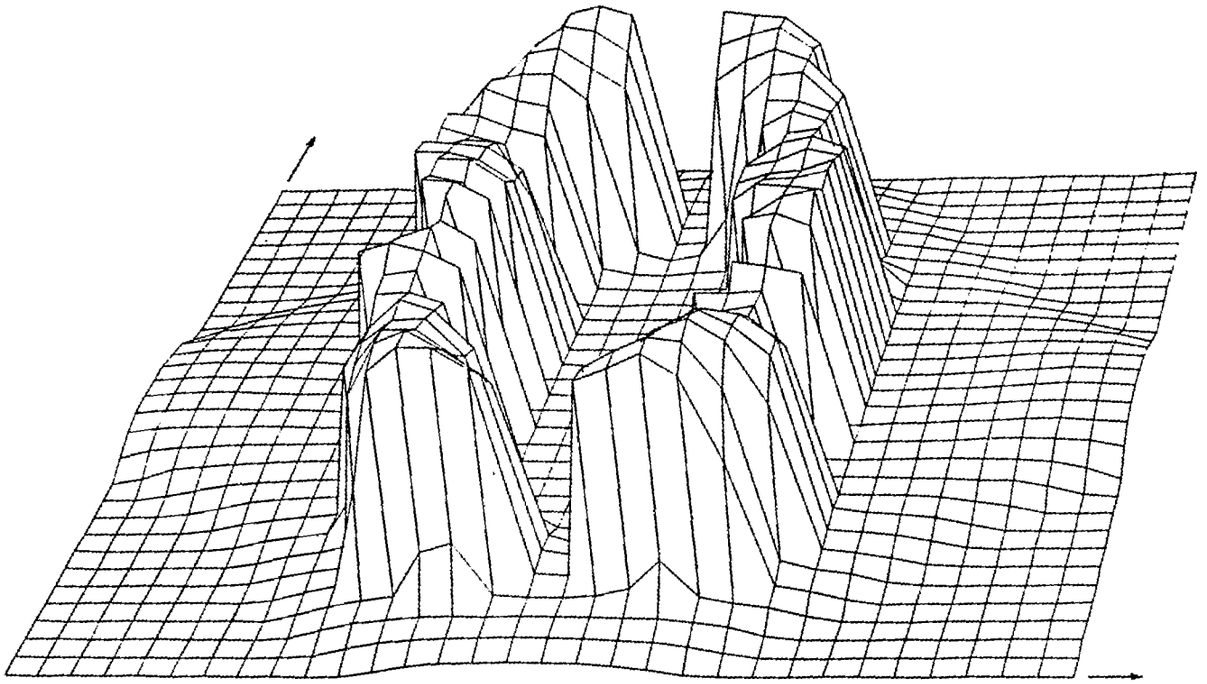


Fig.9.11: Results using ML algorithm for 100 iterations on computer simulated ideal data from source distribution consisting of two point sources of strength 110 units each, separated by 8 voxel elements, and superimposed on a uniform background of strength 1 units.

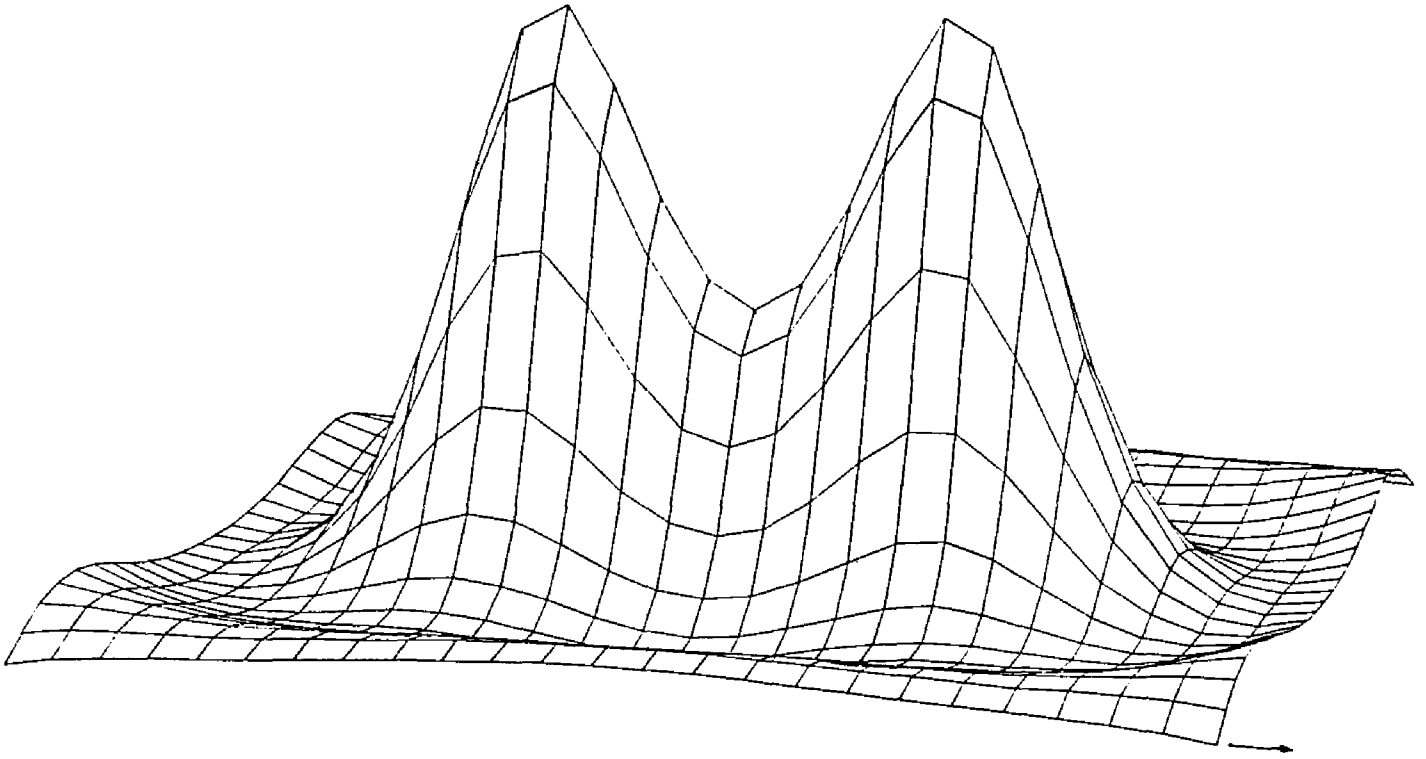
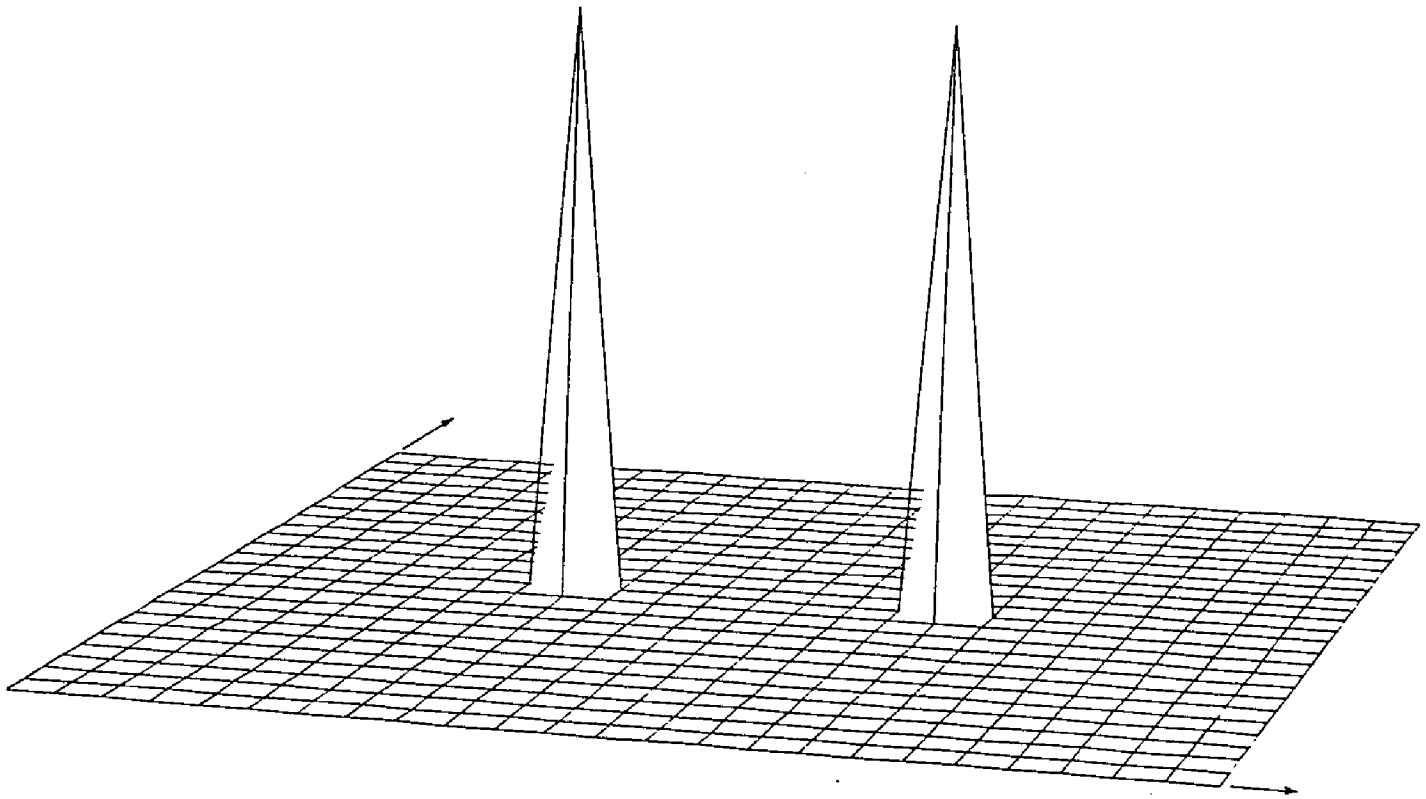


Fig.9.12: Identical results using the two BIP algorithms of fuzzy pattern source constraint and amplitude constraint for 50 iterations on computer simulated ideal data from source distribution consisting of two point sources of strength 110 units each, separated by 8 voxel elements, and superimposed on a uniform background of strength 1 unit. The assumed amplitudes and spatial relation were the same as the source distribution. The weight range was 7 voxel elements.



CHAPTER X

ADDITIVE BAYESIAN IMAGE PROCESSING ALGORITHMS

1). Algorithms for Data Obeying Poisson Statistics

From Bayes' Law of Eq.(1.2) and Eq.(3.48), the ln Bayesian function of Eq.(6.1) is:

$$g(\Phi) \equiv \ln P(\Phi \mid Y) = \ln P(Y \mid \Phi) + \ln P(\Phi) - \ln P(Y)$$
$$= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j) - \ln(Y_i!)] + \ln P(\Phi) - \ln P(Y). \qquad (10.1)$$

1.1). FOR GENERIC CONSTRAINED SOURCES

a). For Uniform $P(\Phi)$:

When

$$P(\Phi) = P_a(\Phi) = \prod_j^J 1/(b_j - a_j)$$

there is:

$$g(\Phi) = \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j) - \ln(Y_i!)] - \sum_j \ln(b_j - a_j) - \ln P(Y)$$
$$= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j)] - [\sum_i \ln(Y_i!) + \sum_j \ln(b_j - a_j) + \ln P(Y)]$$
$$= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j)] - C(...). \qquad (10.2)$$

Using the damping method (Budinger et al 1974), the iterative algorithm is derived below:

From Eq.(10.2), there is:

$$g(\Phi) = \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j)]$$

where the $C(...)$ was dropped since it is independent of Φ. After the n-th iteration of Φ,

$$g(\Phi^{(n)}) = \sum_i [-\sum_j R_{ij}\phi_j^{(n)} + Y_i \ln(\sum_j R_{ij}\phi_j^{(n)})]. \qquad (10.3)$$

The (n+1)th iterated $\{\phi_k\}$ maximizes:

$$g(\Phi^{(n)}, \Phi^{(n+1)}) = \sum_i [-\sum_{j \neq k} R_{ij}\phi_j^{(n)} - R_{ik}\phi_k^{(n+1)} + Y_i \ln(\sum_{j \neq k} R_{ij}\phi_j^{(n)}) + R_{ik}\phi_k^{(n+1)})] \qquad (10.4)$$

that is,

$$\frac{\partial g(...)}{\partial \phi_k^{(n+1)}} = \sum_i [-R_{ik} + \frac{Y_i R_{ik}}{(\sum_{j \neq k} R_{ij}\phi_j^{(n)} + R_{ik}\phi_k^{(n+1)})}] = 0$$

then results in:

$$\sum_i R_{ik} = \sum_i R_{ik} \left[\frac{Y_i}{\sum_j R_{ij}\phi_j^{(n)} + R_{ik}(\phi_k^{(n+1)} - \phi_k^{(n)})} \right]. \tag{10.5}$$

Assuming $| \sum_j R_{ij}\phi_j | \gg | R_{ik}(\phi_k^{(n+1)} - \phi_k^{(n)} |$, and noting that, for $|x| < a$,

$$(a+x)^{-1} = a^{-1} - a^{-2}x + a^{-3}x^2 - \cdots \tag{10.6}$$

Eq.(10.5) becomes:

$$\sum_i R_{ik} = \sum_i R_{ik} [\frac{Y_i}{\sum_j R_{ij}\phi_j^{(n)}} - \frac{R_{ik} Y_i (\phi_k^{(n+1)} - \phi_k^{(n)})}{(\sum_j R_{ij}\phi_j^{(n)})^2}] + O(...)$$

and

$$\phi_k^{(n+1)} = \phi_k^{(n)} + \frac{\sum_i [R_{ik}(Y_i / \sum_j R_{ij}\phi_j^{(n)} - 1)]}{\sum_i [Y_i (R_{ik} / \sum_j R_{ij}\phi_j^{(n)})^2]}. \tag{10.7}$$

From Eq.(10.4), we know that $g(\Phi)$ is maximized for each individual $\phi_k^{(n+1)}$. In such case, the global convergence of Φ cannot be guaranteed and a damping factor $\alpha^{(n)}$ is introduced. Eq.(10.7) becomes:

$$\phi_k^{(n+1)} = \phi_k^{(n)} + \alpha^{(n)} \Delta\phi_k^{(n)}. \tag{10.8}$$

The damping factor $\alpha^{(n)}$ is chosen to maximize $g(\Phi^{(n+1)})$ of Eq.(10.3) at the (n+1)th iteration,

$$g(\Phi^{(n+1)}, \alpha^{(n)}) = \sum_i [- \sum_j R_{ij}\phi_j^{(n)} - \alpha^{(n)} (\sum_j R_{ij}\Delta\phi_j^{(n)}) + Y_i \ln(\sum_j R_{ij}\phi_j^{(n)} + \alpha^{(n)} \sum_j R_{ij}\Delta\phi_j^{(n)})]$$

that is,

$$\frac{\partial g(...)}{\partial \alpha^{(n)}} = \sum_i [- \sum_j R_{ij}\Delta\phi_j^{(n)} + Y_i \frac{\sum_j R_{ij}\Delta\phi_j^{(n)}}{(\sum_j R_{ij}\phi_j^{(n)} + \alpha^{(n)} \sum_j R_{ij}\Delta\phi_j^{(n)})}] = 0$$

and

$$\sum_i (\sum_j R_{ij} \Delta\phi_j^{(n)}) = \sum_i [\frac{Y_i}{(\alpha^{(n)} + \sum_j R_{ij}\phi_j^{(n)} / \sum_j R_{ij}\Delta\phi_j^{(n)})}]. \tag{10.9}$$

By use of Eq.(10.6), for $\alpha^{(n)} \ll \sum_j R_{ij}\phi_j^{(n)} / \sum_j R_{ij}\Delta\phi_j^{(n)}$, there is:

$$\sum_i (\sum_j R_{ij} \Delta\phi_j^{(n)}) = \sum_i [Y_i (\frac{\sum_j R_{ij}\Delta\phi_j^{(n)}}{\sum_j R_{ij}\phi_j^{(n)}}) - \alpha^{(n)} Y_i (\frac{\sum_j R_{ij}\Delta\phi_j^{(n)}}{\sum_j R_{ij}\phi_j^{(n)}})^2] + O(...)$$

and

$$\alpha^{(n)} = \frac{\sum_i [(\sum_j R_{ij}\Delta\phi_j^{(n)})(Y_i / \sum_j R_{ij}\phi_j^{(n)} - 1)]}{\sum_i [Y_i (\sum_j R_{ij}\Delta\phi_j^{(n)} / \sum_j R_{ij}\phi_j^{(n)})^2]}. \tag{10.10}$$

Eqs.(10.8) and (10.10) are the derived iterative algorithm. Another iterative form can be obtained by use of the steepest descent method (Luenberger 1984).

b). For Gaussian $P(\Phi)$:

When

$$P(\Phi) = = P_b(\Phi) = \prod_j^J (2\pi\rho_j^2)^{-1/2} \exp[-(\phi_j - \bar{\phi}_j)^2/(2\rho_j^2)]$$

there is:

$$\begin{aligned} g(\Phi) &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j) - \ln(Y_i!)] \\ &\quad - \sum_j [\tfrac{1}{2}\ln(2\pi\rho_j^2) + (\phi_j - \bar{\phi}_j)^2/(2\rho_j^2)] - \ln P(Y) \\ &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j)] - \sum_j (\phi_j - \bar{\phi}_j)^2/(2\rho_j^2) \\ &\quad - [\sum_i \ln(Y_i!) + \sum_j \tfrac{1}{2}\ln(2\pi\rho_j^2) + \ln P(Y)] \\ &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j)] - \sum_j (\phi_j - \bar{\phi}_j)^2/(2\rho_j^2) - C(...). \end{aligned} \tag{10.11}$$

Using the Picard method (Isaacson et al 1966), the algorithm is formulated following:

$$\frac{\partial g(\Phi)}{\partial \phi_k}\Big|_{\Phi=\Phi^*} = \sum_i [-R_{ik} + \frac{R_{ik}Y_i}{\sum_j R_{ij}\phi_j^*}] - \frac{(\phi_k^* - \bar{\phi}_k)}{\rho_k^2} = 0$$

and

$$\phi_k^* = \bar{\phi}_k + \rho_k^2 \sum_i R_{ik} \left(\frac{Y_i}{\sum_j R_{ij} \phi_j^*} - 1 \right) = F_k^{(1)}(\Phi^*)$$

the iterative algorithm then is:

$$\phi_k^{(n+1)} = F_k^{(1)}(\Phi^{(n)}) = \bar{\phi}_k + \rho_k^2 \sum_i R_{ik} \left(\frac{Y_i}{\sum_j R_{ij} \phi_j^{(n)}} - 1 \right). \tag{10.12}$$

Another iterative form can be obtained by use of the steepest descent method.

c). For Poisson $P(\Phi)$:

When

$$P(\Phi) \Longrightarrow P_c(\Phi) = \prod_j^J (e^{\bar{\phi}_j}) \frac{(\bar{\phi}_j)^{\phi_j}}{(\phi_j!)}$$

there is:

$$\begin{aligned} g(\Phi) &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j) - \ln(Y_i!)] \\ &\quad + \sum_j [-\bar{\phi}_j + \phi_j \ln(\bar{\phi}_j) - \ln(\phi_j!)] - \ln P(\mathbf{Y}) \\ &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j)] + \sum_j [\phi_j \ln(\bar{\phi}_j) - \ln(\phi_j!)] \\ &\quad - [\sum_j \bar{\phi}_j + \ln P(\mathbf{Y})] \\ &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \ln(\sum_j R_{ij}\phi_j)] + \sum_j [\phi_j \ln(\bar{\phi}_j) - \phi_j \ln(\phi_j) + \phi_j] - C(...) \end{aligned} \tag{10.13}$$

where Stirling's approximation (Feller 1968)

$$\ln(\phi_j!) \approx \phi_j(\ln\phi_j - 1)$$

has been used.

Using the Picard method, there is:

$$\frac{\partial g(\Phi)}{\partial \phi_k}\Big|_{\Phi=\Phi^*} = \sum_i [-R_{ik} + \frac{R_{ik} Y_i}{\sum_j R_{ij} \phi_j^*}] + \ln(\bar{\phi}_k) - \ln(\phi_k^*) = 0$$

and

$$\phi_k^* = \bar{\phi}_k \, e^{\sum_i R_{ik}(Y_i / \sum_j R_{ij}\phi_j^* - 1)} = F_k^{(2)}(\Phi^*)$$

the iterative algorithm then has form:

$$\phi_k^{(n+1)} = F_k^{(2)}(\Phi^{(n)}) = \bar{\phi}_k \, e^{\sum_i R_{ik}(Y_i / \sum_j R_{ij}\phi_j^{(n)} - 1)} . \qquad (10.14)$$

Another iterative form can be obtained using the steepest descent method.

In order to accelerate convergence of solution, the algorithm (10.14) can be modified as below (Isaacson et al 1966):

$$\phi_k^{(n+1)} = \theta_k F_k^{(2)}(\Phi^{(n)}) + (1-\theta_k)\phi_k^{(n)} \qquad (10.15)$$

where

$$\theta_k = \frac{1}{1 - F_{kk}^{(2)}(\Phi^{(n)})} , \quad F_{kk}^{(2)} \equiv \frac{\partial F_k^{(2)}}{\partial \phi_k}$$

The same steps can be applied to the $F_k^{(1)}(\Phi)$ of Eq.(10.12).

d). For Correlated $P(\Phi)$:

When

$$P(\Phi) = \prod_{jl}^{J} (2\pi \Xi_{jl})^{-1/2} \exp[-\frac{1}{2} \frac{\chi_{jl}}{\rho_j \rho_l} (\phi_j - \bar{\phi}_j)(\phi_l - \bar{\phi}_l)]$$

there is:

$$\begin{aligned} g(\Phi) &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \, ln(\sum_j R_{ij}\phi_j) - ln(Y_i!)] \\ &\quad - \sum_{jl} [-\frac{1}{2} ln(2\pi \Xi_{jl}) + \frac{1}{2} \frac{\chi_{jl}}{\rho_j \rho_l} (\phi_j - \bar{\phi}_j)(\phi_l - \bar{\phi}_l)] - \ln P(\mathbf{Y}) \\ &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \, ln(\sum_j R_{ij}\phi_j)] - \frac{1}{2} \sum_{jl} \frac{\chi_{jl}}{\rho_j \rho_l} (\phi_j - \bar{\phi}_j)(\phi_l - \bar{\phi}_l) \\ &\quad - [\sum_i ln(Y_i!) + \frac{1}{2} \sum_{jl} ln(2\pi \Xi_{jl}) + \ln P(\mathbf{Y})] \\ &= \sum_i [-\sum_j R_{ij}\phi_j + Y_i \, ln(\sum_j R_{ij}\phi_j)] - \frac{1}{2} \sum_{jl} \frac{\chi_{jl}}{\rho_j \rho_l} (\phi_j - \bar{\phi}_j)(\phi_l - \bar{\phi}_l) - C(...) . \end{aligned} \qquad (10.16)$$

Using the Picard method, we have:

$$\frac{\partial g(\Phi)}{\partial \phi_k} \Big|_{\Phi = \Phi^*} = \sum_i [-R_{ik} + \frac{R_{ik} Y_i}{\sum_j R_{ij} \phi_j^*}] - \frac{\chi_{kk}}{\rho_k^2} (\phi_k^* - \bar{\phi}_k) - \sum_{j \neq k} \frac{\chi_{jk}}{\rho_j \rho_k} (\phi_j^* - \bar{\phi}_j) = 0$$

and

$$\phi_k^* = \bar{\phi}_k + \frac{\rho_k^2}{\chi_{kk}} [\sum_i R_{ik} (\frac{Y_i}{\sum_j R_{ij} \phi_j^*} - 1) - \sum_{j \neq k} \frac{\chi_{jk}}{\rho_j \rho_k} (\phi_j^* - \bar{\phi}_j)] = F_k^{(3)}(\Phi^*)$$

the iterative algorithm is then expressed as:

$$\phi_k^{(n+1)} = F_k^{(3)}(\Phi^{(n)})$$
$$= \bar{\phi}_k + \frac{\rho_k^2}{\chi_{kk}} [\sum_i R_{ik} (\frac{Y_i}{\sum_j R_{ij}\phi_j^{(n)}} - 1) - \sum_{j \neq k} \frac{\chi_{jk}}{\rho_j \rho_k} (\phi_j^{(n)} - \bar{\phi}_j)] . \quad (10.17)$$

1.2). FOR QUANTITATIVE CONSTRAINED SOURCES

Since the quantitative information functions, $P(\Phi)$'s, have very complicated forms, one simple method used to maximize function $g(\Phi)$ of Eq.(10.1) in additive form is described for introduction as below.

By applying the steepest descent method to Eq.(10.1), there are:

$$\phi_k^{(n+1)} = \phi_k^{(n)} + \alpha^{(n)} d_k^{(n)} \quad (10.18)$$

and

$$d_k^{(n)} = -\frac{\partial g(\Phi^{(n)})}{\partial \phi_k^{(n)}} = -[\sum_i R_{ik} (Y_i / \sum_j R_{ij}\phi_j^{(n)} - 1) + q_k^{(n)}] \quad (10.19)$$

where

$$q_k^{(n)} = \frac{\partial \ln P(\Phi)}{\partial \phi_k} |_{\Phi = \Phi^{(n)}} \quad (10.20)$$

and $P(\Phi)$'s were given in sections 2 and 3 in chapter IV.

The step length $\alpha^{(n)}$ may simply be assumed as unit or approximats to Eq.(10.10) with replacement of $\Delta\phi_j^{(n)}$ by $d_j^{(n)}$.

2). Algorithms for Data Obeying Gaussian Statistics

The ln Bayesian function of Eq.(10.1) now, using Eq.(3.49), becomes:

$$\bar{g}(\Phi)=-\sum_i[\,\tfrac{1}{2}\ln(2\pi\sigma_i^2)+(Y_i-\sum_j R_{ij}\phi_j)^2/(2\sigma_i^2)\,]+\ln P(\Phi)-\ln P(\mathbf{Y}). \tag{10.21}$$

2.1). FOR GENERIC CONSTRAINED SOURCES

a). For Uniform $P(\Phi)$:

When

$$P(\Phi)=\!=P_a(\Phi)=\prod_j 1/(b_j-a_j)$$

there is:

$$\begin{aligned}\bar{g}(\Phi)&=-\frac{1}{2}\sum_i[\,(Y_i-\sum_j R_{ij}\phi_j)^2/\sigma_i^2+\ln(2\pi\sigma_i^2)\,]-\sum_j\ln(b_j-a_j)-\ln P(\mathbf{Y})\\&=-\sum_i\frac{1}{2}(Y_i-\sum_j R_{ij}\phi_j)^2/\sigma_i^2-[\,\sum_i\tfrac{1}{2}\ln(2\pi\sigma_i^2)+\sum_j\ln(b_j-a_j)+\ln P(\mathbf{Y})\,]\\&=-\tfrac{1}{2}(\mathbf{Y}-R\Phi)^{\tau}\Lambda^{-1}(\mathbf{Y}-R\Phi)-\mathbf{C}_0\end{aligned} \tag{10.22}$$

where Λ is the covariance matrix for the measured data and is a symmetric positive semi-definite matrix.

One algorithm form can be obtained using the damping method as shown above, the results are:

$$\phi_k^{(n+1)}=\phi_k^{(n)}+\beta^{(n)}\delta\phi_k^{(n)} \tag{10.23}$$

$$\delta\phi_k^{(n)}=\frac{\sum_i[(Y_i-\sum_j R_{ij}\phi_j^{(n)})\frac{R_{ik}}{Y_i}]}{\sum_i(R_{ik}^2/Y_i)} \tag{10.24}$$

and

$$\beta^{(n)}=\frac{\sum_i[(\sum_j R_{ij}\delta\phi_j^{(n)})(Y_i-\sum_j R_{ij}\phi_j^{(n)})/Y_i\,]}{\sum_i[(\sum_j R_{ij}\delta\phi_j^{(n)})^2/Y_i\,]}. \tag{10.25}$$

Here we will use the steepest descent method and the conjugate gradient method (Luenberger 1984) to derive the algorithms which generate the solution iteratively and approach the minimum value of the quadratic function

$$g(\Phi)\equiv-\bar{g}(\Phi)=\tfrac{1}{2}(\mathbf{Y}-R\Phi)^{\tau}\Lambda^{-1}(\mathbf{Y}-R\Phi)+\mathbf{C}_0. \tag{10.26}$$

The steepest descent method gives the iterative forms of

$$\Phi^{(n+1)}=\Phi^{(n)}+\alpha^{(n)}\mathbf{d}^{(n)} \tag{10.27}$$

and

$$\mathbf{d}^{(n)}=-\nabla g(\Phi^{(n)})=R^{\tau}\Lambda^{-1}(\mathbf{Y}-R\,\Phi^{(n)}) \tag{10.28}$$

where the step length $\alpha^{(n)}$ is determined by

$$\frac{\partial g(\Phi^{(n+1)})}{\partial \alpha^{(n)}}=\frac{\partial g(\Phi^{(n)}+\alpha^{(n)}\mathbf{d}^{(n)})}{\partial \alpha^{(n)}}=0 \tag{10.29}$$

and it results in:

$$\alpha^{(n)}=\frac{<\mathbf{d}^{(n)},\mathbf{d}^{(n)}>}{<\mathbf{d}^{(n)},R^{\tau}\Lambda^{-1}R\,\mathbf{d}^{(n)}>}. \tag{10.30}$$

Other two formulas, which are useful for computer programming, are:

$$\mathbf{d}^{(n+1)}-\mathbf{d}^{(n)}=-\alpha^{(n)}R^{\tau}\Lambda^{-1}R\,\mathbf{d}^{(n)} \tag{10.31}$$

and

$$g(\Phi^{(n+1)})-g(\Phi^{(n)})=-\tfrac{1}{2}\alpha^{(n)}<\mathbf{d}^{(n)},\mathbf{d}^{(n)}>. \tag{10.32}$$

The steps of performing the iterative algorithm of Eqs.(10.27), (10.28) and (10.30) are:

(i). $\mathbf{d}^{(0)}=R^{\tau}\Lambda^{-1}(\mathbf{Y}-R\,\Phi^{(0)})$;

(ii). $g(\Phi^{(0)})=\tfrac{1}{2}\Big[(\Phi^{(0)})^{\tau}R^{\tau}\Lambda^{-1}(R\,\Phi^{(0)}-2\mathbf{Y})+\mathbf{Y}^{\tau}\Lambda^{-1}\mathbf{Y}\Big]$;

(iii). $\Omega^{(n)}=R^{\tau}\Lambda^{-1}R\,\mathbf{d}^{(n)}$;

(iv). $\alpha^{(n)}=\dfrac{<\mathbf{d}^{(n)},\mathbf{d}^{(n)}>}{<\mathbf{d}^{(n)},\Omega^{(n)}>}$;

(v). $\Phi^{(n+1)}=\Phi^{(n)}+\alpha^{(n)}\mathbf{d}^{(n)}$;

(vi). $g(\Phi^{(n+1)})=g(\Phi^{(n)})-\tfrac{1}{2}\alpha^{(n)}<\mathbf{d}^{(n)},\mathbf{d}^{(n)}>$;

(vii). $\mathbf{d}^{(n+1)}=\mathbf{d}^{(n)}-\alpha^{(n)}\Omega^{(n)}$;

(viii). go back to (*iii*).

The conjugate gradient method gives the iterative forms of

$$\mathbf{q}^{(n+1)}=\mathbf{d}^{(n+1)}+\beta^{(n)}\mathbf{q}^{(n)}\ ,\qquad \beta^{(n)}=-\frac{<\mathbf{q}^{(n)},R^{\tau}\Lambda^{-1}R\,\mathbf{d}^{(n+1)}>}{<\mathbf{q}^{(n)},R^{\tau}\Lambda^{-1}R\,\mathbf{q}^{(n)}>} \tag{10.33}$$

$$\Phi^{(n+1)}=\Phi^{(n)}+\alpha^{(n)}\mathbf{q}^{(n)}\ ,\qquad \alpha^{(n)}=\frac{<\mathbf{q}^{(n)},\mathbf{d}^{(n)}>}{<\mathbf{q}^{(n)},R^{\tau}\Lambda^{-1}R\,\mathbf{q}^{(n)}>} \tag{10.34}$$

$$\mathbf{d}^{(n+1)}=-\nabla g(\Phi^{(n+1)})=\mathbf{d}^{(n)}-\alpha^{(n)}R^{\tau}\Lambda^{-1}R\,\mathbf{q}^{(n)} \tag{10.35}$$

and

$$g(\Phi^{(n+1)})-g(\Phi^{(n)})=-\tfrac{1}{2}\alpha^{(n)}<\mathbf{q}^{(n)},\mathbf{d}^{(n)}>. \tag{10.36}$$

Other three identities, which are very useful for computer programming, are:

$$<\mathbf{d}^{(m)},\mathbf{d}^{(n)}>=k\delta_{mn}\ ,\qquad \alpha^{(n)}=\frac{<\mathbf{d}^{(n)},\mathbf{d}^{(n)}>}{<\mathbf{q}^{(n)},R^{\tau}\Lambda^{-1}R\,\mathbf{q}^{(n)}>} \tag{10.37}$$

and

$$\beta^{(n+1)}=\frac{<\mathbf{d}^{(n+1)},\mathbf{d}^{(n+1)}>}{<\mathbf{d}^{(n)},\mathbf{d}^{(n)}>}. \tag{10.38}$$

The steps on computer programming are:

(i). $\mathbf{d}^{(0)}=R^{\tau}\Lambda^{-1}(\mathbf{Y}-R\,\Phi^{(0)})$;

(ii). $g(\Phi^{(0)})=\tfrac{1}{2}\Big[(\Phi^{(0)})^{\tau}R^{\tau}\Lambda^{-1}(R\,\Phi^{(0)}-2\mathbf{Y})-\mathbf{Y}^{\tau}\Lambda^{-1}\mathbf{Y}\Big]$;

(iii). $\mathbf{q}^{(0)}=0$, $\quad\beta^{(0)}=0$;

(iv). $\beta^{(n)}=\beta^{(n-1)}<\mathbf{d}^{(n-1)},\mathbf{d}^{(n-1)}>$;

(v). $\mathbf{q}^{(n)}=\mathbf{d}^{(n-1)}+\beta^{(n)}\mathbf{q}^{(n-1)}$;

(vi). $\beta^{(n)}=1/<\mathbf{d}^{(n-1)},\mathbf{d}^{(n-1)}>$;

(vii). $\Omega^{(n)}=R^{\tau}\Lambda^{-1}R\,\mathbf{q}^{(n)}$;

(viii). $\alpha^{(n)}=\dfrac{<\mathbf{d}^{(n-1)},\mathbf{d}^{(n-1)}>}{<\mathbf{q}^{(n)},\Omega^{(n)}>}$;

(ix). $\Phi^{(n)}=\Phi^{(n-1)}+\alpha^{(n)}\mathbf{q}^{(n)}$;

(x). $\mathbf{d}^{(n)}=\mathbf{d}^{(n-1)}-\alpha^{(n)}\Omega^{(n)}$;

(xi). $g(\Phi^{(n)})=g(\Phi^{(n-1)})-\tfrac{1}{2}<\mathbf{d}^{(n-1)},\mathbf{d}^{(n-1)}>$;

(xii). go back to (*iv*).

Angel et al (1978, Appl. Optics) applied the conjugate gradient method to solve the least squares image restoration of Eq.(3.16), in which Λ^{-1} is a unit matrix and R a spatially varying function.

b). For Gaussian $P(\Phi)$:

When

$$P(\Phi) = P_b(\Phi) = \prod_j (2\pi\rho_j^2)^{-1/2} \exp[-(\phi_j - \bar{\phi}_j)^2/(2\rho_j^2)]$$

there is:

$$\begin{aligned}
\bar{g}(\Phi) &= -\frac{1}{2}\sum_i [(Y_i - \sum_j R_{ij}\phi_j)^2/(\sigma_i^2) + ln(2\pi\sigma_i^2)] \\
&\quad -\frac{1}{2}\sum_j [(\phi_j - \bar{\phi}_j)^2/\rho_j^2 - ln(2\pi\rho_j^2)] - \ln P(Y) \\
&= -\frac{1}{2}[\sum_i (Y_i - \sum_j R_{ij}\phi_j)^2/\sigma_i^2 + \sum_j (\phi_j - \bar{\phi}_j)^2/\rho_j^2] \\
&\quad -[\tfrac{1}{2}\sum_i ln(2\pi\sigma_i^2) + \tfrac{1}{2}\sum_j ln(2\pi\rho_j^2) + \ln P(Y)] \\
&= -\frac{1}{2}\sum_i (Y_i - \sum_j R_{ij}\phi_j)^2/\sigma_i^2 - \frac{1}{2}\sum_j (\phi_j - \bar{\phi}_j)^2/\rho_j^2 - C_0(...) \\
&= -\tfrac{1}{2}(Y - R\Phi)^t \Lambda^{-1}(Y - R\Phi) - \tfrac{1}{2}(\Phi - \bar{\Phi})^t \Gamma^{-1}(\Phi - \bar{\Phi}) - C_0
\end{aligned} \tag{10.39}$$

where Γ is the covariance matrix for the voxel values Φ and is a symmetric positive definite matrix.

The Flether-Reeres method (Luenberger 1973) generates the iterative form to approach that solution Φ^* which gives the minimum value of the non-quadratic function of

$$\begin{aligned}
g(\Phi) &\equiv -\bar{g}(\Phi) \\
&= \tfrac{1}{2}(Y - R\Phi)^t \Lambda^{-1}(Y - R\Phi) + \tfrac{1}{2}(\Phi - \bar{\Phi})^t \Gamma^{-1}(\Phi - \bar{\Phi}) + C_0
\end{aligned} \tag{10.40}$$

via the conjugate gradient algorithm, as shown below:

(i) given initial $\Phi^{(0)}$, compute $\mathbf{d}^{(0)} = \nabla g(\Phi^{(0)})$, and set $\mathbf{q}^{(0)} = -\mathbf{d}^{(0)}$;

(ii) for $n = 0,1,2,...,m-1$, set

$$\Phi^{(n+1)} = \Phi^{(n)} + \alpha^{(n)}\mathbf{q}^{(n)} \tag{10.41}$$

where $\alpha^{(n)}$ minimizes function $g(\Phi^{(n)} + \alpha^{(n)}\mathbf{q}^{(n)})$;

(iii) compute $\mathbf{d}^{(n+1)} = \nabla g(\Phi^{(n+1)})$;

(iv) unless $n = m-1$, set

$$\mathbf{q}^{(n+1)} = \mathbf{d}^{(n+1)} + \beta^{(n)}\mathbf{q}^{(n)} \tag{10.42}$$

where

$$\beta^{(n)}=\frac{<\mathbf{d}^{(n+1)},\mathbf{d}^{(n+1)}>}{<\mathbf{d}^{(n)},\mathbf{d}^{(n)}>}; \tag{10.43}$$

(v) when $n=m-1$, replace $\Phi^{(0)}$ by $\Phi^{(n)}$ and go back to (i).

Other methods used before are indicated to derive some different iterative forms.

c). For Poisson $P(\Phi)$:

When

$$P(\Phi)=P_c(\Phi)=\prod_j (e^{-\bar{\phi}_j})\,(\bar{\phi}_j)^{\phi_j}/\phi_j!$$

there is:

$$\begin{aligned}\tilde{g}(\Phi)&=-\sum_i[(Y_i-\sum_j R_{ij}\phi_j)^2/(2\sigma_i^2)+\tfrac{1}{2}\ln(2\pi\sigma_i^2)]\\&\quad+\sum_j[-\bar{\phi}_j+\phi_j\ln(\bar{\phi}_j)-\ln(\phi_j!)]-\ln P(\mathbf{Y})\\&=-[\sum_i(Y_i-\sum_j R_{ij}\phi_j)^2/(2\sigma_i^2)+\sum_j(\phi_j\ln(\phi_j)-\phi_j\ln(\bar{\phi}_j)-\phi_j)]\\&\quad-[\tfrac{1}{2}\sum_i\ln(2\pi\sigma_i^2)+\sum_j\bar{\phi}_j+\ln P(\mathbf{Y})]\\&=-\frac{1}{2}\sum_i(Y_i-\sum_j R_{ij}\phi_j)^2/\sigma_i^2+\sum_j[\phi_j\ln(\bar{\phi}_j)-\phi_j\ln(\phi_j)+\phi_j]-C_0(\ldots).\end{aligned} \tag{10.44}$$

The Flether-Reeres method is applicable to search for that solution which maximizes function $\tilde{g}(\Phi)$.

d). For Correlated $P(\Phi)$:

When

$$P(\Phi)=\prod_{jl}(2\pi\Xi_{jl})^{-1/2}\exp[-\tfrac{1}{2}\frac{\chi_{jl}}{\rho_j\rho_l}(\phi_j-\bar{\phi}_j)(\phi_l-\bar{\phi}_l)]$$

there is:

$$\begin{aligned}\tilde{g}(\Phi)&=-\frac{1}{2}[\sum_i(Y_i-\sum_j R_{ij}\phi_j)^2/\sigma_i^2+\sum_{jl}\frac{\chi_{jl}}{\rho_j\rho_l}(\phi_j-\bar{\phi}_j)(\phi_l-\bar{\phi}_l)]\\&\quad-[\sum_i\tfrac{1}{2}\ln(2\pi\sigma_i^2)+\sum_{jl}\tfrac{1}{2}\ln(2\pi\Xi_{jl})+\ln P(\mathbf{Y})]\\&=-\tfrac{1}{2}(\mathbf{Y}-R\Phi)^{\tau}\Lambda^{-1}(\mathbf{Y}-R\Phi)-\tfrac{1}{2}(\Phi-\bar{\Phi})^{\tau}\Gamma_x^{-1}(\Phi-\bar{\Phi})-C_0\end{aligned} \tag{10.45}$$

where Γ_x is generally not a symmetric matrix and reflects the correlation between neightboring voxels.

The Flether-Reeres method is also applicable in this situation.

2.2). FOR QUANTITATIVE CONSTRAINED SOURCES

The method of Eqs.(10.18), (10.19) and (10.20) is applicable to approach to the maximum value of $\bar{g}(\Phi)$ of Eq.(10.21) with those complicated quantitative a priori information functions $P(\Phi)$'s formulated in chapter IV. The Flether-Reeres method is, of course, valid in these situations.

3). Algorithms for Correlated Data

For correlated measured data, there is:

$$P(Y|\Phi)=\prod_{ir}(2\pi\Sigma_{ir})^{-1/2}\exp[-\frac{1}{2}\frac{\kappa_{ir}}{\sigma_i\sigma_r}(Y_i-\sum_j R_{ij}\phi_j)(Y_r-\sum_j R_{rj}\phi_j)] \tag{10.46}$$

and for the generic correlated voxel values:

$$P(\Phi)=\prod_{jl}(2\pi\Xi_{jl})^{-1/2}\exp[-\frac{1}{2}\frac{\chi_{jl}}{\rho_j\rho_l}(\phi_j-\bar{\phi}_j)(\phi_l-\bar{\phi}_l)]. \tag{10.47}$$

The ln Bayesian function under such conditions is:

$$\begin{aligned}\bar{g}(\Phi)=&-\frac{1}{2}\left[\sum_{ir}\frac{\kappa_{ir}}{\sigma_i\sigma_r}(Y_i-\sum_j R_{ij}\phi_j)(Y_r-\sum_j R_{rj}\phi_j)+\sum_{jl}\frac{\chi_{jl}}{\rho_j\rho_l}(\phi_j-\bar{\phi}_j)(\phi_l-\bar{\phi}_l)\right]\\&-[\sum_{ir}\tfrac{1}{2}\ln(2\pi\Sigma_{ir})+\sum_{jl}\tfrac{1}{2}\ln(2\pi\Xi_{jl})+\ln P(Y)]\\=&-\tfrac{1}{2}(Y-R\Phi)^{\tau}\Lambda_x^{-1}(Y-R\Phi)-\tfrac{1}{2}(\Phi-\bar{\Phi})^{\tau}\Gamma_x^{-1}(\Phi-\bar{\Phi})-C_0\end{aligned} \tag{10.48}$$

where Λ_x is generally not a symmetric matrix and reflects the correlation between the neighboring data elements.

Let

$$\tfrac{1}{2}\Lambda_x^{-1}=a\,A\ ,\quad \tfrac{1}{2}\Gamma_x^{-1}=b\,B+c\,C^{-1} \tag{10.49}$$

and consider that maximizing $\bar{g}(\Phi)$ is equivalent to minimizing $g(\Phi)\equiv-\bar{g}(\Phi)$, then the solution Φ^* maximizes $\bar{g}(\Phi)$ and also minimizes function

$$g(\Phi)=a(Y-R\Phi)^{\tau}A(Y-R\Phi)+(\Phi-\bar{\Phi})^{\tau}(bB+c\,C^{-1})(\Phi-\bar{\Phi})+C_0 \tag{10.50}$$

i.e., the Quadratic optimization resulted, where A is a non-negative definite symmetric matrix, B and C are positive definite symmetric matrices.

a). The Quadratic iterative algorithm (Herman et al 1976, Comput. Graph. Image Proc. and Elfving et al 1979) is given below:

(i) for $b>0$ or $c>0$, and $\Phi^{(0)}$ any initial vector, then,

$$\Phi^{(n+1)}=\Phi^{(n)}+\delta^{(n)}C^{t-1}\left[\ a\ C\,R^{\tau}A(Y-R\,\Phi^{(n)})-(b\ C\,B+c\,I)(\Phi^{(n)}-\bar{\Phi})\right] \qquad (10.51)$$

where t=1 or 2.

(ii) for $b^2+c^2>0$, then,

$$\Phi^{(n+1)}=\Phi^{(n)}+\delta^{(n)}\left[\ a\ C\,R^{\tau}A(Y-R\,\Phi^{(n)})-(b\ C\,B+c\ I)(\Phi^{(n)}-\bar{\Phi})\right] \qquad (10.52)$$

with

$$\delta^{(n)}=\frac{2}{\lambda_l(a\ C\,R^{\tau}A\,R+b\ C\,B)+\lambda_s(a\ C\,R^{\tau}A\,R+b\ C\,B)+2c} \qquad (10.53)$$

where $\{\lambda_i\}$ are the eigenvalues of the matrix

$$W_{\lambda}=a\ C^{1/2}R^{\tau}A\,R\,C^{1/2}+b\ C^{1/2}B\,C^{1/2}+c\ I \qquad (10.54)$$

or

$$W_{\lambda}\Phi=\lambda\Phi$$

and λ_l and λ_s are the largest and smallest positive eigenvalues respectively.

b). The Bayesian method (Hunt 1977) is given by using the steepest descent method to Eq.(10.48) with the assumption of all $\bar{\phi}_j>0$.

c). The improved Bayesian method (Trussell et al 1979) is given by using the Picard method to Eq.(10.48) with the assumption of all $\bar{\phi}_j>0$.

For the quantitative correlated source distributions discussed in sections 2) and 3) in chapter IV, the ln Bayesian function now is:

$$\bar{g}(\Phi)=-\frac{1}{2}\left[\sum_{ir}\frac{\kappa_{ir}}{\sigma_i\sigma_r}(Y_i-\sum_j R_{ij}\phi_j)(Y_r-\sum_j R_{rj}\phi_j)+P(\Phi)\right]$$

$$-\left[\sum_{ir}\tfrac{1}{2}\ln(2\pi\Sigma_{ir})+\ln P(Y)\right] \qquad (10.55)$$

The Flether-Reeres method is indicated in this situation to maximize function $\bar{g}(\Phi)$ of Eq.(4.55).

CHAPTER XI
SUMMARY AND CONCLUSIONS

Bayesian image processing formalism is a method of finding the source distribution which is most likely to have given rise to the measured data, subject to the a priori information about the source distribution and the nature of the statistical variation of the measured data.

In the BIP formalism, the statistical nature of the measured data can be incorporated accurately into the conditional probability density function $P(Y|\Phi)$ and the source probability density function $P(\Phi)$ considers any a priori source distribution information which may be available.

The applicability of two general species of a priori source information have now been tested principally in one dimensional situations with computer generated and measured data from phantoms:

a. Generic information in which the source characteristics are constrained by general properties such as maximum entropy considerations.

b. More restrictive quantitative a priori amplitude and/or spatial constraints arising from the nature of the source field or from supplementary measurement.

Depending upon the restrictiveness and accuracy of the a priori source information, the BIP formalism is capable of striking improvement in image processing results over those obtainable by any of the prior standard image procesing methods (see chapter I).

Mathematically most of the standard methods can be viewed as special cases of Bayesian analysis in which the a priori source information function $P(\Phi)$ is generically restricted. The quadratic optimization can be reformulated in a Bayesian frame for each element of measured data and sources obeying gaussian statistics. The maximum likelihood estimation is a special case of Bayesian analysis when the $P(\Phi)$ is uniformly distributed (Hunt 1977). The least square approach determines in particular the maximum likelihood solution when the measured data obeys gaussian statistics while the SIRT was shown to strongly resemble the least square approach in combination with the Richardson algorithm (Lakshminarayanan et al 1979).

Extension and application of the BIP formalism to multi-dimensional static and dynamic systems are currently being considered.

APPENDIX

1). CONVEX AND CONCAVE FUNCTIONS

Let $f(x)$ be a real-valued function of x, $f(x)\varepsilon IR^n$ and $x\,\varepsilon\Omega$, where $\Omega=IR^n$ or Ω is a subset of IR^n.

a). Convex Function

Definition: A function $f(x)$ defined on Ω is said to be convex if, for every x_a, $x_b\,\varepsilon\Omega$ and every τ, $0\leq\tau\leq 1$, there holds:

$$f(\tau x_a+(1-\tau)x_b)\leq \tau f(x_a)+(1-\tau)f(x_b). \tag{a.1}$$

If, for every $0<\tau<1$ and $x_a\neq x_b$, there holds:

$$f(\tau x_a+(1-\tau)x_b)< \tau f(x_a)+(1-\tau)f(x_b) \tag{a.2}$$

then $f(x)$ is said to be strictly convex (Luenberger 1973).

b). Concave Function

Definition: A function $f(x)$ defined on Ω is said to be concave if the function $-f(x)$ is convex, i.e. for every x_a, $x_b\,\varepsilon\Omega$ and every τ, $0\leq\tau\leq 1$, there holds:

$$f(\tau x_a+(1-\tau)x_b)\geq \tau f(x_a)+(1-\tau)f(x_b). \tag{a.3}$$

The function $f(x)$ is said to be strictly concave if $-f(x)$ is strictly convex, i.e. for every $0<\tau<1$ and $x_a\neq x_b$, there holds:

$$f(\tau x_a+(1-\tau)x_b)> \tau f(x_a)+(1-\tau)f(x_b). \tag{a.4}$$

Proposition 1): Let $f_1(x)$ and $f_2(x)$ be concave functions defined on Ω. Then the function $f_1(x)+f_2(x)$ is concave on Ω.

Proof: Applying Eq.(a.3) to $f_1(x)$ and $f_2(x)$ respectively and adding, then,

$$\begin{aligned} f_1(\tau x_a+(1-\tau)x_b)+f_2(\tau x_a+(1-\tau)x_b)\geq \tau[f_1(x_a)+f_2(x)] \\ +(1-\tau)[f_1(x_b)+f_2(x_b)]. \end{aligned} \tag{a.5}$$

Proposition 2): Let $f_i(x)$, $i=1,2,\ldots,n$, be concave functions defined on Ω. Then the function $\sum_{i=1}^{n}\nu_i f_i(x)$ is concave for all $\nu_i\geq 0$.

Proof: Function $\nu_i f_i(x)$ is concave using Eq.(a.3). Then proof is achieved immediately using the proposition 1,

$$\sum_{i=1}^{n} v_i f_i(\tau x_a + (1-\tau)x_b) \geq \tau \sum_{i=1}^{n} v_i f_i(x_a) \qquad \text{(a.6)}$$

$$+(1-\tau)\sum_{i=1}^{n} v_i f_i(x_b).$$

Proposition 3): Let $f(x)$ be a continuously differentiable function defined on Ω. Then $f(x)$ is concave if and only if

$$f(x_b) \leq f(x_a) + \nabla f(x_a)(x_b - x_a). \qquad \text{(a.7)}$$

Proof: First suppose $f(x)$ is concave to prove the "only if" part. Then for all τ, $0 \leq \tau \leq 1$,

$$f(\tau x_a + (1-\tau)x_b) \geq \tau f(x_b) + (1-\tau) f(x_a)$$

and for $0 < \tau \leq 1$

$$\frac{f(x_a + \tau(x_b - x_a)) - f(x_a)}{\tau} \geq f(x_b) - f(x_a)$$

let $\tau \to 0$, it results in:

$$\nabla f(x_a)(x_b - x_a) \geq f(x_b) - f(x_a).$$

Now assume Eq.(a.7) holds to prove the remainder "if" part. Then for all x_c, $x_d \in \Omega$ and all τ, $0 \leq \tau \leq 1$,

$$f(x_c) \leq f(x_a) + \nabla f(x_a)(x_c - x_a) \qquad \text{(a.8)}$$

and

$$f(x_d) \leq f(x_a) + \nabla f(x_a)(x_d - x_a). \qquad \text{(a.9)}$$

Multiplying Eq.(a.8) by τ and Eq.(a.9) by $(1-\tau)$ and then adding, it results in:

$$\tau f(x_c) + (1-\tau) f(x_d) \leq f(x_a) + \nabla f(x_a)[\tau x_c + (1-\tau)x_d - x_a]$$

and let $x_a = \tau x_c + (1-\tau)x_d$, then,

$$\tau f(x_c) + (1-\tau) f(x_d) \leq f(\tau x_c + (1-\tau)x_d).$$

QED.

If, for $x_a \neq x_b$ and $0 < \tau < 1$, there holds:

$$f(x_b) < f(x_a) + \nabla f(x_a)(x_b - x_a) \qquad \text{(a.10)}$$

then $f(x)$ is strictly concave.

Proposition 4): Let $f(x)$ be a twice continuously differentiable function defined on Ω. Then $f(x)$ is concave if and only if the Hessian matrix $F(x)$ of $f(x)$ satisfies:

$$(x_b - x_a)^T F(x_a + \tau(x_b - x_a))(x_b - x_a) \leq 0 \qquad \text{(a.11)}$$

where the Hessian matrix $F(x)$ is a $n \times n$ matrix with elements of $F_{ij}(x) = \frac{\partial^2 f(x)}{\partial x_i \, \partial x_j}$.

Proof: By Taylor's theorem, there is:

$$f(x_b) = f(x_a) + \nabla f(x_a)(x_b - x_a)$$
$$+ \tfrac{1}{2}(x_b - x_a)^T F(x_a + \tau(x_b - x_a))(x_b - x_a)$$

for some τ, $0 \leq \tau \leq 1$. Clearly, if Eq.(a.11) holds, then using proposition 3, $f(x)$ is concave.

If, for $x_b \neq x_a$ and $0 < \tau < 1$, there holds:

$$(x_b - x_a)^T F(x_a + \tau(x_b - x_a))(x_b - x_a) < 0 \qquad \text{(a.12)}$$

then $f(x)$ is strictly concave.

Alternatively, let $x = x_b - x_a \in \Omega$ and x' be a point within interval $[x_a, x_b]$, then,

$$x^T F(x') x < 0. \qquad \text{(a.13)}$$

Proposition 5): Let $f(x)$ be a concave function defined on Ω. The set $\Gamma_c = \{x : x \in \Omega, f(x) > c\}$ is concave for every real value c.

Proof: Let $x_a, x_b \in \Gamma_c$, then $f(x_a) \geq c$, $f(x_b) \geq c$ and for $0 \leq \tau \leq 1$

$$f(\tau x_a + (1-\tau) x_b \geq \tau f(x_a) + (1-\tau) f(x_b) \geq c \qquad \text{(a.14)}$$

thus

$$\tau x_a + (1-\tau) x_b \in \Gamma_c. \qquad \text{(a.15)}$$

Proposition 6): Let $f(x)$ be a concave function defined on Ω. There is a set $\Gamma_{cm} \in \Omega$ on which $f(x)$ reaches its maximal value c_m.

Proof: Immediate using proposition 5,

$$f(\tau x_a + (1-\tau) x_b) = c_m$$

and

$$\tau x_a + (1-\tau) x_b \in \Gamma_{cm}. \qquad \text{(a.16)}$$

Proposition 7): Let $f(x)$ be a strictly concave function defined on Ω. Then there is one and only one point $x^* \varepsilon \Omega$ at which $f(x)$ reaches its maximal value c_m.

Proof: Suppose that there are two points, x_a, $x_b \varepsilon \Gamma_{cm}$, $x_a \neq x_b$, exist, at which $f(x_a)=c_m$ and $f(x_b)=c_m$, then, for $0<\tau<1$ and $\tau x_a+(1-\tau)x_b \varepsilon \Gamma_{cm}$,

$$f(\tau x_a+(1-\tau)x_b) > \tau f(x_a)+(1-\tau)f(x_b)$$
$$> \tau c_m+(1-\tau)c_m > c_m. \qquad \text{(a.17)}$$

This is contradicted to the fact that c_m is the maximal value of $f(x)$, QED.

c). Discussion

(i) $f(x)$ is locally concave if its Hessian matrix satisfies:

$$x^\tau F x \leq 0 \qquad \text{(a.18)}$$

on a subset of Ω;

(ii) $f(x)$ is locally strictly concave if its Hession matrix satisfies:

$$x^\tau F x < 0 \qquad \text{(a.19)}$$

on a subset of Ω;

(iii) The results that are based on concavity apply even to non-concave problems on a subset near the solution, and conversely, local results apply to a global maximum point;

(iv) Similar results will be obtained for convex function $f(x)$.

The discussion iii) is consistent with the previous intuitions about the concavity of function (7.3) and the unique solution of Eq.(8.2).

The strictly concavity of functions $g(\Phi)$ and $\bar{g}(\Phi)$ of Eqs.(6.1) and (6.8) for the generic constraints Eqs.(4.11), (4.12), (4.13) and (4.14) are shown below.

For $P(\Phi)=P_a(\Phi)$, and by use of Eqs.(6.1), (4.11) and (a.13), the quadratic form of $P(\Phi)$ is:

$$Z^\tau[\nabla^2 g(\Phi)]Z=-\sum_i Y_i(\sum_l R_{il} Z_l/\sum_j R_{ij}\phi_j)^2 < 0 \qquad \text{(a.20)}$$

for non-vanishing Z.

For $P(\Phi)=P_b(\Phi)$, and Eqs.(6.1), (4.12) and (a.13) give:

$$Z^\tau[\nabla^2 g(\Phi)]Z=-[\sum_i Y_i(\sum_l R_{il} Z_l/\sum_j R_{ij}\phi_j)^2+\sum_j (Z_j/\sigma_j)^2] < 0. \qquad \text{(a.21)}$$

For $P(\Phi)=P_c(\Phi)$, the quadratic form of $g(\Phi)$ is:

$$Z^t[\nabla^2 g(\Phi)]Z=-[\sum_i Y_i(\sum_l R_{il}\,Z_l/\sum_j R_{ij}\phi_j)^2+\sum_j Z_j^2/\phi_j^2]<0. \qquad \text{(a.22)}$$

For Eq.(4.14), there is:

$$Z^t[\nabla^2 g(\Phi)]Z=-[\sum_i Y_i(\sum_l R_{il}\,Z_l/\sum_j R_{ij}\phi_j)^2+\tfrac{1}{2}\sum_j \chi_{jj}\,Z_j^2/\rho_j^2]<0. \qquad \text{(a.23)}$$

Similarly, the quadratic form of $\bar{g}(\Phi)$ for the four a priori probability density functions are:

$$Z^t[\nabla^2 \bar{g}(\Phi)]Z=-\sum_i(\sum_j R_{ij}Z_j/\sigma_i)^2<0 \qquad \text{(a.24)}$$

$$Z^t[\nabla^2 \bar{g}(\Phi)]Z=-[\sum_i(\sum_j R_{ij}Z_j/\sigma_i)^2+\sum_j(Z_j/\sigma_j)^2]<0 \qquad \text{(a.25)}$$

$$Z^t[\nabla^2 \bar{g}(\Phi)]Z=-[\sum_i(\sum_j R_{ij}Z_j/\sigma_i)^2+\sum_j Z_j^2/\phi_j^2]<0 \qquad \text{(a.26)}$$

and

$$Z^t[\nabla^2 \bar{g}(\Phi)]Z=-[\sum_i(\sum_j R_{ij}Z_j/\sigma_i)^2+\tfrac{1}{2}\sum_j \chi_{jj}\,Z_j^2/\rho_j^2]<0. \qquad \text{(a.27)}$$

Eqs.(a.20),..., (a.27) satisfy the condition (a.13) whenever the matrix R has at least as many rows as columns and is of full column rank. Therefore if matrix R satisfies the conditions above, functions (6.1) and (6.8) are strictly concave for the four a priori probability density functions $P(\Phi)$ and Eqs.(6.2), (6.4), (6.6), (6.10), (6.11), (6.12), (6.33) and (6.35) (set $\phi_k^{(n+1)}=\phi_k^{(n)}$) respectively determine the unique solutions Φ^*.

2). ENTROPY FORMULA

As shown in the fig.(a.1) below, let's consider a picture element (pixel) of area A, in the case 1, that can assume any of Q equally likely gray levels. Then, as any gray level is displayed, it is said that a definite amount of information is received.

If two pixels are now displayed together (in the case 2), each pixel as in case 1, then, intuitively, it would be said that the information received from the two-pixel image has been doubled.

If the pixel area A in the case 1 is divided into two halves, in the case 3, and each can independently assume any of Q equally likely gray levels, then, intuitively, it would also be said that the information received is doubled from that of the case 1.

If the information is defined as the number of possible gray level combinations in a displayed image, then, for cases 1 and 2, the quantities of information are Q and Q^2 respectively, which thus are not consistent with the previous intuitive notion. Similar arguments also apply to the cases 1 and 3. However, if the information is defined as the logarithm of possible gray levels in a displayed image

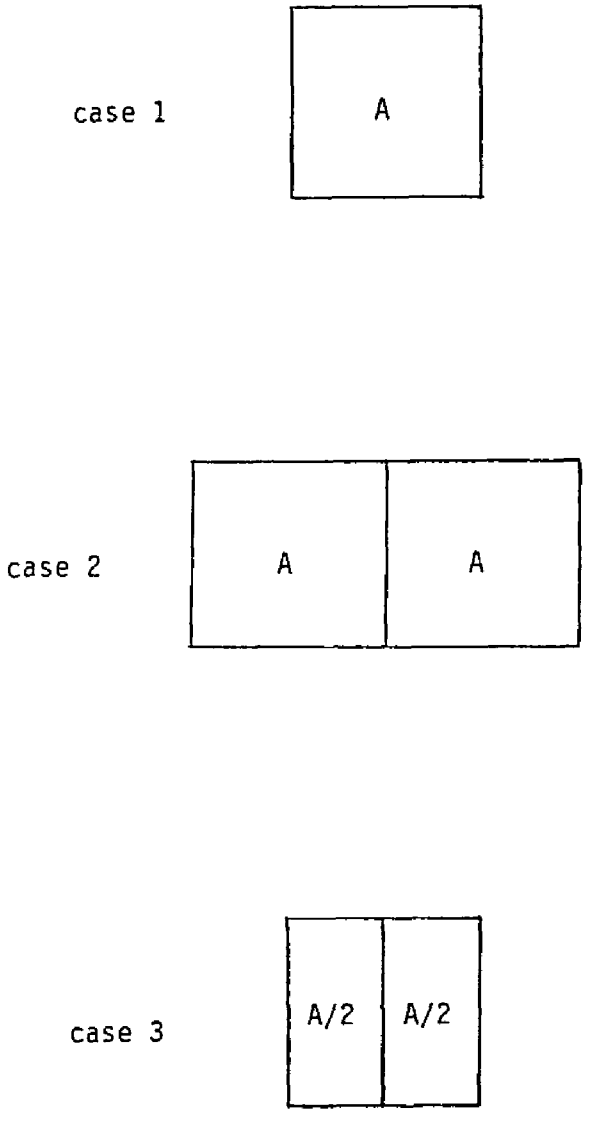


Fig.a.1: A picture element (or pixel) with area A.

(Shannon 1948), then cases 2 or 3 each has information $log(Q^2)=2log(Q)$ which is twice as that of the case 1, just as we intuitively would expect. Using this definition, it is found that the information per pixel for the three cases are identical, but the information per unit area for cases 1 and 2 are each one-half that of the case 3. These observations are also consistent with the previous intuition.

For Q independent equally likely levels, the probability P_i that the i-th gray level occurs is:

$$P_i=\frac{1}{Q}\ ,\quad i=1,2,...,Q\ . \tag{b.1}$$

then according to the above definition, the information, H_i, gained by receiving any gray level in a pixel is:

$$H_i=log(Q)=log(\frac{1}{P_i})=-log(P_i)\ . \tag{b.2}$$

Shannon (1948, Bell Sys. Tech.) generalized the above expression to include unequally likely gray levels and defined the information gained by receiving the i-th gray level as:

$$H_i=-\log_2(P_i) \tag{b.3}$$

i.e. the unit of information is in bite, or binary digits. This definition also suggests that the information can be interpreted as a measure of reduced uncertainty.

Since the i-th gray level occures with the probability P_i, then, on the average, the information gained by receiving the i-th gray level is $-P_i\,log(P_i)$. Consequently, the average information obtained from receiving a large number of gray levels is:

$$H=-\sum_{i=1}^{Q}P_i\,log(P_i)\ , \tag{b.4}$$

it is also called the entropy of the set of probabilities $\{P_i\}$, $i=1,2,...,Q$.

For continuous gray levels, or densities u, Eq(b.4) becomes:

$$H=-\int_0^{\infty}P(u)\log P(u)\,du\ . \tag{b.5}$$

So far we have been concerned with the information gained by receiving only one pixel. If a picture is composed of J independent pixels, then the information gained by receiving the picture is:

$$H_J(u_1,u_2,\ldots,u_J)=-\int\cdots\int P(u_1,u_2,\ldots,u_J)\log P(u_1,u_2,...,u_J)\,du_1,du_2,\ldots,du_J\ . \tag{b.6}$$

The average information gained per pixel from this multi-pixel image is:

$$H_1(u) = \lim_{J \to \infty} \frac{1}{J} H_J . \qquad \text{(b.7)}$$

If the Fourier spectrum of the image is band-limited from $k_x = -\Omega_x$ to Ω_x and $k_y = -\Omega_y$ to Ω_y, then by the sampling theorem (Bracewell 1978) if we sample over the picture area $\tilde{A}$ at intervals $\Delta x = \frac{1}{2\Omega_x}$ and $\Delta y = \frac{1}{2\Omega_y}$, which is bounded by $[-x, x]$ and $[-y, y]$, the $m\,n$ sampled pixel values, where $m = \frac{2x}{\Delta x} = 4x\,\Omega_x$ and $n = \frac{2y}{\Delta y} = 4y\,\Omega_y$, completely specifies the original image on the area $\tilde{A}$.

If the bound-limits Ω_x and Ω_y are getting larger. the picture area can be divided into smaller and smaller pixels. In the limit that $\Omega_x \to \infty$ and $\Omega_y \to \infty$, we have $\Delta x \to 0$ and $\Delta y \to 0$ (or $m \to \infty$ and $n \to \infty$ while maintaining $2x = m\,\Delta x$ and $2y = n\,\Delta y$). Then Eq(b.7) becomes:

$$H_1(u) = \lim_{m,n \to \infty} \frac{1}{(4x\Omega_x)(4y\Omega_y)} H_J \qquad \text{(b.8)}$$

where $J = m\,n = (4x\Omega_x)(4y\Omega_y)$.

The average information gained per unit area of the received picture is:

$$H_A = \lim_{J \to \infty} \frac{1}{\tilde{A}} H_J = \lim_{m,n \to \infty} \frac{1}{4xy} H_J . \qquad \text{(b.9)}$$

So, there is:

$$H_A = (2\Omega_x)(2\Omega_y) H_1 \qquad \text{(b.10)}$$

that is, the information per unit area is dependent upon the number of independent pixels that can be put into the unit area and the information of each pixel. This echoes the previous notion about the case 3 in the figure.

Naturally, if the image is limited by the poor resolution capabilities of an imaging system, then there can be fewer independent pixels in the picture area and consequently the information per unit area carried by the image is less. This is consistent with the intuitions mentioned in chapters II and VI.

3). RANDOM NUMBER GENERATORS

3.1). Uniform Distribution Random Number Generator

a). Mixed Congruential Method

Let T be a large integer and z_0, a and b be three positive integers which are less than T, the sequence values x_i, $x_i = z_i/T$ $(i=1,2,...,n)$, are called a sequence of uniformly distributed pseudorandom values if the integers $\{z_i\}$ are generated recursively, $0 \le z_i < T$, according to the formula (Dahlquist et al 1974)

$$z_{i+1} = mod\,[a \cdot z_i + b\,, T\,] \qquad (c.1)$$

If $T = 2^t$, b is odd and a gives remainder 1 when divided by 4, then the period of the series $\{x_i\}$ is equal to 2^t, where t is a integer.

b). Mixed Multiplicative Congruential Method

It is a mixed congruential method with $b = 0$. If z_0 is odd and a has remainder 3 or 5 when divided by 8, then the period of the sequence $\{x_i\}$ is $2^{(t-2)}$.

(i) example i): If no action is taken on integer overflow in the computer, then let $T = 2^{31} = 2147483648$ and $a = 2^{16} + 3 = 65539$. The subroutine program is given below:

```
c  k is the initial number (or seed) and the previous resulted number
      function random(k)
      k=k*65539
      if(k .lt. 0) k=k+2147483647+1
      random=k*0.4656613e-9
      return
      end
```

(ii) example ii): If integer overflow is prohibited, then let $T = 2^{20} = 1048576$ and $a = 2^{10} + 3 = 1027$, $(a \cdot T = 1.0769e+9 < 2^{31} - 1)$, and choose the initial number $n = 566387$. The subroutine program is shown below:

```
      subroutine random(x)
      data k/1/
      if(k .eq. 0) goto 9
         k=0
         n=566387
9     n=mod(1027*n,1048576)
      x=n*0.9536743e-6
```

return

end

3.2). Gaussian Distribution Random Number Generator

a). The Box-muller's Transformation

This method generats two gaussian distributed random values x_i and x_j from two uniformly distributed random values z_i and z_j, $0 \le z_k \le 1$, at one time.

The relation of $\{x_k\}$ and $\{z_k\}$ is expressed as (Dahlquist et al 1974):

$$x_i = cos(2\pi z_j)\sqrt{-2ln(z_i)} \qquad (c.2)$$

and

$$x_j = sin(2\pi z_j)\sqrt{-2ln(z_i)}$$

Let e be the mean and d the deviation of the gaussian distributed random values, then the two values x_i and x_j with mean 1 and deviation 0 are modefied as:

$$y_i = e + d \cdot x_i \text{ , and } y_j = e + d \cdot x_j \qquad (c.3)$$

where y_i and y_j are gaussian distributed random values with mean e and deviation d.

b). Approximation Method

This method generats one gaussian distributed random value x_i from a sequence of uniformly distributed random values $\{z_i\}$, $i = 1,2,...,n$, in range $0 \le z_i \le 1$.

The formula is given by (Hamming 1962):

$$x_i = \frac{\sum_{j=1}^{n} z_j - \frac{n}{2}}{\sqrt{n/12}} \qquad (c.4)$$

The modefied gaussian distributed random values $\{y_i\}$ with mean e and deviation d is expressed as, in terms of the gaussian distributed random values $\{x_i\}$ with mean 0 and deviation 1,

$$y_i = e + d \cdot x_i \qquad (c.5)$$

(i) example i): A subroutine program written in fortran 77 to obtain the simulation results of gaussian data BIP algorithms shown in chapters VI, VII and VIII using IBM/XT personal computer is shown in the following pages.

3.3). Poisson Distribution Random Number Generator

The method discussed below generats one poisson distributed random number x_i from a sequence of uniformly distributed random values $\{z_i\}$, $i = 1,2,...,k$, in range $0 \le z_i \le 1$.

Let $x_k = z_0 \cdot z_1 \cdot z_2, \ldots, z_k$ be the product of a sequence of $k+1$ uniformly distributed random values. The lowest value of k, which first causes x_k to be less than or equal to $e^{-\lambda}$, will be a random number having the poisson distribution with mean λ (Carnahan et al 1978).

(i) example i): A subroutine program used to obtain the simulation results of poisson data BIP algorithms described in chapters VI, VII and VIII is given in the following pages.

The gaussian and poisson distribution random number generators were tested and the results are shown in the figures (a.2), (a.3), (a.4) and (a.5).

Fig.(a.2) shows the theoretical gaussian (dotted line) and poisson (solid line) distributions with mean value 10.

Fig.(a.3) shows the computer generated results using the gaussian (denoted by crosses) and poisson (indicated by stars) distribution random number generators with mean 10 for 1500 times (or 1500 samples were taken for each distribution).

The theoretical gaussian (dotted line) and poisson (solid line), and computer generated gaussian (crosses) and poisson (stars) distribution with mean 100 are shown in Figs.(a.4) and (a.5) respectively, where 4000 samples were taken to draw each of the distributions of cross and stars.

References (Marsagha 1961) gave some discussion about generating exponential random variable from uniform random variables.

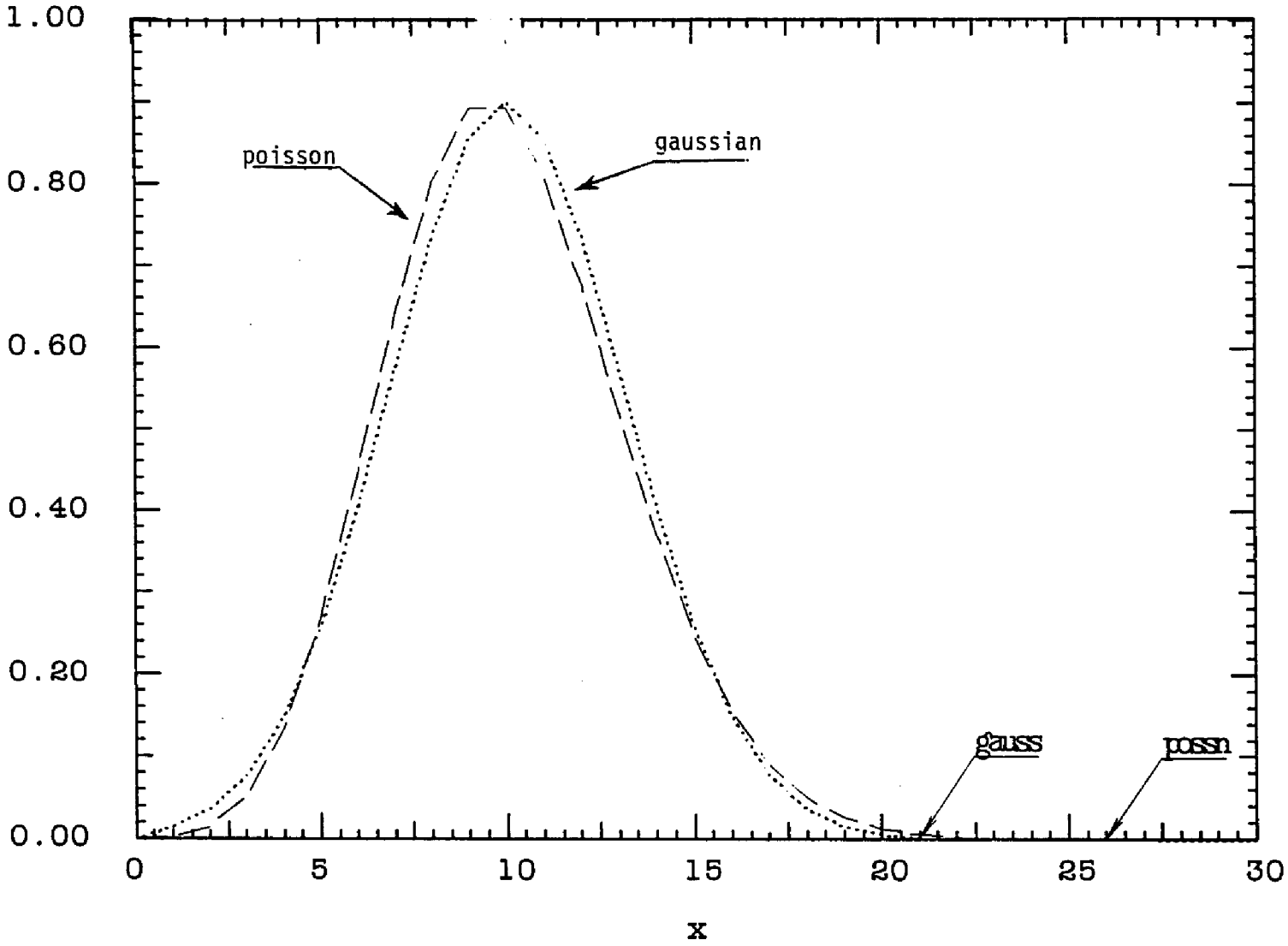


Fig.a.2: Gaussian & Poisson distributions (mean=10)

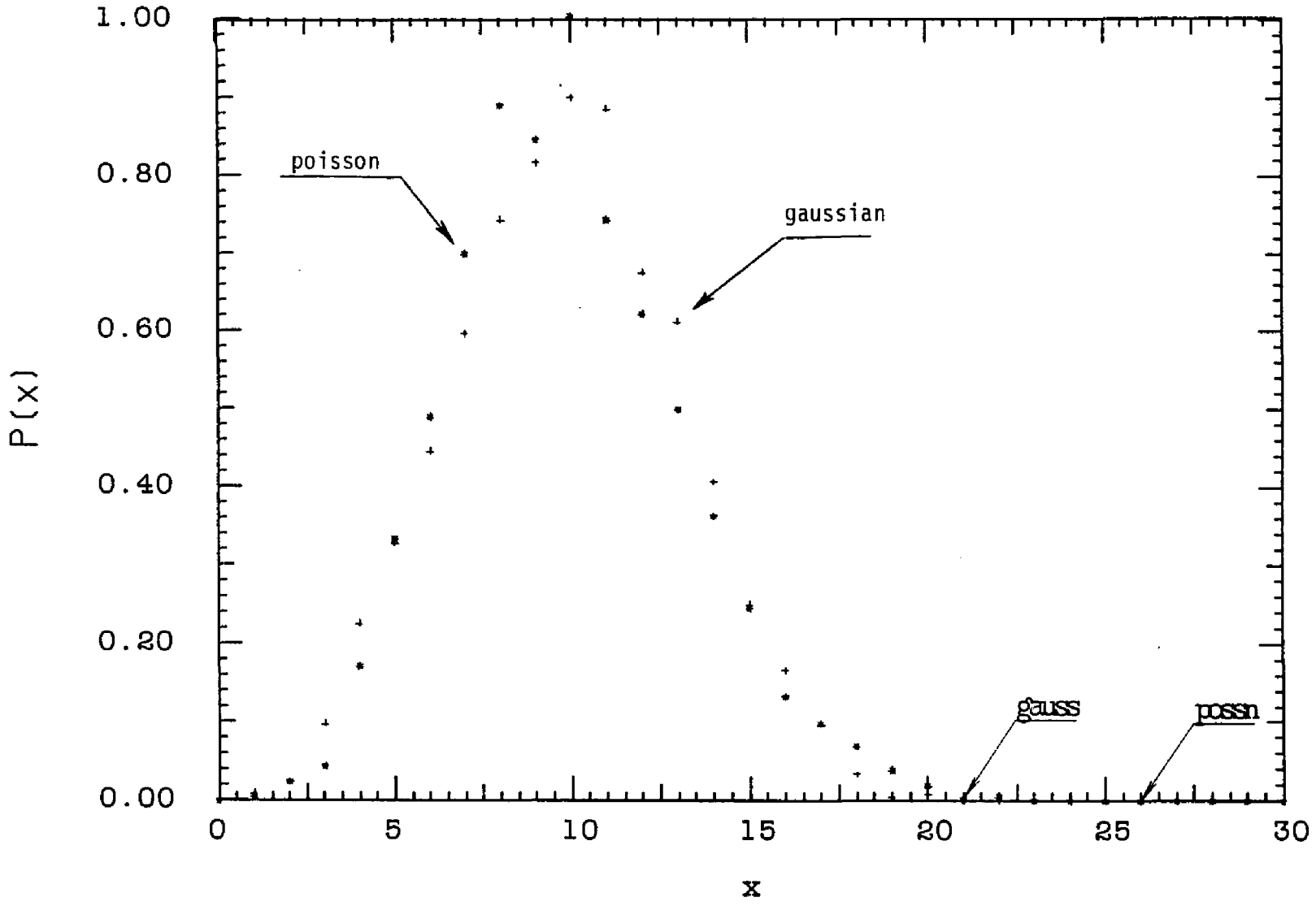


Fig.a.3: Gaussian & Poisson distributions (mean=10,#time=1500)

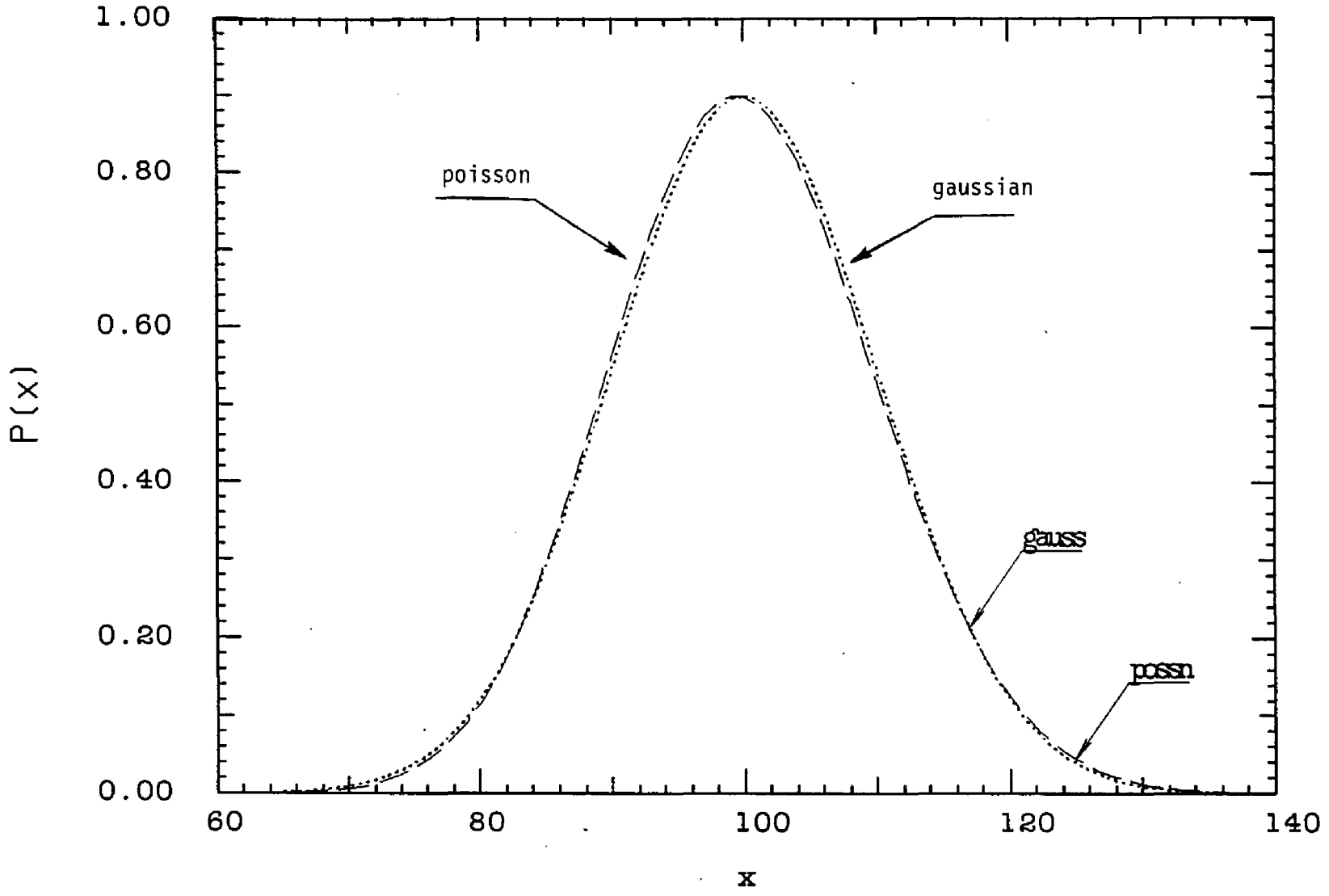


Fig.a.4: Gaussian & Poisson distributions (mean=100)

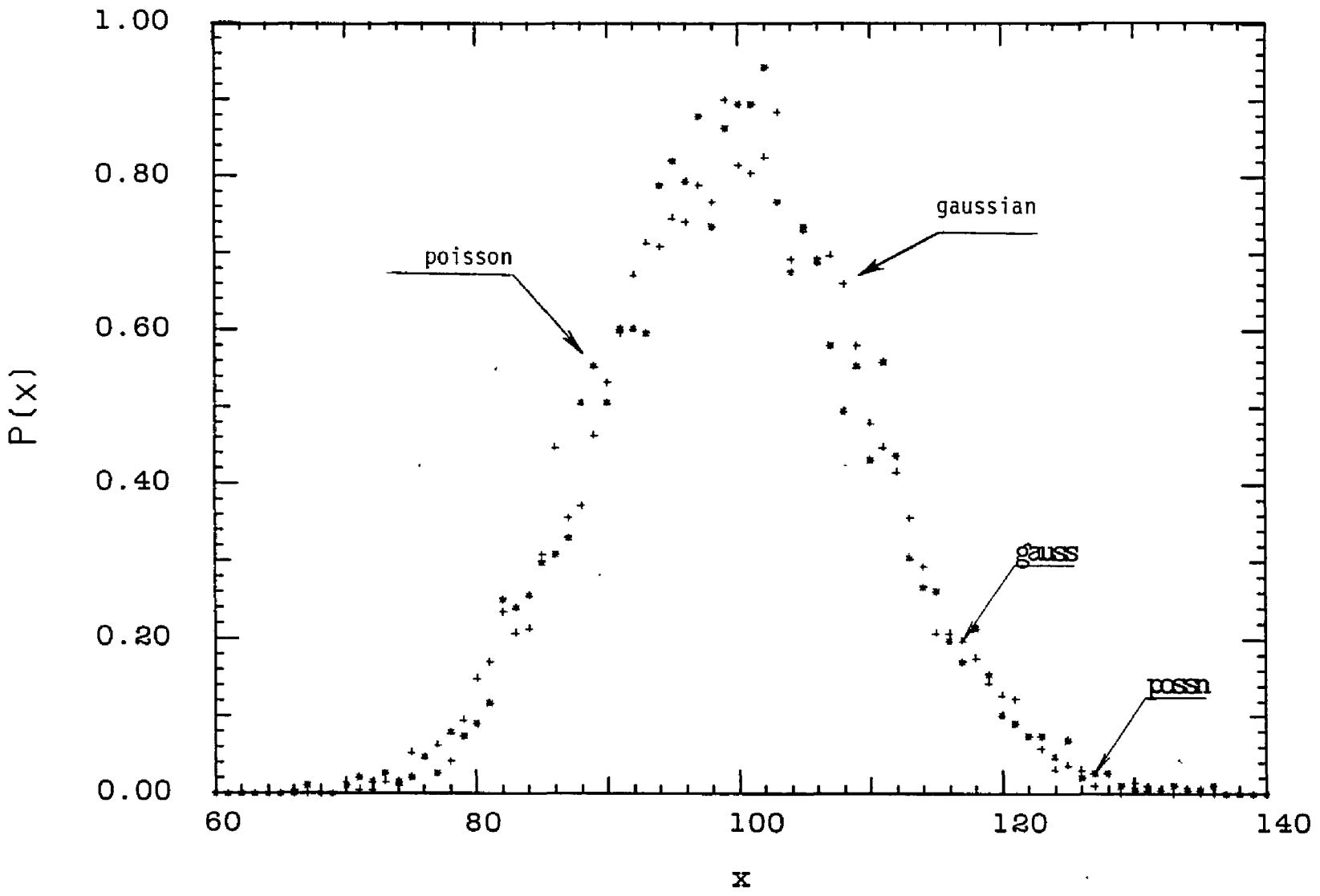


Fig.a.5: Gaussian & Poisson distributions (mean=100,#time=4000)

Poisson Distribution Random Number Generator

```
c    This subroutine is called to generats one sequence of poisson
c    distributed random numbers (k) with mean z. ii is the seed.
         subroutine poissn(z,k,ii)
         double precision x,y,rand
         real z
         integer*2 ii,k
         if(ii .eq. 0) go to 11
            x=rand(ii)
            ii=0
11       y=-dble(z)
         x=0.0d+00
         k=-1
22       if(x.le.y) go to 33
          x=x+dlog(rand(0))
          k=k+1
          go to 22
33       continue
         return
         end
c
c     the function rand generats the uniform distribution random #'s
c
         double precision function rand(i)
         double precision rand
         integer*2 i
         integer*4 j
         if(i .ne. 0) j=566387+2*(i-1)
         j=mod(j*1029+1,1048576)
         rand=dble(j)*0.9536743d-06
         return
         end
```

Gaussian Distribution Random Number Generator

```
c    This subroutine is called to generats one sequence of gaussian
c    distributed random numbers (x) with mean z and deviation sqrt(z).
c    ii is the seed.
         subroutine gauss(z,k,ii)
         double precision x,y,rd
         real z
         integer*2 ii,k,l
         if(ii .eq. 0) go to 11
            x=rd(ii)
            ii=0
11       x=0.0d+00
         do 15 l=1,12
          x=x+rd(0)
15       continue
         x=x-6.0d+00
         y=dble(z)+x*dsqrt(dble(z))
         k=idnint(y)
         return
         end
c
c     the function rd generats the uniform distribution random #'s
c
         double precision function rd(i)
         double precision rd
         integer*2 i
         integer*4 j
         if(i .ne. 0) j=566387+2*(i-1)
         j=mod(j*1029+1,1048576)
         rd=dble(j)*0.9536743d-06
         return
         end
```

4). RECENT CONTRIBUTIONS TO BAYESIAN IMAGE PROCESSING

4.1). Bayesian Analysis for Image Reconstruction

A). Herman et al (1976, Information & Control) presented a Bayesian analysis for image reconstruction, where they used three assumptions:

a.1) the relation of $\mathbf{Y}$ and Φ is linear and the noise is additive:

$$\mathbf{Y} = R\Phi + \mathbf{N}; \tag{d.1}$$

a.2) the source Φ is a random vector with a multivariate Gaussian distribution, whose covariance matrix A is a positive definite symmetric matrix, i.e.,

$$P(\Phi) = (2\pi)^{-J/2} (\det A)^{-1/2} \exp[-\frac{1}{2} (\Phi - \overline{\Phi})^{\tau} A^{-1} (\Phi - \overline{\Phi})] \tag{d.2}$$

where $\overline{\Phi}$ is the expected value of Φ;

a.3) the noise $\mathbf{N}$ is a random vector with a multivariate Gaussian distribution, whose expected value is the zero vector and whose covariance matrix B is a positive definite symmetric matrix, i.e.,

$$P(\mathbf{N}) = (2\pi)^{-I/2} (\det B)^{-1/2} \exp[-\frac{1}{2} (\mathbf{Y} - R\Phi)^{\tau} B^{-1} (\mathbf{Y} - R\Phi)]. \tag{d.3}$$

The iterative algorithm approaching the solution Φ^{*} which maximizes function $\ln P(\Phi | \mathbf{Y}) \equiv \ln P(\mathbf{Y} | \Phi) + \ln P(\Phi) - \ln P(\mathbf{Y})$ is:

$$\Phi^{(0)} = \overline{\Phi}$$
$$\Phi^{(n+1)} = \Phi^{(n)} + \alpha^{(n)} [\overline{\Phi} - \Phi^{(n)} + 2A R^{\tau} B (\mathbf{Y} - R\Phi^{(n)})] \tag{d.4}$$

where $P(\mathbf{Y} | \Phi)$ was assumed as $P(\mathbf{N})$, $\alpha^{(n)} \approx 1$ and the average $\overline{\Phi}$ was chosen either as:

i). vector $\overline{\Phi}$ has all elements of value $\frac{1}{J} \sum_j \phi_j \approx \sum_i Y_i$; or

ii). vector $\overline{\Phi}$ has all elements of value $\frac{1}{J} \sum_j \phi_j^{(n)}$, where n is the iteration index.

The matrices A and B have been assumed as:

$A_{ij} \neq 0$ only if the i-th and j-th voxels coincide with or touch each other and all diagonal elements are same;

The noise in the measurements is uncorrelated with zero mean, i.e., B^{-1} is a diagonal matrix all of whose non-vanishing entries are $(2\sigma)^{-1}$, where σ is the standard deviation of the noise.

They concluded that "Based on such demonstrations (given in this paper), the performance of our algorithms dose not appear to be as good as that of some alternative algorithms; however, it is expected that a choice of parameters reflecting a better knowledge of the statistical properties of the image space will eventually lead to a much improved performance".

B). Herman et al (1977, ed by Terpogossian) analysed the Bayesian concepts for image reconstruction.

In this paper, they emphasized that "It appears that the Bayesian approach, in spite of its great mathematical appeal, must be substantially improved before it can become a serious competitor to other image reconstruction techniques".

C). Herman et al (1979, Information & Control) gave further detailed discussion about Bayesian analysis for image reconstruction, where they mentioned that "The Bayesian approach selects the image Φ^* which has the largest a posteriori probability, where the a posteriori probability distribution depends on assumed a priori probability distributions of the images Φ and the errors N. Previously, we have assumed that both these a priori distributions are multivariate Gaussian. Rochmore and Macovski (1976) stated that 'there seems to be no justification for such a step'. For that reason they advocated maximum likelihood estimation, rather than Bayesian estimation. we shall return to this point below", and they proposed more assumptions:

a.4) there is a picture $\bar{\Phi}$ such that the components of $\Phi - \bar{\Phi}$ are not correlated and each component $\phi_j - \bar{\phi}_j$ is an element of a zero mean Gaussian distribution with a fixed standard deviation ρ;

This assumption was derived in our work as one of several possibilities by applying entropy considerations to the fluctuations of the source field (i.e., Eq.(4.12) of the generic constraint in chapter IV).

a.5) the components of the error N are not correlated and each component N_i is an element of a Gaussian distribution with standard deviation σ.

They said "we have found that choosing $\bar{\Phi}$ as the output of backprojection method (Herman and Rowland 1977) very badly violates the assumption a.4), and the Bayesian estimate so produced is very different from the original picture Φ". To remedy such deficiency, they chose empirical values of ρ and σ for Φ and N respectively in different image situations.

They concluded that "we have shown that, even though the physical assumptions on which the estimate is based are not necessarily satisfied in practice (i.e., violation of assumption a.4)), the procedure can be used to improve the output of standard non-iterative image reconstruction methods (i.e., the $\bar{\Phi}$ is chosen as the output)".

4.2). Bayesian Method for Image Restoration

A). Hunt (1977, IEEE Trans. Computer) formulated a Bayesian method for image restoration, where he used the same three assumptions of Herman et al, i.e., Eqs.(d.1), (d.2) and (d.3).

The iterative algorithms are:

i) Picard's method

$$\text{if} \quad \Phi = F(\Phi), \quad \text{then} \quad \Phi^{(n+1)} = F(\Phi^{(n)}); \tag{d.5}$$

b) Gradient descent method

$$\Phi^{(0)} = \overline{\Phi}$$
$$\Phi^{(n+1)} = \Phi^{(n)} + \alpha^{(n)} \nabla g(\Phi^{(n)}) \tag{d.6}$$

where $g(\Phi) \equiv \ln P(\Phi \mid \mathbf{Y})$ and $\alpha^{(n)} \simeq 1$.

The matrices A and B of Appendix 4.1) are now assumed to be Toeplitz (i.e., $A_{ij} = A_{mn}$ if $i-j = m-n$), since the random processes of Φ and N are assumed as stationary about the ensemble mean.

In this paper, the A and B matrices were chosen as diagonal matrices with non-vanishing entries of $(2\rho^2)$ and $(2\sigma^2)$ respectively and $\overline{\Phi}$ the blurred image (i.e., $\overline{\Phi} \simeq \mathbf{Y}$).

He said "suppose some a priori knowledge about the restored image is known, then the a priori knowledge could be used as the image $\overline{\Phi}$, and the extent of knowledge about the original image would determine how detailed the a priori image became".

B). Trussell et al (1979, IEEE Trans. Computer) used the modified Picard method to maximize the function $g(\Phi)$ instead of Picard's method and the gradient descent method which were used by Hunt (1977).

C). Saleh et al (1983, J. Opt. Soc. Am) applied the Bayesian methods of Hunt (1977) to the restoration of bilinearly distorted images, where

$$\Phi = 3, \quad \text{and} \quad \rho = 1; \quad \overline{N} = 0, \quad \text{and} \quad \sigma = 0.1$$

were chosen and the A and B matrices were assumed as diagonal matrices with non-vanishing entries of $(2\rho^2)$ and $(2\sigma^2)$ respectively.

Papers published before Hunt's and relating to Bayesian concepts for image restoration (Richardson 1972, Nahi et al 1972 and Habbibi 1972) essentially reflect the limitations of Hunt (1977).

New contributions associated with the BIP methods developed in this thesis can be summarized as follows:

a) Herman's assumption a.4) was shown to be a special case of a class of generic constraints arising from entropy considerations.

b) A new class of generic $P(\Phi)$ arising from inter-element continuity considerations (related to high frequency F.T. filtering) has been proposed and applied.

c) The EM technique previously applied by Shepp et al and Lange et al in non-Bayesian formalisms has now been applied to BIP.

d) In carrying out the detailed calculations, a wide variety of relatively novel parametric dependencies were used in obtaining improved images.

e) The techniques referred to in paragraphs a) through d) above were then applied to situations in which a priori quantitative source information might be available. Five classes of such source informations were considered (Eqs.(4.17), (4.20), (4.23), (4.26) and (4.37) in chapter IV).

Depending upon the validity and restrictiveness of the a priori source information, improved image processing was demonstrated.

In many cases, the results obtained using BIP with a priori quantitative $P(\Phi)$'s on computer simulated and experimental phantom imaging data were, to our knowledge, qualitatively superior to all previous methods (see chapters VII, VIII and IX).

5). TABLE OF ACRONYMS AND SELECTED REFERENCES

ART --- Algebraic (image) reconstruction technique (Gordon et al 1970).

BIP --- Bayesian image processing.

CAT --- (X-ray) computer assisted tomography (Ledley 1976).

CT --- Computerized tomography, the same as CAT.

ECT --- Emission computerized tomography, e.g. PET and SPECT below (Phelps 1977).

EM --- Expectation and maximization (technique) (Dempster et al 1977).

FCCS --- Focusing collimator coincidence scanning (Hart et al 1977).

FCSGS --- Focusing collimator single gamma scanning (Hart 1968).

F.T. --- Fourier transform.

FWHM --- Full width half maximum (of a point spread function).

ILST --- Iterative least square technique (Budinger et al 1974).

IMENT --- Iterative maximum entropy technique (pp.45, this Thesis).

IMINT --- Iterative minimum norm technique (pp.46, this Thesis).

I.V. --- Intravenous (adiminstration of a radioisotopic dose).

MART --- Multiplicative algebraic (image) reconstruction technique (Lent 1976).

MENT --- Maximum entropy technique (Minerbo 1979).

ML --- Maximum likelihood (technique) (Shepp et al 1982).

MLEM --- Maximum likelihood expectation and maximization (algorithms) (Lange et al 1984).

MRI --- Magnetic resonance imaging (technique), previously referred to as NMR (Nuclear Magnetic Resonance) (Mansfield et al 1982).

NLM --- Statistical weighted non-local method (Hart 1983).

PET --- Positron emission (computerized) tomography (Phelps et al 1975).

PSF --- Point spread function, or point source response function.

SIRT --- Simultaneous iterative (image) reconstruction technique (Gilbert 1972).

SPECT --- Single photon emission computerized tomography (Budinger 1980).

TCT --- Transmission computerized tomography, e.g. CAT and Ultrasound CT (White et al 1981 and Greenleaf 1983).

BIBLIOGRAPHY

Agmon N., Alhassid Y. and Levine R.D. (1979), "An Algorithm for Finding the Distribution of Maximal Entropy." J. Comput. Phys., vol.30, 250-258.

Andrews H.C. and Hunt B.R. (1977), *Digital Image Restoration.* Prentice-Hall, Inc.

Angel E.S. and Jain A.K. (1978), "Restoration of Images Degraded by Spatially Varying Point Spread Functions by a Conjugate Gradient Method." Appl. Optics, vol.17, 2186-2190.

Artzy E., Elfving T. and Herman G.T. (1979), "Quadratic Optimization for Image Reconstruction, II." Computer Graph. Image Proc., vol.11, 242-261.

Barrett H.H. and Swindell W. (1981), *Radiological Imaging.* Academic Press,Inc.

Beauchamp K.G. and Yuen C.K. (1979), *Digital Methods* for *Signal Analysis.* George Allen and Unwin, LTD.

Bertsekas D.P. (1982), *Constrained Optimization* and *Lagrange Multiplier Methods.* Academic Press, Inc.

Bracewell R.N. (1978), *The Fourier Transiform* and its *Applications.* McGraw-Hill, Inc.

Brooks R.A. and Chiro G.D. (1976), "Principles of Computer Assisted Tomography (CAT) in Radiographic and Radioisotopic Imaging." Phys. Med. Biol., vol.21, 189-732.

Brownell G.L. (1964), *Symposium* on *Medical Radioisotope Scanning.* (I.A.E.A. Vienna), vol.1

Budinger T.F. and Gullberg G.T. (1974), "Three-dimensional Reconstruction in Nuclear Medicine Emission Imaging." IEEE Trans. Nucl. Sci., vol.21, 2-20.

Budinger T.F. (1980), "Physical Attributes of Single-photon Tomography." J. Nucl. Med., vol.21, 579-592.

Cannon T.M. (1974), "Digital Image Deblurring by Non-linear Homomorphic Filtering." Ph.D Thesis, Dept. of Comput. Sci., Univ. of Utah, Salt Lake City, Utah.

Cannon T.M., Trussell H.J. and Hunt B.R. (1978), "Comparison of Image Restoration Methods." Appl. Optics, vol.17, 3384-3390.

Carley A.F. and Joyner R.W. (1979), "The Application of Deconvolution Methods in Electron Spectroscopy - a review." J. Ele. Spect. & Related Phenomena, vol.16, 1-23.

Carnahan B., Luther H.A. and Wilkes J.O. (1978), *Applied Numerical Methods.* John Wiley & Sons, Inc.

Censor Y., et al (1979), "A New Approach to the Emission Computerized Tomographic Problem: Simultaneous Calculation of Attenuation and Activity Coefficients." IEEE Trans. Nucl. Sci., vol.20, 986-991.

Censor Y. (1983), "Finite Series-expansion Reconstruction Methods." Proc. IEEE, vol.71, 409-419.

Chang L.T. (1978), "A Method for Attenuation Correction in Radionuclide Computerized Tomography." IEEE Trans. Nucl. Sci., vol.25, 638-643.

Chang L.T. (1979), "Attenuation Correction and Incomplete Projection in SPECT." IEEE Trans. Nucl. Sci., vol.26, 2780-2789.

Chang L.T. and Herman G.T. (1978), "Filter Selection for the Fan Beam Convolution Algorithm." Tech. Rept. No.MIPG6, Med. Image Processing Group. Dept. of Computer Sci., SUNY at Buffalo, Amherse, N.Y.

Cho Z.H., et al (1974), "Computerized Image Reconstruction Methods with Multiple Photon/X-ray Transmission Scanning." Phys. Med. Biol., vol.19, 511-522.

Cho Z.H., et al (1982), "Fourier Transform Nuclear Magnetic Resonance Tomographic Imaging." Proc. IEEE, vol.70, 1152-1173.

Cormack A.M. (1973), "Reconstruction of Density from Their Projections, with Applications in Radiological Physics." Phys. Med. Biol., vol.18, 195-207.

Coulam, et al (1981), *The Physical Basis* of *Medical Imaging.* Appleton-Century Crofts.

Curry T.S., Dowdey J.E. and Murry R.C. (1984), *Christensen's Introduction* to the *Physics* of *Diagnostic Radiology."* Lea & Febiger.

Dahlquist G. and Bjorck A. (1974), *Numerical Methods.* Prentice-Hall, Inc.

Dempster A.P., Laird N.M. and Rubis D.B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm." JRSS, vol.39, 1-38.

Ekstrom M.P. (1973), "A Numerical Algorithm for Identifying Spread Functions of Shift-invariant Imaging Systems." IEEE Trans. Computers, vol.22, 322-327.

Ekstrom M.P. (1984), *Digital Image Processing Techniques.* Academic Press, Inc.

Ell et al (1982), *Radionuclide Section Scanning.* Grune & Stration.

Feller W. (1968), *An Introduction* to *Probability Theory* and its *Applications.* John Wiley & Sons, Inc.

Franklin J.N. (1970), "Well-posed Stochastic Extensions of Ill-posed Linear Problems." J. Math. Analy. Appl., vol.31, 682-716.

Frieden B.R. (1972), "Restoring with Maximum Likelihood and Maximum Entropy." J. Optical Soc. America, vol.62, 511-518.

Frieden B.R. (1980), "Statistical Models for the Image Restoration problem." Computer Graph. Image Proc., vol.12, 40-59.

Fukunaga K. (1972), *Introduction* to *Statistical Pattern Recognition.* Academic Press, Inc.

Gilbert P. (1972), "Iterative Methods for Three-dimensional Reconstruction of an Object from Projections." J. Theor. Biol., vol.36, 105-117.

Glover G.H and Sharp J.C. (1977), "Reconstruction of Ultrasound Propagation Speed Distributions in Soft Tissue: Time-of-Fight Tomography." IEEE Trans. Son. Ultra., vol.24, 229-234.

Goldman S. (1955), *Information Theoery.* Prentice-Hall, N.Y.

Gordon R., Bender R. and Herman G.T. (1970), "Algebraic Reconstruction Techniques (ART) for Three-dimensional Electron Microscopy and X-ray Photography." J. Theor. Biol., vol.29, 471-481.

Greenleaf J.F., Johnson S.A. and Lent A.H. (1978), "Measurement of Spatial Distribution of Refractive Index in Tissues by Ultrasounic Computer Assisted Tomography." Ultrasound Med. Biol., vol.3, 327-339.

Greenleaf J.F. (1983), "Computerized Tomography with Ultrasound." Proc. IEEE, vol.71, 330-337.

Gullberg G.T. (1979), "The Attenuation Radon Transform: Theory and Application in Medicine and Biology." Ph.D Thesis in Donner Lab. in UCBL.

Gullberg G.T. and Budinger T.F. (1981), "The Use of Filtering Methods to Compensate for Constant Attenuation in SPECT." IEEE Trans. Nucl. Sci., vol.28, 142-157.

Habbibi A. (1972), "Two Dimension Bayesian Estimate of Images." Proc. IEEE, vol.60, 878-883.

Hall E.L. (1979), *Computer Image Processing* and *Recognition.* Academic Press, Inc.

Hamming R.W. (1962), *Numerical Methods* for *Scientists* and *Engineers.* McGraw-Hill, N.Y.

Hamming R.W. (1977), *Digital Filters.* Prentice-Hall, Inc.

Hanson K.M. (1980), "On the Optimality of the Filtered Backprojection Algorithm." J. Comput. Assis. Tomog., vol.4, 361-363.

Hart H. (1983), "Statistical Weighted Non-local Method of Iterative Digital Image Processing." Proc. SPIE, vol.432, 111-118.

Hart H. (1968), "Comparative Resolution of Single Gamma Counting and Coincidence Counting in Focusing Collimator Scanning Systems." Trans. N.Y. Acad. Sci., vol.30, 580-599.

Hart H. and Rudin S. (1977), "Three Dimensional Imaging of Multimillimeter Sized Cold Lesions by Focusing Collimator Coicidence Scanning (FCCS)." IEEE Trans. Biol. Engin., vol.24, 167-177.

Herman G.T. (1979), "On Modifications to the ART." Comput. Biol. Med., vol.9, 271-276.

Herman G.T. (1980), *Image Reconstruction from Projections, the fundamentals of computerized tomography.* Academic Press, Inc.

Herman G.T., et al (1979), "On the Bayesian Approach to Image Reconstruction." Information & Control, vol.42, 60-71.

Herman G.T., Hurwitz H. and Lent A.H. (1977), "A Bayesian Analysis of Image Reconstruction." from "Reconstructions Tomography in Diag. Med. & Nucl. Med." Ed by Terpogossian, Univ. Park Press, Balt. MD., 85-103.

Herman G.T. and Lent A.H. (1976), "Quadratic Optimization for Image Reconstruction, I." Computer Graph. Image Proc., vol.5, 319-332.

Herman G.T. and Lent A.H. (1976), "Iterative Reconstruction Algorithms." Comput. Biol. Med., vol.6, 273-294.

Herman G.T. and Lent A.H. (1976), "A Computer Implementation of a Bayesian Analysis of Image Reconstruction." Information & Control, vol.31, 364-384.

Herman G.T., Lent A.H. and Lutz P.H. (1978), "Relaxation Method for Image Reconstruction." Comm. ACM, vol.21, 152-158.

Herman G.T. and Rowland S.W. (1977), "SNARK77: A Programming System for Image Reconstruction from Projections." Tech. Report No.130, SUNY at Buffalo, Amherst, NY.

Hinshaw W.S. and Lent A.H. (1983), "An Introduction to NMR Imaging: From the Bloch Equation to the Imaging Equation." Proc. IEEE, vol.71, 338-350.

Hoffman E.J., Huang S.C. and Phelps M.E. (1974), "Quantitation in PET: 1. Effect of Object Size." J. Comput. Assis. Tomog., vol.3, 299-308.

Hoffman E.J., et al (1981), "Quantitation in PET: 4. Effect of Accidental Coincidences." J. Comput. Assis. Tomog., vol.5, 391-400.

Hsieh R.C. and Wee W.G. (1976), "On Methods of Three-dimensional Reconstruction from a Set of Radioisotope Scintigrams." IEEE Trans. Systems, Man and Cybernetics, vol.6, 854-861.

Huang S.C., et al (1979), "Quantitation in PET: 2. Effects of Inaccurate Attenuation Correction." J. Comput. Assis. Tomog., vol.3, 804-814.

Huang S.C., et al (1980), "Quantitation in PET: 3. Effect of Sampling." J. Comput. Assis. Tomog., vol.4, 819-826.

Hunt B.R. (1973), "The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer." IEEE Trans. Computers, vol.22, 805-812.

Hunt B.R. (1977), "Bayesian Method in Nonlinear Digital Image Restoration." IEEE Trans. Computers, vol.3, 219-229.

Iinuma T.A. and Nagas T. (1967), "Image Restoration in Radioisotope Imaging System." Phys. Med. Biol., vol.12, 501-509.

Isaacson E. and Heller H.B. (1966), *Analysis* of *Numerical Methods.* John Wiley & Sons, Inc.

Isreal A.B. and Greville T.N.E. (1974), *Generalized Inverses*: *theory* and *applications.* John Wiley & Sons, Inc.

Jansson P.A. (1984), *Deconvolution* with *Applications* in *Spectroscopy.* Academic Press, Inc.

Jaszczak R.J., Coleman R.E. and Lim C.B. (1980), "SPECT: Single Photon Emission Computer Tomography." IEEE Trans. Nucl. Sci., vol.27, 1137-1153.

John F. (1955), *Plane Waves* and *Spherical Means Applied* to *Partial Differential Equations.* Wiley Interscience, N.Y.

Johns H.F. and Cunningham J.R. (1983), *The Physics* of *Radiology.* Charles C Thomas Publisher.

Kennett T.J. and Prestwich W.V. (1979), "On the Deconvolution of Exponential Response Functions." Phys. Med. Biol., vol.24, 1107-1122.

Keyes W.I. (1976), "A Practical Approach to Transverse Section γ-ray Imaging." Br. J. radiol., vol.49, 62-70.

Knoll G.F. (1983), "Single Photon Emission Computed Tomography." Proc. IEEE, vol.71, 320-329.

Kuhl D.E. and Edwards R.Q. (1963), "Image Separation Radioisotope Scanning." Radiology, vol.80, 653-662.

Kuhl D.E. (1983), *Principles* of *Radionuclide Emission Imaging*. Pergamon Press.

Lakshminarayanan A.V. and Lent A.H. (1979), "Methods of Least Squares and SIRT in Reconstruction." J. Theor. Biol., vol.76, 267-295.

Lambert R.F. (1968), *An Introduction* to *Random Signals* and *Communication Theory*. International Textbook Company.

Lange K. and Carson R. (1984), "EM Reconstruction Algorithms for Emission and Transmission Tomography." J. Comput. Assis. Tomog., vol.8, 306-316.

Lathi B.P. (1968), *An Introduction* to *Random Signals* and *Communication Theory*. International textbook Company.

Ledley R.S. (1976), "Introduction to CT." Comput. Biol. Med., vol.6, 239-246.

Lent A.H. (1976), "A Convergent Algorithm for Maximum Entropy Image Restoration, with a Medical X-ray Application." SPSE Conf. Proc., 249-257.

Lewis D.L. Thomas L.J.Jr. and Ter-Pogossian M.M. (1981), "A Mathematical Model for Positron-Emission Tomography Systems Having Time-of-Fight Measurements." IEEE Trans. Nucl. Med., vol.21, 579-592.

Lewitt R.M. (1983), "Reconstruction Algorithms: Transform Methods." Proc. IEEE, vol.71, 390-408.

Llacer J. (1982), "Tomographic Image Reconstruction by Eigenvector Decomposition: Its Limitations and Areas of Applicability." IEEE Trans. Med. Imaging, vol.1, 34-42.

Louis A.K. and Natierer F. (1983), "Mathematical Problems of Computerized Tomography." Proc. IEEE, vol.71, 379-389.

Luenberger D.G. (1973), *Introduction* to *Linear* and *Non-linear Programming*. Addison-Wesley, Inc.

Luenberger D.G. (1984), *Linear* and *Non-linear Programming*. Addison-Wesley, Inc.

Macovski A. (1983), "Physical Problems of Computerized Tomography." Proc. IEEE, vol.71, 373-378.

Mansfield P. and Morris P.G. (1982), *NMR Imaging* in *Biomedicine*. Academic Press, Inc.

Marsaglia B.G. (1961), "Generating Exponential Random Variables." Annals Math. Statis., vol.32, 899-900.

Marsaglia B.G. (1961), "Expressing a Random Variable in terms of Uniform Random Variables." Annals Math. Statis., vol.32, 894-898.

Mazziotta J.C., et al (1981), "Quantitation in PET: 5. Physical-Anatomical Effects." J. Comput. Assis. Tomog., vol.5, 734-743.

Mersereau R.M. (1976), "Direct Fourier Transform Techniques in Three-dimensional Image Reconstruction." Comput. Biol. Med., vol.6, 247-258.

Minerbo G. (1979), "MENT: A Maximum Entropy Algorithm for Reconstructing a Source from Projection Data." Computer Graph. Image Proc., vol.10, 48-68.

Nahi N.E. and Assefi T. (1972), "Bayesian Recursive Image Estimation." IEEEE Trans. Comput., vol.21, 734-738.

Pang S.C. and Genna S. (1979), "The Effect of Compton Scattered Photons on Emission Computerized Transaxial Tomography." IEEE Trans. Nucl. Sci., vol.26, 2272-2274.

Phelps M.E., et al (1975), "Application of Annihilation Coincidence Detection to Transaxial Reconstruction Tomography." J. Nucl. Med., vol.16, 210-223.

Phelps M.E. (1977), "Emission Computed Tomography." Seminars in Nuclear Medicine, vol.7, 337-365.

Phillips D.L. (1962), J. Assoc. Comput., vol.9.

Price G.L. (1982), "Isolation of Instability in the Fredholm Integral Equation of the First Kind: Application to the Deconvolution of noise Spectra." J. Appl. Phys., vol.53, 4571-4578.

Quittner G. (1966), "On Error Propagation in the Resolution Correction." Nucl. Instru. Methods, vol.39, 271-277.

Radon J. (1917), "Uber die Bestimmung von Funktionen durch ihre integral werte langs gewisser mannigfaltigkeiten." Ber. Verb. Saechs. Akad. Wiss., Leipzig, Math. Phys. Kl.69, 262-277.

Rao C.R. (1952), *Advanced Statistical Methods* in *Biometric Research.* John Wiley & Sons, Inc.

Richardson W.H. (1972), "Bayesian Based Iterative Method of Image Restoration." J. Opt. Soc. AM., vol.62, 55-59.

Robbins G.M. and Huang T.S. (1972), "Inverse Filtering for Linear Shift Variant Imaging Systems." Proc. IEEE, vil.60, 862-872.

Rockmore A.J. and Macovski A. (1976), "A Maximum Likelihood Approach to Emission Image Reconstruction from Projections." IEEE Trans. Nucl. Sci., vol.23, 1428.

Rosenfeld A. and Kak A.C. (1982), *Digital Picture Processing.* Academic Press, Inc., vol.1 and 2.

Saleh B.E.A. and Rabbani M. (1983), "Restoration of Bilinearly Distorated Images: II. Bayesian method." J. Opt. Soc. Am., vol.73, 71-75.

Sawchuk A.A. (1974), "Space Variant Image Restoration by Coordinate Transformation." J. Opt. Soc. Am., vol.64, 138-144.

Shannon C.E. (1948), "A Mathematical Theory of Communication." Bell System Tech., Journal 27.

Shaw C.B. (1972), "Improvement of the Resolution of an Instrument by Numerical Solution of an Integral Equation." J. Math. Analy. Appl., vol.37, 83-112.

Shepp L.A. and Logan B.F. (1974), "The Fourier Reconstruction of a Head Section." IEEE Trans. Nucl. Sci., vol.21, 21-43.

Shepp L.A. and Vardi Y. (1982), "Maximum Likelihood Reconstruction for Emission Tomography." IEEE Trans. Med. Imaging, vol.2, 113-122.

Smith P.R., Peters T.M. and Bates T.H.T. (1973), "Image Reconstruction from Finite Numbers of Projections." J. Phys. A: Math. Nucl. Gen., vol.6, 361-382.

Snyder D.L., Thomas L.J. and Ter-pogossian M.M. (1981), "A Mathematical Model for PET Systems Having Time-of-Flight Measurements." IEEE Trans. Nucl. Sci., vol.28, 3575-3583.

Stark H., et al (1985), "An Investigation of CT by Direct Fourier Inversion and Optimum Interpolation." IEEE Trans. Biomed. Engin., vol.28, 496-505.

Tribus M. (1971), *Rational Descriptions, decisions* and *design.* Pergamon, Oxford.

Trussell H.J. and Hunt B.R. (1979), "Improved Methods of Maximum a Posteriori Restoration." IEEE Trans. Comput., vol.27, 57-62.

Vardi Y., Shepp L.A. and Kaufman L. (1985), "A Statistical Model for Positron Emission Tomography." J. American Statist. Association, vol.80, 8-20.

Walkup J.F. and Choens R.C. (1974), "Image Processing in Signal-Dependent Noise." Opt. Engin., vol.13, 258-266.

Walters T.E., et al (1981), "Attenuation Correction in Gamma Emission Computerized Tomography." J. Comput. Assis. Tomog., vol.5, 89-94.

Webb W., et al (1985), "Constrained Deconvolution of SPECT Liver Tomograms by Direct Digital Image Reconstruction." Med. Phys., vol.12, 53-58.

Welkowitz W. and Deutsch S. (1976), *Biomedical Instruments, theory* and *design*. Academic Press, Inc.

Wernecke S.J. and D'addario L.R. (1977), "Maximum Entropy Image Reconstruction." IEEE Trans. Comput., vol.26, 351-364.

White D.R., Speller R.D. and Taylor P.M. (1981), "Evaluating Performance Characteristics in Computerized Tomography." British J. Radiol. vol.54, 221-231.

Wine R.L. (1964), *Statistics* for *Scientists* and *Engineers*. Prentice-Hall, Inc.